

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Biológicas e da Saúde
Programa de Pós-graduação em Genética Evolutiva e Biologia Molecular

Leonardo do Amaral Martins

**MONTAGEM E ANÁLISE TAXONÔMICA-FUNCIONAL DE TRANSCRIPTOMAS DE
ESPÉCIES DE CUPINS BRASILEIROS**

São Carlos - SP

2020

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Biológicas e da Saúde
Programa de Pós-graduação em Genética Evolutiva e Biologia Molecular

Leonardo do Amaral Martins

**MONTAGEM E ANÁLISE TAXONÔMICA-FUNCIONAL DE TRANSCRIPTOMAS DE
ESPÉCIES DE CUPINS BRASILEIROS**

Orientador: Prof. Dr. Caio César de Melo Freire

Dissertação apresentada ao Programa de Pós-graduação em Genética Evolutiva e Biologia Molecular, área de concentração Bioquímica e Biologia Molecular, da UFSCar, como parte dos requisitos para obtenção do título de Mestre.

São Carlos – SP

2020



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Leonardo do Amaral Martins, realizada em 28/02/2020:

Prof. Dr. Caio Cesar de Melo Freire
UFSCar

Profa. Dra. Patricia Domingues de Freitas
UFSCar

Prof. Dr. Daniel Ferreira de Lima Neto
USP

AGRADECIMENTOS

Agradeço, primeiramente, aos meus pais, por terem sempre me amado e cobrado o melhor de mim.

Agradeço meus grandes irmãos de São José do Rio Preto. Aqueles com quem vivi a maior parte da minha vida, e que estão presentes até hoje nela. Nicolas, Pedro, João, Virso e Jorge. Mesmo com a distância, nossos laços nunca perderam as forças. Vocês foram vitais para moldar quem sou hoje.

Agradeço os amigos que fiz durante toda essa jornada em São Carlos. Em especial, Peão, Marília, Tiomat, Fineco, Minnie, Cleitinho, Mineiro e Briza. Vocês são a família que a UFSCar me deu e provaram que não é necessário grandes períodos de tempo para que fortes laços se formem.

Agradeço à banca por terem aceitado o convite para avaliação deste trabalho.

Agradeço a todo grupo de estudo responsável pelo caminhar deste trabalho. Em especial, obrigado Caio por nunca desistir de mim, obrigado Anderson e Francis por todo o respeito que sempre tiveram, obrigado Matheus por estar sempre disposto a me ajudar, mesmo não estando trabalhando conosco mais e, por fim, obrigado Emeline por ser essa pessoa que uniu nosso grupo e agarrou essa pesquisa com todas as forças. Me desculpem se os desapontei em algum momento. Espero poder passar o melhor de cada um adiante, e ser um exemplo para outros, assim como vocês foram para mim.

Lastly, I thank my future wife, Bri. You made me happy every single day of this journey, even the ones I thought I couldn't survive. You were the real reason I kept going on.

Muito obrigado.

RESUMO

Cupins (Isoptera) são insetos eusociais que se destacam por sua organização hierárquica e estrutural, demonstrando comportamentos sociais complexos, numerosas interações simbióticas, e um grande potencial para bioprospecção de enzimas de interesse industrial. Além disso, também são frequentes pragas de monoculturas agrícolas brasileiras. Apesar da importância desta espécie, estudos moleculares sobre os cupins até o momento ainda não foram suficientes para que os mecanismos que governam a plasticidade fenotípica nesses insetos tenham sido completamente elucidados. Neste estudo exploratório, realizamos uma análise comparativa inter e intraespecífica entre transcriptomas de cupins das espécies brasileiras *Heterotermes tenuis*, *Cornitermes cumulans* e *Velocitermes heteropterus*, assim como um levantamento da biodiversidade de suas microbiotas endossimbiontes e prospecção de genes com potencial biotecnológico. Foram encontradas importantes diferenças entre a espécie basal e as superiores quando comparadas quanto a sua diversidade taxonômica, incluindo a surpreendente detecção de simbiontes flagelados nas espécies superiores, com destaque para a grande presença do filo Amoebozoa. Nossa análise de expressão diferencial mostrou que a maioria dos genes anotados eram únicos de uma determinada casta ou espécie, e sua anotação funcional evidenciou diferenças nos processos biológicos, principalmente entre as castas das espécies superiores. Entre os genes mais diferencialmente expressos de cada espécie, encontramos genes com potencial para uso no manejo de cupins como pragas e para aplicações industriais, e.g., hexamerinas, citocromo P450 e celulases diversas. Os resultados gerados e analisados neste estudo avançam os conhecimentos sobre a plasticidade observada nas castas e abrem perspectivas para possíveis aplicações biotecnológicas.

Palavras-chave: Cupim, microbiota, bioinformática, transcriptoma, metatranscriptoma, expressão diferencial.

ABSTRACT

Termites (Isoptera) are eusocial insects that stand out for their hierarchical and structural organization, demonstrating complex social behaviors, numerous symbiotic interactions, and a great potential for bioprospecting enzymes of industrial interest. In addition, they are also frequent pests of Brazilian agricultural monocultures. Despite the importance of this species, molecular studies on termites have not yet been sufficient so that the mechanisms that control phenotypic plasticity in these insects have been completely elucidated. In this exploratory study, we performed a comparative inter and intraspecific analysis between termite transcriptomes of the Brazilian species *Heterotermes tenuis*, *Cornitermes cumulans* and *Velocitermes heteropterus*, as well as a biodiversity survey of their endosymbiotic microbiota and prospecting of genes with biotechnological potential. Important differences were found between the lower and the higher species when compared in terms of their taxonomic diversity, including the surprising detection of flagellate symbionts in the higher species, with emphasis on the great presence of the phylum Amoebozoa. Our analysis of differential expression showed that the majority of the annotated genes were unique to a certain caste or species, and their functional annotation showed differences in biological processes, especially among the castes of higher species. Among the most differentially expressed genes of each species, we found genes with potential for use in the management of termites as pests and for industrial applications, e.g., hexamerins, cytochrome P450 and many cellulases. Together, these results enhance the knowledge about the plasticity observed through the castes and open new perspectives for future biotechnological applications.

Keywords: Termites, microbiota, bioinformatics, transcriptome, metatranscriptome, differential expression.

ÍNDICE DE FIGURAS

Figura 1. Principais hipóteses das relações filogenéticas da infra ordem Isoptera.	2
Figura 2. Exemplo das diferentes castas de cupins.....	3
Figura 3. Vias de desenvolvimento mais comumente relatadas em cupins (A) basais e (B) superiores.	5
Figura 4. Componentes essenciais na digestão de cupins.	7
Figura 5. Exemplos de <i>Heterotermes tenuis</i>	9
Figura 6. Exemplos de <i>Cornitermes cumulans</i>	11
Figura 7. Exemplos de <i>Velocitermes</i> spp.	12
Figura 8. Design experimental do presente estudo.	20
Figura 9. Panorama metatranscriptômico de Ht.	27
Figura 10. Panorama metatranscriptômico de Cc e Vh.	28
Figura 11. Distribuição dos valores de expressão (TPM) pela abundância dos contigs.....	33
Figura 12. Heatmaps e agrupamentos hierárquicos dos DEGs.	35
Figura 13. Relação da classificação de GO para os 20 genes mais diferencialmente expressos de cada casta.	36
Figura 14. Comparação entre castas e espécies de genes comuns e únicos anotadas a partir dos DEGs.	37
Figura 15. Diagrama representando rede de genes e fatores responsivos ao JH.	39

ÍNDICE DE TABELAS

Tabela 1. Descrição das montagens de novo por castas.....	21
Tabela 2. Descrição das três montagens de novo por espécies.	22
Tabela 3. Resumo das análises de read content para as montagens por castas.....	23
Tabela 4. Resumo das análises de read content para as montagens por espécies.	24
Tabela 5. Resultados de completude para as seis montagens por castas.	25
Tabela 6. Resultados de completude para as três montagens por espécies.	25
Tabela 7. Grupos majoritários de endossimbiontes protistas presentes nos cupins estudados.	30
Tabela 8. Filos de endossimbiontes bacterianos presentes nos cupins estudados.....	31
Tabela 9. Resumo da anotação funcional dos DEGs realizada através da ferramenta BLAST2GO.....	34

SUMÁRIO

AGRADECIMENTOS.....	iii
RESUMO.....	iv
ABSTRACT	v
ÍNDICE DE FIGURAS.....	vi
ÍNDICE DE TABELAS.....	vii
SUMÁRIO.....	viii
1 INTRODUÇÃO.....	1
1.1 Cupins e sua complexidade.....	1
1.2 Cupins e seus endossimbiontes	6
1.3 Cupins como inseto-praga e fonte de bioprospecção de genes.....	7
1.4 A espécie <i>Heterotermes tenuis</i>	9
1.5 A espécie <i>Cornitermes cumulans</i>	10
1.6 A espécie <i>Velocitermes heteropterus</i>	11
1.7 Introdução a bioinformática e ciências ômicas.....	12
2 OBJETIVOS	15
2.1 Objetivo geral	15
2.2 Objetivos específicos.....	15
3 METODOLOGIA.....	16
3.1 Obtenção dos insetos e RNA-seq.....	16
3.2 Pré-processamento e montagem do transcriptoma.....	16
3.3 Anotação taxonômica dos transcriptomas de cada casta.....	17
3.4 Análise da expressão diferencial	18
3.5 Anotação taxonômica-funcional dos DEGs.....	19
4 RESULTADOS E DISCUSSÃO	21
4.1 Montagens, avaliação da qualidade e predição das ORFs	21
4.2 Panorama metatranscriptômico de operários e soldados de cupins basais e superiores	27
4.3 Padrões de expressão gênica e anotação funcional dos DEGs.....	33

4.4	Genes ligados à diferenciação de castas e genes com potencial biotecnológico...	38
5	CONCLUSÃO	41
6	REFERÊNCIAS BIBLIOGRÁFICAS.....	42
	APÊNDICE 1	55
	APÊNDICE 2	58

1 INTRODUÇÃO

Cupins são insetos extremamente organizados hierárquica e estruturalmente. Eles demonstram comportamentos sociais complexos, diferenciação de castas e uma classificação filogenética que ainda não é clara e que sofre mudanças até os dias atuais. Ademais, são espécies que realizam diversas interações simbióticas e possuem um grande potencial para bioprospecção de enzimas de interesse biotecnológico. Por outro lado, com o crescimento da monocultura agrícola, também se tornaram praga de diversos tipos de lavouras, além de causar problemas em moradias urbanas.

Mesmo com tantos atrativos, o número de estudos sobre estes artrópodes ainda se mostra baixo. Utilizando-se das tecnologias de sequenciamento de alta cobertura, processamento e análise de dados e novas ferramentas de Bioinformática, somadas ao grande potencial biotecnológico escondido nesses organismos, a busca por novos meios de combate ao inseto-praga e de lacunas no conhecimento sobre sua biologia e genética molecular, o presente estudo visou montar e analisar os transcriptomas de três espécies neotropicais de cupins brasileiros.

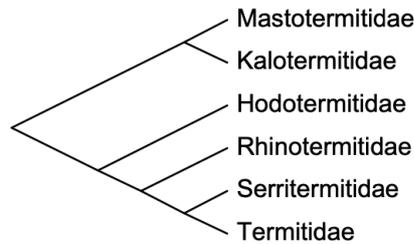
1.1 Cupins e sua complexidade

Cupins são insetos eusociais, hemimetábolos e que vivem em complexas colônias que podem ser modeladas como "superorganismos", onde os indivíduos fazem parte de uma entidade autorreguladora maior (WILSON, 1992). Eles vivem predominantemente em regiões tropicais, onde são considerados os insetos decompositores mais importantes (DAVIES et al., 2003).

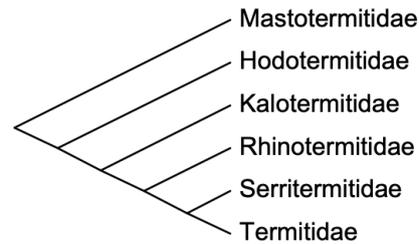
São facilmente distinguíveis de outros grupos de insetos pois exibem um sistema de castas dentro das quais, possuem características morfológicas específicas. Em Biologia, casta se refere a grupos comportamentais especializados criados a partir da divisão do trabalho dentro de uma sociedade animal. Todos os cupins são classificados como insetos eusociais, o que significa que eles possuem divisão de tarefas, organização hierárquica e vivem em colônias, como uma verdadeira comunidade (WILSON, 1971). Apesar da distinção morfológica, a posição filogenética dos cupins no grupo de artrópodes é objeto de debate. A opinião da maioria da comunidade científica especializada é que os cupins fazem parte da infra ordem Isoptera, a qual faz parte da ordem das baratas, Blattodea (INWARD; BECCALONI; EGGLETON, 2007). Ainda assim, dentro da sua infra ordem, os cupins vinham sendo rearranjados constantemente, o que gerava grande confusão até os anos 2000 (Figura 1) e continua abrindo discussões até os dias atuais (LEGENDRE et al., 2008; LEGENDRE;

GRANDCOLAS, 2018; WU et al., 2018). Existem nove famílias, 292 gêneros e aproximadamente 3.000 espécies descritas e estima-se entre 500 a 1.000 ainda não catalogadas (ENGEL; GRIMALDI; KRISHNA, 2009).

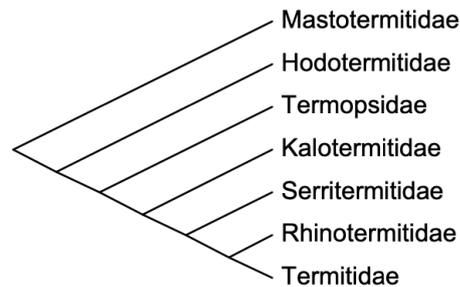
A) Ahmad, 1950 – Krishna, 1970



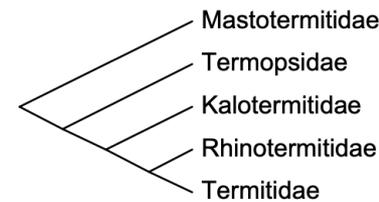
B) Emerson e Krishna, 1975



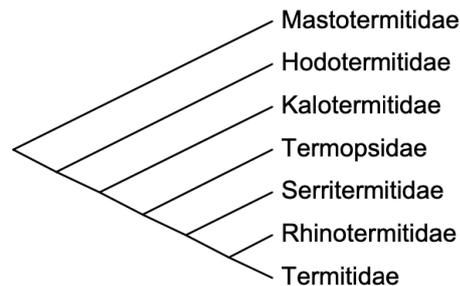
C) Donovan et al., 2000



D) Kambhampati et al., 1996



E) Kambhampati e Eggleton, 2000



F) Thompson et al., 2000

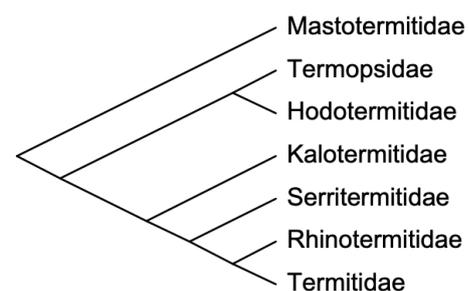


Figura 1. Principais hipóteses das relações filogenéticas da infra ordem Isoptera. A) a F) indicam cada uma das diferentes hipóteses e suas respectivas publicações. Fonte: adaptado de LEGENDRE et al., 2008.

Uma colônia de cupins é uma família em que vivem todos juntos. Ela geralmente é composta por uma parte viva, os cupins, e uma parte inanimada, as estruturas construídas por eles, dentro das quais vivem. Segundo EGGLETON, 2010, a parte viva de uma colônia apresenta indivíduos jovens (imaturos) e tipicamente duas categorias de castas principais: os reprodutores (rainha, rei e alados) e os auxiliares (operários e soldados), ambos representados na Figura 2. A rainha é geralmente o único indivíduo ovopositor da colônia. O

rei é o companheiro da rainha, e sua única tarefa é de se acasalar regularmente com ela, mantendo a taxa de natalidade da colônia estável (KORB, 2008). Os alados são reprodutores dotados de asas que estão se preparando para deixar a colônia, a fim de fazerem enxames, procriar e iniciar novas colônias. Operários, por outro lado, nunca deixam a colônia, exceto para forragearem, garantindo a alimentação, além de serem cegos (pela falta de olhos compostos) em sua grande maioria. Eles são considerados o pilar da colônia e seus papéis são numerosos: *(i)* procuram comida e água (forrageio), *(ii)* constroem e consertam estruturas das colônias, *(iii)* alimentam as demais castas e, em muitas espécies, *(iv)* podem se transformar em reprodutores suplementares no caso da morte dos primários, também conhecidos como reprodutores neotênicos (ROISIN; KORB, 2010). Os soldados têm apenas um trabalho: defender a colônia, em especial a rainha e o rei. Eles geralmente têm cabeças grandes, altamente escleróticas e mandíbulas poderosas, altamente modificadas. Muitas vezes também apresentam defesas químicas, como substâncias irritantes (PRESTWICH, 1984). Essa descrição simples da estrutura das castas se torna mais complexa quando designada à algumas espécies que fogem dessa estrutura hierárquica, no entanto, todas as espécies descritas até hoje possuem pelo menos uma casta estéril pré-determinada durante os estágios de imaturos (BOOMSMA, 2009).

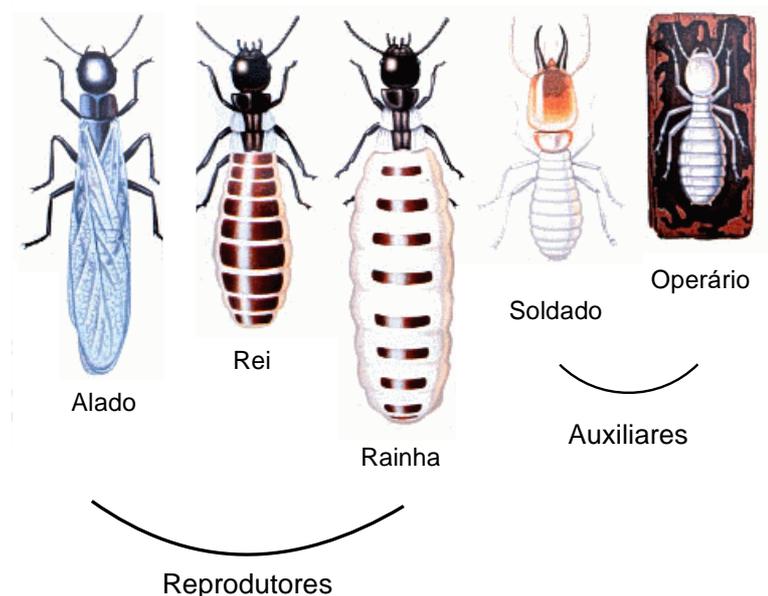


Figura 2. Exemplo das diferentes castas de cupins. Fonte: adaptado de <https://www.doityourselftermitecontrol.com/castes.gif>. Acesso em: 28 de janeiro de 2020.

De acordo com seus hábitos alimentares e sua microbiota endossimbionte, os cupins foram classificados em dois grupos: *(i)* cupins basais, predominantemente alimentam-se de madeira, além de interagir simbioticamente com diversos protistas flagelados, bactérias e arqueias em seu intestino; *(ii)* cupins superiores, os quais consomem uma ampla variedade

de materiais, e.g., madeira, serrapilheira, excremento animal, húmus, matéria orgânica do solo, possuindo apenas espécies de bactérias e arqueias, aparentemente sem protistas flagelados (BREZNAK; BRUNE, 1994; RADEK; TISCHENDORF, 1999; WATANABE; TOKUDA, 2001). O sistema digestório dos cupins é formado por estruturas complexas. Nas famílias basais, ele é muito semelhante ao de baratas, enquanto nas famílias de cupins superiores, é extremamente modificado. Outrora, a digestão da madeira ingerida costumava ser atribuída apenas aos simbiontes, no entanto, pesquisas ao longo das últimas décadas demonstraram que os cupins superiores produzem parte das celulasas de maneira endógena para digerirem a madeira eficientemente (LO; TOKUDA; WATANABE, 2010). Portanto, a digestão geral é dividida entre a produção de celulasas próprias no intestino médio e a fermentação microbiana no intestino grosso.

O ciclo de vida é semelhante na maioria dos cupins, porém o desenvolvimento das castas diverge entre famílias e gera novas descobertas até os dias atuais. As colônias produzem reprodutores com asas (alados), muitas vezes no início da estação chuvosa em habitats mais secos ou sazonais, mas durante todo o ano em habitats mais úmidos e estacionais (GOMES; GOMES; MARTIUS, 1999). Estes reprodutores pousam no chão ou em um pedaço de madeira morta e acasalam. Os pares procuram então uma colônia já estabelecida, um novo solo ou madeira morta. Após a cópula, a rainha coloca os primeiros ovos. O desenvolvimento embrionário dos novos indivíduos pode ser longo, variando entre 24 e 90 dias em algumas espécies. Cupins são considerados insetos hemimetábolos, ou seja, realizam metamorfose incompleta, e seu desenvolvimento depende muito da sua família, ou até mesmo espécie. Cupins basais têm estágios imaturos totipotentes que podem explorar todas as opções de castas, os operários são temporariamente “presos” juvenis que mantêm a plasticidade do desenvolvimento, tendo a capacidade de se diferenciar em soldados desenvolvidos, alados ou reprodutores neotênicos. Esta via de desenvolvimento de castas é conhecida como linear. Por outro lado, cupins superiores têm uma via bifurcante, que se divide em uma linha ninfal que leva a alados e uma linha áptera que leva a operários e soldados, que geralmente é estabelecida no estágio da larva do ovo ou primeiro ínstar (KORB; BELLES, 2017; SCHARF, 2014). Estas duas vias mais presentes no desenvolvimento dos cupins estão melhor esquematizadas abaixo, na Figura 3. Os primeiros indivíduos a se diferenciarem costumam se transformar em operários, que, por sua vez, começam a cuidar dos jovens, construir estruturas do cupinzeiro e forragearem. Um pouco mais tarde, no desenvolvimento da colônia, são gerados os soldados (EGGLETON, 2010). Quando a colônia atinge certa estabilidade, alados são produzidos novamente, e o ciclo continua.

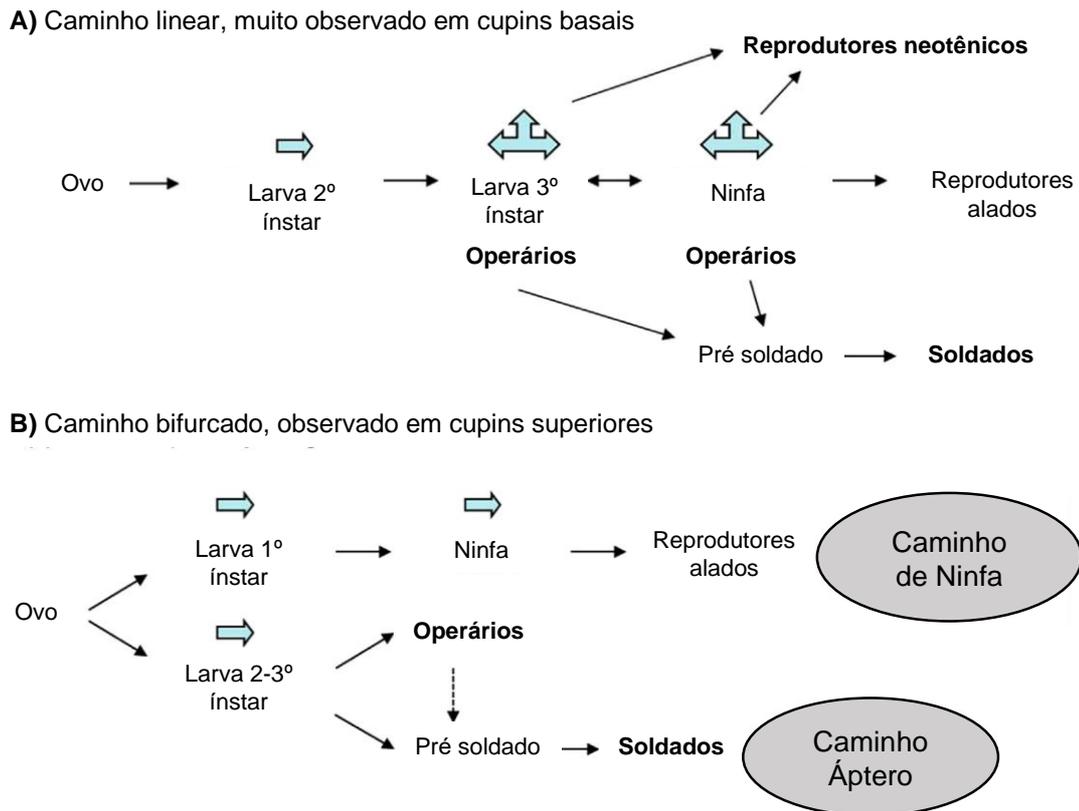


Figura 3. Vias de desenvolvimento mais comumente relatadas em cupins (A) basais e (B) superiores. Fonte: adaptado de KORB; BELLES, 2017.

Com base no desenvolvimento compartilhado de operários e soldados, o que inclui também sua função não-reprodutora dentro de uma típica colônia de cupins, seria natural de se esperar que as duas castas auxiliares tivessem boa parte de seus perfis de expressão gênica sobrepostos. Alternativamente, como já introduzido acima, os soldados de cupins são comportamental e morfologicamente distintos dos operários (MIURA; SCHARF, 2010). O estudo de WU; DHAMI; THOMPSON, 2018 mostra que cupins soldados têm uma expressão gênica fortemente enviesada implicando um grau pronunciado de especialização funcional, diminuindo afinidades funcionais ou de desenvolvimento para a casta operária.

Como observado acima, os mecanismos que governam a diferenciação das castas operário e soldado ainda não foram totalmente elucidados e os perfis de expressão gênica de ambas as castas auxiliares parecem ser divergentes. Assim, muitos estudos tentam relacionar os padrões de expressão gênica observados com os processos de desenvolvimento ou evolução que originam cada casta (BEHL et al., 2018; CORONA; LIBBRECHT; WHEELER, 2016; JONGEPIER et al., 2018). Embasados em outros trabalhos já publicados sabemos, por exemplo, que vias metabólicas dependente de TGF- β (MASUOKA et al., 2018), hormônio juvenil (TOGA et al., 2016) e insulina (HATTORI et al., 2013) estão diferencialmente expressas entre essas castas de uma maneira significativa. De acordo com estudos genômicos sabemos que, até o atual momento, o tamanho de um genoma de cupim pode variar desde os 500 Mb

(mega bases) a mais que 1,3 Gb (giga bases), não havendo um padrão consolidado para genomas de cupins basais ou superiores, e possuir por volta de 15.000 genes codificadores de proteínas (HARRISON et al., 2018; KORB et al., 2015). Porém, esse valor médio não considera amostras de corpo inteiro que incluem o trato digestório dos cupins. Assim, propulsionados pela sede do saber e à luz da ciência, inserimos aqui nosso estudo, com dados inéditos de transcriptomas de espécies brasileiras de cupins.

1.2 Cupins e seus endossimbiontes

Entre todos os organismos que degradam celulose, os cupins são os mais eficientes (BIGNELL; EGGLETON, 2000; ULYSHEN, 2016). Como outros animais que prosperam com uma dieta rica em fibras, os cupins cooperam com um conjunto denso de microrganismos que contribuem com funções que faltam no hospedeiro. A simbiose entre cupins e flagelados é antiga, desde a origem da família Isoptera, e estabilizou-se, em grande parte, pelo desenvolvimento da trofalaxia proctodeal: um comportamento que consiste em trocas de alimento pré-digerido do ânus para a boca, herdado desde os ancestrais comuns entre cupins e baratas, e que é considerado um dos motivos que permitiram a evolução do comportamento social dos cupins (KLASS; NALEPA; LO, 2008; MIRABITO; ROSENGAUS, 2016; NALEPA, 2015). Com o surgimento de genes endógenos de enzimas com capacidade celulolítica e o aparente desaparecimento dos protistas de sua microbiota, o trato digestório dos cupins foi populado por procariotos, gerando uma diversificação de sua dieta (BIGNELL; EGGLETON, 2000).

Esses endossimbiontes são representados por uma alta diversidade de táxons microbianos, provenientes dos três domínios: Archaea, Eubactéria e Eukaryota (representados por protistas flagelados). Eles atuam diretamente no metabolismo dos polímeros de carboidratos, *e.g.*, celulose, hemicelulose, bem como na degradação de compostos aromáticos originados de lignina, no metabolismo de hidrogênio, em sua conversão à acetato, no consumo de oxigênio por bactérias aeróbias (e aeróbias facultativas) e, por fim, servindo de biomassa para suprir a demanda de nitrogênio de seu hospedeiro (BRUNE; DIETRICH, 2015).

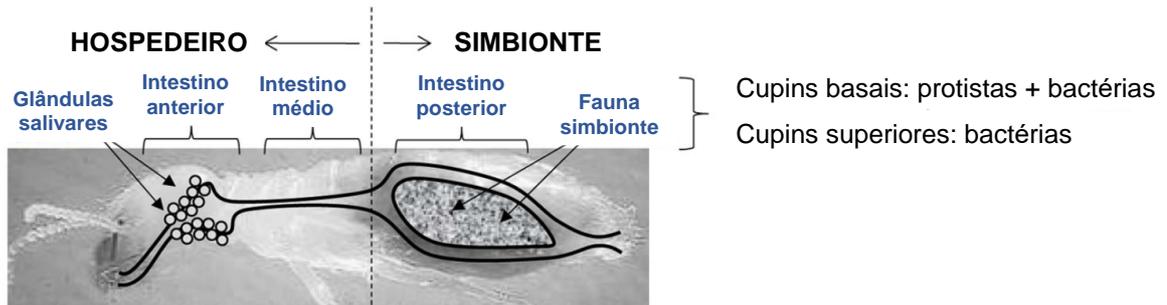


Figura 4. Componentes essenciais na digestão de cupins. Diferentes partes do intestino de cupins realizam diferentes processos de digestão e são muito estudadas para melhor elucidar o papel dos simbioses neste processo. Fonte: adaptado de SCHARF, 2015b.

O trato digestório dos cupins são pequenos ecossistemas com uma ampla gama de micro-habitats que diferem fortemente em seu ambiente abiótico e biótico. Muitas das características ambientais são propriedades intrínsecas do intestino, enquanto outras resultam de atividades fisiológicas do hospedeiro ou dos residentes microbianos no respectivo local (BRUNE; DIETRICH, 2015). Cupins possuem microrganismos ao longo de todo o seu trato digestório, porém a região que realmente se destaca pela microbiota simbiote é a do intestino posterior, como visto na Figura 4 acima. Nela, existe uma menor diversidade de enzimas digestivas do hospedeiro e um maior e mais demorado acúmulo de alimentos, o que beneficia o aumento da comunidade microbiana nessa região.

Enquanto os flagelados são facilmente identificados com base em suas características morfológicas, a caracterização da maioria das comunidades bacteriana e arqueal pode ser realizada com precisão apenas por sequenciamento genético. Combinando dados genômicos, metagenômicos e metatranscriptômicos, estudos vêm mostrando potenciais mudanças da composição simbiote entre castas (POULSEN et al., 2014), além de revelarem novas informações sobre a diversidade de celulase de cupins com diferentes estratégias de simbiose (LIU et al., 2013), de hábitos alimentares (HE et al., 2013), e de funções e distribuição de linhagens menos cultivadas (ŠOBOTNÍK; DAHLSJÖ, 2017).

1.3 Cupins como inseto-praga e fonte de bioprospecção de genes

As espécies de cupins são importantes por razões ecológicas e econômicas. São grandes animais decompositores, mas também, potentes pragas globais da agricultura e estruturas humanas (ROULAND-LEFÈVRE, 2010; SU; SCHEFFRAHN, 2000), causando perdas econômicas a várias décadas (VERMA; SHARMA; PRASAD, 2009) e sendo frequentemente introduzidos em novas áreas onde se tornam invasivos.

Os quatro principais tipos de termiticidas em uso atualmente são pesticidas líquidos de solo, iscas atrativas, tratamentos de madeira e aplicações pontuais (ROULAND-LEFÈVRE,

2010; SCHARF, 2014). Além destes, a utilização de fungos entomopatogênicos (*Beauveria bassiana* e *Metarhizium anisopliae*) diretamente nas colônias também se mostra efetiva (ALMEIDA; ALVES; ALMEIDA, 2000), entretanto ainda dependem da associação com inseticidas em concentrações baixas. Dois aspectos principais da biologia dos cupins os tornam especialmente suscetíveis aos pesticidas de solo e as iscas: sua eusocialidade e a alimentação com base em celulose. Pesticidas de solo e aplicações pontuais dependem do contato físico entre cupins e resíduos de inseticidas. Por outro lado, os tratamentos com iscas e madeira dependem dos cupins forrageiros para encontrar, se alimentar e transferir horizontalmente depósitos residuais de inseticidas pelas colônias.

Apesar das técnicas convencionais mencionadas acima serem de certa forma eficientes, elas acabam, muitas vezes, agredindo o meio ambiente ao seu redor, como solo, plantas e até mesmo outros animais. Com o intuito de um manejo menos agressivo, novas tecnologias de manejo de praga estão surgindo recentemente. Essas tecnologias em desenvolvimento foram amplamente facilitadas por estudos de ômicas e pesquisas integrativas de cupins e incluem: (i) RNA de interferência (RNAi), (ii) inibição digestiva, (iii) controle microbiano aprimorado através da supressão das defesas do hospedeiro, (iv) bactérias intestinais “Trojan horse” e interrupção (v) endócrina ou de (vi) feromônio *primer* da homeostase de castas (SCHARF, 2014).

Cupins têm adaptações evolutivas únicas que os tornam ótimos modelos e alvos para a biotecnologia. Os avanços de ferramentas de bioinformática e de análise de dados possibilitaram gerar informação sobre biologia, fisiologia e bioquímica de cupins, oportunizando o avanço da biotecnologia, seja em novos nichos ou em já existentes. Deste modo, estudos proteômicos em cupins basais identificaram celulases protistas e outras lignocelulases hospedeiras inicialmente identificadas por meio de sequenciamento de metatranscriptoma (SETHI et al., 2013). Através de pesquisas em metabolômica, avanços surgiram na identificação de substâncias tóxicas expelidas por cupins soldados na defesa de suas comunidades (SILLAM-DUSSÈS et al., 2012) e na compreensão da digestão de celulose (MIKAELIAN et al., 2014) e lignocelulose (NI; TOKUDA, 2013).

Vias metabólicas relacionadas à diferenciação de castas não reprodutoras e à digestão de nutrientes também envolvem conjuntos de genes e proteínas que são candidatos naturais para aplicações biotecnológicas, como já mostrado para o bioetanol (SCHARF; BOUCIAS, 2010), indústria de papel e celulose (BRUNE, 2014) e estratégias de controle e manejo de pragas (KRISHNAN et al., 2014). Por exemplo, o uso de RNAi para direcionar RNAs codificadores de enzimas vitais foi uma abordagem para a inibição da digestão de cupins (ZHOU et al., 2008a). Além disso, RNAi também já está sendo aplicado na inibição de genes importantes relacionados ao desenvolvimento e diferenciação de castas, e.g., hexamerinas e receptores de insulina, podendo causar um distúrbio populacional nas colônias

de cupins (ZHOU et al., 2008b). Por fim, outros processos industriais e biotecnológicos inspirados por cupins emergentes na atualidade incluem biomateriais, biorremediação, produção de hidrogênio e substâncias antimicrobianas (SCHARF, 2014).

1.4 A espécie *Heterotermes tenuis*

A espécie *Heterotermes tenuis*, pertencente aos cupins basais da família Rhinotermitidae, é subterrânea, constituída por cupins de pequeno porte e apresenta soldados dimórficos. Os indivíduos dessa espécie são esbranquiçados, com cabeça amarelada, corpo afilado e mandíbulas longas (CONSTANTINO, 1991). É uma espécie nativa do Brasil, com preferência a habitats neotropicais. Algumas das características marcantes do *H. tenuis* são o fato dele não levar terra para o interior das galerias e construir colônias subterrâneas e cilíndricas. Eles também costumam habitar cupinzeiros abandonados de outras espécies. O hábito críptico desse inseto dificulta os estudos e a coleta de dados por parte dos pesquisadores (CAMARGO-DIETRICH; COSTA-LEONARDO, 2000).



Figura 5. Exemplos de *Heterotermes tenuis*. Fonte: extraído de <https://home.czu.cz/sobotnik/termity>. Acesso em: 29 de janeiro de 2020.

Em ambientes naturais, o *H. tenuis* é normalmente encontrado em madeiras com certo grau de decomposição. Porém, com o crescimento drástico da agricultura no Brasil ao longo dos anos, principalmente na forma de monocultura, em conjunto com as preferências alimentares dos cupins em geral, a espécie tornou-se uma potente praga agrícola. Segundo MILL, 1992, a espécie já foi descrita atacando milho envelhecido, eucalipto vivo e, principalmente, cana-de-açúcar, da qual ela é classificada como sendo uma das maiores pragas. Os cupins atacam os toletes (pedaço de cana com uma ou mais gemas), danificando as gemas, o que atrapalha a germinação. Isso acaba provocando perdas e exigindo replantio.

Os prejuízos chegam a atingir 10 toneladas de cana por hectare, anualmente (GALLO et al., 2002).

É necessária a busca por métodos eficazes de controle que não impliquem na utilização de inseticidas clorados ou fosforados, uma vez que estes são proibidos ou têm restrições graves, devido às suas diversas toxicidades. Entre as abordagens não-químicas, alterações nos métodos de cultivo das culturas atacadas não parecem ser garantia de uma proteção duradoura. Portanto, recorrer a uma aplicação química cuidadosamente selecionada para atingir os comportamentos de forrageio dos cupins é uma questão que continua em alta (ROULAND-LEFÈVRE, 2010). Além disso, as estratégias de controle integradas e interdisciplinares, incluindo o controle biológico de pragas, desenvolvimento de cultivares transgênicos resistentes, modificações e silenciamento gênico seriam extremamente oportunas.

O *H. tenuis* é, portanto, um potencial alvo para estudo genéticos, tanto para melhor compreensão de sua biologia, como para prospecção de enzimas celulolíticas de interesse industrial e desenvolvimento de novas estratégias de controle.

1.5 A espécie *Cornitermes cumulans*

O cupim *Cornitermes cumulans*, pertencente a família superior Termitidae, coloquialmente conhecido no Brasil como cupim-de-monte ou cupim-de-montículo é muito comum nas pastagens da região Sul e Sudeste, e em parte nas regiões Centro-Oeste e Nordeste (VALÉRIO, 2006). Esse cupim é encontrado em eucaliptais, beira de estradas, pastagens e em cerrado, além de, em menor quantidade, em pomares e beiras de áreas terraceadas. Sua alimentação consiste em maioria de gramíneas secas, porém com preferências por sementes secas de milho ou *Brachiaria spp.* e toletes de cana-de-açúcar, quando presentes (FERNANDES; ALVES, 1992). Suas colônias formam montes em formatos variados, chegando até um metro de altura, com a parte interna revestida com celulose e terra. Após o estabelecimento de suas colônias, o *C. cumulans* causa pouco dano, sendo mais considerado como uma "praga estética", a qual causa desconforto por parte do produtor e dos grandes compradores agrícolas (ZANETTI et al., 2002).



Figura 6. Exemplos de *Cornitermes cumulans*. Fonte: extraído de <https://home.czu.cz/sobotnik/termity>. Acesso em: 29 de janeiro de 2020.

Podemos dizer, portanto, que *C. cumulans* causa poucos danos diretos a maioria das monoculturas, porém por uma questão intrínseca na cultura brasileira, de que insetos normalmente são relacionados como pragas no campo, existe muito preconceito contra a espécie. Como danos indiretos, por outro lado, o fato desses cupins abandonarem seus montes de tempo em tempo, faz com que estes se tornem abrigo para animais peçonhentos, além de, por serem muito altos, ocupam espaço nas terras e dificultam a passagem de máquinas agrícolas e o manejo da lavoura em áreas muito infestadas (VALÉRIO, 2006). Algo a ser notado é que, o próprio pastoreio de gado e o fogo atado para ciclagem das lavouras, tornam o habitat mais apropriado para estes cupins (REDFORD, 1984).

Embora a importância desses cupins como pragas de pastagem seja discutível, tem sido generalizada a demanda por controle; que, predominantemente, tem sido feito através de inseticidas químicos, *e.g.*, abamectina, fipronil, imidaclopride e fosfeto de alumínio (BIONDO, 1988; PEREZ FILHO et al., 1987) ou por aplicações dos fungos *B. bassiana* e *M. anisopliae* diretamente nas colônias (ALMEIDA; ALVES; ALMEIDA, 2000; NEVES; ALVES, 1999). Portanto há espaço para melhorias nas técnicas de controle e para uma melhor informação da comunidade.

1.6 A espécie *Velocitermes heteropterus*

Velocitermes heteropterus é uma espécie também da família de cupins superiores Termitidae e que está distribuída de maneira ampla, desde o Chaco argentino até as savanas da Venezuela, principalmente em regiões de mata aberta, como cerrado. Habita ninhos de outras espécies, principalmente *Cornitermes spp.* e *Silvestritermes euamignathus* mas também pode construir suas próprias colônias, sendo a única espécie do gênero a realizar tal ato (OLIVEIRA, 2013). Sua alimentação é constituída basicamente de serrapilheira e

gramíneas secas na superfície do solo, com as atividades de forrageamento ocorrendo durante a noite e a céu aberto, sendo que suas trilhas são cobertas gradativamente com a utilização contínua (MATHEWS, 1977). Além disso, também costumam levar resíduos de serrapilheira para o ninho, onde os armazenam nas galerias.



Figura 7. Exemplares de *Velocitermes* spp. Fonte: extraído de <https://home.czu.cz/sobotnik/termiti>. Acesso em: 29 de janeiro de 2020.

Seus cupinzeiros costumam apresentar formato piramidal (COLES DE NEGRET; REDFORD, 1982; MATHEWS, 1977), possuindo uma parte no subsolo, normalmente com cerca de 30 cm de profundidade, e também uma parte acima do solo, a qual varia de região para região e pode chegar a até um metro de altura (OLIVEIRA, 2013). As castas auxiliares dessa espécie são polimórficas, com operários apresentando dimorfismo sexual, sendo as fêmeas grandes e os machos pequenos, e os soldados todos machos, de três tamanhos diferentes, pequenos, médios e grandes. Os operários do sexo feminino realizam atividades como o tunelamento e também, junto aos machos, a digestão de celulose (HAIFIG et al., 2011, 2012).

Diferente das outras duas espécies estudadas neste trabalho, *V. heteropterus* não possui muitos relatos de ataque a lavouras agrícolas de grande importância econômica ou a estruturas humanas.

1.7 Introdução a bioinformática e ciências ômicas

A bioinformática é um campo interdisciplinar que desenvolve métodos e ferramentas computacionais para a compreensão e análise de dados biológicos, principalmente os provindos da genética e biologia molecular (PEVSNER, 2015). Como uma área interdisciplinar da ciência, ela combina áreas do conhecimento como ciência da computação, estatística, matemática e engenharia para analisar e interpretar estes dados.

A obtenção e análise de dados genéticos de células, tecidos e até mesmo indivíduos inteiros, aliados ao vertiginoso avanço da computação e da tecnologia dos instrumentos e maquinário de laboratório, tem permitido à comunidade científica o uso de abordagens integradas e inovadoras no estudo de, por exemplo, estrutura e organização de genomas (ABBY; DAUBIN, 2007), expressão diferencial de genes e proteínas (PATTERSON; AEBERSOLD, 2003), estrutura tridimensional de proteínas (GINALSKI, 2006), biologia de sistemas (FRANCKE; SIEZEN; TEUSINK, 2005), predição e classificação funcional de genes (LEE; REDFERN; ORENGO, 2007) e muito mais. A Bioinformática facilita o uso e compreensão de uma enorme quantidade de dados complexos e variados, possibilitando enfrentar o desafio de decifrar componentes importantes dentro de um universo crescente de informações que são as ciências biológicas, o que promoveu o surgimento de áreas do conhecimento específicas para o estudo e comparação destes dados, conhecidas como ciências ômicas (BINNECK, 2004). Em um sentido amplo, os objetivos subjacentes das ômicas são catalogar, quantificar e caracterizar conjuntos de moléculas biológicas que se traduzem em estrutura, função e processos da vida de um organismo ou, até mesmo, de uma amostra ambiental. Os principais tipos de moléculas biológicas estudados pelas ômicas incluem o DNA, o RNA, as proteínas e os metabólitos, sendo os dados coletados a partir delas, os focos da genômica, transcriptômica, proteômica e metabolômica, respectivamente. Além destas quatro, a análise destes dados podem ser expandidas em muitas outras ômicas, como: epigenômica, farmacogenômica, regulômica, peptidômica, metagenômica e metatranscriptômica (BINNECK, 2004; SCHARF, 2015).

Dentro do estudo das ômicas, especificamente na transcriptômica, o sequenciamento completo de transcriptoma (RNA-seq) utilizando tecnologias de sequenciamento de alta cobertura (HTS) passou a revelar o dinâmico perfil do transcriptoma dos mais diversos organismos com elevada sensibilidade e precisão (MARTIN; WANG, 2011). Transcriptoma é o conjunto completo de transcritos (RNAs) e sua quantidade em uma determinada célula ou tecido, num determinado momento de seu desenvolvimento e/ou sob certo estímulo. O entendimento do transcriptoma é essencial para interpretar os elementos funcionais do genoma, revelar os constituintes moleculares de células e tecidos nos diferentes estágios de desenvolvimento e compreender elementos regulatórios que muitas vezes passam despercebidos. Nele, devem estar compreendidos o RNA mensageiro, RNAs não codificantes e pequenos RNAs. Analisando o transcriptoma podemos (i) elucidar a estrutura das regiões codificadoras, (ii) obter melhor compreensão dos mecanismos pós-traducionais e (iii) quantificar níveis de expressão de genes específicos. A alta resolução e cobertura do RNA-seq é capaz de gerar um perfil praticamente completo de um transcriptoma, incluindo até mesmo transcritos raros que desempenham papéis regulatórios. Por exemplo, o RNA-seq obtém resolução em escala de pouquíssimas unidades de sequências, como também

consegue detectar transcritos muito expressos (OZSOLAK; MILOS, 2011; WANG; GERSTEIN; SNYDER, 2009). Apesar destas vantagens, as leituras de sequência obtidas a partir das plataformas HTS costumam ser muito curtas. Como resultado, é necessário reconstruir os transcritos até seu comprimento total por meio de técnicas de montagem de transcriptoma, utilizando-se de *softwares* de bioinformática. Após a montagem, o conjunto de dados pode então seguir para quaisquer análises necessárias e desejadas pelos pesquisadores (METZKER, 2010).

Por fim, a Bioinformática avança rapidamente na maneira de se gerir, editar e estocar dados. Atualmente, a maior parte da construção e armazenagem na biologia de grandes dados é depositada virtualmente e gerida por grandes bancos de dados online, permitindo que os pesquisadores diminuam os gastos em *hardware* e possibilitando transferência de dados de forma mais rápida e segura. Este novo advento da era da Bioinformática fará com que a nova geração de pesquisadores interessados pela ciência dos dados se preocupe cada vez mais com a melhora da performance das ferramentas de análise e transferência de dados e a produzir novos e significativos conhecimentos, até mesmo, a partir de dados já existentes (MARX, 2013).

2 OBJETIVOS

2.1 Objetivo geral

Buscando por uma melhor compreensão dos mecanismos que governam a diferenciação das castas operário e soldado, e prospecção de genes com potencial biotecnológico, foi realizada neste estudo uma análise comparativa inter e intraespecífica entre transcriptomas de cupins das espécies brasileiras *Heterotermes tenuis*, *Cornitermes cumulans* e *Velocitermes heteropterus*, assim como um levantamento da biodiversidade de suas microbiotas endossimbiontes.

2.2 Objetivos específicos

- 2.2.1** Realizar a montagem *de novo* dos transcriptomas das três espécies de cupins.
- 2.2.2** Observar e comparar os aspectos metatranscriptômicos entre as castas e espécies.
- 2.2.3** Quantificar a abundância dos transcritos e realizar análise da expressão diferencial entre as castas para cada espécie.
- 2.2.4** Anotar funcionalmente os transcritos encontrados.
- 2.2.5** Investigar os resultados em busca de genes relacionados ao mecanismo de diferenciação das castas não-reprodutoras (soldado e operário).
- 2.2.6** Bioprospectar genes com possíveis aplicações biotecnológicas.

3 METODOLOGIA

3.1 Obtenção dos insetos e RNA-seq

Os exemplares dos cupins *Heterotermes tenuis*, *Cornitermes cumulans* e *Velocitermes heteropterus* utilizados neste estudo, foram gentilmente cedidos pela Profa. Dra. Ana Maria Costa Leonardo da Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), que os coletou na região de Rio Claro - SP, especificamente nas coordenadas 22°23'S, 47°31'W (*H. tenuis* e *C. cumulans*) e 21°42'S, 47°28'W (*V. heteropterus*). Para este estudo, estamos investigando dois tipos de castas (soldados e operários). Grupos de 200 cupins, corpo inteiro, de cada casta de cada espécie foram macerados na presença de nitrogênio líquido e seu RNA total foi extraído e purificado usando o Mini Kit RNeasy (QIAGEN - Cat. nº 74106), seguindo os padrões do fabricante.

Para a síntese de cDNA, utilizamos procedimentos previamente descritos por MATZ et al., 1999 e o kit de transcriptase reversa SuperScript II (Invitrogen). O sequenciamento foi realizado a partir de uma biblioteca construída seguindo o protocolo TruSeq RNA™ Sample Preparation para a plataforma HiSeq™2000 (Illumina®), feito na Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ-USP), gerando arquivos fastq *paired-end* com *reads* no tamanho 2x100pb.

3.2 Pré-processamento e montagem do transcriptoma

Para a realização das análises bioinformáticas necessárias ao desenvolvimento deste estudo, o uso de recursos computacionais do servidor em nuvem da UFSCar, Cloud-UFSCar, foi de extrema importância. Antes da montagem dos transcriptomas, foi realizado um pré-processamento dos dados, o qual foi constituído de uma análise da qualidade e posterior "trimagem" dos *reads*. Os arquivos fastq do RNA-seq foram inspecionados quanto à qualidade no programa FastQC v0.11.4 (ANDREWS, 2010), considerando principalmente os dados de qualidade de todas as bases, em termos de Phred Q40, e sua distribuição sobre as leituras, em termos de porcentagem. A ferramenta de *trimming* TrimGalore! v0.4.1 (KRUEGER, 2015) foi então utilizada para remover nove bases da extremidade 5' dos *reads* e uma base da extremidade 3', as quais foram consideradas tendenciosas após inspeção com FastQC. Além disso, possíveis artefatos de baixa qualidade, adaptadores e *reads* que não obtiveram combinação com seu respectivo par, também foram removidos para todas as amostras.

Após o pré-processamento, os *reads* seguiram para montagem *de novo*, ou seja, uma montagem que não utiliza um genoma de referência para apoio, iniciando com a ferramenta de montagem Trinity v2.6.6 (GRABHERR et al., 2011), sob parâmetros padrão, e depois passando para o montador CAP3 (HUANG, 1999), aumentando a "porcentagem máxima de

overhang” para 30 ($h = 30$) a fim de diminuir o número de *contigs* montados que apresentavam redundância, o que é comumente visto em montagens realizadas na ferramenta Trinity (ONO et al., 2015). Primariamente, cada amostra foi montada separada, gerando um total de seis arquivos, um operário e um soldado para cada uma das espécies. Ademais, utilizando a mesma estratégia foram montados outros três transcriptomas, um por espécie, combinando *reads* de ambos operários e soldados. A primeira abordagem foi necessária para analisar o conteúdo endossimbiótico de operários e soldados, e a segunda para a análise da expressão diferencial entre essas castas.

Os dados de conteúdo GC, N50, L50, total de transcritos, maior transcrito, tamanho médio e mediana de transcritos foram coletados para todos os transcriptomas montados usando *scripts* do próprio pacote Trinity e do pacote BBTools (BUSHNELL, 2014). Os *reads* foram também alinhados em cada uma das suas respectivas montagens utilizando-se Bowtie2 v2.3.4 (LANGMEAD; SALZBERG, 2012) a fim de se checar a quantidade de *reads* presentes e concordantemente alinhados em cada um dos arquivos. As métricas de qualidade técnica do sequenciamento foram complementadas pela quantificação da completude dos transcriptomas a partir de medidas baseadas em expectativas evolutivamente informadas do conteúdo de genes de ortólogos de cópia única, ou seja, sequências de cópia única que representam um grupo de ortólogos, realizada pela ferramenta BUSCO v3 (SIMÃO et al., 2015). Cada montagem foi quantificada contra os bancos de dados de ortólogos OrthoDB v9 para Artrópodes ($n=1066$) e para Bactérias ($n=148$), e um banco de dados de ortólogos do ENSEMBL personalizado para Protistas ($n=215$).

Após se observar os resultados obtidos pelas ferramentas descritas acima, os transcriptomas foram levados para o software TransDecoder v5.2.0 (HAAS et al., 2013) no qual as potenciais sequências de codificação (CDSs) em todos os seis quadros de leitura abertos (ORF) foram extraídas seguindo os parâmetros padrões e sem nenhum critério de retenção por homologia. Assim como anteriormente, utilizou-se o montador CAP3 nos ORFs extraídos para diminuir ainda mais o número de possíveis *contigs* redundantes. Todas as montagens foram então traduzidas para aminoácidos através da ferramenta transeq, do pacote EMBOSS (RICE; LONGDEN; BLEASBY, 2000). Por fim, os arquivos de ORFs das montagens por castas seguiram adiante para a anotação taxonômica, enquanto as montagens por espécies foram levadas para quantificação da expressão diferencial e anotação funcional dos transcritos.

3.3 Anotação taxonômica dos transcriptomas de cada casta

Os ORFs das seis montagens por castas foram levados para um *pipeline* de anotação taxonômica realizado manualmente no nosso laboratório. Como estamos investigando o RNA

extraído do corpo inteiro, esperamos encontrar sequências de outros organismos além dos cupins, devido à grande variedade de espécies simbióticas que coexistem em seu trato digestório, como por exemplo protistas e bactérias. Esse *pipeline* se baseou na utilização da ferramenta de alinhamento DIAMOND v0.9.24 (BUCHFINK; XIE; HUSON, 2015) contra o banco de dados RefSeq de proteínas não redundantes do NCBI para as seis amostras, usando o método BLASTp e os parâmetros "*sensitive*" (modo recomendado para alinhamento de sequências mais longas), "*salltitles*" (todos os títulos), "*max-target-seqs*" igual a um (o número máximo de sequências de destino por *query* para reportar alinhamentos) e "*max-hsps*" também igual a um (apenas o par de maior pontuação único de *query/subject* é relatado). Três tipos de saídas foram obtidos: formato BLAST *pairwise*, formato BLAST tabular e arquivo de alinhamento DIAMOND (.daa). Este último tipo de saída foi processado por uma ferramenta de linha de comando chamada Meganizer, que está inclusa no pacote DIAMOND, que permitiu que os dados fossem realocados diretamente na próxima ferramenta para exploração e análise interativas.

Esses arquivos "meganizados" continham todos os *reads* alinhados, alinhamentos, classificações e índices específicos de classificação e foram, então, inspecionados taxonomicamente no programa MEGAN v6.6.3 (HUSON et al., 2007), o qual permite a inspeção visual e interativa de numerosos dados provindos do DIAMOND/BLAST, propiciando análise taxonômica de maneira isolada e comparativa, permitindo a observação do panorama metatranscriptômico das nossas amostras e o debate sobre o papel da biodiversidade microbiana dentro de cada um de nossos cupins. Esse *pipeline* foi considerado eficiente na realização de análises funcionais de conjuntos de dados metagenômicos em larga escala (HUSON et al., 2016), além de ser até 20.000 vezes mais rápido que o BLASTx, mantendo um grau adequado de sensibilidade (BUCHFINK; XIE; HUSON, 2015).

3.4 Análise da expressão diferencial

Para a mensuração dos valores de expressão dos ORFs extraídos dos transcriptomas por espécies, seguimos parte dos protocolos e *scripts* elaborados por HAAS et al., 2013 e disponibilizados juntos com o pacote Trinity. De acordo com o protocolo, novos alinhamentos foram feitos no programa Bowtie2 v2.3.4, mapeando os *reads* pré-processados de cada uma das bibliotecas nos ORFs encontrados de seus respectivos transcriptomas. Em seguida, o pacote RSEM v1.3.0 (LI; DEWEY, 2011) foi utilizada dentro do ambiente R v3.3.1 (DESSAU; PIPPER, 2008) para estimar a contagem de *raw reads* de cada ORF dentro de cada biblioteca, normalizando estes dados para unidades de transcritos por milhão (TPM) e, adiante, gerando matrizes de expressão para comparação entre castas através do método de *trimmed means of M-values* (TMM).

Para identificar genes diferencialmente expressos (DEGs) entre castas, aplicamos testes não paramétricos às contagens normalizadas usando a ferramenta NOISeq-sim (TARAZONA et al., 2011), com condições padrão e o valor limite de significância sugerido pelos autores ($q = 0,9$). Os *heatmaps* dos DEGs foram construídos usando um servidor *web* disponível gratuitamente no Heatmapper (BABICKI et al., 2016), com o método de agrupamento *average linkage* e o método de medição de distância euclidiana.

3.5 Anotação taxonômica-funcional dos DEGs

Assim como para as montagens por castas, a anotação taxonômica-funcional dos ORFs por espécies também foi realizada com a ferramenta DIAMOND, seguindo as mesmas instruções citadas na seção 3.3, porém dessa vez utilizando resultados no formato BLAST tabular. As previsões de ortologia e a anotação funcional dos ORFs também foram inferidos usando o eggNOG-mapper v.2 (HUERTA-CEPAS et al., 2019), considerando um *e-value* de semente de ortólogo de 0,001. Os resultados gerados pelas ferramentas DIAMOND e eggNOG-mapper foram anexados juntos aos valores de níveis de expressão e então filtrados quanto a sua expressão diferencial para a sequência do *pipeline*. A existência de sobreposição nos DEGs entre castas e espécies também foi investigada. Com base na anotação, formamos conjuntos de genes que compreendem genes comuns e únicos entre castas e espécies. As isoformas foram consideradas como um único gene, bem como a mesma anotação genética de diferentes organismos (de cupins ou seus simbiosites). Anotações não específicas, de baixa qualidade, ou referentes a proteínas hipotéticas foram desconsideradas. As interseções resultantes foram plotadas usando o pacote UpSetR (CONWAY; LEX; GEHLENBORG, 2017), também no ambiente R. Por fim, os termos de Ontologia Genética (GO) nos níveis um e dois foram atribuídos usando o Blast2GO v5.2 (CONESA; GÖTZ, 2008) e suas distribuições foram plotadas para os 20 principais hits. Abaixo, na Figura 8, temos o *design* experimental que resume a metodologia realizada neste estudo:

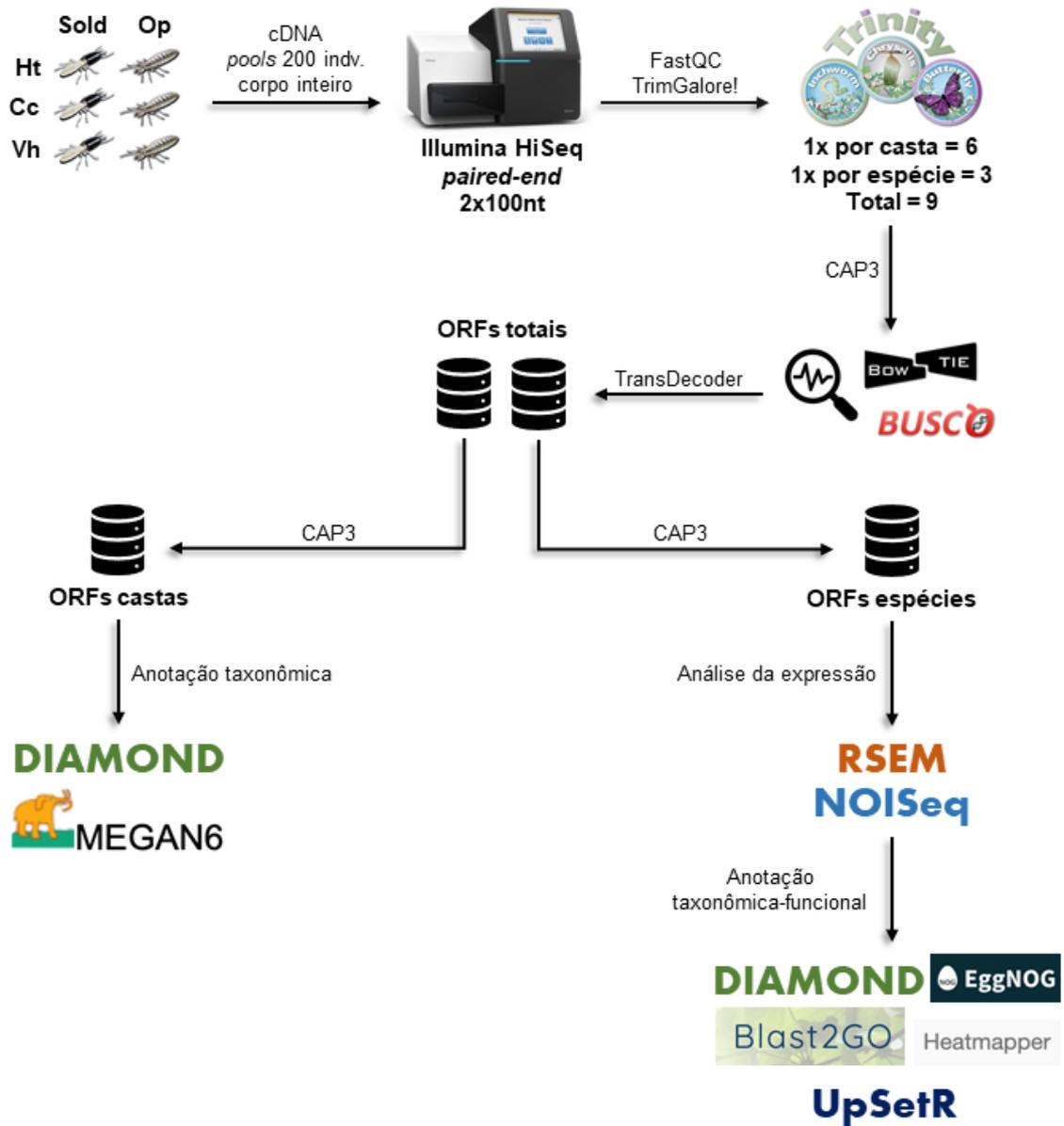


Figura 8. Design experimental do presente estudo.

4 RESULTADOS E DISCUSSÃO

4.1 Montagens, avaliação da qualidade e predição das ORFs

A etapa de pré-processamento dos dados na ferramenta TrimGalore! gerou uma pequena diminuição no tamanho dos *reads*, que passaram de 100 bases para 90 cada, além da filtragem de *reads* de baixa qualidade, o que nos gerou *reads* com uma qualidade superior para a posterior montagem dos transcriptomas. Com o objetivo de investigar o conteúdo endossimbiótico, primeiramente produzimos transcriptomas de ambas as castas, operários e soldados, de cada espécie estudada, totalizando seis perfis de transcriptoma. Além disso, para analisar a expressão gênica das castas, também produzimos um transcriptoma por espécie, combinando os *reads* de operário e soldado na mesma montagem. Para melhor ilustração das tabelas e gráficos reproduzidos nesta dissertação, os dados provindos das espécies *H. tenuis*, *C. cumulans* e *V. heteropterus* estão representados pelas siglas “Ht”, “Cc” e “Vh”, respectivamente, assim como a casta operário e soldado, por “op” e “sold”. Os dados e estatísticas das seis montagens por castas e das três por espécies podem ser inspecionados nas Tabelas 1 e 2, respectivamente.

Tabela 1. Descrição das montagens de novo por castas.

	Ht op	Ht sold	Cc op	Cc sold	Vh op	Vh sold
Sequenciamento e qualidade						
Total de reads paired end	19.608.150	20.757.104	17.964.366	18.206.386	17.991.218	20.562.928
Número de reads filtrados	18.816.712	19.877.680	17.261.756	17.585.414	17.397.290	19.785.684
Estatísticas da montagem						
Total de bases montadas	77.592.524	80.719.165	47.180.262	72.376.483	53.657.320	69.925.670
Total de <i>contigs</i> Trinity	136.055	106.660	61.179	85.144	73.986	86.566
Total de <i>contigs</i> após CAP3	126.272	96.418	53.578	74.712	64.373	73.415
Conteúdo GC (%)	44,52	41,11	41,02	40,62	40,63	40,57
L50	23.212	13.117	7.746	9.795	10.402	10.567
N50	730	1.349	1.340	1.562	1.155	1.422
Média do tamanho de <i>contig</i> (pb)	559	740	762	834	722	811
Maior <i>contig</i> (pb)	9.976	20.084	14.524	52.176	21.733	24.183
ORFs						
Número de ORFs preditas	45.296	28.464	16.637	21.111	18.688	21.187
Número de ORFs preditas após CAP3	42.619	24.796	15.030	17.990	16.414	18.099
Número de ORFs anotadas	27.292	19.744	13.646	15.918	14.682	15.845
ORFs anotadas (%)	64,04	79,63	90,79	88,48	89,45	87,55

Tabela 2. Descrição das três montagens de novo por espécies.

	Ht op+sold	Cc op+sold	Vh op+sold
Sequenciamento e qualidade			
Total de reads paired end	40.365.254	36.170.752	38.553.902
Número de reads filtrados	38.694.392	34.847.170	37.182.974
Estatísticas da montagem			
Total de bases montadas	133.951.862	92.240.773	92.648.075
Total de <i>contigs</i> Trinity	188.478	97.974	105.769
Total de <i>contigs</i> após CAP3	172.946	81.812	86.706
Conteúdo GC (%)	43,61	40,61	40,51
L50	24.214	10.442	12.267
N50	1.155	1.917	1.640
Média do tamanho de <i>contig</i> (pb)	694	931	889
Maior <i>contig</i> (pb)	19.691	52.202	28.360
ORFs			
Número de ORFs preditas	63.066	23.413	24.486
Número de ORFs preditas após CAP3	55.047	18.867	19.272
Número de ORFs anotadas	34.652	16.322	16.424
ORFs anotadas (%)	62,95	86,51	85,22

Como podemos observar nas Tabelas 1 e 2, a maioria de nossas montagens obtiveram valores de N50 acima de 1100 bases, o que, até mesmo para o N50 de 730 de Ht op, são valores satisfatórios (GENG et al., 2018), nos informando que metade das bases montadas dos nossos transcriptomas estão em *contigs* desses valores ou acima. É curioso observar que, existe um certo padrão entre os valores das montagens quando comparamos as castas operário e soldado: no geral, o número de *contigs* e ORFs obtidos para as castas de soldados são sempre maiores que para as castas de operários, o que talvez possa indicar um maior número de proteínas nesta casta em comparação com os operários. Por outro lado, nas montagens das castas da espécie basal o oposto é observado, sendo que o perfil de Ht op apresentou mais *contigs* e o dobro do número de ORFs preditos do que os outros perfis (Tabela 1). Entre as espécies basais, estudos com *Reticulitermes speratus* já demonstraram que o número de protistas intestinais em operários podem ser significativamente maior que em soldados (SHIMADA et al., 2013), o que nos sugere que o maior número de genes expressos em Ht op pode ser devido a expressão gênica de protistas endossimbiontes. Hipótese esta, que foi corroborada pela anotação, como mostraremos em seções seguintes dos resultados.

Comparando somente os transcriptomas por espécies (Tabela 2), nota-se que Ht apresentou o dobro de contigs (172.946) em comparação com Cc e Vh (81.812 e 86.706, respectivamente). Outra comparação similar, realizada entre a espécie basal *Zootermopsis nevadensis* (TERRAPON et al., 2014) e a superior *Macrotermes natalensis* (POULSEN et al., 2014) mostrou que o último possui um tamanho de genoma que é mais do que o dobro do tamanho do primeiro (KORB et al., 2015). Entretanto em ambos os estudos, o tecido intestinal e seu conteúdo não foram considerados, o que sugere que o padrão oposto encontrado para as espécies aqui analisadas pode ser devido a um maior número de genes de endossimbiontes sendo expressos em Ht. Isto também é indicado ao se observar o número de ORF predito, 55.047 para Ht, em contraste com 15.876 e 16.310 para *Z. nevadensis* e *M. natalensis*, respectivamente. Um resultado descrito para o cupim *Reticulitermes flavipes* também corroborou essa ideia, pois 247.174 contigs foram montados a partir de amostras considerando o conteúdo intestinal, mas apenas 13.755 correspondiam a homólogos de insetos (WU; DHAMI; THOMPSON, 2018).

Tabela 3. Resumo das análises de read content para as montagens por castas. Os resultados são referentes aos alinhamentos realizados na ferramenta Bowtie2 com as montagens antes e depois da passagem pelo montador CAP3. Todos os valores estão informados em porcentagem (%).

	Ht op	Ht sold	Cc op	Cc sold	Vh op	Vh sold
Montagem Trinity antes de CAP3						
Taxa de alinhamento geral	86,94	90,46	93,30	91,88	92,80	92,48
Alinhou concordantemente =1x	39,22	36,98	37,93	40,58	42,67	40,58
Alinhou concordantemente >1x	40,14	46,35	45,58	43,63	40,71	41,18
Reads alinhados concordantemente	79,36	83,33	83,51	84,21	83,38	81,76
Montagem Trinity após CAP3						
Taxa de alinhamento geral	86,48	89,89	92,64	91,38	92,27	91,98
Alinhou concordantemente =1x	47,70	53,17	52,02	50,96	53,25	51,16
Alinhou concordantemente >1x	31,42	29,68	31,45	32,99	29,68	30,42
Reads alinhados concordantemente	79,12	82,85	83,47	83,95	82,93	81,58

Tabela 4. Resumo das análises de *read content* para as montagens por espécies. Os resultados são referentes aos alinhamentos realizados na ferramenta Bowtie2 com as montagens antes e depois da passagem pelo montador CAP3. Todos os valores estão informados em porcentagem (%).

	Ht op+sold	Cc op+sold	Vh op+sold
Montagem Trinity antes de CAP3			
Taxa de alinhamento geral (%)	91,10	93,67	94,39
Alinhou concordantemente =1x (%)	39,62	32,47	36,19
Alinhou concordantemente >1x (%)	43,77	52,37	48,04
<i>Reads</i> alinhados concordantemente (%)	83,39	84,84	84,23
Montagem Trinity após CAP3			
Taxa de alinhamento geral (%)	90,51	92,80	93,81
Alinhou concordantemente =1x (%)	49,98	47,51	47,75
Alinhou concordantemente >1x (%)	32,80	36,83	36,40
<i>Reads</i> alinhados concordantemente (%)	82,78	84,34	84,15

Com os resultados dos alinhamentos realizados no Bowtie2 para averiguar o *read content* de nossas montagens (Tabelas 3 e 4), pode-se constatar que ambos os transcriptomas antes e depois da passagem pelo CAP3 mostraram altas taxas gerais de alinhamento, com a soma dos alinhamentos concordantes sempre acima de 79%, e a maioria dos fragmentos mapeados encontrados foram mapeados como pares adequados. Todos esses parâmetros são consistentes com os recomendados pelos desenvolvedores do pacote Trinity (HAAS et al., 2013). Após a remontagem feita pelo CAP3, os transcriptomas apresentaram níveis mais baixos de redundância em comparação as montagens provindas somente do Trinity, o que foi demonstrado pela menor porcentagem de leituras que se alinharam concordantemente mais de uma vez, além de uma pequena diminuição no número de *contigs* totais montados (Tabelas 1 e 2). Estes resultados suportam a estratégia de montagem híbrida escolhida por nossa equipe.

Tabela 5. Resultados de completude para as seis montagens por castas. A soma de BUSCO completos, fragmentados e faltantes é igual a 100%. Os dados de BUSCO completo são divididos em cópia-única e duplicados. Os bancos de ortólogos utilizados foram, respectivamente, Artrópodes, Bactérias e Protistas. Todos os valores estão informados em porcentagem (%).

	Ht op			Ht sold		
	Artrop.	Bac.	Prot.	Artrop.	Bac.	Prot.
BUSCOs completos	67,7	23,6	45,5	82,0	24,4	56,7
Completo e cópia-única	56,0	15,5	39,5	61,4	14,9	46,5
Completo e duplicado	11,7	8,1	6,0	20,6	9,5	10,2
BUSCOs fragmentados	20,4	20,9	1,4	14,4	14,2	0,9
BUSCOs totais	88,1	44,5	46,9	96,4	38,6	57,6
BUSCOs faltantes	11,9	55,5	53,1	3,6	61,4	42,4
	Cc op			Cc sold		
	Artrop.	Bac.	Prot.	Artrop.	Bac.	Prot.
BUSCOs completos	81,8	27,7	60,0	87,8	25,7	64,2
Completo e cópia-única	71,1	22,3	54,4	68,9	18,9	51,2
Completo e duplicado	10,7	5,4	5,6	18,9	6,8	13,0
BUSCOs fragmentados	13,1	9,5	0,0	9,8	10,1	0,9
BUSCOs totais	94,9	37,2	60,0	97,6	35,8	65,1
BUSCOs faltantes	5,1	62,8	40,0	2,4	64,2	34,9
	Vh op			Vh sold		
	Artrop.	Bac.	Prot.	Artrop.	Bac.	Prot.
BUSCOs completos	84,8	23,0	60,5	88,7	25,6	66,0
Completo e cópia-única	71,0	17,6	55,8	71,5	15,5	58,6
Completo e duplicado	13,8	5,4	4,7	17,2	10,1	7,4
BUSCOs fragmentados	11,1	12,8	0,9	9,1	11,5	0,5
BUSCOs totais	95,9	35,8	61,4	97,8	37,1	66,5
BUSCOs faltantes	4,1	64,2	38,6	2,2	62,9	33,5

Tabela 6. Resultados de completude para as três montagens por espécies. A soma de BUSCO completos, fragmentados e faltantes é igual a 100%. Os dados de BUSCO completo são divididos em cópia-única e duplicados. Os bancos de ortólogos utilizados foram, respectivamente, Artrópodes, Bactérias e Protistas. Todos os valores estão informados em porcentagem (%).

	Ht op+sold			Cc op+sold			Vh op+sold		
	Artrop.	Bac.	Prot.	Artrop.	Bac.	Prot.	Artrop.	Bac.	Prot.
BUSCOs completos	89,8	31,7	61,4	94,4	27,0	71,6	95,2	29,0	69,8
Completo e cópia-única	63,2	20,9	45,6	68,9	18,2	56,3	68,7	18,9	57,7
Completo e duplicado	26,6	10,8	15,8	25,5	8,8	15,3	26,5	10,1	12,1
BUSCOs fragmentados	7,9	18,2	1,4	4,7	9,5	0,5	3,7	8,8	0,0
BUSCOs totais	97,7	49,9	62,8	99,1	36,5	72,1	98,9	37,8	69,8
BUSCOs faltantes	2,3	50,1	37,2	0,9	63,5	27,9	1,1	62,2	30,2

Sendo assim, as montagens seguiram adiante para a ferramenta BUSCO, onde foram checadas quanto a completude de seus genes em relação a bancos de dados personalizados de ortólogos de cópia única, medida que enriquece a avaliação da qualidade das montagens em termo de conteúdo gênico (WATERHOUSE et al., 2017). Com os resultados gerados no BUSCO, os quais podem ser observados nas Tabelas 5 e 6 acima, ficou indicado a integridade de nossas montagens, que possuem grande completude com o banco de dados de ortólogos de artrópodes (>88%). Essa análise é importante, pois montagens que parecem ser de boa qualidade com base em medidas de contiguidade não são necessariamente mais completas em termos de conteúdo gênico esperado (SIMÃO et al., 2015). Uma boa variabilidade de genes provindos de protistas (em torno de 60% de completude) também foi relatada até mesmo para os transcriptomas das espécies superiores, dado que será melhor discutido na próxima seção. Novamente, observamos também que Ht op continua divergindo das demais montagens, apresentando níveis de completude diferentes do esperado, como por exemplo, uma menor completude para genes de protistas do que as nossas espécies superiores.

Finalizando o nosso processo de montagem dos transcriptomas, os ORFs foram preditos no programa Transdecoder e levados ao CAP3, o qual mais uma vez ajudou a remover redundâncias. Ao observarmos as quantidades finais de ORFs (Tabelas 1 e 2), é de fácil percepção a diferença de quantidade de sequências codificadoras entre nossos cupins superiores e basais. O número de ORFs encontrados foi de agrado a nossa equipe e também motivo de surpresa, uma vez que Ht op registrou mais que o dobro de quadros abertos de leitura que as montagens dos cupins superiores.

4.2 Panorama metatranscriptômico de operários e soldados de cupins basais e superiores

Os ORFs das montagens por castas foram traduzidos em sequências de aminoácidos e então alinhados com a ferramenta DIAMOND e analisados no software MEGAN, gerando os dados de anotação taxonômica (Figuras 9 e 10). Foi avaliado o conteúdo endossimbiótico de operários e soldados analisando-se a presença dos seguintes grupos taxonômicos: artrópodes, protistas, fungos e bactérias. É importante apontar que, dentro das sequências categorizadas como “Artrópodes”, cerca de 98% se relacionavam de fato com sequências de demais cupins presentes no banco de dados. Os três táxons com maior número de *hits* no alinhamento com o DIAMOND foram os cupins *Cryptotermes secundus* e *Zootermopsis nevadensis* e a barata *Blattella germanica* para Cc e Vh, e *Cryptotermes secundus*, o flagelado *Tritrichomonas foetus* e *Zootermopsis nevadensis* para Ht.

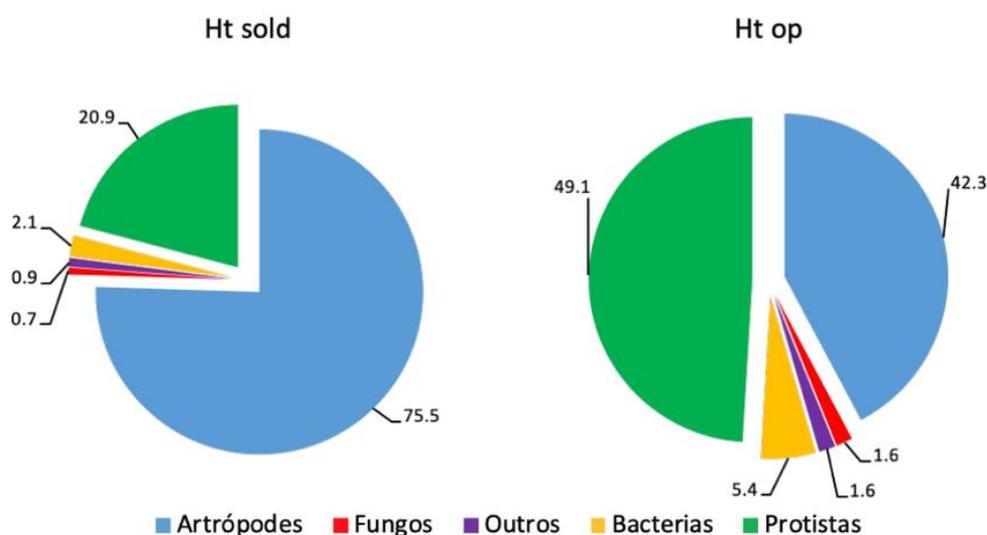


Figura 9. Panorama metatranscriptômico de Ht. Valores em porcentagem (%) da distribuição taxonômica de 19.744 (Ht sold) e 27.292 (Ht op) hits positivos alinhados na ferramenta DIAMOND e analisados no software MEGAN.

Ao se observar os gráficos de classificação taxonômica da espécie *H. tenuis* (Figura 9), a primeira percepção que temos é: estamos investigando um organismo com substanciais interações simbióticas. Em Ht op, foram anotados mais sequências de protistas do que propriamente de cupins, além de mais que o dobro de *contigs* de protistas em comparação com os soldados, o que corrobora a estreita relação de cupins de famílias basais com protistas simbióticos que trabalham, em grande parte, no metabolismo da digestão dos alimentos em seus hospedeiros (LO; EGGLETON, 2010). Protistas, em especial flagelados celulolíticos, foram adquiridos por um ancestral comum de cupins basais e baratas do gênero *Cryptocercus*

(LO et al., 2000), e permitiram uma utilização eficiente da celulose, consequentemente, dando origem ao estilo de alimentação dessas espécies basais de cupins (NAKASHIMA et al., 2002). Além disso, relação essa que, acreditamos ser ainda mais evidenciada em operários do que em soldados, principalmente pelo fato do primeiro ser a casta responsável pelo forrageio e nutrição das demais castas da colônia (BENJAMINO; GRAF, 2016; EGGLETON, 2010).

Ademais, algumas poucas sequências foram anotadas como pertencentes às bactérias e fungos, estes últimos em menor quantidade, para ambas as castas, o que era esperado, uma vez que não existem relatos de grande impacto sobre a relação endossimbiótica de nossa espécie basal com fungos.

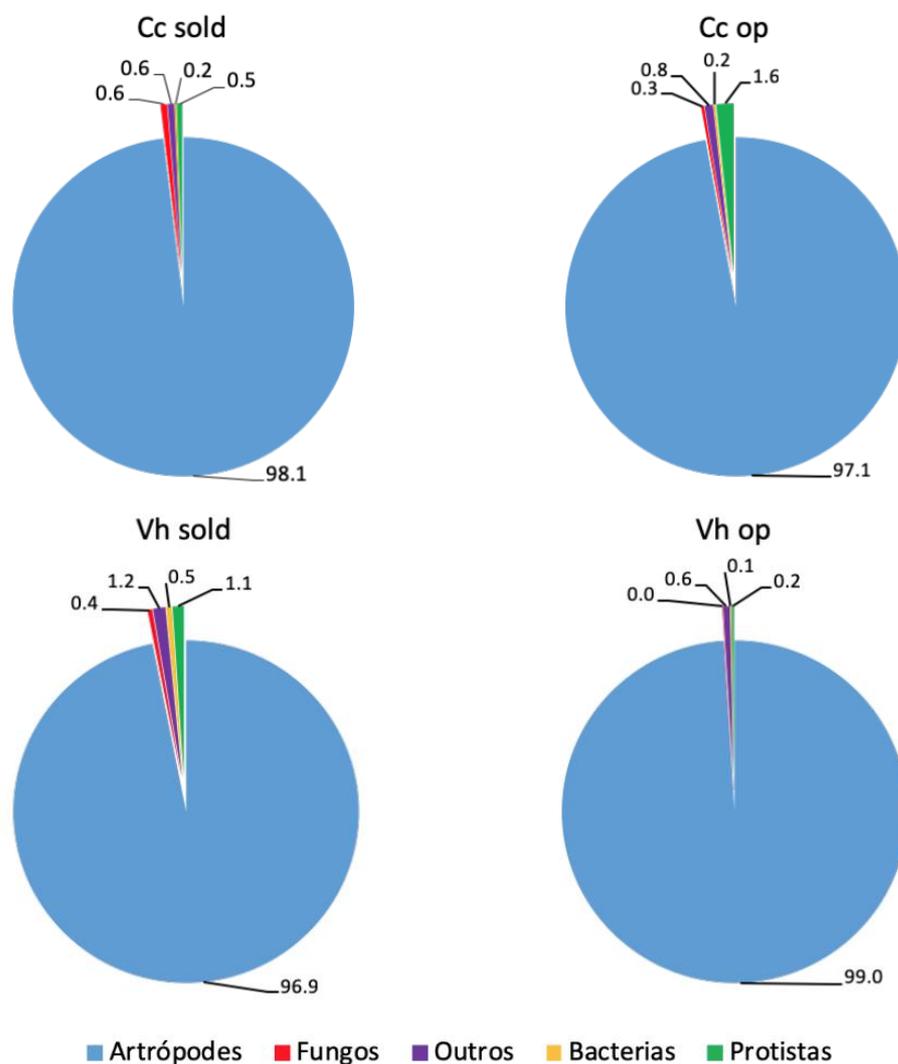


Figura 10. Panorama metatranscriptômico de Cc e Vh. Valores em porcentagem da distribuição taxonômica de 15.918 (Cc sold), 13.646 (Cc op), 15.845 (Vh sold) e 14.682 (Vh op) hits positivos alinhados na ferramenta DIAMOND e analisados no software MEGAN.

Por outro lado, inspecionando a distribuição da diversidade taxonômica das nossas espécies de cupins superiores (Figura 10), podemos constatar que mais de 96% dos ORFs anotados para as quatro diferentes montagens provêm de cupins. Este também é um

resultado esperado, uma vez que cupins superiores são capazes de produzirem suas próprias celulases e digerem a madeira eficientemente (LO; TOKUDA; WATANABE, 2010). O ancestral comum que se derivou nas espécies superiores perdeu simbiontes flagelados, ou acabou realizando transferências horizontais de genes, de modo que não se espera que protistas sejam encontrados no intestino de cupins superiores (DIETRICH; KÖHLER; BRUNE, 2014). No entanto, encontramos alguns genes protistas nos transcriptomas de Cc e Vh. Embora inesperado, isto já havia sido descrito para alguns cupins superiores do gênero *Gnathamitermes*, para os quais foi detectado um filotipo relacionado ao ciliado *Clevelandella* (ABDUL RAHMAN et al., 2015) e para outras espécies superiores, incluindo Cc, que mostraram em sua microbiota intestinal protistas pertencente ao filo Apicomplexa e à ordem Euglenida (GRIECO et al., 2019). Esses resultados indicam que, após a aparente perda de protistas em cupins superiores, esses estão sendo readquiridos em algumas espécies ou, até mesmo, persistiram em números baixos durante essa transição (ABDUL RAHMAN et al., 2015).

Uma vez que cada casta auxiliar tem dietas diferentes e desempenha funções especializadas na colônia, espera-se que o conteúdo de endossimbiontes reflita essas diferenças, mesmo com troca contínua de simbiontes por meio da trofalaxia estomodeal ou proctodeal (OTANI et al., 2019). Encontramos um padrão oposto em Vh, a única espécie em que os operários apresentaram um conteúdo menor de endossimbiontes em comparação com os soldados. Como Cc e Vh têm uma dieta semelhante, sendo ambos colhedores de folhas (CONSTANTINO, 2005), descartamos a hipótese de que essa diferença possa ter ocorrido devido a dietas diferentes. Como o número de ORFs previstos e a porcentagem de ORFs anotados foram muito semelhantes entre os operários de ambas as espécies (Tabela 1), essa diferença provavelmente não aparenta ser devido a artefatos de sequenciamento ou montagem. Assim, supomos que essa diferença se deva a um número maior de genes celulolíticos endógenos em operários de Vh, entretanto estudos complementares são necessários para investigar esse problema em maior detalhe.

Tabela 7. Grupos majoritários de endossimbiontes protistas presentes nos cupins estudados. Os valores informados representam o número de *contigs* anotados para cada um dos grupos.

	Ht op	Ht sold	Cc op	Cc sold	Vh op	Vh sold
Táxon Protista						
Amoebozoa	138	33	114	31	6	29
Apicomplexa	16	5	5	3	1	5
Apusozoa	7	1	1	0	0	3
Choanoflagellida	6	5	3	1	0	1
Ciliophora	69	11	5	1	0	5
Cryptophyta	5	0	0	0	0	4
Euglenida	31	2	1	2	0	3
Fornicata	8	1	19	16	0	2
Haptophyceae	15	1	1	1	0	1
Heterolobosea	28	2	7	2	1	1
Jakobida	1	1	2	1	0	1
Opisthokonta	4	1	3	0	0	4
Oxymonadida	5	0	2	8	21	38
Parabasalia	12634	3904	33	0	1	5
Rhizaria	25	5	2	0	0	1
Stramenopiles	87	21	2	2	2	11
Não classificados	0	0	1	0	0	10
Outros	136	92	1	2	0	3
Total	13215	4085	202	70	32	127

Os protistas intestinais dos cupins geralmente se enquadram em dois grupos principais: o filo Parabasalia e a ordem Oxymonadida (Filo Preaxostyla) (BRUNE; DIETRICH, 2015; NALEPA, 2017). Como pode ser observado (Tabela 7), de fato, a maioria dos *contigs* anotados de protistas de Ht pertence ao filo Parabasalia enquanto poucos *contigs* foram atribuídos à ordem Oxymonadida (cinco *contigs* apenas em operários). Em relação às espécies superiores, em Cc, a maioria dos *contigs* de protistas pertence ao filo Amoebozoa. Destacamos também a presença de Parabasalia em operários, mas não em soldados desta espécie. Em Vh, a maioria dos *contigs* de protistas pertence à ordem Oxymonadida, embora também mostrem um número considerável de *contigs* anotados como Amoebozoa. Levando em conta o nosso conhecimento e a literatura revisada, é a primeira vez que o filo Amoebozoa é mostrado como abundante em amostras de cupins. Além disso, em nossas análises com o BUSCO, observamos porcentagens ligeiramente mais altas de BUSCOs protistas nas espécies superiores (Tabela 5), o que indica que mesmo representando uma parte significativamente menor dos transcriptomas em comparação com Ht, o conjunto de genes

protistas em Cc e Vh, surpreendentemente, parece ser mais diversificado em termos de grupos de ortólogos.

Acredita-se que a perda de simbiontes flagelados em cupins superiores tenha resultado em uma microbiota intestinal rica em bactérias, possibilitando inovações simbióticas e diversificação alimentar (BRUNE; DIETRICH, 2015). Contudo, observamos uma porcentagem maior de sequências de protistas do que de bactérias para as espécies superiores analisadas, apesar de serem pequenas as porcentagens de *contigs* de endossimbiontes (Figura 10).

Tabela 8. Filos de endossimbiontes bacterianos presentes nos cupins estudados. Os valores informados representam o número de *contigs* anotados para cada um dos grupos.

	Ht op	Ht sold	Cc op	Cc sold	Vh op	Vh sold
Táxon Bactéria						
Acidobacteria	19	1	0	0	0	6
Actinobacteria	98	33	2	2	0	2
Bacteroidetes	599	177	2	0	0	6
Elusimicrobia	2	1	0	0	0	2
Firmicutes	344	87	9	9	0	16
Planctomycetes	13	6	0	1	0	1
Proteobacteria	209	55	6	12	11	28
Spirochaetes	65	17	1	0	0	1
Synergistetes	3	1	2	0	0	0
Thermotogae	4	1	0	0	0	2
Verrucomicrobia	13	5	0	0	0	2
Não classificados	34	8	1	1	1	5
Outros	71	17	0	0	1	3
Total	1474	409	23	25	13	74

Analisando a microbiota bacteriana, observamos que operários e soldados de Ht apresentaram grupos dominantes similares (Tabela 8). Nesta espécie, as bactérias pertencentes ao filo Bacteroidetes e Firmicutes foram as mais frequentes, padrão semelhante ao descrito para a espécie basal *Coptotermes formosanus* (SHINZATO et al., 2005), mas diferente do observado nas também espécies basais *Reticulitermes speratus* e *Incisitermes marginipennis*, que possuem a comunidade bacteriana dominada por Spirochaetes (DIETRICH; KÖHLER; BRUNE, 2014; HONGO; OHKUMA; KUDO, 2003). O filo Spirochaetes é característico da microbiota intestinal de cupins que se alimentam de madeira, às vezes, compreendendo até metade dos procariontes totais de certas espécies (BRUNE, 2014), mas que foi consideravelmente sub-representado em Ht.

Embora a microbiota intestinal bacteriana dos cupins basais seja modelada principalmente pelos simbioses específicos de seus flagelados intestinais, as diferenças dentro dos cupins superiores são modeladas principalmente pela filogenia do hospedeiro, além de também poderem ser um reflexo das adaptações funcionais à diferentes dietas do mesmo (ABDUL RAHMAN et al., 2015; BRUNE, 2014). Observamos que Cc e Vh apresentaram predominantemente os mesmos dois grupos de bactérias: Proteobacteria e Firmicutes, o que difere consideravelmente dos resultados relatados para outros cupins superiores. Por exemplo, espécies superiores dos gêneros *Macrotermes* e *Odontotermes* (ambos cupins produtores de fungos) são dominadas pelos Bacteroidetes e Firmicutes, sendo que nessas espécies, subconjuntos de bactérias proliferam diferentemente entre soldados e operários (OTANI et al., 2019). *Nasutitermes takasagoensis* (cupim que se alimenta de madeira) são dominados por Spirochaetes e Fibrobacteres, e *Termes comis* (cupim que se alimenta de húmus) por Firmicutes e Spirochaetes (HONGOHO et al., 2006; THONGARAM et al., 2005). Considerando todas essas descobertas e reunindo nossos resultados para Cc e Vh, sugerimos que exista uma relação entre diferentes dietas e diferentes grupos de bactérias simbioses em cupins superiores.

4.3 Padrões de expressão gênica e anotação funcional dos DEGs

Os dados de expressão foram mensurados em valores de TPM e FPKM, porém o primeiro foi escolhido para a sucessão das análises pois os valores de TPM sempre somam para o mesmo valor independente da amostra (LI; DEWEY, 2011; WAGNER; KIN; LYNCH, 2012), facilitando possíveis comparações.

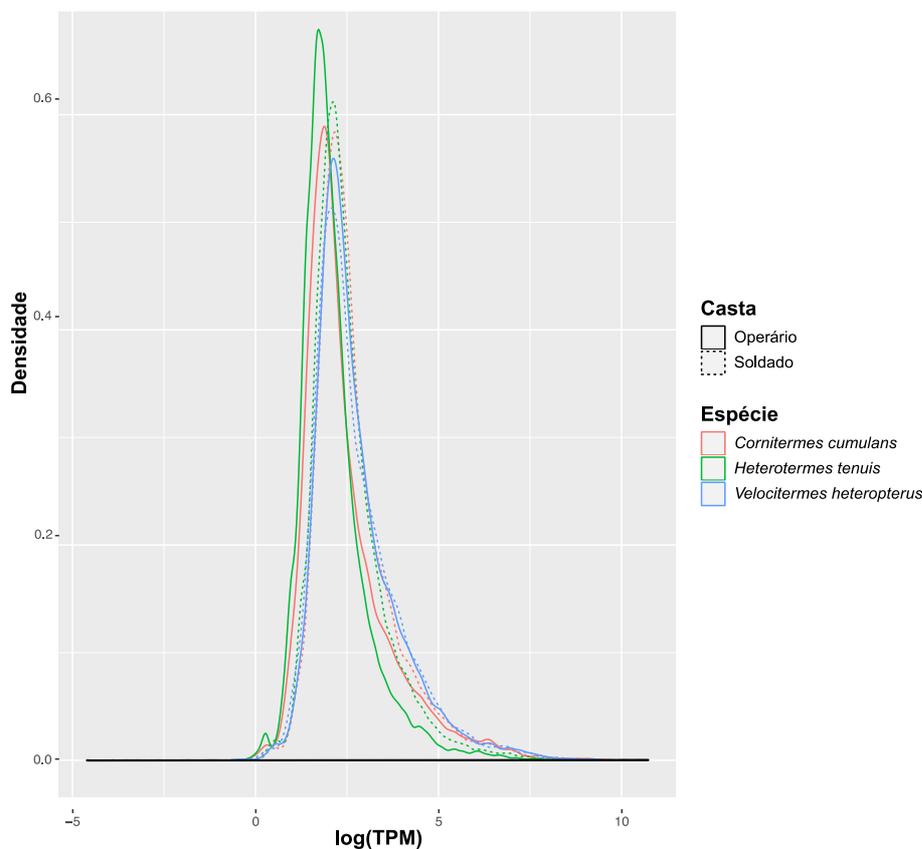


Figura 11. Distribuição dos valores de expressão (TPM) pela abundância dos contigs.

Acima, na Figura 11, está representado a distribuição dos valores de TPM de todos os ORFs extraídos das três montagens por espécies. É visível que todas as nossas amostras apresentam um padrão que culminam em um mesmo pico de abundância, que fica entre os valores de 10 e 1000, inclusive para Ht op, no qual a única diferença é que seu pico tem maior densidade (devido ao seu maior número de ORFs em relação aos demais). Apenas uma pequena parcela dos ORFs se encontra nas extremidades, o que é normal.

Foram encontrados 4821, 367 e 65 genes *up* regulados em operários de Ht, Cc e Vh, respectivamente; e 1447, 767 e 332 genes *up* regulados em soldados de Ht, Cc e Vh, respectivamente (Tabela 9). O total dos DEGs para as espécies correspondeu a 11,39%, 6,01% e 2,06% dos ORFs preditos de Ht, Cc e Vh, respectivamente. Destacamos o grande número de DEGs em Cc em comparação a Vh, embora o número de ORFs extraídos tenha

sido muito semelhante entre eles. Todos os genes inferidos com significância positiva para expressão diferencial apresentaram *fold-change* igual ou maior que oito, sendo os maiores números de *fold-change* observados nas amostras de Cc. Encontramos um número de DEGs substancialmente maior do que os relatados para o cupim *R. flavipes*, espécie na qual de um total de 13.755 genes, 93 foram expressos diferencialmente (comparando soldados, ninfas e operários), destes, 78% *up* regulados para soldado (WU; DHAMI; THOMPSON, 2018).

Tabela 9. Resumo da anotação funcional dos DEGs realizada através da ferramenta BLAST2GO. Os valores informados para "Número de DEGs" representam o número de genes que foram considerados *up* regulados para determinada casta.

	Ht op	Ht sold	Cc op	Cc sold	Vh op	Vh sold
Número de DEGs	4.821	1.447	367	767	65	332
Anotados DIAMOND	1.345	940	197	384	36	168
GOs anotados	2.338	957	343	511	48	300
GOs anotados (%)	48,50	66,14	93,46	66,62	73,85	90,36
Total BP (%)	29,81	29,89	33,24	28,96	31,25	29,33
Total CC (%)	10,52	16,20	9,33	17,03	8,33	15,67
Total MF (%)	59,67	53,92	57,43	54,01	60,42	55,00
Top BP	<i>Cellular process</i>	<i>Cellular process</i>	<i>Metabolic process</i>	<i>Cellular process</i>	<i>Metabolic process</i>	<i>Cellular process</i>
Top MF	<i>Binding</i>	<i>Binding</i>	<i>Catalytic activ.</i>	<i>Binding</i>	<i>Binding</i>	<i>Binding</i>
Top CC	<i>Cell / cell part</i>	<i>Membrane</i>	<i>Cell / cell part</i>	<i>Cell / cell part</i>	<i>Membrane</i>	<i>Organelle / cell / cell part</i>

Uma sobreposição significativa nos perfis de expressão gênica entre cupins operários e soldados é esperada, já que ambas as castas compartilham uma trajetória de desenvolvimento similar, bem como a função não-reprodutiva (KORB et al., 2015). No entanto, uma vez que são comportamental e morfológicamente distintas, os genes envolvidos com sua diferenciação podem apresentar padrões de expressão divergentes, como foi demonstrado para famílias de genes especificamente superexpressas em uma casta específica, assim como outros genes que são estritamente expressos em função da casta, e.g., genes celulolíticos (MITAKA et al., 2016; TERRAPON et al., 2014; WU; DHAMI; THOMPSON, 2018). *Heatmaps* foram gerados através das análises de expressão diferencial comparando operários e soldados de cada espécie e foram observados padrões semelhantes de expressão diferencial para as espécies superiores, uma vez que os soldados mostraram mais genes *up* regulados do que os operários (Figura 12B e 12C), enquanto que nas espécies basais, revelou-se o oposto (Figura 12A).

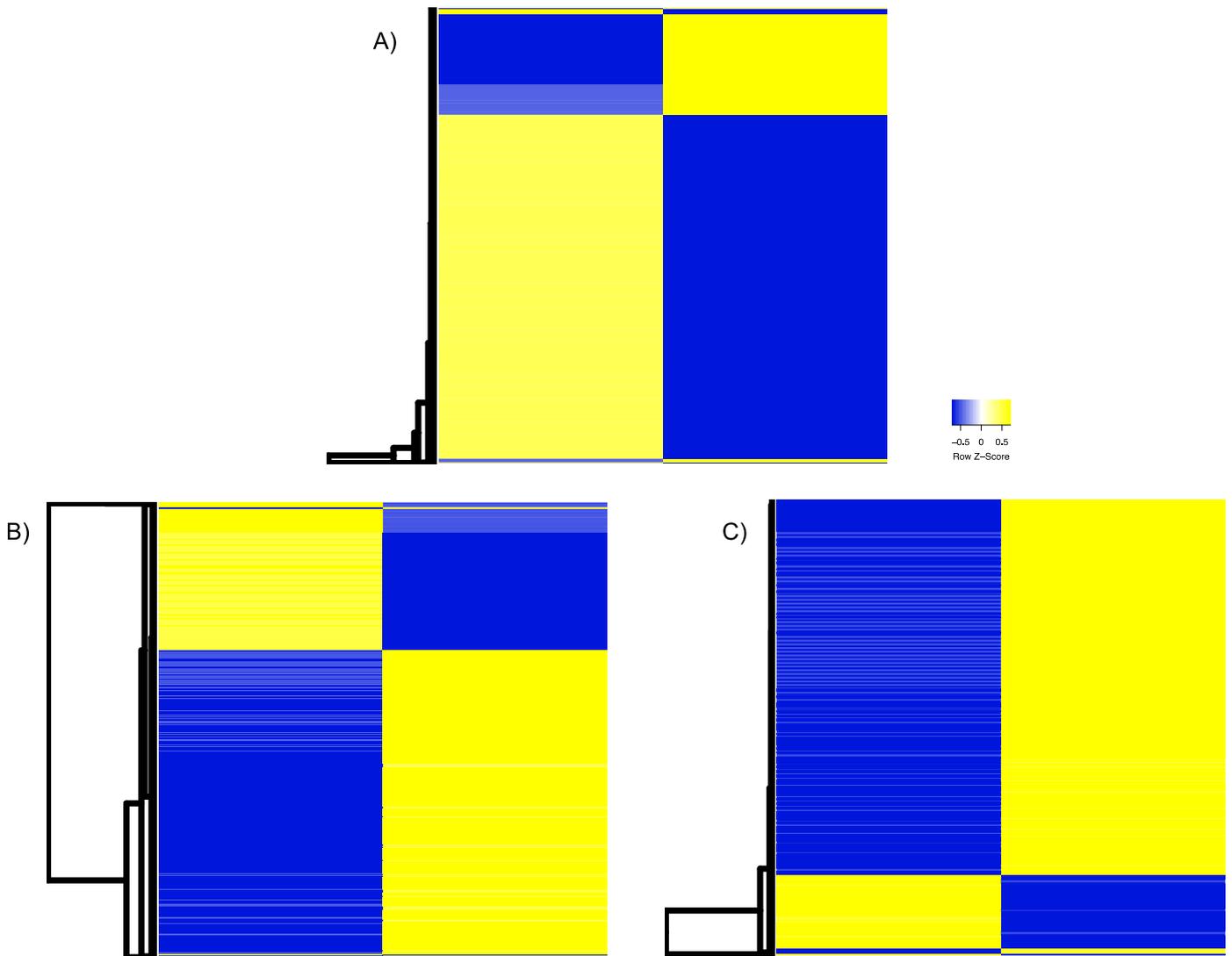


Figura 12. Heatmaps e agrupamentos hierárquicos dos DEGs. Estão representados os DEGs de operários (esquerda) e soldados (direita) de A) Ht, B) Cc e C) Vh. O tom de cor amarelo indica maiores níveis de expressão, enquanto o azul, uma menor expressão.

Como observado (Figura 13 e Tabela 9), as análises de anotação funcional dos DEGs revelaram para as três espécies que, dentro das categorias mais gerais de GO em *Biological Process* (BP), *metabolic process* e *cellular process* possuíram mais genes, enquanto *binding* e *catalytic activity* foram os principais dentro de *Molecular Function* (MF). Os termos GO foram atribuídos a 48,50%, 66,14%, 93,46%, 66,62%, 55,38% e 90,36% dos DEGs de Ht op, Ht sold, Cc op, Cc sold, Vh op e Vh sold, respectivamente. Interessante notar que, entre as duas espécies de cupins superiores, notamos ambos os soldados com foco em *cellular process* ao passo que, nos operários, mais DEGs foram anotados em *metabolic process*, evidenciando o efeito da expressão gênica nas principais funções de cada casta, ou vice-versa.

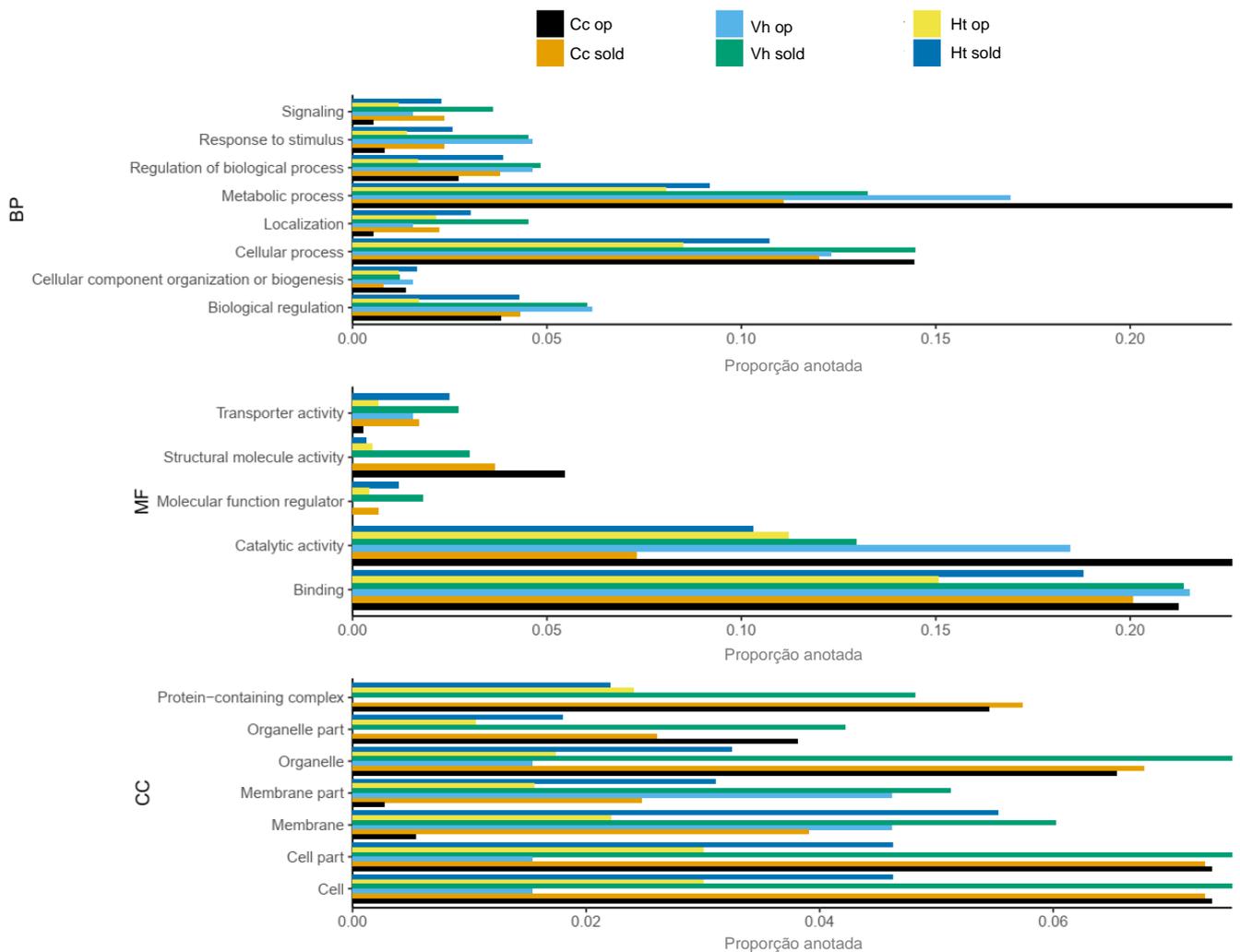


Figura 13. Relação da classificação de GO para os 20 genes mais diferencialmente expressos de cada casta. Estão representados os DEGs de operários e soldados de todas as três espécies de cupins, de acordo com a cor especificada na legenda. O eixo x indica a porcentagem de DEGs anotados enquanto o eixo Y indica os termos de GO de nível dois.

Também investigamos genes comuns e únicos entre castas e espécies (Figura 14). Após a filtragem manual de DEGs com anotações não específicas, de baixa qualidade e referentes a proteínas hipotéticas, obtivemos conjuntos de 505, 458, 119, 197, 27 e 101 DEGs para Ht op, Ht sold, Cc op, Cc sold, Vh op e Vh sold, respectivamente. Foi observado que a grande maioria dos DEGs é expressa exclusivamente em uma determinada casta e espécie, e nenhum gene estava presente simultaneamente nos seis perfis, como esperado. Entretanto, um total de 22 genes se encontram em uma posição onde fazem interseção dentro da própria espécie, ou seja, *up* regulados para ambos operários e soldados, o que é estranho, afinal, eles foram classificados como DEGs. Investigando melhor tais genes, foi notado que a maioria deles são casos onde tivemos a mesma anotação funcional, porém provenientes de diferentes endossimbiontes, ou até mesmo, em uma casta endógeno e em outra, exógeno. Além disso,

uma minoria se tratava de ORFs com CDSs parciais de um mesmo gene, porém uma em operário e a outra, no soldado.

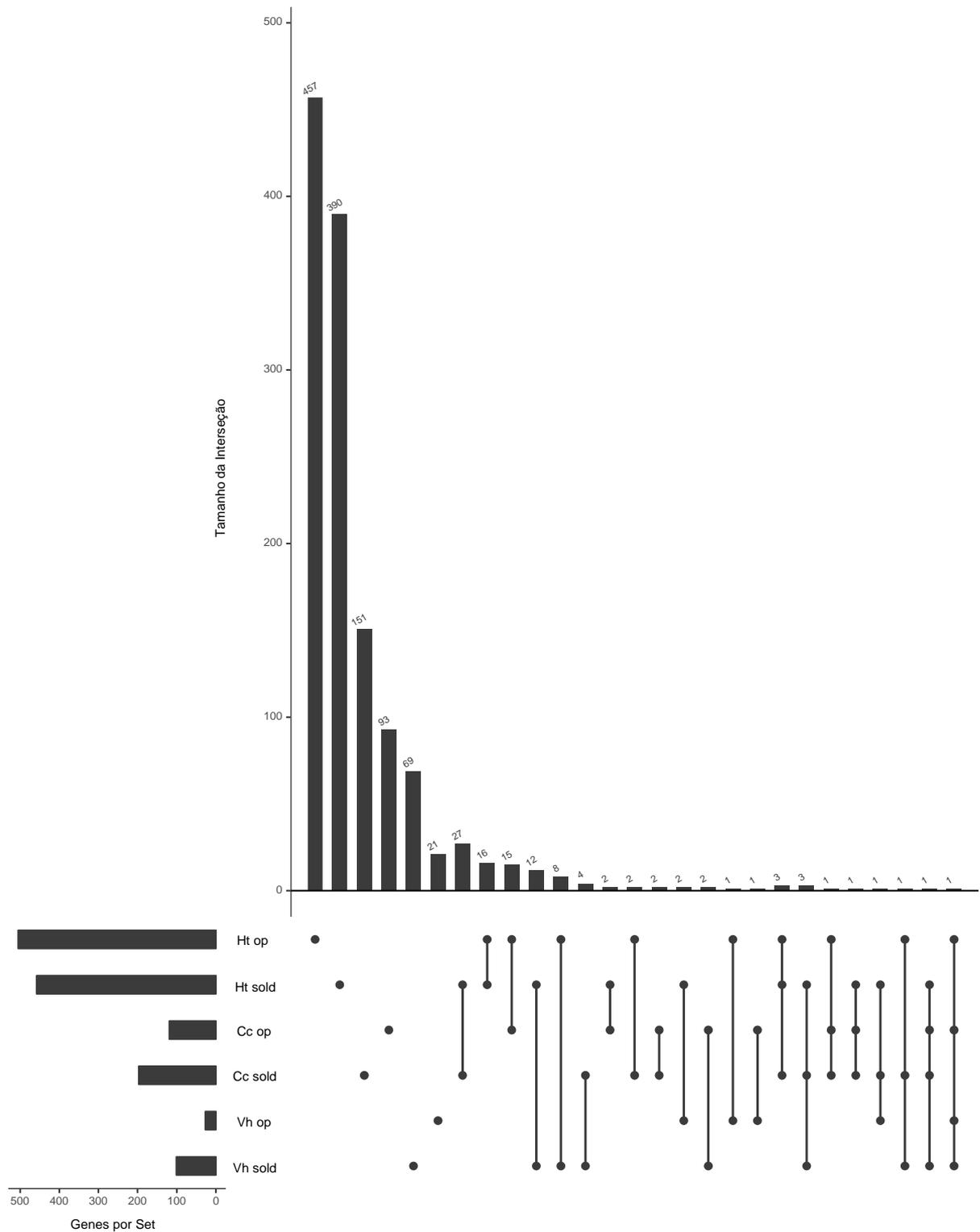


Figura 14. Comparação entre castas e espécies de genes comuns e únicos anotadas a partir dos DEGs. Cada interseção possível está representada na matriz localizada na parte de baixo, enquanto o gráfico de barras a sua esquerda exibe o tamanho de cada conjunto de DEGs. O número de genes em cada interseção é observado no topo, como gráfico principal.

Analisando a mesma casta entre as espécies, foram encontrados quatro genes em comum entre soldados de Cc e Vh, entre eles otoferlina e ferritina, e apenas um entre operários dessas espécies, a hexamerina II, sobre a qual discutiremos adiante. Nas interseções entre castas similares de Ht com Cc e Ht com Vh, tivemos em sua grande maioria, genes estruturais e genes de metabolismo energético básico. Ht e Cc mostraram uma maior sobreposição de genes, embora essas espécies sejam filogeneticamente mais distantes em comparação com Cc e Vh. Este resultado indica que as duas espécies superiores estudadas aqui são consideravelmente diferentes não apenas no conteúdo endossimbionte, mas também no conjunto de gene que é expresso de maneira diferente.

4.4 Genes ligados à diferenciação de castas e genes com potencial biotecnológico.

Foram selecionados os 10 principais DEGs de cada perfil, uma vez que nesses subconjuntos poderiam ser encontrados genes reguladores de castas. Entre esses DEGs, podemos ressaltar o gene hexamerina II (Hex-II) em operários de Cc e Vh e citocromo P450 (CYP) em soldados das três espécies, ambos envolvidos na diferenciação de castas em insetos sociais. Ademais, também foram encontrados genes celulolíticos endógenos de Ht op. A listagem destes genes pode ser melhor observada no Apêndice 1, encontrado ao final deste trabalho.

Os dois genes Hex dos cupins (I e II) são expressos predominantemente nos tecidos adiposos e uma de suas funções mais bem definidas é atuar no mecanismo de regulação de castas moduladas por hormônio juvenil (JH) (CORNETTE et al., 2013; PARK; RAINA, 2004). Com base no modelo proposto por ZHOU et al., 2006 para o cupim basal *R. flavipes*, quando as concentrações de JH excedem a capacidade de ligação/sequestro do sistema Hex, o JH interage com os receptores alvo e ocorre a morfogênese pré-soldado. Alternativamente, a expressão excessiva de Hex em condições com JH disponível resulta em sequestro eficaz do JH no corpo adiposo e, portanto, na inibição da morfogênese pré-soldado, a qual é a condição esperada em condições normais de colônia de status quo, ou seja, a manutenção de altas proporções da casta de operários (ZHOU et al., 2006). Isso é possível porque, nos cupins basais, a casta de operários é um estágio imaturo altamente plástico e totipotente. Não foram realizados estudos em espécies superiores, mas, embora exibam um desenvolvimento relativamente rígido com um status adulto de operários que impede sua diferenciação adicional, a alta expressão de Hex-II em operários de Cc e Vh, ambos amostrados de colônias de cupins bem estabelecidas, sugere que pode haver um mecanismo semelhante de manutenção da estrutura social de cupins envolvendo JH e HEX para as espécies superiores. Outra função bem definida de HEX é o armazenamento de nutrientes, e foi demonstrado que os níveis de HEX se correlacionam significativa e positivamente com o estado nutricional, *i.e.*,

como resultado direto da alimentação com celulose (SCHARF et al., 2007), portanto, faz sentido que esse gene seja *up* regulado em operários.

A regulação da concentração de JH em cupins tem sido associada não apenas à expressão dos genes Hex, mas também à expressão dos genes CYP (CORNETTE et al., 2006; TARVER; COY; SCHARF, 2012). Esses genes pertencem a grandes famílias de genes encontrados em todos os organismos e codificam enzimas oxidativas contendo grupo heme, as quais possuem várias funções em insetos, incluindo a biossíntese de hormônios e o catabolismo de toxinas (FEYEREISEN, 1999). Para os cupins basais *Hodotermopsis sjostedti* e *R. flavipes*, foi demonstrado que os genes CYP estavam envolvidos na biossíntese de JH, sendo significativamente *up* regulados durante a diferenciação do soldado (CORNETTE et al., 2006; HELVIG et al., 2004; TARVER; COY; SCHARF, 2012). Encontramos genes CYP da família 6 em soldados de Vh e Ht (CYP6j1) e da família 4 em soldados de Cc (CYP4c3), sendo essas duas as famílias CYP de insetos mais bem representadas (ZHOU; OI; SCHARF, 2006). Em *R. flavipes*, a expressão de um gene CYP6 foi associada a um baixo nível de desenvolvimento do corpo adiposo (CORNETTE et al., 2006), que é uma característica dos soldados de cupins e uma provável explicação de nossos resultados. Além disso, em *Drosophila melanogaster*, um gene CYP4 participa da última etapa da biossíntese de hidrocarbonetos cuticulares, que são cruciais para a comunicação química (QIU et al., 2012). Este gene também foi encontrado como cópia única nos cupins basais *Z. nevadensis* e *C. secundus*, enquanto que três cópias em *M. natalensis*, o que indica uma maior importância da síntese de hidrocarbonetos cuticulares em cupins superiores (HARRISON et al., 2018). A associação dos genes Hex, CYP e JH estão graficamente resumidas abaixo, na Figura 15.

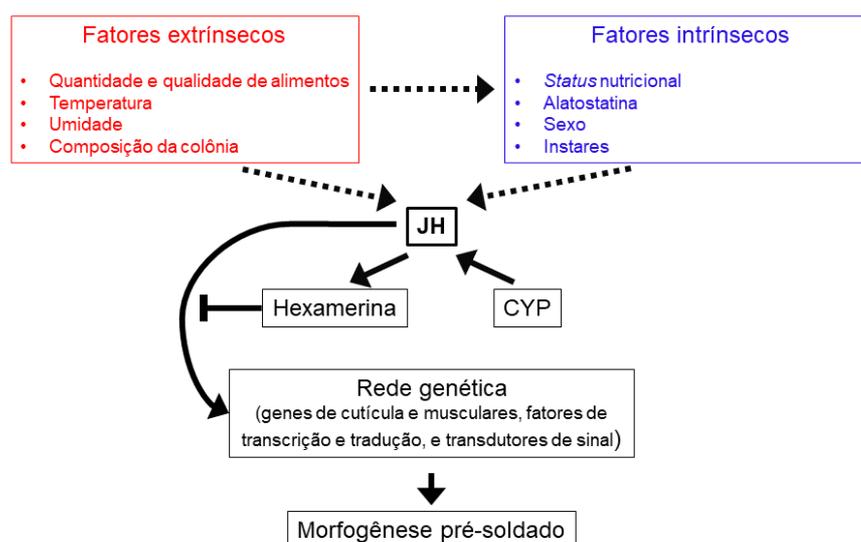


Figura 15. Diagrama representando rede de genes e fatores responsáveis ao JH. Setas sólidas representam o que é conhecido através de pesquisas anteriores e atuais, enquanto setas pontilhadas ilustram os relacionamentos não completamente compreendidos. Fonte: adaptado de ZHOU; TARVER; SCHARF, 2006.

Em operários de Ht, dois dos dez principais DEGs são genes celulolíticos endógenos, uma beta-glicosidase e uma alfa-amilase. A primeira é uma enzima limitadora de taxa para a hidrólise da celulose, que atua na etapa final da conversão da biomassa lignocelulósica para converter a celobiose em glicose, o produto final (SINGHANIA et al., 2013). Já a segunda, hidrolisa ligações alfa de polissacarídeos, como amido e glicogênio, produzindo glicose e maltose. Existe um crescente interesse biotecnológico em ambas as enzimas, principalmente devido ao seu uso na produção de biocombustíveis líquidos. Por exemplo, um dos principais desafios na bioconversão da biomassa lenho-celulósica em biocombustíveis líquidos é encontrar uma beta-glicosidase tolerante à glicose (SINGHANIA et al., 2013). Também devemos destacar o fato de que ambas as enzimas são endógenas da Ht, apesar de seu grande conteúdo de endossimbiontes. Mesmo em cupins basais, as enzimas derivadas do hospedeiro contribuem para a digestão da celulose e já foram descritas como expressas em níveis muito mais elevados do que qualquer enzima microbiana individual (SCHARF, 2014; TARTAR et al., 2009).

Os perfis de soldado mostraram, nos dez principais DEGs, *contigs* que não apresentam homologia com nenhuma proteína conhecida. Como existem dados de várias espécies de cupins disponíveis nos bancos de dados utilizados na anotação funcional, é possível hipotetizar que esses *contigs* não anotados têm maior probabilidade de pertencer à microbiota endossimbionte. A atribuição de genes de endossimbionte de cupins é difícil, pois vários filotipo de intestino de cupins vivem exclusivamente nesse ambiente e nunca foram avaliados, o que significa que suas sequências podem estar sub-representadas nos bancos de dados mais utilizados (SANTANA et al., 2015).

Por fim, mas não menos importante, realizamos uma breve análise exploratória dos genes com maiores níveis de expressão (em TPM) em cada um dos transcriptomas montados, a qual está listada em seis tabelas no Apêndice 2. A partir disto, era facilmente perceptível que genes estruturais são, via de regra, mais expressos, ainda assim como alguns genes do metabolismo e armazenamento de ferro e cálcio. Interessantemente, a enzima 1,4-beta-xylanase está entre as mais expressas de Ht op, enzima na qual existe um amplo interesse biotecnológico principalmente por sua relação com a produção de bioetanol (GRAY; ZHAO; EMPTAGE, 2006; TABKA et al., 2006) e está sendo investigada por nossos colaboradores. Os dados gerados aqui, também, são parte essencial para a progressão de linhas de pesquisa sendo desenvolvidas atualmente no Laboratório de Bioquímica e Genética Aplicada (DGE – UFSCar), como por exemplo, clonagem e expressão heteróloga de diferentes enzimas, e.g., alfa-amilase, endobeta-1,4-glucanase, cellobiohidrolase, betaglucosidase e lacase, todas provenientes de sequências obtidas das montagens de Ht.

5 CONCLUSÃO

Com o objetivo de identificar nos cupins auxiliares genes envolvidos na diferenciação de castas e genes com potencial biotecnológico, montamos os transcriptomas do cupim basal *Heterotermes tenuis* e de duas espécies pertencentes aos cupins superiores, *Cornitermes cumulans* e *Velocitermes heteropterus*. Comparamos o panorama metatranscriptômico de operários e soldados e encontramos diferenças importantes entre espécies basais e superiores, e inesperadamente, detectamos também simbiontes protistas nas duas espécies superiores, com grande destaque ao filo Amoebozoa, o qual não havia sido observado em grande quantidade em estudos prévios. A análise de expressão diferencial entre castas mostrou padrões opostos entre as espécies basais e as superiores, e a maioria dos genes anotados no conjunto de dados de DEGs era única de uma determinada casta ou espécie. Anotação funcional dos DEGs evidenciou diferenças nos processos biológicos com maior enfoque entre as castas operário e soldados das duas espécies superiores. Entre os 10 principais DEGs de cada casta e espécie, encontramos interessantes genes associados à diferenciação de casta e de capacidade hidrolítica, como hexamerinas, citocromo P450, beta-glicosidase e alfa-amilase. Nossos resultados fornecem informação para o desenvolvimento de tecnologias mais eficientes para controle de cupins e descrevem genes relevantes com potencial para aplicações industriais, em especial, para digestão de biomassa e produção de biocombustíveis. Este projeto foi desenvolvido em colaboração e já contribuiu para outros trabalhos, como uma dissertação de mestrado do PPGGEv e trabalhos de monografia.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABBY, S.; DAUBIN, V. Comparative genomics and the evolution of prokaryotes. **Trends in Microbiology**, v. 15, n. 3, p. 135–141, mar. 2007.

ABDUL RAHMAN, N. et al. A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. **Microbiome**, v. 3, n. 1, p. 5, dez. 2015.

ALMEIDA, J. E. M.; ALVES, S. B.; ALMEIDA, L. C. Controle de *Heterotermes tenuis* (Hagen) (Isoptera: Rhinotermitidae) e *Cornitermes cumulans* (Kollar) (Isoptera: Termitidae) com inseticida fipronil associado ao fungo entomopatogênico *Beauveria bassiana* (Bals.) vuill. em isca atrativa na cultura da cana-de. **Arquivos do Instituto de Biociências**, v. 67, n. 2, p. 235–241, 2000.

ANDREWS, S. **FastQC: a quality control tool for high throughput sequence data.**, 2010.

BABICKI, S. et al. Heatmapper: web-enabled heat mapping for all. **Nucleic Acids Research**, v. 44, n. W1, p. W147–W153, jul. 2016.

BEHL, S. et al. Caste-biased genes in a subterranean termite are taxonomically restricted: implications for novel gene recruitment during termite caste evolution. **Insectes Sociaux**, v. 65, n. 4, p. 593–599, nov. 2018.

BENJAMINO, J.; GRAF, J. Characterization of the Core and Caste-Specific Microbiota in the Termite, *Reticulitermes flavipes*. **Frontiers in microbiology**, v. 7, p. 171, 2016.

BIGNELL, D. E.; EGGLETON, P. Termites in Ecosystems. In: **Termites: Evolution, Sociality, Symbioses, Ecology**. Dordrecht: Springer Netherlands, 2000. p. 363–387.

BINNECK, E. As ômicas: Integrando a bioinformação. **Biotecnologia Ciencia e Desenvolvimento**, v. 32, n. janeiro/junho, p. 28–37, 2004.

BIONDO, C. J. Cupim de monte *Cornitermes cumulans* (Kollar, 1832): combate experimental com formulações inseticidas líquidas. **An. Esc. Super. Agric. Luiz de Queiroz**, p. 91–97, 1988.

BOOMSMA, J. J. Lifetime monogamy and the evolution of eusociality. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 364, n. 1533, p. 3191–3207, nov. 2009.

BREZNAK, J. A.; BRUNE, A. Role of Microorganisms in the Digestion of Lignocellulose by Termites. **Annual Review of Entomology**, v. 39, n. 1, p. 453–487, jan. 1994.

BRUNE, A. Symbiotic digestion of lignocellulose in termite guts. **Nature Reviews Microbiology**, v. 12, n. 3, p. 168–180, mar. 2014.

BRUNE, A.; DIETRICH, C. The Gut Microbiota of Termites: Digesting the Diversity in the Light of Ecology and Evolution. **Annual Review of Microbiology**, v. 69, n. 1, p. 145–166, out. 2015.

BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, v. 12, n. 1, p. 59–60, jan. 2015.

BUSHNELL, B. **BBTools software package.**, 2014.

CAMARGO-DIETRICH, C. R. R. DE; COSTA-LEONARDO, A. M. Comportamento intra-específico do cupim *Heterotermes tenuis* (Hagen) (Isoptera, Rhinotermitidae) em condições de laboratório. **Revista Brasileira de Zoologia**, v. 17, n. 2, p. 421–427, 2000.

COLES DE NEGRET, H. R.; REDFORD, K. H. The biology of nine termite species (Isoptera: Termitidae) from the cerrado of Central Brazil. **Psyche: A Journal of Entomology**, v. 89, n. 1–2, p. 81–106, 1982.

CONESA, A.; GÖTZ, S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. **International Journal of Plant Genomics**, v. 2008, p. 1–12, 2008.

CONSTANTINO, R. Termites (Insecta, Isoptera) from the lower Japurá River, Amazonas State, Brazil. **Boletim do Museu Paraense Emílio Goeldi Série Zoologia**, v. 7, n. 2, p. 189–224, 1991.

CONSTANTINO, R. Padrões de diversidade e endemismo de térmitas no bioma Cerrado. **Cerrado: ecologia, biodiversidade e conservação**. Brasília: Ministério do Meio Ambiente, p. 319–333, 2005.

CONWAY, J. R.; LEX, A.; GEHLENBORG, N. UpSetR: an R package for the visualization of intersecting sets and their properties. **Bioinformatics**, v. 33, n. 18, p. 2938–2940, set. 2017.

CORNETTE, R. et al. Caste-specific cytochrome P450 in the damp-wood termite *Hodotermopsis sjostedti* (Isoptera, Termopsidae). **Insect Molecular Biology**, v. 15, n. 2, p. 235–244, abr. 2006.

CORNETTE, R. et al. Differential gene expression in response to juvenile hormone analog treatment in the damp-wood termite *Hodotermopsis sjostedti* (Isoptera, Archotermopsidae). **Journal of Insect Physiology**, v. 59, n. 4, p. 509–518, abr. 2013.

CORONA, M.; LIBBRECHT, R.; WHEELER, D. E. Molecular mechanisms of phenotypic plasticity in social insects. **Current Opinion in Insect Science**, v. 13, p. 55–60, fev. 2016.

DAVIES, R. G. et al. Evolution of termite functional diversity: analysis and synthesis of local ecological and regional influences on local species richness. **Journal of Biogeography**, v. 30, n. 6, p. 847–877, jun. 2003.

DESSAU, R. B.; PIPPER, C. B. [“R”-project for statistical computing]. **Ugeskrift for laeger**, v. 170, n. 5, p. 328–30, jan. 2008.

DIETRICH, C.; KÖHLER, T.; BRUNE, A. The Cockroach Origin of the Termite Gut Microbiota: Patterns in Bacterial Community Structure Reflect Major Evolutionary Events. **Applied and Environmental Microbiology**, v. 80, n. 7, p. 2261–2269, abr. 2014.

EGGLETON, P. An Introduction to Termites: Biology, Taxonomy and Functional Morphology. In: BIGNELL, D. E.; ROISIN, Y.; LO, N. (Eds.). . **Biology of Termites: a Modern Synthesis**. Dordrecht: Springer Netherlands, 2010. p. 1–26.

ENGEL, M. S.; GRIMALDI, D. A.; KRISHNA, K. Termites (Isoptera): Their Phylogeny, Classification, and Rise to Ecological Dominance. **American Museum Novitates**, v. 2009, n. 3650, p. 1, jul. 2009.

FERNANDES, P. M.; ALVES, S. B. Preferência alimentar e danos de Cornitermes cumulans (Kollar, 1832)(Isoptera: Termitidae) às plantas cultivadas em laboratório. **Anais da Sociedade Entomológica do Brasil**, v. 21, p. 125–132, 1992.

FEYEREISEN, R. INSECT P450 ENZYMES. **Annual Review of Entomology**, v. 44, n. 1, p. 507–533, jan. 1999.

FRANCKE, C.; SIEZEN, R. J.; TEUSINK, B. Reconstructing the metabolic network of a bacterium from its genome. **Trends in Microbiology**, v. 13, n. 11, p. 550–558, nov. 2005.

GALLO, D. et al. *Entomologia agrícola*. 2002.

GENG, A. et al. Transcriptome analysis of the digestive system of a wood-feeding termite (*Coptotermes formosanus*) revealed a unique mechanism for effective biomass degradation. **Biotechnology for Biofuels**, v. 11, n. 1, p. 1–14, 2018.

GINALSKI, K. Comparative modeling for protein structure prediction. **Current Opinion in Structural Biology**, v. 16, n. 2, p. 172–177, abr. 2006.

GOMES DA SILVA MEDEIROS, L.; GOMES BANDEIRA, A.; MARTIUS, C. Termite Swarming in the Northeastern Atlantic Rain Forest of Brazil. **Studies on Neotropical Fauna and Environment**, v. 34, n. 2, p. 76–87, ago. 1999.

GRABHERR, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature biotechnology**, v. 29, n. 7, p. 644–52, maio 2011.

GRAY, K. A.; ZHAO, L.; EMPTAGE, M. Bioethanol. **Current Opinion in Chemical Biology**, v. 10, n. 2, p. 141–146, abr. 2006.

GRIECO, M. B. et al. Metagenomic Analysis of the Whole Gut Microbiota in Brazilian Termitidae Termites *Cornitermes cumulans*, *Cyrtillitermes strictinasus*, *Syntermes dirus*, *Nasutitermes jaraguae*, *Nasutitermes aquilinus*, *Grigiotermes bequaerti*, and *Orthognathotermes mirim*. **Current Microbiology**, v. 76, n. 6, p. 687–697, jun. 2019.

HAAS, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. **Nature Protocols**, v. 8, n. 8, p. 1494–1512, ago. 2013.

HAIFIG, I. et al. The size of excavators within a polymorphic termite species governs tunnel topology. **Animal Behaviour**, v. 82, n. 6, p. 1409–1414, dez. 2011.

HAIFIG, I. et al. On the Apterous Line of the Termite *Velocitermes heteropterus* (Isoptera: Termitidae): Developmental Pathways and Cellulose Digestion. **Zoological Science**, v. 29, n. 12, p. 815–820, dez. 2012.

HARRISON, M. C. et al. Hemimetabolous genomes reveal molecular basis of termite eusociality. **Nature Ecology & Evolution**, v. 2, n. 3, p. 557–566, mar. 2018.

HATTORI, A. et al. Soldier Morphogenesis in the Damp-Wood Termite Is Regulated by the Insulin Signaling Pathway. **Journal of Experimental Zoology Part B: Molecular and Developmental Evolution**, v. 320, n. 5, p. 295–306, jul. 2013.

HE, S. et al. Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites. **PLoS ONE**, v. 8, n. 4, p. e61126, abr. 2013.

HELVIG, C. et al. CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata. **Proceedings of the National Academy of Sciences**, v. 101, n. 12, p. 4024–4029, mar. 2004.

HONGO, Y. et al. Phylogenetic Diversity, Localization, and Cell Morphologies of Members of the Candidate Phylum TG3 and a Subphylum in the Phylum Fibrobacteres, Recently Discovered Bacterial Groups Dominant in Termite Guts. **Applied and Environmental Microbiology**, v. 72, n. 10, p. 6780–6788, out. 2006.

HONGO, Y.; OHKUMA, M.; KUDO, T. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). **FEMS Microbiology Ecology**, v. 44, n. 2, p. 231–242, maio 2003.

HUANG, X. CAP3: A DNA Sequence Assembly Program. **Genome Research**, v. 9, n. 9, p. 868–877, set. 1999.

HUERTA-CEPAS, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. **Nucleic Acids Research**, v. 47, n. D1, p. D309–D314, jan. 2019.

HUSON, D. H. et al. MEGAN analysis of metagenomic data. **Genome research**, v. 17, n. 3, p. 377–86, mar. 2007.

HUSON, D. H. et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. **PLOS Computational Biology**, v. 12, n. 6, p. e1004957, jun. 2016.

INWARD, D.; BECCALONI, G.; EGGLETON, P. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. **Biology Letters**, v. 3, n. 3, p. 331–335, jun. 2007.

JONGEPIER, E. et al. Remodeling of the juvenile hormone pathway through caste-biased gene expression and positive selection along a gradient of termite eusociality. **Journal of Experimental Zoology Part B: Molecular and Developmental Evolution**, v. 330, n. 5, p. 296–304, jul. 2018.

KLASS, K.-D.; NALEPA, C.; LO, N. Wood-feeding cockroaches as models for termite evolution (Insecta: Dictyoptera): *Cryptocercus* vs. *Parasphaeria boleiriana*. **Molecular Phylogenetics and Evolution**, v. 46, n. 3, p. 809–817, mar. 2008.

KORB, J. Termites, hemimetabolous diploid white ants? **Frontiers in Zoology**, v. 5, n. 1, p. 15, 2008.

KORB, J. et al. A genomic comparison of two termites with different social complexity. **Frontiers in Genetics**, v. 6, n. MAR, mar. 2015.

KORB, J.; BELLES, X. Juvenile hormone and hemimetabolan eusociality: a comparison of cockroaches with termites. **Current Opinion in Insect Science**, v. 22, p. 109–116, ago. 2017.

KRISHNAN, M. et al. Insect gut microbiome – An unexploited reserve for biotechnological application. **Asian Pacific Journal of Tropical Biomedicine**, v. 4, p. S16–S21, maio 2014.

KRUEGER, F. **Trim galore!**, 2015.

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

LEE, D.; REDFERN, O.; ORENGO, C. Predicting protein function from sequence and structure. **Nature Reviews Molecular Cell Biology**, v. 8, n. 12, p. 995–1005, dez. 2007.

LEGENDRE, F. et al. The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for the evolution of the worker and pseudergate castes, and foraging behaviors. **Molecular Phylogenetics and Evolution**, v. 48, n. 2, p. 615–627, 2008.

LEGENDRE, F.; GRANDCOLAS, P. The evolution of sociality in termites from cockroaches: A taxonomic and phylogenetic perspective. **Journal of Experimental Zoology Part B: Molecular and Developmental Evolution**, v. 330, n. 5, p. 279–287, jul. 2018.

LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC bioinformatics**, v. 12, p. 323, ago. 2011.

LIU, N. et al. Metagenomic Insights into Metabolic Capacities of the Gut Microbiota in a Fungus-Cultivating Termite (*Odontotermes yunnanensis*). **PLoS ONE**, v. 8, n. 7, p. e69184, jul. 2013.

LO, N. et al. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. **Current Biology**, v. 10, n. 13, p. 801–804, jun. 2000.

LO, N.; EGGLETON, P. Termite Phylogenetics and Co-cladogenesis with Symbionts. In: BIGNELL, D. E.; ROISIN, Y.; LO, N. (Eds.). . **Biology of Termites: a Modern Synthesis**. Dordrecht: Springer Netherlands, 2010. p. 27–50.

LO, N.; TOKUDA, G.; WATANABE, H. Evolution and Function of Endogenous Termite Cellulases. In: BIGNELL, D. E.; ROISIN, Y.; LO, N. (Eds.). . **Biology of Termites: a Modern Synthesis**. Dordrecht: Springer Netherlands, 2010. p. 51–67.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature Reviews Genetics**, v. 12, n. 10, p. 671–682, 7 out. 2011.

MARX, V. Biology: The big challenges of big data. **Nature**, v. 498, n. 7453, p. 255–260, 2013.

MASUOKA, Y. et al. A Crucial Caste Regulation Gene Detected by Comparing Termites and Sister Group Cockroaches. **Genetics**, v. 209, n. 4, p. 1225–1234, ago. 2018.

MATHEWS, A. G. A. **Studies on termites from the Mato Grosso state, Brazil**. [s.l.] Academia Brasileira de Ciências Rio de Janeiro, 1977.

MATZ, M. et al. Amplification of cDNA ends based on template-switching effect and step-out PCR. **Nucleic Acids Research**, v. 27, n. 6, p. 1558–1560, 1999.

METZKER, M. L. Sequencing technologies — the next generation. **Nature Reviews Genetics**, v. 11, n. 1, p. 31–46, 8 jan. 2010.

MIKAELIAN, A. et al. The fibre-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). **Environmental Microbiology**, v. 16, n. 9, p. 2711–2722, set. 2014.

MILL, A. E. Termites as Agricultural Pests in Amazônia, Brazil. **Outlook on Agriculture**, v. 21, n. 1, p. 41–46, mar. 1992.

MIRABITO, D.; ROSENGAUS, R. B. A double-edged sword? The cost of proctodeal trophallaxis in termites. **Insectes Sociaux**, v. 63, n. 1, p. 135–141, fev. 2016.

MITAKA, Y. et al. Caste-specific and sex-specific expression of chemoreceptor genes in a termite. **PLoS ONE**, v. 11, n. 1, p. 1–16, 2016.

MIURA, T.; SCHARF, M. E. Molecular Basis Underlying Caste Differentiation in Termites. In: **Biology of Termites: a Modern Synthesis**. Dordrecht: Springer Netherlands, 2010. p. 211–253.

NAKASHIMA, K. et al. Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. **Insect Biochemistry and Molecular Biology**, v. 32, n. 7, p. 777–784, jul. 2002.

NALEPA, C. What Kills the Hindgut Flagellates of Lower Termites during the Host Molting Cycle? **Microorganisms**, v. 5, n. 4, p. 82, dez. 2017.

NALEPA, C. A. Origin of termite eusociality: trophallaxis integrates the social, nutritional, and microbial environments. **Ecological Entomology**, v. 40, n. 4, p. 323–335, ago. 2015.

NEVES, P. J.; ALVES, S. B. Controle associado de *Cornitermes cumulans* (Kollar, 1832) (Isoptera: Termitidae) com *Metarhizium anisopliae*, *Beauveria bassiana* e imidacloprid. **Scientia Agricola**, v. 56, n. 2, p. 305–311, 1999.

NI, J.; TOKUDA, G. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. **Biotechnology Advances**, v. 31, n. 6, p. 838–850, nov. 2013.

OLIVEIRA, D. E. DE. Sistemática do grupo *Velocitermes* (Isoptera, Termitidae, Nasutitermitinae). 2013.

ONO, H. et al. Removal of redundant contigs from de novo RNA-Seq assemblies via homology search improves accurate detection of differentially expressed genes. **BMC genomics**, v. 16, p. 1031, dez. 2015.

OTANI, S. et al. Gut microbial compositions mirror caste-specific diets in a major lineage of social insects. **Environmental Microbiology Reports**, v. 11, n. 2, p. 196–205, abr. 2019.

OZSOLAK, F.; MILOS, P. M. RNA sequencing: advances, challenges and opportunities. **Nature Reviews Genetics**, v. 12, n. 2, p. 87–98, fev. 2011.

PARK, Y. I.; RAINA, A. K. Juvenile hormone III titers and regulation of soldier caste in *Coptotermes formosanus* (Isoptera: Rhinotermitidae). **Journal of Insect Physiology**, v. 50, n. 6, p. 561–566, jun. 2004.

PATTERSON, S. D.; AEBERSOLD, R. H. Proteomics: the first decade and beyond. **Nature Genetics**, v. 33, n. S3, p. 311–323, mar. 2003.

PEREZ FILHO, O. et al. **Observações sobre a estrutura do cupinzeiro de montículo visando o seu controle através da fumigação.** CONGRESSO BRASILEIRO DE ENTOMOLOGIA. **Anais...**1987

PEVSNER, J. **Bioinformatics and functional genomics.** [s.l.] John Wiley & Sons, 2015.

POULSEN, M. et al. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. **Proceedings of the National Academy of Sciences**, v. 111, n. 40, p. 14500–14505, out. 2014.

PRESTWICH, G. D. Defense Mechanisms of Termites. **Annual Review of Entomology**, v. 29, n. 1, p. 201–232, jan. 1984.

QIU, L. et al. Biomimetic superelastic graphene-based cellular monoliths. **Nature Communications**, v. 3, n. 1, p. 1241, jan. 2012.

RADEK, R.; TISCHENDORF, G. Bacterial adhesion to different termite flagellates: Ultrastructural and functional evidence for distinct molecular attachment modes. **Protoplasma**, v. 207, n. 1–2, p. 43–53, mar. 1999.

REDFORD, K. H. The Termitaria of *Cornitermes cumulans* (Isoptera, Termitidae) and Their Role in Determining a Potential Keystone Species. **Biotropica**, v. 16, n. 2, p. 112, jun. 1984.

RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: The European Molecular Biology Open Software Suite. **Trends in Genetics**, v. 16, n. 6, p. 276–277, jun. 2000.

ROISIN, Y.; KORB, J. Social Organisation and the Status of Workers in Termites. In: **Biology of Termites: a Modern Synthesis**. Dordrecht: Springer Netherlands, 2010. p. 133–164.

ROULAND-LEFÈVRE, C. Termites as Pests of Agriculture. In: **Biology of Termites: a Modern Synthesis**. Dordrecht: Springer Netherlands, 2010. p. 499–517.

SANTANA, R. H. et al. The Gut Microbiota of Workers of the Litter-Feeding Termite *Syntermes wheeleri* (Termitidae: Syntermitinae): Archaeal, Bacterial, and Fungal Communities. **Microbial Ecology**, v. 70, n. 2, p. 545–556, ago. 2015.

SCHARF, M. E. et al. Regulation of polyphenic caste differentiation in the termite *Reticulitermes flavipes* by interaction of intrinsic and extrinsic factors. **Journal of Experimental Biology**, v. 210, n. 24, p. 4390–4398, dez. 2007.

SCHARF, M. E. Termites as Targets and Models for Biotechnology. **Annual review of entomology**, n. October, p. 1–26, 2014.

SCHARF, M. E. Omic research in termites: An overview and a roadmap. **Frontiers in Genetics**, v. 6, n. MAR, 2015.

SCHARF, M. E.; BOUCIAS, D. G. Potential of termite-based biomass pre-treatment strategies for use in bioethanol production. **Insect Science**, v. 17, n. 3, p. 166–174, 2010.

SETHI, A. et al. Lignin-associated metagene expression in a lignocellulose-digesting termite. **Insect Biochemistry and Molecular Biology**, v. 43, n. 1, p. 91–101, jan. 2013.

SHIMADA, K. et al. Cellulolytic Protist Numbers Rise and Fall Dramatically in Termite Queens and Kings during Colony Foundation. **Eukaryotic Cell**, v. 12, n. 4, p. 545–550, abr. 2013.

SHINZATO, N. et al. Molecular Phylogenetic Diversity of the Bacterial Community in the Gut of the Termite *Coptotermes formosanus*. **Bioscience, Biotechnology, and Biochemistry**, v. 69, n. 6, p. 1145–1155, jan. 2005.

SILLAM-DUSSÈS, D. et al. Comparative Study of the Labial Gland Secretion in Termites (Isoptera). **PLoS ONE**, v. 7, n. 10, p. e46431, out. 2012.

SIMÃO, F. A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics (Oxford, England)**, v. 31, n. 19, p. 3210–2, out. 2015.

SINGHANIA, R. R. et al. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. **Bioresource Technology**, v. 127, p. 500–507, jan. 2013.

ŠOBOTNÍK, J.; DAHLSJÖ, C. A. L. Isoptera ☆. In: **Reference Module in Life Sciences**. [s.l.] Elsevier, 2017.

SU, N.-Y.; SCHEFFRAHN, R. H. Termites as Pests of Buildings. In: **Termites: Evolution, Sociality, Symbioses, Ecology**. Dordrecht: Springer Netherlands, 2000. p. 437–453.

TABKA, M. G. et al. Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. **Enzyme and Microbial Technology**, v. 39, n. 4, p. 897–902, 2006.

TARAZONA, S. et al. Differential expression in RNA-seq: A matter of depth. **Genome Research**, v. 21, n. 12, p. 2213–2223, dez. 2011.

TARTAR, A. et al. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. **Biotechnology for Biofuels**, v. 2, n. 1, p. 25, 2009.

TARVER, M. R.; COY, M. R.; SCHARF, M. E. Cyp15F1: A NOVEL CYTOCHROME P450 GENE LINKED TO JUVENILE HORMONE-DEPENDENT CASTE DIFFERENTIATION IN THE TERMITE *Reticulitermes flavipes*. **Archives of Insect Biochemistry and Physiology**, v. 80, n. 2, p. 92–108, jul. 2012.

TERRAPON, N. et al. Molecular traces of alternative social organization in a termite genome. **Nature Communications**, v. 5, n. 1, p. 3636, maio 2014.

THONGARAM, T. et al. Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. **Extremophiles**, v. 9, n. 3, p. 229–238, jun. 2005.

TOGA, K. et al. Sexual difference in juvenile-hormone titer in workers leads to sex-biased soldier differentiation in termites. **Journal of Insect Physiology**, v. 87, p. 63–70, abr. 2016.

ULYSHEN, M. D. Wood decomposition as influenced by invertebrates. **Biological Reviews**, v. 91, n. 1, p. 70–85, fev. 2016.

VALÉRIO, J. R. Cupins-de-montículo em pastagens. **Embrapa Gado de Corte-Documentos (INFOTECA-E)**, 2006.

VERMA, M.; SHARMA, S.; PRASAD, R. Biological alternatives for termite control: A review. **International Biodeterioration and Biodegradation**, v. 63, n. 8, p. 959–972, 2009.

WAGNER, G. P.; KIN, K.; LYNCH, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. **Theory in biosciences = Theorie in den Biowissenschaften**, v. 131, n. 4, p. 281–5, dez. 2012.

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

WATANABE, H.; TOKUDA, G. Animal cellulases. **Cellular and Molecular Life Sciences**, v. 58, n. 9, p. 1167–1178, ago. 2001.

WATERHOUSE, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. **Molecular biology and evolution**, v. 35, n. 3, p. 543–548, dez. 2017.

WILSON, E. O. The insect societies. **The insect societies.**, 1971.

WILSON, E. O. The effects of complex social life on evolution and biodiversity. **Oikos**, p. 13–18, 1992.

WU, L.-W. et al. Phylogenetic position of the enigmatic termite family Stylotermitidae (Insecta : Blattodea). **Invertebrate Systematics**, v. 32, n. 5, p. 1111, 2018.

WU, T.; DHAMI, G. K.; THOMPSON, G. J. Soldier-biased gene expression in a subterranean termite implies functional specialization of the defensive caste. **Evolution & Development**, v. 20, n. 1, p. 3–16, jan. 2018.

ZANETTI, R. et al. Manejo integrado de cupins. **Lavras: UFLA**, 2002.

ZHOU, X. et al. Two hexamerin genes from the termite *Reticulitermes flavipes*: Sequence, expression, and proposed functions in caste regulation. **Gene**, v. 376, n. 1, p. 47–58, jul. 2006.

ZHOU, X. et al. Inhibition of termite cellulases by carbohydrate-based cellulase inhibitors: Evidence from in vitro biochemistry and in vivo feeding studies. **Pesticide Biochemistry and Physiology**, v. 90, n. 1, p. 31–41, jan. 2008a.

ZHOU, X. et al. RNA interference in the termite *Reticulitermes flavipes* through ingestion of double-stranded RNA. **Insect Biochemistry and Molecular Biology**, v. 38, n. 8, p. 805–815, ago. 2008b.

ZHOU, X.; OI, F. M.; SCHARF, M. E. Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. **Proceedings of the National Academy of Sciences**, v. 103, n. 12, p. 4499–4504, mar. 2006.

ZHOU, X.; TARVER, M. R.; SCHARF, M. E. Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. **Development**, v. 134, n. 3, p. 601–610, 21 dez. 2006.

APÊNDICE 1 – Lista de tabelas dos 10 genes mais diferencialmente expressos em casa casta dos cupins estudados. Esses dados foram inferidos pela ferramenta NOISeq-sim, onde M = razão log2 das duas condições, D = valor absoluto do valor de expressão entre as condições e Prob. = probabilidade de expressão diferencial.

Tabela 1. Lista dos 10 genes mais diferencialmente expressos de Ht op.

Genes Ht op	Média op	Média sold	M	D	Prob. (%)
serine protease	7264,763	12,817	9,147	7251,946	0,9647
serine protease	2810,806	6,170	8,831	2804,636	0,9647
serine protease	1135,800	1,831	9,277	1133,969	0,9647
beta-glucosidase	267,759	0,222	10,236	267,537	0,9986
ribosomal protein L35Ae	74,965	0,005	14,024	74,961	0,9998
unnamed protein product	36,167	0,005	12,972	36,163	0,9991
hypothetical protein L798_07788, partial	28,527	0,005	12,630	28,523	0,9986
Calmodulin	27,901	0,005	12,598	27,897	0,9985
Nascent polypeptide-associated complex subunit alpha	27,738	0,005	12,590	27,734	0,9985
lipase 3-like	27,661	0,005	12,586	27,657	0,9985

Tabela 2. Lista dos 10 genes mais diferencialmente expressos de Ht sold.

Genes Ht sold	Média op	Média sold	M	D	Prob. (%)
50MGP	0,574	988,986	-10,751	988,412	0,9998
PDZ and LIM domain protein Zasp isoform X7	0,005	69,113	-13,907	69,109	0,9997
elongation of very long chain fatty acids protein AAEL008004-like isoform X2	0,005	62,594	-13,764	62,590	0,9997
PDZ and LIM domain protein 3 isoform X3	0,005	57,089	-13,631	57,085	0,9996
5'-AMP-activated protein kinase subunit gamma-2	0,005	39,630	-13,104	39,626	0,9992
serine protease, partial	0,005	32,306	-12,810	32,302	0,9989
PDZ and LIM domain protein Zasp isoform X7	0,009	26,999	-11,551	26,990	0,9984
ion transport peptide isoform X2	0,005	22,871	-12,311	22,867	0,9979
uncharacterized protein LOC111865514	0,005	22,357	-12,279	22,353	0,9979
cytochrome P450 6j1-like	0,005	21,938	-12,251	21,934	0,9977

Tabela 3. Lista dos 10 genes mais diferencialmente expressos de *Cc op.*

Genes <i>Cc op</i>	Média <i>op</i>	Média <i>sold</i>	M	D	Prob. (%)
Cr-P11 allergen, partial	19204,137	19,384	9,952	19184,753	0,9859
uncharacterized protein LOC110839463	14232,340	15,841	9,811	14216,499	0,9859
major allergen Bla g 1.0101, partial	10605,337	5,464	10,923	10599,873	0,9992
major allergen Cr-P11	4802,732	2,424	10,952	4800,308	0,9992
Cuticle protein 2	2931,932	0,925	11,630	2931,007	1,0000
hexamerin II	960,019	0,373	11,330	959,646	0,9999
ejaculatory bulb-specific protein 3	345,144	0,004	16,589	345,141	0,9998
uncharacterized protein LOC110839463	338,492	0,004	16,561	338,489	0,9998
cuticle protein 7-like	188,281	0,004	15,715	188,278	0,9995
cuticle protein 7-like	176,707	0,004	15,624	176,704	0,9994

Tabela 4. Lista dos 10 genes mais diferencialmente expressos de *Cc sold.*

Genes <i>Cc sold</i>	Média <i>op</i>	Média <i>sold</i>	M	D	Prob. (%)
uncharacterized protein LOC111870392	0,004	415,321	-16,857	415,318	0,9998
alpha-tocopherol transfer protein-like	0,004	405,761	-16,823	405,758	0,9998
unknown	0,004	107,482	-14,906	107,479	0,9990
phospholipid transfer protein C2CD2L isoform X6	0,004	55,032	-13,941	55,029	0,9977
uncharacterized protein LOC111861799 isoform X2	0,004	50,593	-13,819	50,590	0,9974
uncharacterized protein LOC110829859	0,004	32,902	-13,199	32,899	0,9950
serine protease gd-like isoform X2	0,004	28,534	-12,993	28,531	0,9939
serine protease	0,004	27,846	-12,958	27,843	0,9935
cytochrome P450 4c3-like	0,004	25,766	-12,846	25,763	0,9928
high affinity copper uptake protein 1-like isoform X1	0,004	23,859	-12,735	23,856	0,9919

Tabela 5. Lista dos 10 genes mais diferencialmente expressos de *Vh op.*

Genes <i>Vh op</i>	Média <i>op</i>	Média <i>sold</i>	M	D	Prob. (%)
hexamerin II	2146,969	0,769	11,447	2146,200	1,0000
uncharacterized protein LOC110839463	39,389	0,005	13,096	39,385	0,9964
endonuclease V-like isoform X1	19,963	0,005	12,115	19,959	0,9899
S-phase kinase-associated protein 2	19,391	0,005	12,073	19,387	0,9899
hypothetical protein B7P43_G07257	10,937	0,005	11,247	10,933	0,9724
collagen alpha-2(IV) chain-like isoform X1	10,742	0,005	11,221	10,738	0,9724
transmembrane protein 135-like isoform X1	10,113	0,005	11,134	10,109	0,9724
toll-like receptor 2	9,919	0,005	11,106	9,915	0,9673
uncharacterized protein LOC110831486	8,523	0,005	10,887	8,519	0,9608
hypothetical protein B7P43_G12213	8,226	0,005	10,836	8,222	0,9608

Tabela 6. Lista dos 10 genes mais diferencialmente expressos de Vh sold.

Genes Vh sold	Média op	Média sold	M	D	Prob. (%)
Cytochrome P450 6j1, partial	0,412	265,318	-9,331	264,906	0,9626
cytochrome P450 6j1-like	0,005	256,183	-15,797	256,179	0,9998
uncharacterized protein LOC111871152	0,005	241,017	-15,709	241,013	0,9998
tubulin beta chain-like isoform X3	0,005	59,361	-13,687	59,357	0,9982
PREDICTED: histone H4-like	0,005	38,828	-13,075	38,824	0,9963
Histone H2A	0,005	25,541	-12,471	25,537	0,9932
PDZ and LIM domain protein 3 isoform X3	0,005	23,356	-12,342	23,352	0,9923
hypothetical protein	0,005	21,573	-12,227	21,569	0,9912
zinc metalloproteinase	0,005	17,001	-11,883	16,997	0,9881
piRNA biogenesis protein EXD1	0,005	15,559	-11,756	15,555	0,9854

APÊNDICE 2 – Lista de tabelas dos 10 genes mais expressos, em valores de TPM, para cada uma das seis montagens por castas.

Tabela 1. Lista dos 10 genes mais expressos de Ht op.

Genes Ht op	TPM
U2 small nuclear ribonucleoprotein A'-like	16417
actin	7412
alpha-tubulin 1	6590
transmembrane protease serine 9-like	6354
beta-tubulin	6150
transmembrane protease serine 9-like	5329
NADH dehydrogenase subunit 4 (mitochondrion)	4294
epididymal secretory protein E1-like	4192
1,4-beta-xylanase	3889
profilin	3677

Tabela 2. Lista dos 10 genes mais expressos de Ht sold.

Genes Ht sold	TPM
U2 small nuclear ribonucleoprotein A'-like	22561
myosin regulatory light chain 2	16191
actin related protein 1	15942
myosin light chain alkali isoform X2	11869
trypsin beta-like	9706
ferritin heavy polypeptide-like 17	9006
translationally-controlled tumor protein homolog	7025
ADP,ATP carrier protein	6928
ferritin subunit	6726
lysozyme-like	6618

Tabela 3. Lista dos 10 genes mais expressos de Cc op.

Genes Cc op	TPM
actin related protein 1	13696
transmembrane protease serine 9-like	9784
actin, cytoplasmic 2	8497
58rachyuran-like	6932
transferrin	6902
trypsin alpha-like	6671
von Willebrand factor type A, EGF and pentraxin domain	6239
ejaculatory bulb-specific protein 3-like	4646
tubulin alpha-1 chain	3898
lysozyme C-like	3746

Tabela 4. Lista dos 10 genes mais expressos de *Cc sold*.

Genes Cc sold	TPM
actin related protein 1	33610
myosin regulatory light chain 2	10301
troponin C-like	7539
ferritin subunit isoform X1	7251
ejaculatory bulb-specific protein 3-like	6278
ADP,ATP carrier protein	5890
myosin light chain alkali isoform X1	5559
ferritin heavy polypeptide-like 17	4897
tropomyosin isoform X14	4819
calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type	4656

Tabela 5. Lista dos 10 genes mais expressos de *Vh op*.

Genes Vh op	TPM
ferritin heavy polypeptide-like 17	10146
actin related protein 1	9397
ferritin subunit	8865
15-hydroxyprostaglandin dehydrogenase [NAD ⁺]-like	6843
proteoglycan 4-like	5897
myosin regulatory light chain 2	5391
chondroitin proteoglycan 1-like, partial	5353
tubulin alpha-1 chain	4741
actin, cytoplasmic 2	4662
translationally-controlled tumor protein homolog	4659

Tabela 6. Lista dos 10 genes mais expressos de *Vh sold*.

Genes Vh sold	TPM
ejaculatory bulb-specific protein 3-like	9949
ferritin heavy polypeptide-like 17	9849
ferritin subunit	8375
actin related protein 1	5650
actin, cytoplasmic 2	5625
elongation factor 1-alpha	4665
translationally-controlled tumor protein homolog	4138
tubulin alpha-1 chain	3896
15-hydroxyprostaglandin dehydrogenase [NAD ⁺]-like	3744
ADP,ATP carrier protein	3693