

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



Flávia Costa Leonardo

**“ESTUDO DO CUPIM *COPTOTERMES GESTROI*: ANÁLISE DE
GENES DIFERENCIALMENTE EXPRESSOS ENTRE CASTAS E
BUSCA DE GENES DE IMPORTÂNCIA BIOTECNOLÓGICA”**

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
Flávia Costa Leonardo
e aprovada pela Comissão Julgadora.

Tese apresentada ao Instituto de
Biologia para obtenção do Título de
Doutor em Genética e Biologia
Molecular na área de Genética Animal e
Evolução.

Orientador: Prof. Dr. Gonçalo Amarante Guimarães Pereira

Co-Orientador: Prof. Dr. Fernando Ferreira Costa

Campinas, 2010

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP**

L553e

Leonardo, Flávia Costa

Estudo do cupim *Coptotermes gestroi*: análise de genes diferencialmente expressos entre castas e busca de genes de importância biotecnológica / Flávia Costa Leonardo – Campinas, SP: [s.n.], 2010.

Orientadores: Gonçalo Amarante Guimarães Pereira, Fernando Ferreira Costa.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Termite. 2. Biblioteca subtrativa suprimida (SSH). 3. *Coptotermes gestroi*. 4. Celulase. I. Pereira, Gonçalo Amarante Guimarães. II. Costa, Fernando Ferreira. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Título em inglês: Study of the termite *Coptotermes gestroi*: analysis of differentially expressed genes between castes and search for genes with biotechnological applicability.

Palavras-chave em inglês: Termites; Suppression subtractive hybridization (SSH); *Coptotermes gestroi*; Cellulase.

Área de concentração: Genética Animal e Evolução.

Titulação: Doutora em Genética e Biologia Molecular.

Banca examinadora: Gonçalo Amarante Guimarães Pereira, Zilá Luz Paulino Simões, Klaus Hartmann Hartfelder, Mônica Barbosa de Melo, Alberto Arab.

Data da defesa: 15/12/2010.

Programa de Pós-Graduação: Genética e Biologia Molecular.

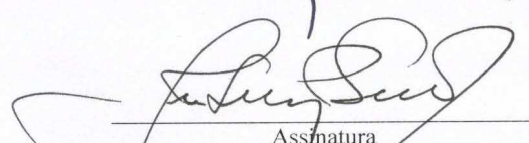
Campinas, 15 de dezembro de 2010

BANCA EXAMINADORA

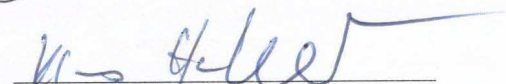
Prof. Dr. Gonçalo Amarante Guimarães Pereira (Orientador)


Assinatura

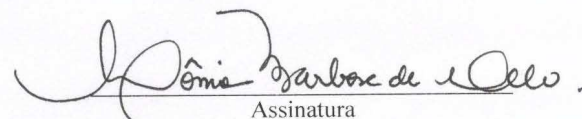
Profa. Dra. Zilá Luz Paulino Simões


Assinatura

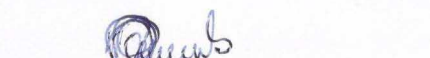
Prof. Dr. Klaus Hartmann Hartfelder


Assinatura

Profa. Dra. Mônica Barbosa de Melo


Assinatura

Dr. Alberto Arab


Assinatura

Profa. Dra. Maria Cristina Arias


Assinatura

Profa. Dra. Odalys García Cabrera


Assinatura

Profa. Dra. Carla Fernanda Franco Penteado


Assinatura

AGRADECIMENTOS

Ao meu orientador Gonçalo pela oportunidade e apoio durante o doutorado;

Ao meu co-orientador tio Fernando, pelo apoio financeiro (Auxílio FAPESP) e oportunidade de aprender em seu laboratório;

Ao Anderson, Dul, Ângela e Ucha por me ensinarem as técnicas básicas de biologia molecular, e por toda a ajuda em experimentos, discussão de resultados, paciência e amizade;

Ao Marcelo pela ajuda com a bionformática e ao João Paulo pelo companheirismo nos experimentos intermináveis de RACE;

À minha mãe Ana Maria por me incentivar a estudar os cupins e também por me fornecer material para a realização da pesquisa;

A todos os amigos que fiz durante todo o período do doutorado tanto no Laboratório de Genômica e Expressão quanto no laboratório de Hemoglobina e Genoma do Hemocentro, obrigado por tudo e pela amizade construída;

À minha família e ao Bernardo por sempre me apoiarem em todos os momentos da minha vida;

À pós-graduação do Instituto de Biologia da Unicamp;

À Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro

RESUMO

Os cupins são insetos pertencentes à Ordem Isoptera, com mais de 3.000 espécies descritas. São importantes agentes de degradação de madeira e compostos celulósicos em geral. Entretanto, essas mesmas características tornam algumas das espécies danosas por destruírem estruturas de edificações, sendo responsáveis por prejuízos da ordem de milhões de dólares. Uma destas espécies-praga é o *Coptotermes gestroi*. Originária do sudeste asiático foi introduzida no Brasil no início do século XX, estabelecendo-se ao longo da região costeira, mas com uma clara expansão para o interior do país. O ciclo de vida de *C. gestroi* inicia com a eclosão do ovo, a larva originada pode se transformar tanto em um operário (linhagem áptera) como em uma ninfa (linhagem ninfal), a qual pode originar um rei ou uma rainha. Além disso, o operário pode se transformar em soldado, sob indução do hormônio juvenil. Assim sendo, o objetivo do presente trabalho foi a avaliação da expressão gênica nas diversas castas e estágios de vida de *C. gestroi* visando a identificação de genes relacionados ao ciclo de vida e com possível aplicação biotecnológica. Para isso, foi feita a análise de uma biblioteca de cDNA de cabeça de operários previamente construída. Para melhor compreender genes diferencialmente expressos entre as castas, outras quatro bibliotecas subtrativas foram confeccionadas. Alguns dos genes identificados foram genes selecionados para a validação pelo método de PCR em tempo Real. O sequenciamento de segunda geração (454) foi realizado com amostras de soldados e operários e mais de 110.000 transcritos foram identificados. Dentre eles destacam-se as Hexamerinas I e II, a epóxido hidrolase do hormônio juvenil e genes relacionados à musculatura. Quanto aos genes com potencial biotecnológico dezenove enzimas envolvidas na quebra da lignocelulose foram identificadas e serão investigadas para confirmar o papel destas neste processo.

ABSTRACT

Termites are insects belonging to the Order Isoptera, with more than 3,000 described species. These insects are important wood and cellulosic compounds degraders. However, these same characteristics make some of them harmful species to mankind, by destroying buildings, accounting for million dollars losses. One of these pest species is *Coptotermes gestroi*. Originated from Southeast Asia, it was introduced in Brazil in the early twentieth century, establishing colonies along the coastal region, but with a clear expansion into the country. The life cycle of *C. gestroi* works as follows: after hatching from the egg, the larvae can become either a worker (apterous lineage) or a nymph (nymphal lineage), which can lead to a king or a queen. In addition, the worker can become a soldier, under induction of juvenile hormone. Therefore, the purpose of this study was the evaluation of gene expression in different castes and developmental stages of *C.gestroi* in order to identify genes related to different life cycles and biotechnological potential. To accomplish this, we analyzed a workers head cDNA library previously prepared. To better understand genes differentially expressed between castes, other four subtractive libraries were constructed. Some of the genes identified were selected for validation by Real Time PCR. Second generation sequencing (454) was performed with samples of workers and soldiers and more than 110,000 transcripts were identified. Among them stand out Hexamerins I and II, the juvenile hormone epoxide hydrolase and genes related to muscle formation. As for genes with biotechnological potential, nineteen enzymes involved in lignocellulose's break down were identified and will be investigated to confirm their role in this process.

LISTA DE ABREVIATURAS

HJ – Hormônio juvenil

AACT – Acetil-CoA tiolase

HMG-S – 3-hidroxi-3-metilglutaril-CoA sintase

HMG-R – 3-hidroxi-3-metilglutaril-CoA redutase

JHE – Esterase do hormônio juvenil

JHEH – Epóxido hidrolase do hormônio juvenil

HEX – Hexamerina

ECS – Extrato de cabeça de soldado

DNA – Ácido desoxirribonucleico

cDNA – DNA complementar

RNA – Ácido ribonucleico

RNA_r – RNA ribossômico

RNA_m – RNA mensageiro

PCR – Reação de polimerase em cadeia

X-GAL – 5-bromo-4-cloro-3-indolil β-D-galactopiranosídeo

DTT – Ditionitreitol

EDTA – Ácido etilenodiamino tetra-acético

dNTPs – Desoxirribonucleotídeos fosfatados

NTC – controle negativo

Ct – cycle threshold

ATP – adenosine- trifosfato

GH – glicosil hidrolase

LISTA DE TABELAS

Tabela 1 – Sequência e tamanho dos fragmentos amplificados de cada par de primer utilizado na amplificação dos genes estudados pela técnica de Real Time PCR.	32
Tabela 2 – Concentrações de primers utilizadas na amplificação dos genes de estudo e eficiência da amplificação obtida. As concentrações foram definidas pela eficiência de amplificação gerada sob as condições testadas.	33
Tabela 3 – Genes escolhidos para a validação das bibliotecas subtrativas 1 e 2.	48
Tabela 4 – Genes escolhidos para a validação das bibliotecas subtrativas 3 e 4.	48
Tabela 5 – Genes diferencialmente expressos entre operários e soldados de <i>C. gestroi</i> através do pirosequenciamento.	58
Tabela 6 – Enzimas identificadas no pirosequenciamento.	61
Tabela 7 – Sequências não completas de enzimas envolvidas na quebra da lignocelulose.	61

LISTA DE FIGURAS

Figura 1 – Esquema geral do ciclo de vida de uma colônia de Isoptera (modificado de Costa-Leonardo, 2002).	4
Figura 2 – Esquema da via do mevalonato. Adaptado de Keeling et al., 2004.....	8
Figura 3 – Diferenciação de castas de <i>Coptotermes gestroi</i>	11
Figura 4 – Morfologia e recursos do trato digestivo de um cupim inferior..	15
Figura 5 – Esquema com os passos envolvidos na construção da biblioteca subtrativa (adaptado do manual do Kit SSH da Clontech).	29
Figura 6 – Resumo ilustrativo do sequenciamento (adaptado do site da ROCHE).....	37
Figura 7 – RNA de <i>C. gestroi</i> i..	43
Figura 8 – Página do projeto cupim do laboratório de genômica e expressão	44
Figura 9 – Gráfico da anotação das seqüências que obtiveram similaridade com o BLAST/ GO das bibliotecas 1 e 2 (larva X linhagem áptera).....	45
Figura 10 – Gráfico da anotação das seqüências que obtiveram similaridade com o BLAST/ GO das bibliotecas 3 e 4 (larva X linhagem ninfal).....	47
Figura 11 – Gráfico do número de vezes que cada gene foi encontrado nas bibliotecas 1 e 2.	49
Figura 12 – Gráfico do número de vezes que cada gene foi encontrado nas bibliotecas 3 e 4.	50
Figura 13 – Gráfico da expressão diferencial dos genes selecionados nas bibliotecas 1 e 2.	51
Figura 14 – Gráfico da expressão diferencial dos genes selecionados nas bibliotecas 3 e 4.	51
Figura 15 – Níveis de expressão relativa do gene que codifica endo- β -1,4 glicosidase nas diferentes castas e estágio de desenvolvimento de <i>C. gestroi</i>	52
Figura 16 – Níveis de expressão relativa do gene que codifica β -glicosidase nas diferentes castas e estágio de desenvolvimento de <i>C. gestroi</i>	53

Figura 17 – Níveis de expressão relativa do gene que codifica Vitelogenina nas diferentes castas e estágio de desenvolvimento de <i>C. gestroi</i>	54
Figura 18 – Níveis de expressão relativa do gene que codifica JHEH nas diferentes castas e estágio de desenvolvimento de <i>C. gestroi</i>	55
Figura 19 – Níveis de expressão relativa do gene que codifica a Hexamerina I nas diferentes castas e estágio de desenvolvimento de <i>C. gestroi</i>	56
Figura 20 – Níveis de expressão relativa do gene que codifica Hexamerina II nas diferentes castas e estágio de desenvolvimento de <i>C. gestroi</i>	57
Figura 21 – Gráfico que mostra a classificação das sequências por resultado do First hit do BlastX .	59
Figura 22 – Gráficos mostrando o resultado da expressão diferencial do pirosequenciamento em comparação aos resultados do estudo de expressão de Real Time PCR de seis genes.	59
Figura 23 – Esquema com todas as enzimas relacionadas com a degradação da madeira encontradas em <i>C. gestroi</i>	60

SUMÁRIO

RESUMO	vii
ABSTRACT	ix
LISTA DE ABREVIATURAS	xi
LISTA DE TABELAS	xiii
LISTA DE FIGURAS	xv
SUMÁRIO	xix
INTRODUÇÃO	1
1 – Diversidade e biologia dos cupins	3
2 –Diferenciação de castas	6
3 – Cupins como insetos praga.....	10
4 – <i>Coptotermes gestroi</i>	10
5 – Alimentação lignocelulósica e microbiota intestinal.....	12
6 – Estudos moleculares com a espécie <i>C. gestroi</i>	15
7 – Potencial biotecnológico do arsenal enzimático dos cupins.....	16
OBJETIVOS.....	19
MATERIAL E MÉTODOS.....	23
1 – Obtenção e identificação dos <i>C. gestroi</i>	25
2 – Extração do RNA	25
3 – Construção das bibliotecas subtrativas.....	26
3.1 – Clonagem, amplificação e sequenciamento dos clones	30
3.2 – Bioinformática e análise dos dados.....	30
4 – Síntese de DNA Complementar (cDNA) para PCR em tempo Real	30
4.1 – Verificação da Síntese de DNA Complementar.....	31
4.2 – Desenho dos Primers para a reação do Real Time PCR	31
4.3 – Padronizações necessárias para Real Time PCR.....	33

4.3.1 – Concentração de Primer	33
4.3.2 – Eficiência de Reação.....	34
4.4 – Real Time PCR	34
4.5 – Análise dos Dados da Real Time PCR.....	35
4.6 – Análise Estatística.....	36
5 – Sequenciamento de segunda geração	36
5.1 – Bioinformática e análise dos dados	38
RESULTADOS.....	41
1 - Extração de RNA	43
2 – Construção das bibliotecas subtrativas	43
3 - PCR em Tempo Real das bibliotecas subtrativas.....	50
4 – Sequenciamento de segunda geração	57
4.1 - Genes com aplicação biotecnológica.....	60
DISCUSSÃO	63
1 - Bibliotecas subtrativas	65
2- Sequenciamento de segunda geração	69
2.1 – Genes relacionados à regulação de castas.....	69
2.2 - Genes com aplicação biotecnológica.....	71
CONCLUSÕES	73
REFERÊNCIAS BIBLIOGRÁFICAS	77
ANEXOS	87
ANEXO I – Artigo aceito para publicação na revista Bulletin of Entomological Research.....	89
ANEXO II – Sequências de todas as enzimas envolvidas na degradação da lignocelulose encontradas no pirosequenciamento.	99

INTRODUÇÃO

1 – Diversidade e biologia dos cupins

Os cupins ou térmitas são insetos hemimetábolos que constituem a Ordem Isoptera, a qual apresenta cerca de 3.000 espécies descritas (Inward *et al.*, 2007). Eles são insetos sociais, sendo que existe intensa interdependência entre os indivíduos que compõem a colônia e ocorre sobreposição de gerações. Uma característica particular desse grupo é que seus reprodutores alados possuem dois pares de asas membranosas iguais. A sociedade dos cupins possui indivíduos morfologicamente diferentes, os quais são adaptados ao trabalho que exercem e baseado neste fator, constituem as diferentes castas presentes nos ninhos, conhecidos como cupinzeiros. Esse é, aliás, um aspecto fundamental para a sobrevivência desses insetos. Os cupins são caracterizados pela fragilidade do seu abdômen, o que torna obrigatória a vida em um ambiente “climatizado”, que os protege da dissecação ou mesmo do contato com predadores. Assim, o microclima necessário para o funcionamento estável do ninho é garantido pela construção das galerias do ninho ou dos túneis utilizados para a locomoção, os quais podem ser subterrâneos, formados pela escavação do solo, ou mesmo superficiais conectando o ninho à fonte alimentar. Seus ninhos constituem obras notáveis de engenharia (Noirot e Darlington, 2000). Essa necessidade de abrigo sugere que a evolução desses organismos se deu a partir de populações que originalmente viviam em madeira apodrecida e só posteriormente adquiriram a capacidade de mover o seu habitat (Abe, 1987).

Adicionalmente, os térmitas são classificados como cupins inferiores ou superiores, baseado na flora intestinal que possuem. Os cupins superiores compreendem as espécies da Família Termitidae que não possuem protozoários em seu trato digestivo. Já os cupins inferiores são os que apresentam protistas simbiotes no trato digestivo e estes compreendem as demais 6 famílias que compõem a ordem Isoptera: Mastotermitidae, Kalotermitidae, Rhinotermitidae, Hodotermitidae, Serritermitidae e Termopsidae (Inward, *et al.*, 2007).

De forma geral, nos ninhos dos cupins são encontrados normalmente um rei e uma rainha, que constituem a casta reprodutiva. Diferentemente dos Hymenoptera (abelhas, formigas e vespas), não ocorre a haplodiploidia, com a determinação do sexo sendo definida pela presença de cromossomos sexuais, com ambos os sexos sendo diplóides (Noirot, 1989). Além da casta reprodutiva, encontram-se também nas colônias operários e

soldados, que constituem a casta neutra (indivíduos não reprodutivos), e ovos, “larvas” e ninfas (Krishna e Weesner, 1969).

O ciclo de vida de uma colônia de cupim pode ser observado Figura 1.

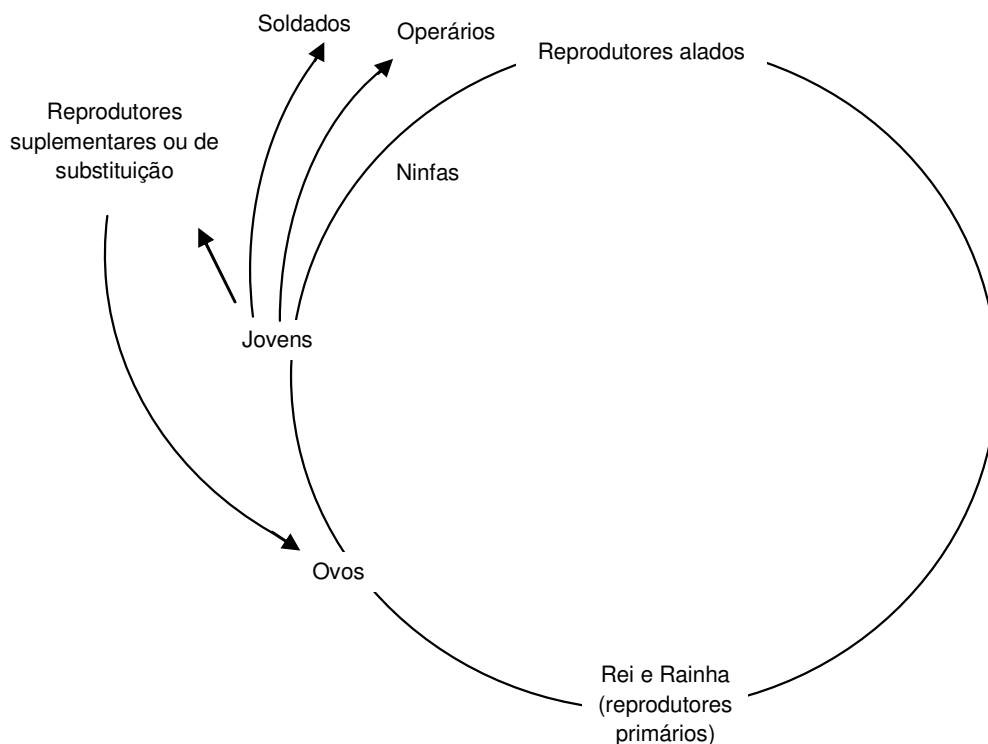


Figura 1 - Esquema geral do ciclo de vida de uma colônia de Isoptera (modificado de Costa-Leonardo, 2002).

Na maioria das espécies de cupins, os operários representam a casta mais numerosa da colônia, são imaturos e não têm asas. Estes indivíduos realizam todas as atividades de forrageamento (busca de alimento), construção e reparo de ninho, além da alimentação das demais castas e cuidados com a cria. Nas espécies desprovidas de soldados, os operários também são ativos na defesa da colônia (Costa-Leonardo, 2002).

Os soldados são indivíduos quase sempre derivados dos operários, que geralmente possuem cabeças maiores e mandíbulas musculosas poderosas para atuarem na defesa da colônia. Os soldados também participam das atividades de forrageamento, em conjunto com os operários. Alguns soldados ejetam da fontanela (orifício da cabeça) um líquido, produzido pela glândula frontal, que é repelente e viscoso, e pode ser tóxico para seus

inimigos (Costa-Leonardo, 2002). Alternativamente, o líquido ejetado pode ter papel físico. Por exemplo, *Coptotermes lacteus* produz um fluido leitoso, que consiste de uma suspensão atóxica de lipídeos e mucopolissacarídeo aquoso, que ao secar se torna viscoso e pode imobilizar pequenos insetos (Moore, 1968).

Os cupins são diplóides e apesar dos operários e soldados constituírem castas imaturas, eles podem ser tanto do sexo masculino quanto do feminino. Entretanto, observa-se a existência de um viés sexual nas castas de algumas espécies. Por exemplo, em Hodotermitidae e na maior parte dos Termitidae todos os soldados são do mesmo sexo, macho ou fêmea, mas os operários são dos dois sexos (Noirot, 1985). Na espécie *Trinevitermes* sp os operários são de um sexo e os soldados são do sexo oposto (Noirot, 1955). Em duas espécies de Rhinotermitidae todos os indivíduos da casta neutra, operários e soldados, são do mesmo sexo (Roisin, 1988). Para a espécie em estudo nada foi descrito em relação à proporção de machos e fêmeas.

Os reprodutores primários ou funcionais consistem de um rei e uma rainha, os quais são assim denominados após a fundação da colônia e ocorrência da cópula. Estes reprodutores se originam de ninfas que se tornam alados na época das revoadas, definidas por condições ambientais e que variam entre as espécies. No estado de São Paulo, a revoada maciça de *C. gestroi* ocorre geralmente nos meses de agosto e setembro. Os alados são atraídos pelas luzes urbanas onde encontram seus pares realizando o vôo nupcial. Logo após o vôo perdem suas asas e fundam novos ninhos. A cópula só ocorre depois do estabelecimento do casal numa “câmara real” determinando então a ocorrência de monogamia absoluta, o que resulta numa colônia formada por irmãos e irmãs. A presença do rei parece ser indispensável durante a vida da colônia porque a rainha precisa ser inseminada regularmente. Em *Macrotermes michaelsoni*, por exemplo, a rainha fisiogástrica só consegue botar ovos fertilizados até 13 dias após a retirada do rei do ninho (Sieber e Leuthold, 1982).

Existe um constante estoque de ninfas nos ninhos, que além de originarem os adultos alados, também podem gerar outros tipos de reprodutores, como os de substituição ou neotênicos (Noirot, 1990). Estes reprodutores são indivíduos que atingem a maturidade sexual antes de atingirem a idade adulta, desenvolvendo órgãos reprodutores funcionais. Entretanto, eles não deixam a colônia parental. Os neotênicos assumem o papel de

reprodutor efetivo apenas nos casos de descontinuidade da colônia; ou seja, quando a colônia sofre uma separação física entre suas partes. Na região desprovida dos reprodutores principais, essas ninfas podem se tornar reprodutores efetivos, o que, em algumas espécies, ocorre após uma forte competição entre esses indivíduos. Ao final, apenas um casal assumirá o papel reprodutivo (Thorne, 1982). Esse processo também ocorre em caso de morte de um dos reprodutores funcionais, sendo uma diferenciação possível em quase todas as espécies de cupim (Myles e Nutting, 1988).

Nos Hymenoptera sociais, apenas os adultos realizam tarefas comunitárias. Nos cupins, os únicos indivíduos considerados adultos dentro de uma colônia são os reprodutores, por terem perdido a capacidade de realizarem mudas (ecdise). A rigor, as demais castas (operários e soldados) não chegam à maturidade, retendo as glândulas protorácicas, que são as glândulas de muda (Noirot, 1969). Isso faz com que os organismos mantenham-se em constante renovação ao longo da vida (desenvolvimento hemimetábolo), sendo talvez essa a característica responsável pela longevidade desses insetos (Wilson, 1971).

2 –Diferenciação de castas

Um importante fator presente nos insetos sociais é a constituição de castas, as quais são formadas por indivíduos com fenótipos diferentes e características específicas para a realização de determinadas tarefas dentro das colônias (Miura, 2004). A determinação de castas nos cupins é fortemente influenciada pelo sistema endócrino (Lüscher, 1958), o qual é constituído pelos “corpora cardíaca”, “corpora allata” e glândulas de muda (Noirot, 1969). Os “corpora allata” secretam hormônio juvenil (HJ), que é responsável pela manutenção das formas imaturas e que também regula a formação de soldados a partir de operários (Lebrun, 1967). Essa função do HJ é muito importante, pois na colônia deve existir uma proporção ideal da casta de soldados, já que um número excessivo pode prejudicar o vigor da colônia, uma vez que os soldados precisam ser alimentados pelos operários e não realizam todas as tarefas essenciais para a sobrevivência da colônia.

O hormônio juvenil presente nos Isoptera é o Hormônio Juvenil III. A via responsável por sua síntese é bastante complexa e envolve produtos intermediários: o mevalonato e o farnesil pirofosfato. A rota de formação do mevalonato envolve a ação das

enzimas acetoacetyl-CoA tiolase (AACT), 3-hidroxi-3-metilglutaril-CoA sintase (HMG-S) e a 3-hidroxi-3-metilglutaril-CoA redutase (HMG-R). O segundo precursor, o farnesil difosfato, envolve a ação das enzimas mevalonato quinase, fosfomevalonato quinase, difosfomevalonato descarboxilase e a isopentenil difosfato isomerase (Figura 2). Estudos de expressão gênica realizados no coleóptero *Ips pini* constataram que a alimentação dos machos aumenta a expressão de sete genes envolvidos nessa via, representando um papel importante na regulação da produção do HJ III (Keeling *et al.*, 2004).

O HJ é degradado por duas enzimas, a esterase do HJ (JHE) e a epóxido hidrolase do HJ (JHEH). A JHE está presente na hemolinfa e nos tecidos, enquanto a JHEH está ligada a tecidos específicos (de Kort e Granger, 1996).

A ação das enzimas JHE e JHEH é influenciada por proteínas que se ligam ao hormônio juvenil e que estariam associadas, teoricamente, a todo o HJ disponível (Touhara e Prestwich, 1994). Em insetos, foram descritos três tipos de proteínas de ligação ao hormônio juvenil: a JHBP (*Juvenile hormone binding protein*) de *M. sexta* (Kramer *et al.* 1976); uma lipoforina, encontrada na hemolinfa de *Periplaneta americana* e *Locusta migratoria* (de Kort e Koopmanschap, 1986; 1987; 1989) e as hexamerinas identificadas em *L. migratoria* (Koopmanschap e de Kort 1988) e posteriormente em *R. flavipes* (Scharf *et al.*, 2005).

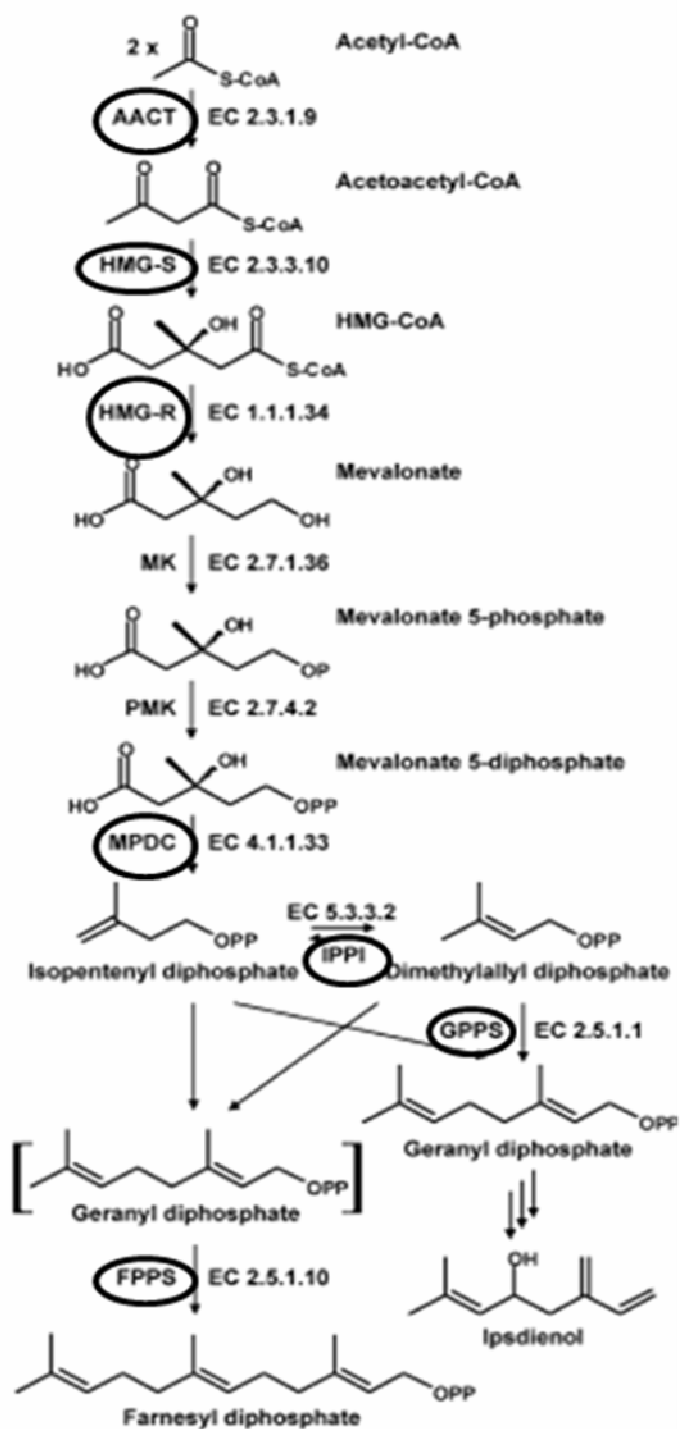


Figura 2 - Esquema da via do mevalonato. Adaptado de Keeling et al., 2004.

Para identificar os genes associados à mudança de castas - de operário para soldado - em *R. flavipes*, foram realizados bioensaios com um juvenóide de HJ III, que é uma substância sintética análoga ao HJ (Scharf, *et al.*, 2005). No referido estudo, foram identificadas quatro proteínas da hemolinfa; vitelogeninas e hexamerinas, cujos níveis de expressão aumentaram após o tratamento com HJ III, indicando um papel importante dessas proteínas na diferenciação de castas em *R. flavipes*. Adicionalmente, utilizou-se a técnica de RNA de interferência para investigar o papel das hexamerinas (HEX-1 e HEX-2) na diferenciação de castas (Zhou, *et al.*, 2006a). Com o silenciamento do gene que codifica HEX-2, o número de pré-soldados aumentou, sugerindo que essa hexamerina tem um papel importante na supressão da diferenciação da casta de soldado.

Além disso, está comprovado que existem outros fatores, diferentes do HJ, que também interferem na diferenciação de castas dos cupins. Todos os ovos de cupins são totipotentes e suportam evidências de que a diferenciação de castas é baseada na expressão de genes que respondem a fatores extrínsecos e intrínsecos. Portanto, o fenótipo dos indivíduos pode mudar. Os fatores que ditam cada uma das mudanças no desenvolvimento dos cupins e como eles interagem ainda não são completamente compreendidos (Tarver *et al.*, 2010). Os fatores intrínsecos incluem níveis de HJ, proteínas de reserva e nutrição. Os fatores extrínsecos, por sua vez, compreendem feromônios modificadores, temperatura, disponibilidade de alimento, presença de companheiros do ninho (soldados ou reprodutores) e estações do ano (Howard e Haverty, 1981; Scharf *et al.*, 2007). Divergências fenotípicas entre operário e soldado podem estar moduladas por múltiplos fatores relacionados ao HJ. Por exemplo, níveis aumentados de HJ em operários estão correlacionados com a diferenciação de pré-soldados (Korb *et al.*, 2009). Além dos níveis de HJ aumentados, a presença de soldados também foi um fator que ajudou a inibir a formação de novos indivíduos dessa casta. Isto sugere que os soldados produzem inibidores que causam a redução de resposta ao HJ ou redução de sua biossíntese nos companheiros do ninho (Park e Raina, 2003). Presume-se que essa inibição seja causada por feromônios modificadores de soldados (Korb *et al.*, 2003). Esses feromônios são definidos como compostos químicos secretados por glândulas exócrinas que são passados para os demais companheiros da colônia e ativam respostas fisiológicas nos indivíduos receptores (Wilson, 1963). No cupim *R. flavipes* foi demonstrado que, quando se aplica HJIII em operários,

ocorre a formação de pré-soldados. No entanto, quando esta aplicação de HJ III é realizada em conjunto com a aplicação do extrato da cabeça de soldados (ECS) a produção de pré-soldados é ainda maior (Scharf *et al.*, 2007). É interessante notar que ECS sozinho não tem impacto algum na formação de pré-soldados (Scharf *et al.*, 2007). Este trabalho mostrou que os soldados vivos podem estar atuando em *feed back* negativo na produção de soldados, ou seja, inibindo a formação de soldados regulando assim genes importantes para a diferenciação de castas.

3 – Cupins como insetos praga

Algumas espécies efetivamente são responsáveis por enormes prejuízos tanto no meio urbano quanto no meio rural. Os cupins subterrâneos são os mais amplamente distribuídos pelo mundo, ocorrendo nas regiões tropicais, subtropicais e temperadas (Eggleton, 2000). De acordo com Edwards e Mill, das 3.000 espécies de cupins existentes, apenas 113 são pragas, sendo que 80% delas são espécies de cupins subterrâneos. O gênero *Coptotermes* (Família Rhinotermitidae) é o de maior importância econômica (Su e Scheffrahn, 1998).

Eles frequentemente atacam construções urbanas ocasionando um impacto econômico considerável. As estimativas mais recentes reportam que foram gastos mais de US\$ 11 bilhões por ano nos EUA para o tratamento de danos provocados pelas espécies de cupins subterrâneos dos gêneros *Reticulitermes* e *Coptotermes*. (Su, 2002). No Brasil, os cupins também são responsáveis por grandes prejuízos nos centros urbanos, principalmente na região sudeste (Fontes e Milano, 2002). Este prejuízo foi amostrado em um levantamento realizado pelo Instituto de Pesquisa e Tecnologia (IPT) no qual foi constatado gasto de US\$ 3,14 milhões em tratamentos curativos só na cidade de São Paulo no ano de 2005 (Romagnano e Nahuz, 2006).

4 – *Coptotermes gestroi*

A espécie *C. gestroi* é proveniente do sudeste asiático, sendo encontrada na Índia, Birmânia, Malásia e Indonésia. Seus primeiros registros no Brasil datam de 1923 no Rio de Janeiro e de 1934 em Santos (Araújo, 1958) e, desde então, está se expandindo para o interior (Costa-Leonardo, 2002). Durante muito tempo esta espécie foi identificada no

Brasil como *Coptotermes havilandi* e só depois, de realizada uma extensa revisão do gênero *Coptotermes*, descobriu-se sua sinonímia e passou a ser denominada *Coptotermes gestroi* (Kirton e Brown, 2003). Aspectos da sua biologia, tais como reprodução, forrageamento, revoadas, ataques a edificações e danos em espécies vegetais, foram registrados por vários pesquisadores (Fontes e Berti Filho, 1998; Zorzenon e Potenza, 1998; Costa-Leonardo *et al.*, 1999; Costa-Leonardo, 2004; Barsotti e Costa-Leonardo, 2005; Costa-Leonardo *et al.*, 2005; Neoh & Lee, 2009).

Em relação ao desenvolvimento, a diferenciação de castas (Figura 3) ocorre da seguinte forma; a larva eclode do ovo, com dois estágios discrimináveis e pode se desenvolver em operário (via áptera) ou ninfa (via ninfal) após vários instares de desenvolvimento. Os operários podem permanecer nessa condição, ou, se tornarem soldados após duas mudas (passa pelo estágio de pré-soldado). As ninfas seguem a via ninfal (real), desenvolvendo-se em alados (rei e rainha: forma reprodutiva primária) ou em reprodutores secundários (reprodutores neotênicos que assumem o papel dos reprodutores primários) (Barsotti e Costa-Leonardo, 2005).

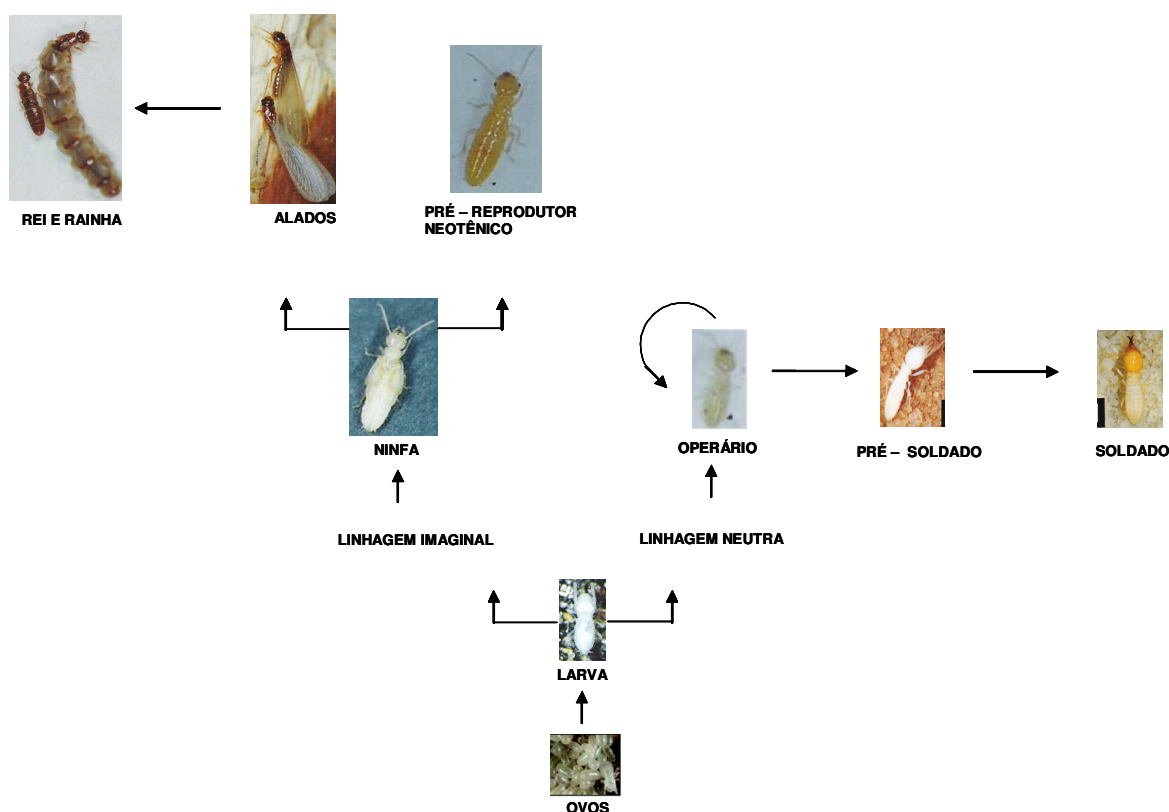


Figura 3– Diferenciação de castas de *Coptotermes gestroi*.

5 – Alimentação lignocelulósica e microbiota intestinal

A base da alimentação dos cupins é lignocelulósica, sendo a madeira o alimento escolhido por muitas espécies. Eles também podem se alimentar de folhas, raízes, sementes, fezes de herbívoros, gramíneas e até restos de animais em decomposição. Os cupins da subfamília Macrotermitinae se alimentam de fungos cultivados em forma de favos (Rouland, 2000).

A lignocelulose é constituída por celulose, hemicelulose e lignina. A celulose é composta por polímeros rígidos de alto peso molecular de glicose com ligações β -1,4 que são mantidos em feixes pela hemicelulose. (Lange, 2007). Esta é composta por pequenos polímeros de açúcares com ligações β -1,4, sendo que a manose é geralmente o açúcar dominante presente nas madeiras ingeridas pelos cupins, com menores quantidades de xilose, galactose, raminose, arabinose, ácido glucurônico, manurônico e galacturônico (Saha, 2003). A lignina não é um carboidrato, e sim um polímero tri-dimensional de compostos fenólicos ligados entre si e à hemicelulose por ésteres (Anderson e Akin, 2008). A lignina é composta por três monômeros; álcool p-coumaril, álcool coniferil e sinapil combinados em diferentes proporções dependendo da espécie da planta (Scharf e Tartar, 2008).

A degradação da lignocelulose em açúcares é um processo complexo que requer uma diversidade de enzimas. Sua degradação começa a partir da mastigação, já que os cupins apresentam uma poderosa armadura bucal com 6 peças móveis (labro ou lábio superior, mandíbulas, maxilas, lábio inferior, epifaringe e hipofaringe) (Krishna e Weesner, 1969). Após a quebra da madeira em partículas menores, a despolimerização da lignina hidrofóbica é extremamente importante para permitir a degradação da hemicelulose. Como a lignina não é um carboidrato, carbohidrolases tais como as celulasas e as hemicelulasas, teoricamente não desempenham papel algum na degradação da lignina. É importante notar que a quebra da lignina necessita de oxigênio, o que suporta a evidência de que o trato digestivo dos cupins não é um ambiente completamente anaeróbico (Brune, 1995). Enzimas envolvidas na oxidação da lignina incluem as lacases e peroxidases. O próximo passo para a degradação da lignocelulose seria a degradação da hemicelulose, estágio crítico para tornar a despolimerização da celulose acessível (Saha, 2003). A biodegradação completa da hemicelulose requer atividade combinada entre endo-exo- β -1,4-xilanases, β -xilosidases,

α -arabinofuranosidases, α -uronidases, e esterases, tais como esterases acetilxilana, ácido ferúlico esterases e ácido p-coumárico esterases. Conversões específicas de cada uma dessas enzimas são as seguintes: (i) endo-xilanases hidrolisam as ligações β -1,4-xilose no suporte principal de xilano; (ii) exo-xilanases hidrolisam as ligações β -1,4-xilose reduzidas liberando xilobiose; (iii) β -xilosidases atuam na xilobiose para liberar xilose e outros oligossacarídeos de cadeias menores; (iv) α -arabinofuranosidases hidrolisam α -arabinofuranose terminais não reduzidas em arabinoxilanos; (v) α -uronidases liberam ácidos α -glucurônicos, manurônicos e galacturônicos e (vi) esterases hidrolisam ligações de ésteres fenólicos, tais como os associados com acetil xilanos, ácidos xilano-felúricos e p-coumárico xilanos (Scharf e Tartar, 2008).

O último passo, a despolimerização da celulose requer a ação de três enzimas primárias que incluem endo- β -1,4-glicanases, exo- β -1,4-glicanases e β -glicosidases (Breznac e Brune, 1995). Conversões específicas realizadas por essas enzimas: (i) endoglicanases hidrolisam as ligações glicosil- β -1,4 no suporte principal da celulose primária liberando glicose, celobiose, celotriose ou outros oligômeros longos; (ii) exoglicanases ou celobiohidrolases agem na região terminal das cadeias poliméricas para liberar glicose ou celobiose, e finalmente (iii) β -glicosidases hidrolisam celobiose e celotriose para liberar monômeros de glicose. Breznak e Brune (1995) observaram que a hidrólise completa da celulose requer que haja uma sinergia entre os três tipos de celulases.

Para a degradação enzimática, um ponto importante é que os cupins precisam do auxílio de simbioses intestinais (Robinson, 1996). Acreditava-se que a habilidade de degradar celulose fosse restrita a bactérias e plantas e que animais que se alimentavam de celulose só conseguiam digeri-la devido à flora intestinal. As evidências que começaram a mudar esse paradigma vieram de estudos bioquímicos que identificaram atividade celulolítica em um ambiente livre de simbioses como as glândulas salivares e extratos digestivos de cupins (Slaytor, 1993). Posteriormente, genes codificadores de endoglicanases foram descobertos no genoma dos cupins, comprovando que eles produzem suas próprias celulases endógenas (Watanabe, 1997). Recentemente, mostrou-se que transcritos de celulases de origem dos simbioses, assim como dos próprios cupins, estão presentes nos tratos digestivos dos cupins e atuam conjuntamente para a degradação da celulose (Nakashima et al., 2002).

A flora digestiva dos cupins é bastante complexa, e de acordo com Bignell (Bignell, 2000), os mutualistas podem ser de três domínios: Archea, Eubactéria e Eucarya (protozoários e fungos). Nos cupins inferiores, os protozoários contribuem com a transformação da celulose e da hemicelulose em acetato, que é aproveitado pelo inseto. Eles também fornecem CO₂ e acetato para as bactérias presentes no trato digestivo. Acredita-se que essas bactérias sejam as responsáveis pela fixação de nitrogênio em amônia, constituindo uma fonte de nitrogênio para os cupins (Slaytor e Chappell, 1994).

Os operários compõem a maior parte dos indivíduos das colônias de cupins e realizam toda a alimentação e digestão da lignocelulose. O trato digestivo de um operário (Figura 4) de cupim inferior é composto por três regiões principais: intestino anterior, intestino médio e intestino posterior (Breznak e Brune, 1995). O intestino anterior é ladeado pelas glândulas salivares, que desembocam na cavidade bucal. Essas glândulas secretam endoglicanases endógenas e outras enzimas relevantes no trato digestivo. O intestino médio é uma região tubular delgada que secreta uma membrana peritrófica em volta do material alimentar e é presumidamente o local em que alguma degradação da lignocelulose ocorre. Análogos aos rins dos vertebrados, os túbulos de Malpighi estão conectados na junção do intestino médio com o posterior e tem função com os processos de excreção. O intestino posterior é constituído por uma câmara que é geralmente anaeróbica, mas que possui uma zona micro-óxida em sua periferia. (Brune et *al.*, 1995). É no intestino posterior que se encontram a maior parte dos simbioses, e é também onde a maior parte da degradação da celulose ocorre, bem como a sua fermentação.

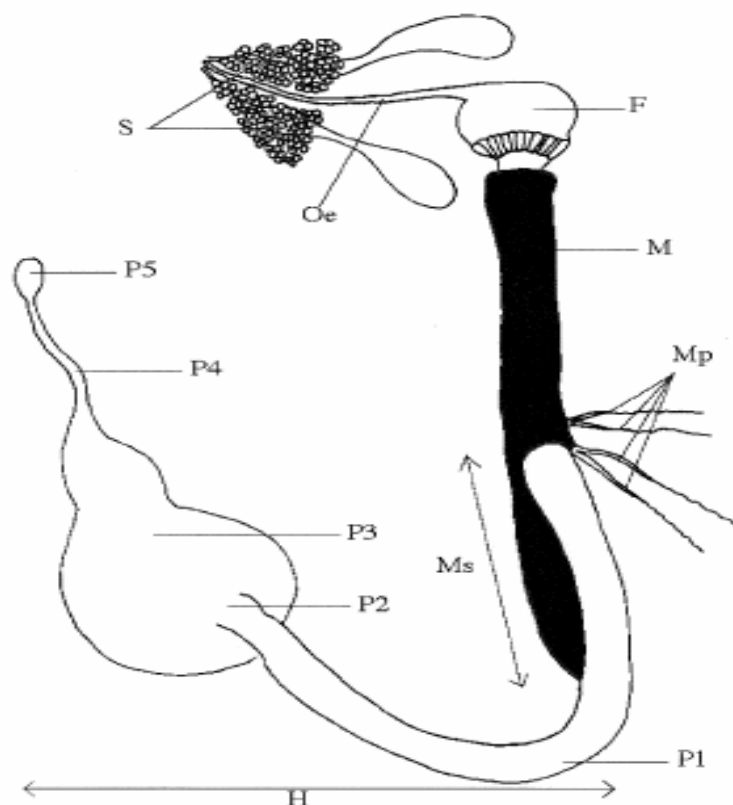


Figura 4 - Morfologia e recursos do trato digestivo de um cupim inferior. O esquema mostra as regiões presentes: **S** – glândulas salivares; **Oe** – esôfago; **F** – intestino anterior; **M** – intestino médio; **Ms** – segmento misto; **Mp** – túbulos de Malpighi; **H** – intestino posterior; **P1** – primeiro segmento proctodeal; **P2** – válvula entérica; **P3** – pança; **P4** – cólon e **P5** – reto.

6 – Estudos moleculares com a espécie *C. gestroi*

Na literatura são encontrados poucos relatos sobre o genoma de *C. gestroi*, sendo, portanto, desconhecida a maioria dos genes expressos para a espécie. Apenas algumas sequências de *C. gestroi* foram depositadas no banco de dados do NCBI (www.ncbi.nlm.nih.gov). Elas são sequências da subunidade II da citocromo oxidase, ou seja, são genes do DNA mitocondrial utilizados para a identificação filogenética das espécies de cupim (Husseneder, *et al.*, 2003), e sequências 12 e 16S rRNA, que também são usados em estudo filogenéticos da espécie (Li *et al.*, 2009). Além dessa sequencias depositadas no NCBI, nosso grupo de pesquisa iniciou um estudo de *C. gestroi* com a construção de uma biblioteca de ESTs a partir de cabeça de operários (Leonardo, 2006). A

princípio, apenas a cabeça foi utilizada com a intenção de evitar contaminação por protozoários/bactérias simbiotes, já que a fauna digestiva dos cupins é bastante complexa com vários protozoários flagelados presentes no trato intestinal. O sequenciamento de 50 placas gerou um total de 3.003 sequências de alta qualidade (Phred $\geq$ 20). As sequências foram agrupadas pelo programa CAP3 (Huang e Madan, 1999) em 245 “contigs” e 450 “singlets”, os quais foram anotados e submetidos ao banco de dados do NCBI. O artigo gerado pela análise dessas sequências está em anexo (ANEXO I).

7 – Potencial biotecnológico do arsenal enzimático dos cupins

Como mencionado acima, os cupins utilizam enzimas próprias e de simbiotes para degradação da madeira. Essas enzimas são principalmente celulasas, as quais, em vista da eficiência de degradação apresentada por esses insetos, podem representar um importante fator de estudo, já abre a possibilidade do emprego dessas enzimas em processos biotecnológicos, em particular, na degradação de compostos celulósicos da palha de cana, que hoje, nas usinas mais eficientes, é queimada para a produção de energia em caldeiras. Contudo, seria um avanço excepcional converter a celulose existente nesses compostos em glicose, a qual poderia ser empregada diretamente na fermentação (Kirk e Farrel, 1987). As avaliações do potencial das enzimas de cupins para essa finalidade têm sido também uma estratégia adotada por outros países, como os EUA, a partir do seu Departamento de Energia (DOE) (Kintisch, 2006), e mostra a importância biotecnológica do estudo molecular dos cupins.

Estudos de metagenômica em diferentes organismos e biomas, visando a análise da atividade enzimática dos micro-organismos, vêm sendo realizados com o intuito de descobrir e caracterizar novas enzimas. Bibliotecas de cDNA e DNA do trato digestivo de cupins *Reticulitermes speratus* (Tokuda *et al.*, 2007) e de uma espécie de *Nasutitermes* (Warnecke *et al.*, 2007), foram construídas para a investigação de todas as enzimas relacionadas com a degradação de celulose. Estes estudos revelaram a presença de várias famílias de glicosil hidrolases, celobiohidrolases, betaglucosidases e xilanases, que são enzimas essenciais no processo de conversão da biomassa, além de genes relacionados à vida microbiana nesse microambiente. Um *digestoma*, ou seja, o estudo do trato digestivo do cupim e de seus simbiotes, da espécie *R. flavipes* foi desenvolvido (Tartar *et al.*, 2008).

Indivíduos inteiros de *R. flavipes* foram seqüenciados, gerando mais de 11.000 ESTs contendo sequências do trato digestivo dos cupins e de seus simbiontes. Dentre elas foram encontradas glicosil hidrolases de 28 famílias e vários genes candidatos a lignases endógenas (Tartar *et al.*, 2009). Recentemente duas isoformas de lacases foram descritas em *R. flavipes*. A função dessas enzimas seria a de atuar na quebra da lignina, e sua produção estaria ocorrendo na glândula salivar (Coy *et al.*, 2010).

O cupim *C. gestroi* foi a espécie escolhida para o desenvolvimento desse trabalho, tendo em vista a falta de estudos moleculares da diferenciação de castas nesta espécie e por se tratar de um inseto xilófago com enzimas lignocelulósicas com grande potencial de uso em processos biotecnológicos.

OBJETIVOS

O objetivo geral deste trabalho foi contribuir para uma melhor compreensão do transcriptoma de *Coptotermes gestroi*.

Os objetivos específicos visaram:

- a identificação de genes diferencialmente expressos entre castas através de estudos de expressão gênica nos diferentes estágios de desenvolvimento com o uso da técnica de Biblioteca Subtrativa Suprimida (SSH), Real Time PCR e Sequenciamento de segunda geração
- identificar enzimas envolvidas na degradação da lignocelulose com possível aplicação biotecnológica, visto que *C. gestroi* é uma espécie praga que degrada compostos celulósicos.

MATERIAL E MÉTODOS

1 – Obtenção e identificação dos *C. gestroi*

Operários e soldados forrageiros de *C. gestroi* foram coletados em armadilhas celulósicas, distribuídas em diversas locais na cidade de Rio Claro-SP. A identificação dos cupins foi realizada de acordo com a colônia e o local de coleta. Os reprodutores alados foram coletados durante as revoadas que ocorreram em agosto/2006. Larvas e reprodutores primários (reis e rainhas) foram coletados de colônias incipientes criadas em laboratório e foram cedidas pela Prof^a. Dra. Ana Maria Costa Leonardo do Instituto de Biociências da UNESP de Rio Claro.

Como o ninho de *Coptotermes gestroi* é subterrâneo, é muito difícil encontrá-lo na natureza. Por isso utilizamos armadilhas celulósicas para a captura de operários, ninfas e soldados forrageiros (que saem do ninho para coletar alimento e explorar o ambiente) distribuídas em locais infestados na cidade de Rio Claro - São Paulo. Os alados foram coletados na época da revoada (agosto e setembro de 2007) em Campinas e em Rio Claro. Estes também foram utilizados para a montagem de colônias incipientes mantidas em laboratório da seguinte forma: um alado macho e um alado fêmea são separados (a morfologia é identificada com o auxílio de lupa) e colocados em uma placa de *petry* de 7 cm de diâmetro com pó de serragem úmida. Neste local eles estabelecem a colônia e começam a se proliferar. Como estas colônias eram muito recentes, a quantidade de cupins ainda era pequena. Por isso, para a coleta de casal real e larvas, utilizamos outras colônias mais antigas mantidas no Laboratório de Cupins da Profa. Dra. Ana Maria Costa Leonardo.

2 – Extração do RNA

Para a extração do RNA, os cupins foram macerados em *potters* de 2mL. Para cada 100mg do material macerado (aproximadamente 1/3 do tubo *eppendorf* de 1,5mL) foi adicionado 1mL de TRIzol® (Invitrogen). As amostras foram homogeneizadas através de várias pipetagens até que se tornassem bastante fluidas.

Durante a homogeneização, o reagente TRIzol® mantém a integridade do RNA enquanto rompe as células e dissolve os componentes celulares. As amostras foram centrifugadas por 10 minutos a 14.000rpm em centrífuga refrigerada a 4°C.

O sobrenadante foi transferido para um novo tubo *eppendorf* de 1,5mL e foi incubado por 5 minutos à temperatura ambiente para a completa dissociação dos complexos

nucleoproteicos. Acrescentou-se às amostras 200µL de clorofórmio (CHCL₃) que foram em seguida vigorosamente agitadas. Nova incubação foi realizada por mais 3 minutos à temperatura ambiente e as amostras foram centrifugadas por 15 minutos a 14.000rpm em centrífuga refrigerada a 4°C. A fase aquosa, onde fica o RNA, foi recuperada e transferida para um novo tubo de microcentrífuga e a partir desta etapa os Kits de extração RNeasy® Mini Kit (QIAGEN) ou RNeasy® Micro Kit (QIAGEN) foram utilizados seguindo as recomendações do fabricante para a obtenção de um RNA purificado.

A concentração de RNA foi determinada em espectrofotômetro NanoDrop ND – 1.000 através de leitura da densidade óptica no comprimento de onda de 260nm. A integridade e qualidade das amostras de RNA total extraídas das várias castas de *C. gestroi* foi verificada através da eletroforese em gel de agarose 1,2% com brometo de etídeo. As amostras de RNA foram estocadas em freezer -80°C.

Vários operários e alados de *C. gestroi* foram utilizados para a extração de RNA pelo método de TRIzol. A extração de RNA dos soldados foi realizada com o auxílio do RNAeasy mini kit (Qiagen). As larvas foram colocadas no TRIzol e com a ajuda de um sonicador, os tecidos foram rompidos. Foram utilizados três pulsos de 15 segundos cada para a extração de RNA pelo método de TRIzol, no entanto o material não foi macerado..

3 – Construção das bibliotecas subtrativas

A técnica utilizada foi a Hibridização Subtrativa Suprimida que permite comparar duas populações de RNA mensageiro (RNAm) e obter clones de genes que estão sendo expressos diferencialmente em dois tecidos ou células, no nosso caso, diferentes castas do cupim *Coptotermes gestroi*. A larva foi utilizada como controle para ser subtraída das duas linhagens; a áptera e a ninfal. Para representar a linhagem áptera foi feito um *pool* de cDNA de operários e soldados. A linhagem ninfal foi representada por um *pool* de cDNA de ninfas, alados e rainhas. Portanto, quatro bibliotecas subtrativas foram construídas: (1) larva subtraída da linhagem áptera, (2) linhagem áptera subtraída da larva, (3) larva subtraída da linhagem ninfal e (4) linhagem ninfal subtraída da larva.

Inicialmente, as duas populações de RNAm foram convertidas em cDNA, de modo que uma população foi denominada de “tester” e a outra população de “driver”. As duas amostras de cDNA, “tester” e “driver” foram digeridas com a enzima de restrição *Rsa* I

para formação de extremidades abruptas, permitindo a ligação de adaptadores. Para que esta ligação seja feita, a amostra de mRNA “tester” é dividida em duas alíquotas e cada uma delas é ligada a um adaptador diferente. Esses adaptadores tratam-se de moléculas de DNA que serão utilizados nas etapas finais de amplificação por PCR, conforme discutido adiante. Cada uma das alíquotas de cDNA “tester” foi hibridizada separadamente com cDNA “driver” em excesso (Figura5A). Desta maneira as moléculas que são expressas apenas na população de cDNA “tester” ficarão na condição de fita simples (Figura 5Ba) , enquanto as moléculas comuns às duas populações sofrerão hibridização, com a consequente formação de estruturas de fita dupla (Figura5Bb e c) . Cabe lembrar que as moléculas que são específicas da população “driver” também estarão na forma de fita simples, mas as mesmas não terão os adaptadores ligados às suas extremidades e por isso não serão amplificadas por PCR nas etapas posteriores (Figura5Figura 5Bd). Nesta etapa as moléculas diferencialmente expressas na população “tester” encontram-se na condição de fita simples, porém com adaptadores diferentes em cada uma das duas alíquotas em que foram subdivididas inicialmente. O segundo passo de hibridização consiste em misturar as duas alíquotas de cDNA “tester”, sem a etapa de desnaturação, com a adição de mais cDNA de driver denaturado, de maneira que apenas as moléculas fita simples (e específicas) ligar-se-ão para a formação de estruturas em fita dupla (Figura5Ce). Seguida a hibridização, as extremidades de todas as moléculas fita dupla foram tratadas com a enzima DNA polimerase, de maneira a obterem-se extremidades abruptas (Figura5D). A posterior amplificação por PCR ocorreu de forma exponencial apenas para as moléculas específicas com estrutura em fita dupla, pois são as que possuem os adaptadores diferentes em suas extremidades (Figura5Ee). As moléculas que possuem adaptadores iguais em suas extremidades têm a tendência de formarem estruturas em grampos e assim não foram amplificadas (Figura5Eb). A outra estrutura em fita dupla sofreu amplificação linear por possuir apenas um dos adaptadores ligados (Figura5Ec). As outras estruturas não foram amplificadas por estarem presentes em fita simples (Figura5Da e d). Todo esse procedimento permite um enriquecimento da população de cDNAs específicos da amostra “tester”. Outra biblioteca foi construída (biblioteca “reverse”) onde a amostra que foi usada como “tester” foi a de operários e soldados. Assim na biblioteca “forward”, serão identificados os transcritos presentes apenas nas larvas e na “reverse” os transcritos

presentes apenas em operários e soldados. A biblioteca subtrativa foi construída utilizando-se o kit SSH da Clontech (Clontech Laboratories, Inc., Palo Alto, CA), de acordo com as instruções do fabricante.

Os fragmentos resultantes da subtração foram clonados no vetor pGEM-T (Invitrogen), o qual é específico para clonagens provenientes de amplificações por PCR. A verificação da inserção foi feita utilizando-se o reagente X-GAL. Assim, as colônias que receberam o plasmídeo com inserto, cresceriam em placas com o antibiótico ampicilina (marca de resistência do plasmídeo) e não teriam a ativação do gene da beta galactosidase e por isso não teriam a capacidade de degradar o X-GAL, fazendo com que as colônias permanecessem brancas. Estas colônias foram inoculadas em meio líquido LB (1% Peptona, 0,5% de Extrato de Levedura e 0,5% de NaCl), contendo ampicilina (100ug/ml) e crescidas durante a noite sob agitação de 120 rpm.

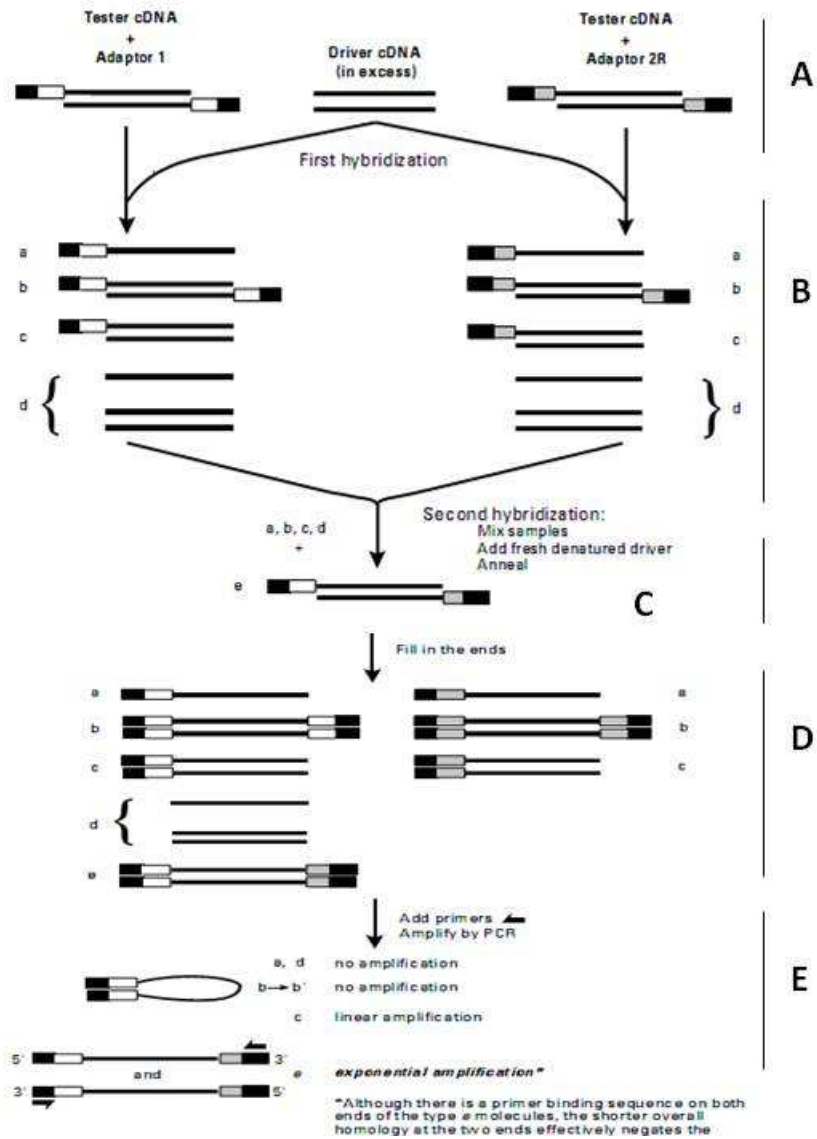


Figura 5 - Esquema com os passos envolvidos na construção da biblioteca subtrativa (adaptado do manual do Kit SSH da Clontech).

3.1 – Clonagem, amplificação e sequenciamento dos clones

Para o sequenciamento dos clones, foram realizadas reações de PCR para amplificar os fragmentos clonados utilizando 2µl do sobrenadante da cultura crescida (plasmídeo); 1,5µl de dNTPs 1,25mM; 0,5µl de MgCl₂ 50mM; 0,35µl de cada *primer* M13 “sense” e “antisense” 5pmol; 1U de Taq DNA polimerase (Invitrogen); 8,6µl de H₂O e 1,5µl de tampão 10X. As temperaturas utilizadas foram as seguintes: denaturação inicial de 94°C por 2 minutos seguidos de 35 ciclos de 94°C por 20 segundos, 57°C por 15 segundos e 72°C por 1 minuto. Os fragmentos amplificados foram seqüenciados utilizando o *primer* M13 “sense” ou “anti-sense” no aparelho MEGA BACE 1000 DNA Analysis System (Molecular Dynamics/ Amersham - Life Science) seguindo as recomendações do fabricante.

3.2 – Bioinformática e análise dos dados

As seqüências geradas foram submetidas à análise de qualidade de bases pelo programa PHRED (Ewing e Green, 1998), clusterizadas e trimadas pelo programa CAP3 (Huang e Madan, 1999) e anotadas automaticamente pelo BLAST N, X e tBLAST X (www.ncbi.nlm.nih.gov/blast) contra o banco de dados. Além disso, as seqüências foram agrupadas a partir do “first hit” gerado pelo BLAST com uma nova ferramenta desenvolvida para agrupar seqüências de bibliotecas subtrativas. Os “reads” estão disponíveis para a mineração, a partir do momento que forem gerados, através do uso do software GeneProjects (Carazzolle, Araújo, Formighieri, Digiampietri, Pereira, protocolo no INPI em 09.01.2004, nº 0057290).

4 – Síntese de DNA Complementar (cDNA) para PCR em tempo

Real

A leitura do RNA total foi realizada em espectrofotômetro (NanoDrop Technologies, Inc., Rockland, DE). As amostras de RNA obtidas foram submetidas à síntese de DNA complementar (cDNA) utilizando-se o kit Superscript III RTTM (Invitrogen).

Para a transcrição, utilizou-se 1 µg de RNA que foi tratado com a enzima DNase I (Invitrogen) para remoção de DNA contaminante. Este tratamento consistiu na adição de 1,0µl de 1u/µl DNase I; 1,0µl de 10x DNase I *Reaction Buffer* (200mM Tris-HCl, 20 mM

MgCl₂, 500 mM KCl₂) e água suficiente para um volume final de 10,0µl de reação. As amostras foram incubadas por 15 minutos em temperatura ambiente e a reação foi bloqueada com a adição de 1,0µl de 25 mM de EDTA e posterior incubação por 5 minutos à 65°C. Após essa etapa, iniciou-se a síntese do cDNA complementar com a adição de 1,0µl de 50 µM oligo (dT)₂₀ e 1,0µl de 10 mM dNTP's. As amostras foram incubadas por 5 minutos à 65°C e em seguida por 1 minuto à 4°C. Para cada amostra, adicionamos 10,0µl da seguinte mistura de reação: 2µl de 10xRT *buffer*; 4,0µl de 25 mM MgCl₂; 2,0µl de 0,1 M DTT; 1,0µl de 40U/µl Rnase OUTTM e 1,0µl de 200 U/µl Superscript III RTTM. A reação ocorreu por 50 minutos à 50°C, seguida de 5 minutos à 85°C.

4.1 – Verificação da Síntese de DNA Complementar

A verificação da síntese de cDNA foi feita através da reação de polimerase em cadeia (PCR) para o gene da beta-actina (β-actina). As reações de PCR foram realizadas com: 1,5µl 10xPCR *buffer* (20 mM Tris-HCl, 500 mM KCl); 0,5µl 50 mM MgCl₂; 1,5µl 1,25 mM dNTP's; 0,35µl 10 mM *primer* BACF (5'–GTTGTCAAATGTAATCCATCC – 3'); 0,35µl de 10 mM de *primer* BACR (5'–CTGACGTACTGTGCGCCTGT – 3'); 0,2µl de Taq DNA polimerase; 1,0µl de cDNA e 9,6µl de água, para um volume final de 15,0µl. O programa foi iniciado por 2 minutos à 94°C, seguido de 35 ciclos: 94°C/15 segundos, 58°C/20 segundos e 72°C/1 minuto, sendo finalizado por 72°C/7 minutos. Os produtos da PCR foram submetidos à eletroforese em gel de agarose 1,5% para verificar a amplificação de um fragmento de 127 pares de base.

4.2 – Desenho dos Primers para a reação do Real Time PCR

Os *primers* utilizados nas reações de Real Time PCR foram desenhados com o uso do site www.invitrogen.com e com o *software* “Primer Express” (Applied Biosystems). Os fragmentos desenhados foram submetidos ao programa *Blast* (www.ncbi.nlm.nih.gov/blast) para confirmar homologia com o gene de interesse. A formação de estruturas como *hairpins* e *dimers* também foram avaliadas através do programa Gene Runner. Os *primers* utilizados neste estudo estão listados na Tabela 1.

Tabela 1 – Sequência e tamanho dos fragmentos amplificados de cada par de primer utilizado na amplificação dos genes estudados pela técnica de Real Time PCR.

Gene	Sequência <i>Primer</i>	Tamanho do fragmento amplificado
AACT – F AACT - R	5´- ACTTGCCTTTTCGTCCTCCTT – 3´ 5´- TTCAGGAGCACGCATAACAACA – 3´	118 pb
α -TUB-F α -TUB-R	5´- CCCTCTCCTTCGCCCTTCAA -3´ 5´-GAGGATCTCGCTGCTCTGGA -3´	71 pb
α -AMILASE-F α -AMILASE-R	5´- ATATGGCTCTTGTCCTCTGCG-3´ 5´- GCACTTCGGCCCTCTACAGT-3´	71 pb
BAC-F BAC-R	5´- GTTGTCAAATGTAATCCATCC-3´ 5´- CTGACGTACTGTGCGCCTGT -3´	127pb
BETAGLICO - F BETAGLICO - R	5´- GATAATAGAGTTGTACCGTGG – 3´ 5´- ACTCCACCGTAGTCCGAGAAAC – 3´	113 pb
CALR-F CALR-R	5´- GCACATTATCAAAGATTGTTCCAG-3´ 5´- GCGTGATGAAATTTGTGCCA-3´	71pb
COX-3-F COX-3-F	5´- TGGGCTCATGTTACGGTTACTC -3´ 5´- ATCCCTACAAGTCCCCTACTAA -3´	70 pb
CHIT-F CHIT-R	5´-GATCATCCCATCTCACATTCCA-3´ 5´-ACAAGAGAGGCAGGCACCC-3´	78pb
EF1A-F EF1A-R	5´- CAACAGTCTGCCTCATGTCACG-3´ 5´- TGTGGAATCTTTCTCGGGAATT-3´	71pb
ENDO-F ENDO-R	5´- GTGCCTTCAACTGGGATAACAA-3´ 5´- TGTATGCCTGCTTGCTTGTGA-3´	61pb
FERRITINA-F FERRITINA-R	5´- CAATTGCATCATCCCAGGTG -3´ 5´- TATGAACAGAGTCGTGAAGGATTT- 3´	73 pb
HEX-1-F HEX-1-R	5´- CATCAGGGTAGAAAGGAGGCAG-3´ 5´- GTTGCCGTGTATTTTGGTAGAGTT-3´	71 pb
HEX-2-F HEX-2-R	5´- TGCTGAGTCCAACGTCTTGC -3´ 5´-AGCTAACTACACTGTGCACACCC-3´	68 pb
HMGS-F HMGS-R	5´- ACAGACATCCACGAAGTAGACAG – 3´ 5´- GAATACATGGAGGAGCAGAGC – 3´	77pb
JHEH-F JHEH-R	5´- TGTCCATGTTGAGAACTTTTCCA-3´ 5´- GGTGTTGCCCTTCGTGATTC – 3´	69 pb
LACASE – F LACASE – R	5´- GTTTGCAGAAGCTCGATGGC – 3´ 5´- TCCGTTTCGAGTCGTCATTTACA – 3´	71 pb
LISOZIMA-F LISOZIMA-R	5´- TGAGCAAATGAGTAAAGCCTCG -3´ 5´- TGAGTAATGGGCAAGTGCGAC -3	71 pb
REGENECTIN-F REGENECTIN-R	5´- CGGCCTCAACGACTTACCTT- 3´ 5´- GTTCTCCTACCACAGTTCCTGTTC-3´	70 pb
PRO-H-CONV-F PRO-H-CONV-R	5´- GAAACCAGATCACCTACGCTT– 3´ 5´- CATCCCTGGCTAGACGCTGT- 3´	68pb
SORBITOL-F SORBITOL-R	5´- TGCATTCCCTTGTCGTGATG – 3´ 5´- ATCCAACAGCTGCCACCACT- 3´	72 pb
VITELOGENIN-F VITELOGENIN-R	5´- CGCTACGTTGCTAGTCTGGTTT-3´ 5´- CCTGCTGGAGTGCCTGGTC -3´	71 pb

4.3 – Padronizações necessárias para Real Time PCR

4.3.1 – Concentração de Primer

Na Real Time PCR (qRT-PCR) a concentração ideal de *primer* a ser utilizada, deve ser a mínima suficiente para permitir a duplicação de todas as cópias do material a cada ciclo de reação. Para isso, utilizando a mesma quantidade de amostra, foram feitas reações contendo cada um dos *primers* (sense e anti-sense) nas concentrações finais de 70nM, 150 nM e 300 nM. A concentração ideal foi aquela em que o gene de interesse obteve o menor valor de Ct (ciclo *threshold*), ou seja, foi a concentração em que o gene foi amplificado mais precocemente. Na Tabela 2 estão listadas as concentrações dos *primers* utilizados na amplificação dos genes de estudo e a eficiência de amplificação obtida.

Tabela 2 – Concentrações de primers utilizadas na amplificação dos genes de estudo e eficiência da amplificação obtida. As concentrações foram definidas pela eficiência de amplificação gerada sob as condições testadas.

Primer	Concentração utilizada	Eficiência do primer
AACT	150nM	100%
α -TUB	150 nM	100%
α -AMILASE	300 nM	100%
BAC	300 nM	100%
BETAGLICO	150nM	100%
CALR	150 nM	100%
COX-3	150 nM	100%
CHIT	150 nM	100%
EF1A	300 nM	100%
ENDO	150 nM	100%
FERRITINA	300 nM	100%
HEX-1	300 nM	100%
HEX-2	150 nM	100%
HMGs	150nM	100%
JUV- H	300 nM	100%
LACASE	70nM	100%
LISOZIMA	150 nM	100%
REGECTINA	300 nM	100%
PRO-H-CONV	300 nM	100%
SORBITOL	150 nM	100%
VITELOGENINA	150 nM	100%

4.3.2 – Eficiência de Reação

A reação de Real Time PCR precisa ser confiável e reprodutiva. Para isso são necessárias condições ótimas de reação, onde as amplificações apresentem 100% de eficiência a cada ciclo da reação de PCR. A eficiência de amplificação é obtida através da fórmula $10^{(-1/slope)}$, onde *slope* é derivado da inclinação da curva, e esse valor quando aplicado a fórmula de eficiência, tem que ser próximo de 2, que significa que a cada ciclo de amplificação o material genômico está sendo duplicado e isso equivale a 100% de eficiência (Meijerink *et al.*, 2001).

4.4 – Real Time PCR

A PCR em tempo real é uma técnica utilizada para a avaliação da expressão gênica. A expressão é dita relativa porque leva em consideração também a expressão de controles endógenos como, beta actina e GAPDH (gliceraldeído-3-fosfato desidrogenase). A técnica consiste no monitoramento óptico da fluorescência emitida durante a reação de PCR (Higuchi *et al.*, 1993), através da ligação de uma sonda específica ou um corante, na fita recém sintetizada.

As amostras de cDNAs utilizadas na reação de qRT-PCR foram quantificadas em espectrofotômetro (NanoDrop Technologies, Inc., Rockland, DE).

Foram feitas três replicas biológicas em duplicata utilizando como sistema de detecção o reagente SYBRTM Green PCR Master Mix[®] (Applied Biosystems, USA), que além de conter todos os reagentes necessários para PCR (dNTP's, MgCl₂, tampão, Taq Ampli-Gold), também contém o corante SYBR Green; uma molécula que emite maior quantidade de sinal fluorescente ao intercalar com DNA dupla fita. Para cada gene analisado foram utilizados *primers* específicos.

A detecção da amplificação em tempo real foi realizada no equipamento ABI PRISM 7300 (Applied Biosystems, USA) em gráficos de fluorescência versus número de ciclos. Quanto maior a expressão de um determinado gene, isto é, quanto mais cópias existirem dele no início da reação, mais precocemente ocorrerá a amplificação e consequentemente, menor será o Ct.

Todas as amostras foram ensaiadas em 12,0µl de volume final; sendo 6,0µl do reagente SYBR Green PCR Master Mix[®], concentrações de cDNA variando de 2ng,

6,32ng, 20ng, 63,26ng, 120ng e 200ng e a concentração ótima de *primer* determinada. Em todas as placas foram pipetados controles negativos (NTC), ou seja, para cada *primer* avaliado foram incluídos poços contendo água estéril em substituição à amostra. As reações foram realizadas em placas de 96 poços (Sorenson, BioScience Inc) com tampas plásticas que permitem a passagem da luz. O programa foi iniciado a 95°C/10 minutos, seguido de 40 ciclos: 95°C/15 segundos e 60°C/1 minuto. Ao final de uma amplificação normal adiciona-se um passo de 20 minutos, no qual a temperatura aumenta gradualmente de 60° para 95°C. À medida que os produtos gerados por PCR desnaturam com o aumento de temperatura, cai o sinal fluorescente do SYBR Green. O gráfico resultante permite verificar se há um ou mais produtos de PCR presentes em cada reação, devido a diferenças de T_m (temperatura de *melting*) entre os produtos de PCR causadas pelo número e composição de bases de cada produto.

4.5 – Análise dos Dados da Real Time PCR

A expressão dos genes de interesse foi determinada de forma relativa, ou seja, os valores de expressão são normalizados em relação a genes chamados de “controles endógenos”. Neste trabalho foi utilizado o gene da β -actina e α - tubulina, que são genes cuja expressão é dita constitutiva, ou seja, apresentam pouca variação entre diversas condições.

Os valores de Ct de cada gene foram obtidos em duplicata, dessa forma, foram calculadas as médias aritméticas destas, para cada amostra que teve um determinado gene estudado. Em seguida foi obtido o valor de delta Ct (ΔCt), que consistiu na subtração do menor Ct de um gene de uma determinada amostra com os demais Cts desse mesmo gene de outras amostras estudadas. O próximo cálculo foi o de “Q” valor, que corresponde a $2^{\Delta Ct}$. Os valores de Q dos genes β -actina e α - tubulina de cada amostra analisada foram submetidos ao programa geNorm (Vandesompele, et al., 2002), que calculou a média geométrica entre eles, valor este denominado de Fator de Normalização da amostra. A expressão normalizada de um dado gene de interesse em uma determinada amostra foi dada pela razão entre o valor de Q do gene de interesse da amostra pelo fator de normalização da amostra. O valor obtido foi expresso em unidades arbitrárias (U.A.) ou valor absoluto de expressão.

4.6 – Análise Estatística

Para avaliar os padrões de expressão foi realizada uma análise de variância (ANOVA) separadamente para cada gene utilizando as castas como fatores fixos (larva, linhagem neutra e linhagem imaginal). O teste comparativo de Tuckey foi realizado para fazer comparações entre pares. As análises estatísticas e os gráficos foram confeccionados utilizando o programa Graph Pad Prism (versão 5.0).

5 – Sequenciamento de segunda geração

Amostras de RNAm de operários e soldados de *C. gestroi* foram sequenciados utilizando a técnica de sequenciamento de segunda geração com o equipamento 454 Life Sciences (ROCHE), a qual foi a primeira plataforma de sequenciamento de nova geração a ser comercializada.

A plataforma 454 realiza o sequenciamento baseado em síntese, o pirosequenciamento (Ronaghi *et al.*, 1998). A leitura da sequência nesse sistema é realizada a partir de uma combinação de reações enzimáticas que se inicia com a liberação de um pirofosfato, oriundo da adição de um desoxinucleotídeo à cadeia. Em seguida, esse pirofosfato é convertido para ATP, pela ATP sulfúrilase, sendo este utilizado pela luciferase para oxidar a luciferina, produzindo um sinal de luz capturado por uma câmera CCD (*charge-coupled device*) acoplada ao sistema. O sistema requer que o DNA seja mecanicamente fragmentado em sequências de 300 – 800pb, transformado em fragmentos abruptos fosforilados e ligado a adaptadores de sequência específica. A biblioteca de DNA da amostra é ligada a adaptadores A e B nas extremidades 3' e 5' dos fragmentos, respectivamente, os quais são utilizados nas etapas posteriores de isolamento dos fragmentos (A-B) e amplificação e nas reações de sequenciamento (Figura 6a). O adaptador B possui biotina ligada à extremidade 5', o que permite o isolamento dos fragmentos ligados ao adaptador A na extremidade 3' e adaptador B na extremidade 5' na amostra. Somente os fragmentos AB são eluídos na reação de purificação e são especificamente ligados às microesferas que carregam várias cópias da sequência complementar exata ao adaptador B de um único fragmento (Margulies *et al.*, 2005). O outro adaptador é utilizado no anelamento do *primer* que inicia a reação de sequenciamento. As microesferas ligadas

aos fragmentos únicos de fita simples são então emulsionadas em uma mistura de água e óleo com reagentes de

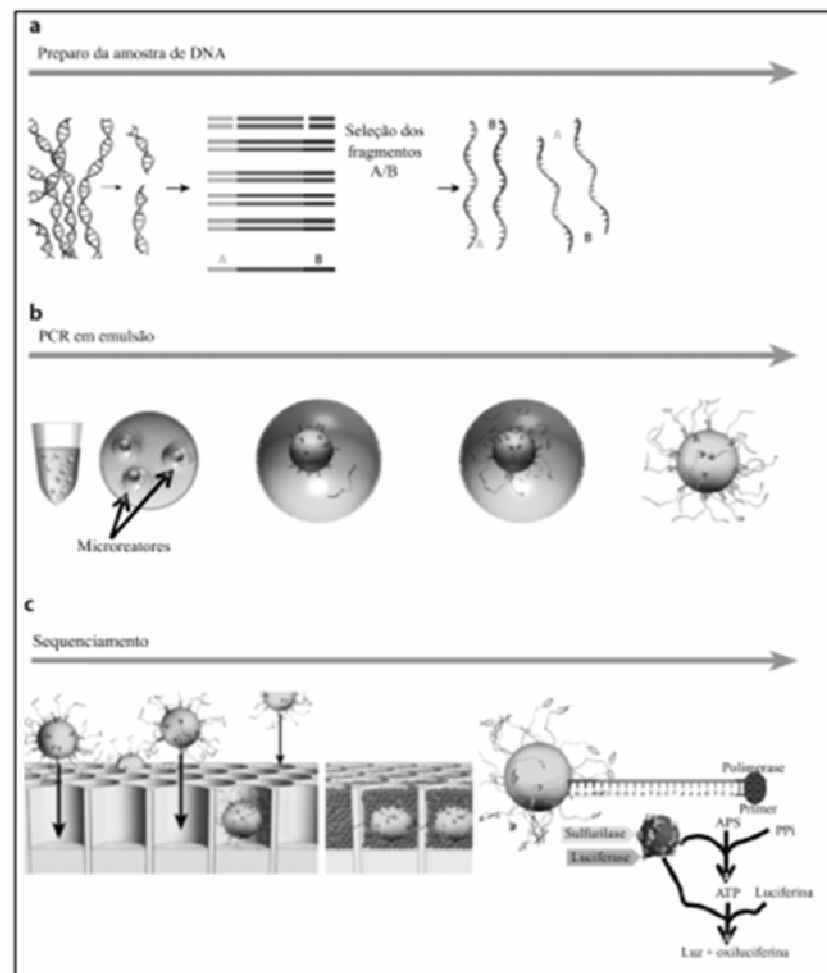


Figura 6 – Resumo ilustrativo do sequenciamento (adaptado do site da ROCHE).

PCR para amplificação clonal do fragmento fita simples em cerca de 1 milhão de cópias. Na PCR em emulsão, o óleo em solução aquosa forma micelas, nas quais as microesferas são capturadas (Figura 6b). Cada micela funcionará como um microrreator, produzindo muitas cópias idênticas de um mesmo fragmento isoladamente em um microsuporte (Dressman *et al.*, 2003). Após a PCR de emulsão, as microesferas ligadas aos fragmentos de fita simples são depositadas em poços distintos em uma placa de sílica onde os reagentes para o sequenciamento são distribuídos. As reações de sequenciamento ocorrem em cada poço, para um único tipo de fragmento ligado à microesfera, não

havendo, portanto, competição por reagentes com outros fragmentos da biblioteca (Figura 6c). A placa de sequenciamento é dividida em 1,6 milhões de poços com diâmetro suficiente para alojar uma única microesfera. Após a placa ser inserida junto ao sistema óptico de leitura no equipamento, os reagentes e as soluções de sequenciamento são então distribuídos a cada ciclo para obtenção do sequenciamento paralelo dos 1,6 milhões de poços. O sequenciamento é realizado em ciclos, sendo que em cada um deles, é adicionado um tipo determinado de nucleotídeo. Caso este seja incorporado à sequência em síntese, um sinal de luz é emitido. Assim, a intensidade desse sinal é reflexo do número de nucleotídeos desse tipo específico que foram sucessivamente incorporados na molécula. Como o nucleotídeo que é adicionado a cada ciclo é conhecido, o sinal de luz emitido pode ser diretamente utilizado como informação de sequência (Ronaghi, 2001). Os fragmentos sequenciados nessa plataforma passam por sistemas de análise de qualidade, onde sequências distintas oriundas do sequenciamento de uma única microesfera são eliminadas, bem como as leituras em que a sequência inicial TCGA (quatro primeiros nucleotídeos dos adaptadores) não aparece. As leituras produzidas possuem geralmente cerca de 400pb, o que representa um comprimento de leitura menor que o produzido pelo sistema de Sanger (~700pb). O maior tamanho das leituras e a grande capacidade de gerar informação tornam o processo de montagem mais fácil num projeto de sequenciamento *de novo* (sequenciamento de genomas desconhecidos) e permite trabalhar com coberturas genômicas mais amplas, favorecendo o processo de montagem. Além disso, pelo fato de não envolver clonagem bacteriana, a representação do genoma é bem mais fiel, de forma que sequências difíceis de clonar e manter em bibliotecas genômicas podem ser acessadas. Com relação às demais tecnologias de sequenciamento da segunda geração, a plataforma 454 é a que produz as maiores leituras e por isso foi a metodologia escolhida para este trabalho, já que ainda não foram publicados genomas completos de cupins para ajudar na montagem das sequências.

5.1 – Bioinformática e análise dos dados

Os adaptadores utilizados no pirosequenciamento foram identificados pelo *software* ssaha2 (Ning, *et al.*, 2001), e a montagem do transcriptoma foi realizada pelo *software* “miraEST assembler” (Chevreux *et al.*, 2004) usando os parâmetros *default* de

sobreposição e identidade. Após os procedimentos de exclusão das sequências de baixa qualidade e das regiões não relevantes (poli A e adaptadores) as sequências foram arranjadas em “contigs”, que são sequências que se sobrepõem para formar uma única sequência de um gene ou fragmento de um gene. Os “contigs” formados foram submetidos a uma análise de expressão diferencial pelo programa DEGseq (Wang *et al.*, 2010) para normalizar e identificar genes diferencialmente expressos entre as duas castas. Todas as sequências foram submetidas à busca de similaridade BLAST X (www.ncbi.nlm.gov/blast).

RESULTADOS

1 - Extração de RNA

Vários operários e reprodutores alados de *C. gestroi* foram utilizados para a extração de RNA pelo método de TRIzol (7a). A extração de RNA dos soldados (7b) foi realizada com o auxílio do RNAeasy mini kit (Qiagen), pois esta casta possui uma enorme glândula que ocupa quase seu corpo inteiro e que produz uma substância lipídica que estava interferindo na extração pelo método de TRIzol. Para a extração do RNA da rainha (7c) também foi utilizado o RNAeasy micro kit (Qiagen) devido à pequena quantidade de material, já que uma rainha criada em laboratório pesa de 7 a 9mg. Algumas larvas foram utilizadas para a extração de RNA pelo método de TRIzol (7d), no entanto o material não foi macerado com nitrogênio. As larvas foram colocadas no TRIzol e com a ajuda de um sonicador, os tecidos foram rompidos. Foram utilizados três pulsos de 15 segundos cada.

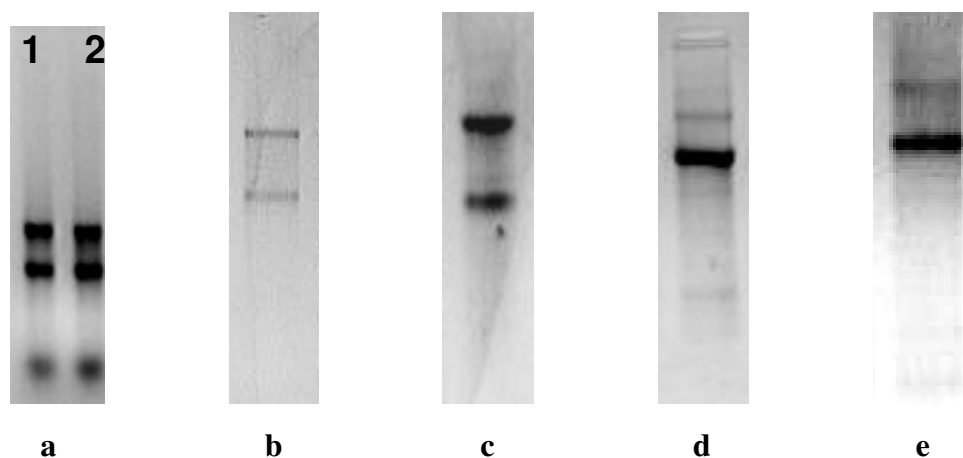


Figura 7 - RNA de *C. gestroi* - **a** - RNA de operário de *C. gestroi* (1) ao lado do RNA de reprodutor alado do cupim *C. gestroi* (2). **b** - RNA de soldado de *C. gestroi*. **c** - RNA de rainha do cupim *C. gestroi*. **d** - RNA de ninfa de *C. gestroi*. **e** - RNA de larvas de *C. gestroi* visualizada em gel de agarose 1,2 % sob luz ultravioleta.

2 – Construção das bibliotecas subtrativas

As bibliotecas subtrativas foram construídas a partir do kit SSH da Clontech (Clontech Laboratories, Inc., Palo Alto, CA). Foram realizadas quatro bibliotecas subtrativas: (1) larva subtraída da linhagem áptera, (2) linhagem áptera subtraída da larva, (3) larva subtraída da linhagem ninfal e (4) linhagem ninfal subtraída da larva. Foram

sequeenciadas 10 placas (960 seqüências) de cada biblioteca e todas as seqüências foram submetidas ao banco de dados do nosso laboratório (www.lge.ibi.unicamp.br/cupim).

Os valores de qualidade de bases produzidos pelo PHRED foram usados por um “script” de trimagem de seqüências similares a vetores, poli-A, poli-T e seqüências contaminantes, como provenientes de *E. coli* (Telles e da Silva, 2001). As seqüências foram submetidas à busca por similaridade através da comparação direta de nucleotídeos (BLAST N) e comparação após tradução dinâmica (BLAST X) contra os bancos de dados “non-redundant” (NR). Após a realização da busca de similaridade, as seqüências foram submetidas a uma análise de biblioteca subtrativa na página de anotação do banco de dados no nosso laboratório (www.lge.ibi.unicamp.br/cupim) (Figura 8). Essa análise agrupa as seqüências que apresentaram o mesmo “*first hit*” no BLAST X. Em seguida esses agrupamentos foram anotados manualmente através dos dados fornecidos pelas buscas de similaridade e também pela anotação automática do programa Gene Ontology (www.godatabase.org/cgi-bin/amigo/go.cgi)



Exibindo resultados de 1 a 50 de 168				
	mRNA	Total de reads	Reads CM01-S3-008	Reads CM01-S3-009
Grupo 168	No hits found	1677	858	819
Grupo 167	gi 133916482 emb CAM36311.1 hypothetical protein [Thermobia domestica]	33	6	27
Grupo 166	gi 89271365 emb CAJ83813.1 CHK1 checkpoint homolog (S. pombe) [Xenopus tropicalis]	5	2	3
Grupo 165	gi 66957915 gb AAU20851.2 hexamerin I [Reticulitermes flavipes]	4	2	2
Grupo 164	gi 124107286 dbj BAF45420.1 arrestin2 precursor [Dianemobius nigrofasciatus]	3	0	3
Grupo 163	gi 48476579 gb AAT44492.1 cytochrome oxidase subunit I [Neotermes koshunensis]	3	2	1
Grupo 162	gi 17946496 gb AAL49280.1 REF74144p [Drosophila melanogaster]	3	1	2

Figura 8 - Página do projeto cupim do laboratório de genômica e expressão

Das bibliotecas de larva x linhagem áptera (1 e 2) foram sequeenciadas 20 placas, que geraram 1.920 seqüências, que depois de “trimadas” totalizaram 1.846 seqüências de

boa qualidade. Após a submissão de busca por similaridade das sequências geradas foram formados dois grupos: as sequências que possuíam alguma similaridade com algum gene conhecido ou as sequências “no hits”. A quantidade de sequências que não apresentaram similaridade com nenhum gene foi elevada (86,67%).

As sequências que obtiveram similaridade com genes conhecidos foram agrupadas em 237 grupos de sequência com o mesmo “*first hit*”, anotadas e classificadas de acordo com a busca por processo biológico no Gene Ontology (Figura 9). Vinte e cinco por cento das sequências foram relacionadas com metabolismo (10% processos metabólicos de proteína, 6% metabolismo de carboidratos, e 9% outros processos metabólicos) seguido por sequências relacionadas à transporte (16%). As demais sequências apresentaram similaridade com outros processos biológicos, tais como; proteínas de reserva da hemolinfa (5%), polimerização de proteína constitutiva (7%), resposta à estímulo (2%), fosforilação de aminoácidos (3%), mitocondrial/ribossomal (11%), sinalização celular (5%), tradução de sinal (3%), desenvolvimento (4%), transcrição (4%), e outros (15%).

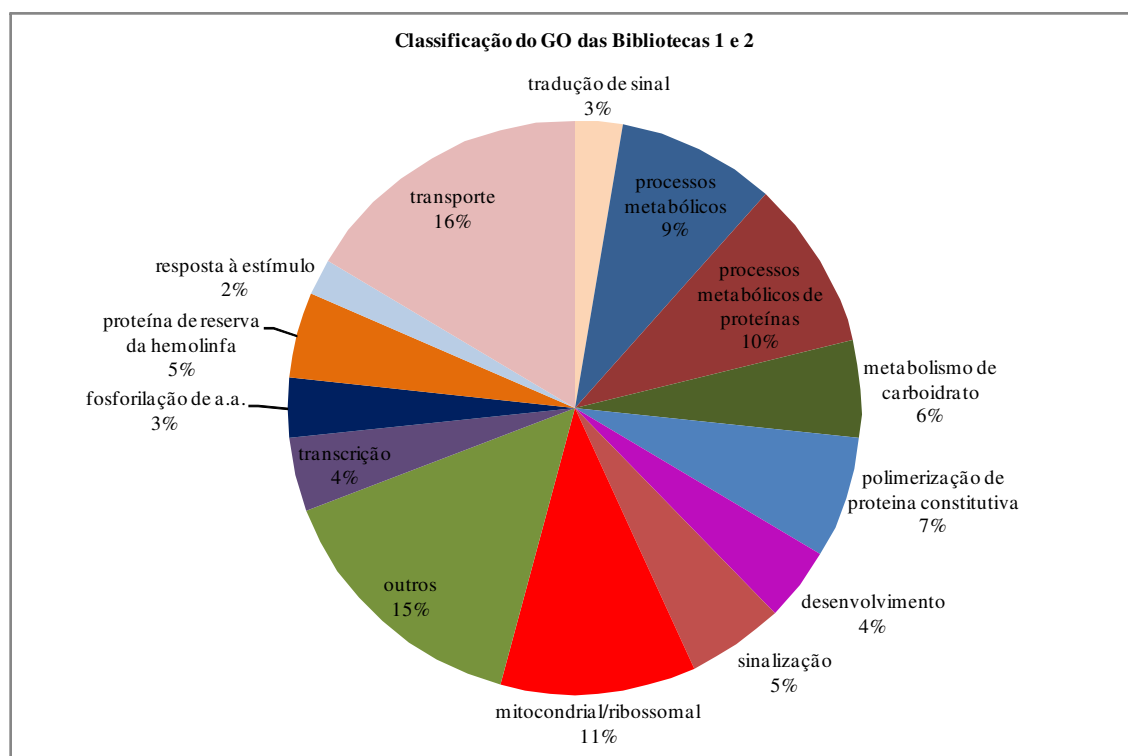


Figura 9 - Gráfico da anotação das sequências que obtiveram similaridade com o BLAST/ GO das bibliotecas 1 e 2 (larva X linhagem áptera).

Das bibliotecas de larva x linhagem ninfal (biblioteca 3 e 4) foram seqüenciadas 20 placas, que após trimadas geraram 1.898 seqüências de boa qualidade, as quais foram submetidas a busca por similaridade das seqüências geradas originando dois grupos: as seqüências que possuíam alguma similaridade com algum gene conhecido ou as seqüências “no hits”. A quantidade de seqüências que não apresentaram similaridade com nenhum gene foi elevada (88,35%).

Logo após à submissão de busca por similaridade, as seqüências foram agrupadas baseadas no seu “first *hit*” do BLASTX. Essa análise gerou 168 grupos, os quais foram anotados manualmente através dos dados fornecidos pelas buscas de similaridade e também pela anotação automática do programa Gene Ontology (Figura 10). Trinta por cento das seqüências foram relacionadas com metabolismo (13% processos metabólicos de proteína, 2% metabolismo de carboidratos, 6% regulação de processos metabólicos, e 9% outros processos metabólicos) seguido por seqüências relacionadas à transporte (18%). As demais seqüências apresentaram similaridade com outros processos biológicos, tais como; proteínas de reserva da hemolinfa (5%), polimerização de proteína constitutiva (7%), resposta à estímulo (2%), fosforilação de aminoácidos (4%), mitocondrial/ribossomal (12%), regulação de transcrição (7%), resposta a estresse (1%), embriogênese (1%), e outros (13%).

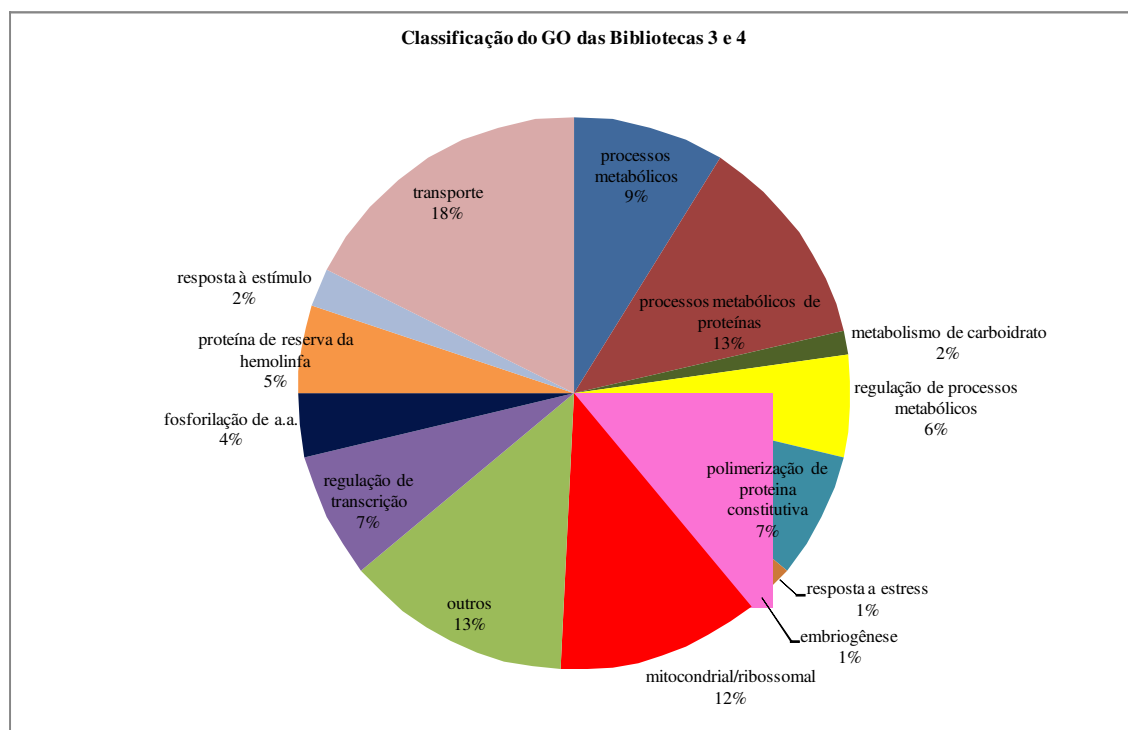


Figura 10- Gráfico da anotação das seqüências que obtiveram similaridade com o BLAST/GO das bibliotecas 3 e 4 (larva X linhagem ninfal).

Para a validação dos dados das bibliotecas subtrativas alguns genes foram selecionados para o experimento de Real Time PCR. Das bibliotecas de larva x linhagem neutra (bibliotecas 1 e 2), 13 genes foram escolhidos para a validação (Tabela 3). Os cinco genes selecionados das bibliotecas de larva x linhagem imaginal (bibliotecas 3 e 4) estão na Tabela 4.

Tabela 3- Genes escolhidos para a validação das bibliotecas subtrativas 1 e 2.

Gene	1º Hit Blast	Provável classificação	Biblioteca
EF1 ^α	elongation factor 1-alpha [Homalodisca coagulata]	Biossíntese protéica	Linhagem áptera/Larva
CALR	Calreticulin [Apis mellifera]	Calreticulina – proteína que se liga ao cálcio	Larva
Regenectina	Regenectina Periplaneta americana		Linhagem áptera
Ferritina	ferritin 2 [Anoplophora glabripennis]	Ferritina – estoque e transporte de ferro	Linhagem áptera
Lisozima	salivary lysozyme [Anopheles stephensi]	Lisozima – enzima envolvida na digestão	Linhagem áptera
Epoxide hidrolase (JHEH)	Juvenile hormone epoxide hydrolase 1 [Drosophila melanogaster]	Degradação de hormônio juvenil	Larva
SDH	Sorbitol dehydrogenase (L-iditol 2-dehydrogenase) [Tribolium castaneum]	A Sorbitol deidrogenase é uma enzima que catalisa a oxidação do sorbitol em frutose	Linhagem áptera
β-actina	Reticulitermes flavipes beta-actin	β-actina – proteína constitutiva utilizada como controle para reação de transcrição	Larva/Linhagem áptera
Quitinase	chitinase [Monochamus alternatus]	Catabolismo de quitina	Linhagem áptera
Endo-β-1,4-glicanase		Hidrólise de ligações β-1,4 de peptídeos	Linhagem áptera
α-amilase	α-amilase Blatella germanica		Linhagem áptera
α-tubulina	alpha-tubulin [Laodelphax striatellus]	Proteína constitutiva	Linhagem neutra/Larva
Pro-Hormônio Convertase	prohormone convertase 1 [Rana catesbeiana]	As pró-convertases são um dos sete membros da família das endoproteases, que ativam proteínas por clivagem.	Linhagem áptera

Tabela 4 - Genes escolhidos para a validação das bibliotecas subtrativas 3 e 4.

Gene	1º Hit Blast	Provável classificação	Biblioteca
Hexamerina I	hexamerin I - Reticulitermes flavipes	Proteína de reserva que se liga a HJ	Linhagem ninfal/Larva
Hexamerina II	hexamerin II Reticulitermes flavipes	Proteína de reserva que se liga a HJ	Larva
Vitelogenina	Vitellogenin precursor (Vg-1)- Periplaneta americana	Proteína de reserva	Linhagem ninfal
β-glicosidase	beta-glucosidase Neotermes koshunensis – Neofem 2 – <i>Cryptotermes secundus</i>	Enzima envolvida na degradação da celulose/feromônio de comunicação	Linhagem ninfal
Citocromo Oxidase III	Cytocrome oxidase subunit III – Reticulitermes speratus	Transporte de elétrons	Linhagem ninfal

A primeira etapa para comparar os níveis de expressão dos genes selecionados foi fazer a anotação de quantas vezes aquele gene foi encontrado nas bibliotecas. A Figura 11 mostra o gráfico com o número de vezes em que a sequência de cada gene foi encontrada nas bibliotecas 1 e 2. As sequências dos genes escolhidos das bibliotecas 3 e 4 estão na Figura 12.



Figura 11- Gráfico do número de vezes que cada gene foi encontrado nas bibliotecas 1 e 2.

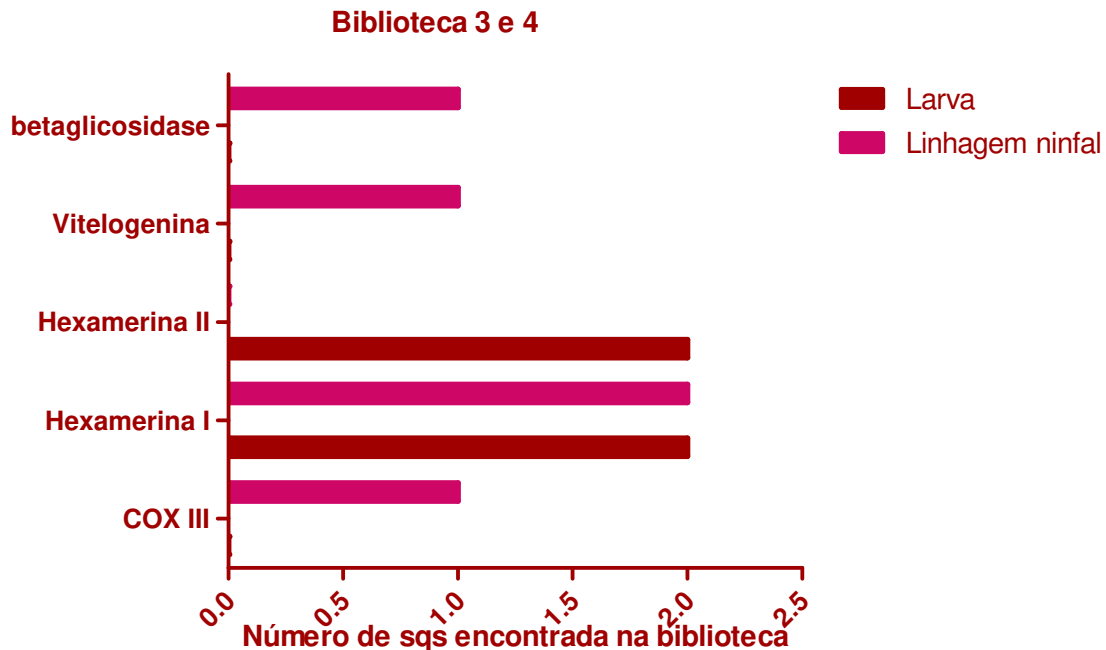


Figura 12 - Gráfico do número de vezes que cada gene foi encontrado nas bibliotecas 3 e 4.

3 - PCR em Tempo Real das bibliotecas subtrativas

A expressão dos 18 genes escolhidos foi avaliada por PCR em tempo real para a validação da biblioteca e verificação da expressão entre as diferentes castas. A expressão relativa dos genes selecionados para as bibliotecas 1 e 2 estão na Figura 13. Os genes selecionados para o estudo de expressão diferencial das bibliotecas 3 e 4 podem ser vistos na Figura 14.

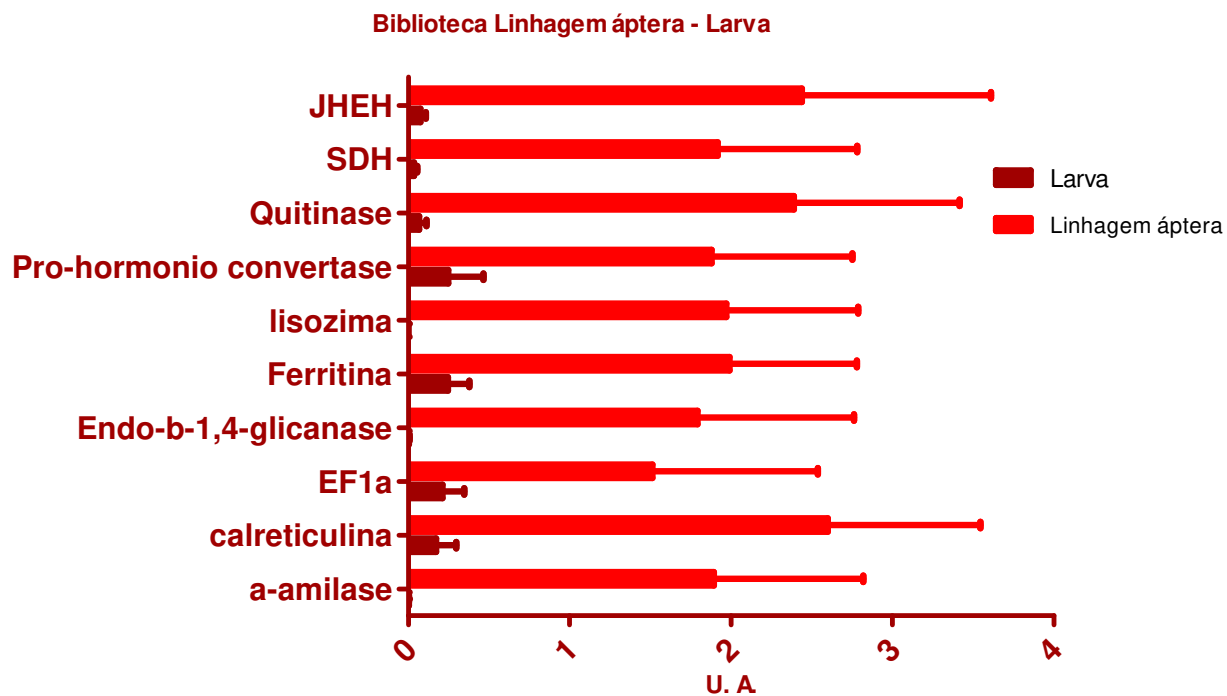


Figura 13 - Gráfico da expressão diferencial dos genes selecionados nas bibliotecas 1 e 2.

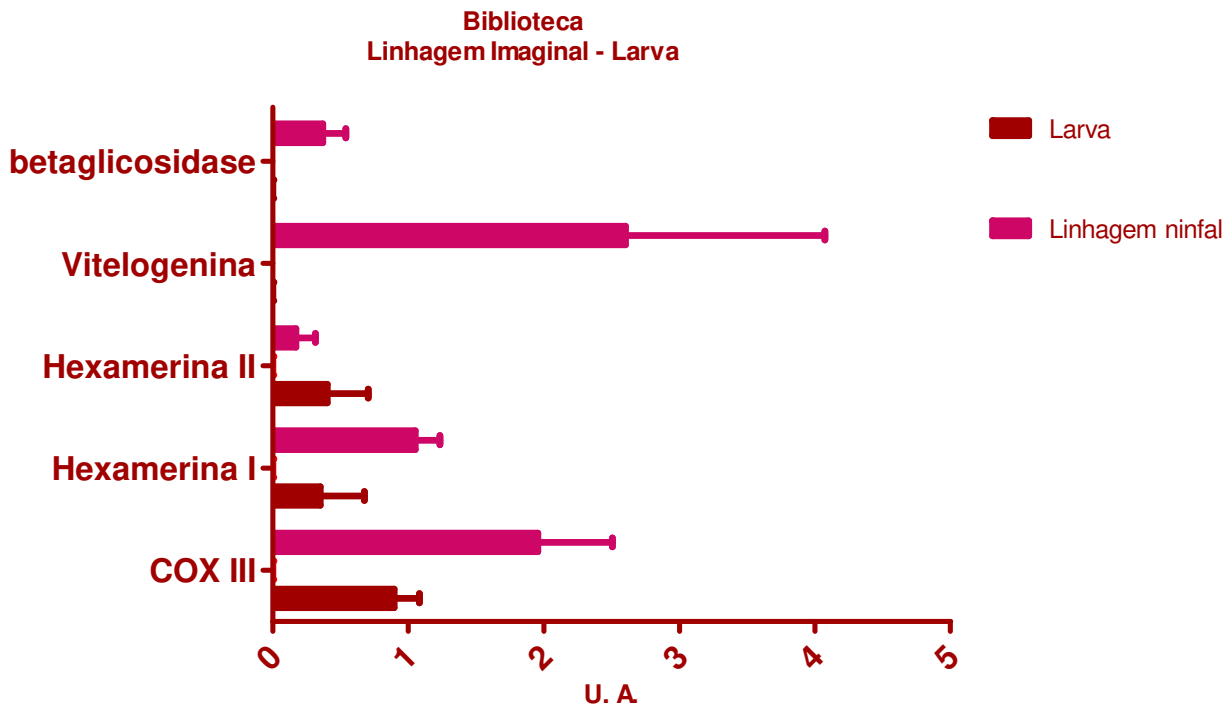


Figura 14- Gráfico da expressão diferencial dos genes selecionados nas bibliotecas 3 e 4.

Dos dezoito genes selecionados para a validação por Real Time PCR apenas três não foram validados, ou seja, 83,34% dos genes foram validados, o que está de acordo com os dados da literatura e da experiência em nosso laboratório (de Andrade *et al.*, 2010).

Dentre os genes selecionados para a validação das bibliotecas, alguns tiveram seu papel em relação à diferenciação de castas investigado. Para isso foram realizados estudos de expressão relativa nas diferentes castas e estágios de desenvolvimento da colônia do cupim *C. gestroi*.

O gene que codifica a enzima endo- β -1,4 glicanase foi encontrado diferencialmente expresso na biblioteca de linhagem neutra subtraída de larvas. Sua expressão se mostrou significativamente elevada em operários em relação aos outros estágios de desenvolvimento (Figura 15). Nos reprodutores alados apesar de não ser significativa a diferença de expressão, ela também se mostra mais elevada em relação às outras castas.

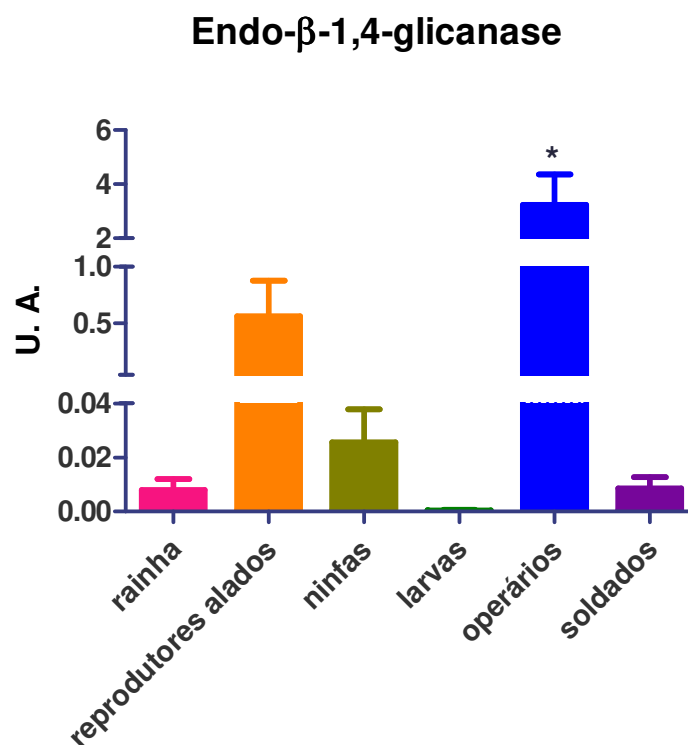


Figura 15 - Níveis de expressão relativa do gene que codifica endo- β -1,4 glicosidase nas diferentes castas e estágio de desenvolvimento de *C. gestroi*. O eixo Y indica a expressão relativa em unidades arbitrárias e os dados são médias de três experimentos independentes. Diferenças significativas estão indicadas pelos símbolos; Teste de Tuckey: * $p < 0,05$).

O gene que codifica a β -glicosidase foi encontrado na biblioteca da linhagem neutra e se mostrou diferencialmente expresso em operários em relação aos outros estágios de desenvolvimento (Figura 16).

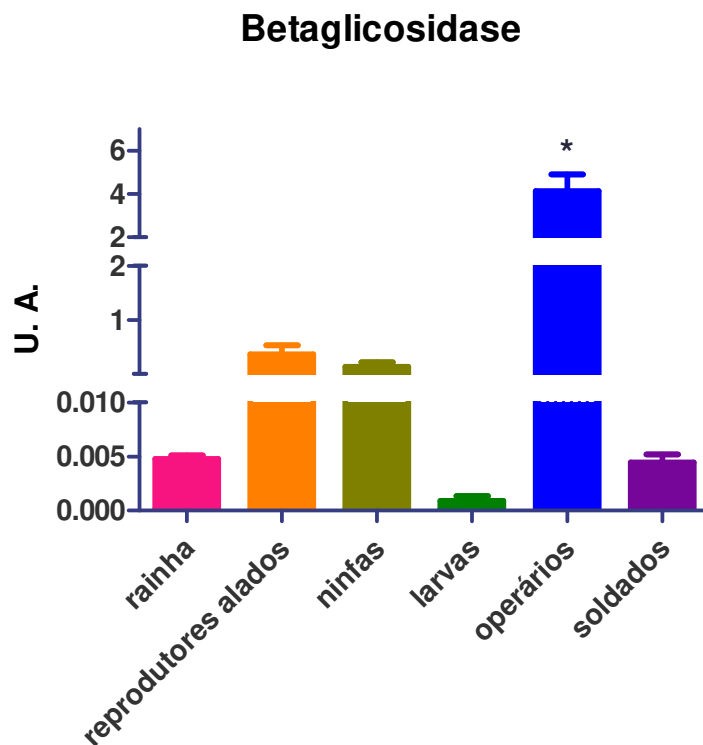


Figura 16 - Níveis de expressão relativa do gene que codifica β -glicosidase nas diferentes castas e estágio de desenvolvimento de *C. gestroi*. O eixo Y indica a expressão relativa em unidades arbitrárias e os dados são médias de três experimentos independentes. Diferenças significativas estão indicadas pelos símbolos; Teste de Tuckey: * $p < 0,05$).

O gene que codifica vitelogenina, encontrado diferencialmente expresso na biblioteca de linhagem imaginal se mostrou mais expresso em reprodutores alados em relação às outras castas e estágios de desenvolvimento (Figura 17).

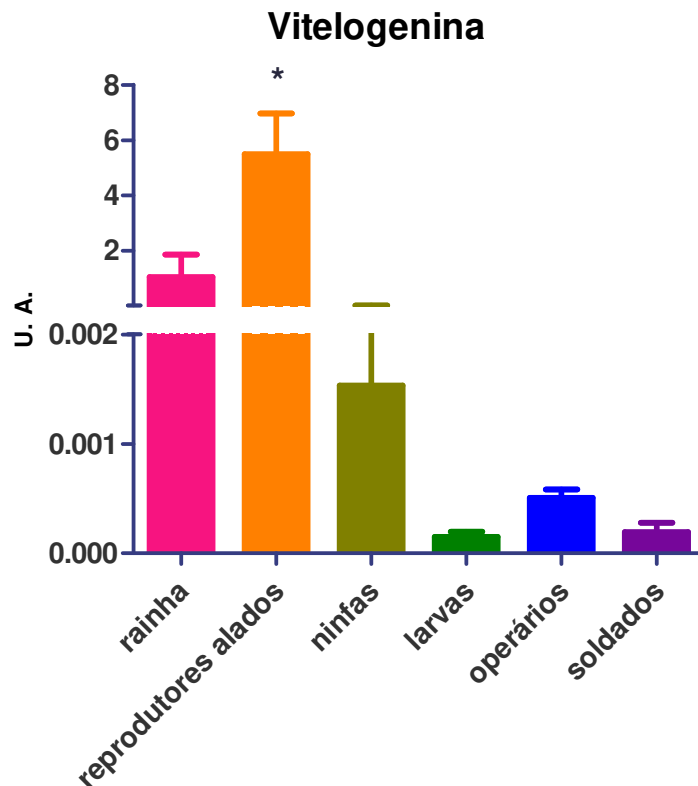


Figura 17 - Níveis de expressão relativa do gene que codifica Vitelogenina nas diferentes castas e estágio de desenvolvimento de *C. gestroi*. O eixo Y indica a expressão relativa em unidades arbitrárias e os dados são médias de três experimentos independentes. Diferenças significativas estão indicadas pelos símbolos; Teste de Tuckey: * $p < 0,05$).

O gene que codifica a enzima epóxido hidrolase, responsável pela degradação do hormônio juvenil foi encontrado diferencialmente expresso na biblioteca de linhagem neutra subtraída de larvas e apresentou níveis elevados de expressão em operários em relação aos demais estágios de desenvolvimento e castas (Figura 18).

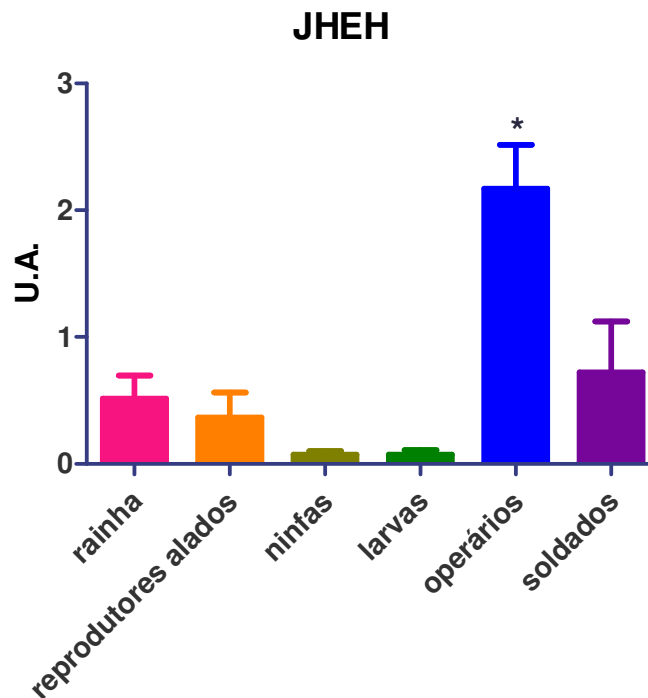


Figura 18 - Níveis de expressão relativa do gene que codifica JHEH nas diferentes castas e estágio de desenvolvimento de *C. gestroi*. O eixo Y indica a expressão relativa em unidades arbitrárias e os dados são médias de três experimentos independentes. Diferenças significativas estão indicadas pelos símbolos; Teste de Tuckey: * $p < 0,05$).

O gene que codifica a Hexamerina I, encontrada diferencialmente expressa na biblioteca de linhagem neutra subtraída de larvas, apresentou altos níveis de expressão na casta dos operários em relação às demais castas (Figura 19).

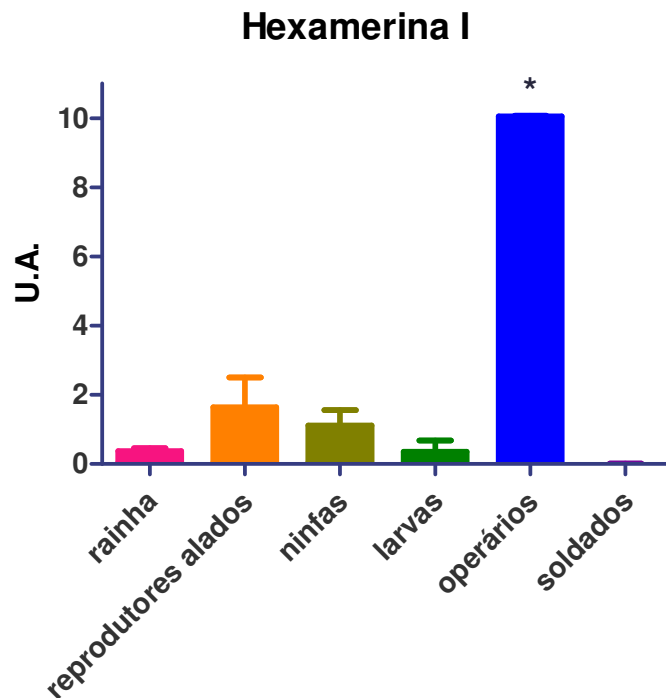


Figura 19 - Níveis de expressão relativa do gene que codifica a Hexamerina I nas diferentes castas e estágio de desenvolvimento de *C. gestroi*. O eixo Y indica a expressão relativa em unidades arbitrárias e os dados são médias de três experimentos independentes. Diferenças significativas estão indicadas pelos símbolos; Teste de Tuckey: * $p < 0,05$).

O gene que codifica a Hexamerina II, encontrado diferencialmente expresso na biblioteca de larvas subtraída de linhagem imaginal, apresentou níveis elevados de expressão em operários em relação aos demais estágios de desenvolvimento e castas (Figura 20).

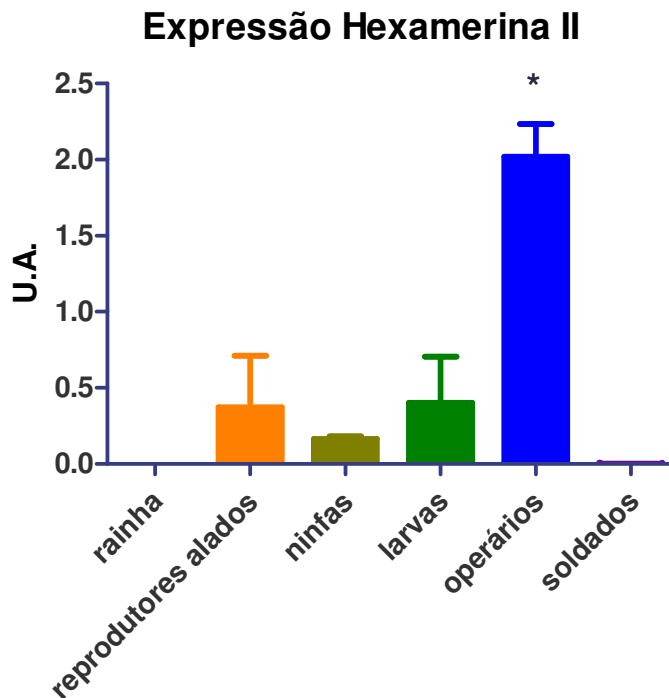


Figura 20- Níveis de expressão relativa do gene que codifica Hexamerina II nas diferentes castas e estágio de desenvolvimento de *C. gestroi*. O eixo Y indica a expressão relativa em unidades arbitrárias e os dados são médias de três experimentos independentes. Diferenças significativas estão indicadas pelos símbolos; Teste de Tuckey: * $p < 0,05$).

4 – Sequenciamento de segunda geração

O pirosequenciamento foi realizado com amostras de RNA mensageiro de operários e soldados. Depois da trimagem das sequências, ou seja, retirada da cauda poli a, sequências de adaptadores utilizados e sequências de baixa qualidade, foi feita a montagem dos “contigs”. Das 1.526,647 sequências de boa qualidade obtidas com o sequenciamento, foram formados 110.899 “contigs”. A análise dos genes diferencialmente expressos feita pelo programa DEGseq gerou 3.921 sequências com diferença estatística ($p < 0,001$) entre operários e soldados. Destes destacam-se os dez mais representativos que estão na tabela a seguir (Tabela 5).

Tabela 5 - Genes diferencialmente expressos entre operários e soldados de *C. gestroi* através do pirosequenciamento.

First Hit	Nº vezes em que a sequência é encontrada em operários	Nº vezes em que a sequência é encontrada em soldados	Valor de p
dbj BAG30740.1 muscle myosin heavy chain [Papilio xuthus]	829	2483	6,30E-223
muscle-specific protein 300 [Tribolium castaneum]	500	1188	3,40E-71
ref XP_002432355.1 paramyosin, long form, putative [Pediculus]	401	1053	3,89E-77
dbj BAA87882.1 soldier-caste specific protein 1 [Hodotermopsis]	1	162	2,89E-58
dbj BAD84176.1 cytochrome P450 [Hodotermopsis sjoestedti]	41	15	1,74E-10
family 4 cytochrome P450 [Coptotermes acinaciformi]	23	8	4,96E-07
Hydroxymethylglutaryl-CoA synthase (HMGs)	2	34	2,31E-10
AACT - Acetoacetyl coenzima A tiolase	33	16	5,05E-06
gb AAU20852.2 hexamerin II [Reticulitermes flavipes]	331	12	0
gb AAU20851.2 hexamerin I [Reticulitermes flavipes]	638	16	0

Após a busca de similaridade constatou-se que do total das “contigs”, 80.640 eram “no hits” (72,71%). As 30.259 sequências que obtiveram similaridade foram divididas em grupos de acordo com o resultado do BLAST X: 977 sequências apresentaram similaridade com bactérias, 24.348 sequências apresentaram *first hit* com sequências de insetos e 4.944 sequências tiveram outros organismos como *first hit* (nestes incluem protozoários vertebrados e fungos) (Figura 21). Os resultados de expressão diferencial do pirosequenciamento foram confirmados por Real Time PCR em 6 genes (Figura 22).

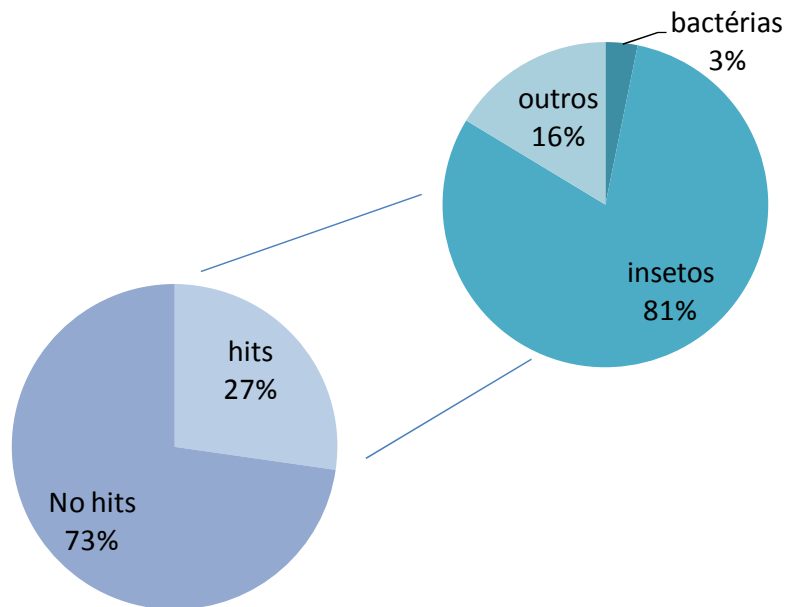


Figura 21 - Gráfico que mostra a classificação das seqüências por resultado do First hit do BlastX

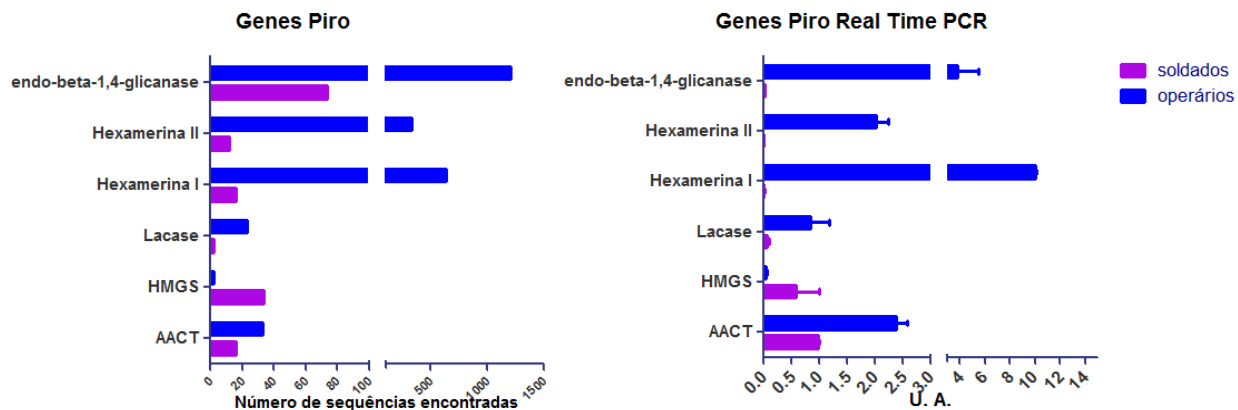


Figura 22 - Gráficos mostrando o resultado da expressão diferencial do pirosequenciamento em comparação aos resultados do estudo de expressão de Real Time PCR de seis genes.

4.1 - Genes com aplicação biotecnológica

Um *screening* em todas as sequências do pirosequenciamento foi realizado para selecionar enzimas relacionadas com a degradação da lignocelulose (Figura 23). As enzimas identificadas foram listadas na tabela abaixo (Tabela 6). As sequências completas destes genes estão em anexo (ANEXO II). Além disso, outras 16 sequências não completas de enzimas também foram identificadas (Tabela 7).

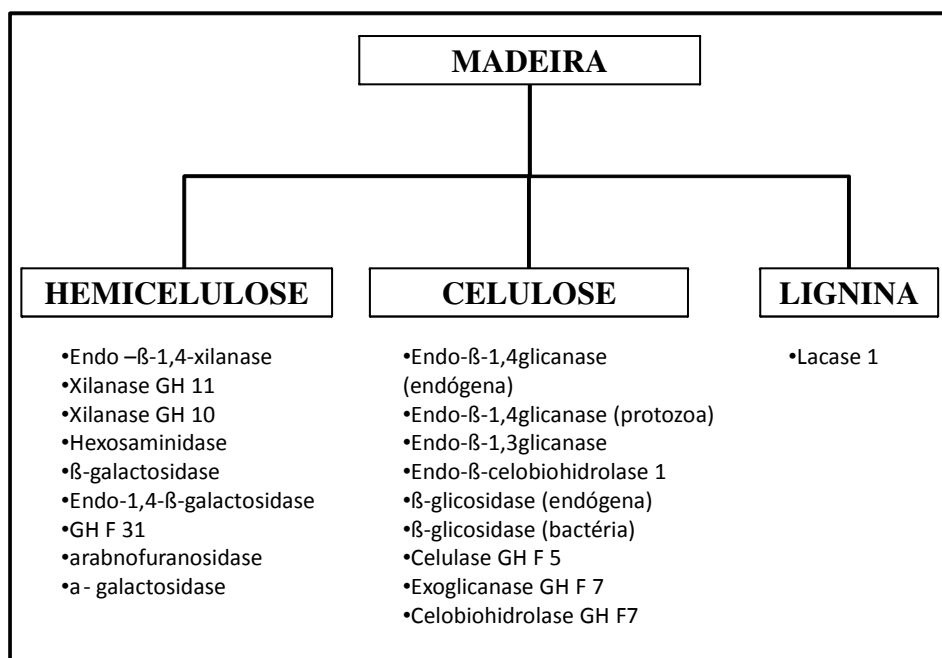


Figura 23 - Esquema com todas as enzimas relacionadas com a degradação da madeira encontradas em *C. gestroi*.

Tabela 6 - Enzimas identificadas no pirosequenciamento.

Enzima	First hit (Blast X)	E-value
Endo- β -1,4-glicanase (endógena)	Endo-beta-1,4-glucanase (<i>Coptotermes formosanus</i>)	0
β -glicosidase (endógena)	ADB23476.1 beta-glucosidase (<i>Coptotermes formosanus</i>)	0
Lacase 1 (endógena)	ACX54558.1 – Laccase 6 (<i>Reticulitermes flavipes</i>)	0
Endo- β -1,3glicanase (endógena)	ABR28480.1 beta-1,3 (4)-glucanase LIC1 (<i>Periplaneta americana</i>)	2e-74
Endo- β -celobiohidrolase 1 (simbionte)	CAK18913.1 – cellulose 1,4 – beta-cellobiosidase precursor (<i>Pleuroptus SP.</i> Florida)	6e-173
Endo- β -1,4xilanase 1 (simbionte)	YP_003873130.1 Beta-1,4-xylanase (<i>Paenibacillus polymyxa</i> E681)	4e-103

Tabela 7- Sequências não completas de enzimas envolvidas na quebra da lignocelulose.

Enzima	Família de Glicosil hidrolase	Provavel origem
Celobiohidrolase	GH F 7	Protozoário simbionte
Exoglicanase	GH F 7	Protozoário simbionte
Celulase	GH F 5	Protozoário simbionte
Xilanase	GH F 11	Bactéria simbionte
α - galactosidase	GH F 59	Bactéria simbionte
Xilanase	GH F 10	Bactéria simbionte
Endo- β -1,4glicanase	GH F 7	Protozoário simbionte
Glicosil hidrolase <i>Trichomonas vaginalis</i>	GH F 31	Protozoário simbionte
β -galactosidase	GH F 42	Bactéria simbionte
Hexosaminidase	GH F 20	Endógena
Endo-1,4- β -galactosidase	GH F 53	Bactéria simbionte
Arabnofuranosidase	GH F 62	Bactéria simbionte
β -glicosidase	GH F 3	Bactéria simbionte

DISCUSSÃO

1 - Bibliotecas subtrativas

O número de sequências gênicas de cupins publicadas é extremamente reduzido. Por este motivo utilizamos a técnica de biblioteca subtrativa suprimida (SSH) para entender as diferenças de expressão gênicas nas diferentes castas. . Este é um método alternativo para se identificar novos genes em organismos não modelo, tal como *C. gestroi*. Como essa técnica permite que apenas duas amostras sejam comparadas, as linhagens imaginal e neutra foram comparadas às larvas. Para isso quatro bibliotecas foram construídas: (1) larva subtraída da linhagem áptera, (2) linhagem áptera subtraída da larva, (3) larva subtraída da linhagem ninfal e (4) linhagem ninfal subtraída da larva.

O sequenciamento das 40 placas gerou 1.851 *reads* das bibliotecas 1 e 2 e 1.846 *reads* das bibliotecas 3 e 4. Após a realização da busca de similaridade com o banco de dados do NCBI constatou-se que o número de sequências que não possuíam similaridade com nenhuma sequência depositada, ou seja, “no hit”, foi muito elevada em todas as bibliotecas. Isto sugere que esses genes devem ser espécie específicos, já que não existem genomas de cupins para comparação no banco de dados.

Dos genes que apresentaram similaridade com sequências do banco de dados, foi realizada a classificação do Gene Ontology em processo biológico. Em todas as bibliotecas observou-se que os processos metabólicos são os melhor representados. Isto sugere que esses genes possuem funções biológicas básicas. Esses dados são semelhantes aos obtidos com a biblioteca construída com ESTs de operárias de *Apis mellifera* (Nunes *et al.*, 2004).

Os 18 genes escolhidos foram investigados em todos os estágios de desenvolvimento através da técnica de Real Time PCR. Dentre estes, alguns se destacaram por apresentar uma variação bem distinta entre as castas e estarem claramente relacionados com funções específicas na diferenciação de castas.

O gene que codifica a enzima endo- β -1,4-glicanase foi encontrado na biblioteca de linhagem áptera subtraída de larvas e também na biblioteca de linhagem ninfal subtraída de larvas. Esta enzima hidrolisa a celulose em celobiose e é encontrada nas glândulas salivares de cupins, tais como *C. formosanus* (Watanabe e Tokuda, 2001). Analisando o experimento de PCR em tempo real, o que se mostrou interessante foi a expressão elevada da endo- β -1,4-glicanase no RNAm de operários, confirmando a hipótese de que esta casta expressa

esta enzima já que necessitam dela para realizar a quebra da celulose para a alimentação dos demais indivíduos da colônia.

Acredita-se que esta enzima também esteja presente nos alados utilizados na confecção da biblioteca da linhagem ninfal, já que eles deixam o ninho de origem e saem para formar o seu próprio, sendo obrigados a cuidar e alimentar sua primeira cria. Para isso, eles carregam alguns endossimbiontes e também as enzimas que auxiliam na degradação da madeira (Scharf *et al.*, 2005). Dados indicam que as ninfas também apresentem celulasas. Scharf e colaboradores (2003) mostraram que os níveis de transcritos para celulasas eram mais elevados em operários e ninfas em relação às outras castas. Apesar da diferença de expressão não ser estatisticamente significativa entre as ninfas e alados, podemos observar que a endo- β -1,4-glicanase é aparentemente mais expressa nestas castas em relação às larvas (Figura 15).

Estudos de expressão diferencial realizados com a espécie de cupim *Cryptotermes secundus* (Weil *et al.*, 2007) identificaram alguns genes que se mostraram importantes na diferenciação de castas. O gene Neofem 2 similar ao gene da β -glicosidase de *Neotermes koshunensis* foi encontrado e as análises do PCR em tempo real revelaram que ele é expresso exclusivamente em fêmeas da linhagem reprodutiva, sugerindo uma função específica ligada ao sexo. RNA de interferência foi realizado com o gene Neofem 2 para confirmar sua importância na supressão da casta reprodutiva em *Cryptotermes secundus*. (Korb *et al.*, 2009). Foi constatado que o Neofem 2 está relacionado com a produção de um hormônio modificador em cupins, o qual faz com que os *pseudergates* (falso operários) saibam que a rainha está presente no ninho. Com o bloqueio desse gene na rainha, os demais indivíduos da colônia se comportaram como se a rainha não estivesse presente, aumentando o comportamento agressivo entre os *pseudergates* para a disputa do “trono” da colônia (Korb *et al.*, 2009). Em *C. gestroi*, no entanto, não foi observado esse aumento de expressão na rainha (Figura 16). Esta enzima parece ter importante papel na degradação de celulose, já que a casta dos operários apresenta maior expressão deste gene em relação aos demais. Em *H. sjostedti* as glândulas salivares dos operários são responsáveis por secretar tanto a endo- β -1,4-glicanase, quanto a β -glicosidase, portanto, ocorre uma redução na expressão de ambos os genes quando se dá a diferenciação de operários em soldados (Fujita *et al.*, 2008). Em *C. gestroi* a mesma redução foi observada. Desta forma, tomando todos os

dados em conjunto podemos hipotetizar que a β -glicosidase esteja atuando na degradação da celulose em *C. gestroi*.

O hormônio juvenil tem um papel importante na regulação do desenvolvimento de insetos. Os processos afetados pelo hormônio juvenil incluem desenvolvimento embrionário, diapausa, desenvolvimento de pupa para a fase adulta, produção de feromônios e processos reprodutivos em insetos adultos (Jones, 1995). O hormônio juvenil pode ter efeito estimulatório ou inibitório, dependendo do hormônio com o qual está atuando em conjunto. A fim de esclarecer sua função, estudos foram desenvolvidos para a melhor compreensão de sua biossíntese, transporte e degradação. A enzima epóxido hidrolase do hormônio juvenil (HJEH) contribui para a degradação desse hormônio, fazendo sua hidrólise (Hartfelder, 2000). Estudos demonstraram que a forma de atuação do hormônio juvenil e sua degradação variam de espécie para espécie, de acordo com o estágio de desenvolvimento no qual o inseto se encontra. Na pulga *Ctenocephalides felis*, a HJEH tem baixa atividade em estágios larvais porque baixos níveis dessa enzima favorecem a alta produção de hormônio juvenil. Em adultos, no entanto, sua atividade é alta principalmente em tecidos ovarianos, indicando que os processos reprodutivos estão diretamente ligados à enzima (Keiser, 2002). Em *C. gestroi* a HJEH foi encontrada na biblioteca de linhagem áptera subtraída de larvas. A determinação de castas nos cupins é fortemente influenciada pelo sistema endócrino (Lüscher 1958), e sabe-se que nestes insetos o hormônio juvenil está intimamente ligado com a mudança de castas de operário para soldados. Estes dados confirmados pela expressão gênica indicam que esta enzima tem sua importância na diferenciação de castas em *C. gestroi*.

Em insetos, a maior parte das proteínas de hemolinfa, tais como vitelogenina, proteínas de reserva (hexamerinas, proteínas ricas em metionina, califorinas e lipoforinas), proteínas responsivas a hormônio juvenil e algumas proteínas associadas à diapausa são primeiramente sintetizadas no corpo gorduroso e secretadas na hemolinfa. Em geral, as principais funções dessas proteínas são reserva nutricional, ligação e transporte. Vitelogeninas de insetos, frequentemente chamadas de proteína de gema, são um grupo grande de fosfoglicoproteínas homólogas. A caracterização molecular do cDNA da vitelogenina de diferentes insetos indica que a estrutura primária desses genes é extremamente conservada (Lee, 2000). Além de prover nutrição para o embrião em

desenvolvimento, as vitelogeninas também funcionam como proteínas carreadoras de lipídeos, fosfatos, vitaminas, metais, carboidratos e hormônios.

Um gene que codifica uma vitelogenina foi identificado diferencialmente expresso na biblioteca de linhagem ninfal subtraída de larvas (Figura 17). Em um estudo de diferenciação de expressão de castas realizado com o cupim *Reticulitermes flavipes* (Scharf *et al.*, 2005) foi descrita uma expressão elevada de vitelogenina em ninfas, alados e reprodutores suplementares. Como esperado, os resultados do estudo de expressão diferencial confirmam os dados obtidos na biblioteca e na literatura.

As hexamerinas dos insetos são importantes proteínas de reserva e também de ligação à hormônios. Zhou e colaboradores (2006) demonstraram, através da técnica de RNA de interferência, uma nova função para as hexamerinas. Estas proteínas de cupim se mostraram importantes na inibição de mudança de castas de operário para soldado. Este mecanismo de regulação de castas resulta em indivíduos temporariamente imaturos (operários). Nesse estudo, com o silenciamento da hexamerina II o número de soldados aumentou, sugerindo um importante papel desta proteína na supressão da diferenciação da casta de soldado. Nossos resultados mostraram que este gene foi encontrado diferencialmente expresso na biblioteca de larvas subtraída da linhagem ninfal, sugerindo que ele pode estar influenciando a mudança de castas das larvas. Nos resultados de expressão gênica o gene da Hexamerina II está mais expresso nas larvas em relação à linhagem ninfal (Figura 20), embora não apresente diferença estatisticamente significativa. Estes dados podem ser explicados porque na linhagem ninfal todas as castas já estão na fase adulta, e aparentemente não existem motivos para a produção desta proteína. Já a hexamerina I que foi encontrada em ambas as bibliotecas, pode representar importante papel de reserva; este fato justifica sua presença em todas as castas. Os dados de expressão gênica mostraram que a hexamerina I é bastante expressa em operários e também em algumas das castas que compõem a linhagem ninfal, como as ninfas e os reprodutores alados (Figura 19).

2- Sequenciamento de segunda geração

2.1 – Genes relacionados à regulação de castas

Os resultados deste trabalho representam o maior número de sequências gênicas de *C. gestroi* já publicados. Após os procedimentos metodológicos para a exclusão das sequências de baixa qualidade e das regiões não relevantes, restaram 1.526,647 sequências (“reads”). Estas foram montadas em 110.899 “contigs”. Este número de sequências é aparentemente elevado para um transcriptoma de uma espécie de cupim, mas não se pode esquecer que *C. gestroi* é um cupim inferior, portanto, possui uma microbiota de simbiontes bem diversificada. As sequências puderam ser divididas em três grupos: similares às sequências de insetos; similares a sequências de bactérias; e outro grupo que incluíram protozoários simbiontes, fungos e também sequências conservadas de vertebrados.

Para saber a diferença de expressão entre as castas de operários e soldados o programa DEGseq foi utilizado. Considerando uma significância estatística elevada ($p < 0,001$), obtivemos 3.921 sequências diferencialmente expressas entre operários e soldados.

A regulação da mudança de castas de operário para soldado em cupins é modulada por vários fatores ligados ao HJ. Baseado nisso foram selecionados genes relacionados a proteínas de reserva, via de produção de HJ e genes exclusivos de uma das castas que apresentam características específicas.

A título de exemplo, na Tabela 5 podem-se observar dez genes relacionados com a mudança de castas. Os quatro primeiros genes tratam-se de sequências mais expressas em soldados, ou seja, genes relacionados a características específicas de soldados. Levando-se em consideração que a dramática mudança morfológica de operário para soldado requer um significativo rearranjo do corpo (Miura, 2004), é justificável que algumas dessas sequências estejam relacionadas a genes presentes na musculatura, provavelmente, devido ao fato desta ser altamente desenvolvida na mandíbula dos soldados.

Outras duas sequências são similares às encontradas no citocromo P450, o qual apresentam um importante papel na biossíntese e metabolismo de hormônios morfogênicos e terpenóides (Anderson *et al.*, 1997). Na barata *Diploptera punctata* a enzima CY15A1, uma P450, degrada o metil farnesoato para produzir HJ III (Helvig *et al.*, 2004). Nos cupins

vários genes P450 da família CYP4 sofreram alterações durante a expressão de HJ induzindo a diferenciação de operários em pré-soldados, sugerindo que esses genes tem papel na mediação ou atenuação da diferenciação de soldados (Zhou *et al.*, 2006b). A expressão mais elevada desses compostos em operários pode indicar que esses genes estejam envolvidos na diferenciação de castas em *C. gestroi*, embora mais estudos sejam necessários para essa confirmação.

Os dois últimos genes encontrados diferencialmente expressos, O AACT e o HMG-S, estão presentes na via do mevalonato. O aumento da expressão de HJ III faz com que os genes envolvidos nessa via aumentem. Analisando um estudo realizado com o Coleoptero *Ips pini*, constatou-se que os machos, ao se alimentarem do floema, produzem HJ III na glândula *corpora allata*, sintetizando ipsdienol pela via do mevalonato (Keeling *et al.*, 2004). Após a produção do feromônio ele é liberado para que outros indivíduos também venham se alimentar, funcionando como um feromônio de agregação. O gene da AACT é mais expresso em operários em relação aos soldados e o gene da HMG-S é mais expresso em soldados em relação aos operários. A PCR em tempo Real confirmou essa expressão diferencial (Figura 22). Este resultado sugere que a variação de expressão dos dois genes esteja influenciando a mudança de castas de operário para soldado, no entanto, ainda são necessários mais experimentos para que essa expressão diferencial seja mais bem compreendida.

Os genes que codificam hexamerinas (Hex I e Hex II), confirmando os dados das bibliotecas subtrativas, foram encontrados diferencialmente expressos no pirosequenciamento. As hexamerinas dos cupins têm cada vez mais se mostrado atuantes na resposta do mecanismo que afeta a atividade do HJ, provavelmente limitando sua disponibilidade (Scharf *et al.*, 2007). Antes de se saber a identidade dessas proteínas solúveis na hemolinfa, elas foram identificadas e completamente sequenciadas em *R. flavipes* (Scharf *et al.*, 2005). Com a técnica de RNA de interferência foi descoberto que o silenciamento de ambas as hexamerinas leva a um aumento significativo da diferenciação de soldados dependente de HJ. Isto forneceu a primeira evidência de que as hexamerinas inibem a formação de pré-soldados, ao invés de induzir essa diferenciação. O significado evolutivo dessa regulação seria o de manter o equilíbrio da colônia com uma proporção ideal entre operários e soldados. A elevada expressão em operários de ambas as

hexamerinas faz com que o HJ III se ligue a essas proteínas fazendo com que esses indivíduos permaneçam como operários, ao passo que os soldados apresentam níveis de expressão muito baixos de ambas as proteínas, como mostrado pelos nossos dados em *C. gestroi* (Figura 22). Sabe-se também que os fatores nutricionais e ambientais interferem nesse mecanismo de diferenciação (Scharf *et al.*, 2007). Experimentos futuros com variações nutricionais e de temperatura devem ser realizados para comprovar se em *C. gestroi* esses fatores realmente interferem na expressão dessas proteínas.

2.2 - Genes com aplicação biotecnológica

A análise preliminar dos resultados do pirosequenciamento foi realizada com o objetivo de identificar enzimas envolvidas na degradação da lignocelulose. Os resultados de busca de similaridade com o banco de dados do NCBI revelaram sequências de seis diferentes genes com aplicação biotecnológica.

Estudos de digestomas de cupins vêm sendo realizados para a identificação de enzimas envolvidas na quebra da lignocelulose e também para compreensão de sua forma de atuação. O sequenciamento da metagenômica dos simbiontes do intestino posterior de *R. speratus* gerou celulases pertencentes a 10 famílias de glicosil hidrolases (GH) (Todaka *et al.*, 2007). Nenhuma lignase foi encontrada e dos nove genes identificados destacaram-se três celobiohidrolases/exoglicanases e 6 endoglicanases. A família GH 7 foi a mais encontrada neste estudo. O sequenciamento de bibliotecas de cDNA preparadas com várias castas de *R. flavipes* revelou celulases endógenas do cupim e de seus simbiontes (Scharf e Tartar, 2008). Mais de 11.000 ESTs foram sequenciadas do transcriptoma do trato digestivo deste inseto o que revelou glicosil hidrolases de 28 famílias diferentes. Além de endoglicanases, exoglicanases e xilanases, várias candidatas a lignases foram identificadas (lacases, peroxidases, epóxido hidrolases, etc). A metagenômica de uma espécie de *Nasutitermes* foi realizada por Warnecke e colaboradores (Warnecke *et al.*, 2007). Como os próprios autores constataram essa metagenômica “mostra como um ambiente de 1µL pode ser tão complexo”. Como o *Nasutitermes* é um cupim superior ele não apresenta protozoários em seu trato digestivo, no entanto secretam suas próprias celulases e abrigam bactéria celulolíticas. Esse estudo combinou técnicas recentes de pirosequenciamento e sequenciamento tradicional (Sanger) para sequenciar o DNA presente nos organismos da

porção posterior do intestino. Foram descritos 12 filos de bactérias presentes, mais de 200 genes relacionados com a degradação da lignocelulose e nenhuma lignase foi encontrada, colaborando com a hipótese de que a lignina não é degradada pelos endossimbiontes.

Com a identificação desse conjunto de enzimas podemos propor os seguintes passos de degradação da lignocelulose em *C. gestroi*; a degradação da lignocelulose tem início com a quebra da madeira em partículas pequenas devido ao aparato bucal desenvolvido para esse tipo de alimentação. Uma lacase, enzima relacionada com oxidação de compostos fenólicos, entraria em ação para quebrar a lignina. A próxima etapa seria a degradação da hemicelulose, estágio crítico para tornar a despolimerização da celulose acessível (Saha, 2003). A biodegradação completa da hemicelulose requer atividade combinada de muitas enzimas, dentre elas, endo-exo- β -1,4-xilanases, β -xilosidases, α -arabinofuranosidases, α -uronidases, e esterases. Nesta etapa as enzimas dos simbiontes do trato digestivo seriam as mais atuantes. Em *C. gestroi* a endo- β -1,4 xilanase, xilanases das famílias de glicosil hidrolases 10 e 11, uma hexosaminidase, uma β -galactosidase, uma endo-1,4- β -galactosidase, uma arabinofuranosidase, α -galactosidase e uma glicosil hidrolase da família 31 estariam atuando nesta etapa. A última etapa do processo de degradação da lignocelulose seria a degradação da celulose. A endo - β -1,4- glicanase produzida na glândula salivar de *C. gestroi* estaria atuando juntamente outra endo - β -1,4- glicanase produzida por simbiontes e também uma endo - β -1,3- glicanase na quebra da celulose em glicose e celobiose. As celobiohidrolase, tais como a endo- β -celobiohidrolase e uma celobiohidrolase (GH F 7) estariam atuando na celobiose transformando-a em celotriose que por sua vez estariam sendo hidrolizadas por β -glicosidases (tanto a endógena do cupim quanto a de origem bacteriana) para a liberação de monômeros de glicose.

Neste trabalho foi gerado o maior número de sequências já publicadas de *C. gestroi*. As análises levadas a efeito até agora possibilitaram avaliar mecanismos de diferenciação de castas que ainda não haviam sido descritos nesta espécie. Além disso, foi possível identificar dezenove enzimas que participam do processo de quebra da lignocelulose.

Tomados em conjunto, nossos resultados podem contribuir de maneira significativa para uma melhor compreensão dos ainda misteriosos mecanismos da vida social dos cupins.

CONCLUSÕES

- Quanto aos genes diferencialmente expressos nas diferentes castas:
 - os genes que codificam endo- β -1,4-glicanase, β -glicosidase, a epóxido hidrolase do hormônio juvenil foram identificados nas bibliotecas de linhagem áptera e confirmaram sua maior expressão na casta dos operários;
 - o gene da vitelogenina foi encontrado na biblioteca de linhagem ninfal e sua maior expressão foi confirmada em rainhas e alados;
 - os genes da Hex-I e Hex-II também se mostraram mais expressos em operários confirmando os resultados obtidos nas bibliotecas subtrativas. Deste modo nossos dados contribuem para indicar uma função muito importante na manutenção da casta de operário, já que sua expressão está extremamente baixa nos soldados;
 - o gene do HMS-G, enzima que faz parte da via do mevalonato, se mostrou mais expresso em soldados em relação aos operários, sugerindo que essa diferença de expressão pode estar relacionada com a diferenciação de castas de operários para soldados em *C. gestroi*, já que esta é uma importante via de produção de HJ III;
 - 3 genes relacionados à musculatura se mostraram mais expressos em soldados em relação aos operários, sugerindo que estes devem estar relacionados com as mudanças morfológicas que os operários sofrem quando se dá a diferenciação em soldados;
 - alguns genes de P450 foram identificados mais expressos em operários, sugerindo papel na mediação ou atenuação da diferenciação de soldados.
- Quanto aos genes de potencial biotecnológico:
 - Dezenove enzimas que atuam na degradação de lignocelulose (Tabela 6) foram sequenciadas, sugerindo uma ação em conjunto das enzimas endógenas de *C. gestroi* e de seus simbiontes para a degradação de lignocelulose.

REFERÊNCIAS BIBLIOGRÁFICAS

Abe, T. (1987) Evolution of life types in termites. Evolution and coadaptation in biotic communities. T. Hidaka. Tokyo, University of Tokio Press: 125 - 148.

Andersen, J.F.; Walding, J.K., et al. (1997) "Substrate specificity for the epoxidation of terpenoids and active site topology of house fly cytochrome P450 6A1." Chemical Research in Toxicology 10: 156-164.

Anderson, W. F. e Akin, D. E. (2008) "Structural and chemical properties of grass lignocelluloses related to conversion for biofuels." Journal of Industrial Microbiology and Biotechnology 35: 355-366.

Araujo, R. (1958) "Contribuições à biogeografia dos térmitas de São Paulo, Brasil (Insecta, Isoptera)". Arquivos do Instituto Biológico 25: 185-217.

Barsotti, R. C. and Costa-Leonardo, A. M. (2005) "The Caste System of *Coptotermes gestroi* (Isoptera:Rhinotermitidae)." Sociobiology 46(1): 87-103.

Bignell, D. (2000) "Introduction to Symbiosis". Termites: evolution, sociality, symbiosis, ecology." T. Abe, D. Bignell and M. Higashi. Dordrecht, Kluwer Academic Publishers: 466.

Breznack, J. A. e Brune, A. (1995) "Role of microorganisms in the digestion of lignocellulose by termites." Annual Review of Entomology 39: 453-487.

Brune, A., Emerson, D. e Breznack, J. A. (1995) "Termite gut microflora as an oxygen sink: microelectrode determinations of oxygen and pH gradients in guts of lower and higher termites." Applied and Environmental Microbiology 61: 2681-2687.

Chevreur, B.; Pfisterer, T. et al. (2004) "Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs." Genome Research 14: 1147-1159.

Costa-Leonardo, A. M. (2002). Cupins-Praga: Morfologia Biologia e Controle. Rio Claro, Divisa. Costa-Leonardo, A. M. (2004). "A new interpretation of the defense glands of neotropical *Ruptitermes* (Isoptera, Termitidae, Apicotermatinae)." Sociobiology 44(2):391-402.

Costa-Leonardo, A. M.; Arab, A. et al. (2005). "Note about the first swarming in a laboratory colony of *Coptotermes gestroi* (Isoptera: Rhinotermitidae)." Sociobiology 45(1): 137-139.

Costa-Leonardo, A. M.; Barsotti, R. C. et al. (1999). "Review and update on the biology of *Coptotermes havilandi* (Isoptera, Rhinotermitidae)." Sociobiology 33(3): 339-356.

Coy, M. R.; Salem, T. Z. et al (2010). "Phenol-oxidizing laccases from the termite gut." Insect Biochemistry and Molecular Biology 1-10.

de Andrade, T.; Moreira, L. et al. (2010) "Expression of New Red Cell Related Genes in Erythroid Differentiation." Biochemical Genetics, 48: 164-171.

de Kort, C.D.A.e Granger, N.A. (1996) "Regulation of JH titers: The relevance of degradative enzymes and binding proteins." Arch. Insect Biochem. Physiol. 33: 1- 26.

de Kort, C.A.D. e Koopmanschap, A.B. (1989) "Binding of juvenile hormone III to lipophorin from the American cockroach, *Periplaneta americana*." Archives of Insect Biochemistry and Physiology 11: 159-172.

de Kort, C.A.D. e Koopmanschap, A.B. (1987) "Specificity of binding of juvenile hormone III to hemolymph proteins of *Leptinotarsa decemlineata* and *Locusta migratoria*." Experientia 43 :904-905.

de Kort, C.A.D.e Koopmanschap, A.B. (1986) "High molecular weight transport proteins for JH-III in insect hemolymph." Experientia 42 : 834-836.

Dressman, D.; Traverso, Y. T. et al. (2003) "Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations." Proceedings of the National Academy of Science of the United States of America 100: 8817-8822.

Ewing, B. e Green, P. (1998) "Base-calling of automated sequencer traces using phred. II. Error probabilities." Genome Research, 8 (3):186-194.

Fontes, L. e Berti Filho, E. (1998) Cupins: O desafio do conhecimento. Piracicaba, Fundação de Estudos Agrários Luiz de Queiroz.

Fontes, L. and S. Milano (2002) "Termites as an urban problem in South America." Sociobiology 40: 103 - 151.

Fujita, A.; Miura, T. e Matsumoto, T. (2008) "Differences in cellulose digestive systems among castes in two termite lineages." Physiological Entomology 33, 73-82.

Gilbert, L.I. (2004) "Halloween genes encode *P450* enzymes that mediate steroid hormone biosynthesis in *Drosophila melanogaster*." Molecular and Cellular Endocrinology 215: 1–10.

Hartfelder, K. (2000) "Insect juvenile hormone: from “status quo” to high society." Brazilian Journal of Medical and Biological Research 33: 157-177.

Helvig, C.; Koener, J.K.et al. (2004) "CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata." Proceedings of the National Academy of Science of the United States of America 101(12): 4024-4029.

Higuchi, R., C. Fockler, et al. (1993) "Kinetic PCR analysis: real-time monitoring of DNA amplification reactions." Biotechnology 11(9): 1026-1030.

Howard, R.W. e Haverty, M.I. (1981) "Production of soldiers and maintenance of soldier proportion by laboratory experimental groups of *Reticulitermes flavipes* (Kollar) and *Reticulitermes virginicus* (Banks) (Isoptera: Rhinotermitidae)." Insectes Sociaux, 28: 32-39.

Howard, R. W. e Haverty, M. I. (1979) "Termites and juvenile hormone analogues: a review of methodology and observed effects." Sociobiology, 4: 269-278.

Huang, X. e Madan, A. (1999) "CAP3: A DNA sequence assembly program." Genome Research 9(9): 868-877.

Husseneder, C., E. L. Vargo, et al. (2003) *Molecular Genetic Methods: New*

Approaches to Termite Biology. Wood Deterioration and Preservation: Advances in Our Changing World. B. Goodnell, D. D. Nicholas and T. P. Schultz, American Chemical Society Sympson Series. 845: 358 - 370.

Inward, D. J. G.; Vogler, A. P.; e Eggleton, P. (2007) "A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology." Molecular Phylogenetics and Evolution 44: 953-967.

Jones, G. (1995) "Molecular mechanisms of action of juvenile hormone." Annual Review of Entomology 40: 147-69.

Keeling, C. I., Blomquist, G. J. e Tittiger, C. (2004) "Coordinated gene expression for pheromone biosynthesis in the pine engraver beetle, *Ips pini* (Coleoptera: Scolytidae)." Naturwissenschaften 91: 324-328.

Keiser, K. C.; Brandt, K. S. et al. (2002) "Cloning, partial purification and in vivo developmental profile of expression of the juvenile hormone epoxide hydrolase of *Ctenocephalides felis*." Archives of Insect Biochemistry and Physiology 50(4): 191-206.

Kintisch, E. (2006) "Biofuels to be focus of DOE centers." Science. 313: 746a.

Kirk, T. K. and R. L. Farrel (1987) "Enzymatic "combustion": The microbial degradation of lignin." Annual Review of Microbiology 41: 465 - 505.

Kirton, L. and V. Brown (2003) "The taxonomic Status of pest Species of *Coptotermes* in Southeast Asia: Resolving the Paradox in the Pest Status of the Termites, *Coptotermes gestroi*, *C. havilandi* and *C. travians* (Isoptera: Rhinotermitidae)." Sociobiology 42(1): 43 - 63.

Koopmanschap, A.B.e de Kort, C.A.D. (1988) "Isolation and characterization of a high molecular weight JH-III transport protein in the hemolymph of *Locusta migratoria*." Archives of Insect Biochemistry and Physiology. 7: 105-118.

Korb, J.; Roux, E.A.; Lenz, M. (2003) "Proximate factors influencing soldier development in the basal termite *Cryptotermes secundus* (Hill)". Insectes Sociaux 50: 299-303.

Korb, J., T. Weil, et al. (2009) "A gene necessary for reproductive suppression in termites." Science 324 (5928): 758.

Koyama, T.; Matsubara, M.; Ogura, K. (1985) "Isoprenoid enzyme systems of silkworm. I. Partial purification of isopentenyl pyrophosphate isomerase, farnesyl pyrophosphate synthetase, and geranylgeranyl pyrophosphate synthetase." Journal of Biochemistry 98: 449-456.

Kramer, K.J.; Dunn, P.E. et al. (1976) "Purification and characterization of the carrier protein for juvenile hormone from the haemolymph of the tobacco hornworm *Manduca sexta* Johannson." The Journal of Biological Chemistry 251: 4979-4985.

Krishna, K. e F. Weesner (1969) *Biology of termites*. New York, Academic Press.

Lange, J.P. (2007) "Lignocellulose conversion: an introduction to chemistry, process and economics." Biofuels Bioproducts and Biorefinery 1: 39-48 .

Lebrun, D. (1967) "La détermination de castes du termite à cou jaune (*Calotermes flavicollis* Fabr.)." Bulletin Biologique de la France et de la Belgique 101: 139 - 217.

Lee, K. and T. Wood (1971) *Termites and soils*. New York, Academic Press.

Lee, J. M.; Hatakeyama, M. et al. (2000) "A simple and rapid method for cloning insect vitellogenin cDNAs." Insect Biochemistry and Molecular Biology 30 (3): 189-194.

Leonardo, F. (2006) Identificação de genes em *Coptotermes gestroi* (Isoptera, Rhinotermitidae). Campinas, Universidade Estadual de Campinas.

Lüscher, M. (1958) "Über die Entstehung der Soldaten bei Termiten." Revue Suisse Zoologie 65: 372 - 376.

Margules, M.; Egholm, M. et al. (2005) "Genome sequencing in open microfabricated high density picoliter reactors." Nature, v.437, n. 7057, p. 376-380.

Meijerink, J., C. Mandigers, et al. (2001) "A novel method to compensate for different amplification efficiencies between patient DNA samples in quantitative real-time PCR." Journal of Molecular Diagnostics 3(2): 55-61.

Miura, T. (2004) "Proximate mechanisms and evolution of caste polyphenism in social insects: From sociality to genes". Ecological Research 19: 141-148.

Moore, J. (1968) "Studies on the chemical composition and function of the cephalic gland secretion in Australian termites." Journal of Insect Physiology 14: 33 - 39.

Myles, T. G. and W. L. Nutting (1988) "Termite eusocial evolution: a re-examination of Bartz' hypothesis and assumptions." Quarterly Review of Biology 63: 1 - 23.

- Nakashima, K., H. Watanabe, et al. (2002) "Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki." Insect Biochemistry and Molecular Biology 32(7): 777-84.
- Neoh, K.B. & Lee, C. Y. (2009) "Flight Activity and Flight Phenology of the Asian Subterranean Termite, *Coptotermes gestroi* (Blattodea: Rhinotermitidae)." Sociobiology 54(2): 521-530.
- Ning, Z.; Cox, A.J. e Mullikin, J.C. (2001) "SSAHA: a fast search method for large DNA databases." Genome Research 11(10): 1725-1729.
- Noirot, C. (1955) "Recherches sur le polymorphisme des termites supérieures (Termitidae)." Annales des Sciences naturelles (Zoologie) 11(17): 399 - 595.
- Noirot, C. (1969) Glands and secretion. Biology of termites. K. KRISHNA and F. WEESNER. London, Academic Press. 1: 89 - 123.
- Noirot, C. (1985) Pathways of caste development in the lower termites. Caste differentiation in social insects. J. A. L. Wartson. Oxford, Pergamon Press: 405
- Noirot, C. (1989) "Social structure in termite societies." Ethology, Ecology & Evolution 1: 1 - 17.
- Noirot, C. (1990) Sexual caste and reproductive strategies in termites. Social Insects - An evolutionary approach to castes and reproduction. W. Engels. New York, Springer - Verlag: 5 - 35.
- Noirot, C. and J. Darlington (2000) Termite nests: architecture, regulation and defense. Termites: Evolution, Sociality, Symbioses, Ecology. T. Abe, D. Bignell and M. Higashi. Dordrecht, Kluwer Academic Publishers: 121 - 139.
- Nunes, F. M., V. Valente, et al. (2004) "The use of Open Reading frame ESTs (ORESTES) for analysis of the honey bee transcriptome." BMC Genomics 5(1): 84.
- Park, Y.I. e Raina, A. K. (2004) "Juvenile hormone III titers and regulation of soldier caste in *Coptotermes formosanus*." Journal of Insect Physiology 50: 561-566.
- Park, Y.I. e Raina, A.K. (2003) "Factors regulating caste differentiation in the formosan subterranean termite with emphasis on soldier formation." Sociobiology 41: 1-12.
- Robinson, W. (1996) Urban entomology: insect and mite pest in the human environment. London, Chapman & Hall.
- Roison, Y. (1988) "The caste system of *Parrhinotermes browni* (Isoptera, Rhinotermitidae)." Sociobiology 14: 21 - 28.
- Romagnano, L. F. T. e Nahuz, M. A. R. "(2006) Controle de cupins subterrâneos em ambientes construídos". Técnica A revista do engenheiro civil IPT, 114: 1.

- Ronaghi, M. (2001) "Pyrosequencing sheds light on DNA sequencing." Genome Research 11(1): 3-11.
- Ronaghi, M. et al. (1998) "A sequencing method based on real-time pyrophosphate." Science, 281: 363–365.
- Rouland, C. (2000) Symbiosis with fungi. Termites: Evolution, Sociality, Symbioses, Ecology. T. Abe, D. Bignell and M. Higashi. Dordrecht, Kluwer Academic Publishers: 289 - 306.
- Saha, B. (2003) "Hemicellulose conversion." Journal of Industrial Microbiology and Biotechnology 30: 279 - 291.
- Scharf, M. E. e Tartar, A. (2008) "Termite digestomes as sources for novel lignocellulases." Biofuels, Bioproducts and Biorefining 6 (2): 540-552.
- Scharf, M. E., Buckspan, C. E. et al (2007) "Regulation of polyphenic caste differentiation in the termite *Reticulitermes flavipes* by interaction of intrinsic and extrinsic factors." The Journal of Experimental Biology 210: 4390-4398.
- Scharf, M., X. Zhou, et al. (2005) "The application of molecular genomics in addressing long-standing questions on termite biology". Proceedings of the 5th International Conference on Urban Pests, Malaysia, Perniagaan Ph'ng @ P&Y Design Network.
- Scharf, M. E., D. Wu-Scharf, et al. (2003) "Caste- and development-associated gene expression in a lower termite." Genome Biology 4(10): R62.
- Scharf, M. E., D. Wu-Scharf, et al. (2005) "Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*." Insect Molecular Biology 14(1): 31-44.
- Sieber, R. and Leuthold (1982) "Repeated copulation and testis enlargement in *Macrotermes michaelseni*." Physiological Entomology 7: 457 - 465.
- Slaytor, M. (1993) "Cellulose digestion in termites and cockroaches: what role do symbionts play?" Comparative Biochemistry and Physiology 103B:775–784.
- Slaytor, M. and D. J. Chappell (1994) "Nitrogen metabolism in termites." Comparative Biochemistry and Physiology 107B: 1 - 10.
- Su, N.Y. (2002) "Novel technologies for subterranean termite control." Sociobiology 40: 95-101.
- Su, N. and R. Scheffrahn (1998) "A review of subterranean termite control practices and prospects for integrated pest management programs." Integrated pest management Reviews 3: 1-13.

Tarver, M. R.; Zhoul, X. et al. (2010) "Environmental and endocrine influences on developmental and caste-regulatory gene expression in the eusocial termite *Reticulitermes flavipes*." BMC Molecular Biology 11:28.

Telles, G. e da Silva, F. (2001) "Trimming and clustering sugarcane ESTs". Genetics and Molecular Biology 24: 17-23.

Thorne, B. L. (1982) Multiple primary queens in termites: phyletic distribution, ecological context, and a comparison to polygyny in Hymenoptera. The biology of social insects. M. C. e. a. Breed. Boulder, Westview Press: 206 - 211.

Todaka, n. Morya, S. et al. (2007) "Enviromental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protistic community of *Reticulitermes speratus*". FEMS Microbiology Ecology 59: 592-599.

Tokuda, G., H. Saito, et al. (2002) "A digestive beta-glucosidase from the salivary glands of the termite, *Neotermes kosshunensis* (Shiraki): distribution, characterization and isolation of its precursor cDNA by 5'- and 3'-RACE amplifications with degenerate primers." Insect Biochemistry and Molecular Biology 32(12): 1681-1689.

Tokuda, G., H. Watanabe, et al. (1997) "Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis* (Shiraki): distribution of cellulases and properties of endo-beta-1,4-glucanase." Zoological Science 14(1): 83-93.

Touhara, K.e Prestwich, G.D. (1994) "Role of juvenile hormone binding protein in modulating function of JH epoxide hydrolase in eggs of *Manduca sexta*". Insect Biochemistry and Molecular Biology 24: 641-646.

Vandesompele, J.; De Preter, K. et al. (2002) "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." Genome Biology 3(7): RESEARCH0034.

Wang, L.; Feng, Z.; Wang, X.; Wang X.; e Zhang, X. (2010) "DEGseq: an R package for identifying differentially expressed genes from RNA-seq data" Bioinformatics 26: 136-138.

Ward, V.K.; Bonning, B.C.; et al. (1992) "Analysis of the catalytic mechanism of juvenile hormone esterase by sitedirected mutagenesis." International Journal of Biochemistry 24: 1933-1941.

Warnecke, F., Luginbühl, P. et al. (2007) "Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite." Nature 450: 560 - 565.

Watanabe, H. and G. Tokuda (2001) "Animal cellulases." Cellular and Molecular Life Sciences 58(9): 1167-1178.

Weil, T., J. Korb, et al. (2009) "Comparison of queen-specific gene expression in related lower termite species." Moecularl Biology and Evolution 26(8): 1841-1850.

Whitfield, C. W.: Band, M. R. et al. (2002) "Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee." Genome Research 12(4): 555-566.

Wilson, E. O. (1971) *The Insect Societies*. Cambridge, MA, The Belknap Press of Harvard University.

Wilson, E.O., Bossert, W. H. (1963) "Chemical communication among animals." Recent Progress in Hormone Research 19: 673-716.

Wu-Scharf, D., M. E. Scharf, et al. (2003) "Expressed sequence tags from a polyphenic *Reticulitermes flavipes* (Isoptera : Rhinotermitidae) cDNA library." Sociobiology 41(2): 479-490.

Zhou, X., F. M. Oi, et al. (2006a) "Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites." Proceedings of the National Academy of Science of the United States of America 103(12): 4499-4504.

Zhou, X.; Song, C. et al. (2006b) "Juvenile hormone and colony conditions differentially influence cytochrome P450 gene expression in the termite *Reticulitermes flavipes*". Insect Molecular Biology 15(6): 749-761.

Zorzenon, F. and M. Potenza (1998) *Cupins: pragas em áreas urbanas*. Boletim Técnico do Instituto Biológico. São Paulo. 10: 40.

ANEXOS

Analysis of the workers head transcriptome of the Asian subterranean termite, *Coptotermes gestroi*

F.C. Leonardo¹, A.F. da Cunha², M.J. da Silva³,
M.F. Carazzolle^{1,4}, A.M. Costa-Leonardo⁵, F.F. Costa⁶
and G.A. Pereira^{1*}

¹Laboratório de Genômica e Expressão, Departamento de Genética Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, CEP: 13083-970, Campinas, São Paulo, Brazil: ²Laboratório de Bioquímica e Biotecnologia molecular, Departamento de Genética e Evolução, Universidade Federal de São Carlos, Via Washington Luis Km 235, CEP:13565-905, São Carlos, São Paulo, Brazil: ³Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Avenida Cândido Rondon, 400, CEP: 13083-875, Campinas, São Paulo, Brazil: ⁴Centro Nacional de Processamento de Alto Desempenho em São Paulo, Rua Saturnino Brito, 45, Cidade Universitária, CEP: 13083-889, Universidade Estadual de Campinas, São Paulo, Brazil: ⁵Departamento de Biologia, Universidade Estadual Paulista, Avenida 24^a, Bela Vista, CEP: 13506-900, Rio Claro, São Paulo, Brazil: ⁶Centro de Hematologia e Hemoterapia, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Rua Carlos Chagas, 480, CEP: 13083-878, Campinas, São Paulo, Brazil

Abstract

The lower termite, *Coptotermes gestroi* (Isoptera: Rhinotermitidae), is originally from Southeast Asia and has become a pest in Brazil. The main goal of this study was to survey *C. gestroi* transcriptome composition. To accomplish this, we sequenced and analyzed 3003 expressed sequence tags (ESTs) isolated from libraries of worker heads. After assembly, 695 uniESTs were obtained from which 349 have similarity with known sequences. Comparison with insect genomes demonstrated similarity, primarily with genes from *Apis mellifera* (28%), *Tribolium castaneum* (28%) and *Aedes aegypti* (10%). Notably, we identified two endogenous cellulases in the sequences, which may be of interest for biotechnological applications. The results presented in this work represent the first genomic study of the Asian subterranean termite, *Coptotermes gestroi*.

Keywords: *Coptotermes gestroi*, ESTs, cDNA library, Isoptera, cellulose

(Accepted 20 August 2010)

Introduction

In recent years, the genomes of key insects such as *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Bombyx mori*, *Tribolium castaneum* and *Acyrtosiphon pisum* have been sequenced. Although novel sequencing methods, such as pyrosequencing (Ronaghi, 2001), promise a new era in

*Author for correspondence
Fax: 0055-19-35216235
E-mail: goncalo@unicamp.br

42	genomics, sequencing of cDNA libraries produced from whole	106
43	body or selected tissues is still an efficient strategy to gain	107
44	insights into genetics and physiology of non-model insects.	108
45	Termites are social insects that can be roughly separated	
46	into two major groups, higher and lower termites, according	
47	to the symbionts found in their gut (Grassé, 1949). Higher	
48	termites harbor only bacteria in the digestive tract and include	
49	only species of the family Termitidae. In addition to bacterial	
50	flora, the lower termites also present symbiotic protozoa in	
51	their gut, which contribute to their digestive processes. A	
52	variety of molecular studies have been conducted to under-	
53	stand caste polymorphism in termites. Wu-Scharf <i>et al.</i> (2003)	
54	described 88 expressed sequence tags (ESTs), generated from a	
55	polyphenic cDNA library from <i>Reticulitermes flavipes</i> (Wu-	
56	Scharf <i>et al.</i> , 2003). Among these sequences, they found genes	
57	that were caste-specifically expressed (Scharf <i>et al.</i> , 2003, 2005),	
58	including two hexamerin encoding genes. With the use of	
59	RNAi, it was discovered that dual silencing of both hexamer-	
60	ins leads to significant increases in JH-dependent presoldier	
61	differentiation (Zhou <i>et al.</i> , 2006).	
62	Since most of the termites are xylophagous, they have the	
63	ability to digest lignocellulose due to lignocellulose enzymes	
64	present in their digestive tract and in their gut symbionts	
65	(Nakashima <i>et al.</i> , 2002; Scharf <i>et al.</i> , 2003). These enzymes	
66	have demonstrated a great potential for bioenergy appli-	
67	cations and led to a new era in research (Scharf & Tartar, 2008).	
68	Digestive genomic studies include symbiont metatran-	
69	scriptomic sequencing of cDNA clones, representing ex-	
70	pressed genes of eukaryotic (protistan) symbionts from the	
71	lower termite, <i>Reticulitermes speratus</i> (Todaka <i>et al.</i> , 2007).	
72	This study revealed diverse cellulase and hemicellulase	
73	sequences in this termite, while another investigation was	
74	conducted to elucidate the symbiont metagenome sequencing	
75	of a higher <i>Nasutitermes</i> species from Costa Rica (Warnecke	
76	<i>et al.</i> , 2007). This study produced over 100 million bases of	
77	DNA sequence and revealed genes that relate to numerous	
78	aspects of microbial life in the gut microenvironment. Most	
79	recently, dual-host symbiont transcriptome sequencing was	
80	accomplished to determine genes expressed in the gut of the	
81	termite, <i>Reticulitermes flavipes</i> . The results increased the	
82	number of genes of the host and symbiotic glycosyl hydrolases	
83	families, supporting previous models of lignin degradation	
84	and host-symbiotic collaboration in hemicellulose digestion	
85	(Tartar <i>et al.</i> , 2009).	
86	Among the Rhinotermitidae, <i>Coptotermes gestroi</i> is marked	
87	as the most destructive pest termite, damaging structural	
88	wood in the urban areas in Southeast Asia (Kirtan, 2005). This	
89	endemic Southeast Asian species has spread in various regions	
90	around the world: Marquesas Islands (Pacific Ocean),	
91	Mauritius and Reunion (Indian Ocean), Brazil, Barbados,	
92	West Indian Islands, southern Mexico and the United States	
93	(Jenkins <i>et al.</i> , 2007).	
94	Although <i>C. gestroi</i> has had an important economic impact,	
95	to date, only a few sequences have been produced from this	
96	species, and most of these correspond to ribosomal or	
97	mitochondrial fragments (Jenkins <i>et al.</i> , 2007). In this study,	
98	we report the identification of 695 <i>C. gestroi</i> ESTs, interpreting	
99	their possible role in the biology of this species. Among these	
100	ESTs, we identified two cellulases: endo- β -1,4-glucanase	
101	and α -glucosidase. Comparative analysis showed that	
102	these enzymes are very similar to counterparts found in	
103	other termite species and, therefore, are likely to represent	
104	genuine termite enzyme coding genes and not enzymes from	
105	the symbiotic fauna of the workers' digestive tract.	
	Quantitative PCR of different developmental stages was	106
	performed to demonstrate the comparative expression of	107
	these cellulases.	108
	Experimental procedures	109
	<i>Biological samples and RNA extraction</i>	110
	Workers of <i>C. gestroi</i> were obtained from a field colony	111
	located in the city of Rio Claro, São Paulo State, Brazil (22°23'S,	112
	47°32'W). In order to avoid contamination by gut symbiotic	113
	protists, only the workers' heads were used in the experiment.	114
	Prior to RNA isolation, the workers had their heads cut off,	115
	and these were frozen at -80°C. Total RNA was isolated from	116
	heads using TRIzol reagent (Invitrogen).	117
	<i>Generation of ESTs (expressed sequence tags)</i>	118
	Aliquots of 1.85 μ g of total RNA were reverse transcribed	119
	utilizing SuperScript II Reverse Transcriptase (Invitrogen),	120
	primer Oligo dT [5'-GGCGGCCGCACAACTTTGTACAAG-	121
	AAAGTTGGGT(T) ₁₉ -3'] and primer Poly G (5'-TCGTCGGG-	122
	GACAACTTTGTACAAAAAAGTTGG-3'). First-strand cDNA	123
	synthesis occurred at 42°C for 60 min in a total volume of 10 μ l.	124
	The cDNAs contained in 2 μ l of each RT-product were	125
	then amplified by long-distance polymerase chain reaction	126
	(LD-PCR), which is a technique that enables the amplification	127
	of large cDNAs from an uncloned pool of mRNA ex-	128
	tension products using the same primers in a PCR mix (5.0 μ l	129
	10 $\times$ Advantage 2 PCR buffer, 1.0 μ l primer Oligo dT, 1.0 μ l	130
	primer poly G, 1.0 μ l enzyme 50 $\times$ Advantage 2 polymerase	131
	mix and 40 μ l distilled water). The amplification protocol	132
	consisted of an initial step at 95°C for 20s, followed by	133
	20 cycles of 95°C for 5s and 68°C for 6 min.	134
	Aliquots of the PCR products (5 μ l) were run on 1% agarose	135
	gels and stained with ethidium bromide. The PCR product	136
	will produce a <i>smear</i> that ranges from 0.1 to 4.0Kb, rep-	137
	resenting the most abundant transcripts. The amplification	138
	products of over 600bp were then extracted from the agarose	139
	gels (Wizard-SV gel and PCR Clean-up System, Promega)	140
	and ligated into a pDONR-222 vector (Invitrogen). Ligation	141
	was transformed by electroporation into competent <i>E. coli</i>	142
	DH10B-cells and selected for kanamycin resistance. Plasmid	143
	DNA was purified and sequenced using M13 primer (5'-	144
	GTAAACGACGGCCAG-3'). High throughput EST sequen-	145
	cing was performed using dye-terminator chemistry on an	146
	ABI 3700 sequencer.	147
	<i>Sequence analysis</i>	148
	The EST sequences were base-called and screened for	149
	vector sequences using PHRED and cross-match software	150
	(Ewing & Green, 1998). Poly-A repeats and low quality	151
	sequences were trimmed (Telles & da Silva, 2001), and the	152
	high quality ones (at least 100bp with Phred>20) were	153
	clustered and assembled by CAP3 (Huang & Madan, 1999).	154
	The clusters were submitted to automatic and manual	155
	annotation, where the sequences were used as queries for	156
	similarity searches by BLASTn, tBLASTx and BLASTx	157
	(Altschul <i>et al.</i> , 1997) against GenBank (nucleotide and	158
	amino acid databases). Database searches were performed	159
	under default settings, with a threshold e-value of lower than	160
	10 ⁻⁵ regarded as significant. In addition, we compared the	161
	<i>C. gestroi</i> sequences with those of the 18,122 Isoptera EST	162

Table 1. Primers used in the relative gene expression experiment.

Primers	Sequence 5–3'	Concentration
β -glucosidase F	GATAATAGAGTTGTACCGTGG	150 nM
β -glucosidase R	ACTCCACCGTAGTCCGAGAAAC	
Endo- β -1,4-glucanase F	GTGCCTTCAACTGGGATAACAA	150 nM
Endo- β -1,4-glucanase R	TGTATGCCTGCTTGCTTGTA	
β -actin F	GTGTGCAAATGTAATCCATCC	300 nM
β -actin R	CTGACGTACTGTGCGCTGT	
α -tubulin F	CCCTCTCTTCGCCCTTCAA	150 nM
α -tubulin R	GAGGATCTCGCTGCTCTGGA	

sequences available in a public database (www.ncbi.nlm.nih.gov) using tBLASTx with a threshold e-value of lower than 10^{-10} .

The clusters that presented 'no database matches' (no hits) were analyzed by the ESTScan program (Iseli *et al.*, 1999), employing the Drosophila model for CDS prediction. GC content of the coding and non-coding sequences and GC percentage at different positions (GC1, GC2 and GC3) were obtained using the predicted CDS and comparing them with CDS of the hits clusters. Functional classification was performed using the Gene Ontology database (www.geneontology.org).

Relative gene expression

Total RNA was extracted from the whole body of different castes and developmental stages (larvae, soldiers, alates, queens and workers) using QIAGEN RNeasy Mini Kit. RNA samples (5 μ g) were incubated with 1U DNaseI (Invitrogen Corp., Rockville, MD, USA) for 15 min at room temperature, and EDTA was added to a final concentration of 2 mM to stop the reaction. The DNaseI enzyme was subsequently inactivated by incubation at 65°C for 5 min. DNaseI-treated RNA samples were then reverse-transcribed with Superscript III and RNaseOut (Invitrogen Corp., Carlsbad, CA, USA) for 50 min at 50°C, 15 min at 70°C. cDNA samples were quantified using a Nanodrop spectrophotometer (ND-1000; Nanodrop Technologies Inc., Wilmington, DE, USA). Synthetic oligonucleotide primers were designed to amplify cDNA for the genes encoding the two cellulases (Primer ExpressTM, Applied Biosystems, Foster City, CA, USA). For primer sequences and concentrations, see table 1. Primers were synthesized by Invitrogen Corp. (Carlsbad, CA, USA) and IDT (Coralville, IA, USA). All samples were assayed in a 12- μ l volume, containing 5 ng cDNA (3 μ l), 6 μ l SYBR Green Master Mix PCR (Applied Biosystems), and 3 μ l-specific primers in a MicroAmp Optical 96-well reaction plate (Applied Biosystems), using the Step one Plus (Applied Biosystems), as described previously. Gene expression analyses were performed with the geNorm program (Vandesompele *et al.*, 2002). Two replicates were run on the plate for each sample, and the data are mean values of three independent preparations. Results are expressed as mRNA levels of each gene studied, normalized according to β -actin and α -tubulin expressions.

Statistical analyses

To evaluate expression patterns (real-time PCR), one-way analyses of variance (ANOVA) were performed separately

for each gene using caste (larvae, nymph, workers and soldiers) as fixed factors. Tukey's multiple comparison test was applied to reveal pair-wise differences. Statistical analyses were performed using Graph Pad Prism (version 5.0).

Results

cDNA library construction and analysis

We generated and analyzed 4697 sequences of a cDNA library from worker heads of the subterranean termite *C. gestroi* (general scheme in fig. 1). From the remaining sequences, 3003 were considered of appropriate quality – presenting at least 100 bases with Phred ≥ 20 . These sequences were clustered by CAP3 into 245 contigs and 450 singlets, and the EST size of the contigs is shown in fig. 1. Sequences were submitted to similarity searches by BLAST against GenBank and, after manual annotation, resulted in 321 (46.19%) hits and 374 (53.81%) no hits. All these sequences were deposited in GenBank's dbEST database and are available using the accession numbers GT566617–GT568042.

From the 321 ESTs, 42 unigenes presented similarity with non-insect genes such as bacteria, protozoa, fungi and plants. The other 271 ESTs were assigned as the termite's sequences and were compared to other insect genomes.

ESTScan analysis

The ESTScan analyses applied to no-hits sequences identified 129 predicted CDS. In order to validate these sequences, a GC content analysis was performed, and the results were compared with GC content from *C. gestroi* hits and sequences of other Isoptera (table 2).

Gene ontology classification

For the putative genes with known function, we assigned biological processes by using level three of the Gene Ontology (GO) classification (Ashburner *et al.*, 2000). There is a clear prevalence for ESTs representing cellular physiological process and metabolism (fig. 2). This result was also found in a previous study of *Apis mellifera*'s ESTs (Nunes *et al.*, 2004).

Genome comparison

We compared *C. gestroi* sequences with those from insects whose genomes have been completely sequenced. The distribution of the respective best hits were with *A. mellifera* (28%) and *Tribolium castaneum* (28%), followed by *Aedes aegypti* (10%), *Anopheles gambiae* (8%) and *Drosophila sp.* (7%).

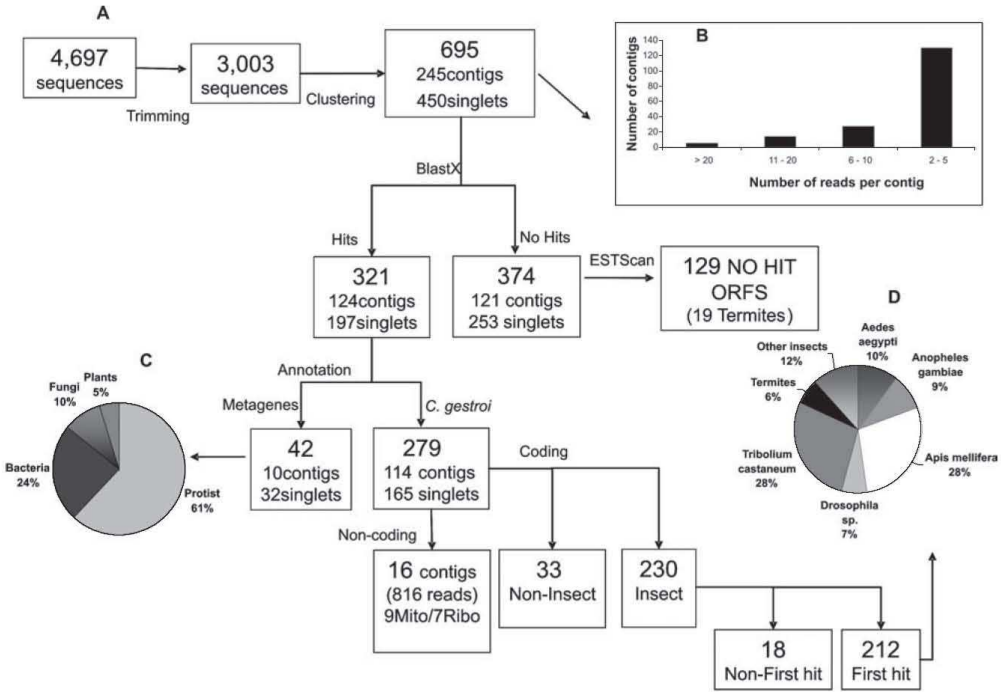


Fig. 1. General scheme representing how the sequences were grouped and their manual annotation. (a) Fluxogram generated by the treatment of the sequences, (b) distribution of reads in clusters, (c) metagenes, and (d) first hit comparison with insect genomes.

Table 2. Base composition (%GC) at different positions for coding sequences of *C. gestroi*'s library and Isoptera sequences available in a public database (GenBank).

GC position	Isoptera	<i>C. gestroi</i> hits	<i>C. gestroi</i> no hits
GC1	52 ± 5	53 ± 6	54 ± 9
GC2	40 ± 7	40 ± 5	43 ± 10
GC3	52 ± 10	49 ± 10	55 ± 8

Even though the number of sequences we obtained for *C. gestroi* is relatively low, an interesting finding was that we obtained ESTs representing two cellulases: endo- β -1,4-glucanase (EC 3.2.1.4) (fig. 3) and β -glucosidase (EC 3.2.1.21) (fig. 4), which have been previously identified in salivary glands of other termites species.

Gene expression analyzes

To validate and analyze the expression of the putative cellulases found in the cDNA library, relative gene expression experiment was performed using RNA samples derived from different castes and developmental stages (larvae, alates, soldiers, queens and workers). Figure 5 shows relative

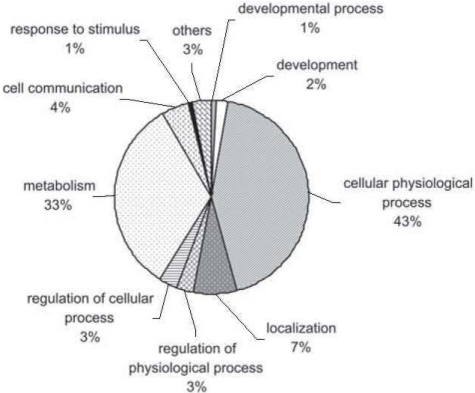


Fig. 2. Gene ontology classification of *C. gestroi* cDNA library sequences.

expression levels of both cellulases in all castes. The statistical analysis showed that the expression of both genes is higher in workers compared to the other developmental stages.

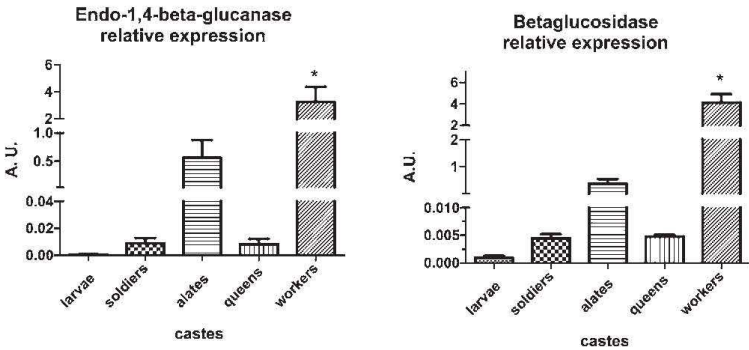


Fig. 5. Relative expression levels of (a) endo- β -1,4-glucanase and (b) β -glucosidase in different castes measured by real-time PCR. The Y-axis indicates gene expression in arbitrary units and data are mean values of three independent preparations. Significant differences are indicated (workers have a significant difference compared to the other castes; Tuckey: *, $P < 0.05$).

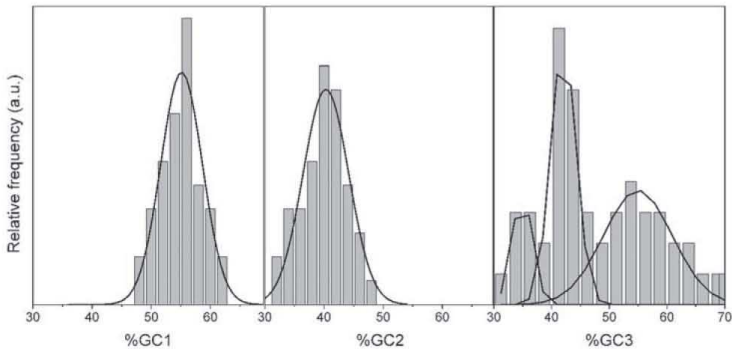


Fig. 6. Distribution of GC content at different positions in the *C. gestroi* hits sequences.

the 266 ORFs of *C. gestroi* reported in this work. This pattern suggests that these sequences are indeed suitable for translation and may be treated as true genes. Another explanation for the apparently large quantity of no hits is that they may represent 'taxonomically restricted genes' (TRGs). Every eukaryotic genome contains 10–20% of genes without any significant sequence similarity to genes of other species that are classified as 'orphans' or TRGs (Wilson *et al.*, 2005). Although such genes have arisen in the genomes of every group of organisms studied so far, they have received comparatively little attention, and their functions remain largely unknown (Khalturin *et al.*, 2009).

The 245 remaining no hits were also 'no ORFs' and might have been considered an artifact. However, non-hypothesis-driven gene expression studies on honeybee (Colonello-Frattini & Hartfelder, 2009) and fruit fly (Tupy *et al.*, 2005) transcripts also identified 'no ORF' ESTs that were clearly expressed in different organs. These sequences are thought to have a potential function as nuclear RNAs (Sawata *et al.*, 2004), micro RNAs (Bartel, 2004) or long non-coding RNAs (Mercer *et al.*, 2009). Therefore, the 'no ORF' *C. gestroi*

sequences should be analyzed further to identify possible functions.

When comparing *C. gestroi* sequences with those of other insects, the distribution of the first (best) hit demonstrated higher similarity to the honey bee than to dipteran species (fig. 1d). At first glance, this could suggest some similarity based on common molecular mechanisms to suit social life. However, the similarity between *C. gestroi* and *A. mellifera* is equivalent to the similarity to the flour beetle, *Tribolium castaneum*, a non-social insect species. The lack of a clear pattern may reflect the relative paucity of sequences from hemimetabolous species in the GenBank database. This kind of developmental profile may have an association with defined genomic structures that are not yet available for comparison.

The sequences were categorized by gene ontology in biological processes (fig. 2). A large portion of the genes (43%) was classified as having a role in cellular physiological process and metabolism, which represent basic functions. This result was also found in a previous study of *Apis mellifera* ESTs (Nunes *et al.*, 2004).

We analyzed the dispersion of GC content at the individual positions from putative hit-ORFs. As expected, at positions GC1 and GC2, the dispersion is lower (Sabater-Muñoz *et al.*, 2006) and presents a homogeneous Gaussian distribution (fig. 6). At GC3, dispersion is wider; but, curiously, we identified three distribution patterns. We hypothesized that these distributions could represent different classes of genes, but the analysis with the available set of genes failed to indicate a clear pattern.

Even though we were careful not to contaminate the samples with gut material, 42 uniESTs clearly derived from non-insect genes were also identified: ten sequences from various types of bacteria, 26 from protist, four from fungi and two from plants. Some bacterial sequences produced positive hits against a databank from marine metagenomics (Kannan *et al.*, 2007), soil fungi and soil bacteria, probably representing DNA from soil particles.

Most of the protist sequences presented *Trichomonas vaginalis* as a first hit. This does not mean, however, that this species is present in *C. gestroi*. This result probably indicates the presence of parabasalian protist, which has been reported in the termite digestive tract (Ohkuma, 2008). The most representative group found in the library was composed of genes with enzymatic activity. Two cellulase-encoding sequences were of particular interest to us because of their potential biotechnology applications. Only around ten years ago the first termite endogenous cellulase was reported (Tokuda *et al.*, 1997), changing the general belief that termite cellulase is produced only by the symbiotic microbes presented in their digestive tract.

The role of these cellulases in *C. gestroi* has not yet been elucidated, but the possibility of mimetizing termites for use of cellulose to convert biomass into energy resources has attracted much interest (Scharf & Tartar, 2008). Termites initiate the digestion of wood by mechanical degradation of the substrate with their powerful mandibles. Unlike some biotechnology processes, these mild conditions allow digestive enzymes to act quickly, with the first cellulolytic attack probably occurring in the mouth. Since we prepared cDNA of workers' heads, the cellulases we identified probably are present in the salivary glands. This would be consistent with previous reports that found a gene similar to the *C. gestroi* endo- β -1,4-glucanase (GT567748.1) gene expressed in the salivary glands of *C. formosanus* (Nakashima *et al.*, 2002) and *R. flavipes* (Zhou *et al.*, 2007).

After this initial digestion step, the substrate passes to the hindgut where a more complex enzymatic combination – possibly a dual cellulolytic system between termites and symbionts – works together to accomplish cellulose degradation and lignin removal (Todaka *et al.*, 2007). The other enzyme, β -glucosidase (GT566634.1), should possibly act in this step. Enzymes responsible for lignin degradation were not found in our library. However, recent studies reported that analyses of cDNA libraries from host termites and symbiotic protists presented genes related to lignin degradation (Kudo, 2009; Tartar *et al.*, 2009). A research conducted with the dampwood termite, *Zootermopsis angusticollis*, demonstrated that there is lignin degradation in the gut of this wood-feeding insect (Geib *et al.*, 2008).

To verify the expression of the cellulases in *C. gestroi* species, a quantitative expression analysis was conducted using RNA samples from different castes (larvae, workers, nymphs, alates, soldiers and queens). Both genes presented a higher expression in workers, compared to the other castes

($P < 0.05$). This result was expected considering that this caste executes the feeding of the colony, and it was consistent with a previous study conducted with *Reticulitermes flavipes* that found high levels of expression of this cellulase in workers, nymphs, alates and supplementary reproductives (Scharf *et al.*, 2005). The β -glucosidase gene is similar to another homolog, the Neofem 2, which is believed to play a role in reproductive suppression in the termite, *Cryptotermes secundus* (Weil *et al.*, 2007). A knockdown of Neofem 2 was conducted in queens, making the workers believe that the colony was queenless (Korb *et al.*, 2009). Conversely, our experiments show that, in *C. gestroi*, the expression of this gene is extremely low in queens, supporting the idea that this enzyme works together with others to efficiently degrade cellulose. Earlier RT-PCR experiments demonstrate that workers of the lower termite, *Neotermes kosshunensis*, are known to express β -glucosidase in the salivary glands (Tokuda *et al.*, 2002), supporting our findings since we generated a cDNA library from workers' heads. A differential study of the digestive cellulase expression among castes in two termite species showed that minor workers of *Nasutitermes takasagoensis* presented higher rates of expression of β -glucosidase, when compared to soldiers, medium and major workers, revealing a division of labor controlled by nutrition (Fujita *et al.*, 2008). Even though the rates were differentially expressed, the site of expression was higher in the midgut for all castes. In the termite *Hodotermopsis sjostesti*, however, the expression was higher in the salivary glands for the worker caste and in the hindgut for the soldiers (Fujita *et al.*, 2008). Further investigation are been carried out to locate the sites of expression of these cellulases in *C. gestroi*.

In summary, this is the first survey of the *C. gestroi* transcriptome, with a focus on the heads of workers, the caste involved in cellulose degradation. As previously discussed, investigations of the termite's lifestyle, combined with molecular analyses of its physiology and symbionts, indicate that integration of mechanical processes and enzymatic treatment of wood may be the key for efficient cellulose utilization. Clearly, these insects deserve further in-depth molecular studies to elucidate the metabolic pathways employed in lignocellulose degradation.

We are currently investigating the properties of these putative cellulase enzymes and their possible utilization in biotechnological processes.

Acknowledgements

We thank Klaus Hartfelder for helpful comments on previous versions of the manuscript and Nicolau Conran, HEMOCENTRO-UNICAMP, for help with English revision. This study was supported by FAPESP, Brazil (No. 06/59086-8 and No. 08/50114-4).

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25(17), 3389–3402.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinsk, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarski, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. & Sherlock, G. (2000) Gene

- ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25(1), 25–29.
- Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2), 281–297.
- Colonello-Frattini, N. & Hartfelder, K. (2009) Differential gene expression profiling in mucus glands of honey bee (*Apis mellifera*) drones during sexual maturation. *Apidologie* 40, 481–495.
- Ewing, B. & Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* 8(3), 186–194.
- Fujita, A., Miura, T. & Matsumoto, T. (2008) Differences in cellulose digestive systems among castes in two termite lineages. *Physiological Entomology* 33, 73–82.
- Geib, S.M., Filley, T.R., Hatcher, P.G., Hoover, K., Carlson, J.E., del Mar Jimenez-Gasco, M., Nakagawa-Izumi, A., Sleighter, R.L. & Tien, M. (2008) Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences of the United States of America* 105, 12932–12937.
- Grassé, P.P. (1949) Ordre des Isoptères ou termites. pp. 408–544 in Grassé, P.P. (Ed.) *Traité de Zoologie*. v. 9, Masson, Paris.
- Huang, X. & Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Research* 9(9), 868–877.
- Iseli, C., Jongeneel, C.V. & Bucher, P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. pp. 138–148 in *Proceedings of the 7th International Conference on Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park, CA, USA.
- Jenkins, T.M., Jones, S.C., Lee, C.Y., Forschler, B.T., Chen, Z., Lopez-Martinez, G., Gallagher, N.T., Brown, G., Neal, M., Thistleton, B. & Kleinschmidt, S. (2007) Phylogeography illuminates maternal origins of exotic *Coptotermes gestroi* (Isoptera: Rhinotermitidae). *Molecular Phylogenetics and Evolution* 42(3), 612–621.
- Kannan, N., Taylor, S.S., Zhai, Y., Venter, J.C. & Manning, G. (2007) Structural and functional diversity of the microbial kinome. *Public Library of Science Biology* 5(3), e17.
- Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. & Bosch, T.C. (2009) More than just orphans: are taxonomically restricted genes important in evolution? *Trends in Genetics* 25(9), 404–413.
- Kirton, L.G. (2005) The importance of accurate termite taxonomy in the broader perspective of termite management. pp. 1–7 in Lee, C.-Y. & Robinson, W.H. (Eds) *Proceedings of the Fifth International Conference on Urban Pests*. P & Y Design Network, Penang, Malaysia.
- Korb, J., Weil, T., Hoffmann, K., Foster, K.R. & Rehli, M. (2009) A gene necessary for reproductive suppression in termites. *Science* 324, 758.
- Kudo, T. (2009) Termite-Microbe Symbiotic System and Its Efficient Degradation of Lignocellulose. *Bioscience Biotechnology and Biochemistry* 73(12), 2561–2567.
- Mercer, T.R., Dinger, M.E. & Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* 10(3), 155–159.
- Nakashima, K., Watanabe, H., Saitoh, H., Tokuda, G. & Azuma, J.I. (2002) Dual cellulose digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. *Insect Biochemistry and Molecular Biology* 32(7), 777–784.
- Nunes, F.M., Valente, V., Sousa, J.F., Cunha, M.A., Pinheiro, D. G., Maia, R.M., Silva, W.A., Araujo, D.D., Costa, M.C.R., Martins, W.K., Carvalho, A.F., Monesi, N., Nascimento, A.M., Peixoto, P.M.V., Silva, M.F.R., Ramos, R.G.P., Reis, L.F.L., Dias-Neto, E., Souza, S.J., Simpson, A.J.G., Zago, M.A., Soares, A.E.E., Bitondi, M.M. G., Espreafico, E.M., Espindola, F.S., Paco-Larson, L., Simoes, Z.L.P., Hartfelder, K. & Silva, W.A. (2004) The use of Open Reading frame ESTs (ORESTES) for analysis of the honey bee transcriptome. *BioMed Central Genomics* 5(1), 84–1.
- Ohkuma, M. (2008) Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends in Microbiology* 16(7), 345–352.
- Ronaghi, M. (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Research* 11(1), 3–11.
- Sabater-Muñoz, B., Legeai, E., Rispe, C., Bonhomme, J., Dearden, P., Dossat, C., Duclert, A., Gauthier, J.P., Ducray, D.G., Hunter, W., Dang, P., Kambhampati, S., Martinez-Torres, D., Cortes, T., Moya, A., Nakabachi, A., Philippe, C., Prunier-Leterme, N., Rahbé, Y., Simon, J.C., Stern, D.L., Wincker, P. & Tagu, D. (2006) Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera). *Genome Biology* 7(3), R21.
- Sawata, M., Takeuchi, H. & Kubo, T. (2004) Identification and analysis of the minimal promoter activity of a novel non coding nuclear RNA gene, AncR-1, from the honeybee (*Apis mellifera* L.). *Rna* 10(7), 1047–1058.
- Scharf, M.E. & Tartar, A. (2008) Termites digestome as sources for novel lignocellulases. *Biofuels, Bioproducts and Biorefining* 6(2), 540–552.
- Scharf, M.E., Wu-Scharf, D., Pittendrigh, B.R. & Bennett, G.W. (2003) Caste- and development-associated gene expression in a lower termite. *Genome Biology* 4(10), R62.
- Scharf, M.E., Wu-Scharf, D., Zhou, X., Pittendrigh, B.R. & Bennett, G.W. (2005) Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. *Insect Molecular Biology* 14(1), 31–44.
- Tartar, A., Wheeler, M.M., Zhou, X., Coy, M.R., Boucias, D.G. & Scharf, M. (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnology for Biofuels* 2, 25.
- Todaka, N., Moriya, S., Saita, K., Hondo, T., Kiuchi, I., Hirotschi, T., Ohkuma, M., Piero, C., Hayashisaky, Y. & Kudo, T. (2007) Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. *Federation of European Microbiological Societies Microbiology Ecology* 59, 592–599.
- Tokuda, G., Watanabe, H., Matsumoto, T. & Noda, H. (1997) Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis* (Shiraki): distribution of cellulases and properties of endo-beta-1,4-glucanase. *Zoological Science* 14(1), 83–93.
- Tokuda, G., Saito, H. & Watanabe, H. (2002) A digestive β -glucosidase from the salivary glands of the termite, *Neotermes kosshunensis* (Shiraki): distribution, characterization and isolation of its precursor cDNA by 5'- and 3'-RACE amplifications with degenerate primers. *Insect Biochemistry and Molecular Biology* 32(12), 1681–1689.
- Telles, G. & da Silva, F. (2001) Trimming and clustering sugarcane ESTs. *Genetics and Molecular Biology* 24, 17–23.
- Tupy, J.L., Bailey, A.M., Dailey, G., Evans-Holm, M., Siebel, C. W., Misra, S., Celniker, S.E. & Rubin, G.M. (2005) Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* 102(15), 5495–5500.

- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B.,
 Van Roy, N., De Paepe, A. & Speleman, F. (2002) Accurate
 normalization of real-time quantitative RT-PCR data by
 geometric averaging of multiple internal control genes.
Genome Biology 3(7), RESEARCH0034.
- Warnecke, F., Luginbuhl, P., Ivanova, N., Ghassemian, M.,
 Richardson, T.H., Stege, J.T., Cayouette, M., McHardy, A.
 C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S.G.,
 Podar, M., Martin, H.G., Kunin, V., Dalevi, D.,
 Madejska, J., Kirton, E., Platt, D., Szeto, E., Salamov, A.,
 Barry, K., Mikhailova, N., Kyrpides, N.C., Matson, E.G.,
 Ottesen, E.A., Zhang, X., Hernández, M., Murillo, C.,
 Acosta, L.G., Rigoutsos, I., Tamayo, G., Green, B.D.,
 Chang, C., Rubin, E.M., Mathur, E.J., Robertson, D.E.,
 Hugenholtz, P. & Leadbetter, P. (2007) Metagenomic and
 functional analysis of hindgut microbiota of a wood-feeding
 higher termite. *Nature* 450(7169), 560–565.
- Watanabe, H. & Tokuda, G. (2001) Animal cellulases. *Cellular and
 Molecular Life Sciences* 58(9), 1167–1178.
- Weil, T., Rehli, M. & Korb, J. (2007) Molecular basis for the
 reproductive division of labour in a lower termite. *BMC
 Genomics* 8, 198.
- Wilson, G.A., Bertrand, N., Patel, Y., Hughes, J.B., Feil, E.J. &
 Field, D. (2005) Orphans as taxonomically restricted
 and ecologically important genes. *Microbiology* 151, 2499–
 2501.
- Wu-Scharf, D., Scharf, M.E., Pittendrigh, B.R. & Bennett, G.W.
 (2003) Expressed sequence tags from a polyphenic
Reticulitermes flavipes (Isoptera: Rhinotermitidae) cDNA
 library. *Sociobiology* 41(2), 479–490.
- Zhou, X., Oi, F.M. & Scharf, M.E. (2006) Social exploitation of
 hexamerin: RNAi reveals a major caste-regulatory factor in
 termites. *Proceedings of the National Academy of Sciences of the
 United States of America* 103(12), 4499–4504.
- Zhou, X., Smith, J.A.M., Oi, F.M., Koehler, P.G., Bennett, G.W. &
 Scharf, M.E. (2007) Correlation of cellulose gene expression
 and cellulolytic activity throughout the gut of the termite
Reticulitermes flavipes. *Gene* 395, 29–39.

ANEXO II – Sequências de todas as enzimas envolvidas na degradação da lignocelulose encontradas no pirosequenciamento.

Endo-Beta 1,4 Xilanase de simbionte de *Coptotermes gestroi*

```
1 atgtatcaagttactcagccctttgcaaacgggaagctaagtttc
  M Y Q V T Q P F A N G K L S F
46 ttaggaaatattattccatcaagcgttccatcaaactgggatact
  L G N I I P S S V P S N W D T
91 tattggaatcaagctacagcagagaatggatgcaagtggggaagt
  Y W N Q A T A E N G C K W G S
136 gttcaaagctcgcaaggatcatttaactggggacaatgtgataca
  V Q S S Q G S F N W G Q C D T
181 gcatataatcatgcaaaaacagcaggtattgcattcaagttccat
  A Y N H A K T A G I A F K F H
226 acattagtttggggaagccaagaaccaggatggattggatcattg
  T L V W G S Q E P G W I G S L
271 tcaggcaataactcaacaagaagcagttgtagctggatcaaagca
  S G N T Q Q E A V V S W I K A
316 gcaggacaaaaatattcaagtcagaattgcttgatgttgtaaat
  A G Q K Y S S A E L L D V V N
361 gaagcacttcatgcaccagcaagtttccgtaattcacttgggtgga
  E A L H A P A S F R N S L G G
406 tcaggaagcacaggatgggattggattgtttggatcatttaaacag
  S G S T G W D W I V W S F K Q
451 gcaaaacaatcatttccaagctcaaaacttttaaatcaatgaatat
  A K Q S F P S S K L L I N E Y
496 ggtatcatcagtgatccaagtgaagctcgtcaatatgttgaaatt
  G I I S D P S E A R Q Y V E I
541 attgatattttgaagagcaatagcttaattgatggatttggaatt
  I D I L K S N S L I D G I G I
586 cagtgccatcaattcaacgtcaacacagtcctcagcatcaacagcc
  Q C H Q F N V N T V S A S T A
631 aaatcagtacttgatacacttggagcaacaggtatttcaatctat
  K S V L D T L G A T G I S I Y
676 gtttcagaatttgatgcaaatggaaatacagaacaagaacaacaa
  V S E F D A N G N T E Q E Q Q
721 gatattttatgcaagagtttttccagttccttgggagcatagttca
  D I Y A R V F P V L W E H S S
766 gttaaaggaatcactcttggggttacatcacaggtcaaactatgg
  V K G I T L W G Y I T G Q T W
811 aaagatggaacaggtattgttgattcaagcggaatgaacgaggt
  K D G T G I V D S S G N E R G
856 gcaatagaatggctcaagcaatacatgtcatctagttaa 894
  A I E W L K Q Y M S S S *
```

β-Celobihidrolases de simbionte de *Coptotermes gestroi*

```

1 atgttttgttctcgtttttcttggttaaacacattcgcttggaaca
  M F V L V F L G L T H S L G T
46 ggaactaaccaagcagagagccatccatcactttcatggcagaac
  G T N Q A E S H P S L S W Q N
91 tgtaaatcaggaggttcatgtacacaaacatctggttcaattgta
  C K S G G S C T Q T S G S I V
136 ttagattcaaaactggagatggacacatgatgattcattaacaaat
  L D S N W R W T H D D S L T N
181 tgttatgatggtaatgaatggtcatcatcactttgtccagatcca
  C Y D G N E W S S S L C P D P
226 aaaacatgttcaaataattgtgtcattgatggagcagattattca
  K T C S N N C V I D G A D Y S
271 ggaacatatggaatcacaacaagtggaaactcacttaaattagtt
  G T Y G I T T S G N S L K L V
316 tttgttacacatggatcgtattcaacaaacattggatcaagagtt
  F V T H G S Y S T N I G S R V
361 tatcttcttaaggatgaaacacattatcagatttttgatttaaag
  Y L L K D E T H Y Q I F D L K
406 aacaaagaattcacatttacagttgatgattctaattcttgactgt
  N K E F T F T V D D S N L D C
451 ngattaaatggtgcactttactttgtttcaatggacgaagatggt
  G L N G A L Y F V S M D E D G
496 ggaacatcaaaatattcaatcaacaagcaggagccaaatatgga
  G T S K Y S I N K A G A K Y G
541 acaggatattgtgatgcacaatgccacatgatattaaattcatc
  T G Y C D A Q C P H D I K F I
586 aatggagaagctaattgttgaattgaaaccacaaacaaatgat
  N G E A N V E N W K P Q T N D
631 gaaaatgcaggcaatggacgatatggagcatgctgcacagaaatg
  E N A G N G R Y G A C C T E M
676 gatatttgggaagctaacaatatgcaacagcatcacaccacac
  D I W E A N K Y A T A Y T P H
721 atttgtacagtgaatggtgaatacagatgcatggaacagaatgt
  I C T V N G E Y R C D G T E C
766 ggagatactgattcaggaaatagatatggaggtgtttgtgataaa
  G D T D S G N R Y G G V C D K
811 gttacagtgtgatttcaattacacagaatgggaaatacaagcttc
  D G C D F N S Y R M G N T S F
856 tggggaccaggattaatcattgatactggaagccagttacagtt
  W G P G L I I D T G K P V T V
901 gttacacaatttgtcacaaaagatggatcagataatggacaatta
  V T Q F V T K D G S D N G Q L
946 tctgaaatcagaagaaaatatgtccaaggtggaaaagttattgaa
  S E I R R K Y V Q G G K V I E
991 aacactgttgttaacattgcaggtatgagttcagtcattcagtt
  N T V V N I A G M S S V N S V
1036 tcagatgatttctgcaatgatcaaaagacagcatttggtgatact
  S D D F C N D Q K T A F G D T
1081 aatgattttgagaagaaaggaggttggcaggacttgaaaagca
  N D F E K K G G L A G L G K A
1126 tttgattatggaatgttctgtttgtcactttgggatgatcat
  F D Y G M V L V L S L W D D H
1171 caggtcaacatgcttggcttgattcaatctatccaactgataaa
  Q V N M L W L D S I Y P T D K
1216 ccagcaagtgaaccaggtgttaagagaggaccttgtgcaacatca
  P A S E P G V K R G P C A T S
1261 tctggagttccatctgatgttgaaagtcagcatcctgatgcact
  S G V P S D V E S Q H P D A S
1306 gtcacattctctgacatcagatttggaccaattgattcaacctac
  V T F S D I R F G P I D S T Y
1351 taa 1353
  *
```

β -Glicosidase de *Coptotermes gestroi*

1 atgaggttccaaacgctttgcctcgtcgtctttgtgacgacagta
M R F Q T L C L V V F V T T V
46 ttccggagatgacgtcgataacgagacccttggttacgttttcctgac
F G D D V D N E T L V T F P D
91 gattttaagttaggtgccgctacagcttcataaccagattgaagga
D F K L G A A T A S Y Q I E G
136 ggatgggatgacggacgaaaggggtcccaacatctgggacactctg
G W D A D G K G P N I W D T L
181 acacacgagcgccacacttagtggtgatcggtcaacaggtgac
T H E R P H L V V D R S T G D
226 gtggcgatgattcgtatcacctgtacctggaagacgttcaactt
V A D D S Y H L Y L E D V Q L
271 ctgaaggacatggggcagaagtttatcgcttctccgtatcttgg
L K D M G A E V Y R F S V S W
316 gctcgcatcctgctgaaggacacgataatagagtgaaccaagcg
A R I L P E G H D N R V N Q A
361 ggcattgagtactataacaaactcatagacgcacttctacagaat
G I E Y Y N K L I D A L L Q N
406 ggaatagagccgatgggttactatgtatcactgggacttaccacag
G I E P M V T M Y H W D L P Q
451 aaactccaagacctaggaggatggccaaataggatattggccaaa
K L Q D L G G W P N R I L A K
496 tacgccgagaattacgcccgcgttttgttagtaactttggggac
Y A E N Y A R V L F S N F G D
541 aggggtcaaactgtggctcaccttcaacgaacctctgacgttatg
R V K L W L T F N E P L T F M
586 gatgcatatgcatctgacacaggaatggctccatcaatcaacaca
D A Y A S D T G M A P S I N T
631 cccggtatcggcgactacctcacggcacatactgtaatccttgcc
P G I G D Y L T A H T V I L A
676 catgccaatatctaccgtttgtatgagaggggaattcagagaggaa
H A N I Y R L Y E R E F R E E
721 cagcaaggacaggttggtattgcactcaatatccactgggtgag
Q Q G Q V G I A L N I H W C E
766 ccggagactggttcgccccaaagacgttgaggcttgtgaaaggtac
P E T G S P K D V E A C E R Y
811 caacagttcaatctgggaatatacgtcatcccatcctcagtga
Q Q F N L G I Y A H P I L S E
856 aacggcgactacccagtggtttgaaaacgagggtagacgcaaac
N G D Y P S V L K T R V D A N
901 agcgcatcggaaggctactcaacatcgcgctaccgaagttcact
S A S E G Y S T S R L P K F T
946 ccagaggaagtagctttcgtcaatggaacatatgatttcctgggt
P E E V A F V N G T Y D F L G
991 ttgaattttctacactgctgtagtcgggagagacggagttgaaggg
L N F Y T A V V G R D G V E G
1036 gaaccgccttcacggtacagagacatgggacgataatcacag
E P P S R Y R D M G A I I S Q
1081 gacccggagtggtggccgagtcgtcctcttcaggtcagagttgta
D P E W P E S A S S W L R V V
1126 ccgtggggattccgcaaggaaactcaactggatcgccaacgaatac
P W G F R K E L N W I A N E Y
1171 ggtaacccccctatatattcataactgaaaacggtttctcgactac
G N P P I F I T E N G F S D Y
1216 ggtggagtcaatgacactaacagagtactttactacactgaacat
G G V N D T N R V L Y Y T E H
1261 ttaaaggaaatgctgaaggcaattcacatagacggagttaacgta
L K E M L K A I H I D G V N V
1306 attggatatacagcttgagcctcatagacaacttcgaatggttg
I G Y T A W S L I D N F E W L
1351 cgaggatatactgagaggttcggtatacatgaagtgaatttcac
R G Y T E R F G I H E V N F I
1396 gacccaagtcgcccacgaactcccaagggaatcgccaggggtactc
D P S R P R T P K E S A R V L
1441 acagagatatctcaacaagacagattccagagcgcttccgagac
T E I F S T R Q I P E R F R D
1486 taa 1488

Laccase de *Coptotermes gestroi*

```

1  atgtggctctgcgtcctgttgctgcgcaattggaattgcttct
M W L C V L L A C A I G I A S
46  gcagcgacatcagtgctgctgaattcatatcttcagcccaacgat
A A T S V L L N S Y L Q P N D
91  gacattgatagaaaacgcgtacctcctaaaaagcacagggcagcaac
D I D R N A Y L L K A Q G S N
136  tgtgcccgtatgtgcaacgcaacagaggcacccaaaatctgctac
C A R M C N A T E A P K I C Y
181  taccaatggacaattgagaactacgtaacagtggtcaggtgcgtgt
Y Q W T I E N Y V T V S G A C
226  ggcaactgtcccgtgaatgcgacagcctgttacacccacagtcg
G N C P V N A T A C Y N P Q C
271  attacagccgatggatatgagcgagctattctttcggtaaacagg
I T A D G Y E R S I L S V N R
316  aagttaccgggaccttccattgaggtgtgcctcagagacagagta
K L P G P S I E V C L R D R V
361  atttgggatataaccaacaacatggctgggaggactactgccatc
I V D I T N N M A G R T T A I
406  cactggcatgggttatttcagaaagggtcccagtaacatggacggc
H W H G L F Q K G S Q Y M D G
451  gttcccattgtaacccagtgctctatacatgaggggtgacacattt
V P M V T Q C S I H E G D T F
496  cggtagcactttatcgctaacaacgagggaactcatttctggcac
R Y D F I A N N E G T H F W H
541  tcccagacggtttgcagaagctcgatggcttggcaggttaacttgg
S H D V C R S S M A W Q V T W
586  tggttagggtatgacgactcgaacggaaaaatgtacgattttgac
W L G Y D D S N G K M Y D F D
631  ctaccagaacacaaaattttcatcgcgactggctacatcttccc
L P E H K I F I G D W L H L P
676  gcagatgaccattttcccggacttccaggggcaaatgcagagacaa
A D D H F P G L R G A N A G Q
721  gatgctgactcctttctcattaacggaagaggaaatccttgatc
D A D S F L I N G R G N T L I
766  ggaactctgtccaccaacacaccgtatgcacagataaaggtgcag
G T L S T N T P Y A Q I K V Q
811  tgggctgagggtaccggcttcgattgtaggagccttgtgcact
W G R R Y R L R I V G A L C T
856  gtgtgccccacacagctcaccattgatgggcacaaaatgaccgtc
V C P T Q L T I D G H K M T V
901  atagccaccgacggcaaccctgtggctcctgccaggatcgactcc
I A T D G N P V A P A R I D S
946  cttatcatttactctggtgaaagatatgatgtcgtgtagaagca
L I I Y S G E R Y D V V L E A
991  actaacacggaaggatcttactggatccacttaaaaggctcgcgc
T N T E G S Y W I H L K G L A
1036  acttgtgttgcaaccagagtttaccagctgggagtgtgcagtat
T C V A T R V Y Q L G V L Q Y
1081  gaaaatacagcaaccaataacctccgtgccctgaccctgacca
E N T A T N N L R A L T P D P
1126  gggtagcagcgattcccgcgaccagcagcctaccgcgtcctgaac
G Y D G F P R P A A Y R V L N
1171  ccagagaacgcaagctgcagcgccggttcgacaggttatgcgtc
P E N A S C S A G S T G L C V
1216  acgcaactggcgaactcggaacccgtgccacgggacatcctaacc
T Q L A N S D P V P R D I L T
1261  cagctcccggaggttaactatcttctcaacttttggatttga 1302
Q L P D V N Y L L N F W I *
```

Endo-β- 1,4 - Glicosidase de *Coptotermes gestroi*

```

1  atgaggggtcttcggtttgccttctgtcggcgctcgcgctttgcca
M R V F V C L L S A L A L C Q
46 gctgcttacgactacaagacagtactgaagaattcgctgcttttc
A A Y D Y K T V L K N S L L F
91 tacgagggctcagcgatcgggacaattgcccgctgatcagaaggctc
Y E A Q R S G Q L P A D Q K V
136 acgtggaggaaggattccgcccttaacgacaagggaacaacggc
T W R K D S A L N D K G N N G
181 gaggatctgacaggggatattatgacgctggtgattttgtgaaa
E D L T G G Y Y D A G D F V K
226 ttcggttttccaatgcggtacacggccaccgctcctggcttggggc
F G F P M A Y T A T V L A W G
271 ctggtggactatgaagcgggctactcctcagcaggcgctctggat
L V D Y E A G Y S S A G A L D
316 gacggtcgcaaagctgttaaatgggccacggactacttctcaag
D G R K A V K W A T D Y F L K
361 gccacacggccgcctcagaattctacggacaagtgggtcaggga
A H T A A S E F Y G Q V G Q G
406 gatgtggaccacgcctactggggacgtccagaagacatgacg
D V D H A Y W G R P E D M T M
451 gacagacctgcctacaagatcgacgcggcgacccagggtcggac
D R P A Y K I D A A H P G S D
496 ctgcgcggcgagacagccgcgccttagctgcagcttccatcggtg
L A G E T A A A L A A A S I V
541 ttccaggggtgtcgactcctcttattccgaaaagctgctcgaccgc
F R G V D S S Y S E K L L D R
586 gccaaagcagcttttcgacttcgcccagacaagaatcgcggaatac
A K Q L F D F A D K N R G K Y
631 agtgattcaatcaccgacgcgaggaatttctacgcgtccggagac
S D S I T D A R N F Y A S G D
676 tacaagatgagttagtagtgggcagccacatggctctacagggcg
Y K D E L V W A A T W L Y R A
721 accaagcagcagcgcctatctaactaaagctgagtcgctgtacaac
T N D D A Y L T K A E S L Y N
766 gaattcggcctcggaattggaacggtgcccttaactgggataac
E F G L G N W N G A L N W D N
811 aaagtatccggtgtacaggttctgctggccagactcacaaaaaag
K V S G V Q V L L A R L T K K
856 caggaatacagggacacgggtacaaggctacgtggattacttgatt
Q E Y R D T V Q G Y V D Y L I
901 aacaatcagcagaagacgcccaggcctcctatacatcgaccag
N N Q Q K T P K G L L Y I D Q
946 tggggcaccttgcgacatgctgccaatgcggtttaattatacta
W G T L R H A A N A A L I I L
991 caggcagcagacctgggtatcagtgtgacagttatcgccagttt
Q A A D L G I S A D S Y R Q F
1036 gccaggaagcagatcgattacgcattgggtgacggtggtcgcagc
A R K Q I D Y A L G D G G R S
1081 tacgtcgtaggatttggcgataaaccacctgtacgccctcaccat
Y V V G F G D N P P V R P H H
1126 agatccagttcatgccctgacgcaccagcagtatgtgactggaac
R S S S C P D A P A V C D W N
1171 acgttcaacagcccggacccaaatttccacgtactcacggagct
T F N S P D P N F H V L T G A
1216 ttggtgggaggtccagatcagaatgacaactacgaggacgatcgc
L V G G P D Q N D N Y E D D R
1261 agtgactacatttccaacgaggtagcgacggtattataacgcaggc
S D Y I S N E V A T D Y N A G
1306 ttccaatcagctgtcgccgctctcgtcacgttggcggttaa 1347
F Q S A V A A L V T L G V *

```

Endo-Beta-1,3 - Glicosidase de *Coptotermes gestroi*

```

1 atggccagaatatttgttcgtatgagcgacggttctgtcgatc
  M A R I F V R M S A T V L S I
46 ctgtccctatgccaggcaggatcagcagagacacttgtctggcag
  L S L C Q A G S A E T L V W Q
91 gacgagttcgacacttttgatctgaacaattggaatcatctcgtc
  D E F D T F D L N N W N H L V
136 acagcttgggggggaggcaatcaggaatttcagtactaccgcaat
  T A W G G G N Q E F Q Y Y R N
181 gacaggaggaacagctacgtaagggaacggaattttatacctcaga
  D R R N S Y V R D G I L Y L R
226 ccgacgtggacttctgctgaatatggagatgacttctgttacagt
  P T W T S A E Y G D D F L Y S
271 ggcagtctctcatttccgggttgcaacatggagccctgtacatca
  G S L S F P G C N M E P C T S
316 tctgtcggaaatgacattgtacatccactgctgtccgcaaggata
  S A G N D I V H P L L S A R I
361 accagcactttcgcttcaagtacggcagggtagaagtttagagct
  T S T F A F K Y G R V E V R A
406 aaagtccgagaggagattggatttggcctgcgatctggatgcta
  K V P R G D W I W P A I W M L
451 ccgagagactctgtctatgggtggctggccccgatctggtgaaatt
  P R D S V Y G G W P R S G E I
496 gatatcatggaaagtaaaggaaatgaaaattatgcagatggaaac
  D I M E S K G N E N Y A D G N
541 ggtgtgcaagtgggtaacactttggtggcttccactttacactgg
  G V Q V G N T L V A S T L H W
586 ggtccagatccggaccacaataactactggaggacacactgggaa
  G P D P D H N N Y W R T H W E
631 aggaacattcaaagttctggtcatagtttcgctgacgacttccat
  R N I Q S S G H S F A D D F H
676 ttgtatggactgacatggaccgacaatcatattatcttactgtc
  L Y G L T W T D N H I I F T V
721 gataatgaacaaattggggacgtatgggctcccgaaaatggcttc
  D N E Q I G D V W A P E N G F
766 tggactacggaggtttccaaaataatccaggtggtaccaacatt
  W Y Y G G F Q N N P G G T N I
811 tggcaagatggaaactggatggcaccatttgaccgagagttccag
  W Q D G N W M A P F D R E F Q
856 ataatcctgaatgtggcgtgggggcactttcttccccgacaac
  I I L N V A V G G T F F P D N
901 attggttaaccgacatggagctgggacggccatcccatagagat
  I G N R P W S W D G H P M R D
946 ttctgggagcggcgacgagtggtaccgacgtggcacgaagaa
  F W E R R S E W L P T W H E E
991 gaagcagccctccagattgattacgtcagggtttatcaggactaa 1035
  E A A L Q I D Y V R V Y Q D *

```