



Universidade de Brasília

Departamento de Biologia Celular

Programa de Pós-graduação em Biologia Molecular

Caracterização de fatores que afetam a detecção de presas através do sequenciamento de alto desempenho do DNA

Renata Velôzo Timbó

Orientador:

Dra. Débora Pires Paula (Embrapa Recursos Genéticos e Biotecnologia - Cenargen)

Co-orientadores:

Dr. Roberto C. Togawa (Cenargen)

Dr. David Andow (Universidade de Minnesota, EUA)

Dr. Marcelo Brígido (UnB)



Universidade de Brasília
Departamento de Biologia Celular
Programa de pós-graduação em Biologia Molecular

Caracterização de fatores que afetam a detecção de presas através do sequenciamento de alto desempenho do DNA

Renata Velôzo Timbó

Tese apresentada ao Departamento de Biologia Celular do Instituto de Ciências Biológicas da Universidade de Brasília – UnB, como requisito à obtenção do grau de Doutor em Biologia Molecular.

“As leis são sempre úteis aos que têm posses e nocivas aos que nada têm.”

Jean-Jacques Rousseau

AGRADECIMENTOS

A Deus por permitir que tivesse saúde para concluir mais essa etapa!

Aos meus **pais** *María Darque Velôzo* e *Francisco Parente Tímbo* que sempre acreditaram na minha capacidade e sempre puxaram minha orelha quando percebiam que o desânimo estava me atormentando!

Ao meu **irmão** *Juan Velôzo Tímbo* (meu Ozzie favorito), que se tornou um exemplo de batalhador e, que mesmo distante, continua a torcer pela irmã mais nova!!!

À minha **família** que ao brincarem sobre minhas filhas joaninhas, sempre levou na brincadeira o que era sério e assim, me deixavam um pouco mais calma durante esse período.

À minha **vó** *Gerarda "vó Adá"* que demonstrou muito orgulho e felicidade ao saber que sua neta estaria concluindo mais esta etapa, mas que infelizmente, Deus a levou antes de poder ver tal conclusão. Assim como Ele fez com minha **tia** *Gercina (Bolhota)*, que não pôde presenciar a conclusão da minha graduação, algo que a deixava super animada ao me ver seguindo a área da saúde, como ela!

À minha **orientadora** *Débora Pires Paula*, por sempre se mostrar uma pessoa maravilhosa seja como orientadora ou como amiga. Sempre acreditando no meu potencial, me ouvindo nos momentos de tristeza, me dando forças para continuar nessa batalha e, me ensinando a nunca desistir! Tenho como meta ser ao menos a metade da pessoa que ela é, pois só com a metade já me tornaria um ser humano melhor!

Ao meu **orientador** *Roberto Togawa*, que de repente se viu com uma nova aluna inexperiente na área de Bioinformática e mesmo assim, me aceitou de braços abertos. Posso dizer que se existe uma versão masculina da minha orientadora Débora, essa versão é você!

Ao meu **orientador** *David Andow*, que apesar da dificuldade da comunicação (esses Americanos :-P), demonstrou uma paciência e um zelo quanto ao trabalho que me fazia acreditar que eu seria mesmo capaz de fazer e concluir mais essa etapa! E assim como os demais orientadores, demonstrou uma paixão sem fim em tudo e com tudo que trabalhava.

Aos meus **amigos**, que de uma forma ou de outra sempre me deram uma ajuda como coletar joaninhas e pulgões no campo quando eu não podia (*Lucas Machado*

de Souza e Bruna Lima), e à minha amiga *Karolina Torenzani* das Abelhas que sempre me presenteou com conversas descontraídas e animadoras.

Às minhas **amigas** lindas e “delicadas” da UnB (*Luciana Freire, Mariana Neiva, Thais Oliveira, Carla Patricia, Tauana Ferreira, Larissa Rodrigues*) que me aguentaram nos meus dias mais estressantes e porque não os mais tristes. Me deram colo e abraços que eu ainda não sei se os merecia mas que sou grata demais! E por sempre me darem motivos para sorrir e continuar batalhando, são mais que amigas, são irmãs, filhas, mães.... Tudo de bom! E que com certeza estarão comigo por toda vida!

À minha **amiga** Pretinha *Greice Maria Rodrigues de Souza Garcia*, que a cada dia me mostra como batalhar e seguir em frente. Estive ao lado dela durante todo o processo de doutorado e eu muito brincava dizendo a ela: “quer ser doutora amiga? Então aguenta”. E hoje, chegou o momento em que ela finalmente as repetiu para mim com aquele sorriso. Amiga, obrigada pelo apoio e por tudo que me ensinou como pessoa!

À minha **amiga** linda e “baixinha” *Iracema Sanches*, que em tão pouco tempo conseguiu conquistar um grande espaço em meu coração, compartilhou suas risadas, riu das minhas piadas podres (rs), ouviu meus desabafos e sempre tinha um tempinho para jogar conversa fora, aliviando assim grandes momentos de estresse e desespero que sempre acompanham esse momento de conclusão de tese. Obrigada por ser uma irmã, conselheira, brigona, correta, justa e por mostrar tanto cuidado às pessoas que você gosta, incluindo eu, lógico ☺!

Ao *Helton Moreira de Andrade* por não me deixar desistir nos momentos de desânimo e ser parte da força impulsora que me moveu e permitiu alcançar esse objetivo. Sem esquecer é claro, das horas que me acompanhou e auxiliou em vários dos bioensaios que aconteciam aos fins de semana e no virar da noite.

À Dra. *Brigida Sousa* da Universidade Federal de Lavras, que atendeu ao pedido do grande **amigo** *Lucas Machado* e nos cedeu alguns ovos de *Chrysoperla externa* e assim pudemos reiniciar nossa colônia!

À **Embrapa** que nos cedeu espaço e tudo o que foi necessário para realização dessa pesquisa, um local que irá ficar sempre em minha memória. Foram 8 anos inesquecíveis!

À **Universidade de Minnesota** que ajudou no financiamento do sequenciamento.

À **UnB** que possibilitou mais essa etapa, assim como a **FAPDF** forneceu auxílio com financiamento para participação em Congressos.

À **CAPES** pelo fornecimento da bolsa.

Os meus mais sinceros agradecimentos por tudo que me proporcionaram nesse período, que foi cheio de alegrias, risos, choros, desespero, mas que graças a cada um de vocês, que de uma forma ou de outra somaram nessa caminhada, hoje estou aqui finalizando mais esta etapa!

Sumário

Lista de Figura	ii
Lista de Quadros	iii
Lista de Tabelas	iv
Abreviações	v
1. INTRODUÇÃO	1
2. REVISÃO BIBLIOGRÁFICA	5
2.1 IMPORTÂNCIA DOS INSETOS PREDADORES E DA IDENTIFICAÇÃO DE SUAS PRESAS	5
2.2 PREDADORES E PRESAS	6
2.3 DETECÇÃO DE PRESAS POR MEIO DE MÉTODOS MOLECULARES	14
2.4 FATORES QUE AFETAM A DETECÇÃO DA PRESA	29
3. OBJETIVOS	35
4. MATERIAL E MÉTODOS	36
4.1 CRIAÇÃO DOS INSETOS	36
4.2 BIOENSAIOS DE ALIMENTAÇÃO	39
4.3 CONSTRUÇÃO DO BANCO DE REFERÊNCIA DE DNA MITOCONDRIAL	45
4.4 DETECÇÃO DAS PRESAS NO CONTEÚDO GASTROINTESTINAL DOS PREDADORES	47
4.5 ANÁLISES ESTATÍSTICAS	51
5. RESULTADOS E DISCUSSÃO	59
5.1 ELUCIDAÇÃO DOS MITOGENOMAS	59
5.2 BIOENSAIOS DE ALIMENTAÇÃO	64
5.3 DETECÇÃO DAS PRESAS EM FUNÇÃO DAS ESPÉCIES DE PREDADOR-PRESA EM INTERAÇÃO, MODO DE ALIMENTAÇÃO E FASE DO CICLO DE VIDA DO PREDADOR.....	65
5.4 DETECÇÃO DAS PRESAS EM FUNÇÃO DO NÚMERO DE PRESA CONSUMIDA.....	82
5.5 DETECÇÃO DA PRESA EM PREDACÃO SECUNDÁRIA EM FUNÇÃO DO NÚMERO DE PRESAS CONSUMIDAS E TEMPO PÓS-PREDACÃO	88
5.6 DETECÇÃO DA PRESA EM FUNÇÃO DA TEMPERATURA AMBIENTE DE DIGESTÃO DA PRESA.....	101
6. CONCLUSÃO	109
BIBLIOGRAFIA	111
ANEXO 1 – Artigo publicado	125
ANEXO 2 – Ensaio piloto para preparação para os bioensaios de alimentação	142
ANEXO 3 – Construção do banco de referência de DNA mitocondrial.....	147
ANEXO 4 – Análises de bioninformática para detecção das presas	153
ANEXO 5 – Códigos SRA dos dados de sequenciamento das bibliotecas dos bioensaios A, B, C e D depositados no <i>GenBank</i>	156
ANEXO 6 – Segundo manuscrito submetido para publicação: Análises para seleção dos parâmetros de filtragem de falsos-positivos.....	159
ANEXO 7 – Tabelas e figuras suplementares.....	197

Lista de Figura

- Figura 1.** Cycloneda sanguinea. A - Adulto (Foto de: Ron Hemberger); B - Larva. (Foto de: Sammy Ekker). 7
- Figura 2.** Hippodamia convergens. A - Adulto; B - Larva (Fotos de: Luis F. Aristizábal). 8
- Figura 3.** Harmonia axyridis. A - Adultos em acasalamento; B - Larva, apresentando as cores características (Fotos de: Sam Bal). 9
- Figura 4.** Chrysoperla externa. A - Adulto; B - Larva (fase predadora) (Fotos de: Harvey Schmidt). 10
- Figura 5.** Pulgão Myzus persicae. Adulto áptero (Foto de: E. B. Radcliffe). 11
- Figura 6.** Mosca-branca Bemisia tabaci. A - Adulto; B - Fase larval 12
- Figura 7.** Spodoptera frugiperda. A - Ovos; B - Lagarta. 13
- Figura 8.** Preparo da dieta artificial para os adultos de Ch. externa. A - Aspecto obtido da mistura do levedo de cevada e do mel. B - Gaiola para manter os adultos onde a parte superior apresenta-se vedada por uma tela e a parte inferior é fechada com tecido e liga. 37
- Figura 9.** Aparelho gastrointestinal dos predadores dissecados, visualizados em aumento de 12 ou 16x em lupa estereoscópica. A - De larva de 3º ínstar da joaninha C. sanguinea; B - De larva de 3º ínstar do crisopídeo Ch. externa; Como etapa final, as cabeças foram removidas antes da armazenagem dos gastrointestino. 48
- Figura 10.** Mitogenomas elucidados e anotados: A - Bemisia tabaci; B - Chrysoperla externa; C - Cycloneda sanguinea; D - Hippodamia convergens; E - Myzus persicae; F - Spodoptera frugiperda; G - Harmonia axyridis. 63
- Figura 11.** Número inicial de reads das presas detectadas no conteúdo gastrointestinal dos predadores normalizado pela cobertura média da biblioteca ($6,2 \times 10^6$ reads) com 95% IC. Cs = C. sanguinea; Ce = Ch. externa; L = larva; A = adulto. 75
- Figura 12.** Taxa de decaimento da detecção das reads para ovos de S. frugiperda e adultos ápteros de M. persicae no conteúdo gastrointestinal dos predadores, normalizado para a cobertura média da biblioteca ($6,2 \times 10^6$ reads) com 95% IC. Cs = C. sanguinea; Ce = Ch. externa; L = Larva; A = Adulto. 77
- Figura 13.** Taxas de decaimento das reads presentes no conteúdo gastrointestinal dos três predadores: (A) Ovo em pré-eclosão de Spodoptera frugiperda e (B) Adulto áptero de Myzus persicae. Linhas representam dados da regressão linear. 80
- Figura 14.** Relação entre o número de reads observadas da presa e métodos de regressão. Linha preta: regressão não linear sobre dados não transformados com nível de confiança de 95%. Linha azul: regressão linear em dados ln-transformados. A) Uma presa de M. persicae; B) Três presas de M. persicae; C) Seis presas de M. persicae. 85
- Figura 15.** Taxa de decaimento das reads das presa M. persicae no conteúdo gastrointestinal de Hi. convergens, de acordo com o número de presas consumidas. Linha pontilhada: 1 M. persicae consumido. Linha tracejada: 3 M. persicae consumidos. Linha sólida: 6 M. persicae consumidos. 86
- Figura 16.** Número de reads de M. persicae detectado em H. axyridis após predação primária ou secundária com IC de 95%. 94
- Figura 17.** Decaimento das reads da presa secundária (M. persicae) em H. axyridis em função da predação primária (linha preta sólida, círculos pretos sólidos) ou predação secundária (linha tracejada, círculos abertos) e tempo pós-predação por H. axyridis. O melhor modelo está apresentando com linha cinza sólida. Círculos observados foram normalizados com IC 95%. 95

- Figura 18.** Decaimento das reads da presa secundária (*M. persicae*) no gastrointestinal da presa primária (*Ch. externa*), em função do período de digestão do predador intraguildda (*H. axyridis*) e o tempo pós-predação por *Ch. externa*. Linhas representam os melhores modelos e os círculos representam a normalização a 95% de IC. Linha pontilhada e círculos abertos: 0 h pós-predação por *H. axyridis*; Linha tracejada e círculos cinza: 3 h pós-predação por *H. axyridis*; Linha sólida e círculos negros: 6 h pós-predação por *H. axyridis*..... 96
- Figura 19.** Decaimento das reads da presa secundária (*M. persicae*) no predador intraguildda (*H. axyridis*), em função do tempo de digestão da presa primária (*Ch. externa*) e do tempo pós-predação por *H. axyridis*. Linhas representam os melhores modelos e os círculos representam a normalização com 95% de ICs. Linha pontilhada com círculos abertos: 0 h pós-predação por *Ch. externa*; Linha tracejada com círculos cinza: 3 h pós-predação por *Ch. externa*; Linha sólida com círculos pretos: 6 h pós-predação por *Ch. externa*..... 97
- Figura 20.** Decaimento das reads da presa secundária (*M. persicae*) na presa primária (*Ch. externa*) e no predador intraguildda (*H. axyridis*), em função do tempo total pós-predação pela *Ch. externa* e por *H. axyridis*. Linhas apresentam os melhores modelos e círculos representam a normalização com 95% de IC. Círculos abertos: 0 h pós-predação por *Ch. externa* e 0 – 6 h pós-predação por *H. axyridis*. Círculos cinza: 3 h pós-predação por *Ch. externa* e 0 – 6 h pós-predação por *H. axyridis*. Círculos pretos: 6 h pós-predação por *Ch. externa* e 0 – 6 h pós-predação por *H. axyridis*..... 99
- Figura 21.** Decaimento das reads das presas primárias (*Ch. externa*) no predador intraguildda (*H. axyridis*), em função da alimentação prévia da presa primária (*Ch. externa*) e do tempo pós-predação por *H. axyridis*. Linhas mostram o melhor modelo e os círculos apresentam os dados normalizados a 95% de ICs. Linha e círculos azuis: sem ingestão prévia de presa secundária. Linha pontilhada e círculos abertos: imediatamente pós-predação da presa secundária. Linha tracejada e círculos cinza: 3 h pós-predação da presa secundária. Linha sólida e círculos negros: 6 h pós-predação da presa secundária. 100
- Figura 22.** Correlação entre o número de reads da presa secundária (*M. persicae*) e da presa primária (*Ch. externa*) detectadas em *H. axyridis*. Círculos abertos: 0 h pós-predação da presa secundária por *Ch. externa*. Círculos cinza: 3 h pós-predação da presa secundária por *Ch. externa*. Círculos negros: 6 h pós-predação da presa secundária por *Ch. externa*..... 101
- Figura 23.** Relação entre o número de predadores agrupados por biblioteca e o erro padrão do número estimado de reads das presas com base na variação estimada entre o número de indivíduos utilizados na construção do pool. 105
- Figura 24.** Arena para teste de alimentação do crisopídeo: A) postura de *S. frugiperda*; B) *M. persicae*; C) Ninfa de *B. tabaci*; D) *Ch. externa*; X) Espaço vazio.....142
- Figura 25.** Ensaio piloto de predação secundária. A) Larva de *C. sanguinea* predando larva de *Ch. externa*. B) Larva de *Ch. externa* predando larva de *C. sanguinea*...145

Lista de Quadros

Quadro 1. Comparação da meia-vida de detecção (T_{50}) da presa entre diferentes técnicas moleculares, baseado em Greenstone <i>et al.</i> (2014).....	31
Quadro 2. Delineamento experimental do decaimento da detecção das presas em função das espécies de predador-presa em interação, modo de alimentação e fase do ciclo de vida do predador.	41
Quadro 3. Delineamento experimental do decaimento da detecção da presa em função do número consumido.	42
Quadro 4. Delineamento experimental do decaimento da detecção das presas em predação secundária em função do número de presas e tempo pós-predação.	
Quadro 5. Delineamento experimental do decaimento da detecção das presas em função da temperatura ambiente de digestão da presa.	45
Quadro 6. Mitogenomas utilizados como referência para a elucidação dos mitogenomas das espécies-alvo pelo programa MITObim (v.1.8).....	47

Lista de Tabelas

Tabela 1. Número de <i>reads</i> obtidas e remanescentes após controle de qualidade (CQ) das bibliotecas TruSeq Nano para sequenciamento por Illumina HiSeq2500 (250 pb, <i>paired-end</i> , tamanho do inserto de 550 pb, 1 <i>lane</i>) das amostras para elucidação dos mitogenomas dos insetos de interesse.	59
Tabela 2. Características dos mitogenomas elucidados.....	59
Tabela 3. Número de <i>reads</i> detectadas para predadores, presas e falsos-positivos no Bioensaio A.	65
Tabela 4. ANOVA da detecção inicial da <i>ln</i> -transformada das presas (n_0) em função do tipo de predador para cada presa e tipo de presa para cada predador.	75
Tabela 5. ANOVA para taxa de decaimento em função do tipo de predador para cada presa e cada tipo de presa para cada predador.	76
Tabela 6. Valores estimados dos parâmetros a 95% de intervalo de confiança (IC).....	78
Tabela 7. Parâmetros do melhor modelo para detecção inicial e decaimento das <i>reads</i> da presa nos três tipos de predadores.	82
Tabela 8. Número de <i>reads</i> detectadas para predadores, presas e falsos-positivos no Bioensaio B.	84
Tabela 9. Valores dos parâmetros para cada tratamento.	86
Tabela 10. Comparação dos modelos testando as hipóteses quantitativas com o número de <i>reads</i> da presa presentes no conteúdo gastrointestinal de <i>Hi. convergens</i>	87
Tabela 11. Parâmetros estimados para o Modelo 2.	88
Tabela 12. Número de <i>reads</i> detectadas para predadores, presas e falsos-positivos no Bioensaio C.	92
Tabela 13. Número de <i>reads</i> detectadas para os predadores, presas e falsos-positivos no Bioensaio D.	102
Tabela 14. Número médio das <i>reads</i> de presas por biblioteca com insertos de 350 pb e 550 pb.....	107

Tabela 15. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio A no <i>GenBank</i>	156
Tabela 16. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio B no <i>GenBank</i>	157
Tabela 17. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio C no <i>GenBank</i>	157
Tabela 18. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio D no <i>GenBank</i>	158
Tabela 19. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio D no <i>GenBank</i>	158

Abreviações

Ad – adulto
 AIC_c – Critério de Informação Akaike Corrigido
 BOD – câmara germinativa
 CeTime – período de tempo de *Harmonia axyridis*
 COX1 – Citocromo c oxidases
 CQ – Controle de qualidade (*Quality Control*)
d – taxa de decaimento
 DGGE – Gel de Poliacrilamida Gradiente Desnaturante
D_{max} – período de detectabilidade
 ELISA – Enzyme *Linked ImmunoSorbent Assay*
F – razão *Fisher* das médias quadráticas
 IC – intervalo de confiança
 L – larva
 QM – quadrado médio
 NGS – *Next Generation Sequencing*
 OTU – *Operational Taxonomic Unit*
 pb – pares de bases
 PCR – Reação em cadeia da polimerase
 ptn – genes codificadores de proteína
 qPCR – PCR quantitativo
reads – leituras de sequenciamento
 T0 – tempo zero ou imediatamente pós-predação
 T3 – 3 horas pós-predação
 T₅₀ – meia vida detecção
 T6 – 6 horas pós-predação
 UR – umidade relativa

Resumo

A identificação de presas consumidas por predadores constitui-se em importante área da Ecologia por possibilitar o estudo das interações intra e interespecíficas em um ecossistema. Esse conhecimento ajuda a avaliar seu funcionamento e como as interações mudam em resposta ao ambiente. Para muitos estudos de predação o ideal é determinar a biodiversidade de presas consumidas por um dado predador a campo, o que ainda não tem sido satisfatoriamente alcançado pelos métodos em voga de detecção de presas devido a limitações diversas. Muitos deles não conferem resolução taxonômica desejável, ou são dependentes do conhecimento prévio das espécies predadas para ser possível detectá-las através do uso de marcadores específicos, ou se baseiam no uso de PCR, o que é limitado à eficiência dos primers utilizados. Este trabalho visou avaliar a meia-vida de detecção, a sensibilidade (número mínimo de presa detectável), a especificidade (resolução taxonômica) e a variabilidade do método de sequenciamento de alto desempenho do DNA do conteúdo gastrointestinal de predadores para identificação de presas consumidas, o qual teoricamente tem potencial para detectar a biodiversidade de presas consumidas. Nesse método, não há etapa intermediária de amplificação de regiões barcodes das presas. As presas são prontamente identificadas através da detecção de fragmentos de seus DNA mitocondriais sequenciados, graças ao uso de banco de referência de DNA mitocondrial de várias espécies, incluindo aquelas com potencial de serem presas do predador em estudo. Adicionalmente, avaliou-se a relação de dependência da detecção da presa em função: das espécies de presa-predador em interação; da fase do ciclo de vida do predador; do seu modo de alimentação; do número de presas predadas; da predação ser primária ou secundária; da temperatura de digestão; do número de unidades experimentais (indivíduo-predador) utilizado por biblioteca de DNA; do tamanho do inserto (550 e 350 pb) utilizado nas bibliotecas. Quatro bioensaios de alimentação foram realizados utilizando como presas as espécies *Myzus persicae*, *Bemisia tabaci*, *Spodoptera frugiperda*, importantes pragas agrícolas, e como predadores *Cycloneda sanguinea*, *Harmonia axyridis*, *Hippodamia convergens*, *Chrysoperla externa*, importantes inimigos naturais de pragas agrícolas. Após controle de qualidade dos dados de sequenciamento e de análises de bioinformática (pipelines customizados) constatou-se que: houve detecção da ingestão de uma unidade experimental da presa pelo menos imediatamente após predação; houve detecção das presas com alta resolução taxonômica (nível de espécie), porém houve também detecção de falsos-positivos na maioria das bibliotecas e de falso-negativo nas bibliotecas de *B. tabaci*; o tempo pós-predação, as espécies de presa-predador em interação e a forma de alimentação do predador afetaram a detecção da presa; o período de detectabilidade variou em função das espécies de presa-predador em interação; houve detecção de predação secundária. Por outro lado, a temperatura de digestão da presa e o número de presas consumidas não afetaram a detecção da presa, no entanto quanto maior o número de presas consumidas maior foi o período de detectabilidade. Portanto, esse método é adequado para identificar a biodiversidade de predadores amostrados a campo, no entanto, devido a possibilidade de detecção de falsos positivos, recomenda-se confirmar o consumo de cada espécie detectada através de método complementar espécie-específico. Adicionalmente, recomenda-se também sempre usar controle negativo (predador sem alimentação) para verificação de potenciais falsos positivos.

Palavras-chave: controle biológico; inimigo natural; inseto predador; interações tróficas; metagenômica; mitogenoma; pragas da agricultura; teia alimentar.

Characterization of factors that affect prey detection through massively parallel DNA sequencing of predator gut contents

ABSTRACT

Identification of prey species is an important area of ecology because it enables the study of intra- and inter-specific interactions in an ecosystem. This knowledge supports the evaluation of ecosystem functioning and how species interactions change in response to the environment. For many field predation studies an ideal situation would be to determine the biodiversity of prey consumed by a given predator, which has not yet been satisfactorily achieved by prevailing methods of prey detection due to diverse limitations. Many of the prevailing methods do not confer desirable taxonomic resolution; or they are dependent on prior knowledge of the prey species to be able to detect them with specific markers; or they are based on PCR, which is limited by primer efficiency. This work aimed to evaluate the detectability half-life, sensitivity (minimum number of detectable prey), specificity (taxonomic resolution) and variability of prey detection (number of reads) using massively parallel DNA sequencing of the predator gut contents. This theoretically has the potential to detect the biodiversity of prey consumed. In this method, there is no intermediate stage of amplification of prey DNA barcodes by PCR. Prey are readily identified by detection of fragments of their mitochondrial DNA sequences, which are matched to a reference mitochondrial DNA database of species, including those likely to be preyed upon by the predator(s) under investigation. In addition, this study evaluated the dependence of prey detection as a function of: prey-predator species in interaction; predator stage; feeding mode; number of prey; primary or secondary predation; environmental temperature of prey digestion; number of experimental units (individual-predator) used per DNA library; insert size (350 and 550 bp) used in the libraries. Four feeding bioassays were performed using as prey species: *Myzus persicae*, *Bemisia tabaci*, and *Spodoptera frugiperda*, which are all important pests, and as predators: *Cycloneda sanguinea*, *Harmonia axyridis*, *Hippodamia convergens*, and *Chrysoperla externa*, which are important natural enemies of pests. After quality control of the sequencing data and bioinformatic analyses (using customized pipelines), the following were found: one prey individual can be detected, at least immediately after predation; false positives were detected in the majority of the libraries and *B. tabaci* was not detected in any of the libraries of the predator that fed on it (the sole false-negative case); elapsed time after consumption, prey-predator species in interaction and predator feeding mode affected prey detection; detectability period varied according to the prey-predator species in interaction; secondary predation was detected, although there was no discrimination of primary versus secondary predation without prior knowledge of the history of the predation events. On the other hand, temperature of prey digestion and number of prey consumed did not affect prey detection, although higher was the number of prey consumed, the longer was the detectability period. Therefore, this method is suitable to identify biodiversity of prey consumed by field-sampled predators. However, due to the possibility of detection of false positives, it is recommended to validate prey consumption using a complementary species-specific detection method (e.g., PCR). In addition, it is also recommended to include a library for negative control (predator in starvation) to check for the potential of false positives.

Keywords: agricultural pests; biological control; food web; insect predator; metagenomic; mitogenome; natural enemy; trophic interactions.

1. INTRODUÇÃO

A identificação de presas consumidas por predadores constitui-se em importante área da Ecologia por possibilitar o estudo das interações entre indivíduos da mesma espécie e de diferentes espécies em um ecossistema. Esse conhecimento ajuda a avaliar o funcionamento de um ecossistema e como as interações intra e interespecíficas mudam em resposta a alterações ambientais (Ings *et al.*, 2009).

Em se tratando de insetos predadores generalistas, a identificação das presas consumidas constitui-se em conhecimento fundamental para garantir o adequado manejo das espécies a fim de incrementar o controle biológico natural das pragas agrícolas na lavoura. Entretanto, a identificação das presas constitui-se em tarefa dificilmente viável pela observação direta do forrageamento e predação no campo. Isso ocorre em virtude de seu pequeno porte, hábitos crípticos, predação imprevisível e, com exceção de predadores sugadores, parasitóides e parasitas, não deixarem “cadáveres” de suas presas (Harper *et al.*, 2005). Adicionalmente, a identificação taxonômica a partir de partes do corpo das presas em digestão é extremamente trabalhosa e imprecisa (Sunderland *et al.*, 1987; Holopainen; Helenius, 1992; Triltsch, 1999; Harwood; Obryck, 2005).

Devido a essas dificuldades, frequentemente a inferência sobre o potencial de predação (ou mesmo de controle biológico) é indireta, pautada na mera co-ocorrência em campo das espécies (presas/pragas e predadores) em determinada área (Root 1973; Harmon e Andow, 2007) e de suas flutuações populacionais (Walde e Murdoch, 1988; Brook e Bradshaw, 2006) e, por isso, inacurada. Para as situações no campo, não raras vezes não se conhece as espécies de fato predadas e a frequência de predação das mesmas.

Nos últimos 50 anos diversos métodos moleculares (por exemplo: ELISA, PCR, isoenzimas) têm sido aplicados para a identificação mais precisa de componentes da

dieta de herbívoros e predadores, tanto no conteúdo gastrointestinal quanto nas fezes (Pendleton e Grundman, 1954; Murray e Solomon, 1978; Scrimgeour *et al.*, 1995; Jarman *et al.*, 2004; Harwood e Obryck, 2005; Pompanon *et al.*, 2012). Desde 2005, com o advento do desenvolvimento das plataformas de sequenciamento de DNA de alta performance (ou *Next Generation Sequencing* – NGS), os métodos de *metabarcoding* e de sequenciamento *shotgun* de DNA também passaram a ser aplicados aos estudos de identificação de presas. Todos esses métodos apresentam vantagens e desvantagens, desde questões relacionadas à sensibilidade (habilidade de evitar falsos-negativos), especificidade (habilidade de evitar falsos-positivos), resolução taxonômica na identificação das presas, custo, praticidade e grau de complexidade da mão de obra necessária, que serão discutidos mais à frente no **Tópico 2.2**.

Para o estudo da biodiversidade de presas consumidas por insetos predadores o mais importante é utilizar um método de detecção que possibilite a identificação do maior número de espécies predadas (incluindo as consumidas em menor abundância), com alta resolução taxonômica (se possível, a nível de espécie), sem falsos positivos (detecções de taxa não predados) e negativos (taxa predados mas não detectados) (Taberlet *et al.*, 2012; Deagle *et al.*, 2014). Paula *et al.* (2015, 2016) desenvolveram um método de detecção que usa o sequenciamento de DNA de alta performance presente no conteúdo gastrointestinal dos predadores, sem haver prévia amplificação de fragmentos do DNA das potenciais presas. Devido a isso, esse método tem potencial para identificar a biodiversidade de presas consumidas, incluindo presas intra e secundária. Os fragmentos de DNA das presas sequenciados (denominados *reads*) são contrastados com um banco de dados (também denominado banco de referência) que contém o DNA mitocondrial (ou mtDNA ou mitogenomas) de várias espécies de potenciais presas, possibilitando a identificação das espécies em processo de digestão no gastrointestinal

dos predadores. Porém, esse método ainda não foi caracterizado em termos de sensibilidade, especificidade e quanto ao poder de resolução taxonômica, por exemplo, ainda a não se sabe: Todas as espécies de presas consumidas são de fato detectadas (há falsos-negativos)? Espécies não consumidas são detectadas (falsos-positivos)? O tempo de detecção da presa pós-predação pelo predador varia em função da espécie da presa e do predador? A detecção de uma mesma espécie de presa varia de acordo com a espécie do predador? Da fase (imatura ou adulta) do ciclo de vida do predador? Do modo de alimentação do predador (mastigador x sugador)? É possível detectar predação secundária por esse novo método e por quanto tempo pós-predação primária e secundária? Quanto maior for o número de presa (massa) consumida (de uma mesma espécie) maior será o tempo de detecção? A temperatura em que o predador faz a digestão da presa pode afetar (encurtar ou prolongar) o tempo de detecção da presa? O tamanho do inserto (pb) escolhido para a construção das bibliotecas de DNA pode influenciar na detecção da presa (número de *reads* e duração da detecção)? Qual a variabilidade na detecção das presas para bibliotecas de DNA diferentes?

Tendo em vista a importância do conhecimento da diversidade de presas consumidas por insetos predadores para o avanço de seu uso como agentes de controle biológico natural de pragas agrícolas, essa tese de doutorado determinou a sensibilidade, a especificidade e o poder de resolução taxonômica da detecção de presas por meio do método de sequenciamento de alto desempenho do DNA presente no conteúdo gastrointestinal de predadores. Como predadores foram utilizadas as joaninhas (Coleoptera: Coccinellidae) *Cycloneda sanguinea*, *Hippodamia convergens*, *Harmonia axyridis* e o crisopídeo *Chrysoperla externa* (Neuroptera: Chrysopidae) devido a sua abundância nos agroecossistemas nacionais (Carvalho e Souza, 2002; Togni, 2014). Como presas foram utilizados o pulgão-verde *Myzus persicae* (Hemiptera: Aphididae),

a mosca-branca *Bemisia tabaci* (Hemiptera: Aleyrodidae) e a lagarta-do-cartucho do milho *Spodoptera frugiperda* (Lepidoptera: Noctuidae). Além de serem predadas por joaninhas e crisopídeo, essas presas foram selecionadas por serem pragas de impacto significativo em culturas de grande valor econômico no agronegócio nacional, como hortaliças, algodão e milho (Cruz e Monteiro, 2004).

2. REVISÃO BIBLIOGRÁFICA

2.1 Importância dos insetos predadores e da identificação de suas presas

Os inimigos naturais são responsáveis pela regulação populacional de espécies pertencentes a um mesmo nível trófico ou de níveis tróficos diferentes, influenciando diretamente na estrutura da população, com consequentes mudanças na comunidade (Alhmedi *et al.*, 2010). Os predadores são considerados, entre os inimigos naturais, a primeira linha de defesa contra insetos fitófagos, muitos deles pragas da agricultura (Oliveira, 2006). São importantes agentes do serviço ecossistêmico de controle de pragas. A utilização de insetos predadores para controle natural de pragas possui como vantagens o baixo impacto ambiental, não ser danoso à saúde dos trabalhadores rurais e seus efeitos se tornarem de longo prazo quando se conserva a comunidade de predadores na lavoura (Junior, 2011). Losey e Vaughan (2006) citam em seu estudo que os predadores locais são responsáveis por aproximadamente 33% do controle anual de pragas, cujos gastos anuais para controle das mesmas são estimados nos Estados Unidos em aproximadamente US\$4,5 bilhões de dólares.

A mera presença de insetos predadores na lavoura não necessariamente garante o controle natural de pragas. Um dos motivos é que como são generalistas, os predadores podem se alimentar de espécies de presas não-pragas. Para estimar o grau de controle natural de uma espécie ou de uma comunidade de predadores é necessário determinar, portanto, as presas consumidas, já que a dedução baseada na mera coocorrência das espécies (Root, 1973) e de suas flutuações populacionais é indireta e, por isso, inacurada (Walde e Murdoch, 1988; Brook; Bradshaw, 2006).

A identificação das espécies predadas pelos insetos predadores possibilita o conhecimento das relações interespecíficas e de sua dinâmica. Esse conhecimento serve, portanto, de orientação para o manejo e conservação das espécies que podem

incrementar o controle natural de pragas em um dado agroecossistema (Winemiller e Jepsen, 1998; Elliott *et al.*, 2007). Adicionalmente, auxilia na determinação da taxa de predação dos insetos predadores sobre as pragas-alvo e do controle biológico natural exercido por diferentes espécies da comunidade de predadores. Com esses dados, passa-se a conhecer qual predador seria mais efetivo para o serviço ecossistêmico de controle de pragas em certos agroecossistemas em estratégias de controle biológico conservativo (Andow comunicação pessoal).

Os agroecossistemas brasileiros possuem grande diversidade de predadores. Dentre os mais comuns estão diferentes espécies joaninhas (Coleoptera: Coccinellidae), besouros de solo (Coleoptera: Carabidae), crisopídeos (Neuroptera: Chrysopidae), tesourinhas (Dermaptera), sirfídeos (Diptera), formigas, aranhas e vespas (Carvalho e Souza, 2002). Serão apresentados com mais detalhes nessa revisão, os predadores e as presas que foram utilizados nessa tese de doutorado.

2.2 Predadores e presas

a) Predadores

Joaninha *Cycloneda sanguinea*

É conhecida como joaninha vermelha (sem pintas pretas) e pertence à ordem Coleoptera e família Coccinellidae. São facilmente encontradas em áreas temperadas da América Central e do Sul (Vandenger e Gordon, 1988), podendo ser encontradas em plantas dos gêneros *Zea*, *Brassica*, *Citrus*, *Bidens*, *Sonchus*, dentre outras (Sicsú, 2013) e, em plantas arbustivas como o margaridão (*Tithonia diversifolia*) (observação pessoal).

Tanto o adulto (Figura 1A) quanto à larva (Figura 1B) são predadores afidófagos, ou seja, as presas preferenciais são pulgões, no entanto, também se alimentam de ovos e

primeiros ínstaes de várias outras espécies (p.ex. mosca-branca, ácaros, cochonilha, ovos de lepidóptero, dentre outros), além de pólen e néctar (Silva *et al.*, 2013; Sujii e Andow, comunicação pessoal). Joaninhas dessa espécie praticam canibalismo e predação de outros predadores (p.ex. joaninhas), dependendo da densidade populacional e da disponibilidade de presa (Michaud, 2002). A larva se alimenta ingerindo o líquido do corpo de suas presas e o adulto devora a presa por inteiro (Santos e Pinto, 1981).



Figura 1. *Cycloneda sanguinea*. A - Adulto (Foto de: Ron Hemberger); B - Larva. (Foto de: Sammy Ekker).

O ciclo de vida constitui em quatro ínstaes na fase larval, pupa e adulto. A duração média, em dias, de cada estágio varia de acordo com o tipo de alimentação. De acordo com Figueira *et al.* (2003), quando *C. sanguinea* se alimenta do pulgão *Schyzaphis graminum*, cada estágio larval dura em média dois dias, a pré-pupa um dia e a pupa quatro dias. A fase adulta pode apresentar variação no tempo de duração, onde adultos alimentados com pulgão-do-pinus *Cinara atlantica* apresentou duração de 125 dias (Oliveira *et al.*, 2004) e adultos que foram alimentados com *Sc. graminum*. Santa-Cecília *et al.* (2001) relatam que quando alimentados por *Sc. graminum* que o período de duração de desenvolvimento de larva a adultos variou de 15,2 dias para machos e 15,5 dias para fêmeas.

Joaninha *Hippodamia convergens*

Essa joaninha também pertence à ordem Coleoptera e família Coccinellidae e é um predador facilmente encontrado nas Américas do Norte e do Sul (Flint e Dreistadt, 2005). No Brasil, é comumente encontrada em hortaliças, notadamente em couve (Embrapa Milho e Sorgo, 2009), em plantas das famílias Asteraceae, Apiaceae, Brassicaceae, dentre outras (Sícsu, 2013).

Os adultos (Figura 2A) e as larvas (Figura 2B) são predadores afidófagos, mas também se alimentam de ovos de outras espécies de insetos, dos primeiros ínstaes de várias outras espécies, de pólen e de néctar (Silva *et al.*, 2013).



Figura 2. *Hippodamia convergens*. A - Adulto; B - Larva (Fotos de: Luis F. Aristizábal).

Kato *et al.* (1999) estudaram o desenvolvimento e a viabilidade da *Hi. convergens* quando alimentada com três diferentes dietas (ovos de *Anagasta kuehniella* [± 2 mg/dia] e os pulgões *Sc. graminum* e *Brachycaudus schwartzi* [± 180 pulgões/dia]). Foi observada que a duração dos três primeiros ínstaes pode variar entre dois e três dias, o quarto ínstar apresentou duração superior a quatro dias, pré-pupa um dia e pupa quatro dias. Quando comparado o período de duração do período larval a adulto, observaram que houve variação de $19 \pm 1,4$ dias quando alimentados por ovos de *A. kuehniella*, $17 \pm 1,4$ dias quando alimentando com *S. graminum* e, $17,9 \pm 0,7$ dias quando alimentados com *B. schwartzi*.

Joaninha *Harmonia axyridis*

Também pertence à ordem Coleoptera, família Coccinellidae (Figura 3) e é originária da Ásia. Essa espécie tem sido introduzida em diversos países como agente de controle de pulgões e foi introduzida acidentalmente no Brasil no início dos anos 2000 (Almeida e Silva 2002) vinda da Argentina, onde foi intencionalmente introduzida para controle de pulgões (Saini, 2004). O primeiro registro de sua ocorrência no Brasil foi no Estado do Paraná em 2002 se alimentando de pulgões presentes em estremosa, *Lagerstroemia indica* Linnaeus (Lythraceae) e em *Pinus* sp. Em estudo realizado em agroecossistemas no Distrito Federal, Sicsú (2013) relata ter observado a presença de *H. axyridis* em plantas dos gêneros: *Zea*, *Ageratum*, *Brassica*, *Citrus* e *Tithonia*.



Figura 3. *Harmonia axyridis*. A - Adultos em acasalamento; B - Larva, apresentando as cores características (Fotos de: Sam Bal).

Santos *et al.* (2009) relatam que a longevidade de fêmeas e machos alimentados com ovos de *A. kuehniella* foi de $74,1 \pm 3,33$ e $67,3 \pm 6,01$ dias respectivamente e, quando alimentados com adultos de *Sc. graminum*, apresentaram longevidade de $76,2 \pm 4$ e $70,3 \pm 7,59$ dias respectivamente, porém não observam diferença significativa entre a longevidade e as diferentes fontes alimentares.

Crisopídeo *Chrysoperla externa*

Pertencem à ordem Neuroptera, família Chrysopidae. Insetos dessa família são conhecidos vulgarmente como bicho lixeiro. Possuem distribuição cosmopolita e são facilmente encontrados em habitats naturais e em agroecossistemas, como diversos tipos

de pomares e lavouras tais como tomate, algodão, trigo, couve, sorgo, dentre outras (Adams, 1987).

É polívoros na fase adulta (Figura 4A) e predador generalista na fase larval (Figura 4B), sendo um dos mais importantes predadores de pragas (Miller *et al.*, 2004; Sattar *et al.*, 2011), tais como ovos e primeiros ínstar de lepidópteros, ácaros (Tetranychidae e Eriophyidae), hemípteros (cochonilhas: Coccidae, Monophlebidae, Pseudococcidae, Eriococcidae e Diaspididae), pulgões, moscas-brancas, psílídeos (Psyllidae), dentre outros (Saminathan *et al.*, 1999; Albuquerque, 2009; Silva *et al.*, 2013).



Figura 4. *Chrysoperla externa*. A - Adulto; B - Larva (fase predadora) (Fotos de: Harvey Schmidt).

O ciclo de vida é dividido em quatro ínstar larvais, pupa e adulto. Estudos indicam que a duração, em dias, de cada fase depende da presa, variando de três a sete dias no 1º ínstar; três a cinco dias no 2º; três a seis dias no 3º; 11 dias \pm 0,4 na fase de pupa; e 60 dias quando adultos (Ru *et al.*, 1975; Aun, 1986; Ribeiro *et al.*, 1988; Neto *et al.*, 2008).

b) Presas

Pulgão *Myzus persicae*

O pulgão *Myzus persicae* (pulgão-verde) (Figura 5) pertence à ordem Hemiptera e família Aphididae. Os pulgões são insetos sugadores que se alimentam exclusivamente

do floema das plantas hospedeiras e, devido a isso, podem afetar o crescimento e o desenvolvimento vegetal, além de transmitir diversos vírus (Godfrey *et al.*, 2000; Leite *et al.*, 2008). É considerado praga de diversas hortaliças no mundo todo, dentre elas de abóbora, berinjela, pimenta, jiló, tomate, couve (Pinto *et al.*, 2008; Silva *et al.*, 2006; Lopes *et al.*, 2007; Pereira *et al.*, 2012). No Brasil são responsáveis por grandes prejuízos em culturas de sistemas protegidos e em campo (Bueno, 1999).



Figura 5. Pulgão *Myzus persicae*. Adulto áptero (Foto de: E. B. Radcliffe).

O ciclo de vida pode ser dividido em quatro instares e em adulto e, de acordo com Cividanes e Souza (2003), a duração de cada fase está diretamente relacionada à temperatura e à planta hospedeira utilizada. Esses pesquisadores relataram que em couve (*Brassica oleracea* var. *acephala*) a 25 °C, *M. persicae* apresentou em média 3,4 ± 0,4 dias no 1º instar, 1,3 ± 0,33 dia no 2º instar, 1,1 ± 0,32 dia no 3º instar, 1,7 ± 0,36 dia no 4º instar e 18,7 ± 1,84 dias como adulto. Em condições climáticas tropicais, a reprodução acontece exclusivamente por partenogênese (Miranda e Suassuana, 2004) e, devido ao curto ciclo de vida e a alta capacidade reprodutiva, podem atingir altas densidades populacionais no campo (Costello e Altieri, 1995).

Os inimigos naturais de pulgões incluem predadores, tais como coccinelídeos afidófagos (Coleoptera) (*C. sanguinea*, *Hi. convergens*, *H. axyridis*, dentre outras joaninhas), crisopídeos na fase larval (*Chrysoperla carnea* e *Ch. externa*), aranhas

(Araneae), e parasitóides da ordem Hymenoptera (Braconidae, Aphelinidae, Aphidiidae) (Steenis, 1993; Pervez e Omkar, 2006; Soares *et al.*, 2007).

Mosca-branca *Bemisia tabaci*

Comumente conhecidos como mosca-branca pertencentes à ordem Hemiptera e família Aleyrodidae, que também se alimentam por sucção do floema das plantas hospedeiras. Devido ao dano de herbivoria e por atuarem como vetores na transmissão de vírus (Wisler *et al.*, 1998), é considerada praga de grande importância em diversas culturas, tais como hortaliças (abóbora, abobrinha, alface, couve-flor, repolho, tomate), leguminosas e algodoeiro (Pinto *et al.*, 2008; Pereira *et al.*, 2012; Carvalho *et al.*, 2013). Possui ampla distribuição geográfica no Brasil (a exceção da região Amazônica) e em diversos países de clima temperado, subtropicais e tropicais como a Austrália, Japão e Canadá, (Haji *et al.*, 2005; Pinto *et al.*, 2008; Poletti e Alves, 2011), assim como na América do Norte, África, América Central, América do Sul e Ásia (CABI, 2017).

O ciclo de vida total de ovo ao final da fase adulta (Figura 6A) dura aproximadamente 25 dias, passando pela fase de ovo, fase larval com três ínstar (Figura 6B), pupa e adulto. De acordo com Salas e Mendonça (1995), possuem duração de quatro dias para o 1º ínstar; 2,7 para o 2º ínstar; 2,5 para o 3º ínstar; 5,8 para pupa e 19 para adultos.

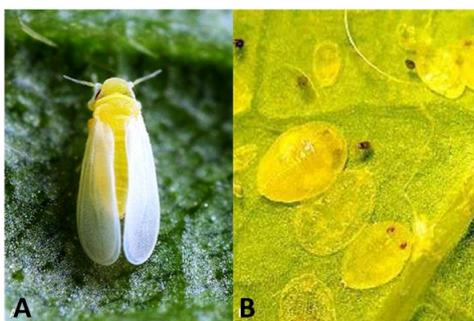


Figura 6. Mosca-branca *Bemisia tabaci*. A - Adulto (Fonte: revistaplantar.com.br/Amosca-branca-ataca-e-prejudica-as-lavouras-de-feijao-de-jatai-go); B - Fase larval (Fonte: Acervo Embrapa Hortaliças: sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Tomate/TomateIndustrial_2ed/Image86.jpg).

A maioria dos inimigos naturais se alimentam de ninfas e dentre eles estão: *Encarsia formosa* e *Encaria lutea* (Hymenoptera: Aphelinidae), *Cycloneda* sp. (Coleoptera, Coccinellidae), *Orius* sp. (Hemiptera: Anthocoridae), *Chrysoperla* sp. (Neuroptera, Chrysopidae), dentre outros (Haji *et al.*, 2005; Silva *et al.*, 2006; CABI, 2017).

Lagarta-do-cartucho do milho *Spodoptera frugiperda*

São conhecidas como lagarta-do-cartucho do milho e pertencem à ordem Lepidoptera, família Noctuidae. A fase imatura é considerada praga de grande importância de culturas como milho, acelga, berinjela, couve, alface, feijão, quiabeiro (Silva *et al.*, 1968; Sarmiento *et al.*, 2002), dentre outras. Apresentam ampla distribuição no Brasil, África, América do Norte, América do Sul, América Central e Europa (CABI, 2017). Na fase larval são mastigadores sendo que, na fase inicial as larvas apenas raspam a superfície das folhas, deixando apenas uma membrana translúcida, quando maiores passam a mastigar todo o tecido das folhas (Moreira e Aragão, 2009).

Possui o ciclo de vida dividido em ovo (Figura 7A), sete ou oito ínstaes larvais (Figura 7B), pupa e adulto. A duração média da fase de ovo é de três dias, 19 dias na fase larval (onde o primeiro e os últimos ínstaes apresentaram duração de três dias e os demais, dois dias), dois dias na fase de pré-pupa, 11 dias na fase de pupa e 11 dias na fase adulta, quando alimentadas com folhas de mandioca (Lopes *et al.*, 2008).



Figura 7. *Spodoptera frugiperda*. A - Ovos (Fonte: <http://bayercontralagartas.com.br/portal/pragas>); B - Lagarta. (Foto de: Marcelo Madalosso). Estas fases são as que são predadas por joaninhas e crisopídeo. Os adultos de *S. frugiperda* possuem outros inimigos naturais.

Os inimigos naturais incluem: *Sceliphronss figulan* (Hymenoptera: Sphecidae); *Dorus luteipes* (Dermaptera: Forficulidae); *Achaetoneura archppivora*, *Archytas incertus*, *Archytas marmoratus* e *Archytas piliventris* (Diptera: Tachinidae); *Calosoma alternans granulatum* e *Calosoma angulata* (Coleoptera: Carabidae); *Coleomegilla maculata* e *C. sanguinea* (Coleoptera: Coccinellidae), dentre outros (Valicente, 1989; Cruz, 1995; Sarmiento *et al.*, 2002; Nava *et al.*, 2009).

2.3 Detecção de presas por meio de métodos moleculares

Os estudos de predação e das interações tróficas iniciou-se pela observação direta no campo, tornando-se impraticável para predadores invertebrados pequenos, de hábitos crípticos (ou subterrâneos), noturnos ou que possuem a combinação desses fatores (Greenstone *et al.*, 2013). Posteriormente tais estudos passaram a ser realizados post-mortem pela coleta dos predadores e tentativa de identificação taxonômica das presas em seus conteúdos gastrointestinais, utilizando-se lupas estereoscópicas e habilidades taxonômicas (Ingerson-Mahar, 2002). No entanto, a correta identificação das espécies a partir de seus restos mortais consistia em tarefa pouco factível, haja vista que a maioria dos predadores terrestres são sugadores dos fluídos das presas (Harwood e Obryck, 2005).

Nos últimos 50 anos passaram-se a utilizar técnicas moleculares para detecção das presas consumidas por insetos predadores, inicialmente por meio: da comparação do perfil eletroforético de isoenzimas (Walrant e Loreau, 1995); do ensaio de imunoabsorção enzimática (do inglês *Enzyme Linked ImmunonoSorbent Assay-ELISA*) (Naranjo e Hagler, 2001; Symondson, 2002; Harwood e Obrycki, 2005; Sheppard e Hardwood, 2005), da eletroforese do DNA em gradiente de gel desnaturante (DGGE) (Harper *et al.*, 2006). Posteriormente, passou-se a utilizar a técnica de reação em cadeia

pela polimerase - PCR (Harper *et al.*, 2005; King *et al.*, 2008; King *et al.*, 2010a,b), radioisótopos (Harwood e Obrycki, 2005) e isótopos estáveis (Harwood e Obrycki, 2005; Coelho, 2008). Mais recentemente passaram a ser aplicadas as técnicas de PCR quantitativo (qPCR) (Zhang *et al.*, 2007; Durbin *et al.*, 2008, Schmidt *et al.*, 2009; Weber; Lundgren, 2009) e de *metabarcoding* (Deagle *et al.*, 2009; Pegard *et al.*, 2009; Soininen *et al.*, 2009; Valentini *et al.*, 2009; Deagle *et al.*, 2010; Bohmann *et al.*, 2011; Kowalczyk *et al.*, 2011; Murray *et al.*, 2011; Rayé *et al.*, 2011). O uso do sequenciamento de alto desempenho de DNA do conteúdo gastrointestinal dos predadores para a identificação de presas foi desenvolvida por Paula *et al.* (2015, 2016). Todas essas técnicas serão apresentadas nos tópicos que se seguem.

Apesar de todo o avanço já alcançado na aplicação de ferramentas moleculares para a identificação de presas, existem ainda limitações importantes para garantir inferências ecológicas robustas quanto às taxas de predação e a força das interações tróficas. Como exemplos pode-se citar a impossibilidade de determinar o número de presas consumidas por unidade de tempo, do tamanho (massa) corporal da presa consumida, do estágio de desenvolvimento da presa (fase imatura ou adulta) e do tempo de predação (Sopp *et al.*, 1992). Adicionalmente, não é ainda possível detectar se a predação foi sobre indivíduos da mesma espécie (ocorrência de canibalismo), se a detecção da presa é fruto de uma predação secundária ou de presa morta (*scavenging*) (Harwood *et al.*, 2001; Calder *et al.*, 2005; Juen e Traugott, 2005; Sheppard *et al.*, 2005). Um melhor entendimento quanto ao canibalismo, predação secundária e *scavenging* é importante, pois têm implicações ecológicas diversas relevantes.

Nos tópicos a seguir, uma breve descrição de algumas metodologias que já foram utilizadas ou ainda são utilizadas na identificação da presa.

a) Perfil eletroforético de isoenzimas

Isoenzimas são enzimas que possuem a mesma função metabólica, mas diferem entre si por alguma propriedade, por exemplo, a massa molecular ou o ponto isoelétrico (Hoffmann e Barroso, 2006). No estudo do perfil de isoenzimas, faz-se a separação de isoenzimas a partir de um campo elétrico (eletroforese) e de um meio de suporte que consiste em gel de poliacrilamida (Hoffmann e Barroso, 2006). Após a eletroforese, o gel é submerso em substratos específicos que reagem com as isoenzimas formando bandas coloridas, indicando a presença de isoenzima específica. De acordo com Symondson (2002), é uma metodologia muito utilizada na detecção da atividade de enzimas e foi utilizada em análises de homogenatos de predadores ou apenas do trato digestivo para identificação de presas. Alguns autores relataram o uso da técnica na análise/identificação da presa no conteúdo gastrointestinal de aranhas e de predadores de ácaros (Dicke e Jong, 1988; Solomon *et al.*, 1996; Ghafoor *et al.*, 2011). Apesar de ser uma técnica fácil e barata, apresenta limitações como a baixa especificidade taxonômica e sensibilidade na identificação das presas (Ferreira e Grattapaglia, 1995; Traugott *et al.*, 2013).

b) Eletroforese do DNA em gradiente de gel desnaturante (DGGE)

A técnica consiste na extração de DNA total de várias amostras, seguida da reação de PCR utilizando universal (iniciadores) das regiões-alvo escolhidas pelo pesquisador. O produto da reação de PCR de cada amostra, por sua vez, é aplicado em gel de poliacrilamida gradiente desnaturante (DGGE) para a separação simultânea dos amplicons de acordo com as sequências de pares de bases (definidos pela proporção G:C e A:T) e, não de acordo com o tamanho dos fragmentos gerados (Muyser e Smalla, 1998; Aboim *et al.*, 2004; Sheppard e Harwood, 2005).

Sheppard e Harwood (2005) relataram a dificuldade em se analisar o conteúdo gastrointestinal de predadores generalistas utilizando iniciadores universais, indicando para esses casos a eletroforese do DNA em gradiente de gel desnaturante. Ainda, de acordo com os autores, a técnica é capaz de separar espécies onde os iniciadores universais não conseguem ou na identificação de presas até a espécie. Tollit *et al.* (2009) utilizaram a técnica para identificar dezenas de espécies de presas contidas em amostras fecais de aves rapinas e relataram que esta técnica é adequada para a análise/identificação de presas contidas em amostras fecais ou conteúdo gastrointestinal de predadores.

A técnica é raramente utilizada para a identificação de presas atualmente. Deagle *et al.* (2005) relataram problema na amplificação de algumas sequências de DNA, sugerindo que tais problemas estariam relacionados à heteroplasmia (mistura de mais de um tipo de genoma mitocondrial), à amplificação de diversos indivíduos geneticamente diferentes ou ainda às mutações resultantes da amplificação por PCR de DNA degradado. Os fragmentos de DNA podem ser agrupados em “Unidades Taxonômicas Operacionais” (do inglês: OTUs - *Operational Taxonomic Units*) e populações comparadas simplesmente em termos de diversidade, mas ressaltaram que a variedade observada pode ser gerada pela existência de haplótipos diferentes. Tais autores citam ainda que embora o DGGE tenha um potencial considerável para a detecção da variação do haplótipo entre uma população de presas, isso pode representar um problema para a identificação de espécies convencionais além de gerar dúvidas quanto à identificação taxonômica, predação secundária e a presença de parasitas e parasitóides nas presas e/ou predadores, problemas também observados nas demais técnicas utilizadas para esse fim.

c) Ensaio de imunoabsorção enzimática (ELISA)

Essa técnica foi utilizada pela primeira vez por Brooke e Proske em 1946 em um estudo sobre predação de invertebrados. A técnica se baseia na produção de anticorpos em cobaias como ratos, galinhas, coelhos, dentre outros, a partir de um extrato protéico total da presa (Symondson, 2002; Sheppard e Harwood, 2005). O anticorpo é utilizado em ensaio imunoenzimático (ELISA) com o extrato do intestino do predador. Podem ser produzidos anticorpos mono e policlonais para (i) ELISA indireto (o mais utilizado), (ii) ELISA direto, (iii) ELISA sanduíche (o que apresenta menor chance de reações inespecíficas) e ainda (iv) o ensaio de Dotblot.

Com o uso de anticorpos monoclonais foi possível identificar e quantificar a presa consumida por determinado predador, sendo utilizado principalmente em estudos de avaliação de inimigos naturais de pragas (Symondson, 2002).

Sheppard e Harwood (2005) citaram diversos estudos envolvendo o uso de anticorpos monoclonais, que utilizaram múltiplos predadores (Hagler e Naranjo, 1994a; 1994b; 2005a, 2005b). Mesmo utilizando predadores em diferentes estágios de desenvolvimento, os autores obtiveram resultados positivos e consequente identificação das presas-alvo presente no conteúdo gastrointestinal dos predadores.

Hagler e Naranjo (1994a) aplicaram pela primeira vez o estudo do conteúdo gastrointestinal de predadores (*Hi. convergens* e *Collops vittatus* - Coleoptera: Melyridae) utilizando múltiplos anticorpos monoclonais para avaliar a frequência com que tais predadores se alimentam das presas: mosca-branca e ovos da lagarta rosada (*Pectinophora gossypiella*). Hagler e Naranjo (1994b) realizaram um estudo com mais de 9.000 predadores, representando seis diferentes gêneros, com o intuito de identificar a presença de antígenos específicos para mosca-branca *B. tabaci* e ovos da lagarta rosada *P. gossypiella*, utilizando anticorpo monoclonal capaz de detectar presas mesmo

após 24 h pós-predação pelo predador. Em 2005, Hagler e Naranjo utilizando anticorpo monoclonal para antígeno de ovos de mosca-branca para avaliar o efeito de diferentes regimes de tratamento inseticida no complexo predador atacando mosca-branca na cultura do algodão, onde foram coletados mais de 30.000 predadores pertencentes a nove taxa distintos, comprovaram, pela primeira vez a predação da mosca-branca por aranhas.

O ELISA permite a quantificação do antígeno a partir de uma curva de calibração do antígeno purificado em concentrações conhecidas e foi a primeira técnica molecular a ser testada para obter parâmetros quantitativos sobre a(s) presa(s) predada(s) no estudo de identificação de presas. É o único método capaz de identificar qual o estágio de desenvolvimento das presas, utilizando anticorpos contra proteínas específicas de cada estágio (Weber e Lundgren, 2009) ou até mesmo fêmeas de *Bemisia* sp (Symondson *et al.*, 1999). Hagler e Naranjo (1994 a,b) ressaltaram que a técnica de ELISA é uma técnica rápida, precisa, barata e sensível, porém citam diversos problemas que podem ocorrer em ensaios imunológicos, tais como reações falsos-positivas devido a interações de terceiro nível trófico, ou ainda por predação secundária e por alimentação de cadáveres (*scavenging*). Citam ainda a dificuldade em quantificar os resultados imunológicos, devido a fatores como temperatura, taxa de digestão do predador, estado metabólico antes da ingestão ou ainda o tamanho da presa. Todos esses problemas na verdade não são exclusivos dessa técnica. De acordo com Sheppard e Harwood (2005), apesar de ser um método que apresenta muitas vantagens, pode se tornar um método caro quando é necessário o desenvolvimento dos anticorpos de interesse; apresentar resultados falsos-positivos (Sunderland *et al.*, 1987; Harwood *et al.*, 2001; Calder *et al.*, 2005); apresentar falsos-negativos e impossibilitar a avaliação do tempo pós-predação,

pois os epítópos podem sofrer a ação da digestão e serem degradados (Symondson, 2002).

d) Rádíoisótopos

O método faz uso de alimentos marcados com rádioisótopos (p.ex. ^{14}C , ^{32}P), os quais são oferecidos para a presa e realiza-se a análise da concentração desses rádioisótopos presentes no predador (Harwood e Obrycki, 2005). Em 1954, Pendleton e Grundmann observaram o acúmulo do rádioisótopo ^{32}P (injetado no tecido de plantas) em três espécies de coccinelídeos, provavelmente devido a predação do pulgão *Anuraphis* sp. Reinfelder e Fisher (1994) estudaram a absorção de diversos rádioisótopos (metais e não metais) por peixes após se alimentarem de crustáceos. Eles relataram que os rádioisótopos não metais (^{14}C , ^{32}P , ^{35}S , ^{75}Se) apresentaram maior taxa de absorção do que os metais (^{109}Cd , ^{57}Co , ^{65}Zn).

Apesar de apresentar bons resultados quanto ao estudo da relação entre presa e predador, a detecção de presas por rádioisótopos é pouco utilizada devido principalmente as questões de segurança biológica e a necessidade de marcar preliminarmente os alimentos com rádioisótopos. Portanto, o método é limitado a estudos em laboratório.

e) Isótopos estáveis naturais

Método que utiliza a análise de fluxo de carbono presente em um ecossistema por meio da comparação da razão dos isótopos $^{13}\text{C}/^{12}\text{C}$ encontrado no intestino do predador e nas presas. O método baseia-se no fato de que alguns isótopos, como o ^{13}C , apresentam elevado grau de conservação e diferentes concentrações entre os organismos (Moore e Semmes, 2008). Ostrom *et al.* (1997) utilizaram os isótopos estáveis de carbono (^{13}C) e nitrogênio (^{15}N) para avaliar o fluxo de energia na teia alimentar da

joaninha *Co. maculata*. Foi possível observar a importância de diferentes plantas (alfafa, trigo e milho) para a base alimentar dessa joaninha nos agroecossistemas estudados. Steffan *et al.* (2001) observaram que a joaninha *Hi. convergens* apresentou enriquecimento de ^{15}N após consumir o néctar de repolhos fertilizados com nitrato de potássio enriquecido com ^{15}N ($\text{KNO}_3\text{-}^{15}\text{NO}_3$). Madeira *et al.* (2014) utilizaram os isótopos estáveis de carbono (^{13}C) e nitrogênio (^{15}N) para analisar a dispersão dos predadores *Orius majusculus* e *Coccinella septempunctata* entre o milho e a alfafa, nos estágios vegetativo e reprodutivo do milho, utilizando o pulgão como presa.

Esse método é mais adequado para estimar o tempo pós-predação das presas, pois não possui resolução suficiente para identificação taxonômica das presas (Moore e Semmes, 2008). Adicionalmente, o método requer a utilização de equipamentos caros e de mão-de-obra muito especializada, não havendo oferta de serviço terceirizado comercialmente para as análises (Harwood e Obrycki, 2005; Coelho, 2008).

f) Reação em cadeia por polimerase (PCR)

É um método muito utilizado devido a sua alta sensibilidade e baixo custo, além da molécula-alvo (DNA) apresentar maior resistência ao processo de digestão do que proteínas (Symondson, 2002). Existem duas abordagens quanto às análises de PCR do conteúdo gastrointestinal: a) utilização de iniciadores específicos para amplificação de determinada região do DNA de determinada espécie de presa; e b) utilização de iniciadores degenerados para amplificação de regiões barcodes do DNA das presas, com subsequente clonagem dos amplicons e sequenciamento de DNA (pelo método de Sanger *et al.*, 1977) para identificação das presas com amplicons de sequências nucleotídicas dos barcodes já determinadas (Pompanon *et al.*, 2012). Regiões barcodes do DNA são fragmentos de genes utilizados para taxonomia molecular por diferirem pouco entre os indivíduos da mesma espécie (baixa variação intraespecífica), mas

diferirem o suficiente ($\geq 2\%$) entre diferentes taxa (alta variação interespecífica) (Hebert *et al.*, 2003). As regiões flanqueadoras das regiões barcodes onde se desenham os iniciadores degenerados devem ter baixa variação de sequência nucleotídica entre diferentes espécies (p.ex. serem conservadas entre taxa) de tal forma que permita o uso de iniciadores para amplificação das regiões barcodes de vários taxa (Ji *et al.*, 2013).

Os genes preferencialmente utilizados para identificação de espécies de animais são de origem mitocondrial. Isso se deve ao fato: a) da maior resistência à degradação devido a sua circularidade; b) ao seu elevado número de cópias nas células comparado ao DNA nuclear; c) possuir regiões barcodes, tal como *citocromo c oxidase 1* (COX1 ou COI) (Agustí *et al.*, 2003; Pompanon *et al.*, 2012). Existem bancos de dados públicos (iBOL, www.barcodeoflife.org, e *GenBank*, <http://www.ncbi.nlm.nih.gov/GenBank/>) onde estão disponíveis as informações sobre os barcodes já obtidos para identificação de várias espécies. Folmer *et al.* (1994) desenvolveram e disponibilizaram uma lista de iniciadores para amplificação da região barcode do gene COX1 em vários taxa de insetos.

Existem vários exemplos na literatura quanto ao uso de PCR para verificação da predação de presas previamente conhecidas, por exemplo, (i) a identificação de nematóides ingeridos por *Collembola* (Read *et al.*, 2006), (ii) avaliação do tempo de digestão de DNA em joaninhas coletadas no campo a partir da avaliação do número de amplicons obtidos por PCR utilizando iniciadores com amplificação para diferentes tamanhos de fragmentos (Hoogendoorn e Heimpel, 2003) ou ainda (iii) análise do conteúdo gastrointestinal de aranhas (Greenstone e Shufan, 2003). No caso da identificação de presas através de regiões barcodes do DNA, a taxonomia se dá pela comparação da similaridade entre as sequências de DNA amplificadas e sequenciadas das presas com sequências barcodes de taxonomia definida, depositadas em um banco

de dados. Pompanon *et al.* (2012) cita diversos exemplos de estudos utilizando DNA barcode na análise do conteúdo intestinal ou fecal de anfípodas e bivalves marinhos (Blankenship e Yayanos, 2005), pinguins (Deagle *et al.*, 2007), golfinhos (Dunshea, 2009), morcegos (Zeale *et al.*, 2011). Pesquisadores relataram ainda a utilização de DNA barcodes no estudo da relação entre os besouros da subfamília Chrysomelinae e plantas hospedeiras (Jurado-Rivera *et al.*, 2009), ou ainda na identificação do hospedeiro utilizado durante a fase larval de parasitóides (Rougerie *et al.*, 2010).

Apesar da alta sensibilidade e eficiência da técnica de PCR, ambas as abordagens vêm sendo criticadas na comunidade científica. Na primeira abordagem, o método está limitado ao conhecimento prévio da espécie de presa que é ingerida pelo predador, o que torna o método inadequado para detectar presas desconhecidas ou para determinar a diversidade de presas consumidas. Adicionalmente, em alguns casos as presas podem apresentar alto grau de proximidade taxonômica com o predador (relação intraguilda) e, mesmo utilizando iniciadores específicos, os resultados podem ser suprimidos devido à maior concentração do DNA do predador quando comparado ao DNA digerido da presa (Vestheim e Jarman, 2008; Deagle *et al.*, 2009).

Na segunda abordagem, embora o método tenha a vantagem de potencialmente poder identificar ampla gama de espécies de presas ingeridas, a abordagem é considerada muito trabalhosa devido as etapas de clonagem e sequenciamento de vários clones. Adicionalmente, pesquisadores têm relatado a dificuldade de se obter dados confiáveis, sem falsos-negativos ou falsos-positivos. Por exemplo, Johanowicz e Hoy (1996) ao utilizarem o barcode 16S para a detecção de uma espécie de presa em duas espécies diferentes de predadores observaram que a presa apenas foi detectada em uma espécie de predador. Além disso, a detecção ocorreu em apenas um indivíduo dentro de um grupo de cinco que ingeriram a presa, mesmo que os predadores tenham sido

congelados imediatamente pós-predação da presa. Falsos-negativos e falsos-positivos podem ser gerados devido: as regiões barcodes não apresentarem variação interespecífica suficiente para distinguir diferentes grupos taxonômicos (Ficetola *et al.*, 2010; Deagle *et al.*, 2014); durante a PCR pode haver erros como substituição, inserção ou remoção de bases; o primer utilizado não deve ser capaz de distinguir espécies estritamente relacionadas; erros no sequenciamento; contaminação das amostras; ou baixa eficiência do primer (Coissac *et al.*, 2012; Pinto e Raskin, 2012; Taberlet *et al.*, 2012; Clarke *et al.*, 2014; Deagle *et al.*, 2014; Ficetola *et al.*, 2014; Piñol *et al.*, 2015).

g) PCR quantitativo (qPCR)

O qPCR estima a quantidade de DNA relativa ou absoluta em uma amostra por meio da captação de fluorescência gerada nas reações de amplificação do(s) gene(s)-alvo por agentes intercalantes do DNA ou por sondas fluorescentes no decorrer dos ciclos de PCR (VanGuilder *et al.*, 2008). Esse método tem sido utilizado principalmente para determinar parâmetros quantitativos relacionados à detecção das presas. Por exemplo, Schmidt *et al.* (2009) avaliaram o tempo de detecção da presa *Myzus persicae* pelo predador *Macrolophus caliginosus* utilizando iniciadores para o gene mitocondrial COX1 e para o gene esterase presente no DNA nuclear. Eles observaram que a presa pôde ser detectada até 9 h pós-predação e que houve maior detecção para o gene mitocondrial do que para o nuclear. Também foi possível os autores avaliarem outros fatores que afetam a detecção das presas, como por exemplo, a temperatura de digestão. Weber e Lundgren (2009) avaliaram a influência de diversos fatores na detecção da presa *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae), utilizando o gene COX1, consumida pelo predador *Co. maculata*. Os fatores estudados foram número de ovos ingeridos; idade dos ovos; a forma em que os predadores foram conservados até o momento da análise; e a presença de alimento anterior à predação.

O qPCR é considerado um método relativamente simples, de alta sensibilidade e barato (Sheppard e Hardwood, 2005). A principal vantagem sobre o PCR convencional é o caráter quantitativo. No entanto, apesar da possibilidade de quantificação do número de cópias dos genes-alvo da(s) presa(s) presente no conteúdo gastrointestinal dos predadores, o qPCR ainda não possibilita a contagem do número de presas ingeridas por predadores coletados no campo. O motivo é que o número de moléculas de DNA, seja nuclear ou mitocondrial, da(s) presa(s) está relacionado à quantidade de massa ingerida. A quantidade de massa ingerida de presa por sua vez está relacionada não apenas ao número de espécimes consumidos, mas também ao seu tamanho corporal e estágio fenológico/idade. Outro fator complicador é o desconhecimento do tempo pós-predação, ou seja, do grau de digestão da presa, ao analisar predadores coletados no campo. Em termos de desvantagem, assim como o PCR, a detectabilidade da presa está diretamente ligada à eficiência dos iniciadores utilizados, podendo gerar falsos-negativos e falsos-positivos (Sint *et al.*, 2014).

h) Metabarcoding

A utilização de regiões específicas do DNA nuclear ou diferentes regiões de organelas utilizadas para amplificação e sequenciamento foram definidas como "DNA *barcoding*" e Metabarcoding consiste no sequenciamento de alto desempenho em paralelo de regiões barcodes do DNA (Hebert *et al.*, 2003), amplificadas por PCR, para acessar a biodiversidade de múltiplas espécies ou taxa presentes em uma amostra de elevado rendimento utilizando o DNA total tipicamente degradado extraído de uma amostra ambiental (isto é, solo, água, fezes, dentre outros) (Taberlet *et al.*, 2012; Ji *et al.*, 2013). Não há etapa de clonagem dos amplicons previamente ao sequenciamento de alto desempenho do DNA. De acordo com a similaridade, as sequências obtidas das presas são categorizadas por unidades taxonômicas operacionais OTUs, as quais

idealmente devem conter somente as sequências de um mesmo táxon. A seguir, a identificação taxonômica das presas é feita pela comparação da similaridade das sequências das OTUs com as depositadas em bancos de dados de barcodes (p.ex. iBOL, *GenBank*).

Esse método tem sido cada vez mais adotado para a identificação de presas (e componentes da dieta de herbívoros), principalmente em mamíferos, devido a sua alta sensibilidade e por possibilitar a análise da presença simultânea de várias presas (Pegard *et al.*, 2009; Soininen *et al.*, 2009; Deagle *et al.*, 2009,2010; Murray *et al.*, 2011; Bohmann *et al.*, 2011; Kowalczyk *et al.*, 2011; Rayé *et al.*, 2011). No entanto, não há ainda estudos com predadores invertebrados e não existem relatos na literatura quanto a sua meia-vida de detecção, bem como de sua sensibilidade em detectar predação secundária e *scavenging*.

As desvantagens do *metabarcoding* para identificação de presas são as mesmas mencionadas quanto ao uso do PCR, com o agravante do alto enriquecimento na amostra com a amplificação do material genético do próprio predador, que por apresentar maior concentração de DNA presente na amostra, pode interferir na amplificação de DNA em menores concentrações, como o da presa (Vestheim e Jarman, 2008). A amplificação do material genético do predador é indesejada, haja vista seu enriquecimento natural na amostra, reduzindo a captação do sinal do material genético das presas. Para minimizar esse problema foi desenvolvido um método para bloqueio da amplificação do material genético do predador. Vestheim e Jarman (2008) desenvolveram uma estratégia em que em uma reação de PCR para amplificação de DNA, além dos iniciadores universais utilizados para a possível identificação da presa, utilizaram também iniciadores modificados, capazes de interagir especificamente com o DNA do predador e interferir com a amplificação desse, agindo como um bloqueador.

Existem ainda outros erros mais comuns visualizados: a degradação do DNA molde; erros durante a amplificação e, erros durante o sequenciamento, sendo que este pode variar de acordo com a tecnologia utilizada, sendo que o erro que pode gerar maior problema durante as análises dos dados seria a baixa qualidade dos bancos de dados de referência taxonômica (Taberlet *et al.*, 2012). Um dos requisitos de grande importância para o uso dessa metodologia é a necessidade de se possuir um banco de dados com referências taxonômicas de alta qualidade (Taberlet *et al.*, 2012). Tais autores relatam ainda que essa tecnologia irá evoluir cada vez mais, ainda mais quando se observa os avanços tecnológicos do sequenciamento de última geração.

i) Sequenciamento direto de alto desempenho do DNA

Esse método consiste no sequenciamento direto de alto desempenho da comunidade de DNA presente no conteúdo gastrointestinal dos predadores sem haver previamente a amplificação de regiões barcodes por PCR, como é feito na técnica de *metabarcoding* (Paula *et al.*, 2015; Srivathsan *et al.*, 2015). Esse método também difere da metagenômica por não haver a montagem de genomas para detecção da diversidade de DNA. A identificação das presas é feita diretamente pela análise de similaridade das *reads* (fragmentos de DNA sequenciados) com as sequências de nucleotídeos do mitogenoma (DNA mitocondrial) das potenciais presas em um banco de dados de referência. O banco de referência não precisa ser mitocondrial apenas, poderia ser do DNA nuclear, no entanto está disponível para menos espécies do que o mitocondrial. O banco de dados de referência de DNA pode ser composto do genoma mitocondrial (ou mitogenoma) das espécies potenciais de presas, obtidos do *GenBank* e/ou providenciado pelo pesquisador.

A metagenômica compreende o estudo da comunidade de DNA em uma amostra ambiental usando sequenciamento de alto desempenho em paralelo do genoma (nuclear

ou mitocondrial, p.ex.) (e não apenas de alguns marcadores moleculares) dos organismos presentes (Thomas *et al.*, 2012). Ela surgiu da necessidade de se estudar a diversidade de populações presentes nas comunidades microbianas, que muitas vezes não são possíveis de serem detectadas por métodos tradicionais de microbiologia (Handelsman, 2004; Marco, 2010). As etapas que normalmente envolvem as análises de metagenômica compreendem o sequenciamento dos genomas da amostra ambiental, a reconstrução (montagem ou assembly) das sequências nucleotídicas dos genomas, a categorização dos grupos de genomas presentes (*binning*), anotação e análises estatísticas ou de diversidade (Thomas *et al.*, 2012). Muitos protocolos de metagenômica mitocondrial foram recentemente desenvolvidos para analisar a diversidade de artrópodes em amostras ambientais contendo mistura de espécies (Zhou *et al.*, 2013; Tang *et al.*, 2014; Andújar *et al.*, 2015). Contudo, não houve aplicação desse método para a identificação de componentes da dieta ou de presas consumidas ainda pela inviabilidade de reconstruir genomas a partir da comunidade de DNA em degradação no gastrointestinal de predadores ou de material fecal. Foran (2006) relata que os loci nucleares de cópia única geralmente não podem ser amplificados a partir de restos degradados, necessitando assim da análise do DNA mitocondrial, que apresenta um maior número de cópias por célula e, que o sucesso na análise de mtDNA geralmente é considerado como resultado do seu maior número de cópias na célula e por fatores como a localização celular ou as características moleculares, que podem influenciar na maior preservação do mtDNA.

Utilizando o método de identificação das presas pelo sequenciamento direto de alto desempenho do DNA do conteúdo gastrointestinal dos predadores, Paula *et al.* (2016) detectaram maior diversidade de presas do que o reportado em outros trabalhos. Utilizando diferentes bancos de referência de DNA (e não apenas mitocondrial das

potenciais presas), detectaram ainda a presença de outras fontes de DNA exógeno como, por exemplo, simbiontes e parasitóides tanto das presas quanto dos predadores. Essa diversidade de DNA exógeno detectada possibilitou a construção de uma rede qualitativa de interações tróficas demonstrando predação extra e intraguilda e predação secundária.

Devido à falta de dados quanto ao potencial do método em gerar falsos-negativos e positivos, ainda não está claro seu real custo-benefício em relação às técnicas de detecção mencionadas anteriormente. Srivathsan *et al.* (2015) fizeram um estudo para comparar as vantagens e desvantagens do uso de *metabarcoding* e do sequenciamento de alto desempenho direto de DNA para análise da dieta de macacos por meio de suas fezes. Os autores constataram que o sequenciamento direto de alto desempenho de DNA detectou sete dos nove componentes da dieta e em nível taxonômico até gênero, enquanto que por *metabarcoding* detectou-se apenas cinco componentes da dieta e em nível taxonômico de menor resolução. No entanto por *metabarcoding* pôde-se identificar componentes da dieta menos abundantes, o que não foi detectado pelo sequenciamento direto de alto desempenho de DNA (falsos-negativos). Usando o sequenciamento direto de alto desempenho de DNA, além dos componentes da dieta, foi também detectado o material genético de diversos parasitas (microbioma) presentes no trato gastrointestinal dos macacos, enquanto que esses não foram revelados por *metabarcoding*.

2.4 Fatores que afetam a detecção da presa

Vários fatores bióticos e abióticos influenciam a detecção das presas, tais como: o método de detecção, o modo de alimentação (mastigação ou sucção), a espécie do predador e o tempo decorrido pós-predação (Greenstone *et al.*, 2014); a quantidade

(número) e a idade da presa ingerida, a presença adjuvante de alimento no gastrointestinal do predador e a forma de conservação/fixação do material a ser analisado (Weber e Lundgren, 2009); a temperatura do ambiente onde o predador digere a presa, a espécie e o tamanho do predador (Hagler e Naranjo, 1997). O relato sobre cada um desses fatores será apresentado a seguir.

Greenstone *et al.* (2014) reportam uma comparação da meia-vida de detecção da presa por várias técnicas moleculares por meio de bioensaios de alimentação em condições controladas (Quadro 1). O trabalho tornou evidente que o decaimento da detecção é função de diversos fatores, mas de maneira geral tem variado na literatura de 0,9 a 193,7 h. Diversos pesquisadores vêm estudando a influência de tais fatores na detecção da presa, avaliando esses fatores de forma independente e por diferentes métodos (PCR, ELISA, dentre outros).

Quadro 1. Comparação da meia-vida de detecção (T50) da presa entre diferentes técnicas moleculares, baseado em Greenstone *et al.* (2014).

Método de detecção	Tamanho (pb)	Presa	Táxon/taxa Presa	Táxon/taxa Predador	Temperatura (°C)	T50 (h)	Bibliografia
ELISA	na	1, 3, 6 ou 10 ovos	Gelechiidae	Cocc	15-35	0,9 e 18	Hagler e Naranjo (1997)
PCR	188 e 271	1-5 ninfas	Psyllidae	Anthocoridae	22	20,9 e 24,1	Agustí <i>et al.</i> (2003)
ELISA	na	3 ou 5 ovos	Cicadellidae	Cocc e Chry	25	11,8 e 34,4	Fournier <i>et al.</i> (2006)
TGGE	332	Afídios a vontade	Afídeo e Lum	Carabidae	25	12	Harper <i>et al.</i> (2006)
PCR	214 e 219	1 ovo	Chry	Cocc e Pent	na	7 e 50,9	Greenstone <i>et al.</i> (2007)
PCR	219	1 ovo ou larva	Scarabaeidea	Carabidae	16	18	Juen e Traugott (2005)
PCR Multiplex	109 e 310	1 lesma	Agri e Arionidae	Carabidae	8	19,7 e 22,4	Hatteland <i>et al.</i> (2011)
PCR	227	1 afídeo	Afídeo e Lum	Lyc e Tetr	24/20	2 e 4,2	Kerzicnik <i>et al.</i> (2012)
PCR	180	1 collembola	Coll e Plat	Lin e Tetr	24	9,5 e 32	Chapman <i>et al.</i> (2013)
PCR	180	Afídios a vontade	Afídeo e Lum	Carabidae	20	12,3	Firlej <i>et al.</i> (2013)
Sequenciamento direto de alto desempenho de DNA	na	1 afídeo	Afídeo	Cocc	25	23,1 e 32,7	Paula <i>et al.</i> (2015)

TGGE: *Temperature Gradient Gel Electrophoresis*; na: Não se aplica; Agri: Agriolimacidae; Chry: Chrysomelidae; Cocc: Coccinellidae; Coll: Collembola; Tetr: Tetragnathidae; Lin: Linyphiidae; Lum: Lumbricidae; Lyc: Lycosidae; Pent: Pentatomidae; Plat: Platygastridae.

Greenstone *et al.* (2007), utilizando PCR, observaram que o tempo máximo de detecção do ovo do besouro *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae) foi de setes horas quando ingerido pela joaninha *Co. maculata* (predador mastigador) e de 50,9 h quando ingerido pelo percevejo *Podisus maculiventris* (Hemiptera: Pentatomidae), predador sugador. Foi a primeira demonstração de que o modo de alimentação do predador influencia no tempo de detecção da presa pós-predação. Os autores atribuíram a diferença no tempo de decaimento da detecção da presa à velocidade diferencial de digestão entre os predadores e hipotetizaram que predadores sugadores tem taxa de digestão mais lenta que predadores mastigadores.

Weber e Lundgren (2009) avaliaram a detecção do gene COX1 por qPCR (amplicon de 214 pb) da presa *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae) no trato gastrointestinal do predador *Co. maculata* (Coleoptera: Coccinellidae), após esses ingerirem ovos da presa. Foram avaliados: o número de ovos ingeridos; idade dos ovos (em 1, 3 e 5 dias); a forma em que os predadores foram conservados até o momento da análise e o estado do predador (predador faminto, predador que se alimentou a vontade de pulgões não ápteros após a ingestão do ovo da presa-alvo e predadores que se alimentaram a vontade de ovos da mesma espécie, após a ingestão do ovo da presa alvo). Como resultado, foram observados que: i) quanto maior o número de ovos ingeridos, maior a quantidade de DNA detectada da presa, ou seja, maior é a quantidade de DNA disponível para ser detectada; ii) quanto maior tempo decorrido pós-predação, menor a quantidade de DNA detectada da presa, ou seja, a medida que a digestão da presa transcorre, menor quantidade de DNA-alvo passível de ser amplificada por PCR/detectada; iii) a idade do ovo da presa ingerida não influenciou na detecção do DNA-alvo, ou seja, aparentemente a quantidade de DNA supostamente diferencial entre as idades (quanto mais velhos os ovos, maior a quantidade de DNA)

não foi suficiente para apresentar significativa diferença na detecção da presa; iv) os predadores que foram conservados em etanol 70% resfriados -20 °C apresentaram uma maior recuperação do DNA-alvo, ou seja, o método de armazenamento da amostra pode influenciar na preservação do DNA da presa e, portanto, em sua detecção pós-predação; v) o decaimento na detecção do DNA da presa-alvo era mais lento em predadores que permaneceram sem se alimentar após a ingestão da presa-alvo (T_{50} = 51 min), do que os predadores que se alimentaram a vontade de pulgões ápteros (T_{50} = 16 min) ou de ovos da mesma espécie do predador (T_{50} = 31 min). A presença coadjuvante de alimento no gastrointestinal influencia na taxa de degradação do DNA da presa ingerida mais recentemente.

Hagler e Naranjo (1997), utilizando ELISA indireta, avaliaram o conteúdo gastrointestinal de três predadores de espécies e tamanhos diferentes [*Orius insidiosus* (Hemiptera: Anthocoridae) – predador sugador, *Geocoris punctipes* (Hemiptera: Lygaeidae) predador sugador e *Hi. convergens* (Coleoptera: Coccinellidae) – predador mastigador], que foram alimentados com diferentes números de ovos da lagarta-rosado-algodão *Pectinophora gossypiella* (Lepidoptera: Gelechiidae) e mantidos por tempos e temperaturas diferentes durante a digestão. Como resultado, observaram que (i) as espécies menores de predadores (*O. insidiosus* e *G. punctipes*) apresentaram maior tempo de detecção das presas, podendo ser por terem taxa de digestão mais lenta já que são sugadores e não mastigadores e também porque foram mais imunoresponsivos à ELISA do que a joaninha; (ii) quanto maior a quantidade de ovos ingerida, maior o sinal de detecção da presa, ou seja, maior a quantidade de antígenos da presa para ser detectada; (iii) quanto maior o tempo decorrido da predação, menor foi sinal de detecção da presa, pois os antígenos da presa vão sendo degradados cada vez mais a medida que o processo de digestão evolui; (iv) quanto maior a temperatura de digestão,

mais curta foi a duração da detecção da presa, ou seja, em temperaturas mais altas a taxa de digestão dos antígenos da presa aumenta.

Não existem ainda relatos na literatura, utilizando como técnica o sequenciamento de alto desempenho do DNA do conteúdo gastrointestinal de predadores, se presas de diferentes espécies, predadas pela mesma espécie do predador, tem meia-vida de detecção diferente e se a taxa de digestão da presa é influenciada pela fase do desenvolvimento do predador (imatura x adulta). O número de fatores que podem influenciar a detecção das presas ingeridas revela a complexidade em realizar inferências sobre predação em insetos coletados a campo e a importância, portanto, de se conhecer melhor sobre a influência dos mesmos, sobretudo para o método de detecção da presa em que determinado estudo será pautado e eventualmente comparado a demais trabalhos científicos.

3. OBJETIVOS

Como objetivo geral pretendeu-se determinar a capacidade do método de sequenciamento direto de alto desempenho de DNA do conteúdo gastrointestinal dos predadores ser aplicado para identificar a diversidade de presas consumidas por predadores generalistas em estudos futuros de predação a campo ou de interações tróficas.

Como objetivos específicos pretenderam-se:

- Determinar a meia-vida de detecção em função da espécie de presa consumida, da espécie de predador, modo de alimentação e fase do ciclo de vida do predador.
- Verificar se uma presa consumida por predador é detectável e se o aumento do número (ou biomassa) de presa consumida aumenta o sinal de detecção. Adicionalmente, se é possível detectar predação secundária e por quanto tempo, bem como se a temperatura de digestão da presa e o tamanho do inserto (pb) utilizado na construção das bibliotecas de DNA influenciam na detecção.
- Avaliar a ocorrência de falsos-positivos e de falsos-negativos;
- Estimar a variabilidade do número de *reads* detectadas para as presas entre bibliotecas do mesmo tratamento.

4. MATERIAL E MÉTODOS

4.1 Criação dos insetos

As colônias de insetos foram mantidas no Laboratório de Ecologia e Biossegurança da Embrapa Recursos Genéticos e Biotecnologia, em Brasília-DF. As coletas de campo foram realizadas sob a autorização do SISBIO número 36950. Os acessos ao patrimônio genético foram realizados sob autorização do IBAMA número 02001.008598/2012-42.

a) Criação dos coccinelídeos *C. sanguinea*, *H. axyridis* e *Hi. convergens*

A criação das três espécies de joaninhas foi realizada de acordo com o protocolo estabelecido no POP “Procedimento de manutenção e manuseio da colônia de *Cycloneda sanguinea*” (código 038.11.02.01.3.009). A criação ocorreu em câmaras climatizadas, do tipo BOD, a $25\pm 4^{\circ}\text{C}$, $60\pm 20\%$ UR e fotoperíodo de 16 h. Os adultos foram mantidos em potes de um litro contendo algodão embebido em água e alimentados com pulgões (*Aphis gossypii*, *Myzus persicae*, *Brevicorine brassicae*) provenientes de colônia já estabelecida ou coletados no campo. A manutenção e a limpeza dos potes foram realizadas em dias alternados. As posturas obtidas foram transferidas para novos potes, a fim de evitar canibalismo pelos adultos, e dois dias após a eclosão dos ovos, as larvas foram individualizadas em copos plásticos de 50 mL também para evitar canibalismo. A manutenção das larvas (alimentação e limpeza do recipiente) seguiu o mesmo procedimento que para os adultos, acrescentado alimentação com *A. kuehniella* além dos pulgões. Periodicamente a colônia recebeu a introdução de adultos coletados no campo.

b) Crisopídeo *Chrysoperla externa*

Os ovos de *Ch. externa* foram obtidos a partir da criação mantida no Laboratório de Entomologia da Universidade Federal de Lavras (UFLA), Lavras, Minas Gerais, coordenado pela professora Dra. Brígida Souza. A criação e manutenção de *Ch. externa* foi realizada em salas climatizadas a $25\pm 4^{\circ}\text{C}$, $60\pm 20\%$ UR e fotoperíodo de 12 h. Os ovos foram mantidos em potes plásticos contendo papel picado e postura de *A. kuehniella*. Após eclosão, as larvas foram individualizadas em copos plásticos de 50 mL e alimentadas com postura de *A. kuehniella* e/ou pulgões (*A. gossypii*, *M. persicae*) até que atingissem a fase de pupa. Os adultos emergidos (aproximadamente 20 indivíduos) foram mantidos em canos (90 mm de diâmetro x 30 cm de altura) forrados com papel (Figura 8), para suporte da postura dos ovos, e selado com telas em um dos lados. O outro lado foi fechado com um tecido, a fim de permitir a manutenção e acesso. Os adultos foram alimentados com uma mistura de mel e levedo de cevada (sem proporção definida por depender da consistência do mel e do levedo utilizados, porém a mistura deve ficar pastosa) e, receberam água através de algodão umedecido. Os papéis contendo as posturas foram recortados e acondicionados em potes plásticos até a eclosão das larvas. A manutenção da alimentação das larvas foi realizada três vezes por semana e dos adultos duas vezes por semana.

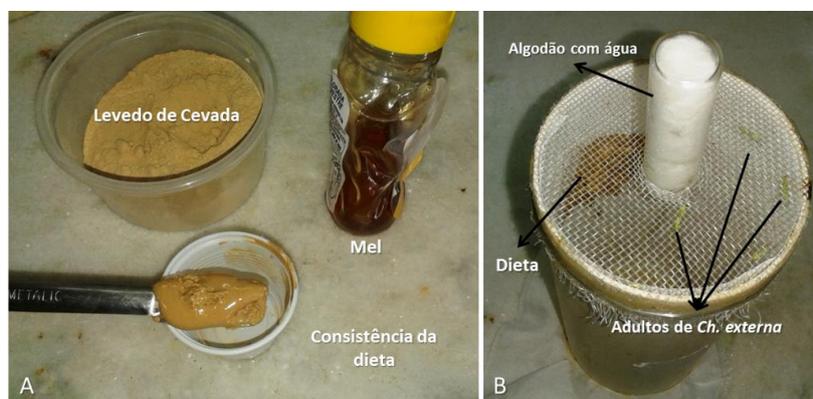


Figura 8. Preparo da dieta artificial para os adultos de *Ch. externa*. A - Aspecto obtido da mistura do levedo de cevada e do mel. B - Gaiola para manter os adultos onde a parte superior apresenta-se vedada por uma tela e a parte inferior é fechada com tecido e liga.

c) Pulgão *Myzus persicae*

A criação e a manutenção da colônia foram realizadas em sala de criação climatizada a $25\pm 4^{\circ}\text{C}$, $60\pm 20\%$ UR e fotoperíodo de 14 h, em mudas de couve, mantidas em potes plásticos de um litro contendo 2/3 de solo fertilizado e 1/3 de substrato de cultivo (BIOPLANT). As mudas foram renovadas a cada 20 dias e regadas manualmente a cada dois dias. Para a renovação da colônia foram utilizados indivíduos coletados no campo. A colônia foi vistoriada diariamente a fim de eliminar qualquer parasitóide encontrado. Os indivíduos coletados no campo permaneceram separados da colônia por pelo menos uma semana, sendo mantidos em folhas frescas de couve, para que fosse possível a manutenção e retirada de parasitóides ou múmias remanescentes.

d) Mosca-branca *Bemisia tabaci*

As moscas foram coletadas no campo e transferidas para gaiolas de voal contendo mudas de algodoeiro, couve ou repolho, utilizadas para alimentação e como suporte para postura de ovos. A criação ocorreu em salas climatizadas a $25\pm 4^{\circ}\text{C}$, $60\pm 20\%$ UR e fotoperíodo de 14 h ou em casa de vegetação. As mudas foram mantidas em potes plásticos contendo 2/3 de solo fertilizado e 1/3 de substrato de cultivo (BIOPLANT). As plantas foram renovadas sempre que necessário e regadas manualmente a cada dois dias. A colônia foi vistoriada diariamente a fim de eliminar qualquer parasitóide encontrado. A população de moscas foi renovada constantemente com moscas coletadas no campo.

e) Lagarta-do-cartucho do milho *Spodoptera frugiperda*

A manutenção da colônia de *S. frugiperda* foi realizada de acordo com Schmidt *et al.* (2001). Para a criação de adultos foram utilizados potes plásticos de dois litros contendo uma folha de papel sulfite cobrindo toda a lateral interna, servindo de suporte

para a oviposição. Os potes continham pequenos furos para permitir a passagem de ar. Os adultos foram alimentados em bebedouro contendo dieta líquida (cada litro aquoso de dieta contém: 50g mel, 100mg ácido ascórbico, 50g sacarose). Para a coleta das posturas, os papéis foram renovados a cada dois dias. Os papéis contendo a postura foram recortados, e os pedaços contendo os ovos lavados por cinco minutos em formaldeído 5%, enxaguados com água autoclavada, e por fim, lavada em solução de sulfato de cobre a 1% por dois minutos. Após o processo de lavagem, as posturas (de dois a cinco pedaços de papel contendo posturas, a depender do tamanho da postura) foram acondicionadas em potes plásticos de 250 mL contendo dois cubos de dieta sólida (cada litro aquoso de dieta contém: 137,5g de feijão carioca; 41,7g de levedo de cevada; 66g de germe de trigo; 4,16g de ácido ascórbico; 1,33g de ácido sórbico; 12,5g de nipagin; 17,1g de ágar; 10% v/v de formol). Ao atingirem o terceiro ínstar, as lagartas foram individualizadas e mantidas até o estágio de pupa a fim de evitar canibalismo.

4.2 Bioensaios de Alimentação

a) Bioensaio A: Avaliação do decaimento da detecção da presa em função das espécies de predador-presa em interação, modo de alimentação e fase do ciclo de vida do predador

Nesse bioensaio foram utilizados como predadores larvas (3º ínstar, 24 h em jejum) e adultos (após 48 h de emergência, em jejum) de *C. sanguinea* e larva (3º instar, 24 h em jejum) de *Ch. externa*. *Cycloneda sanguinea* é um predador mastigador e *Ch. externa* é um predador sugador. Para cada espécime do predador foi oferecido individualmente apenas um espécime de presa (ovo de *S. frugiperda* com 72 h pós postura (pré-eclosão), pulgão adulto de *M. persicae* ou ninfa de 4º ínstar de *B. tabaci*), sendo que para os adultos de joaninha foram utilizados cinco espécimes por sexo, por série temporal, conforme o Quadro 2.

Na Quadro 2 o número de replicatas foi maior ($n=12$) para as amostras indicadas com asterisco porque, ao invés de fazer um pool de 10 indivíduos, foram divididos em três grupos de quatro espécimes para construção de triplicatas de bibliotecas por amostra (teste de variabilidade entre bibliotecas/amostra). O número de bibliotecas para essas amostras foi 12 (e não nove), pois além do teste de variabilidade, também testou-se, em uma amostra, dois tamanhos de inserto das bibliotecas (350 e 550 pb).

Quadro 2. Delineamento experimental do decaimento da detecção das presas em função das espécies de predador-presa em interação, modo de alimentação e fase do ciclo de vida do predador.

Presas	Predador						Tempo (h) pós-predação	Extrações de DNA	Bibliotecas de DNA
	Larva <i>C. sanguinea</i>		Adulto <i>C. sanguinea</i>		Larva <i>Ch. externa</i>				
	<i>N</i>	ID	<i>N</i>	ID	<i>n</i>	ID			
Zero	10	BioA_01	10	BioA_16	10	BioA_31	-1	3	3
Ovo pré-eclosão de <i>S. frugiperda</i> (<i>n</i> = 1)	10	BioA_02	10	BioA_17	10	BioA_32	0	3	3
	10	BioA_03	10	BioA_18	10	BioA_33	3	3	3
	10	BioA_04	10	BioA_19	10	BioA_34	6	3	3
	10	BioA_05	10	BioA_20	10	BioA_35	9	3	3
	12*	BioA_06 BioA_07 BioA_08	12*	BioA_21 BioA_22 BioA_23	12*	BioA_36 BioA_37 BioA_38	0	9*	12*
Adulto áptero de <i>M. persicae</i> (<i>n</i> =1)	10	BioA_09	10	BioA_24	10	BioA_39	3	3	3
	10	BioA_10	10	BioA_25	10	BioA_40	6	3	3
	10	BioA_11	10	BioA_26	10	BioA_41	9	3	3
	10	BioA_12	10	BioA_27	10	BioA_42	0	3	3
Larva de último ínstar de <i>B. tabaci</i> (<i>n</i> =1)	10	BioA_13	10	BioA_28	10	BioA_43	3	3	3
	10	BioA_14	10	BioA_29	10	BioA_44	6	3	3
	10	BioA_15	10	BioA_30	10	BioA_45	9	3	3
Total	132		132		132		45	48	

T0= imediatamente pós-predação; T3= 3 h pós-predação; T6= 6 h pós-predação; T9= 9 h pós-predação; *n*= número de repetições (um espécime de presa/predador); ID= identificação da amostra.

b) Bioensaio B: Avaliação do decaimento da detecção da presa em função do número de presa consumida

Foram utilizadas como predador larvas de 3º ínstar de *Hi. convergens*. Para cada larva foi oferecida uma quantidade conhecida de pulgão adulto áptero de *M. persicae* (0, 1, 3 ou 6) em cada série temporal pós-predação, conforme Quadro 3. No Quadro 3, o número de replicatas foi maior (n=12) para a amostra indicada com asterisco porque, ao invés de fazer um pool de 10 indivíduos, foram divididos em três grupos de quatro espécimes para construção de triplicatas de bibliotecas por amostra (teste de variabilidade entre bibliotecas/amostra). O número de bibliotecas para essas amostras foi quatro (e não um), pois além do teste de variabilidade, foi testado também, em uma amostra, dois tamanhos de inserto das bibliotecas (350 e 550 pb).

Quadro 3. Delineamento experimental do decaimento da detecção da presa em função do número consumido.

<i>M. persicae</i> consumido	Larva <i>Hi. convergens</i>	Tempo (h) pós-predação	Extrações de DNA	Bibliotecas de DNA	ID
Zero	10	-1	1	1	BioB_46
n= 1	12*	0	3*	4*	BioB_47; BioB_48; BioB_49
	10	3	1	1	BioB_50
	10	6	1	1	BioB_51
	10	9	1	1	BioB_52
n= 3	10	0	1	1	BioB_53
	10	3	1	1	BioB_54
	10	6	1	1	BioB_55
	10	9	1	1	BioB_56
n= 6	10	0	1	1	BioB_57
	10	3	1	1	BioB_58
	10	6	1	1	BioB_59
	10	9	1	1	BioB_60
Total	132		15	16	

T0= imediatamente pós-predação; T3= 3 h pós-predação; T6= 6 h pós-predação; T9= 9 h pós-predação; ID= identificação da amostra.

c) Bioensaio C: Avaliação do decaimento da detecção da presa em predação secundária em função do número de presas consumidas e tempo pós-predação

Utilizou-se como presa secundária o pulgão adulto áptero *M. persicae*; como predador primário larva (3º ínstar) de *Ch. externa*; como predador secundário larva (3º ínstar) de *H. axyridis*. Foi oferecido individualmente às larvas de *H. axyridis* uma unidade de larva do crisopídeo, a qual se alimentou de diferentes quantidades de pulgão *M. persicae* (de 0 a 3). As larvas de crisopídeo foram oferecidas para predação pelas larvas de joaninha após terem decorrido 0, 3 e 6 h pós-predação do pulgão. Por sua vez as larvas de joaninhas alimentadas com larvas do crisopídeo também foram coletadas em série temporal pós-alimentação de até 6 h, conforme Quadro 4. No Quadro 4, o número de replicatas é maior (n=12) para a amostra com asterisco porque, ao invés de fazer um pool de 10 indivíduos, dividiu-se em três grupos de quatro espécimes para construção de triplicatas de bibliotecas por amostra (teste de variabilidade entre bibliotecas/amostra). Foram feitas quatro bibliotecas para essas amostras (e não uma), pois além do teste de variabilidade, também testou-se, em uma amostra, dois tamanhos de inserto das bibliotecas (350 e 550 pb).

Quadro 4. Delineamento experimental do decaimento da detecção das presas em predação secundária em função do número de presas e tempo pós-predação.

Presas	Larva <i>Ch. externa</i>	Larva <i>H. axyridis</i>	Tempo (h) pós-predação	Extrações de DNA	Bibliotecas de DNA	ID
Zero (sem pulgão e crisopídeo)	0	10	-1	1	1	BioC_77
1 pulgão, sem crisopídeo	0	12*	0	3*	4*	BioC_78 BioC_79 BioC_80
3 pulgões, sem crisopídeo	0	10	0	1	1	BioC_81
3 pulgões, sem crisopídeo	0	10	3	1	1	BioC_82
3 pulgões, sem crisopídeo	0	10	6	1	1	BioC_83
1 crisopídeo faminto	10	10	0	1	1	BioC_68
1 crisopídeo faminto	10	10	3	1	1	BioC_69
1 crisopídeo faminto	10	10	6	1	1	BioC_70
Pulgões <i>n</i> = 3 (crisopídeo T0)	12	10	0	1	1	BioC_71
Pulgões <i>n</i> = 3 (crisopídeo T0)	10	10	3	1	1	BioC_72
Pulgões <i>n</i> = 3 (crisopídeo T0)	10	10	6	1	1	BioC_73
Pulgões <i>n</i> = 3 (crisopídeo T3)	10	10	0	1	1	BioC_74
Pulgões <i>n</i> = 3 (crisopídeo T3)	10	10	3	1	1	BioC_75
Pulgões <i>n</i> = 3 (crisopídeo T3)	10	10	6	1	1	BioC_76
Pulgões <i>n</i> = 3 (crisopídeo T6)	10	10	0	1	1	BioC_76.1
Pulgões <i>n</i> = 3 (crisopídeo T6)	10	10	3	1	1	BioC_76.2
Pulgões <i>n</i> = 3 (crisopídeo T6)	10	10	6	1	1	BioC_76.2
Total	122	172		19	20	

T0= imediatamente pós-predação; T3= 3 h pós-predação; T6= 6 h pós-predação. Pulgões *M. persicae*; ID= identificação da amostra.

d) Bioensaio D: Avaliação do decaimento da detecção da presa em função da temperatura ambiente de digestão da presa

Utilizou-se como predador larva de 3º ínstar de *C. sanguinea*. Cada larva do predador se alimentou de um pulgão adulto áptero de *M. persicae*. As unidades experimentais foram divididas nas temperaturas ambiente de digestão da presa de 20, 25 e 30 °C. Para cada temperatura foram coletadas 10 larvas de joaninha em série temporal pós-predação de 3, 6 e 9 h e, como controle, foram coletadas larvas que permaneceram em jejum por 24 h (controle -1 h) e larvas coletadas logo após se alimentarem de um pulgão (controle 0 h) (Quadro 5). Os controles foram realizados a 25 °C.

Quadro 5. Delineamento experimental do decaimento da detecção das presas em função da temperatura ambiente de digestão da presa.

Temperatura de digestão da presa	Larva <i>C. sanguinea</i>	Tempo (h) pós-predação	Extrações de DNA	Biblioteca de DNA	ID
Zero	10	-1	*	*	BioA_1
<i>n</i> = 1	10	0	1	1	BioD_61
<i>n</i> = 1 (20 °C)	10	3	1	1	BioD_62
	10	6	1	1	BioD_63
	10	9	1	1	BioD_64
<i>n</i> = 1 (25 °C)	10	3	*	*	BioA_09
	10	6	*	*	BioA_10
	10	9	*	*	BioA_11
<i>n</i> = 1 (30 °C)	10	3	1	1	BioD_65
	10	6	1	1	BioD_66
	10	9	1	1	BioD_67
Total	110		7	7	

Pulgão = *M. persicae*; ID= identificação da amostra. As amostras indicadas com asterisco, foram obtidas do Bioensaio A.

4.3 Construção do banco de referência de DNA mitocondrial

Como quase todas as espécies utilizadas nesse estudo (*Ch. externa*, *C. sanguinea*, *Hi. convergens*, *H. axyridis*, *S. frugiperda*, *B. tabaci*, *M. persicae*) ainda não tinham o mitogenoma elucidado disponível em bancos de dados públicos, fez-se inicialmente a elucidação de seus genomas mitocondriais. Foram feitas extrações de DNA,

individualmente, de 25 mg de massa de cada uma das espécies. Os predadores foram mantidos em jejum de 24 h. O DNA total foi extraído utilizando-se o kit *DNeasy Blood and Tissue* (Qiagen®), de acordo com protocolo do fabricante, macerando-se as amostras com um bastão de vidro em microtubo (de 1,5 mL) contendo o primeiro tampão de extração. A quantidade de DNA obtida nas amostras foi quantificada utilizando-se o kit *dsDNA HS Assay* em *Qubit* (LifeTechnologies®), de acordo com protocolo do fabricante. A quantidade total de DNA por amostra foi normalizada para 1 µg e mantidas a -80 °C. As amostras foram liofilizadas em centrífuga a vácuo para envio para sequenciamento. As amostras foram utilizadas para construção de bibliotecas TruSeq Nano e sequenciamento na plataforma Illumina HiSeq2500 (250 pb, *paired-ended*, *insert size* entre 300 a 600 pb, 1 *lane*). O serviço de construção das bibliotecas de DNA e de sequenciamento foi terceirizado e realizado em parte no Centro Genômico da Universidade de Minnesota (EUA) e em parte no Centro de Biotecnologia Roy J. Carver Biotechnology da Universidade de Illinois (EUA).

Adaptou-se diferentes metodologias para a elucidação dos mitogenomas das espécies-alvo, a qual foi comparada com outras citadas na literatura (Timbó *et al.*, 2017, Anexo 1). A checagem da qualidade dos dados de sequenciamento (formato FASTQ) de cada biblioteca foi realizada utilizando-se o programa FastQC (v.0.11.3) (Andrews, 2010). Quando necessário houve remoção de sequências super-representadas ou index de adaptadores das bibliotecas Illumina utilizando-se os programas Fastqc-mcf (v.1.04.807) (Aronesty, 2011) e Cutadapt (v.1.9.1) (Martin, 2011). A montagem dos mitogenomas foi realizada com o programa MITObim (v.1.8) (Hahn *et al.*, 2013) utilizando mitogenomas de referência conforme indicado no Quadro 6. As anotações foram realizadas pelo programa MITOS (v.763) (Bernt *et al.*, 2013), seguida da curadoria manual (Anexo 3 no programa Geneious (v.9.0.5) (Kearse *et al.*, 2012). Os

mitogenomas anotados foram depositados no *GenBank* utilizando-se a plataforma BankIt (vide código *GenBank* na Tabela 1).

Quadro 6. Mitogenomas utilizados como referência para a elucidação dos mitogenomas das espécies-alvo pelo programa MITObim (v.1.8).

Mitogenoma a elucidar	Mitogenoma de Referência	Código <i>GenBank</i> do mitogenoma de Referência	Tamanho (pb)
<i>B. tabaci</i>	<i>Aleurodicus dugesii</i>	NC_005939.1	15723
<i>Ch. externa</i>	<i>Chrysoperla nipponensis</i>	NC_015093.1	16057
<i>C. sanguinea</i>	<i>Coccinella septempunctata</i>	JQ321839.1	18965
<i>H. axyridis</i>	<i>Coccinella septempunctata</i>	JQ321839.1	18965
<i>Hi. convergens</i>	<i>Coccinella septempunctata</i>	JQ321839.1	18965
<i>M. persicae</i>	<i>Acyrtosiphon pisum</i>	NC_011594.1	16971
<i>S. frugiperda</i>	<i>Spodoptera litura</i>	NC_022676.1	15388

Para a montagem do banco de referência de DNA mitocondrial de insetos visando as análises de detecção das presas consumidas pelos predadores nos bioensaios de alimentação, foram utilizados todos os mitogenomas de insetos disponíveis no *GenBank* (na época, 829 espécies), conforme Paula *et al.* (2015), metodologia descrita no Anexo 3, e os sete mitogenomas sequenciados e montados neste trabalho. O banco de referência final ficou constituído de 836 sequências no formato FASTA.

4.4 Detecção das presas no conteúdo gastrointestinal dos predadores

a) Preparação das amostras para sequenciamento

Os tratos gastrointestinais dos predadores foram dissecados em placas de petri, com auxílio de pinças e tesouras entomológicas e de lupa estereoscópica (Figura 9). Todo o material utilizado (pinças e tesouras entomológicas, placas de petri) foi previamente lavado com sabão Extran (Merck®), incubado em solução de hipoclorito de sódio 0,3% por 10 min e transferido para solução de etanol 70%, permanecendo por 10 min. Em seguida, foram acondicionados em microtubos (de 1,5 mL) contendo o tampão de lise indicado pelo kit *DNeasy Blood and Tissue* (Qiagen®), mantidos em

gelo durante as dissecções. A finalidade da dissecção consiste em reduzir a quantidade de material genético do predador nas amostras e, assim, aumentar as chances de detecção das presas. Os gastrointestinos dissecados foram armazenados em *pool* por tratamento/bioensaio. Ao todo foram realizadas 810 dissecções.

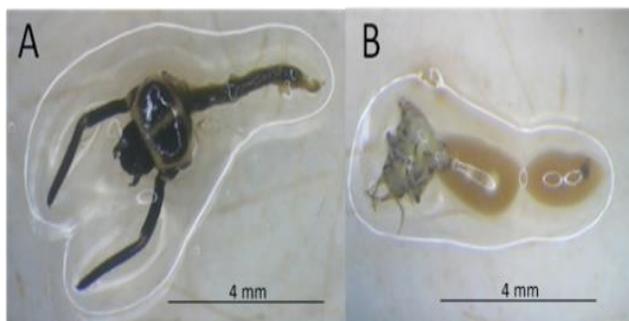


Figura 9. Aparelho gastrointestinal dos predadores dissecados, visualizados em aumento de 12 ou 16x em lupa estereoscópica. A - De larva de 3º ínstar da joaninha *C. sanguinea*; B - De larva de 3º ínstar do crisopídeo *Ch. externa*; Como etapa final, as cabeças foram removidas antes da armazenagem dos gastrointestino.

A extração do DNA total do *pool* de gastrointestinos por tratamento/bioensaio foi realizada utilizando-se o kit *DNeasy Blood and Tissue* (Qiagen®) de acordo com o protocolo do fabricante, com alteração apenas na etapa de recuperação do DNA, em que foram feitas duas lavagens com 106 µL com o tampão AE. A qualidade do DNA total obtido foi avaliada por eletroforese em gel de agarose 1%. A quantidade de material genético obtida foi estimada por fluorescência utilizando-se o kit *dsDNA HS Assay* em *Qubit* (LifeTechnologies®), de acordo com protocolo do fabricante. As amostras foram estocadas a -80 °C. A quantidade total de DNA por amostra foi normalizada para 1 µg e liofilizados em centrífuga a vácuo para envio para sequenciamento. As amostras de cada tratamento/bioensaio foram utilizadas para construção de bibliotecas TruSeq Nano e sequenciamento na plataforma Illumina HiSeq2500 (250 pb, *paired-ended*, *insert size* entre 350 e 550 pb, 2 *lanes*). Para as bibliotecas com tamanho de inserto de 350 pb, 100 ng de DNA por amostra foi usado para construir a biblioteca; e para as bibliotecas com tamanho de inserto de 550 pb, 200 ng de DNA por amostra foi usado para construir a

biblioteca. Essa quantidade de DNA utilizada para construir as bibliotecas de acordo com o tamanho do inserto é dada pela Illumina (<http://www.illumina.com/products/truseq-nano-dna-library-prep-kit.html>). O serviço de construção das bibliotecas de DNA e de sequenciamento foi terceirizado e realizado no Centro Genômico da Universidade de Minnesota (EUA) para o Bioensaio A e no Centro de Biotecnologia Roy J. Carver Biotechnology da Universidade de Illinois (EUA) para os Bioensaios B, C e D.

b) Análises de bioinformática

As análises de bioinformática foram realizadas no Laboratório de Bioinformática da Embrapa Recursos Genéticos e Biotecnologia. Todos os scripts e comandos estão apresentados no Anexo 4. Os dados brutos do sequenciamento de cada biblioteca foram depositados no repositório SRA do *GenBank* (Anexo 5). Os dados de sequenciamento (formato FASTQ) de cada biblioteca foram checados quanto a qualidade por meio do programa FastQC (v.0.11.3) (Andrews, 2010). Caso necessário, sequências de baixa qualidade ou adaptadores das bibliotecas Illumina foram removidos pelos programas Fastqc-mcf (v.1.04.807) (Aronesty, 2011) e Cutadapt (v.1.9.1) (Martin, 2011). A seguir, utilizando o programa SeqTK (v1.2) (<https://github.com/lh3/seqtk>), as *reads* resultantes de cada biblioteca foram convertidas para formato FASTA. As identificações taxonômicas dos mitogenomas de referência foram obtidas do *GenBank* por meio do link ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/gi_taxid_nucl.dmp.gz (Srivathsan *et al.*, 2015). Os dados de cada biblioteca dos bioensaios (agora no formato FASTA) foram então alinhados contra o banco de dados de DNA mitocondrial utilizando o programa o BLASTn (v2.2.27+) (Altschul *et al.*, 1997), utilizando os parâmetros: E-value <1e-5; maximum target sequences 3; no dust. A atribuição taxonômica dos matches obtidos no

alinhamento no BLASTn foi feita por meio de comando customizado em linguagem Python desenvolvido por Srivathsan *et al.* (2015).

A eliminação de falsos-positivos, por biblioteca, foi realizada inicialmente pela filtragem dos resultados (outputs) dos alinhamentos que não tivessem porcentagem de identidade $\geq 98\%$ em ≥ 150 pb utilizando os programas `parse_by_id.py` e `masterfile_tax.py` (Srivathsan *et al.*, 2015). Para a determinação destes parâmetros de filtragem por porcentagem de identidade e de tamanho sobreposição dos matches fez-se um estudo detalhado, o qual foi submetido para publicação e está apresentado no Anexo 6. A segunda etapa para continuar a eliminar falsos-positivos ainda remanescentes foi verificar se um mesmo *read* apresentou matches com mitogenomas de diferentes espécies. Verificando se para cada match de uma *read* R1 existe um match para a *read* R2 correspondente, sendo para isso foi utilizado um *script* escrito na linguagem de programação perl. A terceira etapa da eliminação de falsos-positivos consistiu no mapeamento (*mapping*) das *reads* relativas a esses *matches* nos mitogenomas de referência. *Reads* mapeadas em regiões não conservadas interespecies (p.ex. tRNA, rRNA ou região controle do mtDNA) foram eliminadas, pois essas regiões não conferem resolução taxonômica para identificação das presas consumidas. Para o mapeamento, as sequencias FASTA foram recuperadas a partir da lista de *reads* dos matches remanescentes das etapas anteriores (Anexo 4), utilizando-se os programas `cdbfasta` (<https://sourceforge.net/projects/cdbfasta/>) e `cdbyank` (<http://nebc.nerc.ac.uk/bioinformatics/docs/cdbyank.html>). Esse arquivo FASTA foi utilizado no programa Geneious (v.9.0.5), onde o mapeamento (*Map to Reference*) foi executado de acordo com os parâmetros: *custom sensitivity; do not trim; save list of used reads; save list of unused reads; not allow gaps; minimum overlap 150 bp; minimum identity 98%; not search more thoroughly for poor matching reads.*

Após eliminação dos falsos-positivos, o número restante de *reads* por táxon foi contabilizado e fez-se a análise estatística para comparar os resultados de detecção das presas obtidos para as bibliotecas de cada bioensaio.

4.5 Análises estatísticas

Para os bioensaios A, B e C o número inicial de *reads* detectadas do predador (n_0) e a taxa de decaimento (d) foram estimados como processo de primeira ordem de decaimento (ou taxa constante de decaimento) como descrito a seguir. As análises estatísticas foram realizadas pelo co-orientador professor David Andow.

4.5.1 Bioensaio A: Avaliação do decaimento da detecção da presa em função das espécies de predador-presa em interação, modo de alimentação e fase do ciclo de vida do predador

O número de *reads* brutas após o controle de qualidade foram filtradas de acordo com os métodos descritos por Paula *et al.* 2017 (manuscrito submetido a revista *Molecular Ecology Resources* – Anexo 6). Resumidamente, as *reads* que correspondem à base de dados de mtDNA da presa foram filtradas para sobreposição em pelo menos 150 pb e 98% de identidade. As *reads* filtradas foram mapeadas no mtDNA correspondente e, as *reads* não mapeadas ou mapeadas em regiões não codificantes, assim como *reads single-end* e *reads* que apresentaram match com mais de uma espécie, foram descartadas. As *reads* restantes apresentaram mapeamento em regiões codificantes do mtDNA. O número inicial de *reads* das presas detectadas nos predadores (n_0) e a taxa de decaimento de *reads* da presa (d) foram estimados como processo de decaimento de primeira ordem (ou taxa de decaimento constante) da seguinte maneira.

O número de *reads* das presas observadas em uma biblioteca foi assumido como uma amostra aleatória (Poisson) das *reads* de todo o conteúdo intestinal a partir do qual a biblioteca foi criada. Tais dados foram utilizados para estimar a distribuição Gamma posterior do número de *reads* das bibliotecas, utilizando o método Bayesiano. Cem mil valores aleatórios, dessa distribuição, foram sorteados e corrigidos para toda a amostra do conteúdo gastrointestinal, normalizando a cobertura da biblioteca após o controle de qualidade, seguida pela multiplicação pela média da cobertura para fins de apresentação e transformação por \ln . Esses valores foram usados para calcular a média e a variância da distribuição de \ln -transformado observado no número de *reads*. Em seguida, para cada tratamento, calculou-se uma regressão linear ponderada, utilizando o tempo pós-predação para prever a média da \ln das *reads* observadas, com $1/\text{variância}$ como peso. A interceptação da regressão é o n_0 , o número inicial de *reads* da presa presente no predador e a inclinação é a taxa de decaimento (d). Os intervalos de confiabilidade (IC) de 95% para esses parâmetros foram estimados a partir dos resultados da regressão. A meia-vida de detecção das *reads* das presas foram estimadas utilizando $1/d \cdot D_{max}$. O período de detecção foi estimado com a simulação de Monte Carlo. A distribuição Gamma posterior foi utilizada para a simulação experimental dos dados. A interceptação (b) e a curva dessa regressão (m) da regressão foram utilizadas para estimar o D_{max} , que seria o tempo de uma leitura ($=-b/m$). A simulação foi repetida 200 mil vezes para gerar a distribuição D_{max} , a média e os IC 95% foram calculados a partir da distribuição. Todos os cálculos foram realizados utilizando o programa *Mathematica* 8.0 (Wolfram, 2010).

Para determinar as melhores estimativas de parâmetros para o número de *reads* inicial (n_0) e a taxa de decaimento (d) para todo o conjunto de dados, todos os modelos possíveis foram ajustados utilizando a regressão linear ponderada nos números de *reads*

estimados no conteúdo gastrointestinal dos predadores, permitindo que o número inicial de *reads* dependa do tipo de predador ou das espécies de presas (ou de nenhum dos dois), e permitindo que a taxa de decaimento dependa de forma semelhante entre o tipo de predador e as espécies de presas (ou nenhum dos dois ou ambos). Foi utilizado o Critério de Informação Akaike corrigido (AIC_c) para eliminar modelos com probabilidade relativa <0,05, e calculou-se a média dos modelos restantes usando a probabilidade relativa como pesos. As análises foram realizadas no SAS 9.4 (2014).

4.5.2 Bioensaio B: Avaliação do decaimento da detecção da presa em função do número de presa consumida

A hipótese de que o número de *reads* da presa presente no predador seja proporcional à quantidade de presa consumida foi testada por comparação de dois modelos de AIC_c. Em ambos modelos foi assumido que a taxa de decaimento foi a mesma para todos os tratamentos, o que foi confirmado após análises descritas no parágrafo anterior. O Primeiro Modelo assumiu que o número de *reads* das presas foi diferente nos três tratamentos de número de presa (1, 3 ou 6 presas consumidas), mas não necessariamente proporcional a quantidade de presa consumida. Para esse modelo utilizou-se o tratamento por presa como categoria variável. O Segundo Modelo assumiu que o número de *reads* das presas era proporcional ao número de presas consumidas. Especificamente, se havia três vezes mais presas consumidas então o Segundo Modelo assumia que haveria três vezes mais *reads* detectadas em cada ponto de tempo. Se a hipótese fosse falsa, o Segundo Modelo apresentaria dados significativamente piores que o Primeiro Modelo. Esse modelo assumiu que o número de *reads* seria maior em $\ln-3$ e $\ln-6$ do que nos tratamentos com 1, 3 e 6 *M. persicae*, respectivamente. Como os dados foram transformados por \ln , o número previstos de *reads* pode ser devido a função linear desses valores. As análises foram realizadas no SAS 9.4 (SAS, 2014).

4.5.3 Bioensaio C: Avaliação do decaimento da detecção da presa em predação secundária em função do número de presas consumidas e tempo pós-predação

Para testar se o número de *reads* das presas secundária foi o mesmo entre a predação primária e secundária, comparou-se as bibliotecas onde *H. axyridis* alimentaram diretamente de *M. persicae* com as bibliotecas onde *H. axyridis* alimentaram de *Ch. externa* que havia previamente se alimentado de *M. persicae* (Wolfram, 2010). Nesses casos, a única diferença foi a de que ou o *M. persicae* foi presa primária ou secundária de *H. axyridis*. As *reads* de *M. persicae* transformadas por *ln* foram comparadas por ANOVA, utilizando as variâncias agrupadas para estimar o quadrado médio de erro.

Para testar se a taxa de decaimento de presas secundária no predador intraguildda foi o mesmo, seja diretamente ou indiretamente ingerido pelo predador secundário, vários modelos de regressão foram ajustados, usando os dados do parágrafo anterior. Permitiu-se que a intercepção e a inclinação dependessem da existência de predação primária ou secundária (ou não) em todas as combinações possíveis para quatro modelos diferentes. O AIC_c não pôde ser calculado para o modelo completo com intercepção e inclinação estimados separadamente para predação primária e secundária porque haviam muitos parâmetros no modelo, o que fez com que tal modelo não fosse desconsiderado.

Para testar se a taxa de decaimento da presa secundária consumido pelo predador primário foi a mesma independentemente do período de digestão no predador secundário, utilizou-se as nove bibliotecas com três períodos de tempo para *H. axyridis*, cada um com três períodos de tempo para *Ch. externa*. Utilizou-se o período de tempo de *H. axyridis* como um fator categórico (HaTime), para ajustar quatro modelos de

regressão. Os modelos permitiram que a interceptação e a inclinação dependessem de HaTime (ou não).

Para testar se a taxa de decaimento de presas secundárias no predador secundário foi a mesma independentemente do período de digestão no predador primário, utilizou-se as mesmas nove bibliotecas descritas no parágrafo anterior, entretanto desta vez utilizou-se os períodos de tempo de *Ch. externa* como um fator categórico (CeTime), para caber quatro modelos de regressão. Os modelos permitiram que a interceptação e a inclinação dependessem de CeTime (ou não).

Para determinar se a taxa de decaimento de presas secundárias foi a mesma no predador secundário e no primário, comparou-se as taxas de decaimento estimadas das duas análises anteriores, usando um teste *t* nas taxas de decaimento estimadas e erros-padrão. Como os erros-padrão apresentaram semelhanças, esses foram reunidos.

Para testar se a taxa de decaimento de presas secundária apresentava dependência de alguma interação complexa entre o tempo de digestão no predador secundário e o tempo de digestão no predador primário, foram utilizados modelos que assumiram que a taxa de decaimento era uma função complexa de ambos. No entanto, devido aos resultados das análises anteriores, não foram considerados todos os modelos possíveis onde poderia haver seis taxas únicas de decaimento. Em vez disso, foram considerados quatro modelos com um único valor de interseção e com inclinação que dependiam de HaTime, CeTime ou ambos. Além disso, foi considerado um modelo onde a taxa de decaimento de presas secundária dependia apenas do tempo total de digestão, somando o HaTime e o CeTime.

Para testar se o decaimento da presa primária no predador intraguilda dependia da alimentação prévia da presa secundária, foram comparadas a decaimento das *reads* do predador primário em 12 bibliotecas, sendo que nove com três períodos de tempo para

H. axyridis cada um com três períodos de tempo para *Ch. externa* que havia se alimentado de *M. persicae* e três em que *H. axyridis* havia consumido larvas de *Ch. externa* em jejum. Foi utilizado como fator categórico CeTime e quatro modelos de regressão foram ajustados, o que permitiu que a interceptação e a inclinação dependessem de CeTime (ou não).

Finalmente, para testar se o decaimento da presa secundária e do predador primário foi a mesma no predador secundário, as taxas de decaimento estimadas foram comparadas usando um teste *t* e a correlação de Pearson entre as *reads* do predador primário e da presa secundária foi calculada. As análises foram realizadas no SAS 9.4 (SAS, 2014).

4.5.4 Bioensaio D: avaliação do decaimento da detecção da presa em função da temperatura ambiente de digestão da presa

Nesse bioensaio pretendeu-se avaliar se a temperatura do ambiente poderia influenciar no decaimento da detecção da presa pelo fator Boltzmann ou pela relação de Van't Hoff-Arrhenius. A normalização dos dados foi feita com 95% de CIs. A sobreposição dos CIs indicam que não há diferença significativa entre os números de *reads* detectadas.

4.5.5 Análise estatística sobre a variação do número de indivíduos agrupados numa mesma biblioteca e tamanho do inserto da biblioteca

Efeito do agrupamento e variação do número de indivíduos na construção da biblioteca: Cada uma das bibliotecas de 350 pb foi formada por quatro predadores agrupados, de modo que cada uma destas bibliotecas era uma estimativa de $\mu_{ind} \pm \sigma_{ind} / 4$ (1/2). Assim, o desvio padrão destas três bibliotecas, $s_{350} = \sigma_{ind} / 4$ (1/2) e assim σ_{ind} pôde ser estimado. Utilizamos a simulação de Monte Carlo para estimar o s_{350}

(desvio padrão). Os números observados de *reads* das presas foram utilizados para estimar a distribuição de probabilidade de possíveis números de *reads* observadas usando métodos Bayesianos com um Jeffries anterior e normalizado à profundidade média da biblioteca após o *quality control*, $4,1 \times 10^6$ de *reads*. As distribuições posteriores foram usadas para gerar aleatoriamente números de *reads* para cada biblioteca 2×10^6 vezes, que foram ln-transformadas. A média, m350 e s350, para as três bibliotecas de 350 pb foram calculadas para cada aleatorização, e a dependência entre elas foi examinada. Σ ind foi estimado a partir da média calculada para cada simulação. Calculou-se, separadamente para os cinco tipos de predadores alimentados com *M. persicae* no tempo de 0 h, a relação funcional entre o número de indivíduos utilizados na construção do *pool* e o erro do número de *reads* estimado.

Efeito do tamanho do inserto (350 e 550 bp) usado nas bibliotecas na detecção das *reads* das presas. Para o tamanho do inserto de 350 pb, 100 ng de amostra foram sequenciados, e para 550 pb, foram sequenciados 200 ng de amostra. A amostra de biblioteca de 550 pb foi reunida a partir das três amostras de 350 pb e para a junção foi adicionada a mesma concentração de cada uma das três amostras. Assim, se a detecção das *reads* fosse igual nas duas bibliotecas, seria esperado duas vezes mais *reads* nas bibliotecas de 550 pb como nas bibliotecas de 350 pb. Isto foi testado usando um teste de permutação. Os números observados de *reads* de presas foram utilizados para estimar a distribuição de probabilidade de possíveis números de *reads* observados usando métodos Bayesianos com um Jeffries anterior e normalizado à profundidade média da biblioteca após *quality control* de $4,1 \times 10^6$ *reads*. As distribuições posteriores foram usadas para gerar aleatoriamente números de *reads* para cada biblioteca. Calculou-se a média, m350, e desvio padrão, s350, para as três bibliotecas de 350 pb. A distribuição de $(2 * m350 - m550) / s350$ foi compilada a partir de 2×10^5 interações e

utilizada para calcular a probabilidade de se detectar o dobro de *reads* nas bibliotecas de 550 pb como nas bibliotecas de 350 pb. A distribuição de $(m_{350}-m_{550}) / s_{350}$ foi utilizada para calcular a probabilidade de que mais *reads* fossem detectadas na biblioteca de 550 pb.

5. RESULTADOS E DISCUSSÃO

5.1 Elucidação dos mitogenomas

Após a análise de controle de qualidade e retirada das sequências super-expressas e dos adaptadores, apenas 0,09% (em média) das *reads* foram eliminadas (Tabela 1). Os mitogenomas das espécies de insetos utilizados nesse trabalho apresentaram dois RNAs ribossomais (*rrnL* – subunidade maior 16S e *rrnS* – subunidade menor 12S), 22 genes de RNA transportador (tRNA), 13 genes codificadores de proteína, com um tamanho variando entre 15 e 19 mil pares de bases (Tabela 2) (Figura 10). De acordo com Li *et al.* (2011) o mitogenoma de insetos apresenta entre 14 e 20 kb de tamanho e o mesmo número de genes encontrados aqui.

Tabela 1. Número de *reads* obtidas e remanescentes após controle de qualidade (CQ) das bibliotecas TruSeq Nano para sequenciamento por Illumina HiSeq2500 (250 pb, paired-end, tamanho do inserto de 550 pb, 1 lane) das amostras para elucidação dos mitogenomas dos insetos de interesse.

Amostra	# <i>Reads</i> obtidas	# <i>Reads</i> após CQ
<i>Bemisia tabaci</i>	1.707,954	1.706,373
<i>Chrysoperla externa</i>	6.215,444	6.211,816
<i>Cycloneda sanguinea</i>	4.587,115	4.569,238
<i>Myzus persicae</i>	1.860,816	1.859,633
<i>Harmonia axyridis</i>	3.032,949	3.032,949
<i>Hippodamia convergens</i>	2.827,085	2.827,085
<i>Spodoptera frugiperda</i>	1.608,864	1.608,404

Tabela 2. Características dos mitogenomas elucidados.

Amostra	Tamanho (pb)	rRNA	tRNA	ptn	Código GenBank
<i>B. tabaci</i>	16207	2	22	13	KU877168
<i>Ch. externa</i>	16510	2	22	13	KU877169
<i>C. sanguinea</i>	18713	2	22	13	KU877170
<i>H. axyridis</i>	19330	2	22	13	KU877028.1
<i>Hi. convergens</i>	18419	2	22	13	KU877030
<i>M. persicae</i>	17099	2	22	13	KU877171
<i>S. frugiperda</i>	16346	2	22	13	KU877172

rRNA: RNA ribossomal (subunidades maior e menor); tRNA: RNA transportador; ptn: genes codificadores de proteína.

G

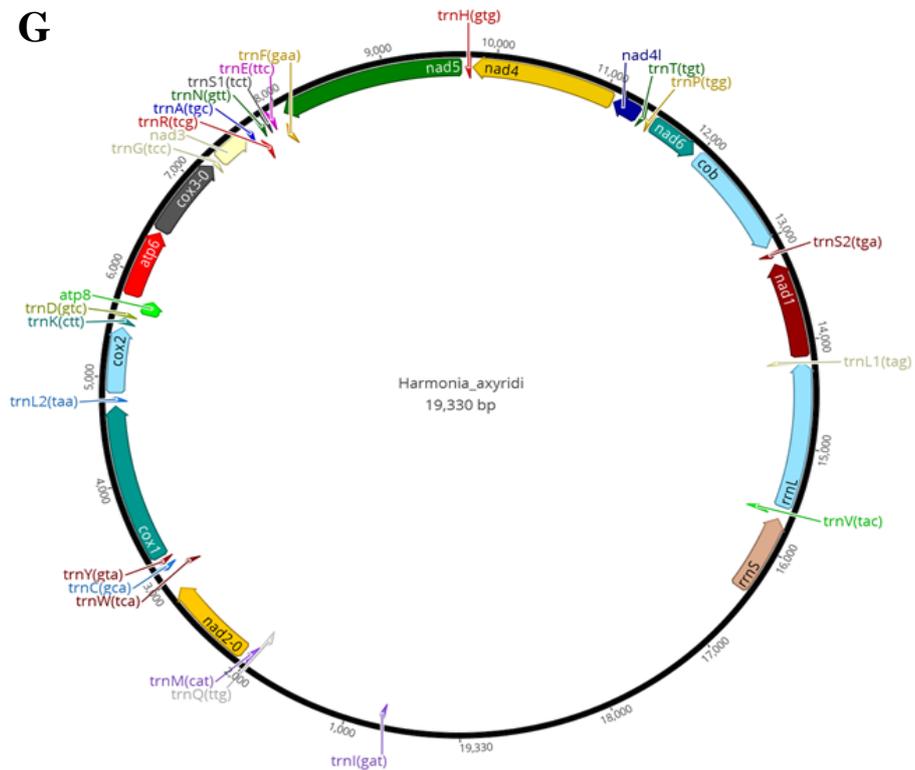


Figura 10. Mitogenomas elucidados e anotados: A - *Bemisia tabaci*; B - *Chrysoperla externa*; C - *Cycloneda sanguinea*; D - *Hippodamia convergens*; E - *Myzus persicae*; F - *Spodoptera frugiperda*; G - *Harmonia axyridis*.

5.2 Bioensaios de alimentação

Previamente aos bioensaios de alimentação, fez-se ensaios pilotos para determinar:

- a) o tempo de jejum para as larvas de joaninha *C. sanguinea* (14 ou 24 h) e do crisopídeo (24 ou 48 h);
- b) o tempo médio que adultos da joaninha *C. sanguinea*, recém-emergidos, iniciam a alimentação (24 ou 48 h);
- c) a viabilidade de larvas de 3º ínstar de joaninha *C. sanguinea*, *H. axyridis* e *Hi. convergens* predarem, em menos de 1 h, larva de 3º ínstar de crisopídeo;
- d) verificar o número máximo médio de pulgões que larvas de crisopídeos de 3º ínstar podem se alimentar em uma hora.

A metodologia e os resultados desses bioensaios pilotos estão descritos no Anexo 2. Foram determinados:

- a) o tempo de jejum de 24 h para larvas de 3º ínstar da joaninha *C. sanguinea* e do crisopídeo;
- b) o tempo de 48 h de jejum para os adultos recém-emergidos de joaninhas *C. sanguinea*;
- c) larva de 3º ínstar de *H. axyridis*, pois foi a única espécie de joaninha capaz de preda, no tempo máximo de 1 h, larva de 3º ínstar do crisopídeo. As demais espécies de joaninha foram predadas pela larva de 3º ínstar do crisopídeo;
- d) larvas de crisopídeo de 3º ínstar é capaz de se alimentar de oito pulgões adultos ápteros no prazo de 1 hora.

Ao todo foram realizados quatro bioensaios de alimentação para avaliação da meia-vida (T_{50}) de detecção do DNA das presas em função:

- a) das espécies do predador e da presa em interação, modo de alimentação e da fase do ciclo de vida do predador (Bioensaio A);
- b) do número de presas consumidas (Bioensaio B);
- c) da predação secundária (Bioensaio C);
- d) da temperatura ambiente de digestão da presa (Bioensaio D).

A metodologia de cada bioensaio está descrita nos tópicos subsequentes. Os bioensaios A, B e C foram feitos com joaninhas de espécies diferentes para que, ao final, os dados de detecção das presas entre as espécies de coccinelídeos fossem comparados. Todos os bioensaios foram realizados a 25°C e 16 h de fotoperíodo em câmaras climatizadas (BOD), a não ser quando diferentemente especificado. Para todos os bioensaios, os tempos de avaliação foram: antes da alimentação (predador faminto, controle negativo), denominado de -1 h; imediatamente após alimentação (0 h); 3, 6 ou 9 horas após alimentação. Para cada série temporal foram coletados 10 espécimes do predador e imediatamente armazenados em etanol 95% a -20°C. As larvas dos predadores foram utilizadas no 3º instar por apresentarem maior capacidade predatória. Os adultos dos predadores foram utilizados 48 h após emergência por conterem o conteúdo gastrointestinal naturalmente vazio. Para realização dos bioensaios de alimentação, as presas foram diretamente oferecidas ao predador para minimizar a variação do tempo de procura da presa entre as unidades experimentais.

5.3 Detecção das presas em função das espécies de predador-presa em interação, modo de alimentação e fase do ciclo de vida do predador

Neste bioensaio investigou-se se o tempo de detecção de determinada presa, pelo método de sequenciamento direto de alto desempenho de DNA do conteúdo

gastrointestinal do predador, é influenciado pela espécie do predador, pelo seu modo de alimentação e pela fase de seu ciclo de vida (imatura x adulta), bem como se o decaimento da detecção da presa é função da espécie de presa consumida, ou seja, se uma espécie de presa estaria mais sujeita a ser detectada por maior tempo após predação do que outra espécie. Também foram avaliados: se houve a detecção de falsos-negativos (presas consumidas, mas não detectadas) e de falsos-positivos (presas não consumidas, mas detectadas); a variabilidade na detecção (número de *reads*) de dada presa/predador entre bibliotecas de DNA do mesmo tratamento; se o tamanho do inserto utilizado para construir as bibliotecas de DNA influencia na detecção (número de *reads* e duração da detecção).

O número total de *reads* por biblioteca após controle de qualidade e a detecção das presas constam na Tabela 3. De maneira geral, pôde-se notar um número muito inferior de *reads* sendo detectadas para as presas do que para os predadores, ainda que os conteúdos gastrointestinais dos predadores tenham sido dissecados para eliminar extração de DNA de outros tecidos do corpo. Nenhuma *read* de presas ou de falsos-positivos foi detectada em qualquer das bibliotecas utilizadas como controles. A presa mais detectada foi o pulgão *M. persicae*. A quantidade de *reads* detectada por presa/predador reduz com o tempo de digestão. O predador com maior detecção de presa foi o *C. sanguinea*, onde a detecção foi maior na fase adulta (mastigadora) que na larval (sugadora). Houve baixa detecção de presa no predador *Ch. externa*, independentemente do tipo de presa consumida. Greenstone *et al.* (2007) em estudo utilizando a técnica de PCR para avaliar o modo de alimentação dos predadores *Co. maculata* (mastigador) e *P. maculiventris* (sugador) na detecção da presa *L. decemlineata*, relataram que a meia-vida de detecção da presa no conteúdo gastrointestinal dos predadores foi de 7 h e 50 h, respectivamente. Diferentemente do

predador *Ch. externa*, o predador *L. decemlineata*, mesmo sendo um inseto sugador, apenas inicia a digestão da presa quando o conteúdo é completamente absorvido (Cohen, 1990; Greenstone *et al.* 2007). De um total de 45 bibliotecas, houve detecção de falsos-positivos em 37 delas, sendo de até três espécies falsamente detectadas. A diferenciação entre os tipos de falsos-positivos foi amplamente discutida em Paula *et al.* (2017; Anexo 6). A variabilidade e o tamanho do inserto serão discutidos em tópico separado, pois foi realizada uma análise integrada dos dados de todas as bibliotecas, independente do bioensaio.

Tabela 3. Número de reads detectadas para predadores, presas e falsos-positivos no Bioensaio A.

Predador <i>Cycloneda sanguinea</i>, larva de 3º instar															
Biblioteca	# Reads pós CQ	Presa	# Presa consumida	Tempo pós-predação (h)	Tamanho do inserto (pb)	# Reads do predador	# Reads da presa	<i>Aphis gossypii</i>	<i>Chrysoperla externa</i>	<i>Ephesia kuehniella</i>	<i>Hippodamia convergens</i>	<i>Myzus persicae</i>	<i>Spodoptera litura</i>	# Espécies falso positivas	# Reads falso positivas
BioA1	4760877	-	0	-1	350	35342	0	0	0	0	0	0	0	0	0
BioA2	4899374	Sp	1	0	350	40034	18	0	0	0	0	2	0	1	2
BioA3	5106868	Sp	1	3	350	55658	0	0	2	0	0	2	0	2	4
BioA4	4874681	Sp	1	6	350	52856	0	0	0	0	0	0	0	0	0
BioA5	5005875	Sp	1	9	350	53382	0	0	0	0	0	0	0	0	0
BioA6	2163276	Mp	1	0	350	12844	30	0	8	0	0	0	0	1	8
BioA7	5028655	Mp	1	0	350	33900	60	0	0	0	0	0	0	0	0
BioA8	4673168	Mp	1	0	350	29808	218	0	0	0	2	0	0	1	2
BioA6B	2186703	Mp	1	0	550	13674	106	0	0	2	0	0	0	2	26
BioA9	4797188	Mp	1	3	350	42578	4	2	0	0	0	0	0	1	2
BioA10	4975496	Mp	1	6	350	44144	0	0	0	0	0	0	0	0	0
BioA11	4888879	Mp	1	9	350	40764	0	0	0	0	4	0	0	1	4
BioA12	4702259	Bt	1	0	350	32492	0	0	0	0	0	0	0	0	0
BioA13	4778075	Bt	1	3	350	31430	0	0	0	2	0	4	0	2	6
BioA14	4923251	Bt	1	6	350	26080	0	0	0	0	0	2	0	1	2
BioA15	4716568	Bt	1	9	350	41806	0	0	0	2	0	4	0	1	6

Predador *Cycloneda sanguinea*, adulto

Biblioteca	# Reads pós CQ	Presa	# Presa consumida	Tempo pós-predação (h)	Tamanho do inserto (pb)	# Reads do predador	# Reads da presa	<i>Aphis gossypii</i>	<i>Chrysoperla externa</i>	<i>Ephesia kuehniella</i>	<i>Harmonia axyridis</i>	<i>Hippodamia convergens</i>	<i>Myzus persicae</i>	<i>Spodoptera litura</i>	# Espécies falso positivas	# Reads falso positivos
BioA16	4871206	-	0	-1	350	32104	0	0	0	0	0	2	4	0	1	6
BioA17	4342471	Sp	1	0	350	30314	352	0	0	0	0	0	8	0	1	8
BioA18	4608944	Sp	1	3	350	24186	254	0	2	0	0	0	0	0	1	0
BioA19	4592446	Sp	1	6	350	31438	74	0	0	0	0	0	0	0	0	2
BioA20	48289721	Sp	1	9	350	383122	580	4	8	2	0	2	2	2	6	20
BioA21	4871758	Mp	1	0	350	33716	1224	0	0	0	0	0	0	0	0	0
BioA22	4852099	Mp	1	0	350	23220	1688	0	0	0	0	0	0	0	0	0
BioA23	4352190	Mp	1	0	350	11008	352	0	0	0	0	19882	0	0	1	19882
BioA21B	4871758	Mp	1	0	550	6472	350	0	8	0	0	4700	0	0	2	4708
BioA24	4658501	Mp	1	3	350	46932	10	0	12	0	0	0	0	0	1	6
BioA25	6589879	Mp	1	6	350	50366	2	0	2	2	2	0	0	0	1	12
BioA26	5658045	Mp	1	9	350	33298	0	0	6	0	0	0	0	0	0	6
BioA27	5404718	Bt	1	0	350	50542	0	0	4	0	0	0	4	0	1	8
BioA28	5401794	Bt	1	3	350	36284	0	0	6	0	0	0	0	0	1	6
BioA29	5331194	Bt	1	6	350	34710	0	0	10	42	0	0	0	0	2	52
BioA30	5206632	Bt	1	9	350	28008	2	0	6	2	0	0	0	0	3	10

Predador *Chrysoperla externa*, larva de 3^o instar

Biblioteca	# Reads pós CQ	Presa	# Presa consumida	Tempo pós-predação (h)	Tamanho do inserto (pb)	# Reads do predador	# Reads da presa	<i>Cycloneda sanguinea</i>	<i>Ephesia kuehniella</i>	<i>Hippodamia convergens</i>	<i>Myzus persicae</i>	<i>Spodoptera litura</i>	# Espécies falso positivas	# Reads falso positivos
BioA31	6012026	-	0	-1	350	15960	0	0	0	0	0	0	0	0
BioA32	5921713	Sp	1	0	350	25220	16	6	0	0	0	0	1	6
BioA33	5672122	Sp	1	3	350	26784	0	26	0	0	0	0	1	26
BioA34	6058781	Sp	1	6	350	28212	0	22	2	0	0	0	3	28
BioA35	5144757	Sp	1	9	350	26548	0	2	2	0	0	0	2	4
BioA36	6080153	Mp	1	0	350	17256	0	6	2	0	0	0	2	8
BioA37	6230288	Mp	1	0	350	29946	4	36	2	0	0	0	3	60
BioA38	5835554	Mp	1	0	350	12506	0	8	0	0	0	0	1	8
BioA36B	314306	Mp	1	0	550	9312	2	12	0	0	0	0	1	12
BioA39	6165987	Mp	1	3	350	15240	0	12	0	0	0	0	1	12
BioA40	5897930	Mp	1	6	350	17268	0	20	2	2	0	0	3	24
BioA41	5998691	Mp	1	9	350	18212	0	54	0	0	0	0	1	54
BioA42	5689442	Bt	1	0	350	7500	0	310	4	2	2	0	3	318
BioA43	5542817	Bt	1	3	350	7698	0	12	0	0	0	0	2	14
BioA44	7493457	Bt	1	6	350	13714	0	5546	0	0	10	0	1	5556
BioA45	6215761	Bt	1	9	350	8730	0	22	0	0	0	0	2	24

CQ: controle de qualidade; Cs: *C. sanguinea*; Ce: *Ch. externa*; L: larva; Ad: adulto; Sp: *S. frugiperda*; Mp: *M. persicae*; Bt: *B. tabaci*.

Entre 0 e 1688 *reads* de *M. persicae* foram recuperadas imediatamente após a alimentação, sendo mais detectadas em joaninhas adultas do que nos demais predadores. Entre 0 e 10 *reads* de *M. persicae* foram detectadas 3, 6 e 9 h pós-predação. Entre 0 e 352 *reads* de *S. frugiperda* foram recuperadas imediatamente após a alimentação, sendo mais detectadas em joaninhas adultas do que nos demais predadores. Entre 0 e 254 *reads* de *S. frugiperda* foram detectadas 3, 6 e 9 h pós-predação, sugerindo um decaimento na detecção mais lento que o observado com *M. persicae*. Nenhuma *read* de *B. tabaci* foi encontrada em nenhuma das bibliotecas em que os predadores se alimentaram dessa presa, ou seja, houve falso-negativo (espécie de presa sabidamente consumida, mas não detectada). Hagler e Blackmer (2013) utilizando a técnica de PCR para detecção de presas no conteúdo gastrointestinal de predadores, observaram que não houve detecção da presa *B. tabaci* em bioensaios realizados em laboratório, apesar de que em amostras coletadas a campo foi possível a detecção da presa. Eles sugeriram que os predadores coletados em campo podem ter se alimentado de um maior número de presas de *B. tabaci* e que isso possa ter influenciado na detecção da presa, o que poderia explicar a não detecção da presa *B. tabaci* nos bioensaios realizados nesse trabalho. Outra possibilidade para não detecção das presas *B. tabaci* seria a utilização de biótipo de mosca-branca nos bioensaios diferente do biótipo que consta no banco de referência de DNA mitocondrial utilizado para as análises de detecção, devido a periódica renovação da colônia com adultos coletados em campo. Em geral, esses resultados sugerem que a biomassa da presa é um fator determinante na detecção. Dentre as três presas utilizadas (*B. tabaci*, *M. persicae*, *S. frugiperda*), *B. tabaci* foi a menor das presas. Como *B. tabaci* apresenta diversas espécies (complexo formado por 41 populações diferentes) (Silva *et al.*, 2009), é possível que a sequência de referência do genoma mitocondrial para *B. tabaci* seja diferente da sequência nos indivíduos

utilizados nesse trabalho. No entanto, ao se comparar os genomas mitocondriais completos de *B. tabaci* disponíveis no *GenBank*, há aproximadamente 77% de identidade entre eles (*GenBank*: AY521259.2 GI: 51944936; KU877168.1 GI: 1033204568; NC_006279.1 GI: 52220940). Então a diferença de 23% de identidade não explicaria a falta total de detecção de *reads* de *B. tabaci*, sugerindo que a melhor explicação seja a quantidade insuficiente de biomassa ingerida.

Os falsos-positivos podem ser contaminantes advindos da fase de obtenção das amostras de sua preparação para sequenciamento [p.ex. troca de amostras, tempo insuficiente de jejum dos predadores (como pode ser o caso do BioA9 que apresentou a detecção da *Aphis gossypi* que foi utilizado na manutenção das colônias dos predadores), dentre outros] ou de falsa identificação na análise de bioinformática devido a erro na elucidação dos mitogenomas referenciais ou do insuficiente poder discriminatório do método para espécies próximas geneticamente relacionadas. O caso de troca de amostras pode ter ocorrido, por exemplo, para as bibliotecas A23 e consequentemente da biblioteca A21B (formada pela junção das bibliotecas A21, A22, A23), onde foram detectadas muitas *reads* para *Hi. convergens* que não foi utilizada no bioensaio, mas havia colônia de criação no laboratório na mesma época em que havia colônia de criação de *C. sanguinea*. O caso de tempo insuficiente de jejum pode ter ocorrido, por exemplo, para todas as amostras de todas as bibliotecas em que houve detecção de *E. kuehniella* (nome científico sinônimo a *Anagasta kuehniella*), pois os ovos dessa espécie foram utilizados para manutenção das colônias dos predadores. O caso do insuficiente poder discriminatório do método para discernir entre espécies geneticamente próximas pode ter ocorrido, por exemplo, nas bibliotecas em que houve detecção de *Ch. nipponensis* (A30, A34, A37, A43 e A45), pois essa espécie não foi utilizada nos bioensaios, mas sua detecção pode estar associada à detecção da espécie

do mesmo gênero *Ch. externa*. Outro caso seria o da biblioteca A44, em que se observa um elevado número de *reads* de *C. sanguinea*, contaminação que pode ter acontecido nas etapas de preparação do DNA, pois de acordo com o número de *reads* observadas nas bibliotecas A42 a A45 nota-se uma mesma proporção entre elas, porém a amostra em questão apresenta 1/3 a mais de *reads*, sugerindo assim uma maior concentração de DNA que as demais.

Hagler e Naranjo (1997), utilizando ELISA indireta, ao analisarem o conteúdo gastrointestinal de três predadores (*Orius insidiosus*, *G. punctipes* – sugadores; *Hi. convergens* – mastigador), pós alimentação com diferentes números de ovos da lagarta-rosada-do-algodão *Pectinophora gossypiella* (Lepidoptera: Gelechiidae) e, foram mantidos por tempos e temperaturas diferentes durante a digestão, observaram que (i) as espécies menores de predadores apresentaram maior tempo de detecção das presas, sugerindo que apresentem taxa de digestão mais lenta já que são sugadores e por serem mais imunoresponsivos à ELISA do que a joaninha; (ii) quanto maior a quantidade de ovos ingerida, maior o sinal de detecção da presa, ou seja, maior a quantidade de antígenos da presa para ser detectada; (iii) quanto maior o tempo decorrido da predação, menor foi sinal de detecção da presa, pois os antígenos da presa vão sendo degradados cada vez mais à medida que o processo de digestão evolui.

Em estudo utilizando como predador a larva do besouro *Poecilus versicolor* (Carabidae: Coleoptera) e como presa a larva de primeiro instar ou pedaços frescos de larvas de segundo e terceiro instar do besouro *Melolontha melolontha* (Scarabaeidae: Coleoptera), relatam que após a ingestão de pelo menos 2 mg de presa, o predador foi coletado após 0, 2, 4, 8, 16, 24 e 32 h e, que foi possível detectar a presa em todos os tempos pós alimentação utilizando a técnica de PCR. Sheppard *et al.* (2005) em estudo de predação por PCR, observaram que a detecção da presa *Sitobion avenae* (Hemiptera:

Aphididae) foi positiva nos predadores *Tenuiphantes tenuis* (Araneae: Linyphiidae) e *Pterostichus melanarius* (Coleoptera: Carabidae) e que o tempo de detecção da presa foi superior em *T. tenuis* (entre 2 e 2,5 dias), sugerindo que essa detecção tenha sido possível por possuírem um metabolismo mais lento, cerca de 50-70%, do que os demais vertebrados e que predadores como os besouros tendem a ter um metabolismo mais acelerado devido ao gasto de energia à procura de presas.

5.3.1 Tipo de Predador

Detecção inicial (*n*₀). Houve diferenças significativas no número inicial das *reads* das presas para os três tipos de predadores (Tabela 4, Figura 11). Para as presas de *S. frugiperda* e *M. persicae*, o número inicial de *reads* foi maior para *C. sanguinea* (larvas e adultos) do que para *Ch. externa* (Tabela 4, $P= 0,0306$ e $2,01 \times 10^{-7}$). Para *S. frugiperda*, o número inicial de *reads* foi maior em adulto do que em larva de *C. sanguinea*. Embora não seja estatisticamente significativo, os adultos também apresentaram um maior número de *reads* iniciais que as larvas para *M. persicae* (Figura 11). Tais resultados podem ter relação com a forma de alimentação dos predadores, uma vez que os adultos de *C. sanguinea* ingerem por completo a presa, enquanto que as larvas de *C. sanguinea* podem apenas sugar o conteúdo da presa e as larvas de *Ch. externa* apenas sugam o conteúdo da presa. Portanto, o modo de alimentação pode influenciar na massa consumida da presa, ou seja, pode afetar o número inicial de *reads* detectadas das presas. Adicionalmente, o predador crisopídeo possui uma digestão extraoral da presa, o que pode aumentar a degradação do DNA da presa antes da ingestão, explicando o por quê de o crisopídeo apresentar o menor número de *reads* detectadas dentre os três predadores.

Tabela 4. ANOVA da detecção inicial da ln transformada das presas (n_0) em função do tipo de predador para cada presa e tipo de presa para cada predador.

Efeito	QM	F	P'
Tipo de presa para larva de <i>C. sanguinea</i>	25,1135	91,938	4,13E-08
Tipo de presa para adultos de <i>C. sanguinea</i>	53,8291	197,06	1,76E-11
Tipo de presa para larva de <i>Ch. externa</i>	7,34692	26,896	1,05E-04
Tipo de predador para <i>S. frugiperda</i>	3,81172	13,954	5,60E-04
Larva versus adulto de <i>C. sanguinea</i>	5,40948	19,804	8,24E-04
<i>C. sanguinea</i> versus <i>Ch. externa</i>	2,21395	8,1050	0,0306
Tipo de predador para <i>Myzus persicae</i>	11,1776	40,920	6,50E-07
Larva versus adulto de <i>C. sanguinea</i>	1,24940	4,5739	0,2084
<i>C. sanguinea</i> versus <i>Ch. externa</i>	21,1058	77,266	2,01E-07
Erro	0,273157		

QM: quadrado médio; F: razão Fisher das médias quadráticas; P': -valor-p do procedimento Ryan-Holm step-down Bonferroni.

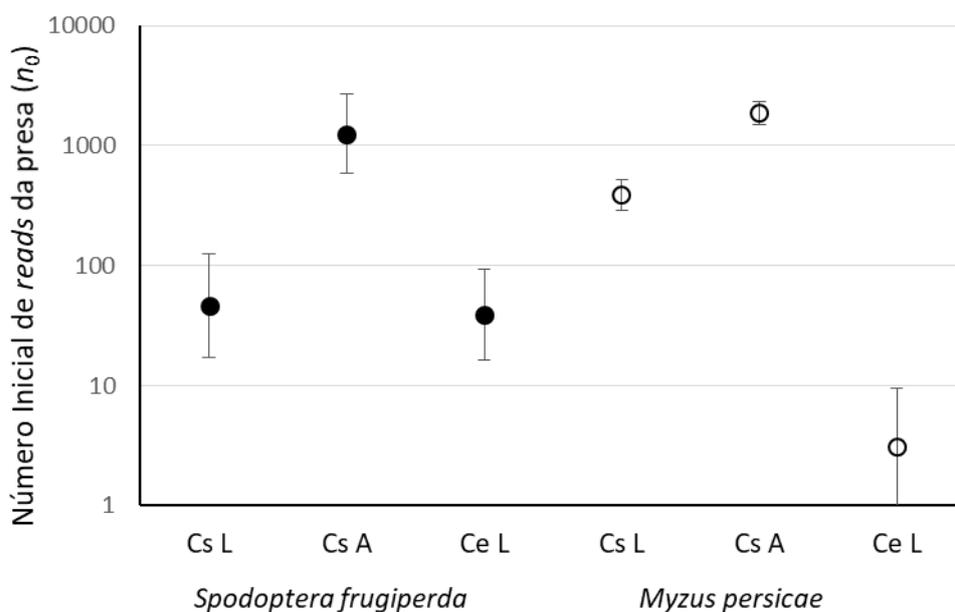


Figura 11. Número inicial de reads das presas detectadas no conteúdo gastrointestinal dos predadores normalizado pela cobertura média da biblioteca ($6,2 \times 10^6$ reads) com 95% IC. Cs = *C. sanguinea*; Ce = *Ch. externa*; L = larva; A = adulto.

Em estudo utilizando a técnica de PCR, Eitzinger *et al.* (2014) analisaram a influência do tamanho do corpo do predador e da qualidade nutricional da presa. Observaram que o tamanho corporal do predador não interferiu na detecção da presa, porém relataram que larvas de 4º ínstar de *Co. maculata* apresentaram maior eficiência

de digestão quando comparadas a larvas de ínstares menores. A observação deles pode estar relacionada ao resultado observado nesse trabalho, ou seja, uma fase do desenvolvimento ter taxa de digestão mais acelerada que outra. Quanto à análise da qualidade nutricional das presas, eles não conseguiram observar diferença entre as taxas de detecção das presas no conteúdo gastrointestinal.

Taxa de decaimento (*d*). Houve pouca diferença significativa na taxa de decaimento entre os tipos de predadores para as espécies presas (Tabela 5). Para a presa *S. frugiperda*, os adultos de *C. sanguinea* tiveram uma taxa de decaimento mais lenta do que as larvas de *C. sanguinea* ($P=0,0247$, Figura 12).

Tabela 5. ANOVA para taxa de decaimento em função do tipo de predador para cada presa e cada tipo de presa para cada predador.

Efeito	QM	F	P'
Tipo de presa para larva de <i>C. sanguinea</i>	0,847285	3,1018	0,3271
Tipo de presa para adultos de <i>C. sanguinea</i>	0,854416	3,1279	0,4433
Tipo de presa para larva de <i>Ch. externa</i>	0,283548	1,0380	0,3211
Tipo de predador para <i>S. frugiperda</i>	0,0547154	0,20031	0,9671
Larva versus adulto de <i>C. sanguinea</i>	0,0859727	15,242	0,0247
<i>C. sanguinea</i> versus <i>Ch. externa</i>	0,0234581	0,041989	0,9743
Tipo de predador para <i>M. persicae</i>	0,250487	0,91701	0,8015
Larva versus adulto de <i>C. sanguinea</i>	0,0154177	0,027597	0,9978
<i>C. sanguinea</i> versus <i>Ch. externa</i>	0,485558	0,86914	0,7414
Erro	0,558667		

QM: quadrado médio; F: razão Fisher das médias quadráticas; P': -valor-*p* do procedimento Ryan-Holm *step-down* Bonferroni.

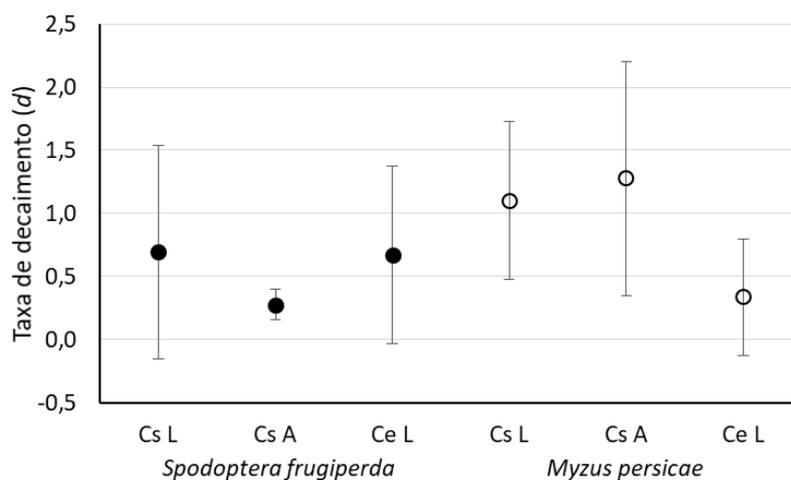


Figura 12. Taxa de decaimento da detecção das reads para ovos de *S. frugiperda* e adultos ápteros de *M. persicae* no conteúdo gastrointestinal dos predadores, normalizado para a cobertura média da biblioteca ($6,2 \times 10^6$ reads) com 95% IC. Cs = *C. sanguinea*; Ce = *Ch. externa*; L = Larva; A = Adulto.

Período de detectabilidade (D_{max}). Os ovos de *S. frugiperda* podem ser detectados 25 h pós-predação em adultos de *C. sanguinea*, porém por menos de 8 h pós-predação nos demais predadores (Tabela 6). A presa *M. persicae* foi detectável por menos de 8 h pós-predação para todos os tipos de predadores. Para ambas as espécies de presas, *Ch. externa* apresentou o menor e *C. sanguinea* adulto o maior período de detectabilidade.

Tabela 6. Valores estimados dos parâmetros a 95% de intervalo de confiança (IC).

Predador	Presa	n_0		d (h^{-1})		D_{max} (h)		T_{50} (h)	
		Média	95% IC	Média	95% IC	Média	95% CI	Média	95% IC
Cs L	Sf	46,7	[17,3; 126,3]	0,694	[-0,153; 1,541]	5,6	[0,2; 12,7]	1,4	[0,6; --]
Cs Ad	Sf	1253,5	[588,8; 2668,6]	0,279	[0,162; 0,397]	25,0	[23,9; 26,3]	3,6	[2,5; 6,2]
Ce L	Sf	39,1	[16,4; 93,6]	0,674	[-0,028; 1,376]	5,4	[0,3; 11,8]	1,5	[0,7; --]
Cs L	Mp	385,2	[288,0; 515,2]	1,101	[0,475; 1,728]	6,8	[4,2; 9, 7]	0,9	[0,6; 2,1]
Cs Ad	Mp	1871,5	[1505,7; 2326,2]	1,277	[0,349; 2,205]	7,7	[5,3; 9,8]	0,8	[0,5; 2,9]
Ce L	Mp	3,1	[0,99; 9,43]	0,336	[-0,336; 0,798]	3,3	[--; --]	3,0	[1,3; --]

n_0 : reads iniciais da presa; Cs: *C. sanguinea*; Ce: *Ch. externa*; Sf: *S. frugiperda*; L: larva; Ad: adulto; d : taxa de decaimento; D_{max} : período de detectabilidade; T_{50} : meia-vida; - não estimável.

5.3.2 Tipo de presa

Detecção inicial (n_0). Para os três tipos de predadores, o número inicial de *reads* das presas diferiu para as duas presas (*M. persicae* e *S. frugiperda*) (Tabela 4, Figura 11). Para larvas e adultos de *C. sanguinea* foram detectadas mais *reads* de *M. persicae* do que *reads* de *S. frugiperda* ($P'= 4,12 \times 10^{-8}$ e $1,76 \times 10^{-11}$). Sendo detectadas 8,2 vezes mais em larvas e 1,5 vezes em adultos. No entanto, para larvas de *Ch. externa*, o oposto foi verdadeiro ($P'= 1,05 \times 10^{-4}$), sendo 12,6 vezes mais *reads* de *S. frugiperda* do que *M. persicae*.

Taxa de decaimento (d). Para os três tipos de predadores, a taxa de decaimento não diferiu para as duas presas (*M. persicae* e *S. frugiperda*) (Tabela 5, $P'= 0,3271, 0,4433$ e $0,3211$). Para larvas e adultos de *C. sanguinea*, a taxa de decaimento foi mais lenta para *S. frugiperda*, mas não significativamente diferente (Figura 12), o que pode ter sido influenciado pelo restrito número de pontos de tempo pós-predação. Não houve diferenças significativas nas taxas de decaimento de presas para *Ch. externa* (Figura 12). A meia-vida de detecção das presas variou entre 0,8 - 3,6 h. Hagler e Blackmer (2013) relataram que presas maiores permanecem por mais tempo no conteúdo gastrointestinal do predador, o que pode influenciar no período maior de detecção da presa, resultado contraditório ao observado nesse estudo.

As taxas de decaimento dependeram das espécies de presas. O decaimento do número de *reads* de *M. persicae* foi significativamente mais rápido do que o número das *reads* da presa *S. frugiperda*. Embora os motivos para isso sejam desconhecidos, esse resultado mostra que a interpretação da análise do conteúdo intestinal deve considerar as taxas de decaimento específicas das espécies de presas, semelhante ao que foi sugerido para a análise por PCR dos conteúdos intestinais (Greenstone *et al.*, 2014). As taxas de

decaimento não dependeram significativamente do tipo de predador. Embora os dados indicassem alguma variação devido ao tipo de predador (Figura 12), a alta variação associada a alguns dos pontos de dados diminuiu a significância de tal variação. É possível que se houvessem mais pontos de tempos de observação pós-predação, as diferenças relacionadas com predadores na taxa de decaimento fossem reveladas. No entanto, os resultados observados sugerem que essa variação, se presente, é mais fraca do que o efeito das espécies de presas na taxa de decaimento.

A degradação do mtDNA de *M. persicae* e *S. frugiperda* nos três tipos de predadores ajustou muito bem o modelo de decaimento de primeira ordem para todos os três tratamentos de presas (Figura 13 e Figura S1, Anexo 7). Os valores de r^2 foram uniformemente elevados, variando de 0,83 a 0,98, embora devido ao pequeno número de pontos de tempo, algumas regressões não foram significativas.

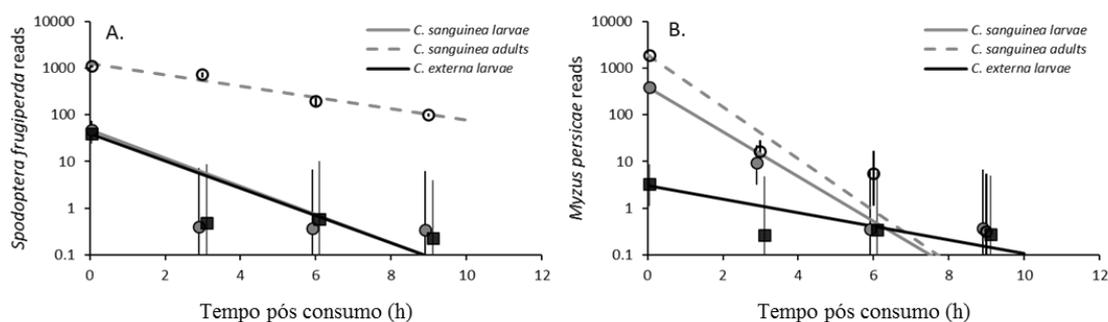


Figura 13. Taxas de decaimento das reads presentes no conteúdo gastrointestinal dos três predadores: (A) Ovo em pré-eclosão de *Spodoptera frugiperda* e (B) Adulto áptero de *Myzus persicae*. Linhas representam dados da regressão linear.

Período de detectabilidade (D_{max}). O período de detectabilidade está diretamente relacionado a detecção inicial e a taxa de seu decaimento (Tabela 6), e não com as espécies das presas. A presa *S. frugiperda* apresentou o período de detectabilidade mais longo em adultos de *C. sanguinea*, que apresentaram a maior detecção inicial e a menor taxa de decaimento. Para todos os outros, o período de detectabilidade foi inferior a 8 h devido a baixa detecção inicial ou a alta taxa de decaimento. No geral, o DNA dessas

pequenas presas foi degradado rapidamente (geralmente <8 h pós-predação). Isso implica que a detecção nesses predadores coletados no campo pode ser efetiva somente para presas recém-ingeridas ou para presas ingeridas em maior quantidade do que aquelas usadas em nossos bioensaios.

5.3.3 Detecção e decaimento das presas nos três tipos de predadores

Vinte modelos foram avaliados, exceto para os quatro modelos com um número constante de *reads* iniciais (n_0), todos foram ajustados para $r^2 > 0,989$. Cinco modelos foram retidos usando um critério de verossimilhança relativa baseado em AIC_c (Tabela S1, Anexo 2). Todos os modelos retidos permitiram que n_0 variasse com o tipo de predador e apenas um modelo exigia n_0 dependente de uma interação entre o tipo de predador e as espécies de presas. As taxas de decaimento em todos os modelos retidos não dependeram do tipo de predador, mas três dos modelos necessitaram da taxa de decaimento para depender das espécies de presas.

O melhor modelo (Tabela 7) mostra que a detecção inicial das *reads* das presas dependia diretamente do tipo predador e indiretamente das espécies de presas. Adultos de *C. sanguinea* apresentaram o maior número de *reads* iniciais e as larvas de *Ch. externa* o menor. A taxa de decaimento dependeu das espécies presas, porém não das espécies predadoras. O decaimento das *reads* de *S. frugiperda* foi cerca de três vezes mais lento do que o decaimento das *reads* de *M. persicae*. Fournier *et al.* (2008), em estudo comparando as técnicas PCR e ELISA, relataram que a detecção da presa variou tanto devido ao fator predador/presa quanto pela técnica e que o período de meia-vida de detecção pelo método de PCR foi maior que o observado pelo método ELISA, em que as presas foram detectadas após 11 h, 17,5 h e 51 h, no conteúdo gastrointestinal dos predadores *Ch. carnea* (larva de 3º instar), *H. axyridis* (adulto) e de *Zelus renardii*

(adulto) respectivamente, que se alimentaram de ovos ou adultos de *Homalodisca vitripennis* (Germar) (Hemiptera:Cicadellidae). Diversos autores (Hagler e Naranjo, 1997; Harper *et al.*, 2005; Fournier *et al.*, 2006; 2008) relataram que a meia-vida de detecção da presa apresentou uma elevada variação entre diferentes espécies de predadores. Em estudo realizado com sete espécies de Carabidae, três espécies de Staphylinidae e duas espécies de Linyphiidae, Sopp e Sunderland (1989), utilizando a técnica de ELISA, relataram que a taxa de decaimento do antígeno para os estafilínídeos foi muito mais rápida do que para os carabídeos ou para as aranhas. Tais resultados corroboram com o observado nessa tese.

Tabela 7. Parâmetros do melhor modelo para detecção inicial e decaimento das reads da presa nos três tipos de predadores.

Predador	Presa	$\ln n_0$	EP	d	EP
<i>C. sanguinea</i> – larvas	<i>S. frugiperda</i>	5,67	0,49	0,301	0,041
<i>C. sanguinea</i> – larvas	<i>M. persicae</i>	6,09	0,61	1,025	0,348
<i>C. sanguinea</i> – adultos	<i>S. frugiperda</i>	7,30	0,24	0,301	0,041
<i>C. sanguinea</i> – adultos	<i>M. persicae</i>	7,66	0,52	1,025	0,348
<i>Ch. externa</i> – larvas	<i>S. frugiperda</i>	3,12	1,40	0,301	0,041
<i>Ch. externa</i> – larvas	<i>M. persicae</i>	3,41	1,59	1,025	0,348

EP: erro padrão; d : taxa de decaimento.

5.4 Detecção das presas em função do número de presa consumida

Nesse bioensaio investigou-se se quanto maior a quantidade de presa ingerida, maior o número de *reads* detectadas para a presa e o tempo de detecção da presa pós-predação. Também foram avaliados: se houve a detecção de falsos-negativos (presas consumidas, mas não detectadas) e de falsos-positivos (presas não consumidas, mas detectadas); a variabilidade na detecção (número de *reads*) de dada presa/predador entre bibliotecas de DNA do mesmo tratamento; se o tamanho do inserto utilizado para construir as bibliotecas de DNA influência na detecção (número de *reads* e duração da detecção). O número total de *reads* obtidos por biblioteca após controle de qualidade e a

detecção das presas constam na Tabela 8. De maneira geral, pôde-se notar um número muito inferior de *reads* sendo detectadas para as presas que para os predadores, o que já era esperado. Nenhuma *read* de presas ou de falsos-positivos foi detectada em qualquer das bibliotecas utilizadas como controles. A detecção da presa decaiu com o tempo de digestão. O número de *reads* recuperadas da presa *M. persicae* variou entre 70 e 1240 em joaninhas que se alimentaram de mais de uma presa e imediatamente foram congeladas e variou entre 0 a 28 *reads* em joaninhas coletadas 3, 6 e 9 h de pós-predação, apresentando assim um rápido decaimento na detecção. De um total de 16 bibliotecas, houve detecção de falsos-positivos em 15 delas, sendo de até três espécies falsamente detectadas. A variabilidade e o tamanho do inserto serão discutidos em tópico separado, pois foi realizada uma análise integrada dos dados de todas as bibliotecas, independente do bioensaio.

Tabela 8. Número de reads detectadas para predadores, presas e falsos-positivos no Bioensaio B.

Biblioteca	# Reads pós CQ	Predador	Fase do predador	Presa	# Presa consumida	Tempo pós-predação (h)	Tamanho do inserto (pb)	# Reads do predador	# Reads da presa	<i>Adalia bipunctata</i>	<i>Chrysoperla externa</i>	<i>Cycloneda sanguinea</i>	<i>Ephesia kuehniella</i>	<i>Harmonia axyridis</i>	# Espécies falso positivas	# Reads falso positivas
BioB46	2827085	Hc	L	Mp	0	-1	350	24979	0	0	0	0	0	0	0	0
BioB47	2715140	Hc	L	Mp	1	0	350	22	6	0	508	2	0	34326	3	0
BioB48	2454150	Hc	L	Mp	1	0	350	15524	266	2	0	0	4	4	3	4
BioB49	2835714	Hc	L	Mp	1	0	350	18328	472	2	0	2	2	0	3	2
BioB47B	3506378	Hc	L	Mp	1	0	550	15652	350	0	156	0	6	7686	3	6
BioB50	2982288	Hc	L	Mp	1	3	350	17848	0	0	0	2	0	4	2	0
BioB51	2678374	Hc	L	Mp	1	6	350	18448	8	2	0	0	2	0	2	2
BioB52	2875610	Hc	L	Mp	1	9	350	13110	0	0	0	2	0	2	2	0
BioB53	2467086	Hc	L	Mp	3	0	350	12268	484	0	0	2	0	4	2	0
BioB54	2473628	Hc	L	Mp	3	3	350	18292	4	2	0	2	0	0	2	0
BioB55	2755179	Hc	L	Mp	3	6	350	27900	28	0	0	0	0	2	1	0
BioB56	2798613	Hc	L	Mp	3	9	350	26474	2	2	0	0	0	2	2	0
BioB57	2343983	Hc	L	Mp	6	0	350	15608	1240	0	0	0	2	6	2	2
BioB58	2577512	Hc	L	Mp	6	3	350	22190	8	0	0	4	4	6	3	4
BioB59	2908835	Hc	L	Mp	6	6	350	21142	14	2	0	2	0	0	2	0
BioB60	2585855	Hc	L	Mp	6	9	350	13564	4	0	0	4	0	2	2	0

CQ: controle de qualidade; Hc: *Hi. convergens*; Ha: *H. axyridis*; Ce: *Ch. externa*; L: larva; Mp: *M. persicae*

Análises preliminares do decaimento da detecção do mtDNA indicam que o modelo linear utilizando dados lineares transformados foi mais adequado do que o modelo não-linear usando dados não transformados para os tratamentos de 3 e 6 h (Figura 14), não apresentando diferença na regressão pré-definida para o bioensaio com uma presa *M. persicae* (Figura 14A). No entanto, o modelo não-linear apresentou declínio mais acentuado do que os dados os bioensaios com três (Figura 14B) e seis *M. persicae* (Figura 14C), assim o modelo linear foi utilizado nas análises subsequentes.

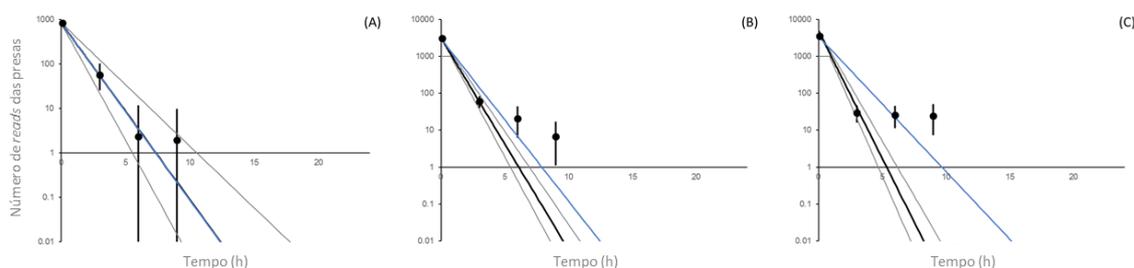


Figura 14. Relação entre o número de reads observadas da presa e métodos de regressão. Linha preta: regressão não linear sobre dados não transformados com nível de confiança de 95%. Linha azul: regressão linear em dados ln-transformados. A) Uma presa de *M. persicae*; B) Três presas de *M. persicae*; C) Seis presas de *M. persicae*.

O modelo de decaimento de primeira ordem para a decomposição do mtDNA de *M. persicae* em larvas de *Hi. convergens* apresentou-se adequado para todos os três tratamentos de presas (Tabela 9, Figura 15). O r^2 apresentou valores elevados, exceto para a regressão do bioensaio de 6 h devido ao pequeno número de pontos. Analisando o ajuste dos modelos, é possível que a degradação do DNA da presa desacelerasse com o tempo, já que todos os pontos de dados, incluindo o 9 h pós-predação, estavam acima dos valores previstos (Figura 15). A taxa de decaimento do mtDNA variou entre 0,842 a 1,013 h^{-1} (Tabela 9), apresentando meia-vida de detecção de 0,99 (0,82 e 1,23) a 1,19 (0,91 e 1,72) h (com erro padrão). Não houve, porém, relação entre a taxa de decaimento ou meia-vida de detecção com o número de presas consumidas, sendo que os três tratamentos com presas não apresentaram variação significativa nas taxas de

decaimento ($F_{2,9} = 0,196$, valor- $p=0,8251$) e o período de detectabilidade (D_{max}) foi maior quando mais presas foram consumidas, variando de 7,33 a 9,66 h.

Tabela 9. Valores dos parâmetros para cada tratamento.

# presas de <i>M. persicae</i>	Inclinação (h^{-1})	$\ln n_0$	n_0 (reads)	r^2	$F_{1,2}$	valor- p	D_{max} (h)
1	-0,915 (0,062)	6,713 (0,022)	888,6	0,991	219,37	0,0045	7,33
3	-1,013 (0,201)	7,996 (0,209)	2134,9	0,927	25,24	0,0374	7,89
6	-0,842 (0,260)	8,138 (0,198)	3711,5	0,840	10,52	0,0833	9,66

n_0 não foi transformado. O erro-padrão está indicado entre parêntesis.

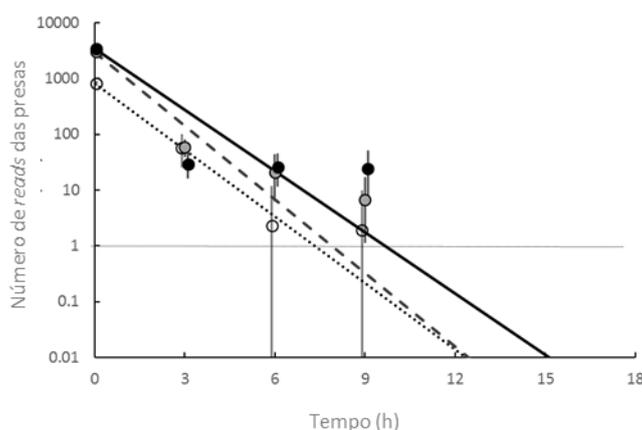


Figura 15. Taxa de decaimento das reads das presa *M. persicae* no conteúdo gastrointestinal de *Hi. convergens*, de acordo com o número de presas consumidas. Linha pontilhada: 1 *M. persicae* consumido. Linha tracejada: 3 *M. persicae* consumidos. Linha sólida: 6 *M. persicae* consumidos.

O modelo com o número de presas como a categoria variável (Modelo 1) deveria, em teoria, ajustar melhor os dados do que o modelo de regressão (Modelo 2), pois o Modelo 1 necessita de um parâmetro adicional. Os valores apresentados pelo r^2 foram de 0,931 para o Modelo 1 e 0,905 para o Modelo 2, ou seja, o Modelo 1 apresentou um melhor resultado. O AIC_c foi usado como parâmetro adicional do Modelo 1, como um método objetivo na discriminação entre os dois modelos com diferentes números de parâmetros. Quanto menor o AIC_c , melhor o modelo explica os dados. $\text{Exp}(\Delta AIC_c/2)$ estima a probabilidade relativa dos modelos avaliados. O Modelo 2 apresentou menor

AIC_c do que o Modelo 1 (Tabela 10), sugerindo que seria 1.8 vezes mais provável que o Modelo 1.

Tabela 10. Comparação dos modelos testando as hipóteses quantitativas com o número de reads da presa presentes no conteúdo gastrointestinal de *Hi. convergens*.

Modelos	<i>K</i>	<i>n</i>	AIC _c
1	3	11	81,12
2	2	11	79,89

Modelo 1: Análise categórica do número de presas;
 Modelo 2: Análise de regressão do número de presas.
K é o tamanho da amostra; AIC_c é o menor Critério de Informação de Akaike da amostra.

Consequentemente, a hipótese de que o número de *reads* de mtDNA de *M. persicae* está quantitativamente relacionado ao número inicial de presas consumidas é a melhor explicação para os resultados observados do que qualquer outra alternativa. Além disso, se o número de *reads* está diretamente relacionado ao número de presas consumidas, então o coeficiente de regressão (m_1 , Tabela 11) é predito igual a 1. Assim, para essa hipótese, $t_9 = 1,646$ e valor- $p = 0,1343$, que mostra que a inclinação de m_1 é ligeiramente menor que 1, não sendo assim significativamente diferente de 1. Tomando assim todos os dados em conjunto, temos que a taxa de decaimento foi de $0,895 \pm 0,131$ (h^{-1} , erro padrão), com uma meia-vida de detecção de 1,12 h {0,97;1,31}. Weber e Lundgren (2009) em estudo semelhante, observaram que a detecção da presa (ovos de *L. decemlineata*) por PCR no conteúdo gastrointestinal do predador *Co. maculata* (larva de 4º instar) estava diretamente relacionado ao número de ovos consumidos e não pela idade dos ovos e que a meia-vida de detecção da presa foi de pouco mais de 59 min para os predadores que estavam em jejum. Paula *et al.* (2015), utilizando sequenciamento *shotgun* de alta performance, avaliaram a detecção da presa *Acyrtosiphon pisum* (Hemiptera: Aphididae), no conteúdo gastrointestinal do predador adulto *H. axyridis* e relataram a detecção da presa nos tempos imediatamente (0 h) e 3h pós-predação.

Relataram ainda que o número de *reads* foi maior no tempo 0, enquanto que nesse estudo, utilizando a larva de 3º instar de *Hi. convergens* como predador, foi possível a detecção da presa *M. persicae* 6 h pós-predação e o maior número de *reads* também aconteceu no tempo 0 pós-predação. Hagler e Naranjo (1997), utilizando ELISA indireta, avaliaram o conteúdo gastrointestinal de três predadores *O. insidiosus*, *G. punctipes* e *Hi. convergens* que se alimentarem de diferentes números de ovos da lagarta-rosada-do-algodão *P. gossypiella* e observaram que os predadores sugadores (*O. insidiosus* e *G. punctipes*) apresentaram maior tempo de detecção das presas, sugerindo que podem apresentar taxa de digestão mais lenta que os predadores mastigadores. Além disso, relataram que quanto maior a quantidade de ovos ingerida, maior o sinal de detecção da presa, ou seja, maior a quantidade de antígenos da presa para ser detectada e que quanto maior o tempo decorrido da predação, menor foi sinal de detecção da presa, pois os antígenos da presa vão sendo degradados cada vez mais a medida que o processo de digestão evolui.

Tabela 11. Parâmetros estimados para o Modelo 2.

Parâmetro	Média (erro-padrão)	<i>t</i>	valor- <i>p</i>
n_0	6,790 (0,166)	40,86	$1,57 \times 10^{-11}$
m_1 , núreimo de presas	0,798 (0,123)	6,49	$1,12 \times 10^{-4}$
m_2 , tempo	-0,895 (0,131)	-6,83	$7,62 \times 10^{-5}$

$r = n_0 + m_1 N + m_2 T$, em que r = número de reads da presa *ln*-transformado; N = é o número de presas *ln*-transformado utilizadas na alimentação do predador, T = o tempo transcorrido após alimentação; m_1 e m_2 são os intervalos da regressão; n_0 = o número *ln*-transformado de reads das presas no tempo 0 para uma presa; t = estatística de *Student* para a hipótese nula em que o parâmetro é 0.

5.5 Detecção da presa em predação secundária em função do número de presas consumidas e tempo pós-predação

Nesse bioensaio investigou-se: se é possível detectar presa secundariamente consumida, ou seja, consumida por um predador (primário ou direto) que virou presa de

um segundo predador (secundário ou indireto); quanto tempo pós-predação primária e secundária a presa pode ser detectada de acordo com o número (biomassa) consumido e se o decaimento da detecção independe da predação ser primária (direta) ou secundária (indireta); se a taxa de decaimento da detecção da presa no predador primário independe do período de digestão do predador secundário; se a taxa de decaimento da detecção da presa depende da interação entre o tempo de digestão no predador primário e o tempo de digestão no predador secundário; se o decaimento da detecção do predador primário consumido pelo predador secundário (nesse caso o predador primário é considerado presa primária do predador secundário) depende de sua alimentação prévia de presa secundária; se é possível distinguir se uma dada presa detectada foi predada primariamente ou secundariamente. Também foram avaliados: detecção de falsos-negativos (presas consumidas, mas não detectadas) e de falsos-positivos (presas não consumidas, mas detectadas); variabilidade na detecção (número de *reads*) de dada presa/predador entre diferentes bibliotecas de DNA; relação entre o tamanho do inserto utilizado para construir as bibliotecas de DNA na detecção das presas (número de *reads* e duração da detecção).

O número total de *reads* por biblioteca após controle de qualidade e a detecção das presas constam na Tabela 12. De maneira geral, pôde-se notar um número muito inferior de *reads* sendo detectadas para as presas secundária e primária que para o predador secundário. A quantidade de *reads* detectada por presa/predador reduz com o tempo de digestão. Nenhuma *read* de presas ou de falsos-positivos foi detectada em qualquer das bibliotecas utilizadas como controles da predação intraguilda. No entanto, de três bibliotecas controle para predação intraguilda (C68 a C70) houve a detecção de presa secundária *M. persicae* em duas bibliotecas, embora ambos os predadores primário (*Ch. externa*) e secundário (*H. axyridis*) tenham permanecido em jejum por 24

h antes do início do bioensaio. A detecção das *reads* de *M. persicae* como falso-positivo nesse bioensaio levantou a incerteza de que outras bibliotecas não tenham sido contaminadas com *M. persicae*, ou se o tempo de 24 h em jejum foi suficiente para esvaziamento total do conteúdo gastrointestinal dos predadores antes de serem utilizados nos bioensaios, uma vez que para a manutenção das colônias dos predadores (joaninha e crisopídeo) foram utilizadas diferentes espécies de pulgões (*M. persicae*, *A. gossypii*, *Uruleucon* sp.). Outro incidente com falsos positivo foi a biblioteca C74 ter apresentado *reads* (n=3254) para *Hi. convergens*. A contaminação pode ter acontecido durante o bioensaio em que um indivíduo de *Hi. convergens* pode ter sido inadvertidamente adicionado ao *pool* de predadores utilizado na construção dessa biblioteca. Outra alternativa seria a contaminação ter ocorrido no momento da dissecação dos intestinos do predadores, uma vez que os quatro bioensaios foram realizados de forma concomitante.

A detecção de predação secundária é factível por esse método, sendo a presa primária mais passível de detecção do que a secundária. Entre 2 e 2112 *reads* da presa primária (*Ch. externa*) foram recuperadas, sendo mais detectadas imediatamente após a alimentação do predador, *H. axyridis*. Entre 0 e 350 *reads* de *M. persicae* foram recuperadas como presas secundárias no predador intraguilha (presa secundária – *M. persicae* – consumida pela presa primária – *Ch. externa* – e depois consumida por um predador intraguilha – *H. axyridis*). De um total de 20 bibliotecas, houve detecção de falsos-positivos em 18 delas, sendo de até três espécies falsamente detectadas. A predação secundária é sim detectada por esse método, o que impossibilita a distinção entre a predação primária e a secundária. A detecção inicial e o decaimento das *reads* da presa secundária no geral apresentou a mesma tendência que o observado nos demais bioensaios. Inicialmente a detecção das *reads* da presa primária (*Ch. externa*) no

predador intraguildd (*H. axyridis*) dependeu de seu histórico alimentar , ou seja, da alimentação prévia de *M. persicae*, o que não era esperado.

Tabela 12. Número de reads detectadas para predadores, presas e falsos-positivos no Bioensaio C.

Biblioteca	# Reads pós CQ	Predador	Fase do predador	Presa primária	Presa secundária	# Presa primária consumida	# Presa secundária consumida	Tempo pós predação secundária	Tempo pós predação primária	# Reads predador	# Reads presa primária	# Reads presa secundária	<i>Aphis gossypii</i>	<i>Cycloneda sanguinea</i>	<i>Ephesia kuehniella</i>	<i>Helicoverpa zea</i>	<i>Hippodamia convergens</i>	# Espécies falso positivas	# Reads falso positivos
BioC68	3233855	Ha	L	Ce	Mp	1	0	-	0	17462	2112	4	0	0	20	0	0	1	20
BioC69	2566038	Ha	L	Ce	Mp	1	0	-	3	31884	376	2	2	2	4	0	0	3	8
BioC70	2762291	Ha	L	Ce	Mp	1	0	-	6	44802	116	0	0	0	56	2	2	3	60
BioC71	2296493	Ha	L	Ce	Mp	1	3	0	0	8440	228	6	0	0	2	0	2	2	4
BioC72	2973215	Ha	L	Ce	Mp	1	3	0	3	41224	92	8	0	0	4	0	6	2	10
BioC73	2864639	Ha	L	Ce	Mp	1	3	0	6	55024	6	8	0	2	6	0	2	3	10
BioC74	2898985	Ha	L	Ce	Mp	1	3	3	0	22180	10	70	2	0	0	0	3254	2	3256
BioC75	2809740	Ha	L	Ce	Mp	1	3	3	3	25706	76	0	0	2	12	0	0	2	14
BioC76	2966448	Ha	L	Ce	Mp	1	3	3	6	46236	64	6	0	0	2	0	0	1	2
BioC76.1	2290736	Ha	L	Ce	Mp	1	3	6	0	36104	4	0	0	0	0	0	2	1	2
BioC76.2	2608065	Ha	L	Ce	Mp	1	3	6	3	28334	2	0	0	0	0	0	0	0	0
BioC76.3	2484257	Ha	L	Ce	Mp	1	3	6	6	24610	24	0	0	0	2	0	0	1	2
BioC77	3032949	Ha	L	Ce	Mp	0	0	-	-1	78714	0	0	0	2	2	0	0	2	4
BioC78	2877323	Ha	L	Ce	Mp	0	1	0	-	32400	0	512	0	0	0	0	2	1	2
BioC79	2930724	Ha	L	Ce	Mp	0	1	0	-	25240	0	492	0	0	4	0	0	1	4
BioC80	2853434	Ha	L	Ce	Mp	0	1	0	-	19870	0	52	0	0	0	0	2	1	2
BioC78B	3585431	Ha	L	Ce	Mp	0	1	0	-	30622	0	72	0	0	2	0	0	1	2
BioC81	2676830	Ha	L	Ce	Mp	0	3	0	-	23746	0	146	0	0	18	0	2	2	20
BioC82	3493106	Ha	L	Ce	Mp	0	3	3	-	47824	0	108	0	0	0	0	0	0	0
BioC83	2623573	Ha	L	Ce	Mp	0	3	6	-	22506	0	2	0	0	8	0	0	1	8

CQ: controle de qualidade; Ha: *H. axyridis*; Ce: *Ch. externa*; L: larva; Mp: *M. persicae*; Inserto: 350pb.

Harwood *et al.* (2001) utilizando ELISA avaliaram a predação secundária do pulgão *Sitobion avenae* (Hemiptera: Aphididae) consumido pela aranha *Lepthyphantes tenuis* (Araneae: Linyphiidae), por sua vez consumida pelo besouro *Poecilus* (=Pterostichus) *cupreus* (Coleoptera: Carabidae). Eles observaram que a detecção do pulgão só foi possível em besouros (predador secundário) que foram imediatamente congelados após se alimentarem de aranhas (predador primário do pulgão e presa primária do besouro) que por sua vez foram imediatamente congeladas após se alimentarem do pulgão. Sheppard *et al.* (2005) por meio da técnica de PCR, observaram ser possível a detecção do pulgão *S. avenae*, que inicialmente foi consumido pelo predador primário *T. tenuis*, no conteúdo gastrointestinal do predador secundário *P. melanarius*. De acordo com eles a detecção da presa secundária foi possível no predador secundário: em até 8 h pós-predação do predador primário, que havia imediatamente consumido o pulgão; e em até 4 h pós-predação do predador primário 4 h após este ter predado o pulgão. Apesar de o predador primário (*Ch. externa*) apresentar uma taxa de metabolismo mais acelerado do que o observado em aranhas (Greenstone e Bennett, 1980), os resultados obtidos nesse trabalho apresentaram um intervalo de detecção superior ao estudo citado em aranhas, o que sugere que a detecção da presa está relacionada às espécies envolvidas e as técnicas utilizadas para a detecção.

5.5.1 Detecção de presa secundária em predações primária e secundária

O número de *reads* de *M. persicae* detectadas em *H. axyridis* não foi significativamente diferente quando foi consumido por predação primária ou por secundária (Tabela S2, Anexo 7). Embora não tenha havido sobreposição dos ICs de 95% durante 3 h pós-predação da presa secundária (Figura 16), a elevada sobreposição a 0 e 6 h fez a diferença global não ser significativa.

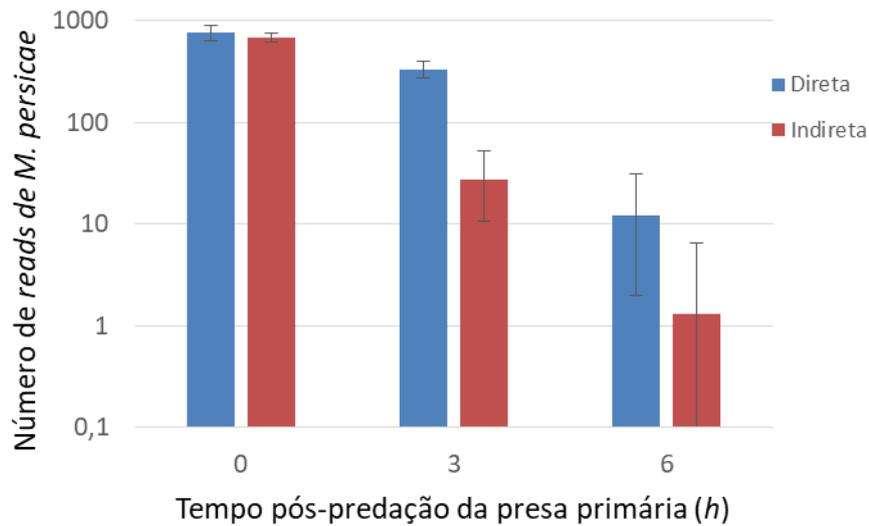


Figura 16. Número de reads de *M. persicae* detectado em *H. axyridis* pós-predação primária ou secundária com IC de 95%.

5.5.2 Taxa de decaimento da presa secundária em predações primária e secundária

A taxa de decaimento (d) da detecção de *M. persicae* consumido por *H. axyridis* não diferiu da predação ter sido primária ou secundária (Tabela S3, Anexo 7, Teste 1). O r^2 ajustado para este modelo não foi alto (0,5466), provavelmente devido ao grande número de *reads* observadas para predação primária até 3 h pós-predação (Figura 17). Apesar disso, a comparação do modelo completo e do melhor modelo mostrou que os ICs de 95% se sobrepuseram para o número inicial de *reads* (n_0), para a taxa de decaimento (d) e para o período de detecção (D_{max}) (Tabela S4, Anexo 7, Teste 1). A distinção entre a predação primária e secundária também não foi possível mesmo pela análise da taxa de decaimento.

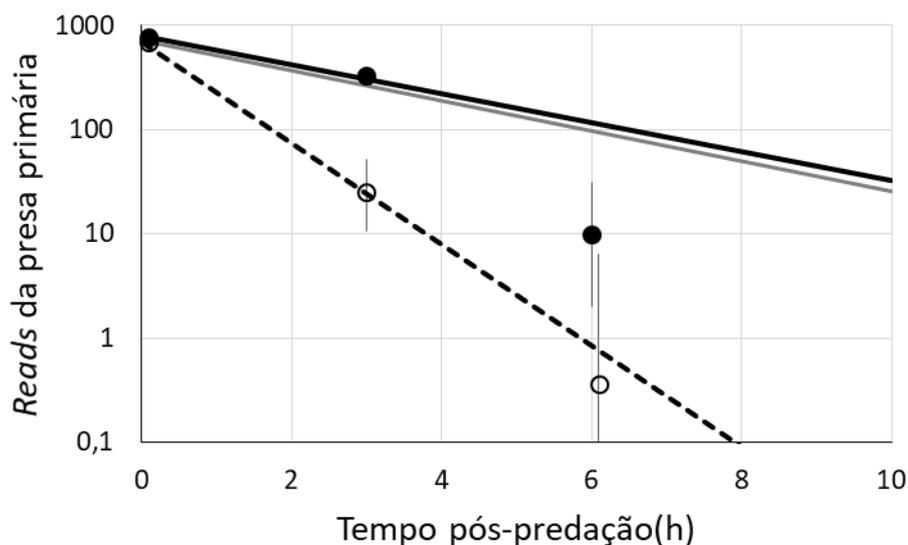


Figura 17. Decaimento das reads da presa secundária (*M. persicae*) em *H. axyridis* em função da predação primária (linha preta sólida, círculos pretos sólidos) ou predação secundária (linha tracejada, círculos abertos) e tempo pós-predação por *H. axyridis*. O melhor modelo está apresentando com linha cinza sólida. Círculos observados foram normalizados com IC 95%.

5.5.3 Taxa de decaimento das reads da presa secundária no conteúdo gastrointestinal da presa primária em relação ao tempo de digestão no predador intraguilda

O melhor modelo apresentou a detecção inicial (n_0) dependente do período de digestão no predador intraguilda (*H. axyridis*) e uma taxa de decaimento comum (Tabela S2, Anexo 7, Teste 2). O modelo com números de taxas de detecção iniciais e taxa de decaimento comuns apresentou um AIC_c semelhante, mas apresentou um r^2 baixo (0,4197). O modelo selecionado apresentou melhor ajuste dos dados com um r^2 elevado (0,9975). O modelo completo diferiu ligeiramente do melhor modelo (Tabela S4, Anexo 7, Teste 2) e os números de detecção iniciais foram quase idênticos. Reads foram mais detectadas no momento inicial pós-predação (0 h) pelo predador intraguilda que após 3 ou 6 h pós-predação (Tabela S4, Anexo 7, Figura 18). Os ICs de 95% para a taxa de decaimento para o modelo completo no tempo de 6 h não se sobrepôs ao do modelo completo a 0 h pós-predação (Tabela S4, Anexo 7), mas a AIC_c foi muito maior para o modelo completo (Tabela S2, Anexo 7). A AIC_c penaliza um modelo com mais

parâmetros, de modo que o modelo de taxa de decaimento comum foi superior. A taxa de decaimento comum da presa secundária foi de 0,799 no grupo controle (tempo h⁻¹) na presa primária, *Ch. externa* e o período de detectabilidade (D_{max}) variaram de 3,3 a 6,2 h.

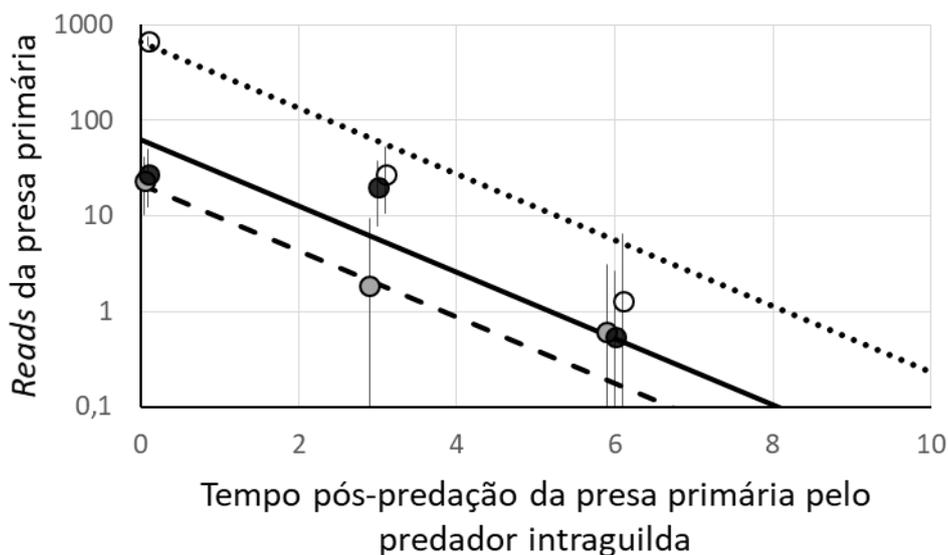


Figura 18. Decaimento das reads da presa secundária (*M. persicae*) no gastrointestinal da presa primária (*Ch. externa*), em função do período de digestão do predador intragilda (*H. axyridis*) e o tempo pós-predação por *Ch. externa*. Linhas representam os melhores modelos e os círculos representam a normalização a 95% de IC. Linha pontilhada e círculos abertos: 0 h pós-predação por *H. axyridis*; Linha tracejada e círculos cinza: 3 h pós-predação por *H. axyridis*; Linha sólida e círculos negros: 6 h pós-predação por *H. axyridis*.

5.5.4 Taxa de decaimento das reads da presa secundária no conteúdo gastrointestinal do predador intragilda em relação ao tempo de digestão na presa primária

O melhor modelo apresentou a detecção inicial (n_0) dependente do período de digestão na presa primária e uma taxa de decomposição comum (Tabela S3, Anexo 7, Teste 3). O modelo com números de taxas de detecção iniciais comuns e taxa de decaimento tiveram um AIC_c muito baixo, assim como um r^2 (0,6251). O modelo selecionado apresentou melhor ajuste dos dados com r^2 alto (0,9935). O modelo completo, por sua vez, apresentou um r^2 ligeiramente mais elevado, mas um AIC_c muito alto (Tabela S3, Anexo 7). Os números iniciais de detecção foram qualitativamente

semelhantes entre o modelo completo e o modelo selecionado (Tabela S4, Anexo 7, Teste 3), com mais *reads* de presas secundária detectadas imediatamente pós-predação (tempo 0 h) pela presa primária e menos *reads* detectadas 6 h pós-predação (Tabela S4, Anexo 7, Figura 19). Tal resultado é esperado devido a degradação das *reads* da presa secundária no conteúdo gastrointestinal da presa primária. As taxas de decaimento para o modelo completo variaram consideravelmente entre os três períodos de digestão de presas primária, mas suas ICs de 95% se sobrepuseram (Tabela S4, Anexo 7). A variação deveu-se provavelmente ao pequeno número de presas secundária observadas nos períodos de digestão de presas primária mais tardios. A taxa de decaimento comum da presa secundária foi de $0,514 \text{ h}^{-1}$ no predador primária, *H. axyridis*. O período de detectabilidade (D_{max}) não pôde ser estimado para todos, mas apresentou valores de 8,9 e 12,6 h após 3 e 0 h, após a presa primária consumir a presa secundária, respectivamente.

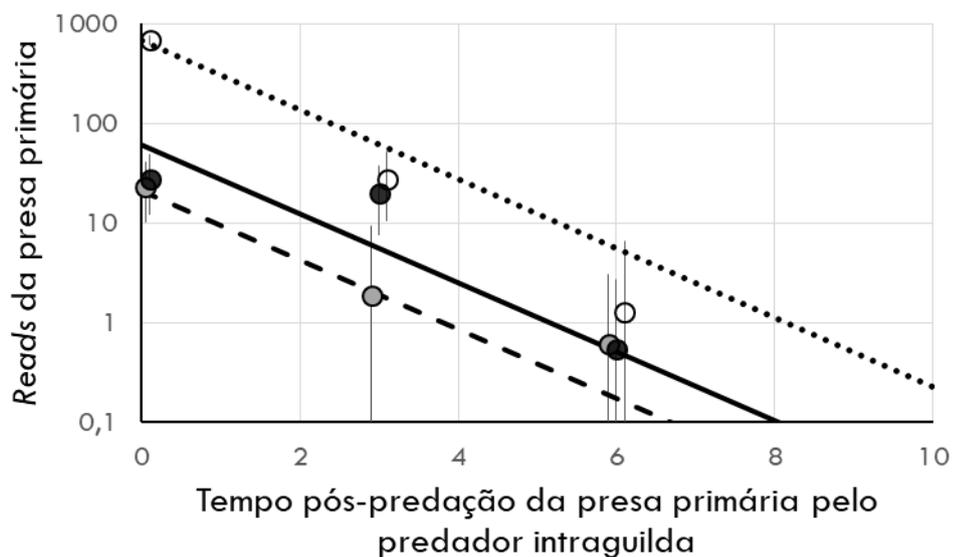


Figura 19. Decaimento das *reads* da presa secundária (*M. persicae*) no predador intraguilda (*H. axyridis*), em função do tempo de digestão da presa primária (*Ch. externa*) e do tempo pós-predação por *H. axyridis*. Linhas representam os melhores modelos e os círculos representam a normalização com 95% de ICs. Linha pontilhada com círculos abertos: 0 h pós-predação por *Ch. externa*; Linha tracejada com círculos cinza: 3 h pós-predação por *Ch. externa*; Linha sólida com círculos pretos: 6 h pós-predação por *Ch. externa*.

5.5.5 Taxa de decaimento das *reads* da presa secundária no conteúdo gastrointestinal da presa primária e no do predador intraguilha

Comparando as taxas de decaimento das *reads* da presa secundária no conteúdo gastrointestinal da presa primária (*Ch. externa*) e no do predador intraguilha (*H. axyridis*) observou-se a diferença de 0,285, não significativa ($t_5 = 1,58$; valor- $p = 0,1747$). Desta forma, pôde-se observar que não houve diferença na taxa de decaimento da presa secundária, independente se foi consumida pela presa primária ou pelo predador intraguilha.

É possível que a decomposição da presa secundária dependa da relação do tempo de digestão no predador intraguilha e do tempo de digestão da presa primária. Teoricamente, dado o desenho experimental desenvolvido, poderiam haver seis diferentes taxas de decaimento, três associadas com decaimento da presa primária, que diferem dependendo do período de digestão no predador intraguilha, e três associadas com decaimento no predador intraguilha, que diferem dependendo do período de digestão na presa primária. Os resultados anteriores, porém, sugerem que este nível de complexidade seja improvável, pois não houve diferença nas taxas de decaimento nas presas primárias que dependessem do predador intraguilha e nenhuma diferença nas taxas de decaimento no predador intraguilha que dependesse da presa primária. Além disso, a análise anterior indicou que a taxa de decaimento na presa primária e no predador intraguilha pode ser a mesma, portanto, considerou-se a hipótese de que a decomposição da presa secundária estava relacionada com o tempo total de digestão, independentemente de estar na presa primária ou no predador intraguilha. O melhor modelo de análise foi o modelo de tempo total (Tabela S3, Anexo 7, Teste 4), apoiando esta hipótese (Tabela S2, Anexo 7, Teste 4). A taxa de decaimento estimada foi de $0,563 \text{ h}^{-1}$ e o período de detectabilidade foi de 8,7 h (Figura 20), sugerindo que a presa

secundária (neste caso, três adultos ápteros de *M. persicae*) pode ser detectada direta ou indiretamente durante 8,7 h após a primeira ingestão.

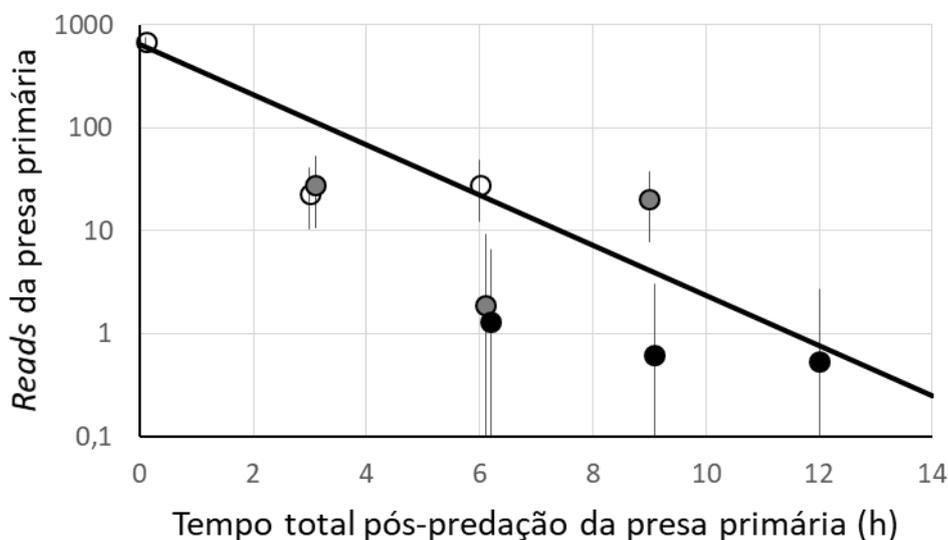


Figura 20. Decaimento das reads da presa secundária (*M. persicae*) na presa primária (*Ch. externa*) e no predador intraguilda (*H. axyridis*), em função do tempo total pós-predação pela *Ch. externa* e por *H. axyridis*. Linhas apresentam os melhores modelos e círculos representam a normalização com 95% de IC. Círculos abertos: 0 h pós-predação por *Ch. externa* e 0 - 6 h pós-predação por *H. axyridis*. Círculos cinza: 3 h pós-predação por *Ch. externa* e 0 - 6 h pós-predação por *H. axyridis*. Círculos pretos: 6 h pós-predação por *Ch. externa* e 0 - 6 h pós-predação por *H. axyridis*.

5.5.6 Decaimento da presa primária no conteúdo gastrointestinal do predador intraguilda em relação à alimentação prévia pela presa primária

O melhor modelo apresentou a detecção inicial (n_0) dependente de a história alimentar prévia da presa primária e de uma taxa de decaimento comum (Tabela S3, Anexo 7, Teste 5). O modelo selecionado ajustou bem os dados com um r^2 alto (0,9988). O modelo completo apresentou um r^2 ligeiramente mais elevado, mas um AIC_c mais elevado (Tabela S3, Anexo 7). Os números de reads iniciais de detecção foram qualitativamente semelhantes entre o modelo completo e o modelo selecionado (Tabela S4, Anexo 7, Teste 5), com mais reads de presas primárias detectadas para presas primárias que não se alimentaram e menos detectadas para presas primárias que se alimentaram de presas secundária com um período de ingestão de até 6 h (Tabela S4, Anexo 7, Figura 21), não estando claro, porém, como a presença de presas secundárias

interferem na detectabilidade da presa primária. As taxas de decaimento para o modelo completo variaram entre os quatro grupos de alimentação de presas de intraguilda, mas seus ICs de 95% se sobrepuseram (Tabela S4, Anexo 7). A variação deveu-se provavelmente ao pequeno número de presas primárias utilizadas nos bioensaios de alimentação de presas primárias, sendo que a taxa de decaimento comum da presa primária foi de $0,484 \text{ h}^{-1}$ no predador intraguilda, *H. axyridis*. O período de detectabilidade (D_{max}) foi entre 7,5 e 18,2 h e geralmente mais longo do que para a presa secundária.

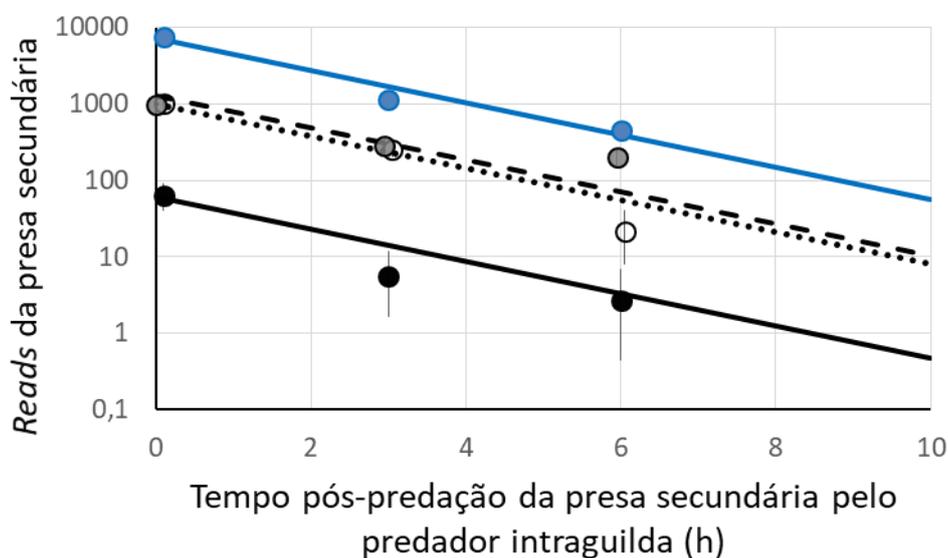


Figura 21. Decaimento das reads das presas primárias (*Ch. externa*) no predador intraguilda (*H. axyridis*), em função da alimentação prévia da presa primária (*Ch. externa*) e do tempo pós-predação por *H. axyridis*. Linhas mostram o melhor modelo e os círculos apresentam os dados normalizados a 95% de ICs. Linha e círculos azuis: sem ingestão prévia de presa secundária. Linha pontilhada e círculos abertos: imediatamente pós-predação da presa secundária. Linha tracejada e círculos cinza: 3 h pós-predação da presa secundária. Linha sólida e círculos negros: 6 h pós-predação da presa secundária.

5.5.7 Taxa de decaimento das presas primária e secundária no predador intraguilda

Para testar se a taxa de decaimento foi a mesma para a presa secundária (*M. persicae*) e a presa intraguilda (*Ch. externa*) no predador intraguilda (*H. axyridis*), as taxas de decaimento estimadas nos resultados anteriores foram comparadas. A diferença

nas taxas de decaimento foi de 0,030, não sendo significativa ($t_{12} = 0,28$, valor- $p = 0,7823$), indicando assim que a taxa de decaimento da presa secundária e da presa intraguilida foram semelhantes no predador intraguilida. O número de *reads* das presas primária e secundária numa biblioteca foi positivamente correlacionado ($r^2 = 0,7095$, $n = 9$, valor- $p = 0,0323$, Figura 22). Seria de se esperar que houvesse relativamente mais *reads* de presas secundária imediatamente pós-predação pelas presas primárias do que 6 h pós-predação. Os dados são consistentes com esta expectativa, mas o número de *reads* das presas primárias covariadas com a presa secundária foi maior do que o esperado.

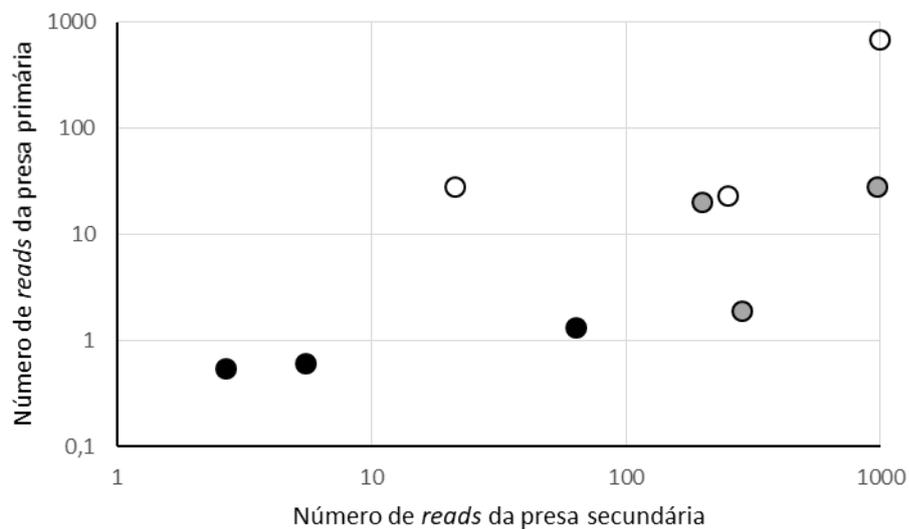


Figura 22. Correlação entre o número de *reads* da presa secundária (*M. persicae*) e da presa primária (*Ch. externa*) detectadas em *H. axyridis*. Círculos abertos: 0 h pós-predação da presa secundária por *Ch. externa*. Círculos cinza: 3 h pós-predação da presa secundária por *Ch. externa*. Círculos negros: 6 h pós-predação da presa secundária por *Ch. externa*.

5.6 Detecção da presa em função da temperatura ambiente de digestão da presa

Nesse bioensaio investigou-se a influência da temperatura ambiente de digestão na detecção da presa (número de *reads* e tempo de detecção). O número total de *reads* por biblioteca após controle de qualidade e a detecção das presas constam na Tabela 18.

Tabela 13. Número de reads detectadas para os predadores, presas e falsos-positivos no Bioensaio D.

Biblioteca	# Reads pós CQ	Predador	Fase do predador	Presa	# Presa	Tempo pós-predação (h)	Temperatura de digestão (C)	Tamanho do inserto (pb)	# Reads do predador	# Reads da presa	<i>Adalia bipunctata</i>	<i>Aphis gossypii</i>	<i>Harmonia axyridis</i>	<i>Hippodamia convergens</i>	# Espécies falso positivas	# Reads falso positivos
BioA1	4760877	Cs	L	-	-	-1	25	350	35342	0	0	0	0	0	0	0
BioD61	2690267	Cs	L	Mp	1	0	25	350	4774	1714	4	2	2	12866	4	12874
BioD62	2773562	Cs	L	Mp	1	3	20	350	24750	6	0	0	2	2	2	4
BioD63	2495933	Cs	L	Mp	1	6	20	350	18950	4	0	0	2	0	1	2
BioD64	2548934	Cs	L	Mp	1	9	20	350	27724	0	0	0	6	0	1	6
BioA9	4797188	Cs	L	Mp	1	3	25	350	42578	0	0	2	0	0	1	2
BioA10	4975496	Cs	L	Mp	1	6	25	350	44144	0	0	0	0	0	0	0
BioA11	4888879	Cs	L	Mp	1	9	25	350	40764	0	0	0	0	4	1	4
BioD65	2951181	Cs	L	Mp	1	3	30	350	23190	2	0	0	6	0	1	6
BioD66	2687841	Cs	L	Mp	1	6	30	350	22332	0	0	0	0	0	0	0
BioD67	2770406	Cs	L	Mp	1	9	30	350	27174	2	0	0	10	0	1	10

CQ: controle de qualidade; Cs: *C. sanguinea*; Mp: *M. persicae*; L: larva.

Não houve influência da temperatura na detecção do número de *reads* das presas. Houve sobreposição de todos os ICs 95% em todos os tratamentos 6 e 9 h pós-predação. No tempo 3 h pós-predação da presa, o tratamento a 20°C apresentou maior número de *reads* do que o tratamento a 25°C, porém não apresentou diferença quando comparado ao tratamento a 30°C. É importante observar o elevado número de *reads* contaminantes (n= 12866) do predador *Hi. convergens*, o que sugere que, assim como no bioensaio A, em alguma etapa do bioensaio D houve a contaminação dessa amostra, seja na dissecação do inseto ou na própria extração de DNA.

Von Berg *et al.* (2008), utilizando a técnica de PCR, avaliaram a influência da temperatura (12, 16 e 20°C) na detecção de presas no conteúdo gastrointestinal dos predadores *Nebria brevicolis* (Coleoptera: Carabidae) e *P. melanarius*. Quando comparada a detecção entre as duas espécies, os autores sugerem que exista uma diferença na capacidade de digestão entre esses predadores, pois não observaram influência da temperatura ou do tamanho do fragmento na detecção da presa. A temperatura apenas interferiu na detecção das presas no predador *P. melanarius*, que apresentou queda no número de detecção à 20°C quando comparado às temperaturas de 12 e 16°C. Mcmillan *et al.* (2007), utilizando a técnica de PCR, avaliou a detecção da presa *Rhopalosiphum padi* (Hemiptera: Aphididae) no conteúdo gastrointestinal da joaninha *A. bipunctata* em duas temperaturas, 14 e 21°C. Não houve efeito de temperatura sobre a decomposição/detecção das *reads* das presas. Hoogendoorn e Heimpel (2001), utilizando a técnica de PCR, observaram menor número de predadores (*Co. maculata*) com detecção positiva de ingestão de presa ao aumentarem a temperatura ambiente de digestão de 20 para 27°C. Utilizando a técnica de ELISA, Haggler e Naranjo (1997), utilizando ELISA, avaliaram a interferência da temperatura na detecção da presa *P. gossypiella* no conteúdo gastrointestinal do predador *O.*

insidiosus. Ao avaliarem as temperaturas de 15, 20, 25 e 30°C, observaram que não houve diferença na detecção das presas entre as temperaturas de 15, 20 e 25°C. Entretanto, a 30°C observaram significativa redução no número de indivíduos com detecção de ingestão de presa. Relataram ainda que a detecção da presa *P. gossypiella* é influenciada pela temperatura em que ocorre a digestão dos predadores (*O. insidiosus*, *G. punctipes* e *Hi. convergens*) e que quanto maior a temperatura, mais curta foi a duração da detecção da presa, ou seja, em temperaturas mais elevadas a taxa de digestão dos antígenos da presa aumenta. Sopp e Sunderland (1989) utilizaram ELISA para avaliar o efeito da variação da temperatura (7, 10, 13 e 16°C) na digestão de sete espécies de Carabidae, três espécies de Staphylinidae e duas espécies de Linyphiidae. Os dados obtidos foram semelhantes aos resultados em temperatura constante e, embora a taxa de decomposição do antígeno tenha sido ligeiramente mais rápido que em temperatura constante, a variabilidade nos dados foi muito grande, não permitindo a detecção de qualquer diferença significativa.

Nesse trabalho não foi possível determinar o efeito da temperatura ambiente de digestão do predador no decaimento da detecção da presa. Isso provavelmente sugere a necessidade de se testar outros tratamentos em que o predador se alimente de mais presas, tratamentos que apresentem temperaturas mais extremas, aumentar o número de predadores utilizados ou ainda coletar os predadores em menores intervalos pós-predação da presa. Por outro lado, testes devem ser feitos para avaliar a possibilidade das larvas de joaninha gerarem calor interno de modo que as taxas de decaimento não sejam tão diferentes como se poderia esperar em diferentes temperaturas ambiente de digestão (Andow, comunicação pessoal).

5.7 Variação do número de indivíduos agrupados numa mesma biblioteca e tamanho do inserto

Variação do número de espécimes utilizadas numa mesma biblioteca e agrupamento: O erro padrão (Figura 23) foi negativamente correlacionado com a média estimada do número de *reads* detectadas para os cinco predadores. O erro padrão mais alto está associado com *Ch. externa*, que apresentou o menor número de *reads* de presas detectadas. Com baixo número de *reads* detectado, a variância dos números de *reads* estimadas é alta, o que contribui para o alto erro padrão estimado.

Estima-se que o erro padrão da média estimada do número de *reads* detectadas seja reduzido à medida que mais indivíduos são agrupados. Se ao invés de utilizarmos apenas um indivíduo, forem utilizados cinco na construção da biblioteca, espera-se haver uma redução do valor do erro padrão para apenas 44% do valor individual. Se 10 indivíduos fossem agrupados em uma mesma biblioteca, o erro padrão seria reduzido para 33% do valor individual. Se 20 indivíduos fossem agrupados, seria reduzido para 22%, mas haveria grande esforço experimental para obtê-los. Assim sendo, considera-se que a utilização de 10 predadores é a melhor estratégia para uma menor variação do número de *reads* detectadas da presa.

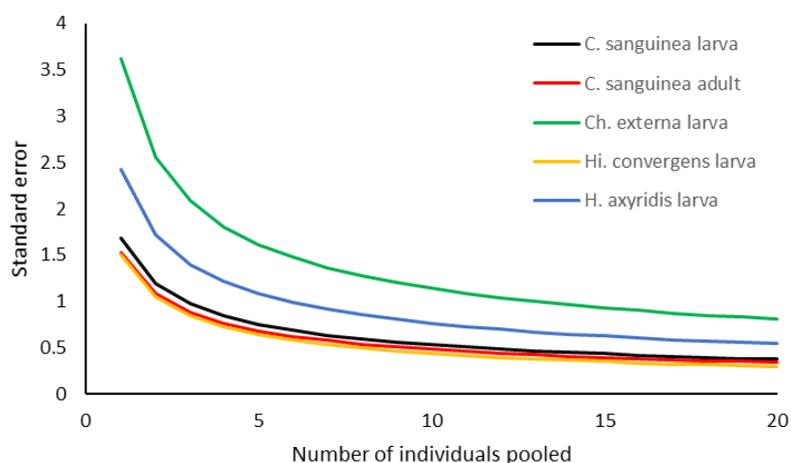


Figura 23. Relação entre o número de predadores agrupados por biblioteca e o erro padrão do número estimado de *reads* das presas com base na variação estimada entre o número de indivíduos utilizados na construção do pool.

Tamanho da biblioteca: Em dois casos, de um total de cinco, a biblioteca com inserto de 550 pb apresentou duas vezes mais *reads* de presas do que a biblioteca com inserto de 350 pb (predadores: larvas de *C. sanguinea* e larvas de *Ch. externa*). No entanto em outros dois casos a biblioteca com inserto de 350 pb apresentou mais *reads* de presas do que a biblioteca com inserto de 550 pb (predadores: adultos de *C. sanguinea* e larvas de *H. axyridis*) (Tabela 19), enquanto que para larvas de *Hi. convergens* não foi observado diferença entre no número de *reads* detectadas utilizando ambos tamanhos de insertos na biblioteca.

Tabela 14. Número médio das reads de presas por biblioteca com insertos de 350 pb e 550 pb.

	Média do número de reads para bibliotecas com inserto de 350 pb	Média do número de reads para bibliotecas com inserto de 550 pb	<i>P</i> 2x	<i>P</i>
Larva de <i>Cycloneda sanguinea</i>	200,6	478,3	0,0035	< 2E-5
Adulto de <i>Cycloneda sanguinea</i>	1054,9	418,0	>0,99999	< 2E-5
Larva de <i>Chrysoperla externa</i>	0,91	47,8	0,0032	0,0007
Larva de <i>Hippodamia convergens</i>	1211,5	1088,9	>0,99999	0,2003
Larva de <i>Harmonia axyridis</i>	2244,8	454,8	>0,99999	< 2E-5

P 2x é a probabilidade de que não haja pelo menos duas vezes mais presas detectadas na biblioteca de 550 pb em comparação com a biblioteca de 350 pb; *P* é a probabilidade de que haja o mesmo número de reads de presas nas bibliotecas de 350 e 550 bp.

O número de espécies contaminantes por biblioteca foi ligeiramente mais baixo nas bibliotecas de 350 pb (1,0 espécie) do que nas bibliotecas de 550 pb (1,4 espécie). O número de *reads* de falsos-positivos foi mais baixo nas bibliotecas com inserto de 350 pb (124 *reads*; 6,2 por biblioteca) do que as bibliotecas de 550 pb (210 *reads*; 42,0 por biblioteca). No entanto, este não foi o caso para as *reads* das presas. Em geral, estes resultados sugerem que a utilização do tamanho de inserto de 350 pb pode ser melhor que a do tamanho de inserto de 550 pb. Comparações adicionais necessitam serem feitas para ter uma resposta definitiva.

6. CONCLUSÃO

Esse trabalho de doutorado avaliou o potencial do método de sequenciamento de alto desempenho do DNA ser utilizado para identificação da biodiversidade de presas consumidas por predadores generalistas coletados a campo, sem que haja etapa intermediária de amplificação de regiões de DNA barcodes das presas. Para isso a sensibilidade, especificidade e variabilidade da detecção da presa foram estudadas a partir de dados coletados de bioensaios de alimentação em laboratório, utilizando três espécies de pragas como presas e quatro espécies de inimigos naturais como predadores. Adicionalmente estudou-se a influência de certos parâmetros bióticos e abiótico na detecção da presa, no decaimento da mesma e no período de detectabilidade. Pôde-se concluir que:

- ❖ O método de detecção de presas testado nesse trabalho possui alto poder de resolução taxonômica e alta sensibilidade, pois todas as presas foram detectadas em nível taxonômico de espécie e com a ingestão de apenas uma unidade de presa ingerida;
- ❖ O método pode gerar falsos-negativos, pois das três presas testadas, a mosca-branca não foi detectada, nem mesmo imediatamente pós-predação pelo predador, bem como falsos-positivos, pois foram detectadas espécies não predadas na maioria das bibliotecas;
- ❖ A detecção da presa decai com o decorrer do tempo pós-predação;
- ❖ As presas apresentam diferentes taxas de decaimento independente da espécie do predador que as consumiu;
- ❖ O período de detectabilidade varia com a biomassa e com as espécies de presa-predador em interação;

- ❖ A forma de alimentação do predador afeta a detecção da presa, sendo maior quando ocorre por mastigação do que por sucção;
- ❖ O número de presas consumidas não afeta a meia-vida de detecção da presa e sua taxa de decaimento, porém o período de detectabilidade é maior quando maior número de presas é ingerida;
- ❖ Predação secundária é detectável, porém não há diferenciação entre presa primária e secundária;
- ❖ A taxa de decaimento da detecção de predação secundária é afetada pelo histórico alimentar do predador primário;
- ❖ As temperaturas de digestão de 20, 25 e 30°C não afetam a detecção da presa;
- ❖ Quanto maior o número de unidades experimentais agrupados por biblioteca, menor o erro-padrão do número de *reads* detectadas para presas por tratamento;
- ❖ Os tamanhos de insertos de 350 e 550 pb não apresentam diferença na detecção das presas, mas sim na detecção de falsos-positivos (menor o inserto, menos falso-positivos detectados);

BIBLIOGRAFIA

- Aboim, M. C. R.; Barbosa, J. C.; Coutinho, H. L. da C.; Rosado, A. S. (2004) Avaliação de diversidade microbiana em amostras de solos técnica do PCR/DGGE (Protocolo Laboratorial) Rio de Janeiro, RJ. ISSN 1517-2627. Empresa Brasileira de Pesquisa Agropecuária Centro Nacional de Pesquisa de Solos Ministério da Agricultura, Pecuária e Abastecimento.
- Adams, P.A. (1987) Studies in neotropical Chrysopidae (Neuroptera) III. Notes on *Nodita amazonica* Navás and *Noenops*, n. sp. *Neuroptera International*, 4: 287-294.
- Agustí, N.; Shayler, S. P.; Harwood, J. D.; Vaughan, I. P.; Sunderland, K. D.; Symondson, W. O. (2003) Collembola as alternative prey sustaining spiders in arable ecosystems: prey detection within predators using molecular markers. *Molecular Ecology*, 12: 3467–3475.
- Albuquerque, G. S. (2009) Crisopídeos (Neuroptera: Chrysopidae). In: Panizzi, A. R.; Parra, J. R. P. (eds) *Biotecnologia e nutrição de insetos: base para o manejo integrado de pragas*. Brasília: Embrapa informação tecnológica, p. 969-1022.
- Alhmedi, A.; Haubruge, É.; Francis, F. (2010) Intraguild interactions implicating invasive species: *Harmonia axyridis* as a model species. *Biotechnologia, Agronomie Société et Environnement, Gembloux*, 14(1):187-201.
- Almeida, L. M.; Silva, V. B. (2002) Primeiro registro de *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae): um coccinélídeo originário da região Paleártica. *Revista Brasileira de Zoologia [online]*. 19(3):941-944.
- 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 Res.* 25(17):3389-3402.
- Andrews, D. (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Andújar, C.; Arribas, P.; Ruzicka, F.; Crampton, Platt, A.; Timmermans, M. J.; Vogler, A. P. (2015) Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology*. 24:3603–3617.
- Aronesty, E. (2011). ea-utils : "Command-line tools for processing biological sequencing data" in: <http://code.google.com/p/ea-utils>.
- Aun, V. (1986) Aspectos da biologia de *Chrysoperla externa* (Hagen, 1861) (Neuroptera: Chrysopidae). p. 65 in: Bortoli, S. A.; Caetano, A. C.; Murata, A. T. Oliveira, J. E. de M. (2006) Desenvolvimento e capacidade predatória de *Chrysoperla externa* (Hagen) (Neuroptera: Chrysopidae) em diferentes presas. *Revista de Biologia e Ciências da Terra*. 6(1):145-152.
- Bernt, M.; Donath, A.; Jühling, F.; Externbrink, F.; Florentz, C.; Fritsch, G.; Pütz, J.; Middendorf, M.; Stadler, P. F. (2013) MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*. 69(2):313-319.
- Blankenship, L.E.; Yayanos, A. A. (2005) Universal iniciadores and PCR of gut contents to study marine invertebrate diets. *Molecular Ecology*. 14: 891–899.

- Bohmann, K.; Monadjem, A.; Lehmkuhl Noer, C.; Rasmussen, M.; Zeale, M. R.; Clare, E.; Jones, G.; Willerslev, E.; Gilbert, M. T. (2011) Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. *PLoS ONE*. 6, e21441.
- Brook, B. W.; Bradshaw, C. J. A. (2006) Strength of evidence for density dependence in abundance time series of 1198 species. *Ecology*. 87:1445-1451.
- Brooke, M. M.; Proske, H. O. (1946) Precipitin test for determining natural insect predators of immature mosquitoes. *Journal of the National Malaria Society*. 5:45-56.
- Bueno, V. H. P. (1999) Protected cultivation and research on biological control of pests in greenhouses in Brazil. Integrated control in glasshouses. *Bull. International Organisation for Biological and Integrated Control/ West Palearctic Regional Section*. 22: 21-24.
- CABI (2017) *Invasive Species Compendium*. Wallingford, UK: CAB International. Link: www.cabi.org/isc.
- Calder, C. R.; Harwood, J. D.; Symondson, W. O. C. (2005) Detection of scavenged material in the guts of predators using monoclonal antibodies: a significant source of error in measurement of predation? *Bulletin of Entomological Research*. 95: 57-62.
- Carvalho, A. D. F.; Amaro, G. B.; Lopes, J. F.; Vilela, N. J.; Filho, M. M.; Andrade, R. (2013) A Cultura do pepino. Circular Técnica 113, Embrapa Hortaliças.
- Carvalho, C. F.; Souza, B. (2002) Potencial de insetos predadores no controle biológico aplicado. In: Parra *et al.* (2002) *Controle biológico no Brasil: parasitoides e predadores*. São Paulo: Manole. Editora Acqua Estúdio Gráfico, 635p.
- Chapman, E. G.; Schmidt, J. M.; Welch, K. D.; Harwood, J. D. (2013) Molecular evidence for dietary selectivity and pest suppression potential in an epigeal spider community in winter wheat. *Biological Control* 65: 7286.
- Cividanes, F. J.; Souza, V. P. (2003) Exigências térmicas e tabelas de vida de fertilidade de *Myzus persicae* (Sulzer) (Hemiptera: Aphididae) em laboratório. *Neotropical Entomology*. 32(3):413-419.
- Clarke, L. J.; Soubrier, J.; Weyrich, L. S.; Cooper, A. (2014) Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*. 14:1160–1170. DOI: 10.1111/1755-0998.12265
- Coelho, R. M. P. (2008) *Fundamentos de ecologia*. Editora Artmed, 256p.
- Cohen, A.C. (1990) Feeding adaptations of some predaceous Hemiptera. *Annals of the Entomological Society of America*. 83:1215–1223.
- Coissac, E.; Riaz, T.; Puillandre, C. (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*. 21:1834–1847.
- Costello, M. J.; Aaltieri, M. A. (1995) Abundance, growth-rate and parasitism of *Brevicoryne brassicae* and *Myzus persicae* (Homoptera: Aphididae) on broccoli grown in living mulches. *Agriculture Ecosystems; Environment*. 52(2-3): 187-196.
- Cruz, I. (1995) A lagarta-do-cartucho na cultura do milho. Circular Técnica. 21, Embrapa/ CNPMS, 45p.

- Cruz, I.; Monteiro, M. A. R. (2004) Controle Biológico da lagarta do cartucho do milho *Spodoptera frugiperda* utilizando o parasitóide de ovos *Trichogramma pretiosum*. Comunicado Técnico. 114, Embrapa Milho e Sorgo.
- Dadd, R. H.; Mitter, T. E. (1996) Permanent culture of an aphid on a totally synthetic diet. *Experientia*. 22, 832.
- Deagle, B. E.; Eveson, J. P.; Jarman, S. N. (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*. 18:2022–2038.
- Deagle, B. E.; Eveson, J. P.; Jarman, S. N. (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*. 11:2039–2048.
- Deagle, B. E.; Gales, N. J.; Evans, K.; Jarman, S. N.; Robinson, S.; Trebilcon, R.; Hindell, M. A. (2007) Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). *PLoS ONE*. 2, e831. DOI:10.1371/journal.pone.0000831.
- Deagle, B. E.; Jarman, S. N.; Coissac, E.; Pompanon, F.; Taberlet, P. (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect matchmatch. *Biology Letters*. 10(9):20140562.
- Deagle, B. E.; Jarman, S. N.; Pemberton, D.; Gales, N. J. (2005) Genetic screening for prey in the gut contents from a giant squid (*Architeuthis* sp). *Journal of Heredity*. 96(4):417-423. DOI:10.1093/jhered/esi036.
- Dicke, M.; Jong, M.de (1988) Prey preference of the Phytoseiid Mite *Typhlodromus pyri* 2. Electrophoretic diet analysis. *Experimental; Applied Acarology*. 4:15-25.
- Douglas, A. E.; van Emden, H. F. (2007) Nutrition and symbiosis. Chapter 5. In: *Aphis as crop pests*. van Emden, H. F.; Harrington, R. p. 105-134. 717p.
- Dunsha, G. (2009) DNA-Based Diet Analysis for Any Predator. *PLoS ONE*. 4(4): e5252
- Durbin, E. G.; Casas, M. C.; Rynearson, T. A.; Smith, D. C. (2008) Measurement of copepod predation on nauplii using qPCR of the *cytochrome oxidase I* gene. *Marine Biology*. 153:699–707.
- Eitzinger, B.; Unger, E. M.; Traugott, M.; Scheu, S. (2014) Effects of prey quality and predator body size in prey DNA detection success in a centipede predator. *Molecular Ecology* 23, 3767–3776.
- Elliott, M.; Whitfield, A. K.; Potter, I.; Blaber, S. J. M.; Cyrus, D. P.; Nordlie, F. G.; Harisson, T. D. (2007) The guild approach the categorizing estuarine fish assemblages: a global review. *Fish and Fisheries*. 8:241-268
- Embrapa Milho e Sorgo (2009) Panomara fitossanitário – Cultura do milho. <http://panorama.cnpms.embrapa.br/insetos-praga/inimigos-naturais/preadores-joaninhas/joaninha-hippodamia-convergens-guerin-meneville-1842-coleoptera-coccinelidae>. Acesso em Março de 2015.
- Ferreira, M. E.; Grattapaglia, D. (1995) Introdução ao uso de marcadores moleculares em análise genética. 3ª ed. Brasília: Embrapa Cenargen. Pp220.
- Ficetola, G. F.; Coissac, E.; Zundel, S.; Riaz, T.; Shehzad, W.; Bessièrè, J.; Tarbelet, P.; Pompanon, F. (2010). An *in silico* approach for the evaluation of DNA barcodes. *Biomed Central Genomics*. 11(1):1.

- Ficetola, G. F.; Pansu, J.; Bonin, A.; Coissac, E.; Giguët-Covex, C.; de Barba, M.; Gielly, L.; Lopes, C. M.; Boyer, F. Pompanon, F.; Rayé, G.; Taberlet, P. (2014) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*. 15:543–556
- Figueira, L. K.; Toscano, L. C.; Lara, F. M.; Boiça Junior, A. L. (2003) Aspectos biológicos de *Hippodamia convergens* e *Cycloneda sanguinea* (Coleoptera: Coccinellidae) sobre *Bemisia tabaci* biótipo B (Hemiptera: Aleyrodidae). *Boletín de Sanidad Vegetal, Plagas*. 29:3-7.
- Firlej, A.; Doyon, J. Harwood, J. D.; Brodeus, J. (2013) A multiapproach study to delineate interactions between carabid beetles and soybean aphids. *Environmental Entomology*. 42:89-96.
- Flint, M. L.; Dreistadt, S. H. (2005) Interactions among convergent lady beetle (*Hippodamia convergens*) releases, aphid populations, and rose cultivar. *Biological Control*. 34:38-46.
- Folmer, O.; Black, M., Hoeh, W.; Lutz, R.; Vrijenhoek, R. (1994) DNA iniciadores for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*. 3(5):294-299.
- Foran, D. R. (2006) Relative degradation of Nuclear and Mitochondrial DNA: an experimental approach. *American Academy of Forensic Sciences*. 51(4):766-770.
- Fournier, V.; Hagler, J. R.; Daane, K. M.; de León, J. H.; Groves, R. L. (2008) Identifying the predator complex of *Homalodisca vitripennis* (Hemiptera: Cicadellidae): a comparative study of the efficacy of an ELISA and PCR gut content assay. *Oecologia* 157:629–640. DOI 10.1007/s00442-008-1095-x.
- Fournier, V.; Hagler, J. R.; Daane, K. M.; de León, J. H.; Groves, R. L.; Costa, H. S.; Henneberry, T. J. (2006) Development and application of a glassy-winged sharpshooter and smoke-tree sharpshooter egg-specific predator gut content ELISA. *Biological Control* 37: 108-118.
- Ghafoor, A.; Mahmood, A.; Iqbal, N.; Khan, M. S.; Aza, F.; Farooq, H. A. (2011) Identification of prey remnants in the gut of spiders by polyacrylamide gel electrophoresis (PAGE) of estera isozymes. *World Applied Sciences Journal* 14(4):586-590.
- Godfrey, L. D.; Rosenheim, J. A.; Goodell, P. B. (2000) Cotton aphid emerges as major pest in SJV cotton. *California Agriculture*. 54:26-29.
- Greenstone, M. H.; Bennett, A. F. (1980) Foraging strategy and metabolic rate in spiders. *Ecology*, 61, 1255–1259.
- Greenstone, M. H.; Shufran, K. A. (2003) Spider predation: species-specific identification of gut contents by polymerase chain reaction. *The Journal of Arachnology*. 31:131–134.
- Greenstone, M. H.; Payton, M. E.; Weber, D. C.; Simmons, A. M. (2013) The detectability half-life in arthropod predator–prey research: what it is, why we need it, how to measure it, and how to use it. *Molecular Ecology*. 23(15)3799-3813.
- Greenstone, M. H.; Rowley, D. L.; Weber, D. C.; Payton, M. E.; Hawthorne, D. J. (2007) Feeding mode and prey detectability half-lives in molecular gut-content analysis:

- an example with two predators of the Colorado potato beetle. *Bulletin of Entomological Research*. 97: 201-209.
- Greenstone, M. H.; Payton, M. E.; Weber, D. C.; Simmons, A. M. (2014) The detectability half-life in arthropod predator-prey research: what it is, why we need it, how to measure it, and how to use it. *Molecular Ecology*. 23: 3799–3813.
- Gullan, P. J.; Cranston, P. S. (2007) *Os insetos: um resumo de entomologia*. 3ª edição, São Paulo: Roca, 2007.
- Hagler, J. R.; Naranjo, S. E. (1994a) Qualitative survey of two coleopteran predators of *Bemisia tabaci* (Homoptera: Aleyrodidae) and *Pectinophora gossypiella* (Lepidoptera: Gelechiidae) using multiple prey gut content ELISA. *Environmental Entomology*. 23:193-197.
- Hagler, J. R.; Naranjo, S. E. (1994b) Determining the frequency of heteroptera predation on sweet-potato whitefly and pink-bollworm using multiple ELISAs. *Entomologia Experimentalis et Applicata*. 72:59-66.
- Hagler, J. R.; Naranjo, S. E. (1997) Measuring the sensitivity of an indirect predator gut content ELISA: detectability of prey remains in relation to predator species, temperature, time, and meal size. *Biological Control*. 9:112-119.
- Hagler, J. R.; Naranjo, S. E. (2005a) Use of a gut content ELISA to detect whitefly predator feeding activity after field exposure to different insecticide treatments. *Biocontrol Science and Technology*. 15:321-339.
- Hagler, J. R.; Naranjo, S. E. (2005b) Identifying inter and intra guild feeding activity of an arthropod predator assemblage. *Ecological Entomology*, 38, 258–271
- Hahn, C.; Bachmann, L.; Chevreux, B. (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Research*. 41(13): 1-9.
- Haji, F. N. P.; Mattos, M. A. A.; Alencar, J. A.; Barbosa, F. R.; Paranhos, B. J. (2005). Manejo da mosca-branca na cultura do tomate. Circular técnica. 81, Embrapa Semi-Árido.
- Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*. 68(4):669-85.
- Harmon, J. P.; Andow, D.A. (2007) Behavioral mechanisms underlying ants' density-dependent deterrence of aphid-eating predators. *Oikos*. 116: 1030-1036.
- Harper, G. L.; King, R. A.; Dodd, C. S.; Harwood, J. D.; Glen, D. M.; Bruford, M. W.; Symondson, O.C. (2005) Rapid screening of invertebrate predators for multiple prey DNA targets. *Molecular Ecology*. 14: 819-827.
- Harper, G. L.; Sheppard, S. K.; Harwood, J. D.; Read, D. S.; Glen, D. M.; Bruford, M. W.; Symondson, W. O. (2006) Evaluation of temperature gradient gel electrophoresis for the analysis of prey DNA within the guts of invertebrate predators. *Bulletin of Entomological Research*. 96:295–304.
- Harwood, J. D.; Phillips, S. W.; Sunderland, K. D.; Symondson, W. O. C. (2001) Secondary predation: quantification of food-chain errors in an aphid-spider-carabid system using monoclonal antibodies. *Molecular Ecology*. 10:2049-2057.
- Harwood, J. G.; Obrycki, J. J. (2005) Quantifying aphid predation rates of generalist predators in the field. *European Journal Entomology*. 102:335-350.

- Hatteland, B. A.; Symondson, W. O. C.; King, R. A.; Skage, M.; Schander, C.; Solhoy, T. (2011) molecular analysis of predation by carabid beetles (Carabidae) on the invasive iberian slug *Arion lusitanicus*. *Bulletin of Entomological Research*. 101:675-686.
- Hebert, P. D. N.; Cywinska, A.; Ball S. L.; Dewaard, J. R. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*. 270:313-321.
- Hoffmann, L. V.; Barroso, P. A. V. (2006) Marcadores moleculares como ferramentas para estudos de genética de plantas. Campina Grande, Embrapa Algodão. Documentos, 147. 35p.
- Holopainen, J. K.; Helenius, J. (1992) Gut contents of ground beetles (Col., Carabidae) and activity of these and other epigeal predators during an outbreak of *Rhopalosiphum padi* (Hemiptera: Aphididae). *Acta Agriculturae Scandinavica, Section B - Soil; Plant Science*. 42:57-71.
- Hoogendoorn, M.; Heimpel, G. E. (2001) PCR-BASED gut content analysis of insect predators: using ribosomal *ITS-1* fragments from prey to estimate predation frequency. *Molecular Ecology*, 10, 2059-2067.
- Hoogendoorn, M.; Heimpel, G. E. (2003) PCR-based gut content analysis of insect predators: a field study. 1st International Symposium on Biological Control of Arthropods.
- Ingerson-Mahar, J. (2002) Relating diet and morphology in adult carabid beetles. in: *The agroecology of carabid beetles* (ed. Holland J), pp. 111–136. Intercept, Andover, UK.
- Ings, T. C.; Montoya, J. M.; Bascompte, J.; Blüthgen, N.; Brown, L.; Dormann, C. F.; Edwards, F.; Figuero, D.; Jacobs, U.; Jones, J. I.; Lauridsen, R. B.; Ledger, M. E.; Lewis, H. M.; Olesen, J. M.; Van Venn, F. J. F.; Warren, P. H.; Woodward, G. (2009) Ecological networks - beyond food webs. *Journal of Animal Ecology*. 78:253–269.
- Jarman, S. N.; Deagle, B. E.; Gales, N. J. (2004) Group-specific polymerase chain reaction DNA-based analysis of species diversity and identify in dietary samples. *Molecular Ecology*. 13:1313-1322.
- Ji, Y.; Ashton, L.; Pedley, S. M.; Edwards, D. P.; Tang, Y.; Nakamura, A.; Kitching, R.; Dolman, P. M.; Woodcock, P.; Edwards, F. A.; Larsen, T. H.; Hsu, W. W.; Benedick, S.; Hamer, K. C.; Wilcove, D. S.; Bruce, C.; Wang, X.; Levi, T.; Lott, M.; Emerson, B. C.; Yu, D. W. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*. 16:1245–1257.
- Johanowicz, D. L.; Hoy, M. A. (1996) *Wolbachia* in a predator-prey system: 16S ribosomal analysis of two phytoseiids (Acari: Phytoseiidae) and their prey (Acari: Tetranychidae). *Annual Entomology Society of America*. 89: 435-441.
- Juen, A.; Traugott, M. (2005) Detecting predation and scavenging by DNA gut-content analysis: a case study using a soil insect predator-prey system. *Oecologia*. 142:344-352.
- Junior, M. E. (2011) Controle biológico de insetos pragas. Seminário mosaico ambiental, n. 1.

- Jurado-Rivera, J. A.; Vogler, A. P.; Reid, C. A. M.; Petitpierre, E.; Gómez-Zurita, J. (2009) DNA barcoding insect-host plant associations. *Proceedings of the royal society B*. 276:639-648.
- Kato, C. M.; Bueno, V. H. P.; Moras, J. C.; Aud, A. M. (1999) Criação de *Hippodamia convergens* Guérin-Meneville (Coleoptera: Coccinellidae) em Ovos de *Anagasta kuehniella* (Zeller). *An. Soc. Entomol. Brasil* 28(3): 455-459.
- Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Mentjies, P.; Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647-1649.
- Kerzicnik, L. M.; Chapman, E. G.; Harwood, J. D.; Peairs, F. B.; Cushing, P. E. (2012) Molecular characterization of Russian wheat aphid consumption by spiders in winter wheat. *Journal of Arachnology*. 40:71-77.
- King, R.A.; Read, D.S.; Traugott, M.; Symondson W. O. C. (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology*. 17:947–963.
- King, R. A.; Moreno-Ripoll, R.; Agustí, N.; Shayler, S. P.; Bell, J. R.; Bohan, D. A.; Symondson, W. O. (2010a) Multiplex reactions for the molecular detection of predation on pest and non-pest invertebrates in agroecosystems. *Molecular Ecology Resources*. 11:270–373.
- King, R. A.; Vaughan, I. P.; Bell, J. R.; Bohan, D. A.; Symondson, W. O. C. (2010b) Prey choice by carabid beetles feeding on an earthworm community analysed using species- and lineage- specific PCR iniciadores. *Molecular Ecology*. 19:1721–1732.
- Kowalczyk, R.; Taberlet, P.; Coissac, E.; Valentini, A.; Miquel, C.; Kamiński, T.; Wójcik, J. M. (2011) Influence of management practices on large herbivore diet— Case of European bison in Białowieża Primeval Forest (Poland) *Forest Ecology and Management*. 261:821–828.
- Leite, M. V. L.; Santos, T. M.; Souza, B.; Calixto, A. M.; Carvalho, C. F. (2008) Biologia de *Aphis gossypii* Glover, 1877 (Hemiptera: Aphididae) em abobrinha cultivar Caserta (*Cucurbita pepo* L.) em diferentes temperaturas. *Ciência e Agrotecnologia, Lavras*. 32(5):1394-1401.
- Li, H.; Gao, J.; Liu, H.; Liu, H.; Liang, A.; Zhou, X.; Cai, W. (2011) The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). *International Journal of Biological Sciences*. 7(6):792-804.
- Lopes, C. A.; Ribeiro, C. S. C.; Cruz, D. M. R.; França, F. H.; Reifschneider, F. J. B.; Henz, G. P.; Silva, H. R.; Pessoa, H. S.; Bianchetti, L. B.; Junqueira, N. V.; Makishima, N.; Fontes, R. R.; Carvalho, S. I. C.; Marouelli, W. A.; Pereira, W. (2007) Pimenta (*Capsicum* spp.). Embrapa Hortaliças Sistemas de Produção, 2 ISSN 1678-880x Versão Eletrônica.
- Lopes, G. da S.; Lemos, R. N. S. de; Machado, K. K. G.; Maciel, A. A. S.; Ottati, A. L. T. (2008) Biologia de *Spodoptera frugiperda* (J. SMITH) (Lepdoptera: Noctuidae)

- em folhas de mandioca (*Manihot esculenta*, CRANTZ). *Revista Caatinga*. 21(3):134-140.
- Losey, J. E.; Vaughan, M. (2006) The economic value of ecological services provided by insects. *BioScience*. 56(4):311-323.
- Madeira, F.; di Lascio, A.; Carlino, P.; Costantini, M. L.; Rossi, L.; Pons, X. (2014) Stable carbon and nitrogen isotope signatures to determine predator dispersal between alfalfa and maize. *Biological Control*. 77:66-75.
- Marco, D. (2010). *Metagenomics: theory, methods and applications*. Caister Academic Press [S.l.] ISBN 978-1-904455-54-7.
- McMillan, S.; Kuusk, A.-K.; Carrel-Lundhagen, Ekbohm, B (2007) The influence time and temperature on molecular gut content analysis: *Adalia bipunctata* fed with *Rhopalosiphum padi*. *Insect Science*. 14:353-358.
- Michaud, J. P. (2002) Invasion of the Florida citrus ecosystem by *Harmonia axyridis* (Coleoptera: Coccinellidae) and asymmetric competition with a native species, *Cycloneda sanguinea*. *Environmental Entomology*. 31(5):827-835.
- Miller, G. L.; Oswald, J. D.; Miller, D. R. (2004) Lacewings and scale insects: a review of predator/prey associations between the Neuropterida and Coccoidea (Insecta: Neuroptera, Raphidioptera, Hemiptera). *Annals Entomology Society of America*. 97:1103-1125.
- Miranda, J. E.; Suassuna, E. D. (2004) Guia de identificação e controle das principais pragas e doenças do algodoeiro. Circular técnica 76. Embrapa, Campina Grande, PB.
- Moreira, H. J. da C.; Aragão, F. D. (2009) Manual de pragas do milho. Campinas, São Paulo.
www.agrolink.com.br/downloads/manual%20de%20pragas%20do%20milho.pdf
- Moore, J. W.; Semmens, B. X. (2008) Incorporating uncertainty and prior information into stable isotope mixing models. *Ecology Letters*. 11: 470–480.
- Murray, D. C.; Bunce, M.; Cannell, B. L.; Oliver, R.; Houston, J.; White, N. E.; Barrero, R. A.; Bellgard, M. I. Haile, J. (2011). DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS ONE*, 6(10), e25776.
- Murray, R. A.; Solomon, M. G. (1978) A rapid technique for analyzing diets of invertebrate predators by electrophoresis. *Annals of Applied Biology*. 90:7-10.
- Muyser, G.; Samalla, K. (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie van Leeuwenhoek*. 73:127-141.
- Naranjo, S. E.; Hagler, J. R. (2001) Toward the quantification of predation with predator gut immunoassays: a new approach integrating functional response behavior. *Biological Control*. 20:175–189.
- Nava, D. E.; Afonso, A. P. S.; Melo, M.; Couto, M.; Anjos e Silva, S. D. (2009) Controle biológico aplicado e natural da lagarta-do-cartucho do milho. Pelotas: Embrapa Clima Temperado.
- Neto, M. P.; Carvalho, C. F.; Reis, P. R.; Canta-Cecília, L. V. C.; Souza, B.; Alcantra, E.; Silva, R. A. (2008) Aspectos biológicos de *Chrysoperla externa* (hagen) predando

- Oligonychys ilicis* (McGregor) e *Planococcus citri* (Risso). *Coffee Science*, Lavras. 3(2):85-93.
- Oliveira, N. C. de; Wilcken, C. F.; Matos, C. A. O. de. (2004). Ciclo biológico e predação de três espécies de coccinelídeos (Coleoptera, Coccinellidae) sobre o pulgão-gigante-do-pinus *Cinara atlantica* (Wilson) (Hemiptera, Aphididae). *Revista Brasileira de Entomologia*, 48(4), 529-533
- Oliveira, J. E. de M. (2006) Desenvolvimento e capacidade predatória de *Chrysoperla externa* (Hagen) (Neuroptera: Chrysopidae) em diferentes presas. *Revista de Biologia e Ciências da Terra*. 6(1):145-152.
- Ostrom, P. H.; Colunga-Garcia, M.; Gage, S. H. (1997) Establishing pathways of energy flow for insect predators using stable isotope ratios: field and laboratory evidence. *Oecologia*. 109:109-113.
- Paula, D. P.; Linard, B.; Andow, D. A.; Sujii, E. R.; Pires, C. S. S; Vogler, A. P. (2015) Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics. *Molecular Ecology Resources*. 15:880–892.
- Paula, D. P.; Linard, B.; Platt, A. C.; Srivathsan, A.; Timmermans, M.; Sujii, E.; Pires, C.; Machado, L.; Andow, D. A.; Vogler, A. (2016) Uncovering trophic interactions in arthropod predators through dna shotgun-sequencing of gut contents. *PLoS ONE* 11, p. e0161841.
- Pegard, A.; Miquel, C.; Valentini, A.; Coissac, E.; Bouvier, F.; François, D.; Taberlet, P.; Engel, E.; Pompanon, F. (2009) Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from feces. *Journal of Agricultural and Food Chemistry*. 57:5700–5706.
- Pendleton, R. C.; Grundmann, A. W. (1954). Use of P32 in tracing some insect-plant relationships of the thistle *Cirsium undulatum*. *Ecology*. 35:187–191.
- Pereira, R. B.; Pinheiro, J. B.; Guimarães, J. A.; Reis, A. (2012) Doenças e pragas do jiloeiro. Circular Técnica 106, Embrapa Hortaliças.
- Pervez, A.; Omkar (2006) Ecology and biological control application of multicoloured Asian ladybird, *Harmonia axyridis*: a review. *Biocontrol Science and Technology*. 16(1/2):111-128.
- Piñol, J.; Mir, G.; Gomez-Polo, P. Agust, N. (2015) Universal and blocking primer mismatchmatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*. 15:819–830.
- Pinto, A. J.; Raskin, L. (2012) PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One*. 7(8)e43093
- Pinto, A. S.; Parra, J. R. P.; Oliveira, H. N. (2008) Guia de campo de pragas e insetos benéficos da soja. Piracicaba: CP2, 64p.
- Poletti, M.; Alves, E. B. (2011) Resistência de mosca-branca a inseticidas. IRAC-BR, Comitê Brasileiro de ação a resistência a inseticidas.
- Pompanon, F.; Deagle, B. E.; Symindson, W. O. C.; Brown, D. S.; Jarman, S. N.; Taberlet, P. (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* 21, 1931–1950.

- Rayé, G.; Miquel, C.; Coissac, E.; Redjadj, C.; Loison, A.; Taberlet, P. (2011) New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study. *Ecological Research*. 26(2):265-276.
- Read, D. S.; Sheppard, S. K.; Bruford, M. W.; Glen, D. M.; Symondson W. O. C. (2006) Molecular detection of predation by soil microarthropods on nematodes. *Molecular Ecology*, 15:1963– 1972.
- Reinfelder, J. R.; Fisher, N. S. (1994) Retention of elements absorbed by juvenile fish (*Menidia menidia*, *Menidia beryllina*) from zooplankton prey. *Limnology Oceanography*. 39(8):1783-1789.
- Ribeiro, M. J. (1988) Biologia de *Chrysoperla externa* (Hagen, 1861) (Neuroptera: Chrysopidae) alimentada com diferentes dietas. Dissertação de mestrado em Agronomia. Lavras – UFLA. p131.
- Root, E. P. (1973) Organization of a plant-arthropod association in simple and diverse habitats: the fauna of collards (*Brassica oleracea*). *Ecological Monograph* 43:95-124.
- Rougerie, R.; Smith, M. A.; Fernandez-Triana, J.; Lopez-Vaamonde, C.; Ratnasingham, R.; Hebert, P. D. N. (2010) Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval host. *Molecular Ecology*. 20(1):179-186.
- Ru, N.; Witcomb, W. H.; Murphey, M.; Carlisle, T. C. (1975) Biology of *Chrysopa lanata* (Neuroptera: Chrysopidae). *Annals of the Entomological Society of America*.
- Saini, E. D. (2004) Presencia de *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae) en la provincia de Buenos Aires. *Aspectos Biológicos Morfológicos* 33: 151--160.
- Salas, J.; Mendonza, O. (1995) Biology of the sweetpotato whitefly (Homoptera: Aleyrodidae) on tomato. *Florida Entomologist*, Winter-Haven. 78:154-160.
- Saminathan, V. R.; Baskaran, R. K. M.; Mahadevan, N. R. (1999) Biology and predatory potential of green lacewing, *Chrysoperla carnea* (Neuroptera: Chrysopidae) on different insect hosts. *Indian Journal of Agricultural Sciences*. 69(7): 502-50.
- Sanger F; Nicklen S; Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74 (12): 5463–5467.
- Santa-Cecília, L.V. C.; Gonçalves-Genário, R. C. R.; Torrês, R. M.; Nascimento, F, R. (2001) Aspectos biológicos e consumo alimentar de larvas de *Cycloneda sanguinea* (Linnaeus, 1763) (Coleoptera: Coccinellidae) alimentadas com *S. graminum* (Rondani, 1852) (Hemiptera: Aphididae). *Ciência e Agrotecnologia*, Lavras, 25:1273-1278.
- Santos, G. P.; Pinto, A. C. Q. (1981) Biologia de *Cycloneda sanguinea* e sua associação com pulgão em mudas de mangueira. *Pesquisa Agropecuária Brasileira*. 16 (4): 473-476.
- Santos, N. R. P. dos; Santos-Cividanes, T. M. dos; Jorge-Cividanes, F.; Anjos, A. C. R. dos; Oliveira, L. V. L. de. (2009) Aspectos biológicos de *Harmonia axyridis* alimentada com duas espécies de presas e predação intraguilda com *Eriopsis conexa*. *Pesquisa Agropecuária Brasileira*. 44(6).

- SAS Institute Inc. 2014. Base SAS® 9.4 Procedures Guide. Cary, NC: SAS Institute Inc.
- Sarmiento, W. de A.; Aguiar, R. W. de S.; Aguiar, R. de A. S. de S.; Vieira, S. M. J.; Oliveira, H. G. de; Holtz, A. M. (2002) Revisão da biologia, ocorrência e controle de *Spodoptera frugiperda* (Lepidoptera, Noctuidae) em milho no Brasil. *Bioscience Journal*. 18(2):41-48.
- Sattar, M.; Abro, G. H.; Syed, T. S. (2011) Effect of different hosts on biology of *Chrysoperla carnea* (Stephens) (Neuroptera: Chrysopidae) in Laboratory Conditions. *Pakistan Journal Zoology*. 43(6):1049-1054.
- Schmidt, F. G. V.; Monnerat, R. G.; Borges, M.; Carvalho, R. da S. (2001) Metodologia de criação de insetos para avaliação de agentes entomopatogênicos. Circular Técnica 11. Embrapa Recursos Genéticos e Biotecnologia, Brasília – DF.
- Schmidt, J. E. U.; Almeida, J. R. M.; Rosati, C.; Arpaia, S. (2009) Identification of trophic interactions between *Macrolophus caliginosus* (Heteroptera: Miridae) and *Myzus persicae* (Homoptera: Aphididae) using real time PCR. *BioControl*, 54:383–391.
- Scrimgeour, C. M.; Gordon, S. C.; Handley, I. I.; Woodford, J. A. T. (1995) Trophic levels and anomalous ¹⁵N of insects on raspberry (*Rubus idaeus* L.) Isotopes in Environmental and Health Studies. 31:107-115.
- Sheppard, S. K.; Bell, J. R.; Sunderland, K. D.; Fenlon, J.; Skirvin, D.; Symondson, W. O. C. (2005) Detection of secondary predation by PCR analyses of the gut contents of invertebrate generalist predators. *Molecular Ecology*. 14: 4461-4468.
- Sheppard, S. K.; Harwood, J. D. (2005) Advances in molecular ecology: tracking trophic links through predator-prey food-webs. *Functional Ecology*. 19(5):751-762.
- Sicsú, P. R. Escolha por sítios específicos para oviposição como fator de estruturação da comunidade de joaninhas (Coleoptera: Coccinellidae) em agroecossistemas no Distrito Federal (2013). 86 p. Dissertação de Mestrado. Universidade de Brasília, Instituto de Biologia Celular, Departamento de Ecologia, Programa de Pós-graduação em Ecologia. Brasília, 05 de abril de 2013.
- Silva, A. G. A.; Gonçalves, C. R.; Galvão, D. M. (1968) Quarto catálogo dos insetos que vivem nas plantas no Brasil, seus parasitos e predadores. Rio de Janeiro: Ministério da Agricultura, 622p.
- Silva, A. C.; Gomes, C. C.; Sacramento, F. Z.; Garcia, G. L.; Schultz, H.; Pian, L. B.; Almeida, L. H. M.; Aguiar, L. A.; Tamashiro, L. A. G. (2013) Guia para o reconhecimento de inimigos naturais de pragas agrícolas. Editora técnica. — Brasília, DF: Embrapa. 47 p.
- Silva, J. B. C.; Giordano, L. B.; Furumoto, O.; Boiteux, L. S.; França, F. H.; Bôas, G. L. V.; Branco, M. C.; Medeiros, M. A.; Marouelli, W.; Silva, W. L. C.; Lopes, C. A.; Ávila, A. C.; Nascimento, W. M.; Pereira, W. (2006) Cultivo de tomate para industrialização. Embrapa Hortaliças Sistemas de Produção, 1 - 2ª Edição. Versão Eletrônica. Acesso em março 2014.
- Silva, M. C.; Lemos, R. N. S.; Lima, L. H. C.; Goulart, L. R.; Pereira, S. R. F. (2009) Variabilidade genética de *Bemisia tabaci* (Gennadius) biótipo B (Hemiptera: Aleyrodidae) em cultivos olerícolas em São Luís, MA. *Neotropical Entomology*, 38(6) 762-768.

- Sint, S.; Niederklapfer, B.; Kaufmann, R.; Traugott, M. (2014) Group-specific multiplex PCR detection systems for the identification of flying insect prey. *PLoS One*. 9(12): e115501.
- Soares, J. J.; do Nascimento, A. R. B.; da Silva, M. V. (2007) Informações sobre *Chrysoperla externa*. Empresa Brasileira de Pesquisa Agropecuária Centro Nacional de Pesquisa de Algodão, Campina Grande, PB. ISSN 0103-0205.
- Soininen, E. M.; Valentini, A.; Coissac, E.; Miquel, C.; Gielly, L.; Brochmann, C. (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology*. 6, 16.
- Solomon, M. G.; Fitzgerald, J.; Murray, R. A. (1996) Electrophoretic approaches to predator/prey interactions. In: *The Ecology of Agricultural pests*. Ed. Symondson, W. O. C.; Liddell, J. E. London, Chapman of Hall, UK.
- Sopp, P. I.; Sunderland, K. D. (1989) Some factors affecting the detection period of aphid remains in predators using ELISA. *Entomologia Experimentalis et Applicata*. 51: 11-20.
- Sopp, P. I.; Sunderland, K. D.; Fenlon, J. S.; Wratten, S. D. (1992) An improved quantitative method for estimating invertebrate predation in the field using an enzyme-linked immunosorbent assay (ELISA). *Journal of Applied Ecology*. 79:295-302.
- Srivathsan, A.; Sha, J. C. M.; Vogler, A. P.; Meier, F. (2015) Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf feeding monkey (*Pygathrix nemaeus*). *Molecular Ecology Resources*. 15:250-261.
- Stenis, M. J. van. (1993). Suitability of *Aphis gossypii* Glov., *Macrosiphum euphorbiae* (Thom.) and *Myzus persicae* Sulz. (Hemiptera: Aphididae) as host for several aphid parasitoid species (Hymenoptera: Braconidae). *Bull. IOBC/WPRS*, XVI/2: 157-160.
- Steffan, S. A.; Daane, K. M.; Mahr, D. L. (2001). ¹⁵N-enrichment of plant tissue to mark phytophagous insects, associated parasitoids, and flower-visiting entomophaga. *Entomologia Experimentalis et Applicata*, 98:173–180.
- Sunderland, K. D.; Crook, N. E.; Stacey, D. L.; Fuller, B. J. (1987) A study of feeding by polyphagous predators on cereal aphids using ELISA and gut dissection. *Journal of Applied Ecology*. 24:907-933.
- Symondson, W. O. C (2002) Molecular identification of prey in predator diets. *Molecular Ecology*. 11, 627–641.
- Symondson, W. O. C.; Gassul, T.; Liddell, J. E. (1999) Rapid identification of adult whiteflies in plant consignments using monoclonal antibodies. *Annals of Applied Biology*. 134: 271-276.
- Taberlet, P.; Coissac, E.; Pompanon, F.; Brochmann, C.; Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*. 21: 2045–2050.
- Tang, M.; Tan, M.; Meng, G.; Yang, S.; Su, X.; Liu, S.; Song, W.; Li, Y.; Wu, Q.; Zhang, A.; Zhou, X. (2014) Multiplex sequencing of *pooled* mitochondrial genomes-a

- crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Research Advance*. 42:(22), e166.
- Timbo, R. V.; Togawa, R. C.; R, Costa, M. M. C.; Andow, A.; Pires, D. P. (2017) Mitogenome sequence accuracy using different elucidation methods. *PLoS ONE* 12(6): e0179971.
- Thomas, T.; Gilbert, J.; Meyer, F. (2012) Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*. 2:3. <http://microbialinformaticsj.com/content/2/1/3>
- Togni, P. H. B. Habitat manipulation for conservation biological control in organic vegetable crops. 2014. 83 f. Tese (Doutorado em Entomologia) – Universidade Federal de Viçosa, Minas Gerais. 2014.
- Tollit, D. J.; Schulze, A. D.; Trites, A. W.; Olesiuk, P. F.; Crockford, S. J.; Gelatt, T. S.; Ream, R. R.; Miller, K. M. (2009) Development and application of DNA techniques for validating and improving pinniped diet estimates (2009). Publications, Agencies and Staff of the U.S. Department of Commerce. Paper 205.
- Traugott, M.; Kamenova, S.; Ruess, L.; Seeber, J. Plantegenest, M. (2013) Empirically characterising trophic networks: what emerging DNA-based methods, stable isotope and fatty acid analyses can offer. In: *Ecological Networks in a Agricultural World*. Academic Press, 524p.
- Triltsch, H. (1999) Food remains in the guts of *Coccinella septempunctata* (Coleoptera: Coccinellidae) adults and larvae. *European Journal Entomology*. 96:355–364.
- Valentini, A.; Miquel, C.; Nawaz, M. A.; Bellemain, E.; Pompanon, E.; Gielly, L.; Cruaud, C.; Nascetti, G.; Wincker, Q.; Swenson, J. E.; Taberlet, P. (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Molecular Ecology Resources*, 9: 51–60.
- Valicente, F. H. (1989) Levantamento dos inimigos naturais de *Spodoptera frugiperda* (J. E. Smith, 1797) (Lepidoptera: Noctuidae) em diferentes regiões do estado de Minas Gerais. *Anais da Sociedade Entomológica do Brasil*. 18(1):119-130.
- Vandenger, N.; Gordon, R. D. (1988) The Coccinellidae (Coleoptera) of South America, part I. A revision of the genus *Erythineda* Timberlake, 1943. *Revista Brasileira de Entomologia*. 32:31-43.
- VanGuilder, H. D.; Vrana, K. E.; Freeman, W. M. (2008) Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* 25th Anniversary. 44(5):619–626.
- Vestheim, H.; Jarman, S. N. (2008) Blocking iniciadores to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*. 5:12.
- Von Berg, K.; Traugott, M.; Symondon, W. O. C.; Scheu, S. (2008) The effects of temperature on detection of prey DNA in two species of carabid beetle. *Bulletin of Entomological Research* 98, 263–269.
- Walde, S. J. & Murdoch, W. W. (1988) Spatial density dependence in parasitoids. *Annual Review of Entomology*. 33:441-466.

- Walrant, A.; Loreau, M. (1995) Comparison of isoenzyme electrophoresis and gut content examination for determining the natural diets of the groundbeetle species *Abax ater* (Coleoptera: Carabidae). *Entomologia Generalis*. 19:253–259.
- Weber, D. C.; Lundgren, J. G. (2009) Detection of predation using qPCR: effect of prey quantity, elapsed time, chaser diet, and sample preservation on detectable quantity of prey DNA. *Journal of Insect Science*. 9:41, 12p.
- Winemiller, K. O.; Jepsen, D. B. (1998) Effects of seasonality and fish movement on tropical river food webs. *Journal Fish Biology*. 53:267-296.
- Wisler, G. J.; Duffus, E.; Liu, K. Y.; Li, R. H. (1998) Ecology and epidemiology of whitefly-transmitted closteroviruses. *Plant Disease*. 82: 270-275.
- Wolfram Research(2010)., Inc., Mathematica, Version 8.0, Champaign, IL
- Zeale, M. R. K.; Butlin, R.; Barker G. L. A. Lees, D. C.; Jones, Gareth. (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*. 11(2): 236-244.
- Zhang, G.-F.; Lu, Z.-C.; Wan, F.-H.; Lovei, G. (2007) Real-time PCR quantification of *Bemisia tabaci* (Homoptera: Aleyrodidae) B-biotype remains in predator guts. *Molecular Ecology*. 7:947–954.
- Zhou, X.; Li, Y.; Liu, S.; Yang, Q.; Su, X.; Zhou L.; Tang, M.; Fu, R.; Li, J.; Huang, Q. (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*. 2:4.



ANEXO 1 – Artigo publicado

Mitogenome sequence accuracy using different elucidation methods

OPEN ACCESS

Citation: Velozo Timbo¹ R, Coiti Togawa R, M. C.

Costa M, A. Andow D, Paula DP (2017) Mitogenome sequence accuracy using different elucidation methods. PLoS ONE 12(6): e0179971.

<https://doi.org/10.1371/journal.pone.0179971>

Editor: Peng Xu, Xiamen University, CHINA

Received: March 30, 2017

Accepted: June 7, 2017

Published: June 29, 2017

Copyright:© 2017 Velozo Timbo¹ et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The mitogenome sequences deposited in *GenBank* (accession codes in [Table 2](#)). The library datasets are deposited at Figshare (DOI: <https://doi.org/10.6084/m9.figshare.3798543.v2>).

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

RESEARCH ARTICLE

Renata Velozo Timbo^{1,2}, Roberto Coiti Togawa¹, Marcos M. C. Costa¹, David A. Andow³, Debora P. Paula^{1*}

1 Embrapa Genetic Resources and Biotechnology, Parque Estaco Biolgica, W5 Norte, Braslia, DF, Brazil, **2** University of Braslia, Campus Universitario Darcy Ribeiro, Braslia, Distrito Federal, Brazil, **3** Department of Entomology, University of Minnesota, 219 Hodson Hall, 1980 Folwell Ave., St. Paul, MN, United States of America

* debora.pires@embrapa.br

Abstract

Mitogenome sequences are highly desired because they are used in several biological disciplines. Their elucidation has been facilitated through the development of massive parallel sequencing, accelerating their deposition in public databases. However, sequencing, assembly and annotation methods might induce variability in their quality, raising concerns about the accuracy of the sequences that have been deposited in public databases. In this work we show that different sequencing methods (number of species pooled in a library, insert size and platform) and assembly and annotation methods generated variable completeness and similarity of the resulting mitogenome sequences, using three species of predaceous ladybird beetles as models. The identity of the sequences varied considerably depending on the method used and ranged from 38.19 to 90.1% for *Cycloneda sanguinea*, 72.85 to 91.06% for *Harmonia axyridis* and 41.15 to 93.60% for *Hippodamia convergens*. Dissimilarities were frequently found in the non-coding A+T rich region, but were also common in coding regions, and were not associated with low coverage. Mitogenome completeness and sequence identity were affected by the sequencing and assembly/annotation methods, and high within-species variation was also found for other mitogenome depositions in *GenBank*. This indicates a need for methods to confirm sequence accuracy, and guidelines for verifying mitogenomes should be discussed and developed by the scientific community.

Introduction

Mitochondrial DNA (mtDNA, mitochondrial genome or mitogenome) is used as a fundamental genetic marker in many research areas, including population genetics, evolutionary biology, phylogenetics and phylogeography, biodiversity studies, and molecular ecology [1–3]. Insect mitochondrial genomes are mostly about 15–22 kbp, circular, double-stranded DNAs that encode a set of 37 genes and a large control region (designated as the A+T-rich region in insects) [3,4]. The genes comprise 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNAs) and two ribosomal RNAs (rRNAs). The A+T-rich region is responsible for regulating

transcription and replication [5] and accounts for much of the mitogenome size variation across insect species [6].

Early elucidation of mitogenomes relied on long range PCR (LR-PCR) plus primer walking coupled with Sanger dideoxy sequencing. More recently, new approaches, using next-generation sequencing (NGS) have become common, including LR-PCR plus NGS, RNA sequencing (RNAseq) plus gap filling, and direct shotgun sequencing [3,7–9].

The assembly and annotation of the early mitogenomes usually relied on BLAST searches to identify protein-coding genes (PCGs) and covariance analyses to identify tRNAs [10], or the use of early assemblers (*e.g.*, TIGR [11]; CAP [12]). Concomitant with the development of NGS, numerous methods for rapidly assembling (*e.g.*, Celera [13]; SOAPdenovo, [14]; MITObim [15]) and annotating (*e.g.*, DOGMA [16]; MOSAS [17]; MITOS [18]) mitogenomes have been developed and used [10, 19–21]. These exciting developments have caused the rate of mitogenome elucidation and deposition to increase tremendously (Fig 1). As a result, there are 4,213 Insecta mitogenomes deposited at *GenBank* (March 28th, 2017).

Public DNA databases (*GenBank*, DDBJ and EMBL) contain mitogenomes obtained through a variety of methodologies, which are often not specified by the depositor and lack an associated publication. Mitogenome sequences have originated from: libraries with pooled genetic material of several species [20,22] or with single species [23]; *de novo* assembly [20] or assembly using a previously sequenced reference genome or individual mito-genes from a target species as ‘seeds’ [15]; and annotation by tRNA identification (*e.g.*, COVE) [24] followed by manual curation [20] or automated pipelines (MITOS) [18].

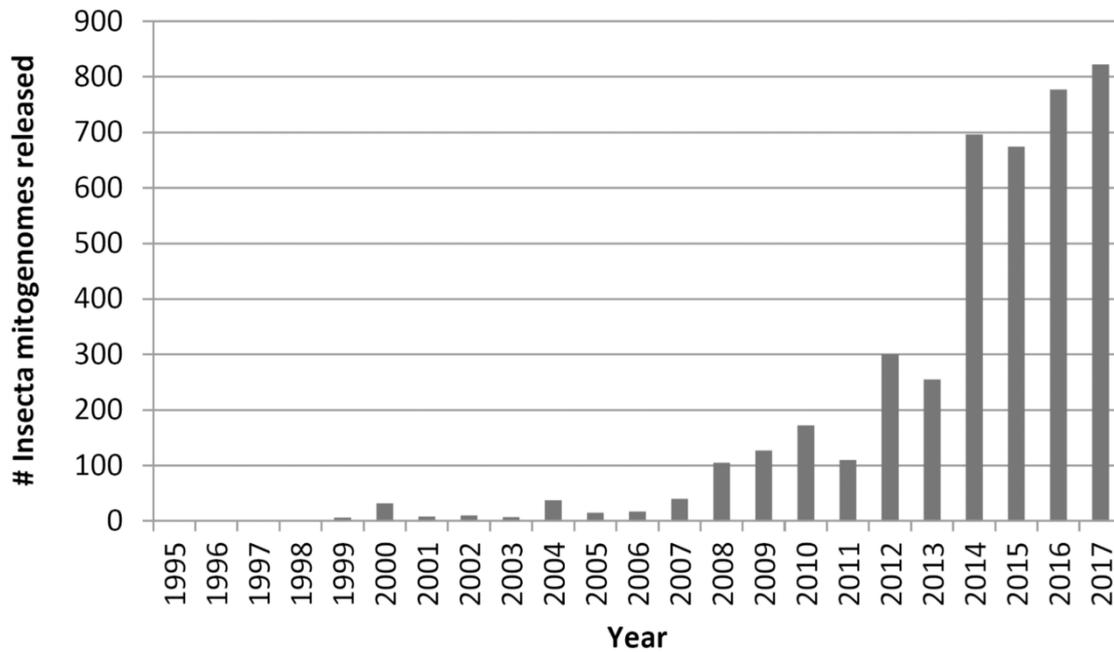


Fig 1. Release of Insecta mitogenome sequences in GenBank since 1995 (2017 covers only the first three months of the year). This information was obtained by searching *insecta[organism] AND mitochondrion[tij] AND genome[tij]* in the Nucleotide database and filtering the results by release date by using the "Release date" search criterion in the left navigation panel.

Despite acknowledgment that assembly and annotation software/methods vary considerably in their quality [25], there has been limited information about the accuracy of the deposited sequences and their annotations, introducing uncertainty about the mitogenome sequences. Public DNA databases are repositories (libraries) and the depositor is expected to provide sequences that meet basic quality criteria regarding the PCG annotations (*e.g.*, presence of start and stop codons, no internal stop codon). The depositor is assumed to have accurate sequences and is not required to provide methodological detail or a proof of quality.

Pooling multiple species in a single library would reduce the cost of library construction compared to single species libraries [20,22], but may increase the probability of interspecific chimeras, resulting in less accurate sequences. *De novo* assembly methods may generate more fragmented assemblies, but use of reference genomes may bias the sequence to the reference or propagate errors in the reference. In this work, we examined the effect of using different sequencing methods (number of species in a library, insert size, sequencing platform) and different assembly and annotation methods in the elucidation of mitogenomes of three species of ladybird beetles.

Material and methods

Sample preparation and DNA shotgun-sequencing

Adult *Cycloneda sanguinea*, *Harmonia axyridis* and *Hippodamia convergens* (Coleoptera: Coccinellidae) were collected from agricultural fields in the Federal District, Brazil. All collections were authorized by SISBIO (authorization number 36950), and access to the genetic heritage and transportation of biological material were authorized by IBAMA (authorization number 02001.008598/2012-42). These species are widespread and important natural enemies of agricultural pests in Brazil and other countries. Three specimens of each

species (about 25 mg of body weight) were used to extract the total DNA separately from each species by DNAeasy Blood and Tissue Kit (Qiagen, Germany), following the manufacturer's instructions. The purified DNA was quantified with Qubit HS (Invitrogen, USA). Two sequencing methods were used. In the first, the three species were pooled without individual tagging after the DNA in the samples was normalized to an equal concentration of 1 µg to make one TruSeq Nano library (Illumina, USA). The library was sequenced on the MiSeq (2×250 bp, insert size 850 bp, 500 cycle kit) using 15% of the flowcell (Illumina, USA). In the second, single species libraries were used with 1 µg of DNA per sample to construct individual TruSeq Nano libraries and sequenced with HiSeq 2500 Rapid Mode (2×250 bp, insert size 550 bp, 250 cycle kit) using 2% of the flowcell (Illumina, USA). We switched to the HiSeq platform due to its much higher sequencing depth and cost-effectiveness.

Mitogenome assembly

Two mitogenome assembly methods were used: *de novo* and using a reference mitogenome. The first method consisted of mining reads in the dataset with similarity to Insecta mitogenomes available at *GenBank*, using the BLASTn algorithm (BLAST v2.2.27; E-value 1e-5; maximum target sequences 1; DUST filtering disabled [26]), according to Crampton et al. [20]. The mined mito-reads were assembled using Celera v7.0 (CA) [13] and IDBA-UD v1.1.1 [27].

For Celera, we used largely default settings without any additional pre-processing of the reads

(doToggle = 1; toggleUnitigLength = 500; unitgger = bogart). Prior to assembly with IDBAUD, the reads were passed through a quality control step with PrinSEQ-lite v0.19.2 [28] to trim low quality bases and remove short sequences (minimum length 150 bp; trim 3' bases below Q20; minimum mean quality Q25; no Ns). Pairs where both reads passed quality control were extracted using a cdbfasta pipeline [20] and converted to Fasta format. The scaffolds generated by both assemblers were concatenated in Geneious v7.0.5 [29] using the parameters: no gaps allowed, minimum overlap 150, maximum mismatches per read 0, minimum overlap identity 99%, maximum ambiguity 1.

To assign the assembled mitogenomes to particular species in the pooled samples, we used barcodes COX1 and *cytb* as 'baits'. Fragments of the 5' portion ('barcode region') of the mitochondrial genes *cytochrome oxidase 1* (COX1) and *cytochrome oxidase b* (*cytb*) were determined independently for each species by sequencing the PCR amplicons generated respectively by the mitochondrial universal primer pairs Jerry; Stev-Pat_R and Sytb-F; Sytb_R, according to Timmermans et al. [30]. The amplicons were checked on 1% agarose gel, purified using QIAquick PCR Purification kit (Qiagen), and sequenced by Sanger methodology (ABI 3700 sequencer). The barcode baits were assigned to the assembled mitogenomes using the Custom BLAST tool (E-value = 0) in Geneious v7.0.5 [29], selecting hits with a minimum length of 95% of the query coverage with at least 98% identity.

In the second method, the quality of the raw dataset was evaluated by FastQC v0.11.3 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and the adaptors were trimmed by Fastqc-mcf v1.04.807 [31] and Cutadapt v.1.9.1 [32] before assembly. The mitogenomes were assembled by MITObim v1.8 [15] using as reference the mitogenome of the coccinellid *Coccinella septempunctata* (*GenBank*: JQ321839.1) and the 5' portion of the

genes COX1 and *cytb* of each target species as baits. The same reference mitogenome was used for the three target coccinellid species because it was the most closely related species to all three targets that was available at *GenBank* at the time of the analysis. The reference and all three targets are in the subfamily Coccinellinae and tribe Coccinellini, and no other species in the tribe was available at *GenBank* to use as a reference, except for *H. axyridis*. However, we considered that using *H. axyridis* as a reference would unduly bias the assembly of the mitogenomes.

The sequencing coverage [33] for each of the four elucidated mitogenomes for each species was obtained by mapping the reads (Q20<90%) in a source library to the respective mitogenome with Geneious v7.0.5 [29] using the parameters: no gaps allowed, minimum overlap 30, maximum mismatches per read 0, minimum overlap identity 99%, maximum ambiguity 4. For methods 2 and 4, in which the assembly method was MITObim, we mapped the reads retrieved in the final iteration.

Annotations

Two annotation methods were used. The first was used with the *de novo* assembly method and consisted of annotating the obtained mitogenome scaffold for the expected tRNA sequences using COVE v2.4.4 [24] with covariance models of Coleoptera [30] according to Crampton et al. [20]. The annotated tRNAs were imported to Geneious v7.0.5 [29] and the target mitogenome was aligned with the annotated mitogenome of the coccinellid *Co. septempunctata* (*GenBank*: JQ321839.1), which was used again as a reference, using standard parameters in MAFFT v7.017 [34]. After this, manual curation was performed to check: a) the expected order of the PCGs and tRNAs; b) the match of the predicted open reading frames (invertebrate mitochondrial genetic code) with the pattern of the reference mitogenome (COX1 does not start with an ATG as methionine and some *nad* genes start with ATT); c) the presence of stop codons in the middle of the coding regions of the target sequence; d) any deletions or insertions that would change the translation frame; e) the stop codons at the end of the coding sequence (the mitochondrial stop codons are TAA, TAG, TA or just T); and f) the orientation of the genes (*e.g.*, some *nad* genes are reverse-oriented). Then, the annotation was transferred to the target mitogenome (selection "Annotation">"Copy to all selected region"), and the annotated target mitogenome was extracted from the alignment. Mitogenome completeness was checked when the length was longer than 16 kb and when both *nad2* and 12S rRNA were present by searching for overlap between the starting and ending regions by aligning the first and last few 100 bp using MAFFT v7.017 [34].

For the second method, which was used with the MITObim assembly method, mitogenome annotations were carried out by MITOS (v763), available for online use at <http://mitos.bioinf.uni-leipzig.de/> [18]. The sequences (Fasta format) of the mitogenomes were uploaded individually for each species to the MITOS server, using the default parameters and invertebrate genetic code. The GFF files generated by MITOS were imported to Geneious v7.0.5 and combined with the original Fasta sequences.

The annotated mitogenomes generated using the combination of all these methods (Table 1) were deposited at *GenBank* (KJ778883, KJ778886, KJ778888, KU877027, KU877028, KU877030, KX755335, KX755334, KX755332, KU877170, KX755330 and KX755331) using the BankIt submission tool.

Mitogenome comparisons

The mitogenomes from the four different methods (two sequencing methods each with two assembly/annotation methods) were compared regarding number of genes obtained and sequence identity using MAFFT alignment v7.017 [34] with the algorithm auto, scoring matrix BLOSUM62, gap opening penalty 1.53, and offset value of 0.123 within Geneious v7.0.5. Annotations were compared regarding gene location/position, length, and order, number of genes annotated and gene transcriptional direction.

Results

The number of raw reads after quality control for the pooled library was 1,663,759 and for the single libraries: *C. sanguinea* 4,587,115; *H. axyridis* 3,233,855; *Hi. convergens* 2,827,085. The lower number of raw reads in the pooled library might have been related to pooling of the species, the longer insert size (850 versus 550 bp) and the use of the MiSeq platform. There was variation in the number of reads mapped to the assembled mitogenomes across pipelines (Table 2), which was directly related to the coverage (Fig 2), but not related to the number of raw reads. Different methods conferred different average coverage. The mitogenomes of *C. sanguinea* and *Hi. convergens* were elucidated for the first time.

Mitogenome completeness

The mitogenomes of the three ladybird beetle species differed across pipelines within species regarding completeness (number of genes) (Table 2). Overall, the coverage was good (>10 [35]) except for parts of *C. sanguinea* method 2 and a few nucleotides in the other mitogenomes (Fig 2). The mitogenomes assembled *de novo* with customized annotation generally were incomplete, lacking some PCGs and/or tRNAs. Overall for all the species, assembly using MITObim (methods 2 and 4) resulted in the complete number of mito-genes, whether species

Table 1. Methods used to obtain and annotate the coccinellid mitogenomes.

	Method 1	Method 2	Method 3	Method 4
Sequencing method	Pooled DNA samples, library insert size 850 bp, Illumina MiSeq platform	Pooled DNA samples, libarey insert size 850 bp, Illumina MiSeq platform	Single DNA sample, library insert size 550 bp, Illumina HiSeq platform	Single DNA sample, library insert size 550 bp, Illumina HiSeq platform
Assembly	<i>de novo</i>	MITObim	<i>de novo</i>	MITObim
Annotation	customized	MITOS	customized	MITOS

Table 2. Coccinellid mitogenomes obtained using different methods.

	<i>Cycloneda sanguinea</i>	<i>Harmonia axyridis</i>	<i>Hippodamia convergens</i>
Method 1 (Pooled_de novo_customized annotation)			
Length (bp)	15,137	15,319	12,618
Protein-coding genes (PCGs)	13	13	12
tRNAs	22	19	15
rRNAs	2	2	1
Number of reads mapped	4,884	4,857	3,807
Mean coverage and std dev	71.8±68.3	74.3±44.2	68.2±39.8
<i>GenBank</i>	KJ778883	KJ778886	KJ778888
Method 2 (Pooled_MITObim_MITOS)			
Length (bp)	7,826	19,464	15,293
PCGs	9	13	13
tRNAs	14	22	22
rRNAs	0	2	2
Number of reads mapped	16,797	12,030	12,652
Mean coverage and std dev	489.5±679.9	134.0±86.2	200.4±46.0
<i>GenBank</i>	KU877027	KU877028	KU877030
Method 3 (Single_de novo_customized annotation)			
Length (bp)	7,382	11,307	16,023
PCGs	7	12	13
tRNAs	11	14	19
rRNAs	0	1	2
Number of reads mapped	87,672	47,512	94,994
Mean coverage and std dev	2,266.7±495.4	816.8±329.1	1,239.9±366.3
<i>GenBank</i>	KX755335	KX755334	KX755332
Method 4 (Single_MITObim_MITOS)			
Length (bp)	18,715	19,330	18,419
PCGs	13	13	13
tRNAs	22	22	22
rRNAs	2	2	2
Number of reads mapped	38,213	20,022	36,543
Mean coverage and std dev	474.7±161.19	222.2±83.8	411.5±404.4
<i>GenBank</i>	KU877170	KX755330	KX755331

were pooled or not in the libraries. The only exception was for the *C. sanguinea* mitogenome obtained by method 2, which for unknown reasons had no or low coverage in several regions (Fig 2). However, even for methods 2 and 4, it is not clear if the length is complete, as none of the sequences could be circularized.

Mitogenome sequence identities across pipelines

In addition to differences in completeness, there were differences in the sequence identities across pipelines for a same species. Table 3 presents the % identity for all pairwise comparisons of the entire mitogenomes obtained by the four methods for all species, and Fig 3 presents the regions of dissimilarity and gaps among the four methods for each species. Comparing methods 1 vs 2 and 3 vs 4 would reveal the effect of the assembly and annotation methods; comparing methods 1 vs 3 and 2 vs 4 would reveal the effect of the sequencing method (species pooling, library insert size and sequencing platform); and comparing methods 1 vs 4 and 2 vs

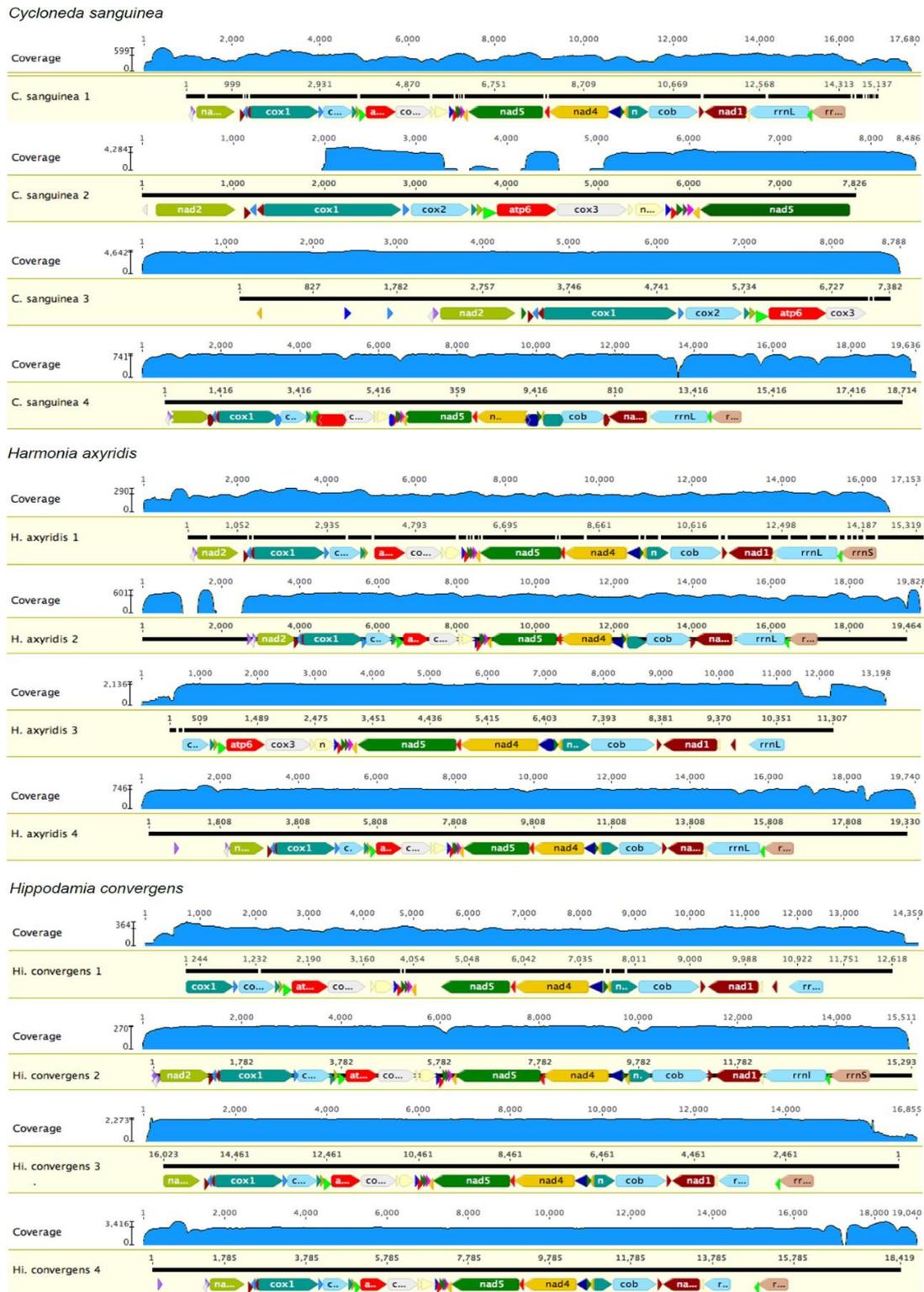


Fig 2. Coverage of the mitogenomes obtained for each species using the four methods. The minimum and maximum (presented in log-scale) is indicated on the ordinate of each figure and the coverage for each nucleotide position is indicated by the height of the blue line. The black bar represents the consensus nucleotide sequence and horizontal thin lines represent gaps. In each figure, the number after the species mitogenome name indicates the method used to assemble and annotate.

3 would reveal the combined interaction effects of the two sets of factors assembly and annotation method with sequencing method. The low identities (Table 3) found among the

mitogenome sequences are related in part to gaps in one of the assemblies, but there are substantial differences in the sequences even in regions without gaps (Fig 3).

The effect of the assembly and annotation method differed depending on the species. The mitogenome sequences of a species from the pooled library were more similar (1 vs 2) than from the separate species libraries (3 vs 4), suggesting that the different assembly and annotation methods generated greater differences in the sequences when the species were in individual libraries than when pooled in the same library. However, mitogenomes from *C. sanguinea* showed the opposite; being more similar from the individual species library.

Similarly, the effect of sequencing method differed depending on the assembly and annotation method and the species. For two of the species, the mitogenome sequences of a species were more similar when assembled by MITObim (2 vs 4) than when done *de novo* (1 vs 3). This indicated that the effect of sequencing method was greater when using *de novo* assembly than MITObim. Again, mitogenomes from *C. sanguinea* showed the opposite; being more similar using *de novo* assembly than MITObim.

Finally, for all species, mitogenome sequences of a species were more similar comparing pooled *de novo* assembly with single-species MITObim assembly (1 vs 4) than methods 2 vs 3.

Thus, the effects of the two sets of factors depended on each other.

The gray and black bars above the annotated and aligned sequences in Fig 3 show the points of dissimilarity among the four methods for each species mitogenomes. While some of the dissimilarity may have been due to low coverage (*e.g.*, *C. sanguinea* method 2 for *nad2* through the first half of COX1), dissimilarity persisted even when coverage was high (*e.g.*, *C. sanguinea* method 2, second half of COX1 to the first half of *cox2* and the second half of *cox3* to the first half of *nad5*). Overall, there was no clear relation between low coverage and dissimilarity among the sequences.

A more detailed examination of the differences generated by the four pipelines is shown in Fig 4 for *C. sanguinea*. For example, at position 2,347 of the consensus sequence of the *nad2* gene, methods 1 and 3 resulted in an adenine (A), but methods 2 and 4 resulted in a guanine (G), indicating that different assembly methods gave different sequences. At position 5,288, methods 1 and 2 had an A and methods 3 and 4 had a G, indicating that different sequencing methods gave different sequences. Finally at positions 2,379 and 4,795, methods 1 and 4 resulted in the same base (thymine (T) or G) while methods 2 and 3 resulted in a different base (A), indicating that both sequencing and assembly methods matter. These differences were not associated with insufficient (considered to be <10-fold redundancy [35]) or differences in coverage (Table 4).

Mitogenome annotations

As expected, the annotations of a species varied across different combinations of libraries and assembly/annotation methods, even when the same annotation method was used

Table 3. Identity (%) among the mitogenomes generated by the different methods.

Mitogenomes	Identity (%)					
	Assembly and annotation method		# species pooling		Interaction of assembly, annotation and species pooling	
	1 vs 2	3 vs 4	1 vs 3	2 vs 4	1 vs 4	2 vs 3
<i>Cycloneda sanguinea</i>	40.17	66.55	52.87	38.19	90.10	51.57
<i>Harmonia axyridis</i>	91.06	72.85	73.93	83.41	87.38	74.06
<i>Hippodamia convergens</i>	93.60	41.15	43.70	96.71	93.51	44.37

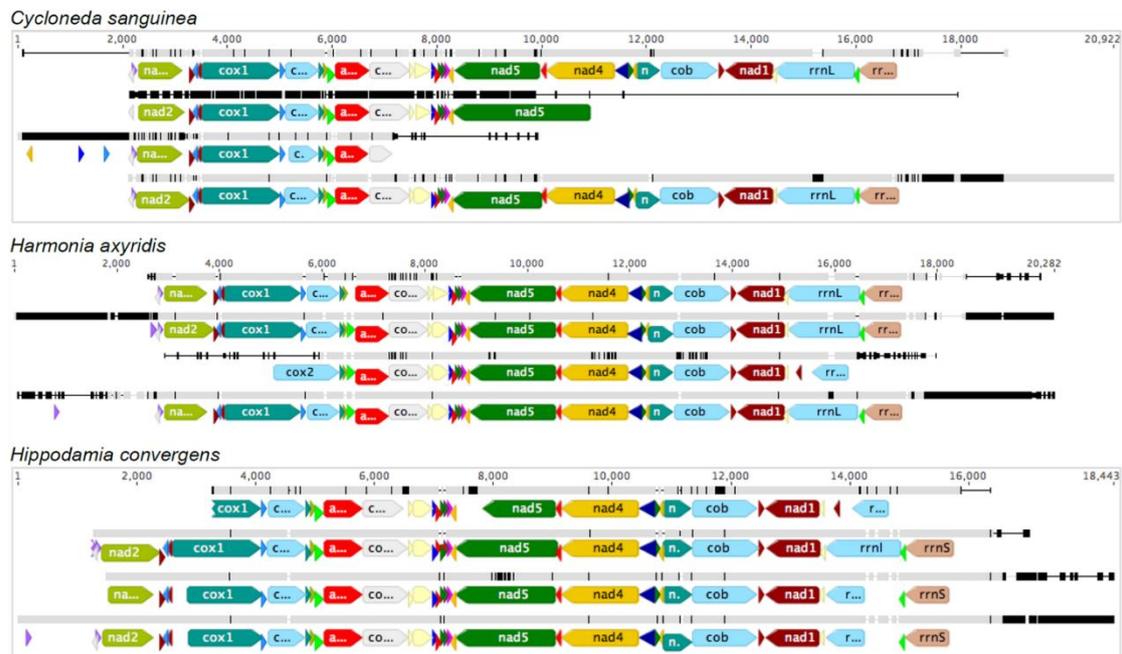


Fig 3. Alignment and comparison of the annotations and similarity of the mitogenomes of *C. sanguinea*, *H. axyridis* and *Hi. convergens* using the four methods. The annotated mitogenomes are in color, with arrows representing the transcriptional sense of the gene, and small arrows representing tRNA genes. Grey bars represent similar nucleotides and vertical black bars represent dissimilar nucleotides in the target sequence compared to the consensus sequence, and horizontal black lines represent gaps in the target sequence. The four methods are displayed in order from the top for each species.

(Figs 2 and 3). Methods 1 and 4 provided the most similar annotations for *C. sanguinea*, going from the gene *trnQ* to *rrnS*, although with a length difference in the *nad2* gene. These were the only methods providing a sequence with all of the expected mito-genes. Methods 2 and 3 provided shorter mitogenomes with fewer annotated genes. For *H. axyridis*, methods 2, and 4 provided sequences with all of the expected mito-genes. These were equivalent to that from method 1, except for the length of the *nad2* gene, which in method 1 was split into two pieces, and in the presence and position of *trnI*. Method 3 provided the most incomplete annotation (missing initial tRNAs and COX1, and *rrnS* at the other end) and had different gene size annotations for *cox2*, *rrnL*, *rrnS*, although it was the only one that presented *trnW*

before *rrnL* gene. The gene *nad2* is also smaller and presented immediately after *rrnL*. For *Hi. convergens*, methods 2 and 4 provided sequences with all of the expected mito-genes, starting from *trnI* to *rrnS*. The genes *nad2*, COX1 and *rrnL* differed in size according to the pipeline used. Method 1 provided a shorter mitogenome sequence and an incomplete annotation, although it was the only one that presented *trnW* before *rrnL* gene. Overall, method 4 (use of MITObim to assemble and MITOS to annotate) seemed to be the one that provided the most complete annotation for all the species.

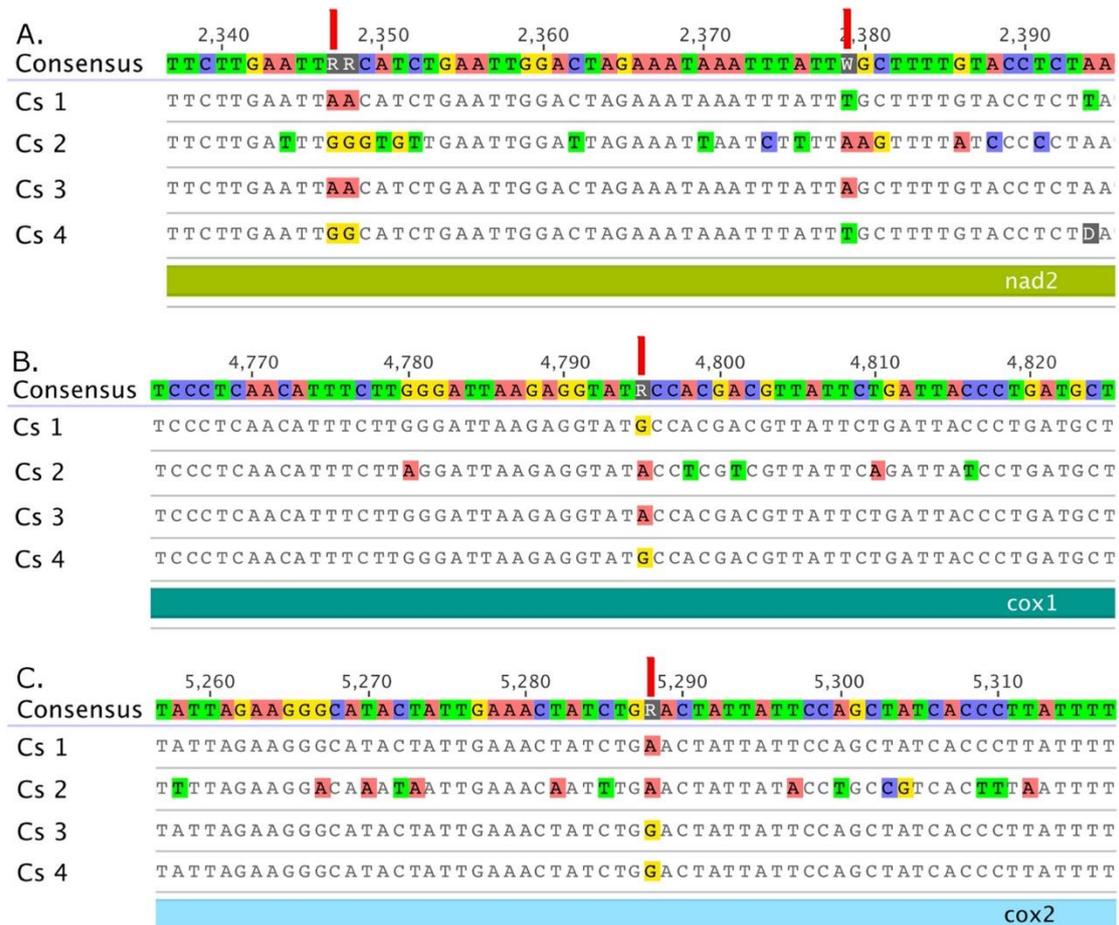


Fig 4. Examples of dissimilarities observed in the mitogenomes of *C. sanguinea* (Cs) generated by different pipelines. The alignments are parts of (A) *nad2*, (B) COX1 and (C) *cox2* genes, with the consensus sequence on the top and the four methods below (Cs1, Cs2, Cs3, and Cs4).

Table 4. Coverage for *C. sanguinea* mitogenomes at the indicated consensus positions for the four methods.

Gene	<i>nad2</i>	<i>nad2</i>	COX1	COX1	<i>cox2</i>
Consensus position (bp)	2,347	2,379	3,796	4,795	5,288
Method 1	45	41	72	122	90
Method 2	1	1	1	1,039	433
Method 3	2,324	2,282	1,924	2,290	2,515
Method 4	495	423	518	600	563

Discussion

Our main result is that for three related species, different sequencing and assembly/annotation methods produced dissimilar mitogenome sequences within each of the three species. In many cases, different library data sets resulted in different nucleotides at the same position when the assembler was the same; the same library data set resulted in different nucleotides when the assembler was different; and different library data sets and assemblers resulted in different nucleotides. In addition, the effects of this methodological variation differed among the species.

This sensitive dependence on sequencing and assembly/annotation method raises concerns. First, if we had relied on only one method, we would not know to question the veracity of the resulting mitogenome sequence. Three of the four methods produced at least one complete mitogenome sequence for the three species, and the remaining method might also produce complete sequences if we had used it on more species. Second, good coverage (>10-fold redundancy) did not eliminate dissimilarity, but might provide illusory confidence in the accuracy of the resulting sequence. Third, the methodological effects differed among the three species, suggesting that finding a “best” method may be challenging, possibly because the mitogenomes of some species may be more difficult to elucidate than others. Although MITObim often provided the mitogenomes with the full mito-genes (Table 2), it did not always do so (*e. g.*, *C. sanguinea* with the pooled library).

We expected that the individual species libraries on the HiSeq platform with smaller insert size would provide higher mitogenome completeness because of the greater sequencing depth and elimination of complications due to the relatedness of the species. However, the mitogenomes of all three species obtained by method 3 (single libraries and *de novo* assembly) were incomplete, despite the high coverage and sequencing depth. Pooling different species in a same library without tagging did not seem to affect significantly the quality of the resulting sequences. Other researchers have successfully pooled species [22,36,37]. According to Tang *et al.* [22] species as close as congeners can be pooled and mitogenomes successfully assembled with more than 50 species in a single lane of HiSeq 2000 (with 100 ng genomic DNA). Although the creation of chimeras is a concern, especially among closely related species, in our case, their effect was minimal as the annotations were equivalent across sequencing pipelines (Figs 2 and 3), with no differences in gene order or transcriptional direction, and only a few differences in gene length. We compared the identity of the mitogenome sequences of *H. axyridis* obtained in this work with the *H. axyridis* KR108208 mitogenome deposited in *GenBank* [38]. The KR108208 sequence is likely to be a high quality sequence, although it is missing *trnI* and *trnQ*, because it was elucidated using LR-PCR and primer walking coupled with Sanger sequencing, and BLAST search in *GenBank* and tRNAscan-SE to assemble and annotate the mitogenome (personal communication, SJ Wei). There was 91.07, 97.81, 74.05 and 89.62 % identity, respectively with our four sequences (Fig 5). The dissimilarities mostly occurred in non-coding regions such as the A+T-rich region and in the terminus. Method 2 was most similar to this comparator and perhaps provided better results for *H. axyridis* mitogenome elucidation and annotation. However, method 4 differed mainly in the A+T-rich region that was missing from the others.

Our results led us to question the reliability of the mitogenome sequences deposited in *GenBank*. In general, these sequences do not have information about the sequencing, assembly or annotation method, so it is not possible to evaluate independently the

completeness or accuracy of the deposited sequences. We examined the 2,469 complete Insecta mitogenomes deposited at *GenBank* and found 60 species with mitogenomes provided by different research groups on species that did not have subspecies or biotypes. Of those, eight species had mitogenome sequence identity <90%. For example, *Bemisia afer*, KR819174 (15,300 bp) and

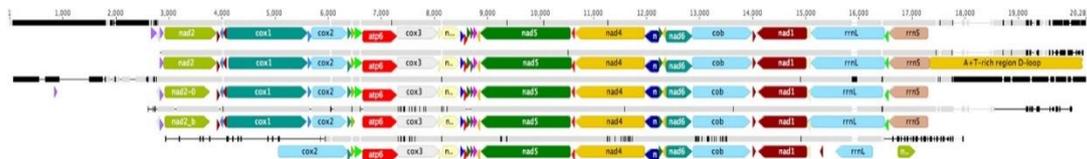


Fig 5. Comparison of the mitogenome sequences of *H. axyridis* obtained in this work with *H. axyridis* (KR108208) [38] using standard parameters in MAFFT v7.017 [34]. The KR108208 has 16,387 bp, 13 PCGs, 20 tRNAs (missing *trnI* and *trnQ*) and 2 rRNAs. The mitogenomes sequenced here are displayed with highest similarity on top. From the top to the bottom are: the mitogenome obtained by method 2; KR108208; followed by methods 4, 1 and 3, respectively. The annotated mitogenomes are in color, with arrows representing the transcriptional sense of the gene, and small arrows representing tRNA genes. Grey bars represent similar nucleotides and vertical black bars represent dissimilar nucleotides in the target sequence compared to the consensus sequence, and horizontal black lines represent gaps in the target sequence.

NC_024056 (14,968 bp), had only with 67.94% similarity. The other seven species are: *Apis florea*; *Choaspes benjaminii*; *Eurema hecabe*; *Gastrimargus marmoratus*; *Luehdorfia chinensis*; *Papilio helenus*; and *Rondotia menciana*. In addition, Lv et al. [39] found mitogenomes of *Nilaparvata lugens* (14,364–14,371 bp) that were much smaller than that reported previously (15,190 bp) [40]. Unfortunately, in many cases, the methods used to elucidate and annotate the mitogenomes were not provided in the *GenBank* deposition or in the associated publication (when available). While it is possible that these dissimilarities might be due to misidentification of the species, several of the eight species are easy to identify, so this is an unlikely explanation for all of the large intraspecific variations in sequence similarity in *GenBank*. Studies on intraspecific variation in parts of the mitogenome typically reported identities >97% [8,39,41]. Hence, similarities below 90% are much lower than expected for intraspecific variation, and are more characteristic of interspecific or intergeneric variation.

We suggest that there is a need to adopt additional reporting requirements and a validation method for mitogenome deposition at public DNA databases. Based on our results, at minimum, information about the sequencing, assembly and annotation methods are relevant, and should be associated with a Sequence Read Archive (SRA) deposition. In addition, coverage should be reported, although this should not be relied on to validate the sequence. As biological inferences in several biological fields have been made and are being made relying on these deposited sequences, a check on accuracy is an issue that urgently needs to be addressed by the scientific community as more and more sequences are expected to be provided (Fig 1) through different bioinformatics pipelines [42].

More broadly, we suggest that guidelines need to be developed by the scientific community to verify the accuracy of mitogenome sequences prior to deposition, perhaps modeled on those for qPCR analysis (The MIQE guidelines) [43]. Until then, one alternative to verify accuracy before database deposition might be based on an approach similar to that used by Tang et al. [22] to assure sequence quality when 49 species were pooled in a library. This alternative could include the use of more than one assembly method (a hybrid

combination of *de novo* and reference genome assemblers might be an interesting choice), concatenate their scaffolds in another assembler, and map the reads to identify regions with lower coverage. Sanger sequencing could be used to verify regions that are assembled only by one of the assemblers with lower coverage or that exhibit dissimilarity or gaps among assemblers.

Conclusions

Mitogenome completeness and sequence accuracy was affected by a number of methodological factors in the experimental design. The scientific community urgently needs to address the issue related to the lack of information regarding the quality of the mitogenomes sequences deposited considering that they have been generated by different methods for sequencing, assembly and annotation, among other factors. The deposition rate of mitogenomes has exploded due to the technical advancements in DNA sequencing methods and bioinformatics pipelines. Guidelines for verifying and validating mitogenome sequences should be discussed and established to assure mitogenome quality for future studies.

Author Contributions

Conceptualization: Marcos M. C. Costa, De' bora P. Paula.

Data curation: Roberto Coiti Togawa, David A. Andow, De' bora P. Paula.

Formal analysis: Roberto Coiti Togawa, David A. Andow, De' bora P. Paula.

Investigation: Renata Velozo Timbo', Roberto Coiti Togawa, Marcos M. C. Costa, David A. Andow, De' bora P. Paula.

Methodology: Renata Velozo Timbo', Roberto Coiti Togawa, De' bora P. Paula.

Project administration: De' bora P. Paula.

Resources: Roberto Coiti Togawa, Marcos M. C. Costa, De' bora P. Paula.

Software: Roberto Coiti Togawa, Marcos M. C. Costa, De' bora P. Paula.

Validation: David A. Andow, De' bora P. Paula.

Writing – original draft: Renata Velozo Timbo', Roberto Coiti Togawa, Marcos M. C. Costa, David A. Andow, De' bora P. Paula.

References

1. Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin PA. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS ONE* 2011; 6(11): e27138. <https://doi.org/10.1371/journal.pone.0027138> PMID: 22073275
2. Dettai A, Gallut C, Brouillet S, Pothier J, Lecointre G, Debruyne R. Conveniently pre-tagged and prepackaged: extended molecular identification and metagenomics using complete metazoan mitochondrial genomes. *PLoS ONE* 2000; 7(12): e51263.
3. Cameron SL. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu Rev Entomol.* 2014; 59: 95–117. <https://doi.org/10.1146/annurev-ento-011613-162007> PMID: 24160435
4. Boore JL. Animal mitochondrial genomes. *Nucl Acids Res.* 1999; 27:1767–1780. PMID: 10101183
5. Wolstenholme DR. Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol.* 1992; 141: 173 – 216. PMID: 1452431
6. Gissi C, Iannelli F, Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 2008; 101: 301–320. <https://doi.org/10.1038/hdy.2008.62> PMID: 18612321
7. Hu M, Jex AR, Campbell BE, Gasser RB. Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing. *Nat Protoc.* 2007; 2: 2339–2344. <https://doi.org/10.1038/nprot.2007.358> PMID: 17947975

8. Jeong HC, Kim JA, Im HH, Jeong HU, Hong MY, Lee JE, Han YS, Kim I. Mitochondrial DNA sequence variation of the swallowtail butterfly, *Papilio xuthus*, and the cabbage butterfly, *Pieris rapae*. *Biochem. Genetics* 2009; 47(3–4): 165–178.
9. Timmermans MJ, Dodsworth S, Culverwell CL, Bocak L, Ahrens D, Littlewood DTJ, et al. Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucl Acids Res.* 2010; 38: e197. <https://doi.org/10.1093/nar/gkq807> PMID: 20876691
10. Huang X. A contig assembly program based on sensitive detection of fragment overlaps. *Genomics* 1992; 14: 18–25. PMID: 1427824
11. Sutton GG, White O, Adams MD, Kerlavage AR. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci Tech.* 1995; 1(1): 9–19.
12. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Res.* 1999; 9: 868–877. PMID: 10508846
13. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, et al. A whole-genome assembly of *Drosophila*. *Science* 2000; 287: 2196–2204. PMID: 10731133
14. Luo R, Liu B, Xie Y. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 2012; 1: 18. <https://doi.org/10.1186/2047-217X-1-18> PMID: 23587118
15. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic nextgeneration sequencing reads—a baiting and iterative mapping approach. *Nucl. Acids Res.* 2013; 41(13): e129. <https://doi.org/10.1093/nar/gkt371> PMID: 23661685
16. Wyman SK, Jensen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 2004; 20: 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352> PMID: 15180927
17. Sheffield NC, Hiatt KD, Valentine MC, Song H, Whiting MF. Mitochondrial genomics in Orthoptera using MOSAS. *Mitochondrial DNA* 2010; 21: 87–104. <https://doi.org/10.3109/19401736.2010.500812> PMID: 20795780
18. Bernt M, Donath A, Ju“hling F, Externbrink F, Florentz C, Fritzsche G, et al. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Mol Phylogen Evol.* 2013; 69(2): 313–319.
19. Gan HM, Schltz MB, Austin CM. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evol Biol.* 2014; 14: 19. <https://doi.org/10.1186/1471-2148-14-19> PMID: 24484414
20. Crampton-Platt A, Timmermans MJTN, Gimmel ML, Sujatha NK, Timothy DC, et al. Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. *Mol Biol Evol.* 2015; 32(9): 2302–2316. <https://doi.org/10.1093/molbev/msv111> PMID: 25957318
21. Lounsberry ZT, Brown SK, Collins PW, Henry RW, Newsome SD, Sacks BN. Next-generation sequencing workflow for assembly of nonmodel mitogenomes exemplified with North Pacific albatrosses (*Phoebastria* spp.). *Mol Ecol Resour.* 2015; 15(4): 893–902. <https://doi.org/10.1111/1755-0998.12365> PMID: 25545584
22. Tang M, Tan M, Meng G, Yang S, Xu S, Shanlin L, et al. Multiple, sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mito-metagenomics. *Nucl Acids Res.* 2014 ; 42(22): e166. <https://doi.org/10.1093/nar/gku917> PMID: 25294837
23. Doyle SR, Griffith IS, Murphy NP, Strugnell JM. Low-coverage MiSeq next generation sequencing reveals the mitochondrial genome of the Eastern Rock Lobster, *Sagmariasus verreauxi*. *Mitochondrial DNA* 2015; 26(6): 844–845. <https://doi.org/10.3109/19401736.2013.855921> PMID: 24409935
24. Eddy S, Durbin R. RNA sequence analysis using co-variance models. *Nucl Acids Res.* 1994; 22: 2079 – 2088. PMID: 8029015
25. Salzberg SL. Genome re-annotation: a wiki solution? *Genome Biol.* 2007; 8(1): 102. <https://doi.org/10.1186/gb-2007-8-1-102> PMID: 17274839
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, et al. Basic local alignment search tool. *J Mol Biol.* 1990; 215: 403–410.
27. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012; 28: 1420–1428. <https://doi.org/10.1093/bioinformatics/bts174> PMID: 22495754
28. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011; 27: 863–8664. <https://doi.org/10.1093/bioinformatics/btr026> PMID: 21278185
29. Kears M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012; 28(12): 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
30. Timmermans MJTN, Vogler AP. Phylogenetically informative rearrangements in mitochondrial genomes of Coleoptera, and monophyly of aquatic elateriform beetles (Dryopoidea). *Mol Phylogen Evol.* 2012; 63: 299–304.
31. Aronesty E. ea-utils: command-line tools for processing biological sequencing data. <https://github.com/ExpressionAnalysis/ea-utils/blob/wiki/FastqMcf.md>. Accessed 28 August 2016.
32. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011; 17(1): 10–12.

33. Lander ES, Waterman MS. Genomic mapping by fingerprint random clones: a mathematical analysis. *Genomics* 1988; 2(3): 231–239. PMID: [3294162](#)
34. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl Acids Res.* 2002; 30: 3059–3066. PMID: [12136088](#)
35. Bouck J, Miller W, Gorrell JH, Muzny D, Gibbs RA. Analysis of the quality and utility of random shotgun sequencing at low redundancies. *Genome Methods.* 1998; 8: 1074–1084.
36. Rubinstein ND, Feldstein T, Shenkar N, Botero-Castro F, Griggio F, Mastrototaro F, et al. (2013) Deep sequencing of mixed total DNA without barcodes allows efficient assembly of highly plastic ascidian mitochondrial genomes. *Genome Biol Evol*; 5: 1185–1199. <https://doi.org/10.1093/gbe/evt081> PMID: [23709623](#)
37. Dettai A, Gallut C, Brouillet S, Pothier J, Lecointre G, Debruyne R (2012) Conveniently pre-tagged and pre-packaged: extended molecular identification and metagenomics using complete metazoan mitochondrial genomes. *PLoS One*; 7: e51263. <https://doi.org/10.1371/journal.pone.0051263> PMID: [23251474](#)
38. Niu FF, Zhu L, Wang S, Wei SJ. The mitochondrial genome of the multicolored Asian lady beetle *Harmonia axyridis* (Pallas) and a phylogenetic analysis of the Polyphaga (Insecta: Coleoptera). *Mitochondrial DNA* 2016; 27(4): 2725–2727. <https://doi.org/10.3109/19401736.2015.1046165> PMID: [26057015](#)
39. Lv L, Peng X, Jing S, Liu B, Zhu L, He G. Intraspecific and interspecific variations in the mitochondrial genomes of *Nilaparvata* (Hemiptera: Delphacidae). *J Econ Entomol.* 2016; 108(4): 2021–2029.
40. Zhang KJ, Zhu WC, Rong X, Zhang YK, Ding XL, Liu J, et al. The complete mitochondrial genomes of two rice planthoppers, *Nilaparvata lugens* and *L. striatellus*: conserved genome rearrangement in Delphacidae and discovery of new characteristics of *atp8* and *tRNA* genes. *BMC Genomics* 2013; 14: 417. <https://doi.org/10.1186/1471-2164-14-417> PMID: [23799924](#)
41. Sarswat M, Dewan S, Fartyal RS. Mitochondrial DNA sequence variation in Drosophilid species (Diptera: Drosophilidae) along altitudinal gradient from Central Himalayan region of India. *J Genetics.* 2016 ; 95: 357–367.
42. Dettai A, Gallut C, Brouillet S, Pothier J, Lecointre G, Debruyne R. Extended molecular identification and metagenomics using complete metazoan mitochondrial genomes. *PLoS ONE* 2012; 7(12): e51263. <https://doi.org/10.1371/journal.pone.0051263> PMID: [23251474](#)
43. Bustin SA, Benes V, Garson JA, Hellems J, Huggett J, Kubista M, et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin Chem.* 2009; 55: 4.

ANEXO 2 – Ensaio piloto para preparação para os bioensaios de alimentação

a) Determinação do tempo de jejum para as larvas do crisopídeo *Chrysoperla externa*

Larvas de 3º ínstar de crisopídeo foram privadas de alimentação por um período de 24 h (10 larvas) e 48 h (10 larvas). As larvas de crisopídeo foram transferidas para o centro de placas de petri 90 x 15 mm contendo uma ninfa de 4º ínstar de mosca-branca *B. tabaci*, um ovo de *S. frugiperda* e um pulgão adulto áptero de *M. persicae* em pontos equidistantes, escolhidos aleatoriamente (Figura 24). Para manter as presas fixas no lugar escolhido, elas foram aderidas à superfície por meio da aplicação de uma fina camada de cola branca (diluída na proporção de 1:10 em água).

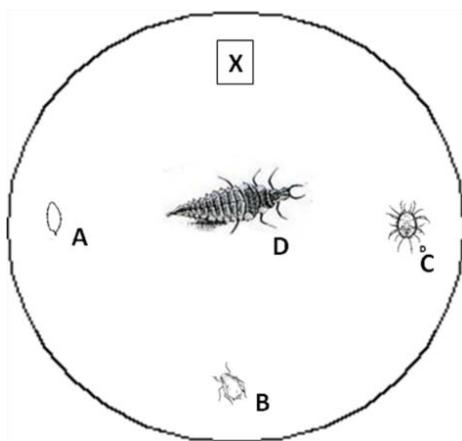


Figura 24. Arena para teste de alimentação do crisopídeo: A) postura de *S. frugiperda*; B) *M. persicae*; C) Ninfa de *B. tabaci*; D) *Ch. externa*; X) Espaço vazio.

Não houve diferença entre o tempo de predação de todas as presas para o jejum de 24 h e no tempo de jejum 48 h para as presas *B. tabaci* e *S. frugiperda* (Tabela 15). O tempo de predação de *M. persicae* foi maior quando a larva do crisopídeo permaneceu em jejum por 48 h. Os crisopídeos levaram mais tempo para se alimentar de um pulgão *M. persicae* ($4,6 \pm 1,26$ min e $8,2 \pm 4,8$ min, para 24 e 48 h de jejum respectivamente), do que de *B. tabaci* ($2,2 \pm 0,42$ min e $2,57 \pm 1,13$ min, para 24 e 48 h de jejum

respectivamente) e do ovo de *S. frugiperda* ($2,33 \pm 0,7$ min e $4 \pm 3,26$ min, para 24 e 48h de jejum respectivamente).

Tabela 15. Tempo pós-predação das presas *B. tabaci*, *M. persicae* e *S. frugiperda* por larvas de 3º ínstar do crisopídeo *Ch. externa* após 24 e 48 h de jejum. Análise estatística por ANOVA e Teste de Tukey.

Comparação entre os grupos				
Grupos	T (min)	Grupos	T (min)	valor-p
<i>M. persicae</i> 48 h	$8,2 \pm 4,80$	<i>B. tabaci</i> 24 h	$2,2 \pm 0,42$	<0,001
<i>M. persicae</i> 48 h	$8,2 \pm 4,80$	<i>S. frugiperda</i> 24 h	$2,3 \pm 0,7$	<0,001
<i>M. persicae</i> 48 h	$8,2 \pm 4,80$	<i>B. tabaci</i> 48 h	$2,6 \pm 1,13$	<0,001
<i>M. persicae</i> 48 h	$8,2 \pm 4,80$	<i>S. frugiperda</i> 48 h	$4,0 \pm 3,26$	0,01
<i>M. persicae</i> 48 h	$8,2 \pm 4,80$	<i>M. persicae</i> 24 h	$4,6 \pm 1,26$	0,047
<i>M. persicae</i> 24 h	$4,6 \pm 1,26$	<i>B. tabaci</i> 24 h	$2,2 \pm 0,42$	0,65
<i>M. persicae</i> 24 h	$4,6 \pm 1,26$	<i>S. frugiperda</i> 24 h	$2,3 \pm 0,7$	0,93
<i>M. persicae</i> 24 h	$4,6 \pm 1,26$	<i>B. tabaci</i> 48 h	$2,6 \pm 1,13$	1
<i>M. persicae</i> 24 h	$4,6 \pm 1,26$	<i>S. frugiperda</i> 48 h	$4,0 \pm 3,26$	1
<i>S. frugiperda</i> 48 h	$4,0 \pm 3,26$	<i>B. tabaci</i> 24 h	$2,2 \pm 0,42$	1
<i>S. frugiperda</i> 48 h	$4,0 \pm 3,26$	<i>S. frugiperda</i> 24 h	$2,3 \pm 0,7$	1
<i>S. frugiperda</i> 48 h	$4,0 \pm 3,26$	<i>B. tabaci</i> 48 h	$2,6 \pm 1,13$	1
<i>B. tabaci</i> 48 h	$2,57 \pm 1,13$	<i>B. tabaci</i> 24 h	$2,2 \pm 0,42$	1
<i>B. tabaci</i> 48 h	$2,57 \pm 1,13$	<i>S. frugiperda</i> 24 h	$2,3 \pm 0,7$	1
<i>S. frugiperda</i> 24 h	$2,33 \pm 0,7$	<i>B. tabaci</i> 24 h	$2,2 \pm 0,42$	1

Pôde-se observar que os crisopídeos que permaneceram em jejum por 48 h apresentaram comportamento de forrageamento reduzido (Gullan e Cranston, 2007). Optou-se então pelo jejum de 24 h para a realização do Bioensaio A.

b) Determinação do tempo de jejum para adultos da joaninha *C. sanguinea* recém-emergidos e larvas de 3º ínstar

Adultos recém-emergidos de *C. sanguinea* permaneceram sem alimentação por um período de 24 h (n= 10) e 48 h (n= 10). Após esse período, os adultos foram colocados em copos plásticos de 50 mL contendo 10 pulgões adultos ápteros de *M. persicae* e o comportamento dos adultos foi acompanhado por 60 min. Foi observado que adultos que permaneceram 24 h sem alimentação permaneceram imóveis durante quase todo o período em que estavam no mesmo ambiente com a presa, e conseqüentemente não se alimentaram. Os adultos que permaneceram 48 h sem alimentação apresentaram grande movimentação e consumiram todos os pulgões (n= 10) em 58 ± 3 min após compartilharem o ambiente com a presa. Optou-se por deixar os adultos recém-emergidos por 48 h sem alimentação para realização do Bioensaio A por apresentarem o comportamento de forrageamento (Gullan e Cranston, 2007). Já as larvas de terceiro ínstar (n=10) que permaneceram em jejum por apenas 14 h não apresentaram o comportamento de forrageamento, diferentemente das que permaneceram 24 h em jejum, sendo assim optou-se pelo jejum de 24 h para as larvas.

c) Análise da viabilidade de larvas de 3º ínstar de joaninha predarem larva de 3º ínstar de crisopídeo

Larvas de *C. sanguinea* (n= 14) foram privadas de alimentação por 24 h. Após esse período, foram transferidas para copos plásticos de 50 mL com tampa de acrílico contendo uma larva de 3º ínstar do crisopídeo alimentada previamente com ovos *A. kuenhiella*.

Apesar da larva de joaninha de *C. sanguinea* ser maior que a larva de crisopídeo e, estar em jejum por 24 h, foi observado que: a) apenas quatro larvas de joaninha (de um total de 10) se alimentaram das larvas de crisopídeo; b) seis larvas de crisopídeo se alimentaram das larvas de joaninha (Figura 25). Os crisopídeos apresentaram maior voracidade quando comparados às joaninhas.

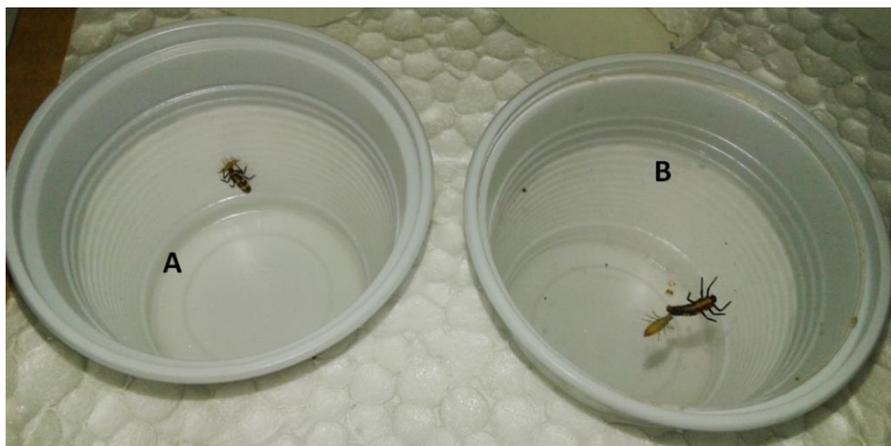


Figura 25. Ensaio piloto de predação secundária. A) Larva de *C. sanguinea* predando larva de *Ch. externa*. B) Larva de *Ch. externa* predando larva de *C. sanguinea*.

O bioensaio teve uma duração de aproximadamente 10 h, provavelmente pelo não “interesse” em predar a larva de crisopídeo pela joaninha e, pelo crisopídeo não estar em jejum e, conseqüentemente, não predou de imediato a larva de joaninha. Dessa forma, optou-se por avaliar larvas de 3º ínstar de duas outras espécies de joaninha a *Hi. convergens* e a *H. axyridis* para observar se estas espécies seriam capazes de predar com maior êxito e menos tempo larvas de 3º ínstar do crisopídeo. Os bioensaios com as joaninhas *Hi. convergens* e a *H. axyridis* foram realizados com o mesmo procedimento descrito anteriormente para a joaninha *C. sanguinea*.

Assim como a *C. sanguinea*, a joaninha *Hi. convergens* apresentou menor voracidade que o crisopídeo, sendo então predada por esse (de n=10, apenas quatro larvas de joaninha predaram a larva de crisopídeo), com tempo de duração de 10 h. Já a *H. axyridis* apresentou voracidade superior às demais espécies de joaninha, tendo predado todas as larvas de crisopídeo (n=10) em menos de 15 min após o início do bioensaio. Dessa forma, optou-se por utilizar a larva de 3º ínstar de de *H. axyridis* para o bioensaio de detecção de presas em predação secundária (Bioensaio C).

d) Análise da viabilidade de larva de crisopídeo de 3º ínstar se alimentar de oito pulgões em 1 h

Larvas de 3º ínstar (n=8) do crisopídeo foram privadas de alimentação por 24 h. Em seguida, as larvas famintas foram individualmente transferidas para tubos de acrílico de 2,5 x 2,5 cm, em que uma das extremidades foi vedada com parafilme (Douglas e van Emden, 2007), contendo oito pulgões adultos ápteros de *M. persicae*, os quais se alimentavam de dieta líquida (Dadd e Mittler, 1966). As larvas foram observadas por 60 min. Seis larvas consumiram todos os pulgões oferecidos em $42,8 \pm 8,58$ min e, os demais se alimentaram de seis e sete pulgões em 1 h. Esses resultados confirmaram que seria possível realizar o Bioensaio C, delineado para larva de crisopídeo consumir três pulgões em menos de 1 h.

ANEXO 3 – Construção do banco de referência de DNA mitocondrial

O processo foi dividido nas seguintes etapas:

- a) elucidação dos mitogenomas das espécies-alvo utilizadas no trabalho e que o mitogenoma ainda não estava elucidado e disponível em banco de dados públicos;
- b) construção do banco de referência de DNA propriamente dito.

a) Elucidação de mitogenomas

Consistiu nas etapas:

- i) verificação da qualidade e presença de adaptadores e *iniciadores* (programa **FastQC**), retirada dos adaptadores e/ou *iniciadores* (programas **Fastqc-mcf** e **Cutadapt**), nova verificação de qualidade e presença de adaptadores e *iniciadores*;
- ii) fazer a montagem dos mitogenomas através do programa **MITObim**, que por sua vez é dividida em duas etapas: **i)** a primeira que utiliza o programa MIRA 4, que irá montar *contigs* baseando-se em um mitogenoma de referência; **ii)** a montagem obtida a partir do MIRA 4, será utilizada interativamente pelo programa **MITObim**, que por sua vez montará *scaffolds* para então formar o mitogenoma de interesse;
- iii) anotação dos genes mitocondriais através do programa **MITOS**.

Os comandos (em Unix) para todo o processo estão descritos a

seguir e indicados pelo símbolo \$. Em vermelho ou entre os símbolos < > foram indicadas informações que devem ser fornecidos pelo usuário (p.ex. nome de arquivos, e outros). O símbolo # indicam explicações nos comandos.

FASTQC

```
$ fastqc arquivo_R1.fastq arquivo_R2.fastq
```

Fastq-mcf

Primeiramente foi criado um arquivo contendo as sequências nucleotídicas de todos os adaptadores conhecidos para as bibliotecas da plataforma Illumina (nome do arquivo: ALL_ADAPTORS.fasta), obtidos do site Illumina.

```
$ fastq-mcf -U -P 33 -w 4 -q 15 -x 5  
../ALL_ADAPTORS.fasta arquivo_R1.fastq  
arquivo_R2.fastq -o arquivo_R1_clipped.fastq -o  
arquivo_R2_clipped.fastq
```

Cutadapt-1.1

```
$ cutadapt-1.1/bin/cutadapt arquivo_R1_clipped.fastq  
-a sequencia do primer >  
arquivo_R1_clipped_cutadapt.fastq
```

```
$ cutadapt-1.1/bin/cutadapt arquivo_R2_clipped.fastq  
-a sequencia do primer >  
arquivo_R2_clipped_cutadapt.fastq
```

MITObim

- i) Criar um subdiretório MITObim para cada biblioteca da espécie a elucidar o mitogenoma;
- ii) Adicionar o mitogenoma de referência, em formato FASTA;
- iii) Criar arquivo txt manifest.conf:

```
project = initial-mapping-Bt-91-to-Bemisia_tabaci_mt  
# Bt-91 representa um nome qualquer do arquivo FASTq  
da espécie-alvo e Bemisia_tabaci_mt o arquivo FASTA  
do genoma de referência em questão
```

```
job=genome,mapping,accurate
```

```
parameters = -NW:mrnl=0 -AS:nop=1 SOLEXA_SETTINGS -  
CO:msr=no
```

```
readgroup
```

```
is_reference
```

```
data = Bemisia_tabaci_mt.fa
```

```
strain = Bemisia_tabaci
```

```
readgroup = reads
```

```
data = arquivo_R1_clipped.fastq  
arquivo_R2_clipped.fastq
```

```
technology = solexa
```

```
strain = Bt-91
```

- iv) Rodar o comando para MIRA 4:

```
$ export  
PATH=/lbi/acgt/bioinfo/ASSEMBLY/MIRA/mira_4.0.2_linu  
x-gnu_x86_64_static/bin/:$PATH
```

```
$ mira manifest.conf
```

```
$ /lbi/acgt/bioinfo/ASSEMBLY/MITOBIM/MITObim-  
master/MITObim_1.8.pl -start 1 -end 15 -sample Bt-91  
-ref Bemisia_tabaci_mt -readpool  
arquivo_R1_clipped.fastq arquivo_R2_clipped.fastq -  
maf initial-mapping-Bt-91- to-  
Bemisia_tabaci_mt_assembly/ initial-mapping-Bt-91-  
to-Bemisia_tabaci_mt_d_results/initial-mapping-Bt-  
91-to-Bemisia_tabaci_mt_out.maf &> log
```

MITOS

Fazer o *upload* da sequência do mitogenoma elucidado pelo MITObim, em formato FASTA, no site <http://mitos.bioinf.uni-leipzig.de/index.py>. Selecionar o código genético para invertebrados. O resultado é enviado por e-mail.

Curadoria manual da anotação dos mitogenomas pelo MITOS

Ao final da análise pelo MITOS, importar os arquivos com extensão “.gff” e FASTA para o programa Geneious, juntamente com o mitogenoma de referência (em formato gb). Selecionar o(s) mitogenoma(s) de referência e o mitogenoma-alvo e fazer o alinhamento (p.ex. no MAFFT, usando parâmetros default). Após o alinhamento, selecionar o mitogenoma de referência para ser e usar a opção “*Set it as reference sequence*”. Para a curadoria manual verificou-se os seguintes aspectos:

- i) Se há codon iniciador no início das sequências codificadoras (normalmente metionina). Os codons iniciadores alternativos para

- invertebrados podem ser AUA, AUU, AUC, GUG. O gene COI não se inicia com ATG (metionina); e alguns genes NDs iniciam com codon ATT (www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi);
- ii) Se não há codons de terminação no meio da sequência codificadora;
 - iii) Se não há deleções ou inserções que mudariam a frame de tradução;
 - iv) Se há codons de terminação no final das sequências codificadoras (em mitogenomas os codons de terminação são TAA, TAG, TA ou apenas T);
 - v) A completude e circularidade do mitogenoma através da verificação de extremidades que se sobrepõem em mais de 25 pb (Tang *et al.*, 2014).
 - vi) Se há sequências repetidas nas extremidades, principalmente em *scaffolds* > 20 kpb, devido a incapacidade de alguns *assemblers* não reconhecerem as extremidades dos genomas.

Deposição dos mitogenomas no *GenBank*

Após a curadoria manual, os mitogenomas elucidados e anotados foram depositados em Batch no *GenBank* através do BankIt. Para isso, três arquivos foram gerados:

- i) FASTA;
- ii) Organism table (tab-delimited text file);
- iii) Five-column, tab-delimited table of feature locations and qualifiers.

b) Geração do banco de referência de DNA mitocondrial de insetos

Consistiu nas seguintes etapas:

- i) Acessar o banco de dados de genomas de organelas do NCBI, através do link:
<http://www.ncbi.nlm.nih.gov/genomes/OrganelleResource.cgi?taxid=50557>;
- ii) Realizar o *download* da lista do número de acesso;
- iii) Acessar o *Batch Entrez* do NCBI através do link <http://www.ncbi.nlm.nih.gov/sites/batchentrez> e informar no campo “escolher arquivo” o arquivo salvo contendo o banco de acessos e em seguida clicar em “*Retrieve*”.
- iv) Clicar em “*Retrieve records for UID(s)*” e, neste momento uma página do NCBI será aberta contendo todas as informações dos mitogenomas listados;
- v) Clicar em “*send to*” e escolher “*file*” format FASTA;
- vi) Inclusão das sequências (Formato FASTA) dos mitogenomas das espécies elucidadas nesse trabalho de doutorado.

ANEXO 4 – Análises de bioninformática para detecção das presas

Os comandos (em Unix) para todo o processo estão descritos a seguir e indicados pelo símbolo \$. Em vermelho ou entre os símbolos < > foram indicadas informações que devem ser fornecidos pelo usuário (p.ex. nome de arquivos, e outros). O símbolo # indicam explicações nos comandos.

BLASTn

```
$ makeblastdb -in <refdb.fasta> -out <refdb> -dbtype nucl

$ blastn -db <banco de dados formado pelor mitogenomas
elucidados mais os obtidos no GenBank> -query
BioA_17_S18_ALL.fa -out BioA_17_S18_ALL_vs_848_insectamito.txt
-task blastn -evaluate 1e-5 -max_target_seqs 3 -outfmt 6 -dust no
-num_threads 55
```

Atribuição taxonômica dos *matches* e filtragem de falsos-positivos

Para se obter todas as identificações taxonômicas (taxid) contidas no *GenBank*, fez-se o *download* de *gi_taxidn_nucl.dmp* do *ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/gi_taxid_nucl.dmp.gz*. No caso dos mitogenomas elucidados e depositados no *GenBank*, fez-se a classificação 'doméstica' do taxid para eles e os incorporouem arquivo (txt) em uma das subpastas contendo os taxid obtidos do *GenBank*.

Foram utilizados os seguintes programas, escritos em Python (Srivathsan *et al.*, 2015):

- `matchtax_2_from_blastout_lowmem.py` (para associação das taxid);
- `parse_by_id.py` (para filtragem pela sobreposição do alinhamento);
- `masterfile_tax.py` (para filtragem pela identidade do alinhamento);

Os comandos utilizados foram:

```
$ Python /usr/local/bin/matchtax_2_fromblastout_lowmwm.py
BioA_17_S18_ALL_vs_848_insectamito.txt # o output será o
arquivo com extensão *.withgi
```

```
$ python /usr/local/bin/parse_by_id.py # seguido da indicação
do arquivo que passará pelo alinhamento e o número de bases
desejado:
```

```
$ infilename: BioA_17_S18_ALL_vs_848_insectamito.txt_withgi
```

```
$ length cutoff: 150# # o output será o arquivo com extensão
*.byid
```

```
$ python /usr/local/bin/masterfile_tax.py
BioA_17_S18_ALL_vs_848_insectamito.txt_withgi_byid 98 # 98 é a
indicação da similaridade desejada, nesse caso, e o o output será o
arquivo com extensão *. cat_con_todiv_final
```

Recuperação das sequências FASTA das *reads* remanescentes

i) Os programas 'cdbfasta' (<https://sourceforge.net/projects/cdbfasta/>) e 'cdbyank' (<http://nebc.nerc.ac.uk/bioinformatics/docs/cdbyank.html>) foram utilizados.

ii) Fazer a indexação de todas as *reads* obtidas em uma dada biblioteca para o formato

FASTA:

```
$ cdbfasta BioA_17_S18_ALL.fasta
```

Será criado o arquivo **BioA_17_S18_ALL.fasta.cidx**

iii) Criar um arquivo texto (retrieve_list.txt) com o nome das *reads* (uma read por linha)

iv) Tendo o arquivo retrieve_list.txt com o nome das *reads*, executar:

```
$ cat retrieve_list.txt |cdbbyank BioA_17_S18_ALL.fasta.cidx >  
retrieve.fasta
```

v) Será criado um arquivo retrieve.fasta

**ANEXO 5 – Códigos SRA dos dados de sequenciamento das bibliotecas
dos bioensaios A, B, C e D depositados no *GenBank***

Tabela16. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio A no *GenBank*.

Biblioteca	Acesso
BioA_1	SRR5342712
BioA_2	SRR5342716
BioA_3	SRR5342715
BioA_4	SRR5342713
BioA_5	SRR5342714
BioA_6	SRR5342674
BioA_6b	SRR5342673
BioA_7	SRR5342672
BioA_8	SRR5342671
BioA_9	SRR5342710
BioA_10	SRR5342709
BioA_11	SRR5342708
BioA_12	SRR5342696
BioA_13	SRR5342695
BioA_14	SRR5342694
BioA_15	SRR5342693
BioA_16	SRR5342711
BioA_17	SRR5342670
BioA_18	SRR5342669
BioA_19	SRR5342668
BioA_20	SRR5342667
BioA_21	SRR5342707
BioA_21b	SRR5342706
BioA_22	SRR5342705
BioA_23	SRR5342704
BioA_24	SRR5342666
BioA_25	SRR5342665
BioA_26	SRR5342664
BioA_27	SRR5342663
BioA_28	SRR5342662
BioA_29	SRR5342661
BioA_30	SRR5342660
BioA_31	SRR5342645
BioA_32	SRR5342644

BioA_33	SRR5342643
BioA_34	SRR5342642
BioA_35	SRR5342641
BioA_36	SRR5342640
BioA_36b	SRR5342639
BioA_37	SRR5342638
BioA_38	SRR5342637
BioA_39	SRR5342636
BioA_40	SRR5342635
BioA_41	SRR5342634
BioA_42	SRR5342633
BioA_43	SRR5342632
BioA_44	SRR5342631
BioA_45	SRR5342630

Tabela 17. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio B no *GenBank*.

Biblioteca	Acesso
46	SRR5342629
47	SRR5342628
47B	SRR5342627
48	SRR5342626
49	SRR5342625
50	SRR5342624
51	SRR5342623
52	SRR5342622
53	SRR5342621
54	SRR5342620
55	SRR5342619
56	SRR5342618
57	SRR5342617
58	SRR5342616
59	SRR5342615
60	SRR5342614

Tabela 18. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio C no *GenBank*.

Biblioteca	Acesso
-------------------	---------------

68	SRR5350656
69	SRR5350655
70	SRR5350654
71	SRR5350653
72	SRR5350652
73	SRR5350651
74	SRR5350650
75	SRR5350649
76	SRR5350648
77	SRR5350644
78	SRR5350643
79	SRR5350641
80	SRR5350640
81	SRR5350639
82	SRR5350638
83	SRR5350637
76_1	SRR5350647
76_2	SRR5350646
76_3	SRR5350645
78B	SRR5350642

Tabela 19. Códigos de acesso aos dados de sequenciamento das bibliotecas do Bioensaio D no *GenBank*.

Biblioteca	Acesso
61	SRR5342703
62	SRR5342702
63	SRR5342701
64	SRR5342700
65	SRR5342699
66	SRR5342698
67	SRR5342697

**ANEXO 6 – Segundo manuscrito submetido para publicação: Análises
para seleção dos parâmetros de filtragem de falsos-positivos**

Molecular Ecology Resources

**MOLECULAR ECOLOGY
RESOURCES**

**Eliminating false prey identifications from predator gut
contents detected by DNA shotgun sequencing**

Journal:	<i>Molecular Ecology Resources</i>
Manuscript ID	Draft
Manuscript Type:	Resource Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Paula, Debora; Embrapa Genetic Resources and Biotechnology, Biological Control Andow, David; University of Minnesota, Department of Entomology Veloze Timbo, Renata; Empresa Brasileira de Pesquisa Agropecuaria; Universidade de Brasilia Coiti Togawa, Roberto; Embrapa Genetic Resources and Biotechnology, Biological Control
Keywords:	false positives, high-throughput-sequencing (HTS), Insecta, prey detection

SCHOLARONE™
Manuscripts

1 Molecular Ecology Resources

2 **Eliminating false prey identifications from predator gut contents detected by DNA**
3 **shotgun sequencing**

4

5 Débora P. Paula¹, David A. Andow², Renata V. Timbó^{1,3}, Roberto Togawa¹

6

7 **Keywords:** false positives, high-throughput-sequencing (HTS), Insecta, prey detection.

8

9 ¹Embrapa Genetic Resources and Biotechnology, Parque Estação Biológica, W5 Norte, P.O.

10 Box 02372, Brasília, DF, 70770-917, Brazil; ²Department of Entomology, University of

11 Minnesota, 219 Hodson Hall, 1980 Folwell Ave., St. Paul, MN 55108, USA; ³University of

12 Brasília, Campus Universitário Darcy Ribeiro, Brasília, Distrito Federal, Brazil

13

14 Author for correspondence: D. P. Paula, Fax: +55(61) 34484673;

15 E-mail: debora.pires@embrapa.br

16 **Running title:** False prey detection elimination from gut data

17 **Abstract**

18 False detection of a taxon in environmental samples is one of the major concerns
19 associated with DNA-based methods because it can mislead derivative ecological
20 inferences. This work demonstrates the feasibility of removing certain kinds of false
21 positives from shotgun sequencing of the DNA community in predator gut contents,
22 without jeopardizing the identification of true positive prey species. Four feeding
23 bioassays using different insect predators and prey were conducted. The gut content of
24 the predators was dissected and used to extract the total DNA for constructing 53
25 TruSeq Nano libraries and sequenced by Illumina HiSeq 2500. Good quality reads were
26 converted from to FASTA and matched to an Insecta mitogenome reference database
27 (969 species) using BLASTn. As a first step to eliminate false prey detections, a set of
28 Python customized scripts was used to filter the output matches using nine
29 combinations of thresholds based on identity and overlap length (98, 99 or 100% with
30 either 100, 150 or 200 bp). False positives were classified as true false positives,
31 mismatches and contaminants. Considerable reduction in false positives was achieved,
32 however downstream processing was essential. Eliminating non-mapping reads, reads
33 mapping to RNAs and control region regions, and single-end reads all true false
34 positives and nearly all (>98%) of mismatches were removed. Contaminant reads were
35 rarely completely removed. The threshold with the best tradeoff was 98% identity in 150
36 bp, combined with downstream processing, but 99% identity and PCR validation might
37 be needed to remove any remaining false positives.

1

38 Introduction

39 Detection and identification of species in the diet of predators or herbivores or in any
40 environmental sample have been greatly enhanced by molecular methods. Considering
41 just DNA-based methods, several highly successful methods rely on specific molecular
42 markers, such as a single-locus barcode in the nuclear DNA (*e.g.*, ITS, 18S, rRNA, *tmL*)
43 or mitochondrial or chloroplast DNA (*e.g.*, COI, Cytb, 16S, *rbcl*, *matK*) (Hebert *et al.*
44 2003; Hollingsworth *et al.* 2009). Examples of these methods include conventional and
45 quantitative PCR and metabarcoding (*e.g.*, Valentini *et al.* 2009; Pompanon *et al.* 2012;
46 De Barba *et al.* 2014) and are now widely used in ecological studies. However, the
47 PCR-based methods can only detect targeted, previously identified species and cannot
48 be used to identify unknown species in a sample. Metabarcoding addresses this
49 problem by amplifying simultaneously the barcode sequence for multiple species,
50 sequencing these amplicons, and identifying species by matching the sequences to a
51 database of known barcode sequences. While this method has been successful for
52 identifying unknown species (Taberlet *et al.* 2012; Yu *et al.* 2012), methods that do not
53 require amplification are desired (Yu *et al.* 2012; Deagle *et al.* 2014).

54 Newer methods, such as mito-metagenomics and DNA shotgun sequencing, do
55 not rely on single barcode sequences for prey identification, but instead they identify
56 species through a marker-free, multi-locus approach and may not require DNA sample
57 amplification. These methods typically identify species by sequencing and detecting
58 multiple mitochondrial or chloroplast coding genes (Taberlet *et al.* 2012; Tang *et al.*

1

59 2014; Srivathsan *et al.* 2015; Paula *et al.* 2015,2016) and matching these sequences to
60 a database of known mitochondrial or chloroplast genome sequences.

61 The underlying principle of many of these methods based on high-throughput
62 sequencing (metabarcoding, mito-metagenomics, DNA shotgun sequencing) is to
63 identify taxa by matching the sequenced fragments (referred as reads) to a reference
64 DNA database. This requires demonstration that the matches correctly identify taxa. For
65 barcode sequences, correct identification is assumed as barcode sequences are
66 expected to be species-specific. Nevertheless, for the marker-free methods, such as
67 DNA shotgun sequencing, this remains to be demonstrated. For this method (Srivathsan
68 *et al.* 2015; Paula *et al.* 2015,2016), the first step for prey identification consists of using
69 a non-stringent BLASTn analysis (E-value <1e-5) or default MEGABLAST settings (word
70 size 28) of the sequencing data against a DNA reference database. The BLAST output
71 data are filtered according to thresholds of percent identity in a specified overlap length
72 and then filtered against additional downstream criteria. However, the best thresholds
73 and downstream analyses for eliminating false positive exogenous taxa, and at the
74 same time, preserving true positive exogenous taxa are not known. For ecological
75 studies in which field samples are used to infer food webs or trophic interactions of
76 target species, it is essential that the exogenous taxa are credibly identified in the
77 samples to enable robust ecological inferences.

78 Previously, the percent identity and overlap length thresholds have been 95% or
79 98% or 99% in 225 bp (Paula *et al.* 2015,2016), i.e. < 10 bp mismatches, or even hit

2

80 length > 50 bp with identity \geq 98% (Srivathsan *et al.* 2015), i.e. 1 bp mismatch. These
81 thresholds have been adopted heuristically based on the assumption that species
82 diverge genetically from each other by at least 3%. However, according to Herbert *et al.*
83 (2003) “there is no simple formula that can predict the length of sequence that must be
84 analyzed to ensure species diagnosis, because rates of molecular evolution vary
85 between different segments of the genome and across taxa”. It is important to validate
86 these thresholds for taxon assignment in environmental samples, especially when using
87 a multi-locus identification method (Rubinoff *et al.* 2006).

88 In this work we conducted feeding bioassays in the laboratory with known insect
89 predator and prey species to evaluate percent identity and overlap length thresholds and
90 downstream processing. The benefits and costs of the thresholds and downstream
91 processing were quantified, considering benefits to be the elimination of false positive
92 reads, and costs to be the elimination of true positive reads. The aim was to eliminate as
93 many false positives as possible while keeping as many true positives as possible. In
94 addition, we demonstrated that downstream processing was essential to supplement the
95 thresholds for removing false positives.

96

97 **Materials and methods**

98 *Feeding bioassays*

99 We conducted four feeding bioassays to test filtering parameters to remove false
100 detection of not preyed upon species in the gut content of predators. In bioassay A
101 (BioA), the insect predators *Cycloneda sanguinea* (3rd instar and adult) and *Chrysoperla*
102 *externa* (3rd instar) were fed individually on one *Spodoptera frugiperda* egg (72 h-old, at
103 pre-eclosion stage), one *Myzus persicae* aptera, one *Bemisia tabaci* nymph (last instar),
104 or not fed on any prey (unfed control group). In bioassay B (BioB), *Hippodamia*
105 *convergens* (3rd instar) was fed on one, three or six *M. persicae* apterae, or not fed on
106 any prey (unfed control group). In bioassay C (BioC), *Harmonia axyridis* (3rd instar) was
107 fed one *Chrysoperla externa* nymph (2nd instar) that had fed or not on three *M. persicae*
108 apterae, or not fed on any prey (unfed control group). In bioassay D (BioD), *C.*
109 *sanguinea* (3rd instar) was fed on one *M. persicae* aptera, and the predators were
110 allowed to digest the prey three hours at 20 or 30°C. All predators were killed
111 immediately after feeding or three hours after feeding in 95% ethanol and stored at -
112 20°C until DNA extraction. In all bioassays 10 individual predators of the same species
113 and stage and with the same feeding history were pooled. Eggs of *Ephestia kuehniella*
114 were used as a maintenance food for all the predator species before the bioassays. The
115 immature predators were starved for 24 h before the feeding bioassays. The adult
116 predators were deprived of food for 48 h after adult emergence. A total of 53 libraries
117 were examined.

118

119 *DNA sequencing*

4

120 The guts of the predators were dissected using clean forceps under a stereomicroscope.
121 Guts from the same treatment were pooled immediately after dissection in the first buffer
122 of the DNA extraction kit using a DNA-free microtube placed in ice. The total DNA was
123 extracted for each sample using DNAeasy Blood and Tissue Kit (Qiagen, Germany),
124 and quantified by Qubit (Invitrogen), following the manufacturer's instructions. The
125 purified DNA was normalized to an equal concentration of 1 µg to make TruSeq Nano
126 libraries (Illumina, USA) for each sample. All the libraries were sequenced on the HiSeq
127 2500 Rapid Mode (2×250 bp, insert size 350 bp, 250 cycle kit) (Illumina, USA) using
128 each 2% of the flow cell. BioA libraries (27 samples) were sequenced at the Genomic
129 Center of the University of Minnesota (USA), and the libraries for the other bioassays
130 (26 samples) were sequenced at the Roy J. Carver Biotechnology Center of the
131 University of Illinois (USA). We found no differences in our results for the two
132 sequencing centers, so our results summarize both together.

133

134 *Taxon assignment*

135 All the programs and settings used in the bioinformatics analyses were based on Paula
136 *et al.* (2015). The quality assessment for each dataset was done using FastQC v.011.3.
137 Library index adapters were trimmed using Fastqc-mcf (v.1.04.807) e Cutadapt (v.1.9.1).
138 After quality control, the retained reads from each library were converted to FASTA
139 format using SeqTK (v1.2) and matched against a DNA reference database using

5

140 BLASTn v2.2.27+ (E-value <1e-5; maximum target sequences 3; no dust). The DNA
141 reference database constituted of the mitogenomes (in FASTA format) of 969 insect
142 species available at the time in GenBank, including the species used in the bioassays
143 (Table S1, Supporting Information). Using a set of Python customized scripts to
144 eliminate false taxon identifications (Paula *et al.* 2015; Srivathsan *et al.* 2015,2016), the
145 matched reads in the output BLASTn were filtered for the threshold of minimum overlap
146 length and identity forming 9 combinations: 100, 150, or 200 bp overlap with 98, 99, or
147 100% identity. In each library, the remaining reads per matched species in each filter
148 combination were counted using an in-house script. We did not consider any overlap
149 length <100 bp as taxonomic information becomes poor (Meusnier *et al.* 2008)

150

151 *Data analysis*

152 True positive species were those that were the prey associated with each of the
153 bioassays, and false positive species were any other species that was neither prey nor
154 predator. We categorized the false positive species detected in our samples in three
155 categories: true false positives, mismatches and contaminants. True false positives
156 comprise the species that could not possibly be in a sample, because they were not
157 present in the bioassays, or even in the laboratory, and in some cases, in the country.
158 They arise because a read matches a mitogenome sequence in the reference database
159 during BLASTn. Mismatches comprise matches to species in the reference database

6

160 that are related and may have genetic similarity to the species used in the experiments.
161 Contaminants comprise a real detection of an external source of DNA not planned by
162 the researcher, and contracted before or during the bioassay, in the sample DNA
163 extraction, or later in the sequencing process. It is difficult to determine if a false positive
164 is a true false positive, a mismatch, or a contaminant. Consequently, we considered
165 distinguishing criteria as follows: 1) a true false positive is a species that is not related to
166 any of the species in the experiment, and is not present in colonies or samples near the
167 laboratory; 2) a mismatch is a species that is closely related to one of the species in the
168 experiment, (at the subfamily level for coccinellids and at the family level for aphids and
169 lacewings), and is not present in colonies or samples near the laboratory; 3) a
170 contaminant is a species that was used in the experiment, but should not have been
171 detected in the library.

172

173 *Downstream processing*

174 To remove remaining false positives after threshold filtering, we mapped the reads with
175 matches at 98 and 99% identity and 150 bp overlap to the mitogenomes of the matched
176 species. To conduct the mapping, FASTA sequences of the matched reads were
177 retrieved through the CAT command, using the programs cdbfasta
178 (<https://sourceforge.net/projects/cdbfasta/sta>) and cdbbyank (<http://nebc.nerc.ac.uk/bioinformatics/docs/cdbbyank.html>). The FASTA sequences of the reads

7

180 of a library and the matched mitogenomes were transferred to the Geneious program
181 (v.9.0.5), where the reads were mapped to a mitogenome reference according to the
182 parameters: custom sensitivity; do not trim; save list of used reads; save list of unused
183 reads; not allow gaps; minimum overlap 150 bp; minimum identity 98% or 99%; not
184 search more thoroughly for poor matching reads. Non-mapping reads, reads mapping to
185 non-coding mitochondrial regions (hyper-variable control region) and mapping to tRNA
186 and rRNA genes were discarded, because these regions do not provide taxonomic
187 resolution. In addition, we eliminated all single-end read matches, because library
188 preparation and sequencing can produce an artifactual excess of R1 reads (Chen *et al.*
189 2017). Finally, reads that matched with more than one taxon in the reference database
190 were eliminated.

191

192 **Results**

193 *False positives*

194 Using a less stringent threshold (overlap length of 100 bp and 98% identity) to filter the
195 BLASTn output read matches, 25 taxa were detected as false positives (Table 1) in the
196 53 libraries from all four bioassays out of 969 possible species in the DNA reference
197 database.

198

199 *a) True false positives*

8

200 The primary benefit of more stringent thresholds is the removal of true false positive
201 read matches (Fig. 1, Table S2, Supporting Information). Ten species were detected as
202 true false positives, presented in decreasing order of abundance of reads with matches
203 in the libraries: *Adoxophyes honmai* [Lepidoptera: Tortricidae], *Diadegma semiclausum*
204 [Hymenoptera: Ichneumonidae], *Atrocalopteryx atrata* [Odonata: Calopterygidae],
205 *Callimorpha dominula* [Lepidoptera: Erebidae], *Anopheles epiroticus* [Diptera:
206 Culicidae], *Trilophidia annulata* [Orthoptera: Acrididae], *Orseolia oryzae* [Diptera:
207 Cecidomyiidae], *Dophla evelina* [Lepidoptera: Nymphalidae], *Evania appendigaster*
208 [Hymenoptera: Evaniidae], and *Heliconius clysonymus* [Lepidoptera: Heliconidae].

209 Some of these true false positive species were detected in almost every library at
210 the less stringent thresholds of overlap length and identity, but were eliminated at more
211 stringent threshold combinations (Fig. 1 and Table S3, Supporting Information).

212 All of the true false positive reads were eliminated at a threshold of 100% identity
213 and 200 bp overlap length, and nearly all were eliminated at 98-99% identity and 200 bp
214 overlap. The thresholds at 99% 150 bp and 100% 150 bp also removed most of true
215 positive reads (Tables S2 and S3, Supporting Information). Overall, all of the 100 bp
216 overlap length thresholds were poor at removing true false positive reads (Fig. 1).

217

218 *b) Mismatches*

219 Four species of coccinellids, three species of chrysopids and one species of aphid were
220 identified as mismatched false positives (Table S4, Supporting Information). All of the
221 *Co. maculata* mismatches were associated with *Hi. convergens* as the predator. The *A.*
222 *bipunctata* mismatches occurred for all three coccinellid predators, *C. sanguinea*, *H.*
223 *axyridis* and *Hi. convergens*. All of the chrysopid mismatches occurred when *Ch.*
224 *externa* was the predator.

225 None of the thresholds removed all mismatched reads (Fig. 2, Tables S4 and S5,
226 Supporting Information). This was especially true of the mismatched coccinellids and
227 *Ch. carnea*, none of which were entirely removed by any of the thresholds. However all
228 of the mismatched aphids were removed at some threshold.

229 The average removal of mismatched reads was not as high as for true false
230 positives (Fig. 2). The most stringent threshold (100% id 200 bp) removed only a few of
231 the mismatched species that were present in the 98% 100 bp threshold (Table S4,
232 Supporting Information), as indicated by the triangular distribution of points in Fig. S1,
233 Supporting Information. Clearly there was a need for downstream processing to try to
234 remove all of the reads from these mismatched species.

235 To determine the “best” threshold for eliminating mismatched reads, we ranked
236 the thresholds for each species (with ties having the average rank), and summed the
237 ranks for each threshold (Table 2). The thresholds with the best ranks were 100% 150
238 bp and 100% 200 bp, but these were not much better than 100% 100 bp and 99% 200

239 bp. In addition, 99% 150 bp and 98% 200 bp were also similar. The remaining three
240 thresholds were considerably poorer.

241

242 *c) Contaminants*

243 Seven species were detected as contaminants in the four bioassays (Table S6,
244 Supporting Information), and some of them were also considered as mismatches (Table
245 S4, Supporting Information). This happened because the bioassays were done during
246 different seasons, and therefore some species were not present in the laboratory at the
247 time of a bioassay. None of the thresholds completely eliminated the reads for any of the
248 contaminant species (Fig. 3, Table S6, Supporting Information). Overall, it appears that
249 the identity thresholds remove more contaminant reads than the overlap length
250 thresholds. The rank order of the thresholds (Table 2) shows this clearly.

251 In general, the thresholds did not remove contaminant reads or species from the
252 libraries very well (Fig. 3, Table S7). Overall, the identity thresholds removed some
253 contaminant species at 100% identity, but the overlap length thresholds did not.

254 Effective removal of contaminant reads generally reduced the number of
255 contaminant species in the library, but not always, as indicated in the triangular
256 distribution of points in Fig. S2 with many points on the x-axis. This result also meant
257 that we should consider downstream methods that remove all of the contaminant

258 species from all of the libraries, rather than focusing on removal of a high proportion of
259 reads.

260

261 **True positives**

262 True positives occurred for all the prey species, although not in all of the libraries; *M.*
263 *persicae* (Bio A: 9 libraries; Bio B,C,D: 17 libraries), *S. frugiperda* (6 libraries), *B. tabaci*
264 (4 libraries), and *Ch. externa* (7 libraries) (Table S8). Similar to the contaminants, none
265 of the species was eliminated from all of the libraries by any of the thresholds, although
266 the more stringent thresholds had fewer prey reads, in some cases half of the least
267 stringent one (Fig. 4, Table S9).

268 Overall, the thresholds did not affect the detection of prey except in a small
269 number of libraries. They eliminated a true prey from at least one library for four of the
270 five prey species at 100% identity (Table S9). Because of the high rate of read
271 elimination and the loss of true positive species at 100% identity, this threshold is not
272 preferred. In addition, the 200 bp threshold lost some true positive species, so this was
273 also not preferred. Based on the results for the thresholds, we focused additional
274 analysis on the 98 and 99% identity and 150 bp overlap length thresholds.

275

276 **Downstream processing to remove remaining false positives**

277 The remaining true false positives were completely eliminated from the data through
278 mapping, for both the 98 and 99% identity and 150 bp overlap length thresholds (Table
279 3). This happened because none of these true false positive reads could be mapped to
280 the mitogenome of the taxon previously matched by BLASTn.

281 Nearly all of the remaining mismatched false positive reads were removed by
282 downstream processing (Table 4). For the 99% identity and 150 bp overlap length
283 threshold, five of seven species were completely removed. For three of these species,
284 none of the mismatched reads could be mapped to their respective mitogenomes. The
285 other two species were removed because most of the mismatched reads were in the
286 rRNA region (*Co. maculata*) or were single-end reads (*Hi. convergens*). The remaining
287 mismatched species were for *Ap. gossypii* (4 reads, 1 library) and *H. axyridis* (2 reads, 1
288 library). For the 98% identity and 150 bp overlap length threshold, five of nine species
289 were completely removed. Three species had no mapped reads, one had most of the
290 reads in the rRNA region (*Co. maculata*) and one had about half the reads in the tRNA
291 or rRNA region (*A. bipunctata*). The remaining mismatched reads were for *Ap. gossypii*
292 (9 reads, 4 libraries), *H. axyridis* (2 reads, 1 library), *Ch. nipponensis* (4 reads, 2
293 libraries) and *Hi. convergens* (4 reads, 2 libraries). No mismatched reads were
294 eliminated because they mapped to the control region or were identified to more than
295 one species.

296 The remaining contaminant false positives were not easily eliminated, but the
297 number of reads of each contaminant species could be reduced in the libraries (Table

298 5). Generally, when there were fewer contaminant reads of a species in a library, it was
299 more likely that they could be eliminated by downstream processing. However, even
300 after downstream processing, there was a considerable number of reads remaining that
301 matched to mitochondrial coding genes of the contaminant species (Fig. 5). No
302 contaminant false positive reads were eliminated because they were identified to more
303 than one species, and very few were eliminated because they mapped to the control
304 region.

305 We also checked the effect of the downstream processing on the number of true
306 positive reads. Because reads that mapped to the control, tRNA and rRNA regions were
307 discarded, the number of reads was reduced considerably (Table 6). However, there
308 was also substantial numbers of reads that did not map to the mitogenome of the
309 identified species or that were single end reads. Overall, 41.8 - 46.7% of the reads after
310 filtering were eliminated by downstream processing for the 98% identity, 150 bp overlap
311 threshold. For the 99% threshold, between 38.3 - 45.7% of the true positive reads were
312 eliminated. There were 30% fewer libraries with true positives after downstream
313 processing for the 99% identity threshold than the 98% identity threshold. Nevertheless,
314 in most libraries the true prey could be identified because of the high number of
315 remaining reads (hundreds).

316

317 **Discussion**

318 *Benefits and costs of more stringent thresholds*

319 The benefit of each threshold combination was related to the type of false positives. All
320 true false positives were removed by the most stringent threshold (100% identity, 200 bp
321 overlap length), and nearly all reads were eliminated for all of the 200 bp overlap
322 thresholds. All of the 150 bp overlap length thresholds also eliminated a large proportion
323 of these species and reads.

324 For mismatches, there was a greater reduction in the number of reads and
325 species by increasing the identity thresholds than by increasing the overlap length
326 thresholds, except for *A. bipunctata* and *Co. maculata*. Because the mismatches are
327 with related species, they may be related to the genetic similarity among the species, so
328 mismatches might be more readily eliminated by a higher identity threshold than by
329 merely increasing overlap length. The fact that nearly all mismatches were associated
330 with related predator species (*e.g.*, coccinellid mismatches when the predator was a
331 coccinellid; chrysopid mismatches when the predator was a chrysopid) implies strongly
332 that these false positives result from taxonomic similarity of the species. The most
333 stringent threshold removed all of the mismatched aphid reads, possibly because there
334 were so few of these reads and the species were not so closely related to the aphid
335 prey, *M. persicae*.

336 Although the number of contaminant reads was reduced at more stringent
337 thresholds, none were completely eliminated, and the number of species contaminating

338 libraries was only slightly reduced. This would be expected if the reads were really due
339 to contamination of the samples, as the contaminant DNA would be present and should
340 be detected. Considering this, we should expect that the thresholds would affect these
341 contaminants similar to the way they would affect both true prey and predator
342 detections. More stringent identity thresholds removed more reads than more stringent
343 overlap length thresholds, except for *H. axyridis* and *Hi. convergens*. At 98% identity and
344 150 bp overlap length, 49.7% of contaminant reads were eliminated, and at 99% identity
345 and 150 bp overlap, 55.2% were eliminated.

346 The primary cost of more stringent thresholds is the undesired removal of true
347 positive reads. True positives are associated with the prey used in the bioassays. There
348 were four species used as prey, *M. persicae*, *S. frugiperda*, *B. tabaci*, and *Ch. externa*
349 (as intraguild prey). The predator reads are also true positives, and it is possible that
350 loss of predator reads could be a cost, but this will not be considered here, as there
351 were large numbers of reads for the predators, and their DNA was not degraded by
352 digestion. The cost can be characterized by the loss of prey species or the loss of prey
353 reads. Because the read number is crucial for estimating the decay rates of prey
354 detection in the predator gut, we prioritized avoiding the loss of reads. Generally, the
355 identity thresholds removed more reads than the overlap length thresholds (except for *B.*
356 *tabaci*), similar to the contaminants. Unlike the contaminants, at 98% identity and 150 bp
357 overlap length, only 8.0% of true positive reads were eliminated, and at 99% identity and
358 150 bp overlap, only 23.9% were eliminated.

359 Balancing the benefits and costs of using different thresholds, it seems
360 reasonable to adopt the thresholds of 98% or 99% identity and 150 bp overlap length.
361 These removed most, but not all, true false positive reads and many mismatched reads,
362 but did not remove too many true positive reads. The 100% identity thresholds removed
363 more false positives, but using 100% identity would not allow detection of the expected
364 amount of intraspecific genetic variation that commonly occurs. The 100 bp overlap
365 length thresholds generally did not remove very many false positive reads or species
366 and the 200 bp thresholds removed too many true positive species detections. Clearly,
367 there are so many remaining false positive reads with the 98% and 99% identity and 150
368 bp overlap length thresholds that downstream processing is needed to improve data
369 quality.

370

371 *Benefits and costs of downstream processing*

372 Downstream processing of the results from the two best thresholds was done by
373 mapping the remaining reads and eliminating unmapped reads and reads mapping to
374 conserved regions of the mitogenome, and eliminating unpaired reads and those that
375 matched to more than one taxon in BLASTn results. Overall, the greatest number of
376 reads was removed because they did not map to the mitogenome. Indeed, all remaining
377 true false positive reads were completely removed from both thresholds for this reason.

17

378 Mismatches also were efficiently removed, especially using the 99% identity
379 threshold. Only 19 mismatched reads (98.4% removal) associated with nine libraries
380 (four different species), and 6 reads (99.3% removal) associated with two libraries (two
381 different species) remained from the 98% and 99% identity thresholds, respectively. The
382 highest number of mismatched reads in a library was only four, possibly suggesting a
383 minimum threshold of four reads to avoid wrongly identifying a mismatch species. Three
384 different patterns of mismatches occurred: a) mismatches did not map to their respective
385 mitogenome, but instead to the closely related predator species used in the feeding
386 bioassay; b) mismatches mapped to their respective genome, but not to the closely
387 related predator or prey mitogenomes used in the bioassay; and c) mismatches that
388 mapped to both the mismatched species and the closely related predator species used
389 in the bioassay, exactly in the same mitogenome region.

390 The first case was illustrated by the mismatches with *Ch. carnea*, *Ch.*
391 *nipponensis*, and *H. axyridis* for the 99% identity and 150 bp overlap length threshold.
392 Similar results were observed for the 98% identity threshold. For *Ch. carnea*, we
393 observed that none of the 701 mismatched reads could be mapped to the mitogenome
394 of *Ch. carnea*. However 667 of these reads mapped to the mitogenome of *Ch. externa*.
395 All of the 701 reads were from libraries with *Ch. externa* as an intraguild prey. Similarly,
396 the five reads that mismatched to *Ch. nipponensis* did not map to this mitogenome, but
397 instead, three of these reads mapped to the *Ch. externa* mitogenome, and these came
398 from libraries with *Ch. externa* either as the primary predator or intraguild prey. These

399 reads (667+5) probably should have been matched to the mitogenome of *Ch. externa* in
400 the first place, but in the BLASTn analysis they matched to other *Chrysoperla* species.

401 The second case was illustrated by the mismatches with *Co. maculata* and *A.*
402 *bipunctata*. The reads mapped to the mismatched species *Co. maculata* and *A.*
403 *bipunctata*, but did not map to the mitogenome of the closely related predator used in
404 the assays, *Hi. convergens*, except for two mismatched *A. bipunctata* reads that
405 mapped to the rRNA of *Hi. convergens*. These reads were removed for other reasons,
406 as they mapped in a conservative region of the mitogenome. This case was also
407 illustrated by the mismatches to *Ap. gossypii*, in which six of nine reads mapped to its
408 mitogenome, as previously identified by BLASTn, but not to the mitogenome of the prey
409 *M. persicae* used in the bioassays. Mismatches to *Hi. convergens* were another
410 example, in which five out six reads indeed mapped to this mitogenome, but none to the
411 closely related predator used in the bioassay, *C. sanguinea*.

412 The final case is illustrated by the mismatches with *H. axyridis* from a bioassay
413 using the closely related predator *C. sanguinea*. Two mismatched reads from one
414 library mapped to both the *H. axyridis* and *C. sanguinea* mitogenomes. The reads
415 mapped to the same region of the *atp6* gene for both species. This mismatch pattern
416 might suggest that closely related species will be difficult to discriminate when few reads
417 are available. Experiments specifically designed to evaluate this possibility are needed.

418 Downstream processing was essential to remove mismatches from the
419 sequencing data after filtering by identity and overlap length. It is essential to curate all
420 reads by mapping them to the mitogenome of the indicated species and removing all
421 reads that do not map and to check to see if the reads map to the predator mitogenome.
422 In addition, it is important to remove reads mapping to conserved regions of the
423 mitogenome. In our data, no mismatched reads mapped to the control region, but many
424 mapped to *tRNA* and *rRNA* genes. Finally, as library preparation and sequencing can
425 induce mutations (Chen *et al.* 2017), it is important to exclude single-end reads. More
426 broadly, mismatches might be avoided by having an accurate DNA reference database.
427 The mismatches to *A. bipunctata* and *Co. maculata* may be related to uncertainty
428 associated with their mitogenomes deposited in GenBank, as they are both classified as
429 unverified.

430 False positives due to contamination were more common than expected and they
431 were not efficiently removed from the samples even after additional downstream
432 processing. In few cases, it was possible to remove all of the reads of a contaminant
433 species when their number was small (*e.g.*, four for *Bemisia tabaci*) or when using an
434 identity threshold of 99% (*e.g.*, for *Chrysoperla externa*). From the 53 libraries analyzed
435 in this study, after all steps of downstream processing, 37 (69.8%) at 98% identity and
436 25 (47.2%) at 99% identity still had contaminants. Seven libraries had three different
437 contaminants, 12 had two and 18 had one (98% identity threshold).

438 *Ephestia kuehniella* was the most common contaminant, although it was not used
439 in any of the feeding bioassays. Its eggs were used as a rearing food for all the
440 predators before the bioassays. Its reads were only detected in assays with 3rd instar
441 larvae of *Hi. convergens* or *H. axyridis*, and not for 3rd instar *C. sanguinea* or *Ch.*
442 *externa* or adult *C. sanguinea*, suggesting stage or species differences. All immature
443 predators were deprived of *E. kuehniella* eggs 24 h before the feeding bioassays and
444 adults were not fed. This deprivation period may not have been sufficient to empty the
445 gut content of the immature predators of two of the species. Adults empty their gut
446 contents during metamorphosis (Chapman 2013), so unfed adults should not contain
447 any *E. kuehniella* reads. *Cycloneda sanguinea* was the second most common
448 contaminant, which might be related to the fact that it was the most used predator in the
449 bioassays, used in the bioassays for 37 libraries out of 53. *Hippodamia convergens* was
450 the third most common contaminant, but had the highest number of contaminant reads.
451 We consider it likely that *Hi. convergens* immatures were mistakenly used in some of the
452 bioassays, instead of the intended predator. Excluding *Hi. convergens*, there were 529
453 reads (average of 9.6 reads/ instance of contamination) and 317 reads (average of 9.3
454 reads/ instance of contamination) remaining from the 98% and 99% identity thresholds,
455 respectively.

456 Consequently, the false positive reads remaining after downstream processing
457 were probably from true contaminants. Contamination could have occurred from many
458 possible sources, for example, taxon misidentification, insufficient starvation time of the

459 predator before the bioassay (the detection of *E. kuehniella* might illustrate this
460 possibility), cross-contamination in the starting materials (e.g., during the bioassay or
461 DNA sample preparation), sample transportation (e.g., loose plate cover), library
462 construction (on the set of barcode adapters or mistakenly identified barcode adapters),
463 and sample carryover between sequencing runs. While downstream processing
464 removed substantial numbers of contaminant reads, the best way to avoid them is
465 preventing them, e.g. avoiding the preparation of multiple samples or multiple libraries at
466 the same time, and isolating the bioassays from each other, as we did previously (Paula
467 *et al.* 2015).

468 When conducting laboratory bioassays using known predators and prey, it is
469 possible to check for cross-contamination in the starting materials by performing PCR on
470 aliquots of the samples used for library construction. If there is no amplicon for the
471 contaminant taxa in the sample, the researchers can discard those contaminant reads.

472 Downstream processing also removed substantial numbers of true positive reads.
473 However, it eliminated all true positive reads from only 2 of 34 libraries for the 98%
474 identity threshold, but 10 of 34 libraries for the 99% identity threshold. Thus, if the
475 concern is retaining as many libraries with true positives as possible, the 98% identity
476 and 150 bp overlap length threshold would be preferred. However, even with
477 downstream processing, there were 19 mismatched reads in 9 libraries and 529
478 contaminant reads in 31 of the 53 libraries at this threshold (excluding *Hi. convergens*).
479 If it were more important to remove false positives, the 99% identity and 150 bp overlap

22

480 length threshold would be preferred. There were only 6 mismatched reads in 2 libraries
481 and 317 contaminant reads in 21 of 53 libraries at this threshold.

482 For laboratory experiments with known prey, the 98% identity threshold might be
483 preferred because all mismatches and contaminants can be inferred from the
484 experimental design. For field samples, this is not possible, and the 99% identity
485 threshold might be preferred. However, even for field samples, we suggest as a first
486 step, the 98% identity threshold be used and compared to the results from the 99%
487 identity threshold. Some mismatched species can be identified from field samples, if the
488 species does not occur in sampled region. For species that are present after processing
489 the 98% identity threshold data, but eliminated with the 99% identity threshold, PCR on
490 aliquots of the samples used for library construction could be used to verify the
491 presence/absence of the species. PCR validation of true positive prey identifications in
492 the predator gut content when it they are supported just by few reads should be
493 preferred over the use of threshold number of reads, as is often done with
494 metabarcoding. For example, Yu *et al.* (2012) eliminated singletons from their 454 data,
495 and Clare *et al.* (2014) eliminated <4 reads from their Ion Torrent data. As shown in this
496 work, a true positive can have a smaller or equal number of reads (Table 6) as a true
497 contaminant (Table 5). Small number of reads detected for true positive species might
498 be related to feeding on a small amount of biomass or a long elapsed time since
499 predation. Without complementary evidence, such as field sampling data or PCR
500 validation of the predator gut content, it would be advisable to be conservative and risk

501 rejecting a true positive than accepting a false positive, because the dietary data would
502 be incomplete, but not wrong. Considering the high biodiversity existing in many
503 ecosystems and the unlikelihood of having a comprehensive mitogenome database for
504 all the species in the studied ecosystem, it may be preferable to underestimate the diet
505 composition of predators or herbivores than possibly to erroneously consider it to be
506 wider.

507

508 **Conclusions**

509 There is no threshold that maximizes benefits and minimizes costs simultaneously.
510 Downstream processing is essential to remove the remaining true false positives and
511 nearly all mismatches. Contaminants cannot be removed using thresholds or
512 downstream processing, except when there are only a few reads. They need to be
513 prevented. Assuming contamination can be limited, filtering at 98% identity and 150 bp
514 overlap length followed by downstream analysis has a good cost-benefit tradeoff as it
515 efficiently eliminates true false positives and mismatches, and does not eliminate true
516 positive taxa. For field samples, it may be useful to compare the results from the 98 and
517 99% identity and 150 bp overlap length thresholds and downstream processing. PCR
518 validation of prey presence/absence could be used on questionable identifications, such
519 as those not supported at both 98 and 99% identity or those supported by only few
520 reads.

521

522 **References**

- 523 Chapman RF (2013) *The Insects: Structure and Function*, 5th edition. Cambridge
524 University Press, Cambridge, UK
- 525 Chen L, Liu P, Evans TC, Ettwiller LM. (2017) DNA damage is a pervasive cause of
526 sequencing errors, directly confounding variant identification. *Science* 355, 752-756.
- 527 Clare EL, Symondson WOC, Fenton MB. (2014) An inordinate fondness for beetles?
528 Variation in season dietary preferences of night-roosting big brown bats (*Eptesicus*
529 *fuscus*). *Mol Ecol.* 23, 3633-3647.
- 530 Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. (2014) DNA
531 metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match.
532 *Biol Lett.* 10(9): 20140562.
- 533 De Barba M, Miquel C, Boyer F et al. (2014) DNA metabarcoding multiplexing and
534 validation of data accuracy for diet assessment: application to omnivorous diet. *Mol*
535 *Ecol Resour.* 14, 306-323.
- 536 Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes.
537 *Proc. R. Soc. Lond. B Biol. Sci.* 270, 313-321.
- 538 Hollingsworth PM, Forrest LL, Spouge JL et al. (2009) A DNA barcode for land plants.
539 *PNAS* 106, 12794-12797.
- 540 Meusnier S, Singer GAC, Landry J-F, Hickey DA, Hebert PDN, Hajibabaei M. (2008) A
541 universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* 9, 214.
- 542 Paula DP, Linard B, Andow DA, Sujii ER, Pires CS, Vogler AP (2015) Detection and
543 decay rates of prey and prey symbionts in the gut of a predator through
544 metagenomics. *Mol Ecol Resour.* 15, 880-892.
- 545 Paula DP, Linard B, Crampton-Platt A, Srivathsan A, Timmermans MJTN, Sujii ER, et
546 al. (2016) Uncovering trophic interactions in arthropod predators through DNA
547 shotgun-sequencing of gut contents. *PLoS ONE* 11(9): e0161841.
- 548 Pompanon F, Deagle BE, Symondson WOC et al. (2012) Who is eating what: diet
549 assessment using next generation sequencing. *Mol. Ecol.* 21, 1931-1950.
- 550 Rubinoff D, Cameron S, Will K (2006) A genome perspective on the shortcomings of
551 mitochondrial DNA for "barcoding" identification. *Journal of Heredity* 97, 581-594.

25

- 552 Srivathsan A, Sha JCM, Vogler AP, Meier R (2015) Comparing the effectiveness of
553 metagenomics and metabarcoding for diet analysis of a leaf feeding monkey
554 (*Pygathrix nemaeus*). *Mol Ecol Resour.* 15: 250-261.
- 555 Srivathsan A, Ang A, Vogler AP, Meier R (2016) Fecal metagenomics for the
556 simultaneous assessment of diet, parasites, and population genetics of an
557 understudied primate. *Front Zool.* 13:17.
- 558 Valentini A, Miquel C, Nawaz MA et al. (2009) New perspectives in diet analysis based
559 on DNA barcoding and parallel pyrosequencing: the *trnL* approach. *Mol Ecol Resour.*
560 9, 51-60.
- 561 Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-
562 generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045-
563 2050.
- 564 Tang M, Tan M, Meng G, et al. (2014) Multiplex sequencing of pooled mitochondrial
565 genomes-a crucial step toward biodiversity analysis using mito-metagenomics.
566 *Nucleic Acids Res.* 42 (22): e166.
- 567 Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z. (2012) Biodiversity soup:
568 metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring.
569 *Methods Ecol. Evol.* 3, 613-623.

570

571 **Data Accessibility**

572 NCBI SRA libraries: BioA SRR5342632, SRR5342633, SRR5342636-SRR5342638,
573 SRR5342640, SRR5342643-SRR5342645, SRR5342662, SRR5342663, SRR5342666,
574 SRR5342669-SRR5342672, SRR5342674, SRR5342695, SRR5342696, SRR5342704,
575 SRR5342705, SRR5342707, SRR5342710-SRR5342712, SRR5342715, SRR5342716;
576 BioB SRR5342616, SRR5342617, SRR5342620, SRR5342621, SRR5342624-
577 SRR5342626, SRR5342628, SRR5342629; BioC SRR5350638-SRR5350641,
578 SRR5350643, SRR5350644, SRR5350646, SRR5350647, SRR5350649, SRR5350650,
579 SRR5350652, SRR5350653, SRR5350655, SRR5350656; and BioD SRR5342699,
580 SRR5342702, SRR5342703.

581

582 **Author contributions**

583 Design of study: DPP, DAA
584 Prepared the DNA samples: RVT
585 Bioinformatic analyses: RVT, RCT, DPP
586 Statistical descriptive analysis: DAA, DPP
587 Wrote the manuscript: DPP, DAA, RCT, RVT

26

588 All authors read and approved the final manuscript.

589

590 **Supporting Information**

591 Additional Supporting Information may be found in the online version of this article:

592 **Table S1** Insecta mitogenomes used in the DNA reference database

593 **Table S2** Number of reads of true false positives per species at each threshold

594 **Table S3** Number of libraries (out of 53) with true false positives per species at each
595 threshold

596 **Table S4** Number of reads of mismatches per species at each threshold

597 **Table S5** Number of libraries (out of 53) with mismatched species at each threshold

598 **Table S6** Number of reads of contaminants at each threshold per species

599 **Table S7** Number of libraries (out of 53) with contaminant species at each threshold

600 **Table S8** Number of reads of true prey positives at each threshold per prey species

601 **Table S9** Number of libraries with true prey positives detected of a certain species at
602 each threshold

603 **Fig. S1** Relation between removal of mismatched reads and libraries with mismatches.

604 **Fig. S2** Relation between removal of false positive reads and libraries with
605 contaminants.

606 **Tables and Figures**

607

608 **Table 1** Taxa with false positive detection and the number of libraries (out of 53) that
 609 had false positives of the indicated species initially after filtering BLASTn output read
 610 matches using 100 bp and 98% identity and after filtering using 150 bp overlap length
 611 and either 98% or 99% identity with subsequent downstream processing. Number
 612 before "/" is for mismatches, and after is for contaminants

Species with false detection	Number of libraries		
	Initial	98%	99%
<i>Adoxophyes honmai</i>	45	0	0
<i>Diadegma semiclausum</i>	36	0	0
<i>Atroclopteryx atrata</i>	26	0	0
<i>Callimorpha dominula</i>	26	0	0
<i>Anopheles epiroticus</i>	24	0	0
<i>Trilophidia annulata</i>	15	0	0
<i>Dophia evelina</i>	7	0	0
<i>Evania appendigaster</i>	3	0	0
<i>Orseolia oryzae</i>	3	0	0
<i>Heliconius clysonymus</i>	2	0	0
<i>Adalia bipunctata</i>	18	0	0
<i>Chrysoperla nipponensis</i>	8	2	0
<i>Coleomegilla maculata</i>	11	0	0
<i>Chrysoperla cameo</i>	8	0	0
<i>Diuraphis noxia</i>	5	0	0
<i>Aphis gossypii</i>	4	4	1
<i>Chrysopa pallens</i>	1	0	0
<i>Acyrtosiphon pisum</i>	1	0	0
<i>Cycloneda sanguinea</i>	27	13	9
<i>Ephestia kuehniella</i>	26	17	12
<i>Hippodamia convergens</i>	23	2 / 10	0 / 5
<i>Harmonia axyridis</i>	13	1 / 5	1 / 4
<i>Myzus persicae</i>	7	5	4
<i>Chrysoperla externa</i>	7	5	0
<i>Bemisia tabaci</i>	2	0	0

613

616 **Table 2** Average ranks of mismatched and contaminant reads. Lower rank indicates
 617 removal of more reads

Threshold (% id_overlap bp)	Mismatches	Contaminants
98_100	8.3	8.8
99_100	6.7	6.5
100_100	4.3	3.3
98_150	6.9	8.0
99_150	5.2	5.8
100_150	3.1	2.8
98_200	4.8	5.3
99_200	3.4	3.6
100_200	2.3	1.1

618

619

620 **Table 3** Number of reads of true false positives cleaned per species in the downstream
 621 analysis of the output BLASTn data filtered by 98 or 99% identity and 150 bp overlap

Criteria used to clean false positives	<i>Adoxophyes honmai</i>		<i>Atrocalopteryx atrata</i>		<i>Orseolia oryzae</i>	
	98% id	99% id	98% id	99% id	98% id	99% id
Not mapping	437	166	18	1	4	4
Mapping in the control region						
Mapping in a tRNA region						
Mapping in a rRNA region						
Single-end read match						
Match with more than one species						
Previous number of matched reads	437	166	18	1	4	4
Remaining reads	0	0	0	0	0	0

622

623

624

Table 4 Number of reads of mismatches cleaned per species in the downstream analysis of the output BLASTn data filtered by 98 or 99% identity and 150 bp overlap

Criteria used to clean false positives	<i>Chrysoperla carnea</i>		<i>Coleomegilla maculata</i>		<i>Adalia bipunctata</i>		<i>Chrysoperla nipponensis</i>		<i>Aphis gossypii</i>		<i>Hippodamia convergens</i>		<i>Diuraphis noxia</i>		<i>Hamonia axyridis</i>	
	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id
Not mapping	827	701	15	10	25	24	30	5	3	3	2	1				
Mapping in the control region																
Mapping in a tRNA region					19				3	2						
Mapping in a rRNA region			201	160	6		1									
Single-end read match							9				5	3				
Match with more than one species																
Previous number of matched reads	827	701	225	170	56	24	44	5	17	9	22	5	1	0	2	2
Remaining reads	0	0	0	0	0	0	4	0	9	4	4	0	0	0	2	2
Libraries with remaining mismatches after downstream analysis							2 at total (A31, A34)		4 at total (C69, C74, D61, D62)	D62	2 at total (A15, A42)				A24	A24

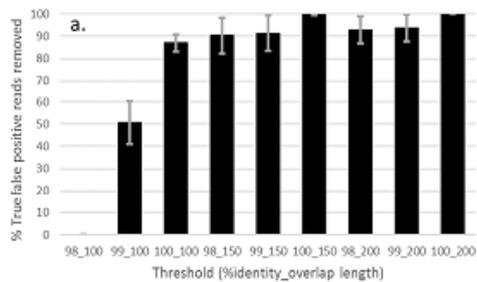
625 **Table 5** Number of reads of contaminants eliminated at each threshold per species in the downstream analysis of the
626 output BLASTn data filtered by 98 or 99% identity and 150 bp overlap

Criteria used to clean false positives	<i>Harmonia axyridis</i>		<i>Cycloneda sanguinea</i>		<i>Hippodamia convergens</i>		<i>Ephesttia kuehniella</i>		<i>Chrysoperla externa</i>		<i>Myzus persicae</i>		<i>Bemisia tabaci</i>	
	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id	98% id	99% id
Not mapping	478	440	58	57	10506	7715	14	11	3	7	2	5	4	4
Mapping in the control region	13			54		9			2					
Mapping in a tRNA region	9	8	41	16	619		18	19	9		1	7		
Mapping in a rRNA region	7	6	121	42	183		13	13	2		5	4		
Single-end read match	17	17	127	11	241	6	9	9	19	35	6	6		
Match with more than one species														
Previous number of matched reads	536	476	747	427	15636	7741	123	98	56	42	41	41	4	4
Remaining reads	12	5	400	247	4072	11	69	46	21	0	27	19	0	0
Libraries with remaining contaminants after downstream analysis	5 at total (B48, B50, B53, B57, B58, D65)	4 at total (B50, B57, B58, D65)	13 at total (A32, A33, A42, A43, A47, B49, B50, B53, B54, B58, C69, C75, C77)	9 at total (A32, A33, A42, A43, A49, B50, B53, B54, C77)	10 at total (A23, C71, C72, C74, C78, C80, C81, D61, D62)	5 at total (C72, C78, C81, C76.1, D62)	17 at total (A13, A24, A36, A37, A42, B48, B49, B57, B58, C68, C69, C71, C72, C75, C77, C79, C81)	12 at total (A13, A24, A36, A42, B48, B49, B57, B58, C68, C69, C71, C72)	5 at total (A03, A06, A24, A27, A28)		5 at total (A02, A13, A16, C68, C69)	4 at total (A13, A16, C68, C69)		

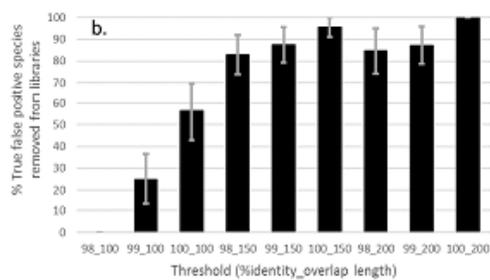
627 **Table 6** Number of reads of true positives reads eliminated at each threshold per
 628 species in the downstream analysis of the output BLASTn data filtered by 98 or 99%
 629 identity and 150 bp overlap

Criteria used to eliminate positives	<i>Myzus persicae</i>		<i>Chrysoperla externa</i>		<i>Spodoptera frugiperda</i>	
	98% id	99% id	98% id	99% id	98% id	99% id
Not mapping	3483	3384	602	588	97	95
Mapping in the control region	664	408	313	16	26	23
Mapping in a tRNA region	1042	899	805	480	55	49
Mapping in a rRNA region	978	726	574	111	30	28
Single-end read match	447	352	247	97	36	35
Match with more than one species						
Previous number of matched reads	15794	14031	5441	3370	524	503
Remaining reads	9180	8262	2900	2078	280	273
Libraries with remaining true positives after downstream analysis	27 at total (A06-A09, A21-A24, A37, B47-B49, B53, B54, B57, B58, C71, C72, C74, C78-C82, D61, D62, D65)	19 at total (A06-A08, A21-A23, B48, B49, B53, B57, C71, C72, C74, C78-C82, D61)	8 at total (C68, C69, C71, C72, C74, C75, C76, 1, C76, 2)	6 at total (C68, C69, C71, C72, C74, C75)	3 at total (A17, A18, A32)	2 at total (A17, A18)

630

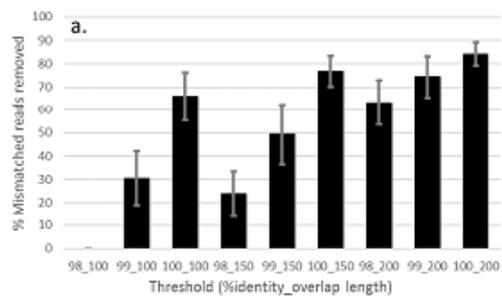


631

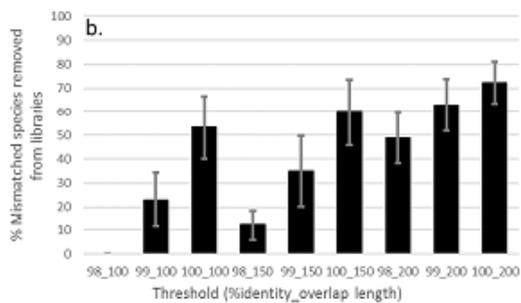


632

633 **Fig. 1** Removal of true false positives by the thresholds: a) removal of reads; b) removal
634 of true false positive species in libraries.



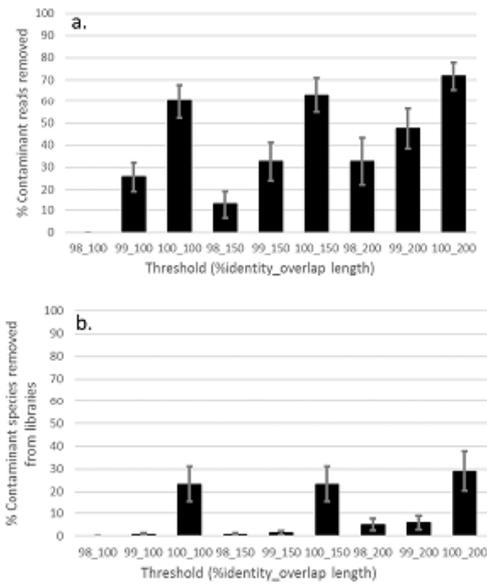
635



636

637 **Fig. 2** Removal of mismatched false positives by the thresholds: a) removal of reads; b)
638 removal of mismatched species in libraries.

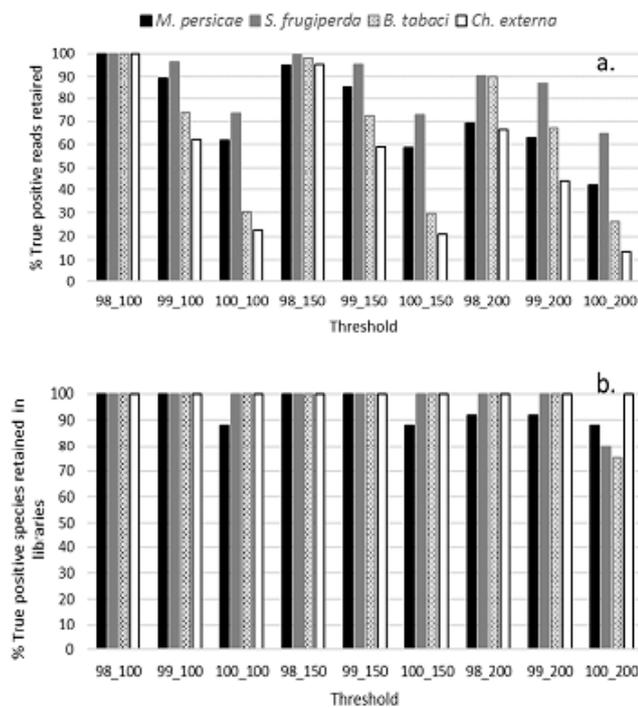
639



640

641 **Fig. 3** Removal of contaminant false positives by the thresholds: a) removal of reads; b)
642 removal of contaminant species in libraries.

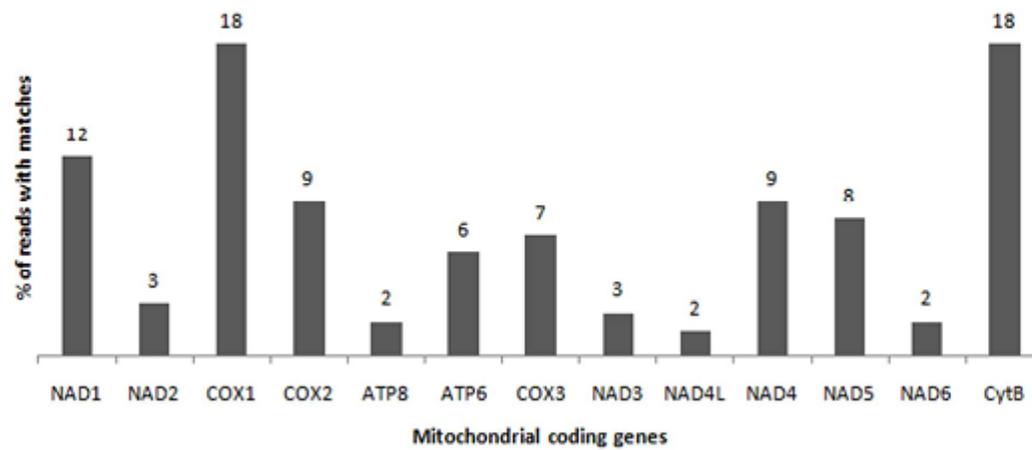
643



644

645 **Fig. 4** Retention of true positives by the thresholds for each prey species: a) retention of
646 reads; b) retention of true positive species in libraries.

647



648

649 **Figure 5** Proportion of all remaining reads that matched in mitochondrial coding genes
650 of the contaminants at 99% identity and 150 bp overlap filtering.

ANEXO 7 – Tabelas e figuras suplementares

Bioensaio A

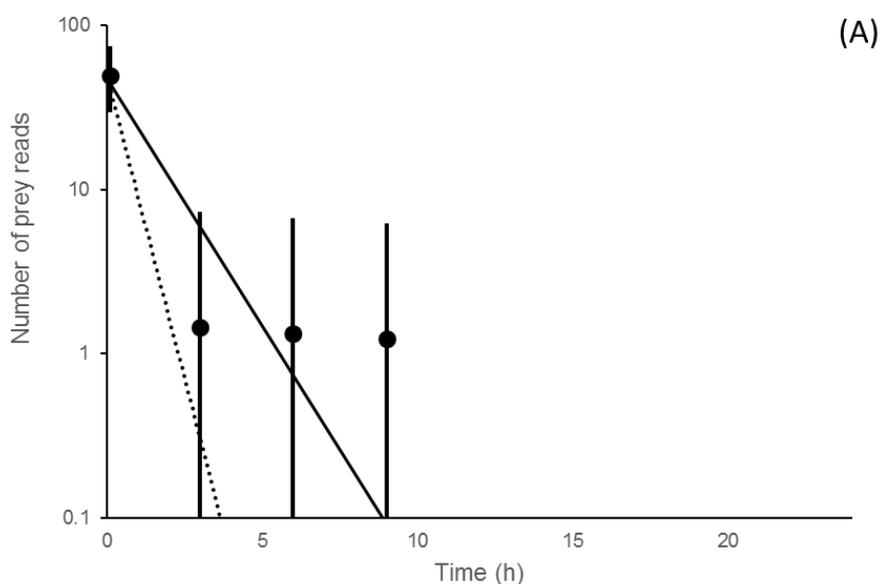


Fig. S1. Relação observada entre as reads da presa *S. frugiperda* consumida pela por larvas de *C. sanguinea* pelo método de regressão linear. Linha pontilhada: regressão não linear de dados não transformados. Linha sólida: regressão linear de dados ln transformado.

Tabela S 1. Características do modelo de retenção usando a seleção AICc. n0: detecção inicial; d: taxa de decaimento; Pred = predador.

Fatores afetando n_0	Fatores afetando d	# Parâmetros	r^2 ajustado	AICc	Probabilidade Relativa
Pred Presa Pred*Presa	Tempo*Presa	8	0,9953	157,27	0,07341
Pred Presa	Tempo	5	0,9944	155,29	0,19707
Pred Presa	Tempo*Presa	6	0,9953	152,05	1,00000
Pred	Tempo	4	0,9949	155,00	0,22814
Pred	Tempo*Presa	5	0,9956	153,20	0,55998

Bioensaio C

Tabela S2. Teste contendo o número diferencial de reads de *M. persicae* detectado no conteúdo gastrointestinal de *H. axyridis* após predação primária ou secundária.

	Graus de liberdade	QM	<i>F</i>	valor- <i>p</i>
Predador	1	8,70177094	9,26	0,0931
Tempo	2	19,6138301	20,87	0,0457
Erro	2	0,93971193		

Predador= efeito da predação primária versus predação secundária. Tempo= tempo pós-predação. Ambos os fatores são categóricos. QM: quadrado médio; *F*: razão Fisher dos quadrados médios.

Tabela S3. O modelo selecionado usando AICc e r^2 ajustado. O modelo selecionado está destacado em cinza. O intercepto do modelo é um fator categórico e o declive do modelo é uma variável contínua.

Teste	Fatores afetando n_0	Fatores afetando d	Parâmetros	r^2	valor-p	AIC _c
1	DI	TimeHa	3	0,9961	0,0001	66,5574
1	Int	TimeHa*DI	3	0,832	0,032	59,8398
1	Int	TimeHa	2	0,5466	0,0569	37,5215
2	TimeHa	TimeHa*TimeCe	6	0,9992	0,0001	127,55853
2	TimeHa	TimeCe	4	0,9975	0,0001	46,80791
2	Int	TimeHa*TimeCe	4	0,1966	0,2905	65,6879
2	Int	TimeCe	2	0,4197	0,0352	46,58851
3	TimeCe	TimeCe*TimeHa	6	0,9953	0,0003	143,73376
3	TimeCe	TimeHa	4	0,9935	0,0001	55,28737
3	Int	TimeCe*TimeHa	4	0,5201	0,0889	61,05025
3	Int	TimeHa	2	0,6251	0,0068	42,65552
4	Int	TimeHa TimeCe	3	0,7758	0,0048	43,84285
4	Int	TimeCe	2	0,4197	0,0352	46,58851
4	Int	TimeHa	2	0,6251	0,0068	42,65552
4	Int	TimeT	2	0,7977	0,0007	37,1039
5	CeFeed Hist	CeFeedHist* TimeHa	8	0,9991	0,0001	141,22801

5	CeFeedHist	TimeHa	5	0,9988	0,0001	72,51814
5	Int	CeFeedHist* TimeHa	5	0,227	0,2319	106,3749
5	Int	TimeHa	2	0,3781	0,0197	88,24491

Teste 1 = predação primária versus secundária; Teste 2 = decaimento em presas primárias (TimeCe); Teste 3 = decaimento no predador *H. axyridis* (TimeHa), Teste 4 = decaimento em relação ao tempo total (TimeT); Teste 5 = decaimento das reads da presa primárias no predador *H. axyridis*. DI = primário VS. secundário; TimeHa = tempo de digestão em *H. axyridis*; TimeCE = tempo de digestão em *Ch. externa*; CeFeedHist = história de alimentação de *Ch. externa* em presas de secundárias; Int = intercepto comum. As intercepções são todos fatores categóricos. * Denota um termo de interação.

Tabela S 4. Estimativa dos parâmetros lineares para o decaimento das reads. n_0 = número inicial de reads detectadas; d = taxa de decaimento; D_{max} = período de detectabilidade. O melhor modelo deve apresentar baixo AICc e um elevado r^2 .

Modelo		n_0	< IC 95%	> IC 95%	d (h^{-1})	< IC 95%	> IC 95%	D_{max} (h)	< IC 95%	> IC 95%
Teste 1		Parâmetro da presa secundária para consumos direto ou indireto								
Completo	Direto	790,7	312,5	2001,0	0,320	-0,141	0,780	10,0	7,4	13,0
Completo	Indireto	685,3	375,8	1249,4	1,116	-0,333	2,564	5,1	3,4	6,9
Melhor	Ambos	715,4	345,8	1479,9	0,335	-0,208	0,878	6,7	4,8	8,4
Teste 2		Tempo de digestão no predador intraguilda	Parâmetros da presa secundária na presa intraguilda							
Completo	0	685,3	585,8	801,6	1,116	0,737	1,494	5,6	3,5	8,0
Completo	3	21,5	7,6	60,7	0,893	-0,105	1,892	3,7	0,9	7,7
Completo	6	28,7	10,2	80,5	0,233	-0,263	0,728	6,4	3,2	11,7
Melhor	0	673,7	508,0	893,4	0,799	0,279	1,319	6,2	3,8	8,8
Melhor	3	21,0	3,3	132,3	0,799	0,279	1,319	3,3	0,9	5,3
Melhor	6	62,0	12,8	300,8	0,799	0,279	1,319	4,6	2,6	6,5
Teste 3		Tempo de digestão no predador intraguilda	Parâmetro da presa secundária no predador intraguilda							
Completo	0	673,8	458,8	989,6	0,663	0,274	1,052	11,0	9,0	13,3
Completo	3	24,0	1,3	448,6	0,053	-0,640	0,747	--	--	--

Completo	6	0,3	--	--	0,148	-3,642	3,939	3,3	--	--
Melhor	0	654,5	417,0	1027,231	0,514	0,115	0,913	12,6	--	--
Melhor	3	95,4	6,4	1432,2	0,514	0,115	0,913	8,9	--	--
Melhor	6	1,0	0,0	--	0,514	0,115	0,913	--	--	--
Teste 4		Tempo total pós-predação da presa secundária								
Melhor		655,6	452,5	949,7	0,563	0,330	0,796	8,7	6,5	11,3
Teste 5		Histórico de predação da presa intraguilda	Parâmetros da presa intraguilda no predador intraguilda							
Completo	Sem presa secundária	7436,0	6006,9	9205,2	0,540	0,415	0,665	18,7	17,5	19,8
Completo	Presa secundária 0 h	1005,4	649,1	1557,3	0,504	0,171	0,838	10,9	9,0	13,1
Completo	Presa secundária 3 h	920,9	487,4	1739,9	0,288	0,064	0,511	25,5	21,8	30,2
Completo	Presa secundária 6 h	59,6	8,1	436,6	0,644	-0,399	1,687	7,1	4,9	10,1
Melhor	Sem presa secundária	7148,3	5606,4	9114,2	0,484	0,363	0,605	13,6	11,8	15,4
Melhor	Presa secundária 0 h	994,9	618,6	1600,2	0,484	0,363	0,605	15,2	13,5	17,0
Melhor	Presa secundária 3 h	1274,8	673,9	2411,6	0,484	0,363	0,605	7,5	5,8	9,1
Melhor	Presa secundária 6 h	59,6	7,2	493,9	0,484	0,363	0,605	18,2	16,0	20,4