

**UNIVERSIDADE FEDERAL DE ALFENAS**

**DENIS PEREIRA**

***PIPELINE PARA A ANÁLISE IN SILICO DE GENES DIFERENCIALMENTE  
EXPRESSOS DURANTE O DESENVOLVIMENTO DE ABELHAS APIS MELLIFERA  
– O CASO DO KNOCK-DOWN DA ESTERASE DO HORMÔNIO JUVENIL***

**Alfenas/MG**

**2021**

DENIS PEREIRA

*PIPELINE PARA A ANÁLISE IN SILICO DE GENES DIFERENCIALMENTE  
EXPRESSOS DURANTE O DESENVOLVIMENTO DE ABELHAS APIS MELLIFERA  
– O CASO DO KNOCK-DOWN DA ESTERASE DO HORMÔNIO JUVENIL*

Dissertação de Pesquisa apresentada como requisito parcial para a obtenção do Título de Mestre no Programa de Pós-Graduação em Biociências Aplicadas à Saúde, Área de Concentração: Neurociências e Comportamento - Universidade Federal de Alfenas.

Orientador: Angel Roberto Barchuk  
Coorientadora: Flávia Cristina de Paula Freitas

Alfenas/MG

2021

Dados Internacionais de Catalogação-na-Publicação (CIP)  
Sistema de Bibliotecas da Universidade Federal de Alfenas  
Biblioteca Central – Campus Sede

Pereira, Denis  
P436p *Pipeline* para a análise *in silico* de genes diferencialmente expressos durante o desenvolvimento de abelhas *apis melifera* – o caso do *knock-down* da esterase do hormônio juvenil. / Denis Pereira. – Alfenas, MG, 2021.  
67 f.: il. –

Orientador: Angel Roberto Barchuk.  
Dissertação (Mestrado em Biociências Aplicadas à Saúde) – Universidade Federal de Alfenas, 2021.  
Bibliografia.

1. Expressão gênica. 2. Hormônios. 3. Desenvolvimento animal.  
4. bioinformática. 5. *Pipeline*. I. Barchuk, Angel Roberto. II. Título.

CDD-570.7

Ficha Catalográfica elaborada por Marlom Cesar da Silva  
Bibliotecário-Documentalista CRB6/2735

**Denis Pereira**

Pipeline para a análise in silico de genes diferencialmente expressos durante o desenvolvimento de abelhas Apis mellifera – O caso do knock-down da esterase do hormônio juvenil

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Biociências Aplicadas à Saúde pela Universidade Federal de Alfenas. Área de concentração: Neurociências e Comportamento

Aprovada em: 11 de novembro de 2021

Prof. Dr. Angel Roberto Barchuk  
Instituição: Universidade Federal de Alfenas  
UNIFAL-MG

Prof. Dr. Paulo Muniz de Ávila  
Instituição: Instituto Federal de Educação, Ciências e tecnologia do Sul de MG.  
IFSULDEMINAS

Profa. Dra. Joseana Vieira  
Instituição: Instituto de Criminalística do estado de São Paulo



Documento assinado eletronicamente por **Angel Roberto Barchuk, Presidente**, em 11/11/2021, às 16:32, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Joseana Vieira, Usuário Externo**, em 11/11/2021, às 16:36, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Paulo Muniz de Ávila, Usuário Externo**, em 11/11/2021, às 16:38, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://sei.unifal-mg.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0616964** e o código CRC **DDC609EA**.

Dedico este trabalho a Deus, à Virgem Maria, aos meus pais (Altamiro Pereira e Maria das Graças santos) aos meus irmãos, minha esposa Lismari e meu filho Denis Pereira Junior, ao meu orientador prof. Angel Roberto Barchuk e coorientadora Flávia Cristina de Paula Freitas pela confiança e paciência e todos que me apoiaram nessa período.

## **AGRADECIMENTOS**

- ❖ Ao meu Deus (YESHUA) que é o Alfa e o Ômega, o principio e o fim de todas as coisas.
- ❖ À Virgem Maria, Santa mãe de Deus, pelo cuidado maternal que sempre me ofertou.
- ❖ Aos meus pais em especial minha mãe, pelo cuidado que sempre teve por nós e por ser uma mulher guerreira e dócil ao mesmo tempo.
- ❖ À esposa que Deus me deu para amar e cuidar e meus três filhos, dois deles no céu, e comigo aqui na terra o júnior que tanto me tem humanizado.
- ❖ Ao meu Orientado e co-Orientadora, pela paciência, e pelo imenso aprendizado obtido nesses dias.
- ❖ Ao grupo de pesquisa BIOLOGIA DA SOCIALIDADE, por me acolher e por compartilhar tanto conhecimento.
- ❖ Ao programa PPGB, por fornecerem condições para a realização desse trabalho.
- ❖ Aos Meus irmãos e amigos e todos que de alguma forma contribuíram para esta conquista.
- ❖ O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

O senhor é meu rochedo, minha fortaleza e meu libertador. Meu Deus é minha rocha onde encontro o meu refúgio, meu escudo, força de minha salvação e minha cidadela

(SALMO 18:3)

## RESUMO

Diante do grande aumento da disponibilidade de dados de sequenciamento de ácidos nucleicos, uma área interdisciplinar chamada bioinformática tem contribuído muito para as pesquisas em análise computacional de sequências de DNA, RNA e proteínas. A bioinformática tornou mais ágil a análise de RNA para identificação da expressão diferencial, possibilitando a criação de *pipelines* cada vez mais eficientes. Em abelhas *Apis mellifera*, a determinação de castas começa com a alimentação diferencial no terceiro *instar* (fase do desenvolvimento) larval. As larvas fêmeas destinadas a se tornarem rainhas são alimentadas com geléia real e as operárias recebem uma proporção menor de geléia real misturada com mel e pólen. Essa dieta ativa diferentes respostas endócrinas nas castas, resultando no estímulo de vias distintas de expressão gênica que culmina no desenvolvimento de rainhas e operárias. Uma das respostas endócrinas moduladas pela alimentação diferencial são os títulos de hormônio juvenil (HJ) que são mais altos nas larvas destinadas a se tornarem rainhas. As diferenças nos títulos de HJ entre rainhas e operárias são reguladas tanto pelo nível de síntese deste hormônio nos *corpora allata*, quanto pela sua degradação mediada pela esterase do hormônio juvenil (JHE). Com o objetivo de determinar os mecanismos moleculares dependentes de jhe direta e indiretamente no desenvolvimento larval, membros de nosso grupo de pesquisa realizaram ensaios funcionais em operárias de *A. mellifera*. Os experimentos consistiram no silenciamento do gene jhe por RNA de interferência no estágio larval L2. Foram coletadas amostras de RNAs de indivíduos tratados (dsRNA-jhe) e controles (*naïve*) no 5º *instar* larval e sequenciadas. Os arquivos contendo as sequências foram disponibilizadas. O objetivo de nosso trabalho foi desenvolver um *pipeline* de análise *in silico* para a identificação de genes diferencialmente expressos entre as abelhas controle e as abelhas tratadas com dsRNA-jhe. Para isto, primeiramente, usamos o programa FASTQC para determinar a qualidade das *reads*, que apresentaram altos valores *phred score*. As sequências de adaptadores identificadas pelo FASTQC e as *reads* com comprimento menor que 10 nt foram removidas usando o programa CUTADAPT. Em seguida, as bibliotecas foram mapeadas no genoma da *A. mellifera* com a ferramenta BWA. As bibliotecas apresentaram um mapeamento em torno 85% das *reads* gerando como saída arquivos no formato SAM. Posteriormente, estes arquivos foram convertidos em



BAM usando o programa SAMTOOLS, os quais foram utilizados pelo programa CUFFLINKS para a identificação de genes diferencialmente expressos (GDEs). Foram encontrados 458 GDEs, sendo destes 151 menos expressos e 307 genes mais expressos nas abelhas tratadas (dsRNA-jhe) em relação às abelhas controle. Os resultados revelaram genes que respondem ao aumento nos títulos endógenos de HJ em resposta ao silenciamento da jhe durante o desenvolvimento larval de *A. mellifera*. Estes resultados se assemelham aos resultados obtidos pela aplicação tópica de HJ em larvas de operárias, previamente publicados por nosso grupo de pesquisa. Além disto, os resultados demonstram a viabilidade do *pipeline* desenvolvido para a análise da expressão gênica diferencial em sistemas biológicos.

Palavras-chave: Expressão gênica; hormônios; desenvolvimento animal; bioinformática; *pipeline*.

## ABSTRACT

Given the great increase in the availability of nucleic acid sequencing data, an interdisciplinary area called bioinformatics has contributed greatly to research in computational analysis of DNA, RNA and protein sequences. Bioinformatics offers the means to deftly identify differentially expressed genes and enables the creation of increasingly efficient pipelines. In the honeybee *Apis mellifera*, caste determination is triggered by differential feeding in the third larval instar. The female larvae destined to become queens are fed royal jelly and the workers receive a lower proportion of royal jelly mixed with honey and pollen. This diet activates different endocrine responses and gene expression pathways in the castes that culminate in the development of queens and workers. One of the endocrine responses regulated by the differential feeding is the levels of juvenile hormone (JH) that are higher in the developing queen. The differences in the JH titers between queen and worker are regulated both by the synthesis of JH by the corpora allata and by its degradation mediated by juvenile hormone esterase (JHE). To determine the molecular mechanisms directly and indirectly dependent on *jhe* in the larval development, members of our research group performed functional assays in *A. mellifera* workers. The experiments consisted of knocking-down the *jhe* gene by interference RNA (iRNA) in the second larval instar. RNA samples were collected from treated (dsRNA-*jhe*) and control (naïve) individuals at the 5th larval instar and sequenced. The aim of our work was to develop an in silico analysis pipeline to identify differentially expressed genes between both control and treated bees. First, we used the FASTQC to determine the quality of the *reads*, which showed high phred score values. Adapter sequences identified by FASTQC and *reads* shorter than 10 nt in length were removed using the CUTADAPT tool. Next, the libraries were mapped on the *A. mellifera* genome with the BWA tool. The mean of mapping percentage of *reads* for all libraries was approximately 85%. The SAM files produced in the mapping step were converted in BAM files by SAMTOOLS, which were used by CUFFLINKS to identify differentially expressed genes (DEGs). A total of 458 GDEs were found, of which 151 were down regulated and 307 up regulated in the treated bees (dsRNA-*jhe*). The results revealed the genes that respond to the increase in endogenous JH titers in response to the knock-down of *jhe* in the larval development of *A. mellifera*. These results are similar to previous data obtained by the topic application of JH in developing worker

larvae previously published by our research group. In addition, the results demonstrate the feasibility of the developed pipeline for the analysis of differential gene expression in biological systems.

Keywords: Gene expression; hormones; animal development; bioinformatics; *pipeline*.

## LISTA DE FIGURAS

Figura 1 -	Modelo geral de diferenciação de castas.....	19
Figura 2 -	Títulos de hormônio juvenil .....	20
Figura 3-	Níveis relativos de transcrição do gene codificador da esterase do hormônio juvenil.....	22
Figura 4-	Fluxograma da abordagem computacional para a identificação de expressão gênica diferencial ( <i>pipeline</i> ).....	26
Figura 5 -	Distribuição dos valores de qualidade ( <i>Phred Score</i> ) ao longo do comprimento das <i>reads</i> . .....	30
Figura 6 -	Linha de comando para execução do programa CUTADAPT.....	31
Figura 7 –	Linhas que compõem a biblioteca “controle 1” .....	32
Figura 8 -	Script para calcular a distribuição do comprimento das <i>reads</i> .....	32
Figura 9 -	Função utilizada no <i>script</i> para salvar o tamanho das <i>reads</i> em arquivo de texto. ....	33
Figura 10 -	Distribuição da quantidade de <i>reads</i> por tamanho de pares de bases .....	34
Figura 11 -	Linha de comando para uso do programa de mapeamento BWA...35	
Figura 12 -	Número de <i>reads</i> mapeadas por bibliotecas .....	35
Figura 13 -	Linha de comando para uso do programa SAMTOOLS, .....	36
Figura 14 -	Linha de comando para uso do programa CUFFDIFF .....	36
Figura 15 -	Gráfico vulcão .....	37
Figura 16 -	Análise de enriquecimento funcional .....	40

## LISTA DE TABELAS

Tabela 1-	Número total de <i>reads</i> nas bibliotecas antes e depois dos cortes com CUTADAPT .....	34
Tabela 2-	Total de genes significativos .....	37

## **LISTA DE ABREVIATURAS E SIGLAS**

HJ	Hormônio Juvenil
JHE	Gene da Esterase do Hormônio Juvenil
KD	knock-dow
GFF	General Feature Format

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>16</b>
1.1	Bioinformática.....	16
1.2	Biologia geral de <i>Apis mellifera</i> .....	17
1.2.1	Diferenciação de castas.....	18
1.2.2	Expressão gênica dependente de HJ no desenvolvimento de castas.....	20
1.2.3	Títulos de hormônio juvenil e sua regulação pela esterase do hormônio juvenil.....	21
1.2.4	Expressão gênica em larvas de operárias <i>knock-down</i> para esterase do hormônio juvenil – motivação e justificativa.....	22
<b>2</b>	<b>OBJETIVOS.....</b>	<b>24</b>
2.1	Objetivo geral.....	24
2.2	Objetivos específicos.....	24
<b>3</b>	<b>MATERIAL E MÉTODOS.....</b>	<b>25</b>
3.1	Análise das bibliotecas de transcriptoma.....	25
3.2	Pré-processamento das bibliotecas de RNA.....	27
3.3	Processamento das bibliotecas de mRNA na identificação dos genes codificadores de proteínas e diferencialmente expressos....	27
<b>4</b>	<b>RESULTADOS E DISCUSSÃO.....</b>	<b>29</b>
4.1	Análise de qualidade das bibliotecas de mRNA.....	29
4.2	Limpeza das bibliotecas de mRNA.....	31
4.3	Bibliotecas após pré-processamento.....	34
4.4	Identificação de genes diferencialmente expressos.....	35
4.5	Análise funcional dos genes diferencialmente expressos. ....	38
<b>5</b>	<b>CONCLUSÃO.....</b>	<b>42</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>43</b>
	<b>APÊNDICE A – Tabela de genes.....</b>	<b>47</b>
	<b>APÊNDICE B – Tutorial para instalação dos programas utilizados</b>	<b>56</b>
	<b>APÊNDICE C – Análise FASTQC.....</b>	<b>64</b>
	<b>APÊNDICE D – Software C#.....</b>	<b>66</b>
	<b>APÊNDICE E – Análise de enriquecimento funcional.....</b>	<b>67</b>

# 1 INTRODUÇÃO

## 1.1 BIOINFORMÁTICA

As milhares de sequências de nucleotídeos que compõem o DNA passaram a ser decifradas mediante experimentos inicialmente bastante laboriosos e desenvolvidos em pequena escala (Sanger *et al.*, 1977). Posteriormente, com a automação computacional, este processo passou a ser realizado em grande escala, como o empreendimento que soletrou a sequência completa do genoma humano (Collins, 2003). Surge, assim, a bioinformática, um campo interdisciplinar que utiliza a aplicação de técnicas de informática para estudos da biologia (como os relativos às moléculas DNA, RNA e proteínas), realizando análise computacional de grande quantidade de dados para uma interpretação biológica (Counsell, 2003). A bioinformática visa aprimorar o processo de análise molecular, oferecendo diferentes ferramentas para análise biológica e diversas estruturas de linhas de comando para a execução de programas que permitem realizar o processamento de dados. Uma sequência ordenada de algoritmos informáticos elaborada com fins específicos para a análise de dados biológicos é comumente chamada de *pipeline* (Roy *et al.*, 2018). Diferentes técnicas de seqüenciamento genômico, com múltiplas estratégias, passaram a ser elaboradas, mas todas fundamentadas no processamento massivo de fragmentos de DNA, entre elas, plataformas como o *MiSeq* e *HiSeq* da *Illumina*; a *Solid* e *Ion Torrent* da *Life Technologies*; *PacBio* da *Pacific Bioscience* entre outras (Glenn, 2011, Quail *et al.*, 2012). A plataforma *Illumina* liga adaptadores aos fragmentos de DNA na região 5' e 3', esses fragmentos se prendem a células que possuem adaptadores correspondentes anexados a sua superfície, estes fragmentos são então amplificados por PCR e purificados em gel, até que se aglomerem em *clusters*, criando muitas cópias de um único modelo de DNA. O seqüenciamento se inicia com a adição de DNA polimerase, primers conectores e 4dNTP à reação. Os dNTP usados são bloqueados em 3'-OH, garantindo que apenas uma base seja adicionada por vez eles também possuem um radical fluorescente incorporado à seqüência de nucleotídeos contendo quatro cores diferentes correspondentes ao nucleotídeos A, T, G, C. O sinal de fluorescência é excitado por laser e o sinal de luz é capturado, finalmente, o sinal óptico é convertido em base de seqüenciamento (*reads*), o terminador e o fluoróforo são então



removidos e outro ciclo de incorporação é iniciado (Bentley, 2006). Esses adaptadores devem ser removidos, pois as suas sequências podem ser incorporadas na montagem dos genes, gerando uma interpretação errada das sequências. *Softwares* de pré-processamento são usados para remover as sequências de adaptadores e aplicar filtragem de qualidade, preparando as bibliotecas para serem utilizadas na pesquisa. Dentre esses *Softwares* podemos citar o CUTADAPT, trimmomatic, trim\_galore, AdapterRemoval e o FASTQC (Andrews, 2013). Cada programa utiliza algoritmos e métodos diferentes para análise de dados, com ferramentas estratégicas para a filtragem de qualidades das *reads*, ou remoção de adaptadores como os programas utilizados para aparar (trimar) as *reads*.

Conjuntos completos de transcritos (RNAs mensageiros, RNAs ribossômicos, RNAs transportadores e os microRNAs) são utilizados em análises transcriptômicas para a identificação dos respectivos genes e seus níveis de transcrição. Esta abordagem possibilita responder às muitas questões biológicas, como, por exemplo, as alterações fenotípicas decorrentes de variações ambientais. Um caso particular deste tipo de problema biológico é o surgimento de fenótipos diferentes, rainhas e operárias, a partir de genótipos equivalentes durante o desenvolvimento de fêmeas das abelhas *Apis mellifera* (Barchuk *et al.*, 2007).

## 1.2 BIOLOGIA GERAL DE *APIS MELLIFERA*

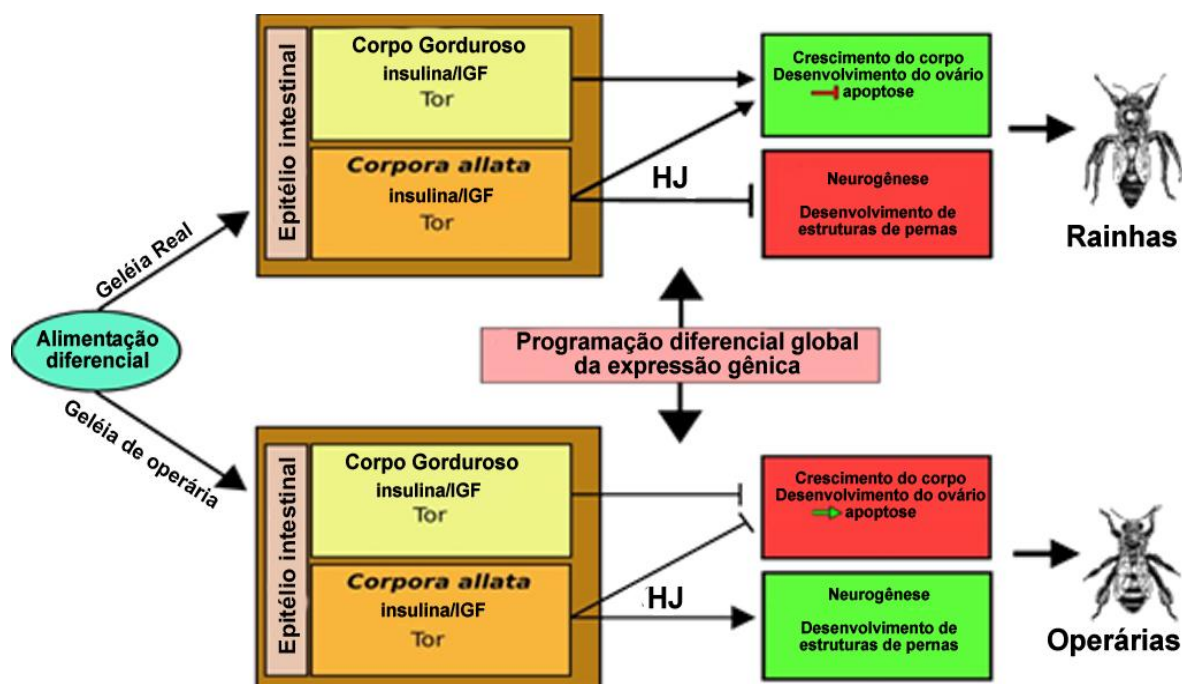
As abelhas da espécie *A. mellifera* pertencem à família *Apidae*, ordem *Hymenoptera*. Estas abelhas vivem em ninhos cujos membros apresentam divisão do trabalho reprodutivo, sobreposição de gerações e cuidado cooperativo da prole (Wilson, 1971). Esta eusocialidade é dada por uma complexidade na sua organização social em que zangões, rainhas e operárias possuem atividades distintas dentro e fora da colméia. Quando jovens, as operárias cuidam das tarefas internas, por exemplo, da alimentação das larvas e limpeza dos alvéolos. Ao envelhecerem, tornam-se forrageiras com atividades como a busca de alimentos e proteção da colméia. As operárias são facultativamente estéreis e possuem um cérebro maior do que as rainhas, um aparelho coletor de pólen e são importantes polinizadores (Lucht-Bertram 1961; Jay 1963; Almeida e Porto 2014). As rainhas são maiores, vivem mais tempo e possuem um aparelho reprodutor bem desenvolvido, o

que lhes permite colocar centenas de ovos por dia. As rainhas quase não saem da colméia, sua presença é importante, mantendo um equilíbrio organizacional. Um dos poucos vôos que realiza é o vôo para o acasalamento (Breed; Gamboa, 1977; Palmer; Oldroyd, 2000; Moritz *et al.*, 2005).

### 1.2.1 Diferenciação de castas

As abelhas *A.mellifera* possuem um ciclo de vida haplo-diploide em que os machos se desenvolvem de ovos não fecundados e os ovos fecundados dão origem às fêmeas. As larvas fêmeas se dividem em duas castas, rainhas e operárias, que recebem uma alimentação diferenciada. Esta alimentação começa no 3<sup>a</sup> instar larval. As larvas que darão origem às rainhas são alimentadas com geléia real e as futuras operárias passam a receber um volume menor de geléia real misturado com mel e pólen (Wirtz, 1973). O alimento ingerido é reconhecido pelo sistema receptor nas células epiteliais do intestino e o sinal de sua ingestão é transmitido ao cérebro mediante a participação do sistema nervoso estomatogástrico. A via de insulina/fator de crescimento de insulina (*IGF*) e o sistema *TOR (TARGET OF RAPAMYCIN)*, um importante sensor nutricional, são ativados de uma forma mais pronunciada em larvas de rainhas, assim, aumentando os níveis de síntese de hormônio juvenil (HJ) nos *corpora allata* (Figura 1) (Barchuk *et al.*, 2007; Patel *et al.*, 2007).

FIGURA 1 - Modelo geral de diferenciação de castas em abelhas *Apis mellifera*.



Fonte: (modificado de Barchuk *et al.*, 2007).

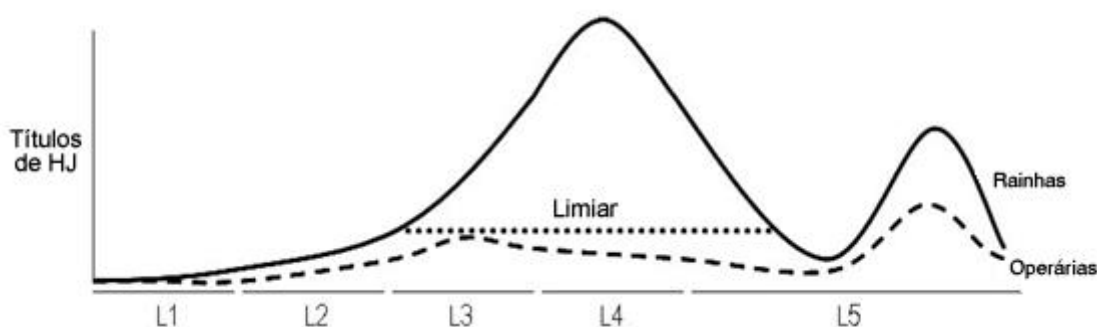
Nota: Rainhas e operárias recebem uma alimentação diferenciada, o alimento ingerido é reconhecido pelo sistema receptor nas células epiteliais do intestino ativando importantes sensores nutricionais de uma forma mais pronunciada em larvas de rainhas, assim aumentando os níveis de síntese de hormônio juvenil (HJ) nos *corpora allata*, regulando a expressão ou repressão de genes com os que atuam no crescimento geral do corpo e desenvolvimento do ovário.

Por meio da diferença nutricional à que são submetidas as larvas, são desencadeadas respostas gênicas distintas, permitindo que um mesmo genótipo possa dar origem a rainhas ou operárias que possuem diferenças marcantes na morfologia, fisiologia e comportamento. As diferenças morfológicas podem incluir alterações no crescimento do corpo ou órgãos específicos e a presença ou ausência de estruturas específicas inteiras. Como exemplo, podemos citar a presença do aparelho coletor de pólen em operárias (Evans; Wheeler, 1999).

O hormônio juvenil (HJ) possui um importante papel no desenvolvimento do dimorfismo de fêmeas em abelhas *A. mellifera*. Sua produção é desencadeada por respostas endócrinas dos *corpora allata* ao diferencial nutricional mencionado antes. O HJ, por sua vez, regula a expressão de genes que atuam no crescimento geral do corpo e desenvolvimento do ovário, regulando negativamente o desenvolvimento de alguns sistemas orgânicos que são características de operárias em *A. mellifera*

tais como o aparelho coletor de pólen. O HJ apresenta-se em níveis diferentes durante o desenvolvimento larval, sendo importante durante a metamorfose (Figura 2). Os títulos de HJ são mais elevados em futuras rainhas, apresentando uma notável diferença entre castas com níveis elevados a partir do terceiro *instar* larval e durante a segunda metade do último *instar*, diminuindo antes da muda da pupa para ambas as castas (Rachinsky *et al.*, 1990; Hartfelder; Engels 1998; Barchuk *et al.*, 2007).

**FIGURA 2 - Títulos de hormônio juvenil durante o desenvolvimento larval (L1-L5) em fêmeas de *Apis mellifera*.**



Fonte: (modificado de Barchuk *et al.*, 2007).

Nota: O eixo y representa os títulos de hormônio Juvenil e o eixo x as fases de desenvolvimento larval de *A. mellifera*.

### 1.2.2 Expressão gênica dependente de hj no desenvolvimento de castas

O HJ e a 20-hidroxi-ecdisona (20E) regulam grande quantidade de atividades biológicas em insetos desempenhando papéis importantes na regulação da expressão gênica. O esteróide 20E é importante para o início de transições de desenvolvimento enquanto o hormônio juvenil previne a metamorfose durante a fase de muda larval. Em abelhas *A. mellifera*, o HJ estimula a síntese de proteínas determinantes da diferenciação de castas durante o estágio larval e, também, controla a evolução do comportamento adulto de operárias, entre outras formas, regulando a neurogênese (Fahrbach; Robinson, 1996). Altos títulos de HJ contribuem para a maturação reprodutiva de fêmeas, promovendo a expressão de genes que previnem a apoptose em ovários durante a fase larval, como *Trap1* e o *TOR*, que é ativado em rainhas (Capella; Hartfelder, 1998; Barchuk *et al.*, 2007; Jindra *et al.*, 2013).

O HJ, semelhante aos ecdisteróides, estimula respostas biológicas por meio de mecanismos moleculares que incluem sua ligação a receptores. Os receptores intracelulares constituem superfamílias de proteínas responsáveis pela regulação dos principais processos de desenvolvimento, e o fazem ligando-se ao *DNA* e alterando a expressão gênica. As proteínas receptoras nucleares são importantes para uma complexa rede reguladora que envolve os genes dependentes de HJ. Um receptor nuclear membro da família de fatores de transcrição *bHLH-PAS*, MET (*Methoprene-tolerant*), localizado nos núcleos da células, possui alta afinidade pelo HJ, e medeia sua atividade (Zhang *et al.*, 2011; Dubrovsky *et al.*, 2011; Jindra *et al.*, 2013).

### **1.2.3 Títulos de hormônio juvenil e sua regulação pela esterase do hormônio juvenil**

Os títulos casta-específicos de HJ podem ser regulados diferencialmente tanto pelo nível de síntese nos *corpora allata* (CA), quanto pela degradação enzimática do HJ mediada pela esterase do hormônio juvenil (*jhe*), presente na hemolinfa e nos tecidos periféricos (De Kort; Granger, 1996; Gilbert *et al.*, 2000; Mackert *et al.*, 2008). Contrastando com o que acontece em abelhas adultas, análises moleculares mostraram que não há diferenças significativas nos níveis de expressão de genes codificadores de enzimas da via de síntese do HJ em larvas de rainhas e operárias. Mas foram encontradas, importantes diferenças nos níveis de expressão do gene que codifica a enzima de degradação de HJ, *jhe* (Bomtorin *et al.*, 2014; Silva, 2012). Estes resultados sugerem que os títulos de HJ durante o desenvolvimento pós-embrionário parecem ser regulados fundamentalmente pelos níveis de degradação e os de abelhas adultas pelos níveis de síntese (Bomtorin *et al.*, 2014).

Análises comparativas de níveis de transcrição do gene *jhe* durante as fases larvais de ambas as castas mostraram alta expressão na quinta fase larval, sendo maior em operárias L5F1 a L5S1 e em rainhas na fase L5S2 (Figura 3) (Silva, 2012). A redução dos níveis de HJ no final do estágio larval (Figura 2) é importante para o processo de muda metamórfica, seguida pelo desenvolvimento *pupal*. O observado declínio dos títulos de HJ está, assim, associado ao aumento dos níveis da JHE, a principal enzima de degradação deste hormônio, sugerindo esta enzima como

reguladora negativa do HJ e apoiando a idéia de que processos de degradação do HJ são fundamentais para a regulação dos títulos deste hormônio durante o estágio larval de abelhas. A jhe é expressa no corpo gorduroso e liberada na hemolinfa e desempenha uma importante atividade hidrolítica com alta especificidade pelo HJ (Hinton; Hammock, 2001).

**FIGURA 3 - Níveis relativos de transcrição do gene codificador da esterase do hormônio juvenil.**



Fonte: (SILVA, 2012).

Nota: Durante o desenvolvimento de fêmeas de abelhas *Apis mellifera*. É possível notar os maiores níveis de transcrição em rainhas durante a primeira parte da quinta fase larval.

#### 1.2.4 Expressão gênica em larvas de operárias *knock-down* para esterase do hormônio juvenil – motivação e justificativa

A participação do HJ como mediador do desenvolvimento de castas foi mostrada em várias espécies de abelhas (Rembold, 1987; Vieira *et al.*, 2006). Porém, estes trabalhos geralmente se baseiam na manipulação hormonal mediante a aplicação exógena de HJ. Doses extras de hormônio podem regular alguns genes alvo dentro de limiares fisiológicos gerando efeitos farmacológicos e, assim, criar um retrato artificial da participação do HJ no desenvolvimento. Com o objetivo de determinar os mecanismos moleculares envolvidos na regulação transcricional do desenvolvimento larval dependente da jhe, membros de nosso grupo de pesquisa realizaram ensaios funcionais em operárias de *A. mellifera* [que, como mencionado

acima (Figura 3), mostram os maiores níveis de expressão *jhe* durante o período larval (Silva, 2012)]. Os experimentos consistiram no silenciamento do gene da *jhe* (*knock-down*, KD) por RNA de interferência (RNAi) durante o estágio larval (L2 – L5). Foram coletadas amostras de RNAs de indivíduos tratados (dsRNA-*jhe*) e controles (sem tratamento, *naïve*) no 5º instar larval e sequenciadas e os arquivos contendo os resultados foram disponibilizados para serem utilizados na identificação de genes codificadores de proteínas e de pequenos RNAs potencialmente envolvidos na regulação da expressão gênica dependente da *jhe* (Silva, 2012). Esta abordagem permitiria obter uma lista com dois grupos de genes: um grupo de genes expressos na presença de *jhe* (em abelhas controle), ou seja, num ambiente com títulos relativamente baixos de HJ, e um grupo de genes expressos na presença de níveis diminuídos de *jhe* (abelhas *knock-down* para *jhe*), ou seja, num ambiente com títulos artificialmente aumentados de HJ endógeno. Este segundo grupo retrataria o perfil de genes que respondem a HJ e são os responsáveis pelo desenvolvimento de características morfofisiológicas que definem o fenótipo de rainhas. Nossa motivação, e a abordagem de nosso trabalho, pressupõe que os níveis diminuídos de mRNA para *jhe* tenham sido, conseqüentemente, responsáveis por diminuições na atividade da respectiva enzima e aumentos equivalentes nos títulos de HJ. Por outro lado, este sistema (KD para *jhe*) constitui um modelo para a elaboração de um *pipeline* para análise da expressão diferencial de genes durante o desenvolvimento larval de abelhas *A. mellifera*.

## 2 OBJETIVO

### 2.1 OBEJTIVO GERAL

O universo computacional é dinâmico e a cada instante apresenta novas abordagens programáticas, reinventando-se constantemente. O mesmo ocorre com as abordagens computacionais de problemas biológicos. Aqui, objetivamos desenvolver um *pipeline* que permita a identificação e quantificação de genes diferencialmente representados em transcriptomas de abelhas *knock-down* para *jhe* (dsRNA-*jhe*) e abelhas controle obtidos mediante trabalhos laboratoriais realizados por Silva (2012). O *pipeline* desenvolvido poderá, posteriormente, ser executado por pesquisadores a fim de identificar genes diferencialmente expressos.

### 2.2 OBEJTIVOS ESPECÍFICOS

Para o desenvolvimento do *pipeline*, objetivamos o ajuste e a utilização de programas de livre acesso que permitam realizar as seguintes atividades:

- a) Analisar a qualidade de bibliotecas de sequenciamento de mRNAs, remover e descartar sequências de baixa qualidade e remover adaptadores;
- b) Analisar a distribuição do comprimento das *reads*;
- c) Mapear as *reads* no genoma de referência;
- d) Identificar a expressão diferencial de genes codificadores de proteínas;
- e) Realizar a análise funcional dos genes identificados;
- f) Elaborar e disponibilizar um tutorial para a análise da expressão gênica diferencial baseado no *pipeline* desenvolvido.



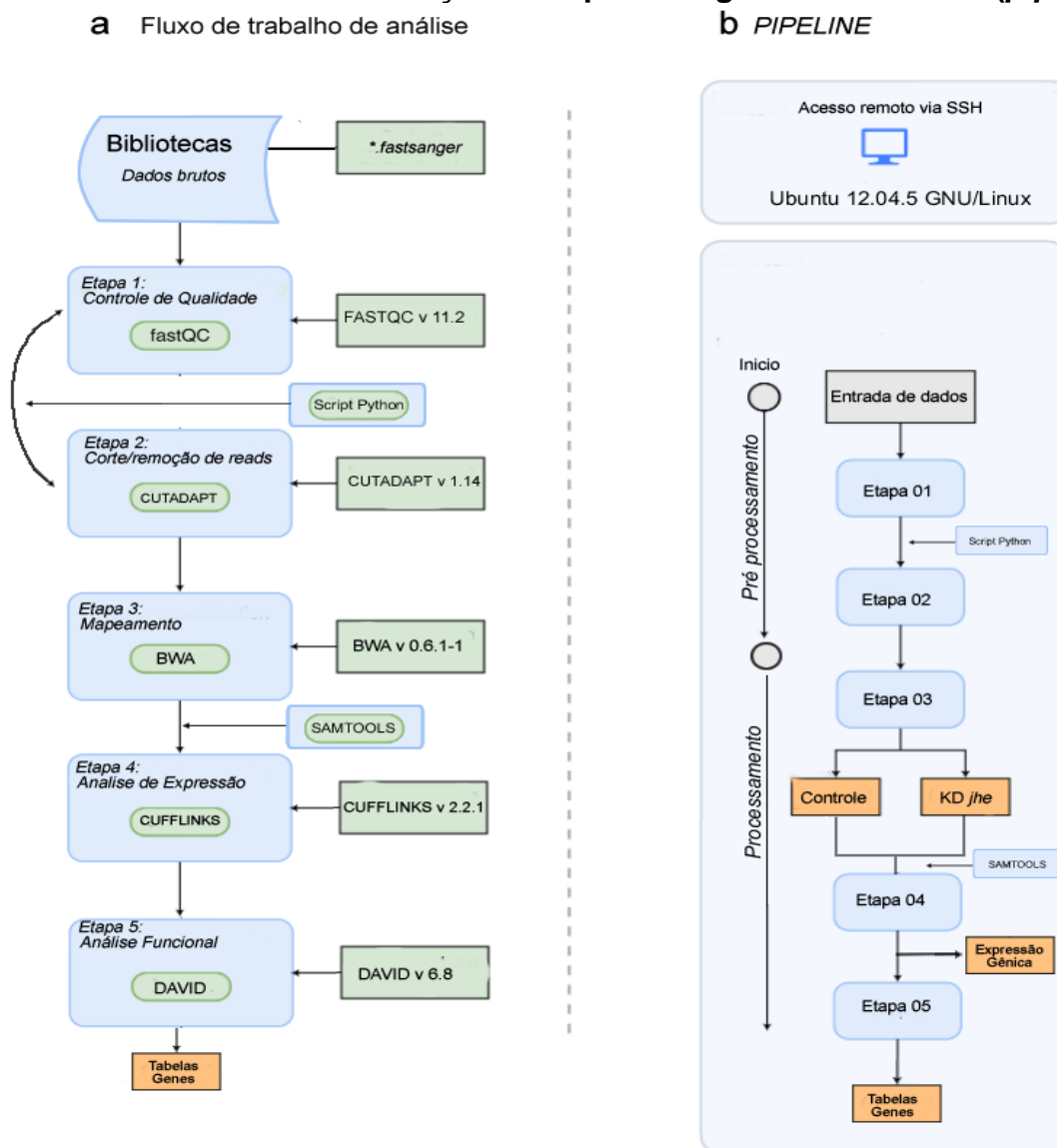
### 3 MATERIAL E MÉTODOS

Para a análise computacional deste trabalho foram utilizados seis arquivos de RNA-seq de leituras de extremidade única com tamanhos de 50 nt obtidos do sequenciamento *Illumina*. Esses dados correspondem às amostras de RNA coletadas a partir do experimento de silenciamento do gene da esterase do hormônio juvenil de *A. mellifera* (*KD* de *jhe*) e submetidas ao sequenciamento de larga escala. Os arquivos correspondentes estiveram disponíveis em um servidor que era possível acessar remotamente via *SSH*, um protocolo de administração remota Linux. Toda a análise dos dados brutos foi realizada no servidor, via terminal Linux sendo possível acessar o diretório de arquivos e realizar a execução de *softwares* de bioinformática, devidamente instalados. O servidor utilizado, com sistema operacional ubuntu 12.04.5 GNU/Linux, conta com a capacidade de Disco Rígido 13.47TB, memória RAM de 188GiB, processador Intel(R) Xeon(R), CPU E5-2620 2.00GHz, itens importantes para uma boa execução de programas de bioinformática e também do *pipeline* desenvolvido.

#### 3.1 ANÁLISE DAS BIBLIOTECAS DE TRANSCRITOMA

Foram analisadas bibliotecas de mRNAs preparadas a partir de RNA total extraídas da fase larval L5 de *A. mellifera*. Foram obtidos do experimento de Silva (2012) seis arquivos do sequenciamento, em formato *fastqsanger*, sendo eles três com dados relativos ao grupo de abelhas controle e três ao grupo de abelhas tratadas (*dsRNA-jhe*). Esses dados brutos foram analisados por *software* de análise de qualidade e remoção de *reads* (Figura 4). Posteriormente, foi realizado o mapeamento de sequências contra o genoma de referência e a análise da expressão diferencial, e, por fim, a identificação dos genes encontrados como diferencialmente expressos entre as duas amostras e suas respectivas características funcionais (Figura 4).

FIGURA 4- Fluxograma da abordagem computacional para a identificação de expressão gênica diferencial (*pipeline*).



Fonte: Autor.

Nota: Em A apresentamos o fluxo de trabalho, que inclui, primeiramente, a obtenção dos dados brutos do sequenciamento, posteriormente a análise computacional das bibliotecas de mRNAs com o programa FASTQC (Andrews, 2013) e a remoção dos adaptadores através do programa CUTADAPT. Em seguida, foi feito o cálculo da distribuição dos comprimentos das *reads* utilizando *scripts* em linguagem *PYTHON* (Van Rossum; Drake, 2011), e a remoção das *reads* menores que 10 nt, repetindo posteriormente a análise com o FASTQC (Andrews, 2013). As próximas etapas foram constituídas pelo mapeamento contra o genoma de referência utilizando o programa BWA, a identificação dos genes diferencialmente expressos através do *software* CUFFLINKS (Trapnell, 2010), os dados foram filtrados em tabelas usando o Excel (Microsoft Office) e a sua análise funcional através da plataforma DAVID (v 6.8), conforme proposto por Huang *et al.*, (2009). Durante todo esse processo o arquivo precisou passar por conversões de compactação e descompactação utilizando o *software* SAMTOOLS. Em B apresentamos a execução deste *pipeline* conforme as etapas descritas.

### 3.2 PRÉ-PROCESSAMENTO DAS BIBLIOTECAS DE RNA

Na primeira etapa da análise computacional das bibliotecas de mRNAs, foram verificados os níveis de qualidade do sequenciamento observando o valor de *score phred* de cada *read* com o programa FASTQC versão 11.2 (Andrews, 2013). Este valor mostra a probabilidade de confiança em cada uma das bases obtidas no sequenciamento. O FASTQC também encontra dentro das *reads* as sequências de adaptadores utilizados no preparo das bibliotecas. Estes adaptadores foram posteriormente removidos utilizando o programa CUTADAPT versão 1.14 (<http://code.google.com/p/cutadapt/>), que também foi utilizado para um refinamento das bibliotecas removendo as pequenas *reads*, menores que 10 nt, que foram identificadas quanto ao seu tamanho utilizando um *script* em PYTHON (Figuras 7 e 8).

### 3.3 PROCESSAMENTO DAS BIBLIOTECAS DE MRNA NA IDENTIFICAÇÃO DOS GENES CODIFICADORES DE PROTEÍNAS E DIFERENCIALMENTE EXPRESSOS

Com as bibliotecas devidamente preparadas, tendo sido removidas as pequenas *reads* e adaptadores, as *reads* foram mapeadas no genoma de *A. mellifera* (versão 4.5, ELSIK *et al.*, 2014) com o programa BWA versão 0.6.1-1 (Li; Durbin, 2009) a fim de obter os arquivos de alinhamentos. A etapa seguinte contou com o cruzamento dos arquivos de alinhamento com os arquivos de anotação do genoma de *A. mellifera* no formato GFF, utilizando a ferramenta CUFFLINKS versão 2.2.1 (Trapnell, 2010). Para a identificação dos genes codificadores de proteínas, os arquivos *GFF* foram recuperados do HymenopteraBase (Elsik *et al.*, 2016).

A análise da expressão diferencial foi realizada utilizando o programa CUFFLINKS. Nesta etapa, foram montados os transcritos e calculado o nível de expressão de cada transcrito nas amostras por meio de algoritmos do próprio programa CUFFLINKS. Posteriormente os dados foram tratados e analisados extraindo os genes diferencialmente expressos com valor de  $p < 0,05$ .

Os genes diferencialmente expressos foram subdivididos em dois subgrupos, um com os mais expressos nas amostras dsRNA-*jhe* e outro com os mais expressos nas amostras controle, apresentados em duas tabelas de anotação gênica (Tabelas

S1 e S2 no apêndice A). As descrições apresentadas nelas foram obtidas mediante consulta ao site NCBI (<https://www.ncbi.nlm.nih.gov/>), via programação, a imagem e link para *download* do executável se encontra no apêndice D - Figura S2.

Utilizando a plataforma DAVID (*Database for Annotation, Visualization and Integrated Discovery*) versão 6.8 (Huang *et al.*, 2009), os genes diferencialmente expressos foram submetidos para análise de Ontologia gênica (em inglês: *Gene Ontology*, GO). Foi criada uma lista com o identificador dos genes diferencialmente expressos, cada identificador começando por “GB” foi inserida no *listbox* do site e a opção de identificador BEEBASE\_ID (modelo de gene de *A. mellifera*), foi selecionada como parte da configuração para início do processo de análise. Os tipos de saídas analisadas foram o agrupamento de anotação funcional, o gráfico de anotação funcional e a tabela de anotação funcional. Não foi usado o limite de contagens de genes mínimos pertencentes a um termo de anotação, sendo aceitos todos os genes que tivessem ao menos um termo GO associado, o que gerou diferentes valores para  $p$  e *Fold Enrichment*, obtendo assim, uma extensa lista com maior possibilidade de análise. Os gráficos foram criados utilizando as categorias de processos biológicos, componente celular e função molecular.

Para uma maior confiabilidade dos dados, a análise foi realizada com todos os genes diferencialmente expressos, 458 genes, e posteriormente para cada grupo, controle com 151 genes e tratadas com 307 genes, obtendo assim uma visão mais ampla e confiável das informações obtidas.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 ANÁLISE DE QUALIDADE DAS BIBLIOTECAS DE MRNA

As bibliotecas foram avaliadas com base no comprimento das sequências e a qualidade de cada uma delas com o programa FASTQC (Andrews, 2013), que apresenta como vantagem uma organização das informações geradas em um mesmo arquivo. Desta forma, a análise torna-se mais eficiente e seus valores de qualidade codificados mais confiáveis uma vez que os arquivos gerados são menores e mais fáceis de trabalhar.

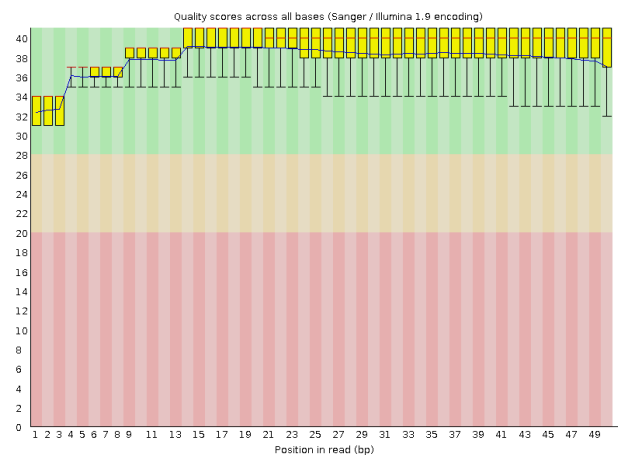
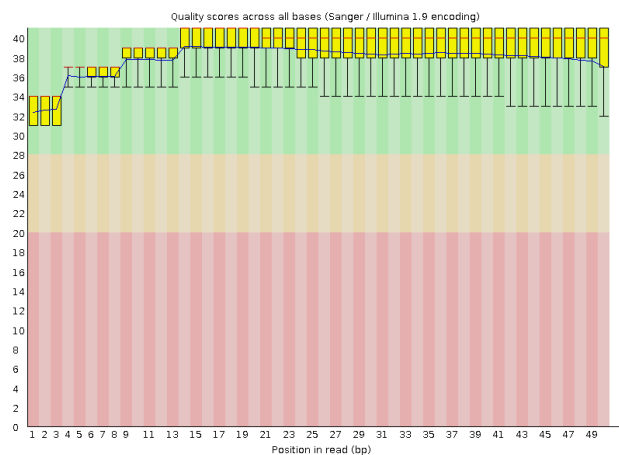
Observou-se que as seis bibliotecas apresentaram altos níveis de qualidade segundo análise FASTQC (Andrews, 2013), com média de *Phred Score* acima de 30 para todas as suas bases de cada uma das *reads* que compõem as bibliotecas.

Todos os fragmento de sequenciamento (*read*) apresentaram comprimento de 50nt. A figura 5 e apêndice C figura S1 mostram os índices de qualidade em todas as bases antes e depois da remoção dos adaptadores *Illumina* e corte das *reads* menores que 10 nt. Foi dividido em duas colunas ao qual observou-se o grupo controle e tratado antes e depois de receberem o devido tratamento para prosseguir com a análise.

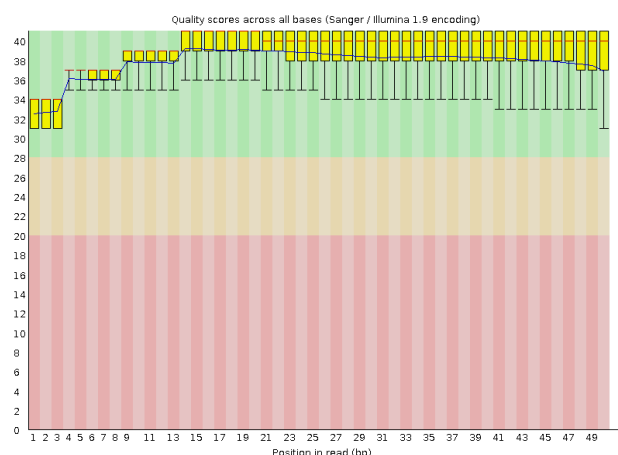
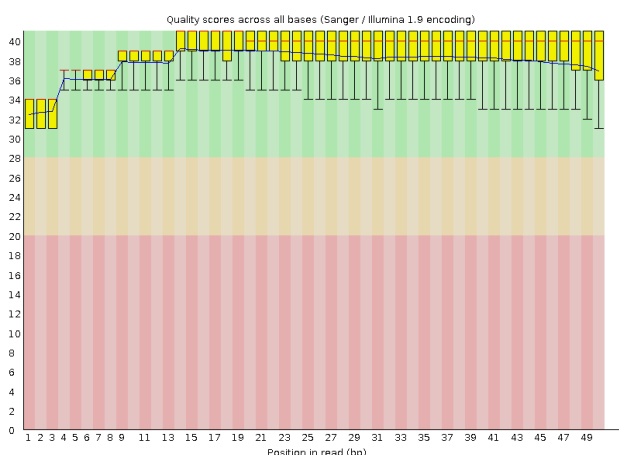
As bibliotecas sofreram poucas alterações, com uma média percentual de 2% a menos no total de *reads* para cada biblioteca após o corte, como é possível observar na Tabela 1 que contém a descrição de cada biblioteca nas linhas e as colunas se referem aos estágios da análise, sendo a primeira os dados brutos, em seguida o número de *reads* após a remoção de pequenas *reads* e adaptadores, e a quarta e quinta coluna desta tabela se refere ao descarte em numeral e percentil.

FIGURA 5 – Distribuição dos valores de qualidade (*Phred Score*) ao longo do comprimento das *reads*.

### A. Controle.



### B. Tratadas.



Fonte: Autor.

Nota: Apresentamos aqui um exemplo de cada biblioteca. Como os resultados foram semelhantes, os gráficos relativos a cada uma das bibliotecas podem ser encontrados no apêndice C(Figura S1). O eixo X representa os pares de base de todas as *reads* que foram analisados e o eixo Y representa os índices de qualidade *Phred* das *reads* analisados pelo programa FASTQC, os *boxplots* em amarelo apresentam o intervalo inter-quartil (25% a 75%) da distribuição dos valores de qualidade e a linha vermelha aponta para o valor da mediana e a linha azul para a média. Em todas as bibliotecas, a distribuição dos valores de qualidade ao longo das *reads* revelou alta qualidade do sequenciamento na identificação de cada base. Foram observadas poucas diferenças entre os arquivos antes e depois do corte e remoção dos adaptadores.

## 4.2 LIMPEZA DAS BIBLIOTECAS DE MRNA

Para garantir a qualidade das *reads* antes do mapeamento, identificamos os adaptadores com a ferramenta FASTQC (Andrews, 2013). Em seguida, os adaptadores foram removidos das *reads* com a ferramenta CUTADAPT, de acordo com a seguinte linha de comando figura 6.

**FIGURA 6 – Linha de comando para execução do programa CUTADAPT.**

```
cutadapt -a “sequência do adaptador” -o C-01-1-cuta.fastq C01-1.fastq
```

Fonte: Autor.

Nota: A sequência do adaptador é fornecida com o comando -a. E entre aspas está o adaptador a ser removido. Com o comando -o é fornecido o nome do arquivo de saída, C-01-1-cuta.fastq, e em seguida o arquivo de entrada C01-1.fastq.

As bibliotecas são um arquivo de texto composto por quatro linhas, em que a primeira e terceira linha informam a identidade da *read*, a segunda linha informa a sequência de bases obtidas no sequenciamento, e quarta linha apresenta os índices de qualidade, (Figura 7). Desta forma, para calcular a distribuição do comprimento das *reads*, foi desenvolvido um *script* em PYTHON que recebe dois parâmetros de entrada, o nome do arquivo e os primeiros caracteres da linha da identidade da *read*, que atribui valores para as variáveis *fastq* e *head*, respectivamente. O algoritmo envia esses dados para um laço de repetição e uma função de condição verifica se a variável *head* foi encontrada na linha, se verdadeiro, é habilitada outra função de condição que analisa a próxima *read* quanto a um dos caracteres A, T, C ou G e após esse processo a função é bloqueada e continua o laço de repetição sempre com esta mesma lógica até acabarem as linhas da biblioteca. Com este *script* (Figura 8) foi quantificado o tamanho das *reads* e observada a presença de pequenas *reads* nas bibliotecas que posteriormente foram removidas utilizando o programa CUTADAPT. A função “*contar*” (Figura 9) chamada neste *script* ordena os dados e gera um arquivo de texto com o nome da biblioteca lida.

FIGURA 7 – Linhas que compõem a biblioteca “Controle 1”.

```

③ @C01-1:274:D1F9AACXX:5:1101:1169:1952 1:
1 ① @C01-1:274:D1F9AACXX:5:1101:1169:1952 1:N:0:ATCACG
2 CTTANATAAAATTTTAAATCTTGAATCTTGTATACTCTCGTACATATAAT
3 +
4 CCCF#2ADHHHHHJJJJIIJJJJJJJEHGGIIIGJJI;FGGGJGIGH
5 ① @C01-1:274:D1F9AACXX:5:1101:2461:1960 1:N:0:ATCACG
6 CAGGNACAAGACATGACATACTCCGTTCAAGGTGTTGGAGTAGTCAATGC
7 +
8 @@<D#22ADH>FB4A<FHGHGDCG?GIHHGID*9?DGGIE?DF?FGIIHG

```

Fonte: Autor.

Nota: Os caracteres marcados com um círculo vermelho são utilizados para o *script* de quantificação das *reads*.

FIGURA 8 – *Script* para calcular a distribuição do comprimento das *reads*.

```

import sys
fastq = sys.argv[1]
head = sys.argv[2]
result = []
resultB = []
header=1
reads = []
with open(fastq, 'r') as f:
    reads = []
    for line in f:
        if header==2:
            if line.startswith("A") or line.startswith("T") or line.startswith("C") or line.startswith("G"):
                result.append(str(len(line)-1))
                header=1
            if line.startswith(head):
                header=2
                reads.append(line)
    contar(result, 'contarA','I',reads)

```

Fonte: Autor.

Nota: Realizado em linguagem de programação *PYTHON*, possui dois parâmetros de entradas definidos nas variáveis *fastq* e *head*, que são o nome do arquivo e os primeiros caracteres da identidade das *reads* respectivamente. Dentro de um laço de repetição verifica a presença da variável *header* na leitura da linha, se verdadeiro a variável *header* habilita a verificação da presença de algum caractere correspondente a um nucleotídeo (A, T, C, G). Se verdadeiro, o tamanho da *reads* é, então, adicionado em um vetor.



**FIGURA 9 – Função utilizada no *script* para salvar o tamanho das *reads* em arquivo de texto.**

```
def contar(result, nome, numero,reads):
    qtd=[]
    cont=1
    #set() remove duplicados
    #sorted classifica
    aux=sorted(set(result))
    aux=sorted(aux)
    x = 0
    for item in aux:
        qtd.append(item)
        qtd.append(result.count( item )) #quantos "item" tem dentro de result[]
    arq1 = open('auxiliar'+numero+'.txt', 'w')
    arq1.write(str(aux))
    arq1.close()
    #salvo apenas a lista de auxiliar
    # salva o resultado pulando linha
    arq = open(nome+head+'.txt', 'w')
    for i in range(len(qtd)):
        arq.write(str(qtd[i])+'-')
        if cont==2:
            cont=0
            arq.write('\n')
        cont+=1
    arq.write(str(len(reads)))
    arq.close
```

Fonte: Autor.

Nota: Esta função é chamada no final do *script* da Figura 8.

As distribuições do comprimento de *reads* que compõem as bibliotecas de mRNAs, é mostrado na Figura 10, onde o cálculo realizado pelo algoritmo mostrou que as *reads* de tamanho igual a 50 nt são a grande maioria (em média 99%), ao final destes processos iniciou-se a etapa de mapeamento.

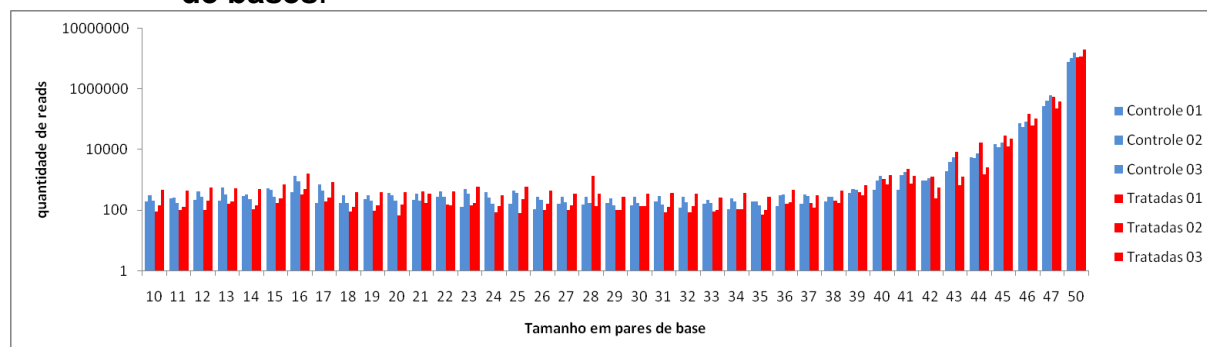
**TABELA 1 – Número total de *reads* nas bibliotecas antes e depois dos cortes com CUTADAPT.**

Amostra	Total de <i>reads</i>	Pós processamento	<i>reads</i> descartadas	% de <i>reads</i> descartadas
C1	8.619.437	8.466.374	153.063	2%
C2	11.586.080	11.305.255	280.825	2%
C3	17.343.307	17.076.287	267.020	2%
T1	12.009.775	11.720.352	289.423	2%
T2	12.414.074	12.324.700	89.374	1%
T3	21.055.880	20.870.101	185.779	1%

Fonte: Autor.

Nota: As linhas separam cada biblioteca onde o grupo C refere-se ao controle e o grupo T às tratadas. Nas colunas, temos da esquerda para a direita: Amostra, que exibe o nome da cada biblioteca, total de *reads*, que é o número total de *reads*; pós processamento, que é o total de *reads* após a remoção de pequenas *reads* (menores que 10 nt) e de adaptadores; *reads* descartadas, que é a quantidade de *reads* removidas; e % de *reads* descartadas, é o percentual deste descarte.

**FIGURA 10 – Distribuição da quantidade de *reads* por comprimento de pares de bases.**



Fonte: Autor.

Nota: No eixo y a quantidade de *reads* em valores de base 10, no eixo x o comprimento em pares de base, em azul o grupo controle e em vermelho o grupo KD para *jhe* onde se observa um leve aumento.

#### 4.3 BIBLIOTECAS APÓS PRÉ-PROCESSAMENTO

O mapeamento foi feito com o programa BWA (Li;Durbin, 2009) no genoma de *A. mellifera* (versão 4.5) recuperada da base de dados Hymenoptera Base (Elsik *et al.*, 2016), (figura 11). Foram mapeadas em torno de 85% das *reads*, como mostrado na figura 12, onde é possível observar os valores mapeados para cada biblioteca.

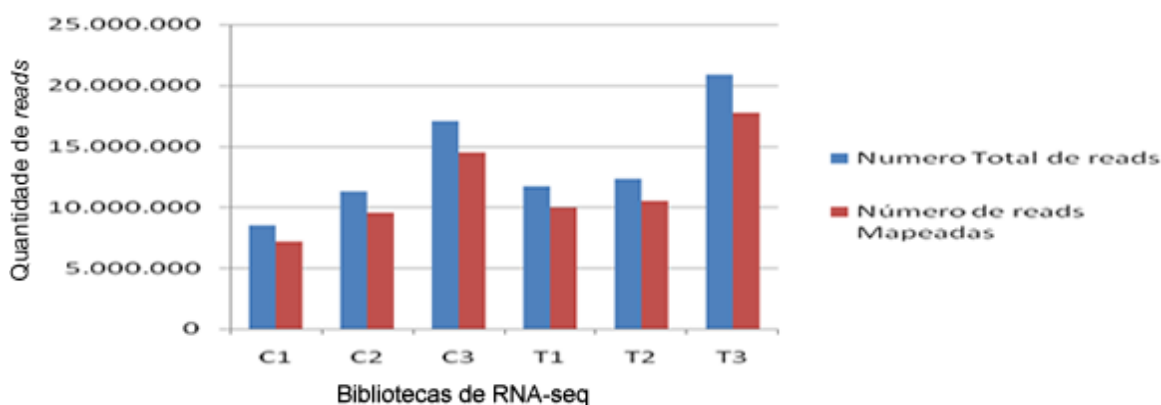
FIGURA 11 – Linhas de comando para uso do programa de mapeamento BWA.

```
bwa index Amel_4.5_scaffolds.fa
bwa aln -t 4 Amel_4.5_scaffolds.fa C01-1_cutadapt.fastq > Mapeamento_bwa_2.sai
bwa samse Amel_4.5_scaffolds.fa Mapeamento_bwa_2.sai C01-1_cutadapt.fastq >
alinhado_com_bwa.sam
```

Fonte: Autor.

Nota: A primeira linha refere-se à indexação do genoma. A segunda linha refere-se a primeira etapa do mapeamento e tem como saída um arquivo no formato \*.sai (arquivo binário). Na terceira linha, o arquivo \*.sai é utilizado de entrada para a segunda etapa do mapeamento em que são gerados arquivos de mapeamento no formato SAM.

FIGURA 12 – Número de *reads* mapeadas por bibliotecas.



Fonte: Autor.

Nota: No eixo y a quantidade de *reads* mapeada e no eixo x os grupos, sendo, controle em C e tratados em T, onde a cor azul refere-se ao total de *reads* que cada grupo contém e a cor vermelha o total de *reads* mapeadas. Aproximadamente 85% das *reads* foram mapeadas para cada biblioteca.

#### 4.4 IDENTIFICAÇÃO DE GENES DIFERENCIALMENTE EXPRESSOS

Para identificar os genes diferencialmente expressos foi utilizado o programa CUFFLNKS versão 2.2.1 (<http://cole-trapnell-lab.github.io/cufflinks/>). Para uso desta ferramenta os arquivos de saída do mapeamento no formato SAM foram convertidos com o programa SAMTOOLS (Li *et al.*, 2009) para o formato BAM. Com estes arquivos convertidos como mostra na linha de comando (Figura 13), foi utilizada a ferramenta Cuffdiff do pacote de ferramentas do CUFFLINKS que é utilizado para encontrar diferenças na expressão gênica

**FIGURA 13 – Linha de comando para uso do programa SAMTOOLS.**

```
samtools view -bS C1.sam > C01.bam
```

Fonte: Autor.

Nota: Converte um arquivo SAM em um arquivo BAM pelo comando “view”, o parâmetro “-bS” altera o formato de saída do padrão de SAM para BAM.

O Cuffdiff recebe como entrada os arquivos BAM, para comparar esses *datasets* quanto aos níveis de expressão, utilizando métodos *FPKMs* de normalizações das *reads* de todas as seis bibliotecas enviadas para análise (Figura 14). A próxima etapa foi de exploração dos dados obtidos como saída analisando os valores como o de  $\log_2$  *Fold Change* e valor de p.

**FIGURA 14 – Linha de comando para uso do programa Cuffdiff.**

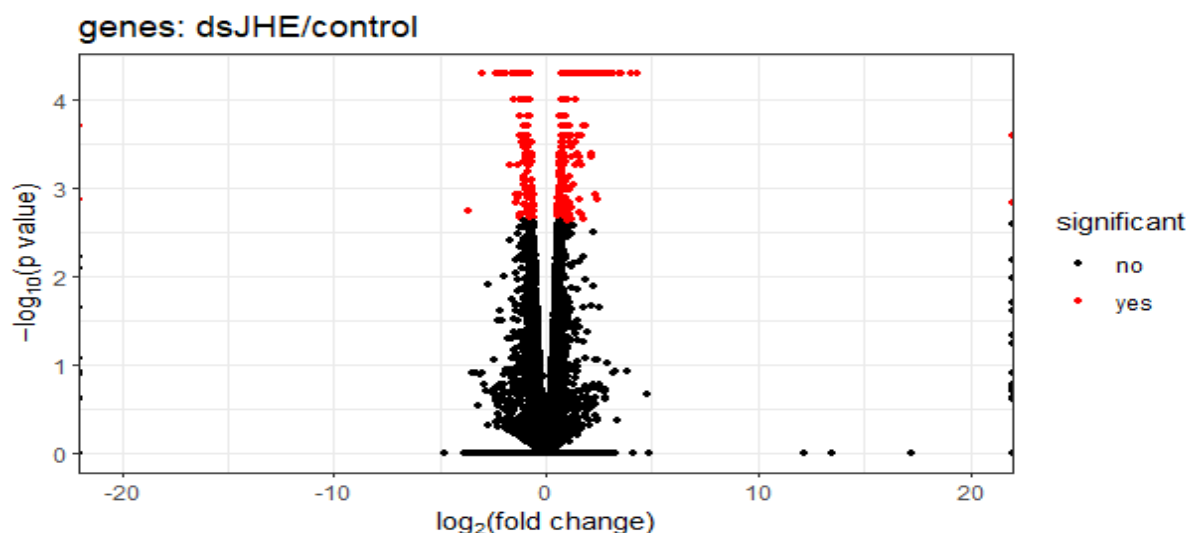
```
cuffdiff -labels control,dsJHE -output-dir cufflink/dados ../data/baixar/my.gtf  
cufflink/C01_sort.bam,cufflink/C02_sort.bam,cufflink/C03_sort.bam  
cufflink/T01_sort.bam,cufflink/T02_sort.bam,cufflink/T03_sort.bam
```

Fonte: Autor.

Nota: É uma ferramenta do pacote CUFFLINKS, recebendo como entrada as seis bibliotecas três controle e três tratadas que estão na pasta “cufflink”, também é passado um arquivo GTF com parâmetro de entrada; o nome “control, dsJHE” corresponde a coluna de cada grupo de bibliotecas é passado no parâmetro “-labels”.

Por meio do gráfico vulcão (Figura 15) é possível obter uma visualização da magnitude da expressão gênica. Foram registrados 458 genes diferencialmente expressos, marcados em vermelho, estes apresentaram o  $-\log_{10}$  (valor de p) maior que 2,5. Podemos observar a diferença entre os grupos controle e KD, dos quais 151 genes foram *down* regulados ou menos expressos nas abelhas tratadas e 307 genes *up* regulados ou mais expressos nas abelhas tratadas (dsRNA-*jhe*), conforme Tabela 2. Assumindo que o experimento de *knock-down* de *jhe* promoveu diminuição da expressão da *jhe* o que, por sua vez, conduziu à diminuição da respectiva atividade enzimática (Silva, 2012), os resultados revelam um grande grupo de genes de resposta a níveis aumentados de HJ.

FIGURA 15 – Gráfico Vulcão mostra os genes diferencialmente expressos entre as amostras dsRNA-*jhe* e controle.



Fonte: Autor.

Nota: Os níveis de expressão significativos estão em vermelho e os não significativos em preto. O eixo y apresenta os valores  $\log_{10}$  do valor de p e o eixo x os valores  $\log_2$  de *fold change* (valores negativos correspondem ao grupo controle e os valores positivos ao grupo dsRNA-*jhe*).

O que se observa é que com as análises realizadas pelos programas e pelas linhas de comando executadas foi possível apresentar um resultado satisfatório, tendo assim uma lista de genes diferencialmente expressos.

O valor de  $\log_2(\text{fold\_change})$  foi utilizado para classificar dois grupos, um com valores maiores que zero, o grupo tratadas, e outro com valores menores que zero, o grupo controle, seguindo o resultado de  $\log_2(\text{fold\_change}) > 0$  ou  $\log_2(\text{fold\_change}) < 0$  conforme se observa nas Tabelas 2.

TABELA 2 – Total de genes significativos.

Genes	Total	Significativos	UP $\log_2(\text{fold\_change}) > 0$	Down $\log_2(\text{fold\_change}) < 0$
			dsJHE	Controle
	15.314	458	307	151

Fonte: Autor.

Notas: Quantidade de genes diferencialmente expressos entre as amostras de abelhas controle e KD para *jhe* valor de  $p < 0,05$  para  $\log_2$ .

Os genes diferencialmente expressos obtidos da análise CUFFLINKS foram organizados em planilha, onde foi possível criar duas as tabelas que possuem

colunas com os valores do GBs do Gene, *Gene symbol*, *Gene description*, valor do *fold change* e valor de *p*, sendo 215 para o grupo tratadas e 124 para o grupo controle, totalizando 339 genes dos 458 encontrados, pois alguns não foram encontrados no site NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

#### 4.5 ANÁLISE FUNCIONAL DOS GENES DIFERENCIALMENTE EXPRESSOS

A análise de enriquecimento funcional é um método estatístico para identificar classes de genes ou proteínas que estão super-representadas, o que possibilita agrupar os genes que estão envolvidos no mesmo processo biológico ou vias de desenvolvimento. Em análise de expressão gênica, os pesquisadores geralmente desejam recuperar um perfil funcional desse conjunto de genes a fim de compreender melhor os processos biológicos subjacentes (Huang *et al.*, 2009).

O DAVID (<https://david.ncifcrf.gov/>) é uma ferramenta disponível *on-line* que fornece um ambiente de anotação funcional de genes. Esta ferramenta permite analisar grandes listas de genes de diversas espécies e retorna os processos biológicos e vias de desenvolvimento enriquecidas. Uma lista de genes de cada grupo (controle, que são os genes *down* regulados ou menos expressos nas abelhas tratadas, e tratadas, que são os genes *up* regulados ou mais expressos nas abelhas tratadas (dsRNA-*jhe*), foi submetida no site DAVID para análise de enriquecimento funcional. Na análise, o total de 458 genes diferencialmente expressos somando os dois grupos (tratadas com 307 genes e controle com 151 genes encontrados), foram submetidos à análise de ontologia gênica. A análise mostrou que, desse total, 116 (95 do grupo tratadas e 23 do grupo controle) estavam enriquecidos em determinadas categorias biológicas (*Fold Enrichment* acima de 1,0). Posteriormente, os genes de cada grupo foram analisados separadamente, atribuindo a estes as categorias de processos biológicos, funções moleculares e componentes celulares. Esta estratégia nos permitiu incluir genes adicionais, mesmo estes apresentando baixos valores de *p* e *Fold Enrichment*.

A maioria dos genes que tiveram sua via molecular caracterizada foram os do grupo tratadas (*KD* para *jhe*) com 139 genes, ou seja 45% dos genes. Já o grupo controle teve 47 dos seus genes encontrados, correspondendo a 31% dos genes, como mostra o apêndice E.

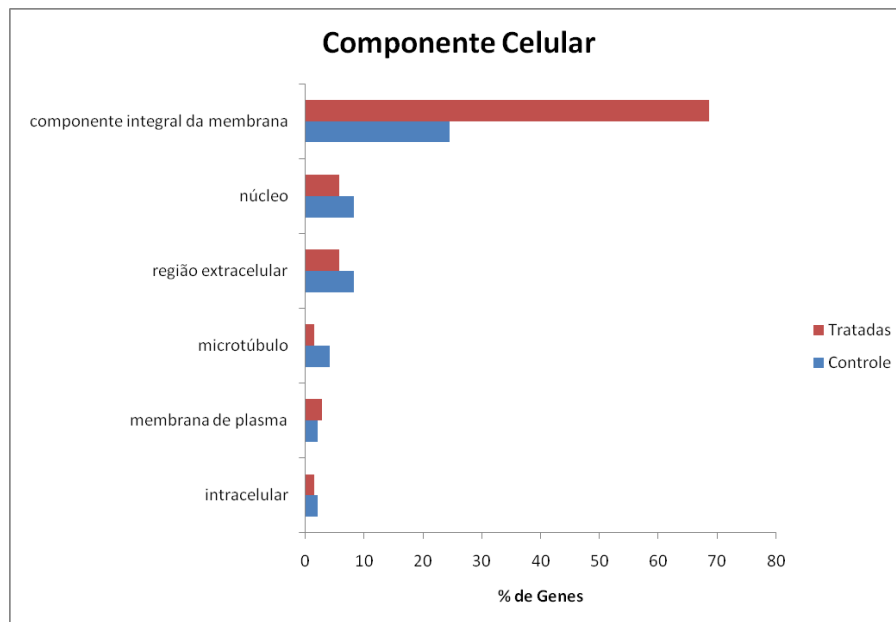
O maior número de genes do grupo *KD* para *jhe* no processo biológico (BP)

foi associado ao processo de transporte transmembrana (11%), processo metabólico (7,4%), processo metabólico da quitina (7,4%), processo metabólico de carboidratos (5,56%), processo metabólico do composto de nitrogênio (5,56%). Em função molecular (MF), os maiores números de genes foram de Ligação de ATP (6,4%), constituinte estrutural da cutícula (4,65%), ligação de íon zinco (4,65%); e em Componente Celular (CC), o maior percentual foi para componente integral da membrana (68,6%). Já no grupo controle, em BP a maioria foram associados à regulação da transcrição modelada por *DNA* (10%), processamento de *rRNA* (7,5%), movimento baseado em microtúbulos (7,5%); em MF, Ligação de *DNA* (8,2%), ligação de íon zinco (11,5%) e ligação de ATP (19,7%); e CC, na região extracelular (8,9%), núcleo (15,5%), componente integral da membrana (28,9%).

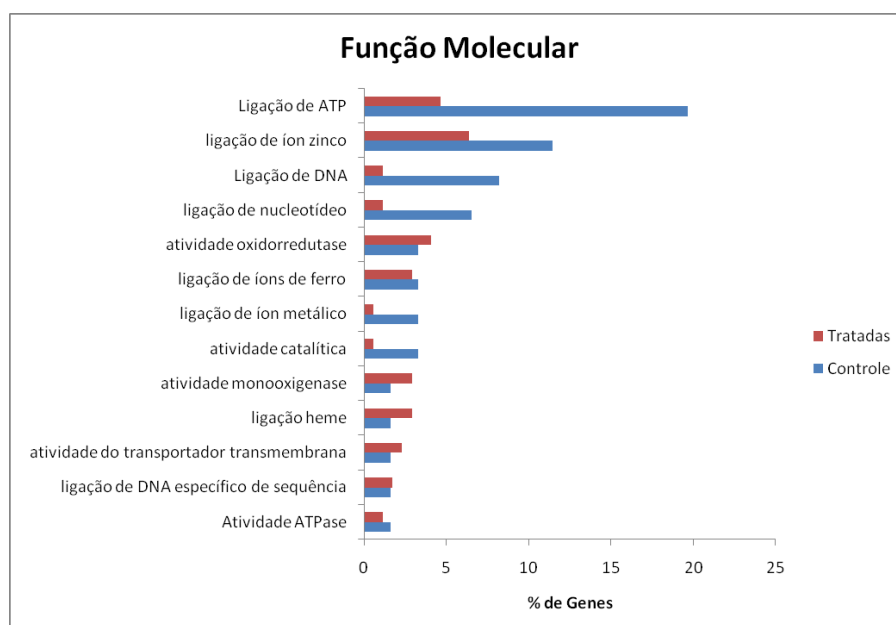
A comparação dos perfis de enriquecimento funcional entre ambos grupos de genes revela diferenças biologicamente relevantes. A análise funcional, através da ontologia gênica (GO) representados na categoria de processos biológicos no grupo tratado, mostrou que 3,7% dos genes foram mapeados para "Regulação da transcrição, modelada por *DNA*" (GO: 0006355), e 7,5% dos genes para "Processos Metabólicos" (GO: 0008152). Estes resultados parecem opor-se aos obtidos com o grupo controle (Figura 16), que obteve 7,5% dos genes mapeados para "Regulação da transcrição, modelada por *DNA*" (GO: 0006355), e 2,5% dos genes para "Processos Metabólicos" (GO: 0008152). Este resultado sugere que o aumento de HJ endógeno pelo silenciamento da *jhe* induziu a expressão de genes associados a processos metabólicos, lembrando, assim, os resultados relatados por Barchuk *et al.*, (2007). Usando aplicação tópica de HJ, estes autores mostraram que enquanto as operárias regulam mais genes de desenvolvimento, nas larvas de futuras rainhas, cujo desenvolvimento é desviado para um crescimento geral do organismo, muitos dos genes expressos induzidos por altos níveis de HJ estão associados a processos fisiometabólicos. Isso favorece o desenvolvimento de estruturas distintas de castas como cérebro, pernas e ovários (Barchuk *et al.*, 2007, Cristino *et al.*, 2006). Outras características funcionais dos genes analisados que chamam a atenção são a associação a estruturas membranosas, no caso dos genes do grupo tratado, e a funções de ligação molecular, no caso dos genes do grupo controle. Ambas as características merecem ser abordadas em análises futuras, para desvendar pormenores das funções dos genes que respondem a títulos elevados de HJ durante o desenvolvimento larval de *A. mellifera*.

FIGURA 16 – Análise de enriquecimento funcional.

A

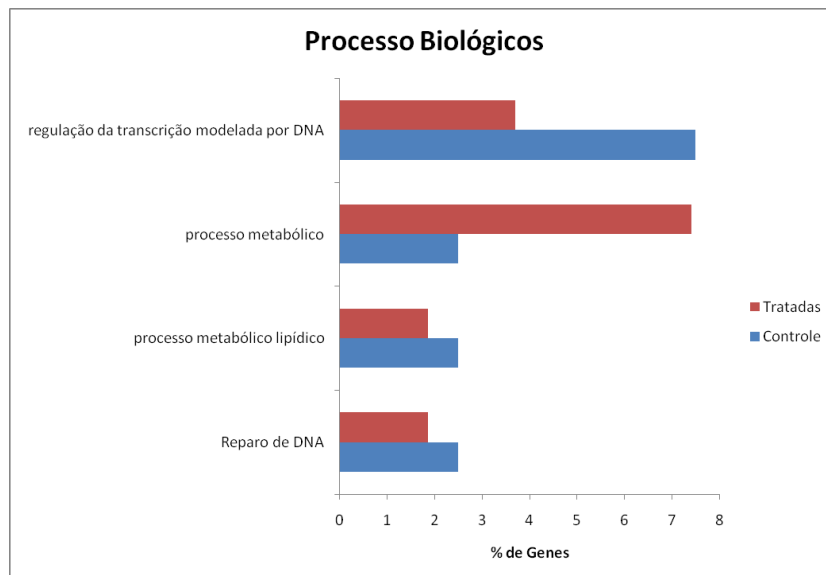


B





C



Fonte: Autor.

Nota: Os gráficos A, B e C exibem as categorias de componentes celulares, funções moleculares e processos biológicos, no eixo X o percentual de expressão dos genes e no eixo Y as categorias correspondentes.

Finalmente, foi criado um tutorial (Apêndice B), que está disponível na *web* através da página: (<https://www.unifal-mg.edu.br/biologiadassociedade/informacoes-e-servicos>). Com explicações sobre a instalação dos programas que foram utilizados para execução deste *pipeline*. As orientações deste tutorial são genéricas, podendo sofrer alterações específicas para cada computador, como atualização do sistema operacional ou instalação ou atualização de algum outro *software* necessário para a perfeita execução dos comandos apresentados neste tutorial.

## 5 CONCLUSÃO

O *pipeline* desenvolvido permitiu atingir satisfatoriamente os objetivos propostos, revelando grupos de genes diferencialmente expressos entre as duas populações de *RNA* analisadas, as provenientes das abelhas dsRNA-*jhe* e as das abelhas controle. Especificamente, foi possível realizar a criação de uma sequência de comandos para coletar os genes diferencialmente expressos em bibliotecas de *RNA-seq*, utilizando programas de análise de transcriptomas, analisando o tamanho das *reads*, seus níveis de qualidade, a quantidades de sequências antes e após o tratamento dos dados e, finalmente, a identificação de genes diferencialmente expressos em cada uma das amostras biológicas, com suas respectivas características funcionais.

A aplicação do *pipeline* às amostras biológicas analisadas revelou muitos genes que respondem ao aumento nos títulos endógenos de HJ (pelo *KD* de *jhe*) durante o desenvolvimento larval de *A. mellifera*.

O aumento de HJ endógeno promovido pelo silenciamento da *jhe* induziu a expressão de genes associados a processos metabólicos, de maneira semelhante ao que acontece quando o aumento artificial nos títulos deste hormônio é possibilitado pela aplicação tópica.

O *pipeline* desenvolvido poderá ser utilizado para mapear quaisquer tipos de transcritos visando destacar os genes diferencialmente expressos entre duas amostras.

## REFERÊNCIAS

A BÍBLIA. **Salmo 18:3**. Editora Ave Maria, 208ª edição. 669 pg. 2016

ANDREWS, S. (2013). **Fastqc**. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (Acesso em : 30 set. 2020)

ALMEIDA, Eduardo AB; PORTO, Diego Sasso. Investigating eusociality in bees while trusting the uncertainty. **Sociobiology**, v. 61, n. 4, p. 355-368, 2014.

BARCHUK, Angel R. et al. Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. **BMC developmental biology**, v. 7, n. 1, p. 1-19, 2007.

BENTLEY, David R. Whole-genome re-sequencing. **Current opinion in genetics & development**, v. 16, n. 6, p. 545-552, 2006.

BOMTORIN, Ana Durvalina et al. Juvenile hormone biosynthesis gene expression in the corpora allata of honey bee (*Apis mellifera* L.) female castes. **PloS one**, v. 9, n. 1, p. e86923, 2014.

BREED, Michael D.; GAMBOA, George J. Behavioral control of workers by queens in primitively eusocial bees. **Science**, v. 195, n. 4279, p. 694-696, 1977.

COLLINS, Francis S.; MORGAN, Michael; PATRINOS, Aristides. The Human Genome Project: lessons from large-scale biology. **Science**, v. 300, n. 5617, p. 286-290, 2003.

COUNSELL, Damian. A review of bioinformatics education in the UK. **Briefings in Bioinformatics**, v. 4, n. 1, p. 7-21, 2003.

CRISTINO, A. S. et al. Caste development and reproduction: a genome-wide analysis of hallmarks of insect eusociality. **Insect molecular biology**, v. 15, n. 5, p. 703-714, 2006.

DUBROVSKY, Edward B. et al. The *Drosophila* FTZ-F1 nuclear receptor mediates juvenile hormone activation of *E75A* gene expression through an intracellular pathway. **Journal of Biological Chemistry**, v. 286, n. 38, p. 33689-33700, 2011.

ELSIK, Christine G. et al. Finding the missing honey bee genes: lessons learned from a genome upgrade. **BMC genomics**, v. 15, n. 1, p. 1-29, 2014.

ELSIK, Christine G. et al. Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. **Nucleic acids research**, v. 44, n. D1, p. D793-D800, 2016.

EVANS, Jay D.; WHEELER, Diana E. Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. **Proceedings of the National Academy of Sciences**, v. 96, n. 10, p. 5575-5580, 1999.

GILBERT, Lawrence I.; GRANGER, Noelle A.; ROE, R. Michael. The juvenile hormones: historical facts and speculations on future research directions. **Insect biochemistry and molecular biology**, v. 30, n. 8-9, p. 617-644, 2000.5

GLENN, Travis C. Field guide to next-generation DNA sequencers. **Molecular ecology resources**, v. 11, n. 5, p. 759-769, 2011.

HARTFELDER, Klaus; ENGELS, Wolf. 2 Social insect polymorphism: hormonal regulation of plasticity in development and reproduction in the honeybee. **Current topics in developmental biology**, v. 40, p. 45-77, 1998.

HINTON, A. C.; HAMMOCK, B. D. Purification of juvenile hormone esterase and molecular cloning of the cDNA from *Manduca sexta*. **Insect biochemistry and molecular biology**, v. 32, n. 1, p. 57-66, 2001.

HUANG, Da Wei; SHERMAN, Brad T.; LEMPICKI, Richard A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. **Nucleic acids research**, v. 37, n. 1, p. 1-13, 2009.

HUANG, Da Wei; SHERMAN, Brad T.; LEMPICKI, Richard A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature protocols**, v. 4, n. 1, p. 44-57, 2009.

JAY, S. Cameron. The development of honeybees in their cells. **Journal of Apicultural Research**, v. 2, n. 2, p. 117-134, 1963.

JINDRA, Marek; PALLI, Subba R.; RIDDIFORD, Lynn M. The juvenile hormone signaling pathway in insect development. **Annual review of entomology**, v. 58, p. 181-204, 2013.

DE KORT, C. A. D.; GRANGER, N. A. Regulation of JH titers: the relevance of degradative enzymes and binding proteins. **Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America**, v. 33, n. 1, p. 1-26, 1996.

LI, Heng; DURBIN, Richard. Fast and accurate short read alignment with Burrows–Wheeler transform. **bioinformatics**, v. 25, n. 14, p. 1754-1760, 2009.

LUCHT-BERTRAM, Elisabeth. Das postembryonale wachstum von hirnteilen bei *Apis mellifica* L. und *Myrmeleon europaeus* L. **Zeitschrift für Morphologie und Ökologie der Tiere**, v. 50, n. 5, p. 543-575, 1961.

MACKERT, Aline et al. Identification of a juvenile hormone esterase-like gene in the honey bee, *Apis mellifera* L.—expression analysis and functional assays. **Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology**, v. 150, n. 1, p. 33-44, 2008.

MORITZ, Robin FA et al. Rare royal families in honeybees, *Apis mellifera*. **Naturwissenschaften**, v. 92, n. 10, p. 488-491, 2005.

PALMER, Kellie A.; OLDROYD, Benjamin P. Evolution of multiple mating in the genus *Apis*. **Apidologie**, v. 31, n. 2, p. 235-248, 2000.

PATEL, Avani et al. The making of a queen: TOR pathway is a key player in diphenic caste development. **PloS one**, v. 2, n. 6, p. e509, 2007.

QUAIL, Michael A. et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. **BMC genomics**, v. 13, n. 1, p. 1-13, 2012.

RACHINSKY, Anna et al. Caste and metamorphosis: hemolymph titers of juvenile hormone and ecdysteroids in last instar honeybee larvae. **General and comparative endocrinology**, v. 79, n. 1, p. 31-38, 1990.

REMBOLD, H.; CZOPPELT, C. H.; RAO, P. J. Effect of juvenile hormone treatment on caste differentiation in the honeybee, *Apis mellifera*. **Journal of insect physiology**, v. 20, n. 7, p. 1193-1202, 1974.

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

VAN ROSSUM, Guido; DRAKE, Fred L. **The python language reference manual**. Network Theory Ltd., 2011.

ROY, Somak et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. **The Journal of Molecular Diagnostics**, v. 20, n. 1, p. 4-27, 2018.

SANGER, Frederick; NICKLEN, Steven; COULSON, Alan R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the national academy of sciences**, v. 74, n. 12, p. 5463-5467, 1977.

SILVA, Aline Carolina Aleixo. **Padrões Diferenciais de Expressão Gênica no Desenvolvimento das Castas de *Apis mellifera*, com Ênfase na Diferenciação das Operárias**. Tese de Doutorado. Universidade de São Paulo.

TRAPNELL, Cole et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature biotechnology**, v. 28, n. 5, p. 511-515, 2010.

VIEIRA, Carlos Ueira et al. Differential gene expression in *Melipona scutellaris* (Hymenoptera, Meliponini): effect of juvenile hormone III. **Bioscience Journal**, v. 22, n. 1, 2006.

WHEELER, D. E.; BUCK, NORMAN; EVANS, Jay D. Expression of insulin pathway genes during the period of caste determination in the honey bee, *Apis mellifera*. **Insect molecular biology**, v. 15, n. 5, p. 597-602, 2006.

WILSON, Edward O. et al. The insect societies. **The insect societies.**, 1971.

WIRTZ, Peter. **Differentiation in the honeybee larva: a histological, electron-microscopical and physiological study of caste induction in Apis mellifera mellifera L.** 1973. Tese de Doutorado. Veenman.

ZHANG, Zhaolin et al. Steroid receptor co-activator is required for juvenile hormone signal transduction through a bHLH-PAS transcription factor, methoprene tolerant. **Journal of Biological Chemistry**, v. 286, n. 10, p. 8437-8447, 2011.

## APÊNDICE A-Tabela de genes

Tabela S1 - Tabela de genes (valores positivos de *fold change*) mais expressos nas abelhas tratadas

Gene	Genesymbol	Genedescription	Foldchange	p-value
GB45950	LOC412056	organiccationtransporterprotein	0,575399	0,0017
GB52656	LOC552154	uncharacterizedLOC552154	0,580853	0,00005
GB54605	LOC409057	uncharacterizedLOC409057	0,581803	0,0018
GB46145	LOC408722	sodium/potassium-transportingATPasesubunitbeta-2	0,613649	0,00075
GB55574	LOC409192	probablephospholipid-transportingATPaseVD	0,622188	0,00065
GB51666	LOC412694	uncharacterizedLOC412694	0,634153	0,00215
GB45613	LOC410654	cholecystokininreceptorlike	0,649106	0,001
GB48321	LOC413631	vaninlikeprotein1	0,650304	0,00005
GB43256	LOC411685	ATPbindingcassettesubfamilyDmember1	0,659048	0,00005
GB50958	LOC408286	2oxoglutaratedehydrogenase,mitochondrial	0,662303	0,00005
GB53141	LOC726980	uncharacterizedLOC726980	0,664534	0,00005
GB48095	LOC725915	MAPKregulatedcorepressorinteractingprotein2	0,665284	0,0003
GB50686	LOC408411	transcriptionfactorSox2	0,669253	0,00005
GB50701	LOC412386	pleckstrinhomologydomain-containingfamilyMmember1	0,672177	0,00055
GB50170	LOC100577783	proteinSkeletalisoformsB/C	0,68208	0,00145
GB55598	TpnI	troponinI	0,685441	0,00055
GB49076	LOC408759	proteinFAM102B	0,685816	0,00005
GB48687	LOC725413	fibrillinIlike	0,691015	0,00005
GB46597	LOC413934	sodiumindependentsulfateaniontransporter	0,694106	0,00005
GB42822	LOC725178	chitooligosaccharidolyticbetaNacetylglucosaminidase	0,696503	0,00075
GB50890	LOC409649	solutecarrierorganicianiontransporterfamilymember5A1	0,696524	0,00005
GB54818	LOC411285	muscleLIMproteinMlp84B	0,705075	0,00025
GB40076	LOC410448	carcininetransporter	0,707805	0,00005
GB45714	Tg	transglutaminase	0,713179	0,00005
GB52432	LOC408697	KNmotifandankyrinrepeatdomaincontainingprotein3	0,713436	0,00005
GB45910	LOC724367	proteinlethal(2)essentialforlife	0,722094	0,00005
GB43143	LOC552804	tubulinspecificchaperonecofactorElikeprotein	0,722714	0,00045
GB48175	LOC551179	methylfarnesoateepoxidase	0,725538	0,00005
GB51165	LOC726470	solutecarrierorganicianiontransporterfamilymember1A5	0,727203	0,0009
GB51653	LOC409843	myosinheavychainmuscle	0,736162	0,0001
GB52185	LOC100578041	uncharacterizedLOC100578041	0,736877	0,00005
GB49331	LOC724772	leucine-richrepeatneuronalprotein1	0,737913	0,0006
GB50932	LOC725389	boxAbindingfactor	0,739027	0,00005
GB48216	TpnCIIB	troponinCtypeIIb	0,740937	0,0002
GB45400	LOC552285	innexinx7	0,748625	0,0005

GB50116	LOC725202	chymotrypsininhibitor	0,759008	0,00005
GB47081	LOC408504	proteinFAM160B1	0,759151	0,0003
GB55328	LOC408583	tropomyosin	0,761525	0,00005
GB47990	LOC408414	tropomyosin1	0,76287	0,0001
GB48842	LOC725838	uncharacterizedLOC725838	0,768045	0,00055
GB45906	LOC410087	uncharacterizedLOC410087	0,771029	0,00005
GB42792	LOC409805	uncharacterizedLOC409805	0,775878	0,00005
GB48066	LOC413048	bifunctional3'-phosphoadenosine5'-phosphosulfatesynthase2	0,779712	0,0001
GB55764	LOC410944	ATPbindingcassettesubfamilyCmemberSur	0,780743	0,0004
GB42800	LOC727028	uncharacterizedLOC727028	0,788183	0,00005
GB53696	LOC412929	androgeninducedgene1 protein	0,789884	0,00005
GB46444	LOC413678	serinepyruvateaminotransferase,mitochondrial	0,789973	0,0005
GB55498	LOC410802	microphthalmiaassociatedtranscriptionfactor	0,790072	0,00005
GB53925	LOC724993	uncharacterizedLOC724993	0,791801	0,00055
GB42567	LOC411229	venomdipeptidylpeptidase4like	0,794742	0,0002
GB54817	LOC408972	muscle-specificprotein20	0,798458	0,00095
GB49017	LOC412007	facilitatedtrehalosetransporterTret1	0,804875	0,00005
GB43029	LOC725851	actin,indirectflightmusclelike	0,805633	0,00005
GB51814	LOC410747	glucosedehydrogenase[FAD,quinone]	0,805726	0,00005
GB54172	LOC410507	sodium-independent sulfateaniontransporter	0,811111	0,00015
GB41310	LOC410075	actin,musclelike	0,815847	0,00005
GB43225	LOC410278	putativeinorganicphosphatecotransporter	0,818048	0,00005
GB54446	Argk	argininekinase	0,820141	0,00045
GB53775	LOC413168	retinoldehydrogenase14	0,821379	0,0002
GB47474	LOC726115	D-2-hydroxyglutaratedehydrogenasemitochondrial	0,824227	0,0003
GB43890	LOC411392	proteinRed	0,824229	0,0001
GB48905	GstS1	glutathioneStransferaseS1	0,825695	0,00005
GB52702	LOC552592	facilitatedtrehalosetransporterTret1	0,828417	0,00005
GB44366	LOC411200	tryptophan5hydroxylase1	0,832947	0,0012
GB53028	LOC724890	laccase1	0,833365	0,00005
GB51787	LOC410058	myosinlightchainalkali	0,837197	0,00005
GB50526	LOC410626	sodiumcoupledd monocarboxylatetransporter1	0,84218	0,00005
GB50630	LOC412316	sodium/potassium/calciumexchanger3	0,843006	0,00035
GB50222	LOC726069	atrialnatriureticpeptidereceptor1	0,843187	0,00005
GB45969	LOC725204	tyrosineaminotransferase	0,846012	0,00035
GB52023	CYP6AQ1	cytochromeP4506AQ1	0,847689	0,00175
GB42616	LOC726818	beta-hexosaminidasesubunitbeta	0,848111	0,00005
GB55602	LOC100578412	uncharacterizedLOC100578412	0,849427	0,00215
GB49147	LOC412923	argininosuccinatesynthase	0,849707	0,00005
GB50262	LOC410578	sodiumandchloridedependentglycinetransporter2	0,854221	0,00215
GB40288	CYP6AS1	cytochromeP4506AS1	0,863077	0,00105



GB49757	Fabp	fattyacidbindingprotein	0,865943	0,00165
GB41720	LOC727121	uncharacterizedLOC727121	0,868265	0,0005
GB41495	LOC100576703	sarcalumenin	0,868523	0,00115
GB48823	CPR2	cuticularprotein2	0,872123	0,0001
GB55864	LOC725106	UDP-glucuronosyltransferase1-8	0,884206	0,00005
GB45973	LOC410638	aromaticLaminoaciddecarboxylase	0,88588	0,00065
GB41413	Tpx4	thioredoxinperoxidase4	0,886143	0,0003
GB52103	LOC100578548	uncharacterizedLOC100578548	0,887691	0,0004
GB42895	LOC100578774	centrosomalproteinof164kDa	0,888989	0,00115
GB42326	LOC412263	endoplasmicreticulumaminopeptidase1	0,891264	0,0013
GB40907	LOC413128	putativeferricchelatereductaseIhomolog	0,898954	0,0005
GB44074	LOC410996	tubulinbetachain	0,90287	0,00215
GB53404	LOC411296	silk glandfactor1	0,905632	0,0019
GB42319	LOC409105	uncharacterizedLOC409105	0,913911	0,00115
GB50123	LOC408532	myophilin	0,91876	0,00045
GB46223	Obp14	odorantbindingprotein14	0,931319	0,00135
GB40240	LOC409881	myosinregulatorylightchain2	0,938445	0,00015
GB52756	LOC408474	apyrase	0,940068	0,00005
GB40284	LOC725159	probablecytochromeP4506a14	0,941672	0,0004
GB54611	LOC413749	serineproteaseinhibitor88Ea	0,948254	0,00005
GB52779	LOC410527	probablecationtransportingATPase13A3	0,954382	0,0004
GB51732	LOC100578928	calciumandintegrinbindingprotein1	0,958499	0,00015
GB51504	LOC410114	solutecarrierfamily23member1	0,966679	0,0002
GB41670	LOC410566	putativeinorganicphosphatecotransporter	0,971391	0,00065
GB51419	LOC100577680	uncharacterizedLOC100577680	0,973575	0,0009
GB46224	Obp15	odorantbindingprotein15	0,976048	0,0006
GB40769	LOC412458	farnesoldehydrogenase	0,976782	0,00005
GB53831	LOC551744	phospholipaseABHD3	0,979716	0,00005
GB55729	LOC727110	yellow-x2	0,992973	0,00005
GB50970	LOC408291	3ketoacylCoAthiolase,mitochondrial	0,996111	0,0003
GB51979	LOC552190	prisilkin-39	1,00317	0,00005
GB43298	LOC725019	pro-resilin-like	1,01341	0,00005
GB55818	LOC412883	AllatostatinCreceptor	1,01411	0,0001
GB46697	LOC408827	carbonicanhydrase1	1,0175	0,00095
GB50434	LOC413117	proton-coupledaminoacidtransporter-likeproteinpathetic	1,02143	0,00005
GB51494	LOC412843	phosphoenolpyruvatecarboxykinase	1,03337	0,00005
GB54675	LOC552784	uncharacterizedLOC552784	1,03684	0,00005
GB50648	LOC409827	serineproteinasesubble	1,04063	0,00005
GB49887	LOC726690	probablecytochromeP4506a14	1,0514	0,00085
GB42985	LOC725646	N-acetylneuraminatelyase	1,05217	0,00005
GB42764	LOC100576640	uncharacterizedLOC100576640	1,05533	0,0002

GB44457	LOC411100	FGGYcarbohydratekinasedomain-containingprotein	1,06373	0,00005
GB44120	LOC409143	venomserineprotease34	1,06833	0,00005
GB50636	LOC724382	cuticularproteinanalogoustoperitrophins3A	1,0734	0,0001
GB41240	LOC552149	aquaporinAQPAn.G	1,07664	0,00105
GB46842	LOC727598	probablecytochromeP4506a13	1,07759	0,00005
GB48289	LOC726292	natterin3	1,09226	0,0013
GB45855	LOC410729	putativeserineproteaseK12H4.7	1,09243	0,00005
GB48435	LOC411691	phosphotriesteraserelatedprotein	1,09302	0,00055
GB53576	LOC406093	apisimin	1,09519	0,00005
GB43713	LOC410492	cytochromeP4509e2	1,10472	0,00005
GB49543	LOC408817	alanineglyoxylateaminotransferase2like	1,10731	0,00005
GB51214	TpnT	troponinTskeletalmuscle	1,11058	0,00005
GB46312	CPR22	cuticularprotein22	1,11483	0,00105
GB49639	LOC726072	putativedefenseprotein3	1,12345	0,00025
GB42550	LOC100577026	mitochondrialamidoximereducingcomponent1	1,1235	0,00165
GB48492	JHBP1	takeoutlikecarrierprotein	1,12721	0,0002
GB52571	LOC100576170	mucin-1	1,13037	0,00005
GB52553	LOC725541	proteinspaetzle5	1,13877	0,00005
GB40478	LOC412439	PDFreceptor	1,14093	0,00005
GB40377	LOC551717	putativeuncharacterizedproteinDDB_G0287265	1,14389	0,002
GB47933	LOC408534	trypsin	1,14999	0,00015
GB44073	LOC410997	organiccationtransporter1	1,15013	0,0008
GB42704	LOC412768	proteintakeout	1,151	0,001
GB49705	LOC410013	calexectin-2	1,17963	0,00005
GB48109	LOC409162	retinoidinducibleserinecarboxypeptidase	1,18788	0,00025
GB52161	CPR28	cuticularprotein28	1,20691	0,0001
GB48850	LOC551094	fatty-acidamidehydrolase2-B	1,22497	0,00035
GB49886	CYP6AS2	cytochromeP4506AS2	1,23384	0,00005
GB50477	LOC100577527	uncharacterizedLOC100577527	1,23747	0,00005
GB51733	LOC726571	venomacidphosphataseAcp1	1,25882	0,0013
GB40997	LOC411021	UDP-glucuronosyltransferase2B18	1,28697	0,00005
GB52757	LOC408473	uncharacterizedLOC408473	1,29722	0,00005
GB42287	LOC100577577	peritrophin1	1,31204	0,00005
GB48656	LOC412969	uncharacterizedLOC412969	1,32734	0,00005
GB47974	LOC726134	carboxylesterase	1,32846	0,00005
GB43285	LOC411599	uridinephosphorylase1	1,37577	0,00005
GB40137	LOC726352	transmembraneproteaseserine11Blikeprotein	1,38072	0,0004
GB50610	LOC100576418	tetra-peptiderepeathomeoboxprotein1	1,38178	0,00005
GB52186	LOC409626	chymotrypsin2	1,39013	0,00005
GB52877	LOC726699	uncharacterizedLOC726699	1,40255	0,00045
GB49930	LOC725069	peptidettransporterfamily1	1,40794	0,00005

GB40157	LOC408421	uncharacterizedLOC408421	1,41472	0,00005
GB52667	LOC552202	monocarboxylatetransporter9	1,43187	0,00005
GB44223	LOC552249	lysosomalalphamannosidase	1,4362	0,00005
GB51107	LOC100578731	uncharacterizedLOC100578731	1,45138	0,0016
GB42515	LOC552066	FAD-dependentoxidoreductasedomain-containingprotein1	1,45448	0,00005
GB43314	LOC724612	thyrotropinreleasinghormonedegradingectoenzyme	1,46165	0,0013
GB54745	LOC100577515	uncharacterizedLOC100577515	1,484	0,00025
GB46297	CPR14	cuticularprotein14	1,48646	0,00005
GB55499	LOC551687	uncharacterizedLOC551687	1,54852	0,00105
GB40713	Fabp	FABP-likeprotein	1,56815	0,00005
GB41912	LOC552024	myoinositol2dehydrogenase	1,58056	0,00045
GB43739	LOC551327	carboxypeptidaseB	1,58914	0,00005
GB49753	LOC410870	protonchannelOtopLc	1,60538	0,00005
GB52101	LOC412570	RYamidereceptor	1,61538	0,00005
GB40635	LOC724418	venomacidphosphataseAcpH-1-like	1,63157	0,00005
GB46299	CPR18	cuticularprotein18	1,63676	0,00005
GB46448	LOC409646	cationicaminoacidtransporter3	1,66789	0,00005
GB50647	LOC100576248	balbianiringprotein3	1,66933	0,00005
GB50026	LOC725250	trypsin-7-like	1,68736	0,00025
GB54554	LOC100576182	pupalcuticleproteinG1A	1,70644	0,00005
GB40299	CPR5	cuticularprotein5	1,71205	0,00025
GB49848	LOC100577143	uncharacterizedLOC100577143	1,7179	0,00025
GB48190	LOC551782	bestrophin4	1,73653	0,0005
GB53354	LOC726934	PIPLCXdomaincontainingprotein1	1,74083	0,00005
GB50650	LOC409826	proteoglycan4	1,76522	0,00005
GB43310	LOC724312	vaninlikeprotein1	1,7756	0,001
GB49700	LOC726762	facilitatedtrehalosetransporterTret1	1,77594	0,00015
GB46286	LOC551524	zinccarboxypeptidase	1,79216	0,00005
GB45771	LOC724825	glycine-richRNA-bindingprotein7	1,80573	0,00005
GB40770	LOC724721	farnesoldehydrogenase	1,82105	0,00005
GB43006	LOC408603	glucosedehydrogenase[FADquinone]	1,83823	0,00005
GB47120	SIFR	SIFamidereceptor	1,84008	0,00005
GB46403	LOC725818	uncharacterizedLOC725818	1,84309	0,00005
GB52203	CPR13	cuticularprotein13	1,8756	0,00005
GB51303	LOC724422	homeoboxproteinHox-B1a	1,87625	0,00005
GB46310	CPR17	cuticularprotein17	1,89981	0,0001
GB41760	LOC411353	lipase3	1,91255	0,00005
GB45565	LOC552301	chymotrypsin-1	1,95791	0,00005
GB49972	LOC100578044	Fbox/LRRrepeatprotein7	2,06844	0,00005
GB45938	LOC410639	alphamethylpaphypersensitiveprotein	2,1328	0,0022
GB42426	LOC551180	aminopeptidaseN	2,28011	0,00005

GB53306	LOC408992	NPCintracellularcholesteroltransporter1homolog1b	2,28642	0,00005
GB53911	LOC725725	peritrophin-1	2,29221	0,00005
GB48391	LOC726592	proteoglycan4	2,31347	0,00005
GB41222	LOC724780	GproteincoupledreceptorMth2	2,36209	0,00005
GB49282	LOC551941	uncharacterizedLOC551941	2,47728	0,00005
GB50901	LOC413924	glyceraldehyde3phosphatedehydrogenase	2,6386	0,0011
GB42053	LOC724386	NPCintracellularcholesteroltransporter2	2,76793	0,00005
GB47975	LOC411223	esteraseFE4	2,79	0,00035
GB54508	LOC725439	uncharacterizedLOC725439	2,93623	0,00005
GB43311	LOC100578995	vanin-likeprotein1	3,02288	0,00005
GB43688	LOC100578392	uncharacterizedLOC100578392	3,32098	0,00005
GB43689	LOC727522	uncharacterizedLOC727522	3,49773	0,00015
GB43690	LOC727344	uncharacterizedLOC727344	3,98991	0,00005
GB51436	LOC727243	uncharacterizedLOC727243	4,28484	0,0002
GB47079	LOC551005	hexokinasetype2	0,82295	0,00005
GB42150	LOC410771	WDrepeatcontainingprotein47	0,06986	0,00145

Tabela S2-Tabela de genes menos expressos (valores negativos de *fold change*) nas abelhas tratadas.

Gene	GeneSymbol	GeneDescription	Fold change	p-value
GB49219	LOC412162	armadillorepeat-containingproteingudu	-3,63231	0,00185
GB49888	LOC413908	cytochromeP4506A1	-3,01775	0,00005
GB52184	LOC100577819	uncharacterizedLOC100577819	-2,00988	0,00005
GB43738	PPO	phenoxidasesubunitA3	-1,86424	0,00005
GB41484	LOC724784	nuclearspecklesplicingregulatoryprotein1	-1,63928	0,00005
GB42170	LOC551646	myb-likeproteinX	-1,51602	0,0001
GB41850	LOC411597	hemocytin	-1,51317	0,00005
GB45382	LOC409358	intraflagellartransportprotein74homolog	-1,4016	0,00145
GB48267	LOC552409	Gpatchdomain-containingprotein4	-1,39644	0,0012
GB47037	LOC410925	hormonereceptor4	-1,38497	0,00005
GB47457	LOC412918	proteinSON-like	-1,35866	0,00005
GB47468	LOC725946	integrinalpha-8	-1,33104	0,00055
GB43562	LOC724593	transcriptionalregulatorATRXhomolog	-1,31916	0,00005
GB45479	LOC725707	centrosomalproteinof135kDa	-1,29683	0,0013
GB45397	LOC100579061	MATHandLRRdomain-containingproteinPFE0570w	-1,25926	0,0012
GB47269	LOC412832	structuralmaintenanceofchromosomesprotein2	-1,25534	0,00005
GB44427	LOC727450	putativeuncharacterizedproteinDDB_G0289041	-1,23035	0,00025
GB48999	LOC726729	helix-loop-helixprotein11	-1,22219	0,002
GB50779	LOC408712	BUD13homolog	-1,20979	0,00005

GB53066	LOC726080	uncharacterizedLOC726080	-1,19924	0,00015
GB41738	LOC410546	uncharacterizedLOC410546	-1,19624	0,0001
GB43968	<i>LOC100578089</i>	uncharacterizedproteinPF11_0213	-1,1828	0,00215
GB43508	LOC727193	lipasememberHA	-1,156286	0,00005
GB48777	LOC413758	U3smallnucleolarRNA-associatedprotein14homologA	-1,14723	0,0003
GB47719	LOC724948	proteinKRI1homolog	-1,14389	0,00005
GB53846	LOC724244	structuralmaintenanceofchromosomesprotein5	-1,13474	0,00005
GB44175	LOC408792	protein4.1homolog	-1,13237	0,00005
GB54593	LOC726262	adenylatecyclasetype8-like	-1,12493	0,00005
GB47543	LOC724831	uncharacterizedLOC724831	-1,12466	0,00005
GB40782	LOC412480	ribosomalRNAprocessingprotein1homolog	-1,120697	0,00005
GB47520	LOC725305	uncharacterizedLOC725305	-1,106655	0,0009
GB41852	LOC409243	A-kinaseanchorprotein9	-1,10289	0,00005
GB54570	LOC551520	golginsubfamilyAmember2	-1,10048	0,00005
GB55417	<i>LOC100577903</i>	uncharacterizedproteinMAL13P1.304	-1,06369	0,0005
GB52555	LOC412878	DNAN6-methyladeninedemethylase	-1,05891	0,00005
GB44285	LOC409107	homeoboxprotein5	-1,05128	0,00005
GB52457	LOC551711	kinesinA	-1,05061	0,0002
GB49758	<i>LOC100578691</i>	rhoGEFdomain-containingproteingxcJ	-1,04638	0,00035
GB55456	LOC726570	CWF19-likeprotein2	-1,02026	0,00005
GB51339	<i>LOC100576554</i>	cilia-andflagella-associatedprotein69	-1,01957	0,0008
GB51263	LOC725251	101kDamalariaantigen	-1,01346	0,0015
GB54992	<i>LOC100578360</i>	dnaJhomologsubfamilyCmember21	-1,01017	0,00075
GB53008	LOC408767	proteindachsous	-1,00299	0,00025
GB51649	LOC413418	structuralmaintenanceofchromosomesprotein4	-0,995756	0,0001
GB51385	LOC412858	hemicentin2	-0,990887	0,00005
GB52316	LOC409414	kinectin	-0,983064	0,00005
GB50114	LOC551562	dyneinbetachainciliary	-0,981988	0,00005
GB43904	LOC725556	intraflagellartransportprotein81homolog	-0,977709	0,00105
GB53910	LOC724240	flapendonucleaseGEN	-0,97671	0,00085
GB48530	LOC551566	uncharacterizedLOC551566	-0,976386	0,00025
GB44305	LOC552061	rhoassociatedproteinkinase2	-0,971723	0,00025
GB46971	<i>LOC100577649</i>	kinesin-likeproteinKlp61F	-0,958939	0,0001
GB46253	<i>LOC100577972</i>	uncharacterizedLOC100577972	-0,958828	0,00005
GB46449	LOC551406	cellgrowth-regulatingnucleolarprotein	-0,950601	0,00155
GB54286	LOC409395	MATHandLRRdomaincontainingproteinPFE0570w	-0,931548	0,00005
GB48948	<i>LOC100576242</i>	cilia-andflagella-associatedprotein91	-0,93137	0,0019
GB41428	Def1	defensin1	-0,924605	0,0004
GB55644	LOC409221	uncharacterizedLOC409221	-0,902653	0,00045
GB51422	LOC409463	WDrepeat-containingprotein63	-0,902327	0,00035
GB47628	LOC725662	uncharacterizedLOC725662	-0,898786	0,0001

GB40021	LOC725387	uncharacterizedproteinDDB_G0283357	-0,894998	0,0003
GB54198	LOC409902	liprin-alpha-1	-0,88887	0,0011
GB43771	<i>LOC100577739</i>	telomere-associatedproteinRIF1-like	-0,888509	0,0004
GB49398	LOC725656	uncharacterizedproteinDDB_G0288805	-0,887142	0,0002
GB48946	LOC726647	uncharacterizedLOC726647	-0,885591	0,0005
GB51271	LOC411953	proteinPeterpan	-0,883107	0,00165
GB50924	LOC413792	nucleoproteinTPR	-0,879351	0,00005
GB55909	LOC725992	uncharacterizedLOC725992	-0,878773	0,00195
GB41283	LOC408864	waprin-Phi1	-0,87215	0,00005
GB52612	LOC724119	ESF1homolog	-0,872004	0,00065
GB45395	LOC412131	kinesin4B	-0,859672	0,00055
GB46250	LOC727239	proteinP54-like	-0,856656	0,00015
GB43510	LOC551268	pancreatictriacylglycerollipase	-0,854123	0,00005
GB48821	LOC727192	structuralmaintenanceofchromosomesprotein6	-0,851989	0,00105
GB40885	PHRF1	PHDandringfingerdomains1	-0,843198	0,00025
GB47285	LOC408348	uncharacterizedLOC408348	-0,830178	0,0005
GB54812	LOC552498	DNArepairproteincomplementingXP-Ccellshomolog	-0,825176	0,00035
GB49601	LOC725269	proteinbarkbeetle	-0,815075	0,0019
GB52308	LOC724560	fattyacyl-CoAreductase1	-0,808137	0,00195
GB50375	LOC724992	RRP15likeprotein	-0,803409	0,00005
GB47470	LOC410009	cadherin-relatedtumorsuppressor	-0,797043	0,001
GB54923	<i>LOC100576132</i>	uncharacterizedLOC100576132	-0,772308	0,0001
GB51223	LOC406142	hymenoptaecin	-0,770337	0,0003
GB50795	LOC410893	transcriptionfactorAP-2-epsilon	-0,769617	0,00225
GB50012	LOC726323	uncharacterizedLOC726323	-0,761202	0,00015
GB52613	LOC725681	uncharacterizedLOC725681	-0,758432	0,0005
GB51706	<i>LOC100578450</i>	histone-lysineN-methyltransferaseSETD2	-0,757037	0,00055
GB47589	LOC409891	celldivisioncycleandapoptosisregulatorprotein1	-0,756523	0,0017
GB42021	LOC409456	AT-richinteractivedomain-containingprotein4B	-0,753976	0,00005
GB48854	LOC413947	sodiumleakchannelnon-selectiveprotein	-0,747464	0,00175
GB54042	LOC412606	ATPasefamilyAAAdomain-containingprotein5	-0,739305	0,0017
GB55264	LOC411910	homeoboxprotein2-like	-0,735549	0,00015
GB53948	LOC410057	uncharacterizedLOC410057	-0,73375	0,0004
GB52261	LOC552280	proteinSDA1homolog	-0,709791	0,0019
GB52568	LOC412688	ribosomebiogenesisproteinNOP53	-0,708924	0,001
GB47488	LOC551109	kinesin5A	-0,692129	0,00215
GB46282	LOC725406	serine/threonine-proteinkinasePRP4homolog	-0,690254	0,0021
GB52708	LOC725754	zincfingerproteinincastorhomolog1	-0,683374	0,00005
GB40808	LOC724634	golginsubfamilyBmember1	-0,682061	0,0004
GB55943	LOC551672	proteinFAM114A2	-0,677432	0,0004
GB52595	LOC551432	zincfingerandBTBdomain-containingprotein20	-0,676702	0,00165

GB44198	LOC408797	nidogen-2	-0,673804	0,0005
GB50951	LOC100576915	asparaginerichprotein	-0,67192	0,00165
GB51998	LOC413753	uncharacterizedLOC413753	-0,663797	0,00205
GB54649	LOC725503	uncharacterizedLOC725503	-0,663437	0,0008
GB50048	Mblk1	transcriptionfactormblk1like	-0,662998	0,0002
GB47831	LOC412227	proteinSPT2homolog	-0,66158	0,00045
GB49918	LOC724728	NFkappaBessentialmodulator	-0,659683	0,00005
GB49262	LOC100577043	uncharacterizedLOC100577043	-0,655163	0,0011
GB46363	LOC409847	15-hydroxyprostaglandinhydrogenase[NAD(+)]	-0,642722	0,0005
GB54622	LOC550695	carnitineOpalmitoyltransferase1,liverisoform	-0,640536	0,002
GB43173	LOC412273	probablechitinase10	-0,638907	0,00005
GB41911	LOC100578909	procollagen-lysine2-oxoglutarate5-dioxygenase3	-0,634317	0,00095
GB48006	LOC724803	neprilysin-4	-0,631233	0,0023
GB53062	LOC409111	serine/threonine-proteinkinaseinibrain	-0,631149	0,0017
GB51619	LOC410671	proto-oncogenetyrosine-proteinkinaseROS	-0,624519	0,0021
GB51545	LOC413968	dystrophinisoformsA/C/F/G/H	-0,620061	0,0023
GB52346	LOC724599	chromodomain-helicase-DNA-bindingprotein1	-0,614566	0,0016
GB51734	LOC412092	myosinheavychainnon-muscle	-0,610997	0,0015
GB49614	LOC725942	glutathioneStransferase11	-0,60267	0,00005
GB48846	LOC412399	organiccationtransporterprotein	-0,601736	0,00045
GB51055	LOC725155	serine/argininerepetitivematrixprotein1	-0,598479	0,0012
GB46618	LOC409474	lamininunitgamma-1	-0,572156	0,00185
GB54851	LOC727479	zinctransporter1	-0,565885	0,00215

## APÊNDICE B–Tutorial Para Instalação Em Uma Distribuição Linux Baseada Em Debian Dos Programas Utilizados Para O *Pipeline*

### INSTALAÇÃO DO FASTQC

Traduzido de <https://pbertinblog.wordpress.com/fastqc-installation/>

Linux: No menu de aplicativos, procure um aplicativo chamado '*Terminal*' ou '*Konsole*'.

Qualquer um deles fornecerá um *shell* utilizável.

No *prompt* de comando, digite '*java -version*' e pressione *Enter*. Você deve ver algo como:

```
java -version  
versão java "1.8.0_60"  
Java (TM) SE Runtime Environment (build 1.8.0_60-b27)  
Servidor VM Java HotSpot (TM) de 64 bits (versão 25.60-b23, modo misto)
```

Se receber um erro, você não tem o *java* instalado. Se a versão listada na primeira linha após o comando for menor que 1,6, então você pode ter problemas ao executar o FastQC.

Faça o *download* e descompacte a pasta (chamada FastQC) em

<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, você deve escolher um local para essa pasta. Por exemplo, no meu computador, a pasta está localizada em **/ Users / pbertin / Bioinformatics / Programs / FastQC .**

O executável FastQC está presente como «fastqc» na pasta baixada. Para poder usá-lo, você deve dar as permissões de execução no último arquivo com o comando:

```
chmod +x fastqc  
sudo apt install fastqc
```

E então use um link simbólico em um diretório PATH (consulte a página de explicações de comandos do *Linux* para saber mais sobre o PATH). Como uma observação, o PATH contém diretórios e os executáveis presentes nesses diretórios



podem ser executados de qualquer lugar em sua máquina.

```
sudo ln -s /Users/pbertin/Bioinformatics/Programs/FasQC/fastqc /usr/local/bin/fastqc
```

Você deve adaptar o comando com seu próprio caminho FastQC

## INSTALAÇÃO DO CUTADAPT

Traduzido de <https://cutadapt.readthedocs.io/en/v3.4/installation.html>

Cutadapt também está incluído em distribuições *Linux* baseadas em *Debian*, como *Ubuntu*. Para instalar o Cutadapt digite esta linha de comando

```
sudo apt install cutadapt
```

ou

```
sudo apt install python3-cutadapt
```

Esteja ciente de que os pacotes de distribuição provavelmente estão desatualizados. Se você encontrar um comportamento inesperado ou precisar de recursos mais novos, use um dos outros métodos de instalação disponíveis do site do desenvolvedor.

## INSTALAÇÃO DO BWA

Traduzido de <https://www.howtoinstall.me/ubuntu/18-04/bwa/>

O pacote de *software* bwa fornece Burrows-Wheeler Aligner,

```
sudo apt-get update  
sudo apt-get install bwa
```

bwa está instalado em seu sistema. Certifique-se de que o pacote bwa foi instalado usando os comandos fornecidos abaixo,

```
sudo dpkg-query -f | grep bwa *
```

Você obterá o nome do pacote bwa , versão, arquitetura e descrição em uma tabela

## INSTALAÇÃO DO SAMTOOLS

Traduzido de [http://cole-trapnell-lab.github.io/cufflinks/getting\\_started/#building-cufflinks](http://cole-trapnell-lab.github.io/cufflinks/getting_started/#building-cufflinks)

Abra um terminal e atualize seu sistema usando os seguintes comandos:

```
sudo apt-get update  
sudo apt-get upgrade
```

### Pré-requisitos de instalação

Para instalar o SAMtools, você precisa instalar algumas bibliotecas necessárias. Cole o seguinte comando no terminal.

```
sudo apt-get install -y libncurses-dev libbz2-dev liblzma-dev  
sudo apt-get update -y  
sudo apt-get install -y samtools
```

### Ou baixando SAMtools

Mude para o diretório onde deseja fazer o *download* do *software* (<https://sourceforge.net/projects/samtools/files/>) ou use o seguinte comando:

```
cd Downloads/  
wget https://github.com/samtools/samtools/releases/download/1.12/samtools-1.12.tar.bz2
```

### Instalando SAMtools

Estamos no mesmo diretório (ou seja, *Downloads*). Agora extraia o pacote baixado usando os seguintes comandos.

```
tar xvf samtools-1.12.tar.bz2
```

Agora, mude para o diretório recém-criado ( *samtools-1.12* ).

```
cd samtools-1.12/
```

Instale usando os seguintes comandos.

```
./configure  
make  
sudo make install
```

Exporte para o caminho usando o seguinte comando:

```
export PATH="$PATH:/home/user/Downloads/samtools-1.12"
```

Agora o adicione ao arquivo *bashrc* conforme mostrado abaixo.

Abrir arquivo *bashrc*

```
sudo gedit ~/.bashrc
```

Adicione a seguinte linha ao final do arquivo:

```
export PATH= "$PATH:/home/user/Downloads/samtools-1.12"
```

Salvar e sair. Depois disso, cole o seguinte comando no terminal:

```
source ~/.bashrc
```

Agora, o SAMtools foi instalado com sucesso em seu sistema. Você pode executá-lo digitando

```
samtools.
```

## INSTALAÇÃO DO CUFFLINKS

Traduzido de [http://cole-trapnell-lab.github.io/cufflinks/getting\\_started/#building-cufflinks](http://cole-trapnell-lab.github.io/cufflinks/getting_started/#building-cufflinks)

## Instalando uma versão binária pré-compilada

Para facilitar a instalação dos Cufflinks, fornecemos alguns pacotes binários para evitar que os usuários passem por processos ocasionalmente frustrantes de construção dos Cufflinks, que exigem a instalação das bibliotecas *Boost*. Para usar os pacotes binários, simplesmente baixe o arquivo para sua máquina, descompacte-o e certifique-se de que os binários cufflinks, cuffdiff e cuffcompare estão em um diretório em sua variável de ambiente PATH.

### Criação de cufflinks a partir da fonte

Para construir os Cufflinks, você deve ter as bibliotecas *Boost C ++* (versão 1.47 ou superior) instaladas em seu sistema. Veja abaixo as instruções sobre como instalar o *Boost*.

#### Instalando Boost

```
sudo apt-get install libboost-all-dev
```

1. Baixe *Boost* e o mecanismo de compilação bjam.
2. Descompacte bjam e adicione-o ao seu PATH.
3. Descompacte o tarball do *Boost* e cd para o diretório de origem do *Boost*. Este diretório é denominado *BOOST\_ROOT* em algumas instruções de instalação do *Boost*.
4. Build *Boost*. Observe que você pode especificar onde colocar o *Boost* com a opção `--prefix`. O diretório de instalação padrão do *Boost* é `/usr/local`. Anote o diretório de instalação do *Boost*, porque você precisará informar ao instalador do Cufflinks onde encontrar o *Boost* mais tarde.

Se você estiver no *Mac OS X*, digite (tudo em uma linha):

```
bjam --prefix=<SEU DIRETÓRIO> --toolset=darwin architecture=x86 address_model=32_64
link=static runtime-link=static --layout=versioned stage install
```

Se você estiver em um sistema *Linux* de 32 bits, digite :

```
bjam --prefix=<SEU DIRETÓRIO> --toolset=gcc architecture=x86 address_model=32 link=static
runtime-link=static stage install
```

Se você estiver em um sistema *Linux* de 64 bits, digite :

```
bjam --prefix=<SEU DIRETÓRIO> --toolset=gcc architecture=x86 address_model=64 link=static
runtime-link=static stage install
```

## Instalando as bibliotecas Eigen

5. Baixar Eigen
6. Descompacte o *tarball* *Eigen* e cd para o diretório de origem do *Eigen*.
7. Copie o subdiretório *Eigen* / em algum lugar do seu sistema onde você mantém os arquivos de cabeçalho (por exemplo, /usr/local/include)

## Construindo Cufflinks

Descompacte o *tarball* de origem do Cufflinks:

```
http://cole-trapnell-lab.github.io/cufflinks/assets/downloads/cufflinks-2.2.1.Linux_x86_64.tar.gz
tar zxvf cufflinks-0.7.0.tar.gz
```

Mude para o diretório Cufflinks:

```
cd cufflinks-0.7.0
```

Configure os Cufflinks. Se o *Boost* estiver instalado em algum lugar diferente de /usr/local, você precisará informar ao instalador onde encontrá-lo usando a opção `--with-boost`. Especifique onde instalar os Cufflinks usando a opção `--prefix`.

```
./configure --prefix=/path/to/cufflinks/install --with-boost=/path/to/boost --with-eigen=/path/to/eigen
```

Se você vir algum erro durante a configuração, verifique se está usando o *Boost*

versão 1.47 ou superior e se o diretório especificado por meio de *-with-boost* contém os arquivos e bibliotecas de cabeçalho do *boost*. Consulte a página de introdução do *Boost* para obter mais detalhes. Se você copiou os binários das ferramentas SAM para algum lugar diferente de `/usr/local/`, pode ser necessário fornecer a opção de configuração *-with-bam*. Finalmente, faça e instale os Cufflinks.

```
make
make install
```

## EXEMPLO DO PIPELINE

```
cutadapt -a ADAPTADOR -o C-01-1-cuta.fastq C01-1.fastq
cutadapt -a ADAPTADOR -o C-01-1-cutadapt.fastq C-01-1-cuta.fastq
cutadapt -m 10 -o C01-1_cutadapt.fastq C-01-1-cutadapt.fastq
bwa index Amel_4.5_scaffolds.fa
bwa aln -t 4 Amel_4.5_scaffolds.fa C01-1_cutadapt.fastq > Mapeamento_bwa_2.sai
bwa samse Amel_4.5_scaffolds.fa Mapeamento_bwa_2.sai C01-1_cutadapt.fastq > alinhado_com_bwa.sam
#=====
cp C1/alinhado_com_bwa.sam ../DEG/C1.sam
cp C2/alinhado_com_bwa.sam ../DEG/C2.sam
cp C3/alinhado_com_bwa.sam ../DEG/C3.sam
cp T1/alinhado_com_bwa.sam ../DEG/T1.sam
cp T2/alinhado_com_bwa.sam ../DEG/T2.sam
cp T3/alinhado_com_bwa.sam ../DEG/T3.sam
#=====
samtools view -bS C1.sam > C01.bam
samtools sort -o cufflink/C01_sort.bam C01.bam
samtools view -bS C2.sam > C02.bam
samtools sort -o cufflink/C02_sort.bam C02.bam
```

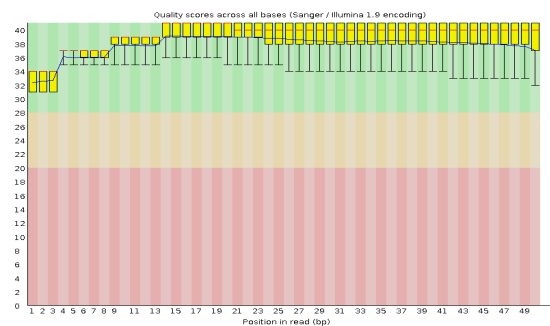
```
samtools view -bS C3.sam > C03.bam  
samtools sort -o cufflink/C03_sort.bam C03.bam  
samtools view -bS T1.sam > T01.bam  
samtools sort -o cufflink/T01_sort.bam T01.bam  
samtools view -bS T2.sam > T02.bam  
samtools sort -o cufflink/T02_sort.bam T02.bam  
samtools view -bS T3.sam > T03.bam  
samtools sort -o cufflink/T03_sort.bam TO3.bam  
  
#=====
```

```
cuffdiff -labels control,dsJHE -output-dir cufflink/dados ../data/baixar/my.gtf  
cufflink/C01_sort.bam,cufflink/C02_sort.bam,cufflink/C03_sort.bam  
cufflink/T01_sort.bam,cufflink/T02_sort.bam,cufflink/T03_sort.bam
```

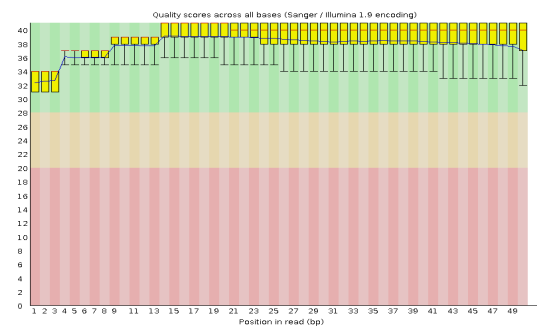
## APÊNDICE C—Análise FASTQC

FIGURAS1-Distribuição dos valores de qualidade (*PhredScore*) ao longo do comprimento das *reads*.

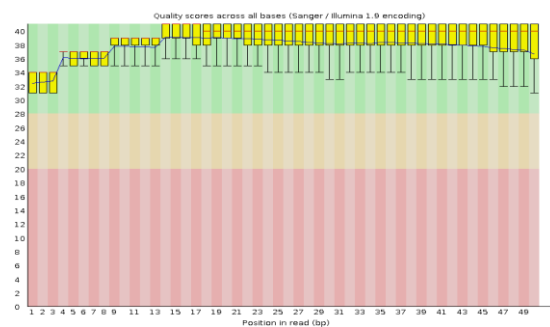
A



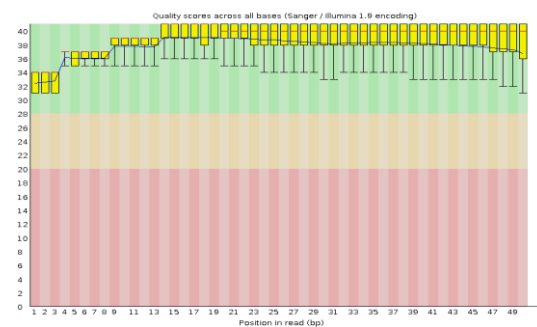
B



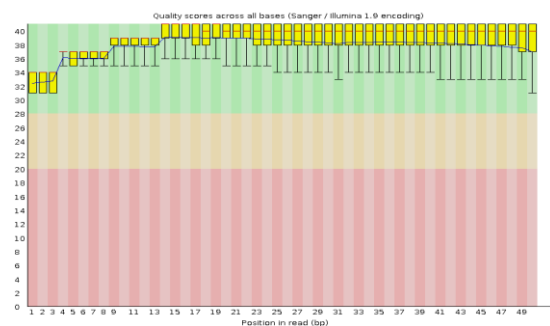
C



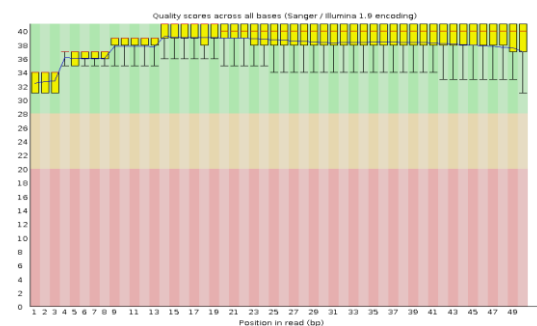
D



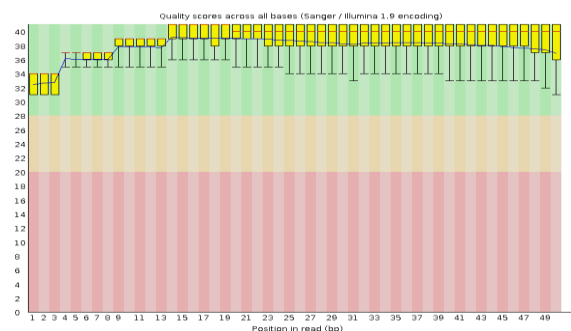
E



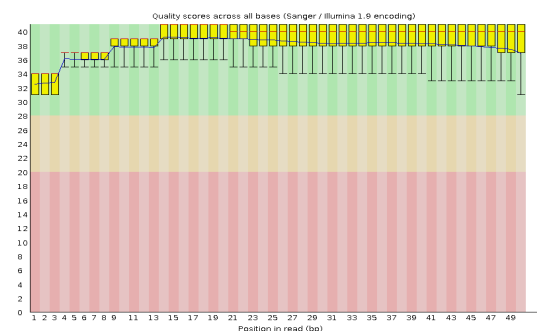
F



G

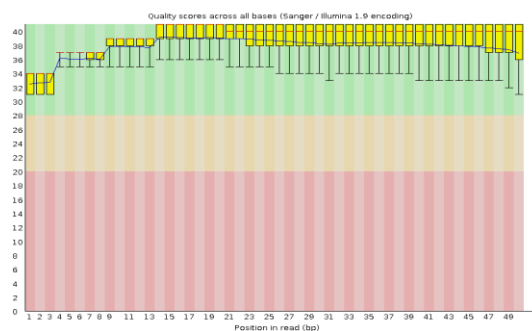


H

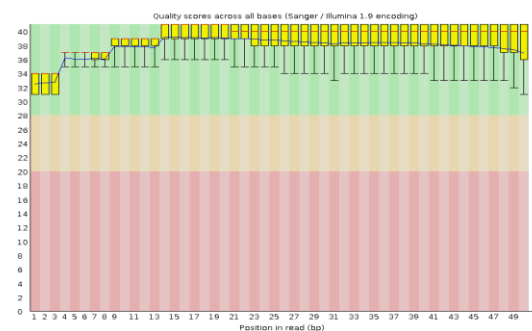




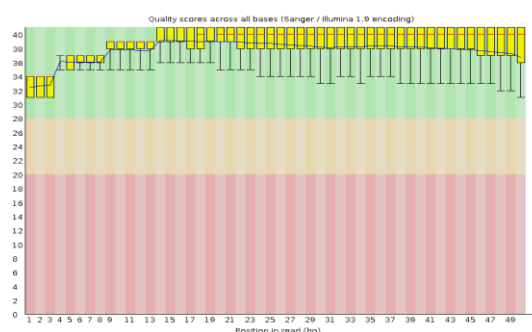
I



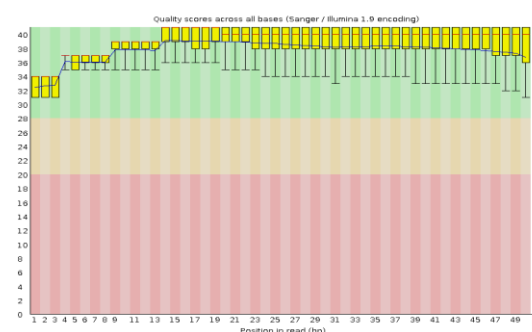
J



K



L

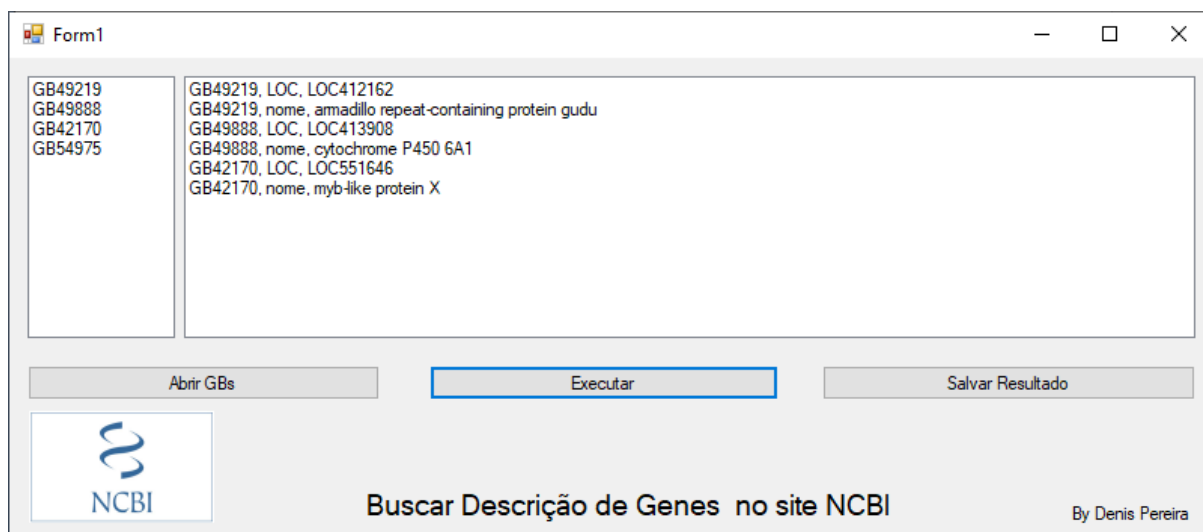


Fonte: Autor.

Nota: As imagens A e B correspondem às bibliotecas controle “C1” antes (dados brutos) e depois de processadas as imagens C e D as biblioteca controle “C2” antes (dados brutos) e depois de processadas, as imagens E e F às biblioteca controle “C3” antes (dados brutos) e depois de processadas, as imagens G e H às biblioteca controle “T1” antes (dados brutos) e depois de processadas, as imagens I e J às biblioteca controle “T2” antes (dados brutos) e depois de processadas, as imagens K e L às biblioteca controle “T3” antes (dados brutos) e depois de processadas, o eixo X representa os pares de base de todas as *reads* que foram analisados e o eixo Y representa os índices de qualidade *Phred* das *reads* analisadas pelo programa FASTQC, os *boxplots* em amarelo apresentam o intervalo inter-quartil (25% a 75%) da distribuição dos valores de qualidade e a linha vermelha aponta para o valor da mediana e a linha azul para a média. Em todas as bibliotecas, a distribuição dos valores de qualidade ao longo das *reads* revelou alta qualidade do esquecimento na identificação de cada base.

## APÊNDICE D–Software C#

FIGURA S2 – Imagem de programa desenvolvido em linguagem de programação C#



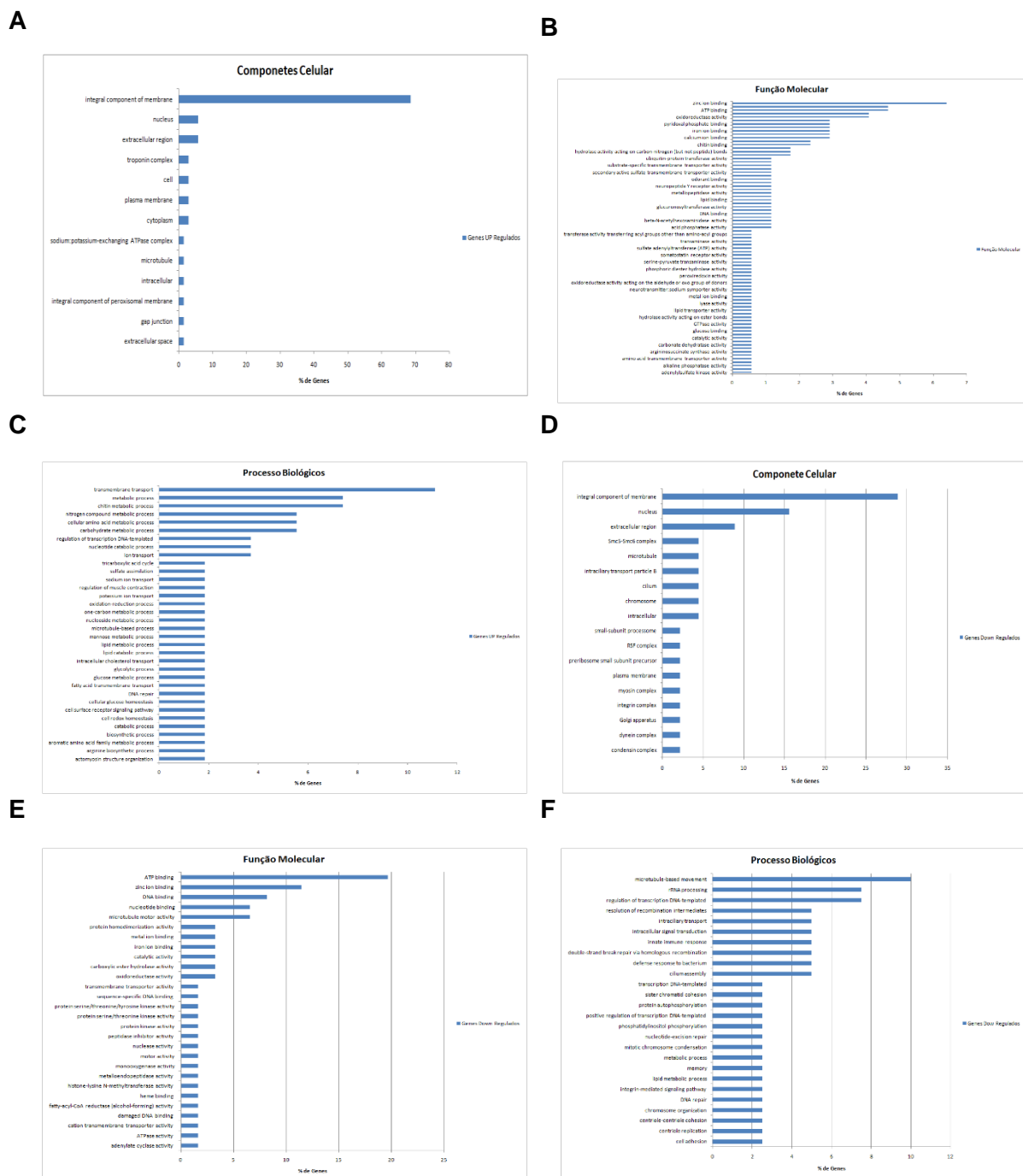
Fonte: Autor.

Nota: Programa desenvolvido em linguagem de programação C# para analisar dentro do site NCBI de informações referentes aos GBs do genes, na caixa da esquerda contem uma lista de todos os GBs de genes lidos o arquivo de texto, e na caixa da direita o resultado da pesquisa, o programa conta com três botões, um para abrir a lista de GBs, o botão para executar e outro para salvar os dados encontrados, o progama está disponível para baixar em:

(<https://mega.nz/file/2boUQDYD#CvuLDpXxjneoy0h54TVG2a3wz9ftQ74MiTBrZBLgkHA>)

## APÊNDICE E– Análise de enriquecimento funcional

### FIGURA S3–Análise de enriquecimento funcional.



Fonte: Autor.

Nota: Os gráficos exibem o percentual de expressão dos dois grupos (controle sendo *Down* reguladas para o grupo *KD* para jhe e tratadas sendo *Up* reguladas para o grupo *KD* para jhe), mostrando as categorias de componentes celulares em A e D, funções moleculares em B e E e processos biológicos em C e F. O eixo X representa o percentual de genes e o eixo y as categorias correspondentes.