

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

ELEMENTOS GENÔMICOS DE REPRODUÇÃO SEXUADA EM  
MICROSPORÍDEOS

DISSERTAÇÃO DE MESTRADO

JULIANO DE OLIVEIRA SILVEIRA

PORTO ALEGRE, JUNHO DE 2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

**ELEMENTOS GENÔMICOS DE REPRODUÇÃO SEXUADA EM  
MICROSPORÍDEOS**

Dissertação submetida ao Programa  
de Pós-Graduação em Biologia  
Celular e Molecular do Centro de  
Biotecnologia da UFRGS como  
requisito parcial para obtenção do  
grau de Mestre.

**JULIANO DE OLIVEIRA SILVEIRA**

KAREN LUISA HAAG – ORIENTADORA

PORTO ALEGRE, JUNHO DE 2017

ESTE TRABALHO FOI DESENVOLVIDO NO LABORATÓRIO DE GENÔMICA EVOLUTIVA, LOCALIZADO NO DEPARTAMENTO DE GENÉTICA DA UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL, SOB ORIENTAÇÃO DA DOUTORA KAREN LUISA HAAG. APOIO FINANCEIRO: COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES) E CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPQ).

## **BANCA EXAMINADORA**

---

### **CHARLEY CHRISTIAN STAATS**

LABORATÓRIO DE FUNGOS DE IMPORTÂNCIA MÉDICA E BIOTECNOLÓGICA  
CENTRO DE BIOTECNOLOGIA - UFRGS

---

### **CLAUDIA ELIZABETH THOMPSON**

UNIDADE DE BIOLOGIA TEÓRICA E COMPUTACIONAL  
CENTRO DE BIOTECNOLOGIA - UFRGS

---

### **NELSON JURANDI ROSA FAGUNDES**

LABORATÓRIO DE GENÉTICA MÉDICA E EVOLUÇÃO  
DEPARTAMENTO DE GENÉTICA - UFRGS

---

### **RODRIGO LIGABUE BRAUN**

(RELATOR E SUPLENTE)

LABORATÓRIO DE PROTEÍNAS TÓXICAS  
CENTRO DE BIOTECNOLOGIA - UFRGS

*DEDICADO ÀS BRASILEIRAS E AOS BRASILEIROS QUE FINANCIAM MEUS ESTUDOS HÁ MAIS DE 10 ANOS, ATRAVÉS DO PAGAMENTO DE IMPOSTOS, SEM O APOIO DOS QUAIS NÃO SERIA POSSÍVEL TER VIVENCIADO AS OPORTUNIDADES QUE TIVE E CHEGADO A ESSE NÍVEL DE FORMAÇÃO.*

---

*"I am because we are."*

- Ubuntu philosophy

---

# SUMÁRIO

SUMÁRIO .....	VI
LISTA DE ABREVIATURAS E SÍMBOLOS .....	VIII
ÍNDICES DE FIGURAS E TABELAS .....	IX
RESUMO .....	XI
ABSTRACT .....	XII
1 INTRODUÇÃO .....	13
2 REVISÃO DE LITERATURA .....	18
2.1 GENOMAS DE MICROSPORÍDEOS.....	18
2.1.1 Aspectos Gerais .....	18
2.2 SEXO EM MICRORGANISMOS EUCARIÓTICOS PATOGÊNICOS .....	20
2.2.1 Aspectos Gerais .....	20
2.2.2 Sexo e Microsporídeos.....	22
2.2.3 Características do gênero <i>Hamiltosporidium</i> ( <i>Hamiltosporidium tvaerminnensis</i> e <i>Hamiltosporidium magnivora</i> ) .....	23
3 JUSTIFICATIVA .....	25
4 OBJETIVOS.....	26
4.1 OBJETIVO GERAL .....	26
4.2 OBJETIVOS ESPECÍFICOS .....	26
5 METODOLOGIA.....	27
5.1 ORIGEM DOS DADOS .....	27
5.2 AVALIAÇÃO E CONTROLE DE QUALIDADE DAS LEITURAS .....	28
5.3 DRAFT ASSEMBLIES DOS GENOMAS DE HAMILTOSPORIDIUM .....	29
5.4 PREDIÇÃO E ANOTAÇÃO DE GENES .....	30
5.5 ANÁLISE DIFERENCIAL E CLUSTERIZAÇÃO DE GENES .....	30
6 RESULTADOS E DISCUSSÃO .....	32
6.1 CONTROLE DE QUALIDADE DOS DADOS DE SEQUENCIAMENTO .....	32
6.2 AVALIAÇÃO DOS DRAFT ASSEMBLIES.....	34
6.2.1 FILTRAGEM DE CONTAMINAÇÕES.....	36
6.3 ANOTAÇÃO DOS GENOMAS DE HAMILTOSPORIDIUM .....	39
6.4 CARACTERIZAÇÃO GENÔMICA .....	40
6.4 TRANSFERÊNCIA HORIZONTAL GÊNICA.....	47
6.5 GENES E SEXO .....	51
7 CONCLUSÕES.....	56
REFERÊNCIAS .....	57

3. APÊNDICES.....	64
APÊNDICE 1: PARÂMETROS E SCRIPT SICKLE .....	64
APÊNDICE 2: PARÂMETROS E SCRIPT SPADES.....	65
APÊNDICE 3: PARÂMETROS E SCRIPT FINDORFS.SH .....	66
APÊNDICE 4: SCRIPT GETTAXID.PL.....	67
APÊNDICE 5: SCRIPT BIN_TAXIDS.PL .....	70
APÊNDICE 6: SCRIPT RUNTAXONOMIZEDBLAST.PL .....	72
APÊNDICE 7: SCRIPT RUN_RNAHAMMER.PL .....	74
APÊNDICE 8: SCRIPT RUN_TRNASCAN.PL.....	75
APÊNDICE 9: PARÂMETROS MAKER.....	76
APÊNDICE 10: CLUSTERS COMPARTILHADOS PELAS 4 LINHAGENS DE MICROSPORÍDEOS.	82
APÊNDICE 11: ANOTAÇÕES FUNCIONAIS DE GENES EXCLUSIVOS DE ILBN-2.....	99
APÊNDICE 12: TABELA DE CATEGORIAS COG .....	110
APÊNDICE 13: ANOTAÇÕES DE GENES COMPARTILHADOS ENTRE ILBN-2 E BEOM-2 A PARTIR DE BLAST BIDIRECIONAL .....	111



## LISTA DE ABREVIATURAS E SÍMBOLOS

<b>AA</b>	aminoácido(s)
<b>HMM</b>	Modelo Oculto de Markov (do inglês, Hidden Markov Model)
<b>KEGG</b>	Kyoto Encyclopedia of Genes and Genomes
<b>NCBI</b>	National Center for Biotechnology Information
<b>nt</b>	nucleotídeo(s)
<b>ORF</b>	Quadro Aberto de Leitura (do inglês Open Reading Frame)
<b>Pfam</b>	Protein Family Database
<b>SMART</b>	Simple Modular Architecture Research Tool

## ÍNDICES DE FIGURAS E TABELAS

FIGURA 1 LINHA DO TEMPO DAS DIFERENTES MUDANÇAS DE CLASSIFICAÇÃO TAXONÔMICA DOS MICROSPORÍDEOS, DESDE SUA DESCOBERTA EM 1857 ATÉ SUA ÚLTIMA MUDANÇA EM 2009. ADAPTADO DE (KEELING, P., 2009) .....	14
FIGURA 2 REPRESENTAÇÃO GERAL DO ESPORO DE MICROSPORÍDEOS (ADAPTADO DE FRANZEN, CASPAR, 2004) .....	16
FIGURA 3 LOCALIZAÇÃO GEOGRÁFICA DAS TRÊS LINHAGENS DE HAMILTOSPORIDIUM UTILIZADAS. EM AZUL HAMILTOSPORIDIUM MAGNIVORA E EM VERDE HAMILTOSPORIDIUM TVAERMINNENSIS.....	28
FIGURA 4 VISUALIZAÇÃO DA QUALIDADE DAS BASES DA BIBLIOTECA DE H. TVAERMINNENSIS (FI-OER-3-3) ANTES (A) E DEPOIS (B) DOS CONTROLES DE QUALIDADE. O EIXO X INDICA A POSIÇÃO DAS BASES NAS LEITURAS E O EIXO Y INDICA OS VALORES DE QUALIDADE PHRED. ....	33
FIGURA 5 TAMANHO DAS MONTAGENS ANTES E DEPOIS DOS CICLOS DE FILTRAGEM. DE FORA PARA DENTRO: I - PRIMEIRA MONTAGEM A PARTIR DE TODAS AS READS JÁ FILTRADAS POR QUALIDADE E TAMANHO; II – MONTAGEM RESULTANTE APÓS PRIMEIRO CICLO DE FILTRAGEM DE CONTAMINANTES; III – MONTAGEM FINAL PARA ANÁLISES SUBSEQUENTES. A: FI-OER-3-3 B: BE-OM-2 C: IL-BN-2 .....	37
FIGURA 6 DIAGRAMA DE VENN DE CLUSTERS ORTÓLOGOS COMPARTILHADAS E EXCLUSIVAS A PARTIR DE ANÁLISE DE CLUSTERS COM ORTHOVENN UTILIZANDO O MICROSPORÍDEO MODELO ENCEPHALITOOZON CUNICULI LINHAGEM EC1 PARA COMPARAÇÃO. NÚMERO DE PROTEÍNAS UTILIZADAS NA ANÁLISE ENCONTRAM-SE NA TABELA 8.....	41
FIGURA 7 DIAGRAMA DE VENN DE CLUSTERS ORTÓLOGOS COMPARTILHADAS E EXCLUSIVAS A PARTIR DE ANÁLISE DE CLUSTERS COM ORTHOVENN. ....	44
FIGURA 8 DIAGRAMA DE VENN DE PROTEÍNAS COMPARTILHADAS E EXCLUSIVAS A PARTIR DE BLAST BIDIRECIONAL. (NÚMERO EM PARÊNTESES REPRESENTA FRAÇÃO ANOTADA FUNCIONALMENTE COM EGGNOG DOS GENES EM CADA CATEGORIA).....	45
FIGURA 9 MAPA DAS CATEGORIAS COG COMPARTILHADAS ENTRE AS LINHAGENS DE HAMILTOSPORIDIUM MAGNIVORA A PARTIR DA ANÁLISE DE BLAST RECÍPROCO E ANOTAÇÃO FUNCIONAL COM EGGNOG. ....	46
FIGURA 10 GRÁFICO DAS CATEGORIAS COG LINHAGEM-ESPECÍFICA ENTRE AS LINHAGENS DE HAMILTOSPORIDIUM A PARTIR DA ANÁLISE DE BLAST RECÍPROCO E ANOTAÇÃO FUNCIONAL COM EGGNOG. BARRAS SOMBREANDO CÓDIGOS COG: LARANJA – PROCESSAMENTO E ARMAZENAMENTO DA INFORMAÇÃO, AMARELA – METABOLISMO,	

VERDE – PROCESSOS CELULARES E SINALIZAÇÃO, MARROM – POUCO CARACTERIZADAS. * QUEBRA DE ESCALA: LINHAGEM ILBN-2 POSSUI 86 INCIDÊNCIAS DA SUB-CATEGORIA [L] REPLICAÇÃO. ....	46
FIGURA 11 SOBREPOSIÇÃO DOS CONTIGS DE BEOM-2 E FIOER-3-3 ONDE FOI ENCONTRADA A PROTEÍNA RNR, RIBONUCLEOSÍDEO DIFOSFATO REDUTASE, DE DAPHNIA MAGNA (MARCADA NA CAIXA VERDE). EMBORA O CONTIG 140 DE HAMILTOSPORIDIUM TVAERMINNENSIS (FIOER-3-3) SEJA MENOR QUE O DE H. MAGNIVORA (BEOM-2), ESTE ALINHA POR COMPLETO. HÁ INCLUSIVE A PRESENÇA DE UMA PROTEÍNA DE MICROSPORÍDEO À MONTANTE DE AMBAS AS SEQUÊNCIAS. ....	50
<hr/>	
TABELA 1 COMPARAÇÃO GERAL ENTRE AS ESPÉCIES DO GÊNERO HAMILTOSPORIDIUM. ....	24
TABELA 2 ORIGEM DOS DADOS: DISTRIBUIÇÃO GEOGRÁFICA E TIPO DE COLETA. ....	28
TABELA 3 MÉTRICAS DE SEQUENCIAMENTO DAS TRÊS LINHAGENS DE HAMILTOSPORIDIUM..	32
TABELA 4 COMPARAÇÃO DE MÉTRICAS DE H. MAGNIVORA (BE-OM-2) OBTIDAS A PARTIR DA MONTAGEM COM OS PROGRAMAS RAY E SPADES. ....	35
TABELA 5 COMPARAÇÃO DE MONTAGENS APÓS PRIMEIRO CICLO DE FILTRAGEM. ....	37
TABELA 6 COMPARAÇÃO DE MONTAGENS APÓS SEGUNDO CICLO DE FILTRAGEM.....	38
TABELA 7 RESULTADOS DA ANOTAÇÃO APÓS CURADORIA MANUAL E BUSCA EM BLASTP TAXONOMIZADO E BUSCA DE ANOTAÇÃO FUNCIONAL NO BANCO DE DADOS EGGNOG. ....	39
TABELA 8 NÚMERO DE PROTEÍNAS, CLUSTERS E SINGLETONS UTILIZADOS NA ANÁLISE DE CLUSTERS ORTÓLOGOS.. ....	42
TABELA 9 POSSÍVEIS EVENTOS DE TRANSFERÊNCIA GÊNICA HORIZONTAL DE DAPHNIA MAGNA PARA HAMILTOSPORIDIUM. ....	49
TABELA 10 GRUPO DE GENES CORE DE MEIOSE HOMÓLOGOS DISTRIBUÍDOS EM MICROSPORÍDEOS E EM LEVEDURA. ....	52

## RESUMO

O valor adaptativo do sexo em patógenos unicelulares eucarióticos é um assunto de intenso debate, e os mecanismos genéticos envolvidos na reprodução sexual desses organismos são pouco desconhecidos. Diferenças nos sistemas de reprodução podem ajudar a explicar as diferenças nos modos de transmissão, amplitude de gêneros hospedeiros e resposta às mudanças ambientais, todas estreitamente relacionadas à patogenicidade e à virulência desses patógenos. Este trabalho tem como objetivo identificar os elementos genômicos que podem estar associados aos modos de reprodução e elucidar o papel dos ciclos sexuais em um microsporídeo. Microsporídeos fazem parte de um filo de fungos patogênicos intracelulares unicelulares. Estes parasitas são encontrados em praticamente todos os tipos de animais, sendo de importância para saúde e agricultura. *Hamiltosporidium tvaerminnensis* é um microsporídeo assexuado que é transmitido vertical e horizontalmente ao seu hospedeiro, o microcrustáceo *Daphnia magna*. *H. tvaerminnensis* possui uma espécie irmã, *Hamiltosporidium magnivora*, que difere por ser sexual e apenas verticalmente transmitida. Nosso trabalho usa como referência um genoma *draft* de *H. tvaerminnensis* de 2009. É um genoma considerado grande para um microsporídeo, com tamanho estimado de 25 Mb. Contamos com sequências de Illumina de duas linhagens recentemente sequenciadas de *H. magnivora*, uma da Bélgica e outra de Israel, bem como um resequenciamento da mesma linhagem de *H. tvaerminnensis*, da Finlândia, usando tecnologia mais recente que a referência. Após a montagem e anotação, os nossos resultados apontam para genes chave de meiose ausentes na linhagem assexuada, bem como um enriquecimento de elementos transponíveis e transferência genética horizontal gênica derivada do hospedeiro nas linhagens sexuais. A presença exclusiva de certos genes da meiose bem como de outros genes de reparo a danos do DNA associados à maior presença de elementos móveis sugere uma possível perda de sexo concomitante à inativação de elementos móveis na linhagem assexuada.

## ABSTRACT

The adaptive value of sex in eukaryotic unicellular pathogens is a matter of intense debate, and the genetic mechanisms involved in sexual reproduction of these organisms are largely unknown. Differences in reproduction systems could help explaining differences in modes of transmission, host range amplitude and response to environmental changes, all tightly related to pathogenicity and virulence of such pathogens. This work aims to identify genomic elements that could be associated with modes of reproduction, and further elucidate the role of sexual cycles in a microsporidian parasite. Microsporidia belong to a phylum of unicellular intracellular pathogenic Fungi. These parasites are found in virtually all types of animals, being of importance to health and agriculture. *Hamiltosporidium tvaerminnensis* is a microsporidian shown to be asexual and being both vertically and horizontally transmitted to its host, the microcrustacean *Daphnia magna*. *H. tvaerminnensis* possesses a sister species, *Hamiltosporidium magnivora*, which differs by being sexual and only vertically transmitted. Our work uses as reference a 2009 assembled draft genome of *H. tvaerminnensis*. It is an unusually large genome for a microsporidian, estimated to contain 25 Mb of sequence. We rely on Illumina sequencing reads from two recently resequenced lineages of *H. magnivora*, one from Belgium and another from Israel, as well as a set from the same *H. tvaerminnensis* lineage as the reference, from Finland, using technology newer than the reference. After assembly and annotation, our results point to missing key meiosis genes in the asexual lineage, as well as an enrichment in transposable elements and host derived horizontal gene transfer in the sexual lineages. The exclusive presence of certain meiosis genes as well as other DNA damage repair genes associated to the larger presence of mobile elements hint to a possible loss of sex concomitant to the inactivation of mobile elements in the asexual lineage.

# 1 INTRODUÇÃO

A evolução do sexo é um dos problemas mais importantes e controversos da biologia evolutiva (TAKAGI, 2010). Embora o sexo seja quase universal em animais e plantas superiores, seus custos inerentes tornam sua manutenção difícil de explicar, especialmente em organismos eucarióticos unicelulares. Um dos principais é o custo dobrado do sexo, uma vez que uma população de fêmeas partenogenéticas é capaz de gerar o dobro do número de descendentes se comparada a uma população sexuada do mesmo tamanho composta de machos e fêmeas na proporção de 1:1. Por outro lado, o sexo introduz uma fonte de variabilidade genética através da recombinação alélica e permite acelerar a taxa de expulsão de alelos deletérios nas populações. Ao longo do século passado, várias hipóteses, juntamente com evidências experimentais para apoiá-las, foram apresentadas para explicar o sexo generalizado, levando em consideração o balanço entre seus custos e benefícios (SCHURKO; NEIMAN; LOGSDON, 2009).

Este balanço é especialmente delicado em organismos eucarióticos patogênicos, que necessitam gerar rápidas adaptações em resposta as defesas do sistema imune do hospedeiro. Além disso, a reprodução sexuada nesses organismos tem implicações em saúde pública, ecologia e agricultura pois facilita a disseminação de mutações relacionadas a resistência a drogas, ou patogenicidade, por exemplo. O conhecimento da estrutura genômica de organismos parasitas é um importante passo para o desenvolvimento de novas terapias ou intervenções e métodos de prevenção, bem como para a compreensão das origens do parasitismo (JACKSON, 2015; JOHNSTON, D. A. *et al.*, 1999). Tal abordagem permite uma melhor compreensão das relações parasita-hospedeiro, a história evolutiva do patógeno, além do status de sua variabilidade genética.

Os microsporídeos (ou Filo Microsporidia) constituem um grupo de fungos parasitas com aproximadamente 1200 espécies identificadas (DIDIER, Elizabeth S., 2005; KEELING, P J; FAST, 2002) classificadas em cerca de 150 gêneros (FRANZEN, Caspar, 2004; MURRAY WITTNER, 1999) capazes de parasitar alguns

grupos de invertebrados e as cinco classes de vertebrados (CANNING *et al.*, 2005; DIDIER, Elizabeth S., 2005). Apesar de terem sido inicialmente classificados como leveduras por Nägeli em 1857, os Microsporídeos foram reclassificados como protozoários e como integrantes do extinto Reino Archezoa. Foi nos últimos anos do século XX que, através de investigações filogenéticas, as evidências começavam a apontar para uma ancestralidade comum à dos fungos, até que em 2004 se propôs definitivamente a origem fúngica dos Microsporídeos (THOMARAT, Fabienne; VIVARÈS, Christian P.; GOUY, Manolo, 2004). Entretanto, contrariamente ao que seria esperado, em 2009 se propôs a exclusão dos Microsporídeos da Internacional Code of Botanical Nomenclature (ICBN) para manter a descrição das espécies seguindo regras de protistas, já que na época existiam aproximadamente 1000 espécies já descritas nesse sistema (Figura 1). Apesar de seguirem as normas zoológicas de taxonomia, está estabelecido que os microsporídeos correspondem a uma linhagem monofilética basal de fungos que evoluiu com uma forte tendência a redução fisiológica e genômica (CAPELLA-GUTIÉRREZ; MARCET-HOUBEN; GABALDÓN, 2012; HAAG, K. L. *et al.*, 2014).

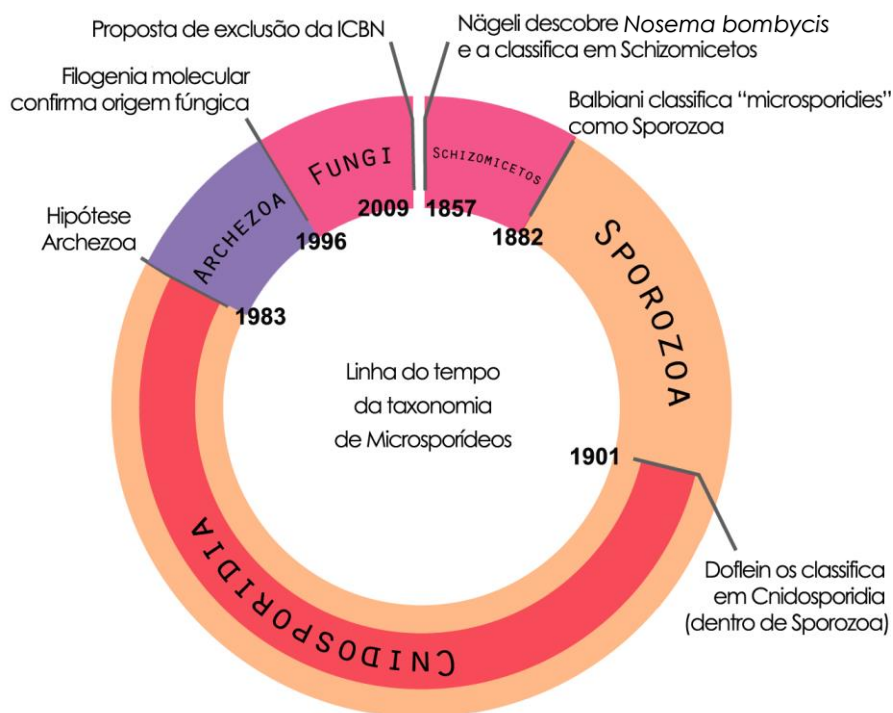
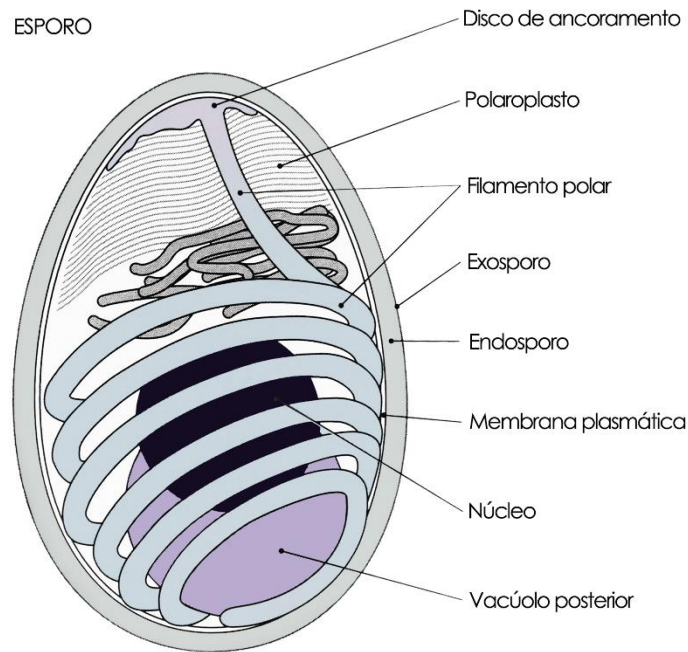


Figura 1 Linha do tempo das diferentes mudanças de classificação taxonômica dos microsporídeos, desde sua descoberta em 1857 até sua última mudança em 2009. Adaptado de (KEELING, P., 2009)

Microsporídeos são eucariotos verdadeiros, ou seja, possuem um núcleo tipicamente eucarioto, um sistema de membranas internas e citoesqueleto, mas revelam características moleculares e citológicas semelhantes a procariotas que incluem o tamanho do genoma, em geral comparável ao de bactérias, a ausência de mitocôndrias identificáveis, de peroxissomos e de um aparelho de Golgi típico. A característica mais peculiar de um microsporídeo é a sua célula infecciosa – o esporo – que é a única fase do ciclo de vida viável fora da célula hospedeira e mais facilmente reconhecível (KEELING, P J; FAST, 2002). É uma célula única, cujas dimensões podem variar entre de 1-40  $\mu\text{m}$  (FRANZEN, C.; MÜLLER, 1999). Há variação também na forma, de arredondada a bastonete, mas a maioria é ovoide (KEELING, P J; FAST, 2002). O esporo é constituído por uma parede grossa que é composta por três camadas: o exósporo (camada externa eletronicamente densa e de composição proteica), o endósporo (camada interna eletronicamente translúcida, e de composição quitinosa) e a membrana plasmática que envolve o citoplasma, o núcleo ou núcleos, vacúolo posterior e o aparelho infeccioso ou de extrusão (BIGLIARDI; SACCHI, 2001). Estes parasitas são especialmente conhecidos por esse mecanismo peculiar de infecção, que é composto de três componentes: o filamento polar, estrutura tubular enrolada em hélice ligada ao ápice do esporo através do disco de ancoramento, o polaroplasto e o vacúolo posterior (Figura 2). O mecanismo infeccioso dos microsporídeos se dá pela extrusão deste filamento polar que inocula o esporoplasma na célula hospedeira (FRANZEN, Caspar, 2004). Com a entrada do esporoplasma do parasita na célula hospedeira, inicia-se a fase proliferativa do ciclo de vida, caracterizada pela ocorrência da merogonia, evento em que o parasita aumenta massivamente o seu número com a formação de células designadas merontes. Segue-se a fase de esporogonia que consiste na formação das células infecciosas que, ao germinarem, reiniciam o ciclo (BIGLIARDI; SACCHI, 2001; FRANZEN, Caspar, 2004).





**Figura 2 Representação geral do esporo de microsporídeos (adaptado de FRANZEN, Caspar, 2004)**

Supôs-se por muito tempo que os microsporídeos eram completamente assexuados. Dessa forma, o sexo teria sido eliminado de seus ciclos de vida pelas mesmas forças que levaram a extremos de miniaturização morfológica e compactação genômica, porém evidências apontam para possíveis eventos de reprodução sexuada ou ciclos parassexuais (IRONSIDE, 2007; LEE, S. C.; WEISS, Louis M.; HEITMAN, 2009). Mesmo assim, a existência de reprodução sexuada em diversos integrantes do grupo ainda é uma pergunta em aberto. Se o sexo é encontrado predominantemente entre microsporídeos como é entre outros eucariotos, isso dá credibilidade à ideia de que o sexo é de alguma forma vital para a sobrevivência ao longo prazo de todos os organismos eucarióticos. Se mesmo microsporídeos que possuem genomas tão compactos mantêm a capacidade de se reproduzir sexualmente, então o sexo deve ser realmente importante. Portanto, é relevante saber não só se os microsporídeos se reproduzem sexualmente, mas como o sexo é distribuído entre os microsporídeos e quais são os efeitos da reprodução sexuada sobre a variação genotípica e fenotípica dos microsporídeos se e quando ela ocorre.

Como a reprodução sexuada de microrganismos eucarióticos tende a ser rara, furtiva ou críptica, encontrar evidência de reprodução sexual torna-se uma tarefa bastante desafiadora (BIRKY, C. W., 2010). Muitas vezes supõe-se que se

não há evidência de reprodução sexual em um organismo, este está se reproduzindo assexuadamente, o que pode ou não ser verdade. Além disso, a reprodução sexual pode ocorrer em um ambiente de laboratório, mas nunca realmente ocorrer naturalmente. A evidência para a reprodução sexual pode ser observada em uma variedade de formas: interações sexuais confirmadas, evidência direta para interações sexuais e evidências indiretas. Uma interação sexual confirmada é a observação confirmada de uma mudança na ploidia de um organismo seguida por uma restauração da ploidia normal. A evidência direta é fornecida ao observar a redução ou restauração da ploidia, bem como a presença de genes da meiose. Evidências indiretas são fornecidas por qualquer número de observações: evidência de recombinação homóloga, fusão citoplasmática, confirmação de um ciclo de vida complexo, ou produção de células reprodutoras em potencial (LAHR *et al.*, 2011). Finalmente, existe a evidência de genes homólogos àqueles em outros organismos que estão envolvidos com a reprodução sexuada. Existem principalmente dois tipos desses genes: os genes *Mating Type* presentes no locus MAT e os genes envolvidos com a meiose (CARPENTER *et al.*, 2012; CERVANTES *et al.*, 2013; HEITMAN, 2010). Portanto, uma maneira indireta de verificar a capacidade de realizar reprodução sexuada é a verificação do conteúdo genômico de uma espécie e analisá-la comparativamente. Neste contexto, o presente estudo teve como objetivo obter amostras dos genomas de *Hamiltosporidium tvaerminnensis* e *Hamiltosporidium magnivora*, espécies irmãs de microsporídeos que diferem quanto ao meio de reprodução, que permitissem caracterizá-los com relação ao conteúdo de genes relacionados a reprodução sexuada, de forma a agregar informações àquelas já existentes do grupo. Através da abordagem utilizada, este trabalho caracteriza a primeira montagem do genoma de *H. magnivora* sequenciado utilizando dados de sequenciamento de alto rendimento.

## 2 REVISÃO DE LITERATURA

### 2.1 GENOMAS DE MICROSPORÍDEOS

#### 2.1.1 Aspectos Gerais

Diversas razões contribuem para o interesse dos pesquisadores no sequenciamento de genomas de microsporídeos. Esses organismos são parasitas intracelulares obrigatórios de diversos grupos de animais. Além das implicações de ordem econômica, na produção de animais para o abate, e de saúde, existem consequências de importância ecológica (VILCINSKAS *et al.*, 2014). Ademais, recentemente, microsporídeos foram considerados patógenos emergentes na cadeia alimentar global (STENTIFORD *et al.*, 2016).

O primeiro genoma completo de microsporídeo a ser sequenciado, *Encephalitozoon cuniculi*, em 2001, foi um marco significativo no estudo da biologia de microsporídeos (KATINKA, M. D. *et al.*, 2001). A organização do genoma e o conteúdo de genes de *E. cuniculi* representam a primeira evidência de redução genômica de microsporídeos - 2,9 Mb e aproximadamente 2000 genes. Comparativamente, a levedura modelo *Saccharomyces cerevisiae*, que possui um genoma relativamente pequeno, possui aproximadamente 6000 genes, o que demonstra a grande redução por que passaram os microsporídeos (GOFFEAU *et al.*, 1996). O genoma de *E. cuniculi* revelou a enorme dependência do hospedeiro, resultado de um estilo de vida intracelular obrigatório, no qual o parasita perdeu muitas vias bioquímicas ou componentes para processos celulares centrais. Incluindo o ciclo do ácido tricarbóxico, a maquinaria para biossíntese do nucleotídeos e proteínas e a maquinaria exigida para a interferência do RNA (RNA pol II, dicer e argonauta) (CORRADI; SLAMOVITS, 2011).

Uma série de projetos de sequenciamento de microsporídeos já foram concluídos, incluindo amostragens de espécies provenientes de diversos hospedeiros, tais como *Nematocida parisii*, um patógeno natural do nematódeo modelo *C. elegans*, o patógeno de abelhas *Nosema ceranae* e *Enterecytozoon*

*bieneusi*, uma espécie que parasita humanos, sequenciada a partir de fezes de doentes (AKIYOSHI *et al.*, 2009; CORNMAN *et al.*, 2009; CUOMO *et al.*, 2012). Essa prospecção revelou uma grande variedade no tamanho do genoma de microsporídeos desde 2,3 Mb do parasita humano *Encephalitozoon intestinalis*, um dos menores genomas nucleares conhecidos, até o genoma de 24 Mb de *Octosporea bayeri* (agora *Hamiltosporidium tvaerminnensis*) (CORRADI *et al.*, 2009, 2010). Curiosamente, esta variação drástica no tamanho do genoma até 10 vezes maior não se traduz em aumento na capacidade codificante e demonstra que a variação na proporção de elementos não codificantes presentes, portanto, reflete o nível de compactação do genoma.

Os microsporídeos adotaram várias estratégias para compactar seus genomas. Em primeiro lugar, eles têm distâncias intergênicas extremamente curtas, que são em média 129 pb em *E. cuniculi*, muito mais curtas do que em *S. cerevisiae* que tem um genoma que é aproximadamente 50% menos denso e onde os espaços intergênicos têm uma média estimada de 500 pb (DUJON, 1996). Esta distância intergênica extremamente curta é causada pela falta de regiões repetitivas e a presença de muito poucos elementos egoístas, como transposons nos representantes mais enxutos (CORRADI *et al.*, 2009). É observado também um grau de compactação do genoma dentro dos próprios genes. Eles possuem muito poucos introns, em alguns casos nenhum, e as proteínas homólogas compartilhadas entre *E. cuniculi* e *S. cerevisiae* são, em média, 15% mais curtas no primeiro (KATINKA, M. D. *et al.*, 2001). Genomas de microsporídeos variam muito em tamanho, principalmente devido a diferenças na proporção de elementos genômicos não codificantes. Eles perderam muitos genes importantes para biossíntese de macromoléculas e adotaram um modo de vida muito dependente do hospedeiro no ambiente intracelular. A prospecção de genomas de microsporídeos tem proporcionado um conhecimento funcional considerável sobre a evolução das relações microsporídeo-hospedeiro nos níveis bioquímico e molecular. O aumento na quantidade de genomas de microsporídeos sequenciados, principalmente a partir do ano de 2014 no qual 8 sequências de genomas foram publicadas, está relacionado ao constante desenvolvimento de tecnologias de obtenção e análise de dados. Essas novas

tecnologias permitem maior rendimento, com menor tempo e custo por base, quando comparadas aos métodos anteriores.

## 2.2 SEXO EM MICRORGANISMOS EUCARIÓTICOS PATOGÊNICOS

### 2.2.1 Aspectos Gerais

Uma das principais inovações nos eucariotos foi a origem da reprodução sexuada através da meiose com suas consequências para efetuar a recombinação do genoma (CAVALIER-SMITH, 2002). A reprodução sexuada é comumente definida como um meio de propagação genética que requer que 2 indivíduos combinem material genético, geralmente unindo 2 células (gametas) contendo cromossomos dos pais, para formar um zigoto. Antes dos gametas serem produzidos, os genomas dos pais sofrem recombinação durante a meiose (KLECKNER, 1996). Portanto, o estudo da evolução do sexo também busca determinar quais condições favorecem a evolução da recombinação, visto que esta é vista como um precursor do aparecimento de sexo obrigatório.

Isso contrasta com a reprodução assexuada, onde um organismo clona seu genótipo para se reproduzir. A assexualidade é muito rara entre os eucariotos, com aproximadamente 0,1% das espécies animais reproduzindo assexuadamente (VRIJENHOEK, 1998). A maioria das linhagens dos assexuais evoluíram recentemente de predecessores sexuais (SIMON *et al.*, 2003; VRIJENHOEK, 1998), embora existam alguns assexuais 'antigos', sendo os rotíferos bdeloídeos os candidatos mais conhecidos (VRIJENHOEK, 1998).

O sexo parece ser ubíquo entre os eucariotos; no entanto, pouco se sabe sobre sexo e meiose fora de animais, plantas e dos chamados "fungos perfeitos". Alguns fungos unicelulares são conhecidos por realizar reprodução sexuada, mas para muitos - incluindo microsporídeos - ainda são poucas as evidências de reprodução sexuada. Os fungos patogênicos parasitas, que já foram considerados estritamente assexuais, parecem ter mantido ciclos sexuais

crípticos, incluindo reprodução unissexuada ou parassexual. Para alguns desses supostos assexuais, uma presunção razoável é que eles perderam a habilidade de realizar reprodução sexuada, tendo sido anteriormente capazes disso.

## 2.2.2 Sexo e Microsporídeos

A reprodução sexuada aqui é definida como modo de reprodução que produz descendentes cuja constituição genética é uma mistura de dois gametas potencialmente diferentes geneticamente, geralmente envolvendo meiose e singamia. Como modo de reprodução, o sexo parece bastante ineficiente, mas é quase onipresente entre os organismos eucarióticos. Os microsporídeos foram considerados uma possível exceção a esta regra. Por muito tempo se acreditou que os microsporídeos eram plenamente incapazes de realizar reprodução sexuada, assim como outros grupos de eucariotos que teriam divergido cedo na evolução dos eucariotos (*i.e.*, Diplomonadas e Thricomonas). Como não se observava mudança de ploidia ou outra evidência de sexo nesses organismos, acreditava-se que a meiose, e conseqüentemente a reprodução sexuada, surgira tardiamente na evolução dos eucariotos. Em 2007, descobriu-se que isolados de *Giardia lamblia* de três linhagens distintas tinham uma diversidade genética extremamente baixa tanto em regiões codificantes como não codificantes, sugerindo que os padrões distintos de variação genética encontrados fossem resultantes de algum tipo de troca genética (TEODOROVIC; BRAVERMAN; ELMENDORF, 2007). Além disso, outro estudo encontrou a presença de cinco homólogos de genes necessários para a meiose (LOGSDON, 2008). Assim, pode-se inferir que algum tipo de troca de material genético estava ocorrendo, mas se essa troca corresponde à reprodução sexuada, não se pode determinar sem mais evidência direta. Embora hoje se saiba que a história evolutiva dos microsporídeos difere da de giárdias, ambos os grupos apresentam representantes diplocarióticos e mitose críptica. Portanto, aqui se mostra importante a busca de evidências genômicas para investigar o modo de reprodução desses organismos.

Hoje se aceita que microsporídeos evoluíram a partir de um mesmo ancestral junto com fungos, portanto é possível supor que este antepassado exibiu desenvolvimento sexual semelhante aos fungos existentes (CAPELLA-GUTIÉRREZ; MARCET-HOUBEN; GABALDÓN, 2012). A reprodução sexual em fungos inclui os três passos cruciais comuns à maioria dos organismos multicelulares

sexuais. Parceiros compatíveis devem se encontrar e reconhecer um ao outro, resultando em plasmogamia (fusão celular). Isto é seguido por cariogamia (fusão nuclear) em que os núcleos dos dois parceiros se fundem para aumentar a ploidia. Então, a meiose ocorre, resultando em recombinação genética entre dois genomas parentais e uma progênie haploide ou hemi-ploidal (HEITMAN, 2011). Algumas espécies de fungos realizam reprodução unissexuada; no entanto, o processo é semelhante ao sexo envolvendo dois parceiros. O sexo de microsporídeos deve seguir de alguma forma estas três etapas essenciais do desenvolvimento sexual, se tiverem realmente um ciclo sexual existente.

Em um nível mais aplicado, a reprodução sexual pode contribuir para a diversidade genética, influenciando a emergência e a disseminação da resistência a medicamentos. Microsporídeos tem sido descritos como agentes etiológicos de doenças oportunistas em humanos, então saber se microsporídeos são sexuados tem, portanto, implicações na nossa capacidade de tratar adequadamente estas infecções oportunistas (JOSEPH; VEMUGANTI; SHARMA, 2005; MATHIS; WEBER; DEPLAZES, 2005).

### **2.2.3 Características do gênero *Hamiltosporidium* (*Hamiltosporidium tvaerminnensis* e *Hamiltosporidium magnivora*)**

Este estudo analisa os parasitas a partir de infecções de *Daphnia magna*, um crustáceo que alterna entre ciclos partenogenéticos e sexuados que é parasitado por uma variedade de bactérias, microsporídeos e outros fungos parasitas (EBERT, D, 2005; GREEN, 1974), que impactam severamente seu *fitness* (ou aptidão) (EBERT, D.; ZSCHOKKE-ROHRINGER; CARIUS, 2000). Diversas características como a facilidade de manipulação devido ao pequeno tamanho e ausência de pigmentação do corpo, curto tempo de gerações e a possibilidade de induzir a partenogênese, facilitando o estabelecimento de linhagens isogênicas, além da grande diversidade de patógenos associados, têm promovido *Daphnia* como modelo para o estudo da coevolução parasita-hospedeiro (EBERT, D., 2011). O microsporídeo *Hamiltosporidium tvaerminnensis* que antes fora denominado *Octosporea bayeri* por Jirovec 1936 é parasita



específico de *D. magna* e é facilmente encontrado em habitats de piscinas rochosas no Mar Báltico (BEN-AMI; RIGAUD; EBERT, D, 2011). *H. tvaerminnensis* infecta principalmente o corpo adiposo e os ovários de *D. magna*, que por sua vez reduz expectativa de vida, fecundidade e habilidade competitiva (BEN-AMI; RIGAUD; EBERT, D, 2011; BIEGER; EBERT, Dieter, 2009). Este parasita é transmitido de duas maneiras, tanto verticalmente (de mãe para prole) quanto horizontalmente (através de esporos ambientais presentes na água que são ingeridos por meio de seu sistema de alimentação por filtração).

O modo de reprodução nestas duas espécies irmãs de microsporídeos foi proposto a partir do estudo de SNPs, em busca de evidências genéticas populacionais de recombinação e reprodução sexual. Após o sequenciamento de 16 loci no genoma de isolados de *H. tvaerminnensis* e *H. magnivora*, observou-se que diferentemente de sua espécie irmã, *H. tvaerminnensis* é assexuado (HAAG, Karen L.; TRAUNECKER; EBERT, Dieter, 2013).

<b>Espécie</b>	<b>Tecido Infectado</b>	<b>Modo de Transmissão</b>	<b>Modo de Reprodução</b>
<i>Hamiltosporidium tvaerminnensis</i>	Tecido adiposo, hipoderme, ovários	Vertical e Horizontal	Assexual
<i>Hamiltosporidium magnivora</i>	Tecido adiposo, hipoderme, ovários	Vertical	Sexual

**Tabela 1** Comparação geral entre as espécies do gênero *Hamiltosporidium*. Adaptado de HAAG et al., 2011

### 3 JUSTIFICATIVA

Em estudos anteriores foram observadas três fontes de evidência de assexualidade em *H. tvaerminnensis*: ausência de recombinação; heterozigose fixa; e divergência entre alelos. Ao mesmo tempo, observaram-se altas taxas de recombinação e distribuição aleatória de SNPs em *H. magnivora*, o que indica sua capacidade de realizar alguma forma de reprodução sexuada. Considerando a carência de informação genômica sobre os organismos do gênero *Hamiltosporidium*, este trabalho visa melhor elucidar o horizonte genômico de duas espécies de *Hamiltosporidium*, o que pode contribuir extensamente com os estudos de evolução do sexo e sua importância em organismos patogênicos, bem como com os estudos de co-evolução parasita-hospedeiro com *Daphnia magna*.

## 4 OBJETIVOS

### 4.1 OBJETIVO GERAL

Montar e anotar três genomas do gênero *Hamiltosporidium* e realizar análise comparativa de genes relacionados a reprodução sexuada nesses organismos.

### 4.2 OBJETIVOS ESPECÍFICOS

- Montar *contigs* dos genomas de *Hamiltosporidium tvaerminnensis* linhagem FI-OER-3-3, e de *Hamiltosporidium magnivora* linhagens IL-BN-2 e BE-OM-2 utilizando leituras obtidas com sequenciador de alto rendimento Illumina;
- Anotar genes a partir de predições automáticas e curadoria manual;
- Classificar genes funcionalmente;
- Identificar grupos de genes ortólogos nos genomas de *Hamiltosporidium*;
- Comparar os perfis gênicos par a par e através de agrupamento ("clusterização") entre as três linhagens com foco em genes relacionados à reprodução sexuada.

## 5 METODOLOGIA

A metodologia empregada neste trabalho consiste em um tratamento híbrido dos dados, tanto através de estratégias automáticas quanto de ajustes e curadoria manuais. O trabalho possui dois componentes metodológicos distintos: montagem e anotação dos genomas, e análise comparativa. A estratégia de montagem e anotação foi realizada sob o planejamento do pesquisador Jean-François Pombert, do Instituto de Tecnologia de Illinois, Chicago, EUA. Parte do treinamento do aluno foi realizado no laboratório do pesquisador (Pombert Lab) durante os meses de janeiro e fevereiro de 2016 e as análises subsequentes foram realizadas à distância utilizando o servidor em Chicago, ou no *laptop* e no servidor do laboratório de Genômica Evolutiva, na UFRGS.

### 5.1 ORIGEM DOS DADOS

Para este trabalho foram utilizados três conjuntos de dados de sequenciamento a partir de esporos purificados de *Hamiltosporidium tvaerminnensis* e *Hamiltosporidium magnivora* isolados de *Daphnia magna* coletadas em diferentes regiões da Europa e da Ásia (Tabela 2) e mantidas em cultivo. Em todos os experimentos as duas extremidades dos insertos das bibliotecas foram sequenciadas, portanto neste trabalho as sequências obtidas foram denominadas pareadas – *paired-end*. O tamanho de inserto utilizado foi de 300 bases. Foi utilizado o sequenciador Illumina HiSeq 2000 no centro de sequenciamento da Universidade da Basileia (Basileia, Suíça).

Espécie	Hospedeiro	Isolado	Tipo de Amostra	Local	Latitude, Longitude
<i>H. tvaerminnensis</i>	<i>D. magna</i>	FI-OER-3-3	Plâncton	Arquipélago	59.78859,
				Tvärminne, Finlândia	23.17414
<i>H. magnivora</i>	<i>D. magna</i>	IL-BN-2	Plâncton	Golan Heights, Israel	33.14318,
				Leuven, Bélgica	35.77409
		BE-OM-2	Plâncton		50.86328,
					4.72137

**Tabela 2 Origem dos dados: distribuição geográfica e tipo de coleta.**



**Figura 3 Localização geográfica das três linhagens de Hamiltosporidium utilizadas. Em azul Hamiltosporidium magnivora e em verde Hamiltosporidium tvaerminnensis.**

## 5.2 AVALIAÇÃO E CONTROLE DE QUALIDADE DAS LEITURAS

O arquivo de saída gerado pelo sequenciador Illumina contém a sequência de nucleotídeos e a representação dos níveis de qualidade para cada base chamada no formato *fastq*. O programa FastQC ("BABRAHAM BIOINFORMATICS", [s.d.]) foi utilizado para extrair relatórios dos conjuntos de sequências a partir dos dados brutos e comparar os resultados das diferentes estratégias de filtragem. Este programa permitiu gerar gráficos com a distribuição das sequências em relação à qualidade média total por base, a qualidade média de todas as leituras e o conteúdo G+C dos dados. Leituras com qualidade inferior ao *phred score* 30 foram descartadas, assim como leituras com tamanho

inferior a 100 bases, utilizando-se a ferramenta Sickle versão 1.233 (JOSHI; FASS, 2011) (Apêndice 1).

### 5.3 DRAFT ASSEMBLIES DOS GENOMAS DE HAMILTOSPORIDIUM

Devido à alta fragmentação da versão disponível do genoma de *H. tvaerminnensis* (CORRADI *et al.*, 2009), foi utilizada uma estratégia de montagem *de novo*. Essa estratégia possibilita a montagem de genomas pela formação de *contigs* e *scaffolds*, sem necessidade de versão prévia e mapeamento de leituras. Após montagem com todas as leituras e comparação entre os *softwares* de montagem Ray, SPAdes e DipSPAdes, o programa SPAdes versão 3.5 (NURK *et al.*, 2013) foi utilizado para obtenção dos rascunhos das montagens *de novo* para as três linhagens de *Hamiltosporidium* (Apêndice 2). Esse programa faz uso de um algoritmo baseado em grafos de Bruijn para realizar a montagem (BANKEVICH *et al.*, 2012).

Para filtragem de *contigs* estritamente pertencentes aos organismos de interesse, e não a contaminantes como bactérias ou o hospedeiro *Daphnia magna*, foi implementada uma metodologia de agrupamento baseada na predição dos quadros abertos de leitura (*Open Reading Frames* – ORFs) de pelo menos 150 aa de tamanho (450 nt) utilizando o script `FindORFs.sh` (Apêndice 3). Os ORFs foram comparados a sequências de bancos de dados de proteínas de microsporídeos, bactérias ou *Daphnia* utilizando BLASTp *output format 6* e *e-value* máximo de  $1.10^{-10}$  (CAMACHO *et al.*, 2009). Os *hits* principais deste BLAST foram utilizados para a obtenção da informação taxonômica de cada ORF utilizando o script `GetTaxIDs.sh` (Apêndice 4). A tabela contendo a identificação taxonômica associada a cada ORF foi utilizada como *input* para o script `BinTaxids.sh` que classifica cada *contig*, baseado no número de ORFs e seus *hits* taxonômicos, como de “*Daphnia*”, ou “Microsporídeo”, ou “Bactéria” ou “Incerto” (*Daphnia*, *Microsporidia*, *Uncertain*) (Apêndice 5). Os *contigs* classificados como “Incerto” foram analisados manualmente para finalizar a classificação. Apenas *contigs* classificados como “Microsporídeo” foram

utilizados subsequentemente. Após a fase de anotação de genes foi utilizado o script `Taxomizedblast.pl` (Apêndice 6) para verificar se ainda haviam *contigs* contaminantes. Nessa fase, *contigs* contendo pelo menos um gene com *hit* em “Microsporídeo” foram mantidos na análise, sendo os outros, excluídos.

#### 5.4 PREDIÇÃO E ANOTAÇÃO DE GENES

RNAs ribossomais e de transferência foram posicionados nos *contigs* de *Hamiltosporidium* utilizando os programas RNAmmer 1.2 (LAGESEN *et al.*, 2007) e tRNAscan-SE 1.3.1 (LOWE; CHAN, 2016), respectivamente (Apêndices 7 e 8). A predição gênica foi realizada utilizando-se da *pipeline* de anotação de genomas MAKER (CANTAREL *et al.*, 2008) (Apêndice 9). Os genes codificadores de proteínas preditos foram posicionados nas sequências com ferramentas incorporadoras do visualizador de genomas WebApollo (LEE, E. *et al.*, 2013). As posições de metioninas iniciais foram então refinadas, e novos potenciais genes anotados utilizando alinhamentos múltiplos de sequências buscadas nos bancos de dados proteicos não-redundantes do NCBI de *Rozella allomycis*, *Encephalitozoon cuniculi* e *Daphnia magna* através de tBLASTn com *e-value* máximo de  $1E10^{-10}$  (CAMACHO *et al.*, 2009). A anotação funcional dos genes preditos foi realizada através da ferramenta EggNOG versão 4.5 (HUERTA-CEPAS *et al.*, 2016), que devido à crescente necessidade de um vocabulário padronizado de anotações funcionais, fornece acesso a mapeamentos nos bancos de dados Gene Ontology (GENE ONTOLOGY CONSORTIUM, 2015), KEGG (KANEHISA *et al.*, 2014), SMART (LETUNIC; DOERKS; BORK, 2015) e Pfam (FINN *et al.*, 2014).

#### 5.5 ANÁLISE DIFERENCIAL E CLUSTERIZAÇÃO DE GENES

Para identificar se existe conteúdo diferencial de genes relacionados à reprodução sexuada entre as linhagens, foi feita uma busca *all-against-all* por meio de BLASTp bidirecional com *e-value* máximo de  $1E10^{-25}$  para traçar o perfil

de cada isolado. A análise genômica de *clusters* ortólogos é um componente importante de estudos de genômica comparativa. A identificação de sobreposição entre *clusters* ortólogos contribui para elucidar a função e evolução das proteínas entre espécies diferentes ou dentro de espécies. Para comparação entre os três genomas, foi utilizado o OrthoVenn (WANG, Y. *et al.*, 2015) (*e-value* máximo de  $1E10^{-25}$ , *inflation value* de 2.0) uma plataforma *online* utilizada para comparações globais do proteoma predito de um genoma e visualização de conjuntos de sequências ortólogas. O OrthoVenn permite a comparação de *clusters* ortólogos com organismos-modelo e também suporta o *upload* de sequências de proteínas personalizadas de espécies definidas pelo usuário. A análise utilizando a plataforma permitiu a identificação de *clusters* ortólogos compartilhados entre os três genomas de *Hamiltosporidium* além de *clusters* de genes de cópia única.



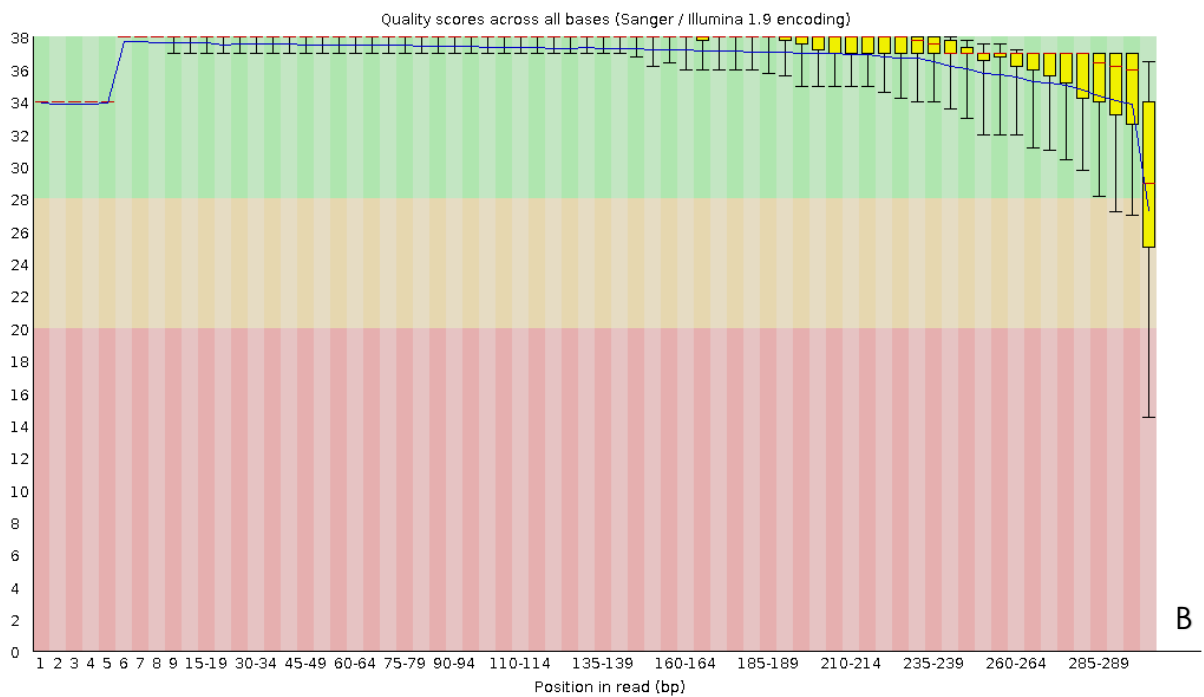
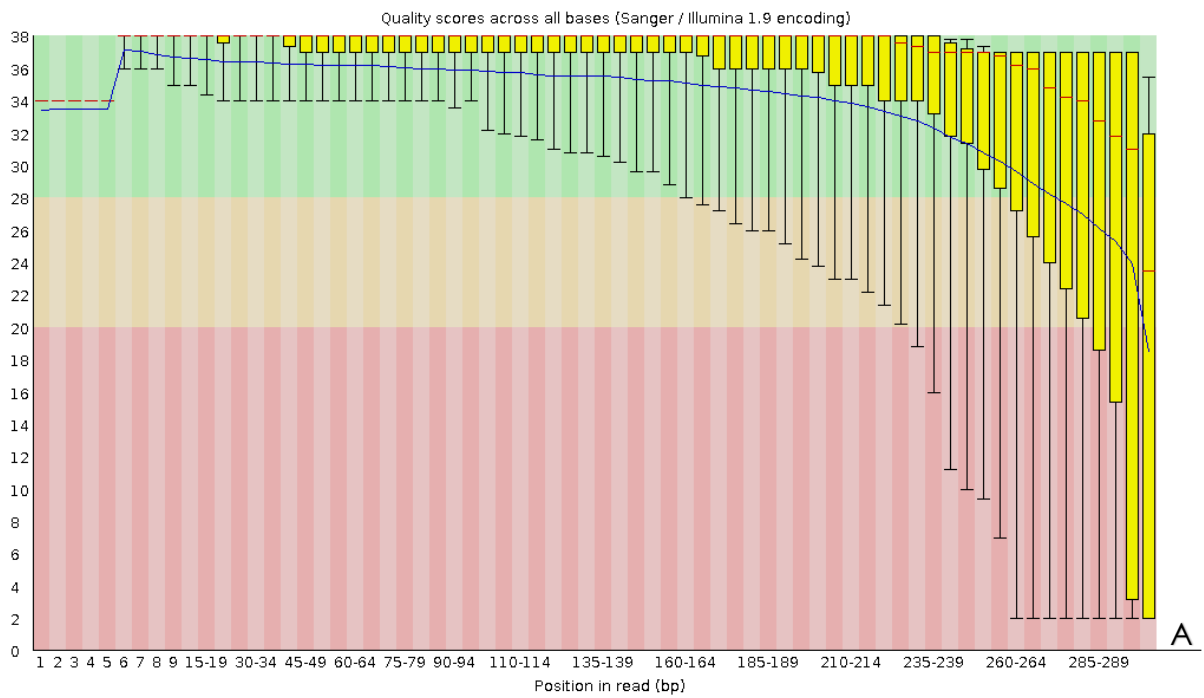
## 6 RESULTADOS E DISCUSSÃO

### 6.1 CONTROLE DE QUALIDADE DOS DADOS DE SEQUENCIAMENTO

Para exemplificar, será feita uma breve descrição dos resultados obtidos com a linhagem de *H. tvaerminnensis* (FI-OER-3-3). Foram sequenciadas 21.6 Gb, distribuídas em 32,807,340 leituras (*reads*) do tipo *paired end*. No controle de qualidade, o volume total de sequências foi reduzido a 15.4 Gb, distribuídas em 26,517,692 leituras (Tabela 3). A visualização das sequências permitiu a observação do impacto da retirada das sequências de baixa qualidade (*trimming*). As sequências que restaram apresentam-se de boa qualidade, com tamanho mínimo de 100 bases, com valores de phred acima de 30 (Figura 4). Como esperado, a retirada de bases de má qualidade e sequências curtas tornou a distribuição mais homogênea para todas as posições das leituras. Apesar da diminuição do volume de dados ocasionada pelo controle de qualidade, este conjunto representa uma aferição mais robusta das leituras.

Biblioteca	Nº total de leituras		Nº total de bases sequenciadas	
	Brutas	Filtradas	Brutas	Filtradas
<b>FI-OER-3-3</b>	32,807,340	26,517,692	21.6 Gb	15.4 Gb
<b>IL-BN-2</b>	16,444,616	12,165,380	10.7 Gb	5.9 Gb
<b>BE-OM-2</b>	14,195,298	10,671,054	9.2 Gb	5.2 Gb

Tabela 3 Métricas de sequenciamento das três linhagens de *Hamiltosporidium*.



**Figura 4** Visualização da qualidade das bases da biblioteca de *H. tvaerminnensis* (FI-OER-3-3) antes (A) e depois (B) dos controles de qualidade. O eixo x indica a posição das bases nas leituras e o eixo y indica os valores de qualidade phred.

## 6.2 AVALIAÇÃO DOS DRAFT ASSEMBLIES

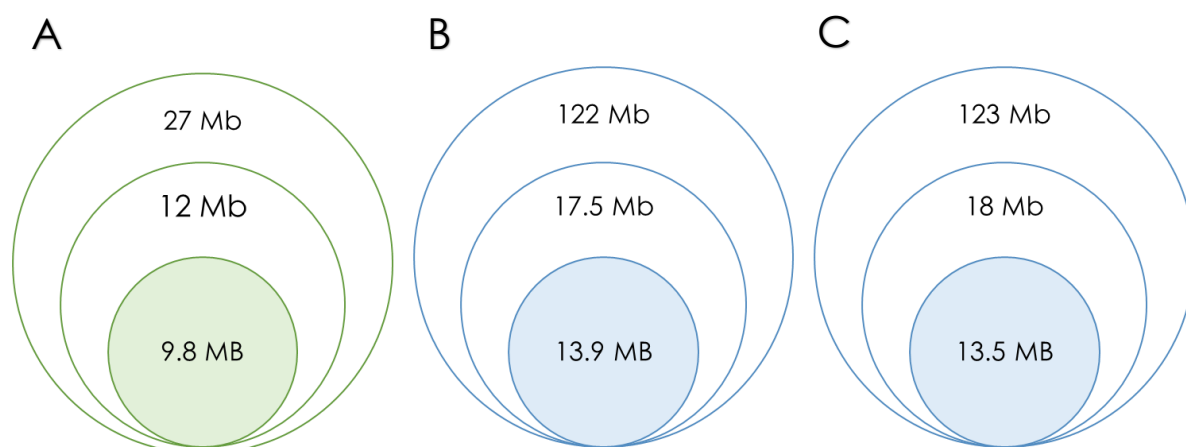
A montagem dos genomas de *Hamiltosporidium* foi realizada através dos montadores SPAdes e Ray, uma vez que o programa dipSPAdes não conseguiu convergir resultados mesmo após 2 tentativas. Foram realizadas diferentes montagens visando encontrar o método adequado para o conjunto de dados (Tabela 4). Os critérios utilizados para avaliação foram o maior valor de N50, o menor número de *contigs* e o tamanho final do genoma. Dessa forma, espera-se um genoma menos fragmentado e mais completo. Com isso a montagem do programa SPAdes que apresentou o maior valor de N50 - 6.2 Kb, 119 Mb de genoma montado e menor número de *contigs* foi escolhida com base na montagem do genoma de *H. magnivora* (BE-OM-2).

Estatísticas	Montagem	
	Ray_k51	SPAdes_trim
# contigs	52052	36917
# contigs (>= 0 bp)	77188	46296
# contigs (>= 1000 bp)	26956	24219
# contigs (>= 5000 bp)	2027	6457
# contigs (>= 10000 bp)	635	2264
# contigs (>= 25000 bp)	49	292
# contigs (>= 50000 bp)	3	43
Maior contig	58507	125691
Tamanho total	83572294	118656002
Tamanho total (>= 0 bp)	91996606	121894824
Tamanho total (>= 1000 bp)	65532055	109899299
Tamanho total (>= 5000 bp)	19532008	68434723
Tamanho total (>= 10000 bp)	10107519	39322477
Tamanho total (>= 25000 bp)	1574736	10914689
Tamanho total (>= 50000 bp)	160995	2690110
N50	1979	6228
N75	1089	2745
L50	9750	4825
L75	24222	12003
GC (%)	36.61	39.05

**Tabela 4** Comparação de métricas de *H. magnivora* (BE-OM-2) obtidas a partir da montagem com os programas Ray e SPAdes.

### 6.2.1 FILTRAGEM DE CONTAMINAÇÕES

Como o tamanho esperado dos genomas do gênero *Hamiltosporidium* é de aproximadamente 24 Mb, buscou-se refinar as montagens para eliminar *contigs* contaminantes. Após o primeiro ciclo de filtragem baseado na predição de ORFs de no mínimo 150 aminoácidos e buscas por BLASTp em conjuntos de dados com proteínas de representantes de bactérias, microsporídeos e do gênero *Daphnia* e agrupamento (*binning*) por grupos taxonômicos, foi possível reduzir consideravelmente a quantidade de *contigs* de bactéria e dáfnia, resultando em um genoma mais purificado (Figura 5, Tabela 5). Após anotação prévia de dos genomas utilizando a *pipeline* MAKER para predição de genes e curadoria manual com predição de novos genes, foi feita uma busca de BLAST taxonomizado para caracterizar as anotações. Nessa fase observou-se ainda bastante contaminação proveniente do hospedeiro. Possivelmente tais genes passaram pela primeira filtragem pois possuem íntrons e, com isso, não obtiveram regiões representativas de suas sequências preditas a partir da simples predição de ORFs. Nesse contexto, um novo ciclo de filtragem foi aplicado, agora utilizando as sequências anotadas a partir de predição gênica e curadoria manual (Tabela 6). As proteínas preditas foram submetidas ao BLAST taxonomizado. *Contigs* com pelo menos um *top hit* de microsporídeo e com conteúdo GC até  $\pm 5\%$  da média dos *contigs* confirmados de microsporídeos foram mantidos nas análises, os demais, descartados.



**Figura 5 Tamanho das montagens antes e depois dos ciclos de filtragem.** De fora para dentro: I - Primeira montagem a partir de todas as reads já filtradas por qualidade e tamanho; II – Montagem resultante após primeiro ciclo de filtragem de contaminantes; III – montagem final para análises subsequentes. A: FI-OER-3-3 B: BE-OM-2 C: IL-BN-2

Linhagem	<i>Hamiltosporidium tvaerminnensis</i>		<i>Hamiltosporidium magnivora</i>	
	FI-OER-3-3 I	FI-OER-3-3 II	BE-OM-2-STRB	IL-BN-2
<b>Genoma montado Mb</b>	13.3	11.87	17.42	18.03
<b>Número de contigs</b>	41,786	3,540	1,392	2,732
<b>Contig N50</b>	442	6,725	23,203	14,395
<b>Contig L50</b>	7,703	385	228	392
<b>Cobertura do genoma(%)</b>	53.3†	47.48†	69.64†	69.88†
<b>Conteúdo G+C (%)</b>	26	27.29	32.78	32.78
<b>Número de genes preditos</b>	2,174#	2.720#	3,650#	3.150#

**Tabela 5 Comparação de montagens após primeiro ciclo de filtragem.** †Baseado no tamanho estimado de 25 Mb. #Inclui ORFs de no mínimo 150 aminoácidos identificados no genoma. FI-OER-3-3 I: 2009 draft genome.

	<i>Hamiltosporidium tvaerminnensis</i>		<i>Hamiltosporidium magnivora</i>	
<b>Linhagem</b>	FI-OER-3-3 I	FI-OER-3-3 II	BE-OM-2-STRB	IL-BN-2
<b>Genoma montado Mb</b>	13.3	9.8	13.9	13.5
<b>Número de contigs</b>	41,786	1,171	746	900
<b>Contig N50</b>	442	17,299	28,476	22,994
<b>Contig L50</b>	7,703	145	161	187
<b>Cobertura do genoma(%)</b>	53.3†	39.2†	55.6†	54†
<b>G+C content (%)</b>	26	26.15	26.22	26.67
<b>Número de genes preditos</b>	2,174#	2,075	2,224	2.248

**Tabela 6 Comparação de montagens após segundo ciclo de filtragem.** †Baseado no tamanho estimado de 25 Mb. #Inclui ORFs de no mínimo 200 aminoácidos identificados no genoma, demais são provenientes de predição gênica utilizando a pipeline MAKER e adição manual de features. FI-OER-3-3 I: 2009 draft genome (CORRADI et al., 2009).

### 6.3 ANOTAÇÃO DOS GENOMAS DE *HAMILTOSPORIDIUM*

O *pipeline* do MAKER realizou a predição de genes codificadores de proteínas a partir da combinação dos resultados utilizando como conjunto de treinamento modelos construídos a partir de genes de *Rozella alomycis* e de *Encephalitozoon cuniculi*. Estas predições gênicas após serem inseridas à plataforma WebApollo, foram confirmadas e novos genes inseridos a partir de buscas tBLASTn de conjuntos de dados de proteínas de *Rozella alomycis*, *Daphnia magna* e *Encephalitozoon cuniculi* contra os genomas montados. A investigação quanto à ortologia nos bancos de dados de eggNOG retornou resultados descritos na Tabela 7. Observa-se que de metade a 1/5 das proteínas não obtiveram *hit* algum com BLASTp e o número de sequências anotadas funcionalmente é de aproximadamente 50% em todas as linhagens. No entanto, essas porcentagens baixas não significam necessariamente que estes genes não sejam homólogos, afinal eles podem apenas refletir a alta taxa de divergência de sequência que ocorre nos microsporídeos (NAKJANG, S. *et al.*, 2013). As proteínas sem *hit* ou anotação podem ser tão diferentes das de outros grupos, próximos ou não, que não alinham com qualidade mínima.

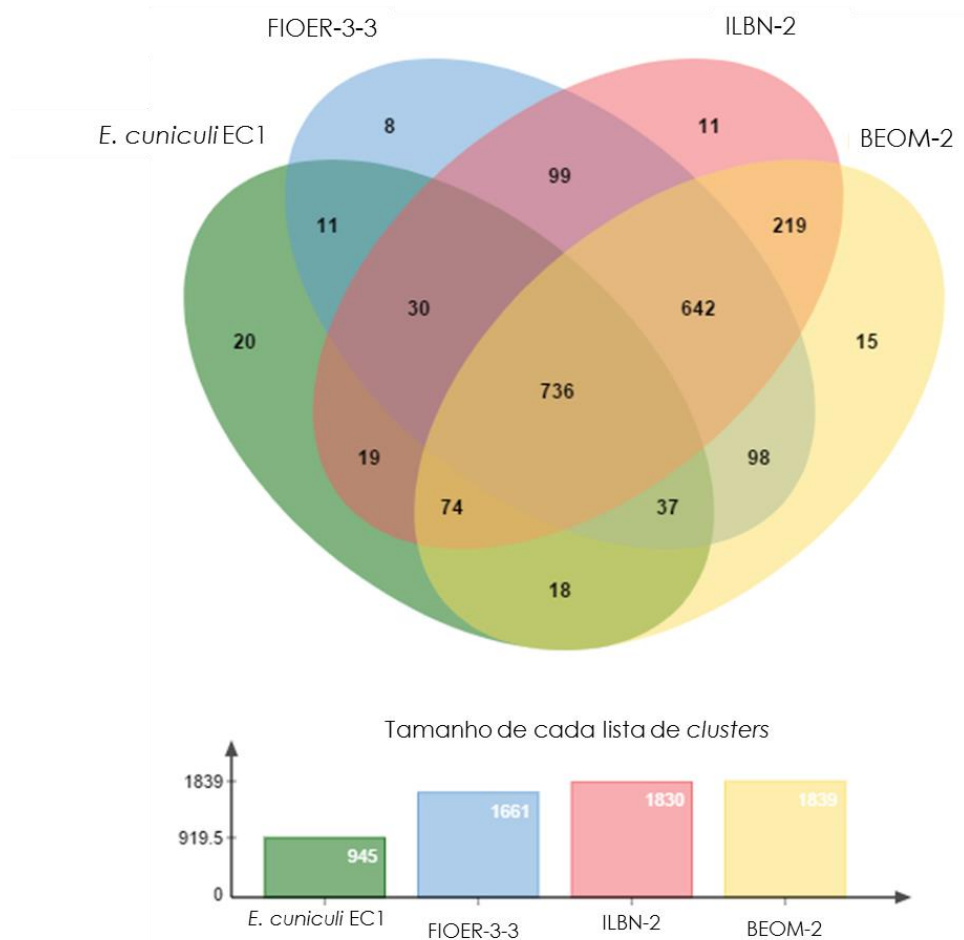
	FI-OER-3-3	BE-OM-2	IL-BN-2
<b>Genes codificadores de proteínas</b>	2,075	2,224	2.248
<b>Com <i>hit</i> em Blast taxonomizado</b>	1,075 (52%)	1,598 (77 %)	1627 (73%)
<b>Com anotação EggNOG</b>	884 (43%)	965 (43 %)	1042 (46%)

**Tabela 7** Resultados da anotação após curadoria manual e busca em BLASTp taxonomizado e busca de anotação funcional no banco de dados eggNOG.



## 6.4 CARACTERIZAÇÃO GENÔMICA

Embora os genomas montados apresentem tamanhos inferiores ao esperado para o gênero *Hamiltosporidium*, observa-se que o número de genes codificantes de proteína preditos é semelhante ao da montagem prévia de 2009, ao passo que os valores de N50 e número de *contigs* demonstram que essa montagem é de maior qualidade devido à baixa fragmentação (CORRADI *et al.*, 2009) (Tabela 6). Além disso, o conteúdo G/C de aproximadamente 26% se conserva em todas as montagens. Na análise de clusterização de ortólogos executada pela plataforma *online* OrthoVenn foram comparados grupos ortólogos compartilhados ou exclusivos nas três linhagens aqui montadas e no microsporídeo modelo *Encephalitozoon cuniculi* linhagem EC1 (Código de acesso NCBI: PRJNA63135). Observa-se que a maioria dos *clusters* agrupa entre os integrantes do gênero *Hamiltosporidium*. É possível perceber, no entanto, um considerável conjunto compartilhado entre as quatro linhagens (736 *clusters*), exemplificando um conjunto de genes *core* de microsporídeos (APENDICE 10). Estes foram encontrados codificando para homólogos de produtos que são essenciais para a sobrevivência de células em organismos modelo, incluindo produtos envolvidos em atividades essenciais para a síntese de proteínas (transcrição e tradução, dobramento de proteínas e transporte) e outros envolvidos na propagação (replicação de DNA, reparação e controle do ciclo celular). Os homólogos de genes adicionais, que são também partilhados pela maioria das microsporídeos, parecem ser aqueles que funcionam no metabolismo *core* de carbono (glicólise e as vias de trealose e pentose fosfato) e os necessários para a formação de esporos e o aparelho de invasão (O-manosila, proteínas da parede do esporo e quitina sintase).



**Figura 6 Diagrama de Venn de clusters ortólogos compartilhadas e exclusivas a partir de análise de clusters com OrthoVenn utilizando o microsporídeo modelo *Encephalitozoon cuniculi* linhagem EC1 para comparação.** Número de proteínas utilizadas na análise encontram-se na Tabela 8.

Os genomas de microsporídeos que possuem dados disponíveis variam em tamanho por uma ordem de magnitude (2,3 a 51.3 Mb) (CORRADI *et al.*, 2009; DESJARDINS *et al.*, 2015), mas essas grandes diferenças parecem refletir diferenças na densidade gênica causadas pela presença variável de elementos transponíveis e sequências repetitivas, presentes em um processo evolutivo dinâmico de ganho, através de transferência gênica horizontal, e perda de genes desnecessários. (HEINZ, Eva *et al.*, 2012; PEYRETAILLADE, Eric *et al.*, 2012). O genoma do microsporídeo *H. tvaerminnensis* possui um tamanho estimado de 25 Mb e deve possuir tal tamanho pelo mesmo motivo, afinal, quando analisamos a quantidade de genes codificantes de proteínas, não são observadas grandes disparidades até mesmo com o modelo de redução genômica *E. cuniculi* (Tabela 8). Embora o tamanho estimado do genoma de *H. tvaerminnensis* (25

Mb) seja mais de 10 vezes maior que o de *E. cuniculi* (2.24 MB), o número de genes codificantes de proteínas preditos é apenas 1.2 vezes maior.

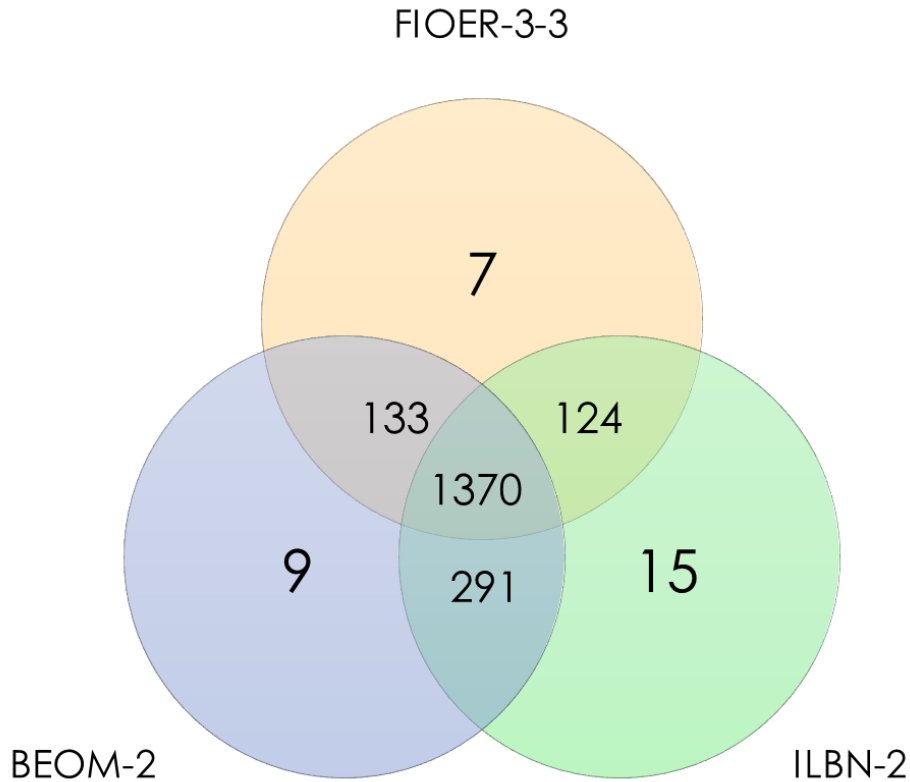
<b>Espécie</b>	<b>Proteínas</b>	<b>Clusters</b>	<b>Singletons</b>
<b><i>E. cuniculi</i> EC1</b>	1847	945	842
<b><i>H. tvaerminensis</i> FI-OER-3-3</b>	2075 (+1.12x)	1661	375
<b><i>H. magnivora</i> ILBN-2</b>	2248 (+1.21x)	1830	341
<b><i>H. magnivora</i> BEOM-2</b>	2224 (+1.20x)	1839	343

**Tabela 8 Número de proteínas, clusters e singletons utilizados na análise de clusters ortólogos.** O proteoma de *E. cuniculi* corresponde à linhagem que possui genoma completo disponível EC1. Entre parênteses ao lado do número de proteínas do proteoma predito de cada linhagem de *Hamilosporidium* encontra-se a variação em relação ao total de proteínas anotadas de *E. cuniculi*.

É possível, no entanto, que tal aproximação no número de proteínas preditas seja devido ao fato de que a montagem da espécie de *Encephalitozoon* está completa, enquanto as aqui apresentadas são *drafts*, existindo a possibilidade de outras proteínas estarem em regiões não acessadas nestas montagens. A montagem de *H. tvaerminensis* OER-3-3 de 2009 (CORRADI *et al.*, 2009), proveniente de sequenciamento de leituras com 30 pb de comprimento, resultou na predição de 2,174 genes codificantes de proteína, enquanto a montagem atual resultou em 2,075 (Tabela 6). Vale ressaltar que o número predito anteriormente considera toda ORF predita com no mínimo 200 AA de comprimento, enquanto a predição gênica neste trabalho foi feita com maior precisão, através da *pipeline* MAKER e busca ativa por indicativos de similaridade de sequência. Espera-se que o tamanho do genoma da espécie irmã, *H. magnivora*, seja semelhante – mesmo que não tenha sido estimado até agora – e talvez o tamanho das montagens das três linhagens representem a quantidade de genoma com sequências distinguíveis pelos *softwares* de montagem, enquanto as regiões faltantes compreendam regiões repetitivas. Sabe-se se que o conteúdo repetitivo de um genoma “quebra” montagens e

tende a subestimar seu real tamanho (TREANGEN; SALZBERG, 2012). Uma possível solução seria a utilização de outro método de sequenciamento, que utilize leituras maiores, que exigem maior quantidade inicial de amostra, o que pode também pode ser um desafio devido à dificuldade de se extrair DNA de *Hamiltosporidium*. Este desafio é comum em parte devido às dificuldades de estudar parasitas intracelulares obrigatórios que geralmente não podem ser mantidos em cultura de tecidos e para os quais não existem ferramentas bem estabelecidas para manipulação genética (PALDI *et al.*, 2010).

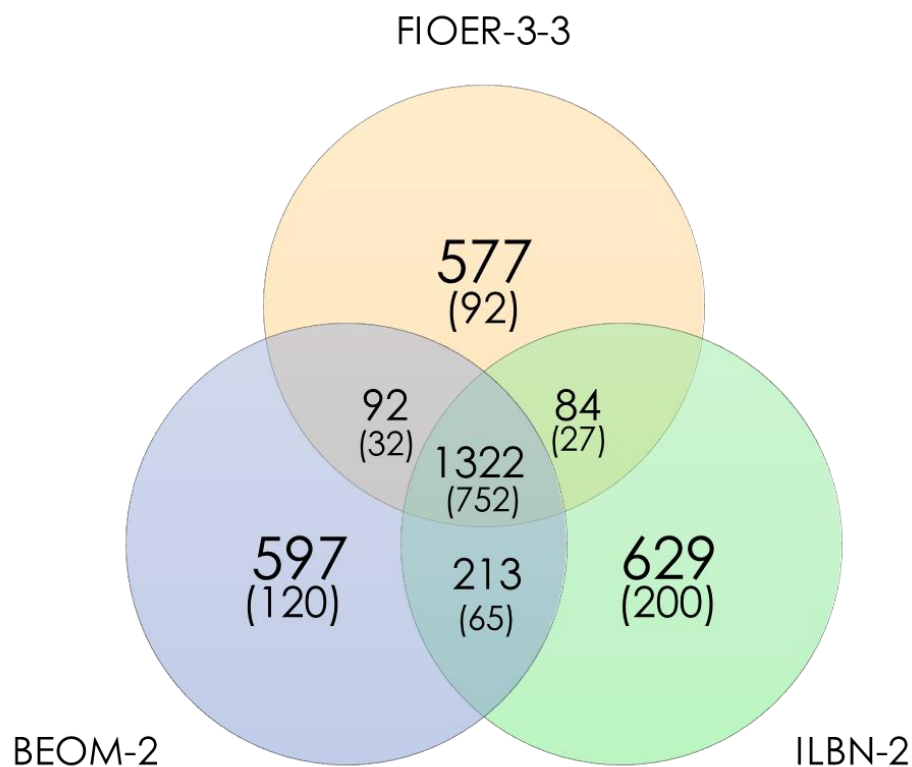
Para investigar o horizonte genômico dos genes relacionados à reprodução sexuada, foi preciso avaliar comparativamente o conteúdo genômico das três linhagens de *Hamiltosporidium*. Primeiramente, foram feitas uma comparação e uma anotação de *clusters* ortólogos entre os genomas montados (Figura 07). Observou-se, como esperado, um maior compartilhamento de *clusters* entre os representantes sexuados, além de um enriquecimento de sequências da enzima transposase no isolado de *H. magnivora* (ILBN-2) (Apêndice 11).



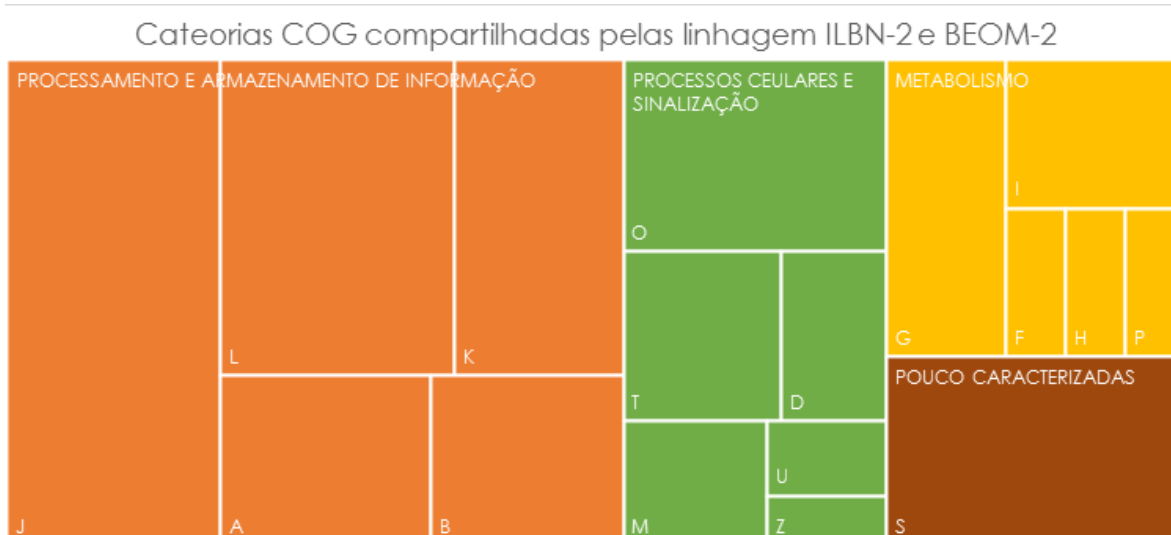
**Figura 7 Diagrama de Venn de clusters ortólogos compartilhadas e exclusivas a partir de análise de clusters com OrthoVenn.**

Para melhor caracterizar os perfis, foi feita uma análise todos-contra-todos de BLASTp recíproco para limitar a análise a proteínas específicas que estavam presentes ou ausentes nos grupos, a fim de eliminar o viés de similaridade semântica de clusterização (WANG, Jing *et al.*, 2010). Nessa análise a quantidade de *singletons* é maior, porém, os genes considerados compartilhados possuem maior confiança (Figura 8). Os resultados demonstram que a maioria das categorias COG (Apêndice 12) das proteínas anotadas compartilhadas entre as linhagens de *Hamiltosporidium magnivora* (ILBN-2 e BEOM-2) correspondem à classe Processamento e Armazenamento de Informação, seguida pela classe Processos Celulares e Sinalização (Figura 9) (Apêndice 13). Entre as categorias em Processamento e Armazenamento de Informação as mais representadas são as classes [J] Tradução, estrutura ribossomal e biogênese, [K] Transcrição, [L] Replicação, recombinação e reparo, e [A] Processamento e modificação de RNA.

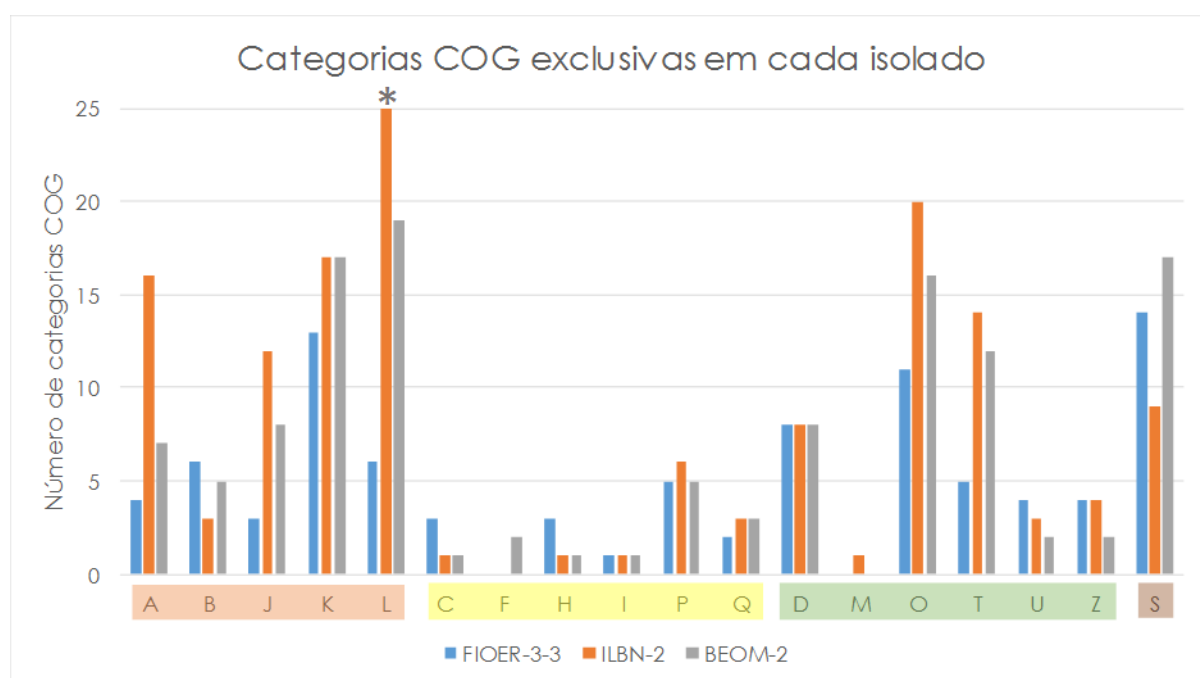
Observa-se em todos os dados comparativos um número maior de genes compartilhados entre os genomas de *H. magnivora*, além de um número maior de genes linhagem-específicos na linhagem ILBN-2. Com o objetivo de investigar melhor a discrepância, buscou-se comparar as categorias COG das sequências exclusivas de cada linhagem (Figura 10). Novamente, as linhagens sexuadas apresentam mais anotações classificadas como Processamento e Armazenamento de Informação, seguida pela classe Processos Celulares e Sinalização. Interessantemente, a linhagem ILBN-2 possui um total de 86 proteínas classificadas na sub-categoria [L], referente à replicação, principalmente devido à presença de várias cópias de genes de transposase de retrotransposon (Apêndice 11).



**Figura 8** Diagrama de Venn de proteínas compartilhadas e exclusivas a partir de BLAST bidirecional. (Número em parênteses representa fração anotada funcionalmente com EggNOG dos genes em cada categoria).



**Figura 9** Mapa das categorias COG compartilhadas entre as linhagens de *Hamiltosporidium magnivora* a partir da análise de BLAST recíproco e anotação funcional com EggNOG.



**Figura 10** Gráfico das categorias COG linhagem-específica entre as linhagens de *Hamiltosporidium* a partir da análise de BLAST recíproco e anotação funcional com EggNOG. Barras sombreando códigos COG: Laranja – Processamento e armazenamento da informação, Amarela – Metabolismo, Verde – Processos celulares e sinalização, Marrom – Pouco caracterizadas. \* Quebra de escala: linhagem ILBN-2 possui 86 incidências da sub-categoria [L] Replicação.

#### 6.4 TRANSFERÊNCIA HORIZONTAL GÊNICA

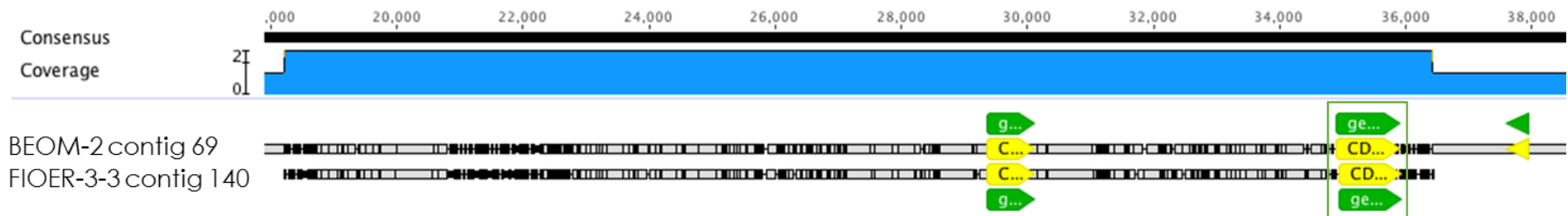
A Tabela 9 apresenta uma visão geral dos possíveis eventos de transferência horizontal gênica (HGT – *horizontal gene transfer*) do hospedeiro *Daphnia magna* para os organismos do gênero *Hamiltosporidium*. Nesse contexto, foram encontradas proteínas com alta similaridade de sequência com proteínas de *Daphnia magna*, inseridas nos genomas. Foram encontradas 9 possíveis HGT nas linhagens de *H. magnivora*, as sexuadas, enquanto na linhagem assexuada, *H. tvaerminnensis*, foi encontrada apenas uma. Todas encontram-se em locus independentes e não sobreponíveis entre si, exceto as sequências BEOM2\_0069p05290 e FIOER33\_0140p06400. Estas além de apresentarem o mesmo *hit* contra o banco de dados não redundante do NCBI, encontram-se em regiões genômicas homólogas (Figura 11).



	DESCRIPTION	MAX SCORE	TOTAL SCORE	QUERY COVER	E VALUE	IDENT	ACCESSION
<b>BEOM2_0069P05290</b>	Ribonucleoside-diphosphate reductase subunit M2 B [Daphnia magna]	500	500	99%	2.00E-175	72%	JAN68171.1
<b>BEOM2_0289P14110</b>	hypothetical protein [Daphnia magna]	405	711	93%	3.00E-134	98%	JAN57812.1
<b>BEOM2_0289P14120</b>	Uncharacterized protein APZ42_020483 [Daphnia magna] {bromodomain containing}	2983	2983	100%	0	97%	KZS14284.1
<b>BEOM2_0297P14390</b>	Replication factor C subunit 4 [Daphnia magna]	200	200	98%	4.00E-65	100%	KZS02750.1
<b>BEOM2_0384P16590</b>	Uncharacterized protein APZ42_023175 [Daphnia magna]	694	694	49%	0	90%	KZS11983.1
<b>BEOM2_0384P16600</b>	Transposase [Daphnia magna]	466	466	84%	5.00E-163	98%	JAN72624.1
<b>BEOM2_0384P16610</b>	Uncharacterized protein APZ42_025355 [Daphnia magna]	491	491	99%	4.00E-174	92%	KZS10248.1
<b>BEOM2_0384P16620</b>	Gdp mannose-4,6-dehydratase [Daphnia magna]	365	365	99%	4.00E-121	67%	JAN81842.1
<b>BEOM2_0630P20900</b>	PREDICTED: uncharacterized protein LOC104607033 [Nelumbo nucifera]	47	47	55%	5.00E-04	40%	XP_010270824.1
<b>ILBN2_0299P12630</b>	Uncharacterized protein APZ42_017517 [Daphnia magna]	447	447	100%	7.00E-159	98%	KZS16892.1
<b>ILBN2_0299P12640</b>	Phospholipase B 2-like protein [Daphnia magna]	1163	1163	100%	0	97%	KZS16891.1
<b>ILBN2_0305P12800</b>	Steroid receptor RNA activator 1 [Daphnia magna]	291	291	100%	5.00E-99	99%	KZS04493.1
<b>ILBN2_0555P17750</b>	Uncharacterized protein APZ42_027940 [Daphnia magna] {Ankyrin like}	651	651	97%	0	91%	KZS08163.1
<b>ILBN2_0555P17770</b>	Polyprotein of retroviral origin [Daphnia magna]	654	654	71%	0	73%	JAN82185.1
<b>ILBN2_0625P18900</b>	Retrovirus-related Pol polyprotein from transposon [Daphnia magna]	124	124	75%	1.00E-32	57%	JAN77544.1
<b>ILBN2_0680P19620</b>	Uncharacterized protein APZ42_010964 [Daphnia magna]	305	398	51%	5.00E-93	92%	KZS22081.1
<b>ILBN2_0820P21590</b>	Polyprotein of retroviral origin [Daphnia magna]	531	531	100%	0	90%	JAN88971.1
<b>ILBN2_0836P21770</b>	Uncharacterized protein APZ42_031262 [Daphnia magna]	344	344	96%	4.00E-116	91%	KZS05528.1

<b>FIOER33_0140P06400</b>	hypothetical protein DAPPUDRAFT_306462 [Daphnia pulex] {Ribonucleoside-diphosphate reductase subunit M2 B [Daphnia magna]}	501	501	99%	4.00E-176	73%	EFX87660.1
---------------------------	--	-----	-----	-----	-----------	-----	------------

**Tabela 9 Possíveis eventos de transferência gênica horizontal de *Daphnia magna* para *Hamiltosporidium*, resultados obtidos a partir de BLASTp taxonomizado e buscas contra as outras linhagens onde não foram encontradas.**



**Figura 11 Sobreposição dos contigs de BEOM-2 e FIOER-3-3 onde foi encontrada a proteína RNR, ribonucleosídeo difosfato redutase, de *Daphnia magna* (marcada na caixa verde).** Embora o contig 140 de *Hamiltosporidium tvaerminnensis* (FIOER-3-3) seja menor que o de *H. magnivora* (BEOM-2), este alinha por completo. Há inclusive a presença de uma proteína de microsporídeo à montante de ambas as sequências.

## 6.5 GENES E SEXO

Os genes *Mating Type* presentes no locus MAT não foram encontrados em nenhum dos genomas amostrados. Os genes específicos de meiose estão presentes na maioria das microsporídeos (*Spo11*, *Rad51*, *Mdn1*), portanto algumas espécies neste grupo são teoricamente capazes de realizar meiose convencional. Para investigar a relação entre o conteúdo de genes de meiose e o modo de reprodução, o conjunto de genes core de meiose proposto por Desjardins *et al.* foram buscados nas montagens para comparação com outros microsporídeos e a levedura *S. cerevisiae* (DESJARDINS *et al.*, 2015) (Tabela 10). Todas as montagens apresentam o gene *Spo11* presente e embora apresentem um padrão semelhante ao de outros microsporídeos, observa-se a ausência dos genes *HOP1* e *MSH6* na linhagem assexuada, de *H. tvaerminnensis* (FIOER-3-3). *HOP1* é uma proteína ligadora de DNA que participa da formação do complexo sinaptonêmico durante a meiose e é de extrema importância para a realização do *crossing over*. Já a proteína *MSH6* é uma proteína de reparo de DNA associada com a religação das fitas de DNA após a quebra, seja durante a meiose ou durante outros eventos de recombinação que envolva quebra da dupla fita.

Gene	<i>S. cerevisiae</i>	<i>A. algerae</i>	<i>E. aedis</i>	<i>E. cuniculi</i>	<i>N. ERTm1</i>	<i>parisii</i>	<i>T. hominis</i>	<i>V. culicis</i>	<i>H. tvaerminnensis</i> FIOER-3-3	<i>H. magnivora</i> ILBN-2	<i>H. magnivora</i> BEOM-2
SPO11	YHL022C	H312_01085; H312_01089	EDEG_02480	ECU04_1110	NEPG_01692	THOM_2627	VCUG_01483	FIOER33_0548p13760	ILBN2_0023p01770	BEOM2_0166p09990	
HOP1	YIL072W	H312_02178	EDEG_03375	ECU02_0890	-	THOM_2996; THOM_2997	VCUG_02610	-	ILBN2_0772p20950	BEOM2_0350p15670	
HOP2	YGL033W	H312_00743	EDEG_02519	ECU01_0960	-	THOM_1118	VCUG_02261	FIOER33_0472p12710	ILBN2_0570p18000	BEOM2_0054p04220	
MND1	YGL183C	H312_01474	EDEG_01364; EDEG_03123	ECU10_1600	NEPG_00016	THOM_0475	VCUG_01234	FIOER33_0803p16850	ILBN2_0692p19830	BEOM2_0005p00640	
REC8	YPR007C	-	-	-	-	-	-	-	-	-	
RAD21/MCD1	YDL003W	-	-	-	NEPG_00917	-	-	-	-	-	
DMC1	YER179W	H312_01180	-	-	-	THOM_1138	VCUG_02488	FIOER33_0055p03270	ILBN2_0119p06970	BEOM2_0246p12710	
RAD51	YER095W	H312_03150	EDEG_00684	ECU11_0820	NEPG_00390	THOM_2061	VCUG_01109	FIOER33_0795p16730	ILBN2_0174p08990	BEOM2_0016p01530	
MSH4	YFL003C	H312_01012	EDEG_02403	-	-	THOM_1300	VCUG_02604	FIOER33_0649p14990	ILBN2_0707p20010	BEOM2_0186p10740	
MSH5	YDL154W	H312_03128	EDEG_02514	-	-	THOM_1865	VCUG_00386	FIOER33_0899p17910	ILBN2_0787p21150	BEOM2_0535p19520	
MSH2	YOL090W	H312_00255	EDEG_04058	ECU03_0540	NEPG_01289	THOM_2653; THOM_2655	VCUG_02338	FIOER33_0384p11330	ILBN2_0498p16850	BEOM2_0350p15680	
MSH6	YDR097C	H312_01018	EDEG_02460	ECU10_0710	NEPG_02026	THOM_0558	VCUG_00115	-	ILBN2_0841p21820	BEOM2_0053p04160	

**Tabela 10 Grupo de genes core de meiose homólogos distribuídos em microsporídeos e em levedura.**

As diferenças principais entre as espécies *Hamiltosporidium tvaerminnensis* e *Hamiltosporidium magnivora* encontradas nesse trabalho são as seguintes:

- O conteúdo de genes core de meiose na linhagem assexuada (*Hamiltosporidium tvaerminnensis*, OER-3-3) apresenta duas proteínas chave (HOP1 e MSH6) faltantes em comparação ao conteúdo das linhagens sexuadas (*Hamiltosporidium magnivora*, BEOM-2 e ILBN-2);

- Existem mais proteínas de elementos móveis, em especial transposases e RNA polimerases, compartilhadas exclusivamente nas linhagens sexuadas, ou presentes como linhagem específica (foram encontradas 86 cópias de gene de transposase de retrotransposon presentes apenas na linhagem ILBN-2);

- Foram encontradas 9 proteínas potencialmente provenientes de transferência horizontal do hospedeiro metazoário *Daphnia magna* nas linhagens sexuadas, enquanto na linhagem assexuada foi encontrada apenas 1.

Muitas espécies eucarióticas podem ter um ciclo de vida predominantemente assexuado com raras ocorrências de sexo. Espécies que realizam reprodução sexuada requerem genes meióticos funcionais, enquanto que em espécies que permanecem assexuadas por um longo prazo não haverá pressão seletiva para reter tais genes. Assim, genes meióticos podem estar presentes como pseudogenes ou completamente ausentes do genoma de assexuados antigos (CARR; LEADBEATER; BALDAUF, 2010). Essa ausência dos dois genes core de meiose na linhagem assexuada de *H. tvaerminnensis*, se confirmada, pode potencialmente justificar sua ausência de recombinação.

O modo de vida intracelular parece ter provocado muitas mudanças redutoras nos genomas da microsporídeos, mas esta interação também oferece oportunidades para que os patógenos adquiram novos genes. Especificamente, os parasitas intracelulares são mais expostos ao contato físico com o material genético do hospedeiro ou de organismos coinfectantes. Selman e Corradi levantaram a hipótese de que essa proximidade tem permitido uma série de transferências genéticas horizontais em microsporídeos, algumas das quais tiveram tremendo impacto na biologia de todo o grupo

(SELMAN; CORRADI, 2011). O evento de transferência horizontal identificado neste trabalho presente em dois genomas (FIOER-3-3 e BEOM-2) indica que a troca entre microsporídeos e dáfias tem ocorrido há muito tempo, anteriormente à divergência das duas espécies de *Hamiltosporidium*. Além disso, observa-se uma proliferação de elementos transponíveis (TE – *transposable element*) nas linhagens sexuadas (ILBN-2 e BEOM-2), associadas a um maior número de possíveis eventos de transferência horizontal a partir do hospedeiro. Estes genes móveis por muito tempo foram ignorados em microsporídeos, mas agora são considerados bastante abundantes e importantes para a evolução do genoma desses parasitas (PARISOT *et al.*, 2014). Análises filogenéticas recentes destes TEs revelaram que muitos podem ter sido trocados entre microsporídeos e uma variedade de linhagens de metazoários, particularmente linhagens de artrópodes (GUO, X. *et al.*, 2014; PARISOT *et al.*, 2014). Ademais, existem evidências de que TEs podem ser transferidos do parasita para outros metazoários, o que significa que os microsporídeos também poderiam atuar como vetores de informações genéticas em ecossistemas terrestres e aquáticos em todo o mundo (PARISOT *et al.*, 2014).

Os mecanismos de inserção de elementos móveis nos genomas de seus hospedeiros são ainda bastante estudados. Recentemente se propôs que proteínas de reparo do DNA do hospedeiro participam na retrotransposição de elementos móveis do tipo LINE (YAMAGUCHI; KAJIKAWA; OKADA, 2015). O gene HTA1 foi encontrado junto dos genes compartilhados entre as linhagens sexuais ILBN-2 e BEOM-2 (APENDICE 10). Este gene codifica uma proteína que é componente do nucleossomo e desempenha um papel central no reparo de quebra de DNA de fita dupla. Possivelmente, a capacidade de realizar recombinação nas linhagens de *H. magnivora* foi o que tornou possível a manutenção de tais elementos móveis. Dessa forma, é provável que a perda do sexo em *H. tvaerminnensis* tenha ocorrido concomitantemente com a inativação dos TEs, ao passo que a presença de genes de meiose e reparo em *H. magnivora*, associada à singamia contribuíram para a proliferação destes elementos nessas linhagens.

A análise de genomas de microsporídeos pode fornecer uma estrutura comparativa para entender como os microsporídeos tornaram-se tão bem-sucedidos como parasitas e ajudam a orientar, e talvez priorizar, o trabalho experimental. Desta forma, estas suposições permanecem em aberto e agora, com os genomas montados e anotados, cabe a investigação mais aprofundada das causas e consequências aqui levantadas.



## 7 CONCLUSÕES

A metodologia de sequenciamento Illumina HiSeq com leituras de 300 pb permitiu a obtenção de uma montagem mais contígua que a montagem disponível de *Hamiltosporidium tvaerminnensis* OER-3-3, com cobertura semelhante em relação ao número de genes codificantes de proteínas.

As montagens dos genomas de *Hamiltosporidium magnivora* representam a primeira exploração do conteúdo genômico desta espécie.

São evidentes indícios genômicos de reprodução sexuada nas linhagens de *H. magnivora* como a presença de um conjunto de genes core de meiose mais completo que o da espécie irmã e de elevada incidência de elementos móveis.

Foi observada também uma maior quantidade de possíveis eventos de transferência horizontal nas linhagens de *H. magnivora*, possivelmente devido ao conteúdo de elementos móveis.

## REFERÊNCIAS

AKIYOSHI, D. E. *et al.* Genomic survey of the non-cultivable opportunistic human pathogen, *Enterocytozoon bieneusi*. **PLoS Pathogens**, 2009. v. 5, n. 1.

BABRAHAM BIOINFORMATICS. [S.l.], [s.d.]. Disponível em: <<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>>. Acesso em: 1º dez. 2016.

BANKEVICH, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal of Computational Biology**, 2012. v. 19, n. 5, p. 455–477.

BEN-AMI, F.; RIGAUD, T.; EBERT, D. The expression of virulence during double infections by different parasites with conflicting host exploitation and transmission strategies. **Journal of evolutionary biology**, jun. 2011. v. 24, n. 6, p. 1307–16. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21481055>>. Acesso em: 8 jan. 2017.

BIEGER, A.; EBERT, D. Expression of parasite virulence at different host population densities under natural conditions. **Oecologia**, 14 maio. 2009. v. 160, n. 2, p. 247–255. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19219457>>. Acesso em: 8 jan. 2017.

BIGLIARDI, E.; SACCHI, L. Cell biology and invasion of the microsporidia. **Microbes and Infection**, 2001. v. 3, n. 5, p. 373–379.

BIRKY, C. W. Giardia Sex? Yes, but how and how much? **Trends in Parasitology**, 2010. v. 26, n. 2, p. 70–74.

BIRKY, W. Giardia Sex? Yes, but how and how much? [s.d.].

CAMACHO, C. *et al.* BLAST+: architecture and applications. **BMC Bioinformatics**, 2009. v. 10, p. 421. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20003500>> <<http://download.springer.com/static/pdf/101/art%253A10.1186%252F1471-2105-10-421.pdf?originUrl=http%253A%252F%252Fbmcbioinformatics.biomedcentral.com%252Farticle%252F10.1186%252F1471-2105-10-421&token2=exp=14>>.

CANNING, E. U. *et al.* *Microgemma vivaresi* n. sp. (microsporidia, tetramicridae), infecting liver and skeletal muscle of sea scorpions, *Taurulus bubalis* (euphrasen 1786) (osteichthyes, cottidae), an inshore, littoral fish. **Journal of Eukaryotic Microbiology**, 2005. v. 52, n. 2, p. 123–131.

CANTAREL, B. L. *et al.* MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. **Genome Research**, 2008. v. 18, n. 1, p. 188–196.

CAPELLA-GUTIÉRREZ, S.; MARCET-HOUBEN, M.; GABALDÓN, T. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. **BMC biology**, 2012. v. 10, n. 1, p. 47. Disponível em: <<http://www.biomedcentral.com/1741-7007/10/47>>.

CARPENTER, M. L. *et al.* Nuclear inheritance and genetic exchange without meiosis in the binucleate parasite *Giardia intestinalis*. **Journal of Cell Science**, 2012. v. 125, n. Pt 10, p. 2523–2532. Disponível em: <<http://jcs.biologists.org/cgi/doi/10.1242/jcs.103879>%5Cn<http://jcs.biologists.org/content/125/10/2523.full.pdf>%5Cn<http://jcs.biologists.org/content/125/10/2523.long>%5Cn<http://www.ncbi.nlm.nih.gov/pubmed/22366460>>.

CARR, M.; LEADBEATER, B. S. C.; BALDAUF, S. L. Conserved meiotic genes point to sex in the choanoflagellates. **Journal of Eukaryotic Microbiology**, 2010. v. 57, n. 1, p. 56–62.

CAVALIER-SMITH, T. Origins of the machinery of recombination and sex Introduction: the fundamental difference in recombination biology of bacteria and eukaryotes. **Heredity**, 2002. v. 88, p. 125–141. Disponível em: <[www.nature.com/hdy](http://www.nature.com/hdy)>.

CERVANTES, M. D. *et al.* Selecting One of Several Mating Types through Gene Segment Joining and Deletion in *Tetrahymena thermophila*. **PLoS Biology**, 2013. v. 11, n. 3.

CORNMAN, R. S. *et al.* Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. **PLoS Pathogens**, 2009. v. 5, n. 6, p. e1000466.

CORRADI, N. *et al.* Draft genome sequence of the *Daphnia* pathogen *Octospora bayeri*: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. **Genome biology**, 2009. v. 10, n. 10, p. R106.

\_\_\_\_\_*et al.* The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. **Nature Communications**, 2010. v. 1, p. 77. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20865802>%5Cn<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4355639>%5Cn<http://www.ncbi.nlm.nih.gov/pubmed/20865802>>.

\_\_\_\_\_; SLAMOVITS, C. H. The intriguing nature of microsporidian genomes. **Briefings in Functional Genomics**, 2011. v. 10, n. 3, p. 115–124.

CUOMO, C. A. *et al.* Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. **Genome Research**, 2012. v. 22, n. 12, p. 2478–2488.

DESJARDINS, C. A. *et al.* Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. **Nature communications**, 2015. v. 6, n. May, p. 7121. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25968466>>.

DIDIER, E. S. Microsporidiosis: An emerging and opportunistic infection in humans and animals. **Acta Tropica**, 2005. v. 94, n. 1, p. 61–76.

DUJON, B. The yeast genome project: what did we learn? **Trends in genetics : TIG**, 1996. v. 12, n. 7, p. 263–270. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/8763498>>.

EBERT, D. Chapter 2 Introduction to Daphnia Biology. **Ecology, Epidemiology, and Parasitism in Daphnia**, 2005. p. 1–25.

EBERT, D. A Genome for the Environment. **Science**, 2011. v. 331, n. 6017, p. 539–540. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1202092>>.

\_\_\_\_\_; ZSCHOKKE-ROHRINGER, C. D.; CARIUS, H. J. Dose effects and density-dependent regulation of two microparasites of *Daphnia magna*. **Oecologia**, 1 fev. 2000. v. 122, n. 2, p. 200–209. Disponível em: <<http://link.springer.com/10.1007/PL00008847>>. Acesso em: 8 jan. 2017.

FINN, R. D. *et al.* **Pfam: The protein families database. Nucleic Acids Research.**

FRANZEN, C. **Microsporidia: How can they invade other cells? Trends in Parasitology.**

FRANZEN, C.; MÜLLER, A. **Molecular techniques for detection, species differentiation, and phylogenetic analysis of microsporidia. Clinical Microbiology Reviews.**

GENE ONTOLOGY CONSORTIUM. Gene Ontology Consortium: going forward. **Nucleic acids research**, 2015. v. 43, n. Database issue, p. D1049–56. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25428369>><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4383973>>.

GOFFEAU, A. *et al.* Life with 6000 Genes. **Science**, 1996. v. 274, n. October, p. 546–567.

GREEN, J. Parasites and epibionts of Cladocera. **The Transactions of the Zoological Society of London**, 1974. v. 32, n. 6, p. 417–515.

GUO, X. *et al.* Evidence of horizontal transfer of non-autonomous Lep1 Helitrons facilitated by host-parasite interactions. **Sci Rep**, 2014. v. 4, p. 5119. Disponível em:

<<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4038834/pdf/srep05119.pdf>>.

HAAG, K. L. *et al.* Cytological and molecular description of *Hamiltosporidium tvaerminnensis* gen. et sp. nov., a microsporidian parasite of *Daphnia magna*, and establishment of *Hamiltosporidium magnivora* comb. nov. **Parasitology**, 2011. v. 138, n. 4, p. 447–462.

HAAG, K. L. *et al.* Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. **Proceedings of the National Academy of Sciences**, 2014. v. 111, n. 43, p. 15480–15485. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.1410442111>>.

HAAG, K. L.; TRAUNECKER, E.; EBERT, D. Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. **Molecular Ecology**, 2013. v. 22, n. 2, p. 314–326.

HEINZ, E. *et al.* The Genome of the Obligate Intracellular Parasite *Trachipleistophora hominis*: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. **PLoS Pathogens**, 2012. v. 8, n. 10.

HEITMAN, J. **Evolution of eukaryotic microbial pathogens via covert sexual reproduction. Cell Host and Microbe.**

\_\_\_\_\_. **Microbial pathogens in the fungal kingdom. Fungal Biology Reviews.**

HUERTA-CEPAS, J. *et al.* EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic Acids Research**, 2016. v. 44, n. D1, p. D286–D293.

IRONSIDE, J. E. Multiple losses of sex within a single genus of Microsporidia. **BMC evolutionary biology**, 2007. v. 7, p. 48. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1853083&tool=pmcentrez&rendertype=abstract>>.

JACKSON, A. P. Preface: the evolution of parasite genomes and the origins of parasitism. **Parasitology**, 2015. v. 142, p. S1–S5.

JOHNSTON, D. A. *et al.* **Genomics and the biology of parasites. BioEssays.**

JOSEPH, J.; VEMUGANTI, G. K.; SHARMA, S. Microsporidia: emerging ocular pathogens. **Indian journal of medical microbiology**, 2005. v. 23, n. 2, p. 80–91. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15928435>>.

JOSHI, N.; FASS, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. **Available at <https://github.com/najoshi/sickle>**, 2011. p. 2011.

KANEHISA, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG - Release 72.1, December 1, 2014 . **Nucleic Acids Res**, 2014. v. 42, n. Database issue, p. D199-205.

KATINKA, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. **Nature**, 2001. v. 414, n. 6862, p. 450–453.

KEELING, P. Five questions about microsporidia. **PLoS Pathogens**, 2009. v. 5, n. 9.

KEELING, P. J.; FAST, N. M. Microsporidia: biology and evolution of highly reduced intracellular parasites. **Annu.Rev.Microbiol.**, 2002. v. 56, n. 0066–4227 (Print), p. 93–116. Disponível em: <c:%5CKarsten%5CPDFs%5CParasitologie-PDFs%5CParasit-2002%5CKeeling - Fast-Microsporidia- biology and evolution of highly reduced intracellular parasites.pdf>.

KLECKNER, N. Meiosis: how could it work? **Proceedings of the National Academy of Sciences of the United States of America**, 1996. v. 93, n. 16, p. 8167–74. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=38641&tool=pmcentrez&rendertype=abstract>>.

LAGESEN, K. *et al.* RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. **Nucleic Acids Research**, 2007. v. 35, n. 9, p. 3100–3108.

LAHR, D. J. G. *et al.* The chastity of amoebae: re-evaluating evidence for sex in amoeboid organisms. **Proceedings. Biological sciences / The Royal Society**, 2011. v. 278, n. 1715, p. 2081–90. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3107637&tool=pmcentrez&rendertype=abstract>>.

LEE, E. *et al.* Web Apollo: a web-based genomic annotation editing platform. **Genome biology**, 2013. v. 14, n. 8, p. R93. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24000942>>.

LEE, S. C.; WEISS, L. M.; HEITMAN, J. **Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer. Communicative and Integrative Biology.**

LETUNIC, I.; DOERKS, T.; BORK, P. SMART: Recent updates, new developments and status in 2015. **Nucleic Acids Research**, 2015. v. 43, n. D1, p. D257–D260.

LOGSDON, J. M. **Evolutionary Genetics: Sex Happens in Giardia. Current Biology.**

LOWE, T. M.; CHAN, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. **Nucleic acids research**, 2016. v. 44, n. W1, p. W54-7. Disponível em:

<<http://www.ncbi.nlm.nih.gov/pubmed/27174935>&http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4987944>.

MATHIS, A.; WEBER, R.; DEPLAZES, P. **Zoonotic potential of the microsporidia. *Clinical Microbiology Reviews*.**

MURRAY WITTNER, L. M. W. *The Microsporidia and Microsporidiosis. ASM Press, 1999. p. 553 Seiten.*

NAKJANG, S. *et al.* Reduction and Expansion in Microsporidian Genome Evolution: New Insights from Comparative Genomics. ***Genome Biology and Evolution***, 2013. v. 5, n. 12, p. 2285–2303. Disponível em: <<https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evt184>>.

NURK, S. *et al.* Assembling genomes and mini-metagenomes from highly chimeric reads. [S.l.]: [s.n.], 2013. V. 7821 LNBI, p. 158–170.

PALDI, N. *et al.* Effective gene silencing in a microsporidian parasite associated with honeybee (*Apis mellifera*) colony declines. ***Applied and Environmental Microbiology***, 2010. v. 76, n. 17, p. 5960–5964.

PARISOT, N. *et al.* Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. ***Genome Biology and Evolution***, 2014. v. 6, n. 9, p. 2289–2300.

PEYRETAILLADE, E. *et al.* Annotation of microsporidian genomes using transcriptional signals. ***Nature Communications***, 2012. v. 3, p. 1137. Disponível em: <<http://www.nature.com/doifinder/10.1038/ncomms2156>>.

SCHURKO, A. M.; NEIMAN, M.; LOGSDON, J. M. Signs of sex: what we know and how we know it. ***Trends in Ecology and Evolution***, 2009. v. 24, n. 4, p. 208–217.

SELMAN, M.; CORRADI, N. Microsporidia: Horizontal gene transfers in vicious parasites. ***Mobile genetic elements***, 2011. v. 1, n. 4, p. 251–255. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3337132&tool=pmcentrez&rendertype=abstract>>.

SIMON, J. C. *et al.* Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. [S.l.]: [s.n.], 2003. V. 79, p. 151–163.

STENTIFORD, G. D. *et al.* **Microsporidia ??? Emergent Pathogens in the Global Food Chain. *Trends in Parasitology*.**

TAKAGI, Y. On the origin of sexual reproduction: a hypothesis. ***Jpn. J. Protozool.***, 2010. v. 43, n. 2, p. 89–93.

TEODOROVIC, S.; BRAVERMAN, J. M.; ELMENDORF, H. G. Unusually low levels of genetic variation among *Giardia lamblia* isolates. **Eukaryotic Cell**, 2007. v. 6, n. 8, p. 1421–1430.

THOMARAT, F.; VIVARÈS, C. P.; GOUY, M. Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. **Journal of Molecular Evolution**, 2004. v. 59, n. 6, p. 780–791.

TREANGEN, T. J.; SALZBERG, S. L. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. **Nature Reviews Genetics**, 2012. v. 13, n. 1, p. 36–46.

VILCINSKAS, A. *et al.* Evolutionary ecology of microsporidia associated with the invasive ladybird *Harmonia axyridis*. **Insect Science**, 2014. v. 0, p. 1–12.

VRIJENHOEK, R. C. Animal clones and diversity. Are natural clones generalists or specialists? **BioScience**, 1998. v. 48, n. 8, p. 617–628.

WANG, J. *et al.* Revealing and avoiding bias in semantic similarity scores for protein pairs. **BMC Bioinformatics**, 2010. v. 11, n. 1, p. 290. Disponível em: <<http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-290>>.

WANG, Y. *et al.* OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. **Nucleic Acids Research**, 2015. v. 43, n. W1, p. W78–W84.

YAMAGUCHI, K.; KAJIKAWA, M.; OKADA, N. LINE retrotransposition and host DNA repair machinery. **Mobile Genetic Elements**, 2015. v. 5, n. 6, p. 92–97. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/2159256X.2015.1096998>>.



### 3. APÊNDICES

#### APÊNDICE 1: PARÂMETROS E SCRIPT SICKLE

Parâmetros utilizados para filtragem de leituras de sequenciamento por qualidade e tamanho.

```
#!/usr/bin/bash

SP1=BE_magnivora
SP2=IL_magnivora

pe1=R1_1.fastq #leituras brutas
pe2=R2_1.fastq #leituras brutas
pe3=R1_q30.fastq #leituras com filtro de qualidade min 30
pe4=R2_q30.fastq #leituras com filtro de qualidade min 30
pe5=R1_q30l100.fastq #leituras filtro de qualidade min 30 e
filtro de tamanho min 100
pe6=R2_q30l100.fastq #leituras filtro de qualidade min 30 e
filtro de tamanho min 100

cd $SP1
sickle pe -q 30 -f $pe1 -r $pe2 -t sanger -o $pe3 -p $pe4 -s
singles.sickle.fastq
cd ../

cd $SP2
sickle pe -q 30 -f $pe1 -r $pe2 -t sanger -o $pe3 -p $pe4 -s
singles.sickle.fastq
cd ../

cd $SP1
sickle pe -l 100 -f $pe3 -r $pe4 -t sanger -o $pe5 -p $pe6 -s
singlesl.sickle.fastq
cd ../

cd $SP2
sickle pe -l 100 -f $pe3 -r $pe4 -t sanger -o $pe5 -p $pe6 -s
singlesl.sickle.fastq
cd ../
```

## APÊNDICE 2: PARÂMETROS E SCRIPT SPADES

Parâmetros para montagem dos genomas com SPAdes.

```
#!/usr/bin/bash

SP1=BE_magnivora
SP2=IL_magnivora

raw=SPAdes_raw
trim=SPAdes_trim

pe1=R1_1.fastq
pe2=R2_1.fastq
pe5=R1_q30l100.fastq
pe6=R2_q30l100.fastq

cd $SP1
cd $raw
/opt/SPAdes-3.5.0-Linux/bin spades.py -k 21,33,55,77,99,127 --
careful --only-assembler --pe1-1 $pe1 --pe1-2 $pe2 -o
spades_output
cd ../
cd $trim
/opt/SPAdes-3.5.0-Linux/bin spades.py -k 21,33,55,77,99,127 --
careful --only-assembler --pe1-1 $pe5 --pe1-2 $pe6 -o
cd ../../

cd $SP2
cd $raw
/opt/SPAdes-3.5.0-Linux/bin spades.py -k 21,33,55,77,99,127 --
careful --only-assembler --pe1-1 $pe1 --pe1-2 $pe2 -o
spades_output
cd ../
cd $trim
/opt/SPAdes-3.5.0-Linux/bin spades.py -k 21,33,55,77,99,127 --
careful --only-assembler --pe1-1 $pe5 --pe1-2 $pe6 -o
cd ../../
```

### APÊNDICE 3: PARÂMETROS E SCRIPT FINDORFS.SH

Parâmetros utilizados para predição de ORFs.

```
#!/usr/bin/bash
SP1=BE_magnivora
SP2=IL_magnivora

trim=SPAdes_trim
cont=spades_trim

cd $SP1
cd $trim
cd $cont
/usr/bin/getorf % getorf -minsize 450 -sequence contigs.fasta -
outseq BE_orfs
cd ../../../../

cd $SP2
cd $trim
cd $cont
/usr/bin/getorf % getorf -minsize 450 -sequence contigs.fasta -
outseq IL_orfs
cd ../../../../
```

## APÊNDICE 4: SCRIPT GETTAXID.PL

Script utilizado para acessar informação taxonômica a parti de blastp durante primeiro ciclo de filtragem

```
#!/usr/bin/perl

### USAGE: perl GetTaxID.pl blastout_file [database] (the sequence
db (nucleotide or protein) the GI is for)
### Requires perl-XML-Twig
### Requires an internet connection to access the taxonomy
information
### Written by Fabien Burki

use Bio::SearchIO;
use Bio::DB::Taxonomy;

my $blast_report = new Bio::SearchIO('-format' => 'blasttable', #
modify this line according to your blast output format (blasttable
if tabular; blast if default, etc)
                                '-file' => $ARGV[0]);

my $db = Bio::DB::Taxonomy->new (-source => 'entrez', -verbose =>
$verbose);

open (OUT, ">taxIDs.txt");

#WARNING: set how many hits and hsp's you want !

my $nbhits=1;
my $nbhsps=1;

print OUT "Query\tAC\tHit description\tScientific
name\tPhylum\tKingdom\tSuperkingdom\tEvalue\tIdentity\n";

while (my $result = $blast_report->next_result) {
    #output "no hits found" if that's the case
    if ($result -> num_hits == 0) {
        print OUT $result->query_name(), "\t";
        print $result->query_name(), "\t";
    #    print $result->query_length(), "\t";
        print OUT "No hits\n";
        print "No hits\n";
    }
    else {
        my $counter=0;
        while (my $hit = $result->next_hit()) {
            if ($counter < $nbhits) {
                while (my $hsp = $hit->next_hsp()) {
                    if ($counter < $nbhsps) {
                        print OUT $result->query_name(),
"\t";
```

```

print $result->query_name(), "\t";
my $hitname = $hit->name();
# print "$hitname\t";
print OUT $hit->accession(), "\t";
if ($hitname =~ /gi\|(\d+)\|/) {
    my @array;
    my $gi = $1;
    print "$gi\t";
    print OUT $hit->description,
"\t";
    my $node;
    eval {$node = $db-
>get_Taxonomy_Node(-gi => $gi, -db => $dbname)};

    if ($?) {
        print "***Unknown taxid***\n";
        print OUT "***Unknown taxid***\n";
        next;
    }
    else {
        print OUT $node-
        print $node-
        my $parent = $db-
        print $parent-
        while (defined $parent
        && $parent->node_name ne 'root') {

            if ($parent->rank
            eq 'superkingdom') {
                push @array,
                $parent->node_name;
            }
            elsif ($parent-
            >rank eq 'kingdom') {
                push
                @array, $parent->node_name;
            }
            elsif ($parent-
            >rank eq 'phylum') {
                push @array,
                $parent->node_name;
            }
            $parent = $db-
            >get_Taxonomy_Node($parent->parent_id);
        }
    }
    my $size = @array;
    if ($size == 1) {

```



## APÊNDICE 5: SCRIPT BIN\_TAXIDS.PL

```
#!/usr/bin/perl

use strict;
use warnings;

my $usage = 'USAGE = bin_taxIDs.pl taxIDs.txt';
die "$usage\n" unless @ARGV;

open IN, "<$ARGV[0]";
open OUT1, ">Microsporidia.txt";
open OUT2, ">Daphnia.txt";
open OUT3, ">Uncertae_sedis.txt";

my $query = 'ABC';
my $microsporidia = 0;
my $daphnia = 0;
my $cutoff = 5; ## Ratio of Microsporidia to Daphnia protein to
be considered valid

while (my $line = <IN>){
    chomp $line;
    if ($line =~
/^ (NODE.*) _(\d+)\t\t\t.*(Microsporidia|Arthropoda)/){
        my $contig = $1;
        my $lineage = $3;
        if ($query eq 'ABC'){
            $query = $contig;
            if ($lineage eq
'Microsporidia'){ $microsporidia++;}
            elsif ($lineage eq 'Arthropoda'){ $daphnia++;}
        }
        elsif ($query eq $contig){
            if ($lineage eq
'Microsporidia'){ $microsporidia++;}
            elsif ($lineage eq 'Arthropoda'){ $daphnia++;}
        }
        elsif ($query ne $contig){
            if (($daphnia >= 0) && ($microsporidia == 0)){
                print OUT2 "$query\n";
            }
            elsif (($daphnia == 0) && ($microsporidia >= 0)){
                print OUT1 "$query\n";
            }
            elsif (($daphnia >= 0) && ($microsporidia >= 0)){
                my $ratio = $microsporidia/$daphnia;
                if ($ratio >= $cutoff){
                    print OUT1 "$query\n";
                }
            }
            else{
                print OUT3 "$query\n";
            }
        }
    }
}
```

```

    }
  }
  ### Working on the new contig
  $query = $contig;
  $microsporidia = 0; ## resetting values
  $daphnia = 0; ## resetting values
  if ($lineage eq
'Microsporidia'){ $microsporidia++;}
    elseif ($lineage eq 'Arthropoda'){ $daphnia++;}
  }
}
### Closure
if (($daphnia >= 0) && ($microsporidia == 0)){
  print OUT2 "$query\n";
}
elseif (($daphnia == 0) && ($microsporidia >= 0)){
  print OUT1 "$query\n";
}
elseif (($daphnia >= 0) && ($microsporidia >= 0)){
  my $ratio = $microsporidia/$daphnia;
  if ($ratio >= $cutoff){
    print OUT1 "$query\n";
  }
  else{
    print OUT3 "$query\n";
  }
}
}

```



## APÊNDICE 6: SCRIPT RUNTAXONOMIZEDBLAST.PL

Script utilizado para acessar informação taxonômica a partir de blastp durante primeiro ciclo de filtragem

```
#!/usr/bin/perl
## Pombert Lab, IIT 2016

## Requirements:
## BLAST 2.2.28+ or later
## NCBI taxonomy database
(ftp://ftp.ncbi.nlm.nih.gov/blast/db/taxdb.tar.gz)
## NCBI NR/NT databases (ftp://ftp.ncbi.nlm.nih.gov/blast/db/)
## NOTE: the NCBI taxdb, nr and nt databases can be downloaded
with the update_blastdb.pl from NCBI
(http://www.ncbi.nlm.nih.gov/blast/docs/update_blastdb.pl)
## The BLASTDB variable must be set in the environmental
variables: export
BLASTDB=/path/to/NCBI/TaxDB:/path/to/NCBI/NR:/path/to/NCBI/NT

use strict;
use warnings;
use Getopt::Long qw(GetOptions);

my $usage = "
USAGE = perl runTaxonomizedBLAST.pl [options]
OPTIONS:
--type          blastn, blastx, blastp... [default = blastp]
--db            nt, nr or custom subset [default = nr]
--threads      2, 4, 8 or more [default = 2]
--evaluate     1e-05, 1e-10 or other [default = 1e-05]
--culling      culling limit [default = 1]
--query        fasta file to be queried
";

die "$usage\n" unless@ARGV;

## Defining options
my $blast_type = 'blastp';
my $db = 'nr';
my $threads = '2';
my $value = '1e-05';
my $culling = '1';
my $query;

GetOptions(
    'type=s' => \$blast_type,
    'db=s' => \$db,
    'threads=i' => \$threads,
```

```
'evaluate=s' => \$evaluate,  
'culling=i' => \$culling,  
'query=s' => \$query,  
);
```

```
## Running BLAST
```

```
system "$blast_type -num_threads $threads -query $query -db $db -  
evaluate $evaluate -culling_limit $culling -outfmt '6 qseqid sseqid  
qstart qend pident length bitscore evalue staxids sscinames  
sskingdoms sblastnames' -out $query.$blast_type";
```

## APÊNDICE 7: SCRIPT RUN\_RNAHAMMER.PL

Script utilizado para predição de RNAs ribossomais.

```
#!/usr/bin/perl

use strict;
use warnings;

my $usage = 'perl run_rnahammer.pl kingdom files.fsa'; ## kingdom
= arc, bac or euk
die "\n$usage\n\n" unless @ARGV;

my $kingdom = shift@ARGV;

while (my $file = shift@ARGV){
    system "rnammer -S $kingdom -m tsu,ssu,lsu -gff $file.gff -h
$file.hmm -f $file.rRNAs < $file";
}
```

## APÊNDICE 8: SCRIPT RUN\_TRNASCAN.PL

Script utilizado para predição de RNAs transportadores.

```
#!/usr/bin/perl

use strict;
use warnings;

while (my $file = shift@ARGV){
    system "tRNAscan-SE $file > $file.tRNAs";
}
```

## APÊNDICE 9: PARÂMETROS MAKER

Arquivo de parâmetros da *pipeline* de predição gênica MAKER utilizando *E. cuniculi* como modelo.

```
#-----Genome (these are always required)
genome=BE_contigs_final.fasta #genome sequence (fasta file or
fasta embeded in GFF3 file)
organism_type=eukaryotic #eukaryotic or prokaryotic. Default is
eukaryotic

#-----Re-annotation Using MAKER Derived GFF3
maker_gff= #MAKER derived GFF3 file
est_pass=0 #use ESTs in maker_gff: 1 = yes, 0 = no
altest_pass=0 #use alternate organism ESTs in maker_gff: 1 = yes,
0 = no
protein_pass=0 #use protein alignments in maker_gff: 1 = yes, 0 =
no
rm_pass=0 #use repeats in maker_gff: 1 = yes, 0 = no
model_pass=0 #use gene models in maker_gff: 1 = yes, 0 = no
pred_pass=0 #use ab-initio predictions in maker_gff: 1 = yes, 0 =
no
other_pass=0 #passthrough anything else in maker_gff: 1 = yes, 0
= no

#-----EST Evidence (for best results provide a file for at least
one)
est= #set of ESTs or assembled mRNA-seq in fasta format
altest= #EST/cDNA sequence file in fasta format from an alternate
organism
est_gff= #aligned ESTs or mRNA-seq from an external GFF3 file
altest_gff= #aligned ESTs from a closly relate species in GFF3
format

#-----Protein Homology Evidence (for best results provide a file
for at least one)
protein= #protein sequence file in fasta format (i.e. from
mutiple organisms)
protein_gff= #aligned protein homology evidence from an external
GFF3 file

#-----Repeat Masking (leave values blank to skip repeat masking)
model_org=all #select a model organism for RepBase masking in
RepeatMasker
rmlib= #provide an organism specific repeat library in fasta
format for RepeatMasker
repeat_protein=/opt/maker/data/te_proteins.fasta #provide a fasta
file of transposable element proteins for RepeatRunner
rm_gff= #pre-identified repeat elements from an external GFF3 file
prok_rm=0 #forces MAKER to repeatmask prokaryotes (no reason to
change this), 1 = yes, 0 = no
```

```

softmask=1 #use soft-masking rather than hard-masking in BLAST
(i.e. seg and dust filtering)

#-----Gene Prediction
snaphmm= #SNAP HMM file
gmhmm= #GeneMark HMM file
augustus_species=encephalitozoon_cuniculi_GB #Augustus gene
prediction species model
fgenesh_par_file= #FGENESH parameter file
pred_gff= #ab-initio predictions from an external GFF3 file
model_gff= #annotated gene models from an external GFF3 file
(annotation pass-through)
run_evm=0 #run EvidenceModeler, 1 = yes, 0 = no
est2genome=0 #infer gene predictions directly from ESTs, 1 = yes,
0 = no
protein2genome=0 #infer predictions from protein homology, 1 =
yes, 0 = no
trna=0 #find tRNAs with tRNAscan, 1 = yes, 0 = no
snoscan_rrna= #rRNA file to have Snoscan find snoRNAs
snoscan_meth= #-O-methylation site file to have Snoscan find
snoRNAs
unmask=0 #also run ab-initio prediction programs on unmasked
sequence, 1 = yes, 0 = no

#-----Other Annotation Feature Types (features MAKER doesn't
recognize)
other_gff= #extra features to pass-through to final MAKER
generated GFF3 file

#-----External Application Behavior Options
alt_peptide=C #amino acid used to replace non-standard amino acids
in BLAST databases
cpus=1 #max number of cpus to use in BLAST and RepeatMasker (not
for MPI, leave 1 when using MPI)

#-----MAKER Behavior Options
max_dna_len=100000 #length for dividing up contigs into chunks
(increases/decreases memory usage)
min_contig=1 #skip genome contigs below this length (under 10kb
are often useless)

pred_flank=200 #flank for extending evidence clusters sent to gene
predictors
pred_stats=0 #report AED and QI statistics for all predictions as
well as models
AED_threshold=1 #Maximum Annotation Edit Distance allowed (bound
by 0 and 1)
min_protein=0 #require at least this many amino acids in predicted
proteins
alt_splice=0 #Take extra steps to try and find alternative
splicing, 1 = yes, 0 = no
always_complete=0 #extra steps to force start and stop codons, 1 =
yes, 0 = no

```

```
map_forward=0 #map names and attributes forward from old GFF3
genes, 1 = yes, 0 = no
keep_preds=0 #Concordance threshold to add unsupported gene
prediction (bound by 0 and 1)

split_hit=10000 #length for the splitting of hits (expected max
intron size for evidence alignments)
min_intron=20 #minimum intron length (used for alignment
polishing)
single_exon=0 #consider single exon EST evidence when generating
annotations, 1 = yes, 0 = no
single_length=250 #min length required for single exon ESTs if
'single_exon is enabled'
correct_est_fusion=0 #limits use of ESTs in annotation to avoid
fusion genes

tries=2 #number of times to try a contig if there is a failure for
some reason
clean_try=0 #remove all data from previous run before retrying, 1
= yes, 0 = no
clean_up=0 #removes theVoid directory with individual analysis
files, 1 = yes, 0 = no
TMP= #specify a directory other than the system default temporary
directory for temporary files
```

Arquivo de parâmetros da *pipeline* de predição gênica MAKER utilizando *Rozella alomycis* como modelo.

```
#-----Genome (these are always required)
genome=BE_contigs_final.fasta #genome sequence (fasta file or
fasta embeded in GFF3 file)
organism_type=eukaryotic #eukaryotic or prokaryotic. Default is
eukaryotic

#-----Re-annotation Using MAKER Derived GFF3
maker_gff= #MAKER derived GFF3 file
est_pass=0 #use ESTs in maker_gff: 1 = yes, 0 = no
altest_pass=0 #use alternate organism ESTs in maker_gff: 1 = yes,
0 = no
protein_pass=0 #use protein alignments in maker_gff: 1 = yes, 0 =
no
rm_pass=0 #use repeats in maker_gff: 1 = yes, 0 = no
model_pass=0 #use gene models in maker_gff: 1 = yes, 0 = no
pred_pass=0 #use ab-initio predictions in maker_gff: 1 = yes, 0 =
no
other_pass=0 #passthrough anything else in maker_gff: 1 = yes, 0
= no

#-----EST Evidence (for best results provide a file for at least
one)
est= #set of ESTs or assembled mRNA-seq in fasta format
altest= #EST/cDNA sequence file in fasta format from an alternate
organism
est_gff= #aligned ESTs or mRNA-seq from an external GFF3 file
altest_gff= #aligned ESTs from a closly relate species in GFF3
format

#-----Protein Homology Evidence (for best results provide a file
for at least one)
protein= #protein sequence file in fasta format (i.e. from
mutiple organisms)
protein_gff= #aligned protein homology evidence from an external
GFF3 file

#-----Repeat Masking (leave values blank to skip repeat masking)
model_org=all #select a model organism for RepBase masking in
RepeatMasker
rmlib= #provide an organism specific repeat library in fasta
format for RepeatMasker
repeat_protein=/opt/maker/data/te_proteins.fasta #provide a fasta
file of transposable element proteins for RepeatRunner
rm_gff= #pre-identified repeat elements from an external GFF3 file
prok_rm=0 #forces MAKER to repeatmask prokaryotes (no reason to
change this), 1 = yes, 0 = no
softmask=1 #use soft-masking rather than hard-masking in BLAST
(i.e. seg and dust filtering)
```



```

#-----Gene Prediction
snaphmm= #SNAP HMM file
gmhmm= #GeneMark HMM file
augustus_species=rozella #Augustus gene prediction species model
fgenesh_par_file= #FGENESH parameter file
pred_gff= #ab-initio predictions from an external GFF3 file
model_gff= #annotated gene models from an external GFF3 file
(annotation pass-through)
run_evm=0 #run EvidenceModeler, 1 = yes, 0 = no
est2genome=0 #infer gene predictions directly from ESTs, 1 = yes,
0 = no
protein2genome=0 #infer predictions from protein homology, 1 =
yes, 0 = no
trna=0 #find tRNAs with tRNAscan, 1 = yes, 0 = no
snoscan_rrna= #rRNA file to have Snoscan find snoRNAs
snoscan_meth= #-O-methylation site file to have Snoscan find
snoRNAs
unmask=0 #also run ab-initio prediction programs on unmasked
sequence, 1 = yes, 0 = no

#-----Other Annotation Feature Types (features MAKER doesn't
recognize)
other_gff= #extra features to pass-through to final MAKER
generated GFF3 file

#-----External Application Behavior Options
alt_peptide=C #amino acid used to replace non-standard amino acids
in BLAST databases
cpus=1 #max number of cpus to use in BLAST and RepeatMasker (not
for MPI, leave 1 when using MPI)

#-----MAKER Behavior Options
max_dna_len=100000 #length for dividing up contigs into chunks
(increases/decreases memory usage)
min_contig=1 #skip genome contigs below this length (under 10kb
are often useless)

pred_flank=200 #flank for extending evidence clusters sent to gene
predictors
pred_stats=0 #report AED and QI statistics for all predictions as
well as models
AED_threshold=1 #Maximum Annotation Edit Distance allowed (bound
by 0 and 1)
min_protein=0 #require at least this many amino acids in predicted
proteins
alt_splice=0 #Take extra steps to try and find alternative
splicing, 1 = yes, 0 = no
always_complete=0 #extra steps to force start and stop codons, 1 =
yes, 0 = no
map_forward=0 #map names and attributes forward from old GFF3
genes, 1 = yes, 0 = no

```

```
keep_preds=0 #Concordance threshold to add unsupported gene
prediction (bound by 0 and 1)

split_hit=10000 #length for the splitting of hits (expected max
intron size for evidence alignments)
min_intron=20 #minimum intron length (used for alignment
polishing)
single_exon=0 #consider single exon EST evidence when generating
annotations, 1 = yes, 0 = no
single_length=250 #min length required for single exon ESTs if
'single_exon is enabled'
correct_est_fusion=0 #limits use of ESTs in annotation to avoid
fusion genes

tries=2 #number of times to try a contig if there is a failure for
some reason
clean_try=0 #remove all data from previous run before retrying, 1
= yes, 0 = no
clean_up=0 #removes theVoid directory with individual analysis
files, 1 = yes, 0 = no
TMP=#specify a directory other than the system default temporary
directory for temporary files
```

## APÊNDICE 10: CLUSTERS COMPARTILHADOS PELAS 4 LINHAGENS DE MICROSPORÍDEOS

Análise do OrthoVenn com as 3 linhagens montadas mais o proteoma completo de *E. cuniculi* EC1.

Cluster ID	Número de proteínas	Swiss-Prot <i>Hit</i>	<i>E-value</i>
cluster455	4	14-3-3-like protein C	9.40E-56
cluster269	4	1-acyl-sn-glycerol-3-phosphate acyltransferase 3	2.50E-11
cluster675	4	26S protease regulatory subunit 6A homolog B	3.70E-47
cluster481	4	26S proteasome non-ATPase regulatory subunit 11 homolog	1.20E-51
cluster433	4	2-oxoisovalerate dehydrogenase subunit beta 2, mitochondrial	7.60E-34
cluster486	4	30S ribosomal protein S11, chloroplastic {ECO:0000255   HAMAP-Rule:MF_01310}	4.60E-06
cluster537	4	30S ribosomal protein S5, chloroplastic	3.00E-11
cluster659	4	30S ribosomal protein S7, chloroplastic	1.30E-06
cluster592	4	31 kDa ribonucleoprotein, chloroplastic	6.00E-08
cluster694	4	31 kDa ribonucleoprotein, chloroplastic	4.70E-03
cluster614	4	3-phosphoinositide-dependent protein kinase 1	6.70E-08
cluster75	4	40S ribosomal protein S11-2	4.40E-24
cluster385	4	40S ribosomal protein S16	5.00E-22
cluster448	4	40S ribosomal protein S23-1	1.30E-35
cluster635	4	40S ribosomal protein S6	1.50E-17
cluster88	4	40S ribosomal protein S8-1	1.10E-07
cluster15	5	4-coumarate--CoA ligase	6.70E-07

cluster361	4	4-coumarate--CoA ligase-like 1	4.00E-15
cluster291	4	50S ribosomal protein L5, plastid	2.80E-03
cluster131	4	5'-3' exoribonuclease 3	1.70E-53
cluster516	4	60 kDa chaperonin, chloroplastic {ECO:0000255   HAMAP- Rule:MF_00600}	2.90E-06
cluster343	4	60S acidic ribosomal protein P0	9.40E-35
cluster478	4	60S ribosomal protein L10	1.50E-33
cluster585	4	60S ribosomal protein L13	6.90E-12
cluster68	4	60S ribosomal protein L13a	2.30E-22
cluster142	4	60S ribosomal protein L15	2.60E-32
cluster763	4	60S ribosomal protein L17-1	2.50E-31
cluster338	4	60S ribosomal protein L18a-3	8.40E-22
cluster460	4	60S ribosomal protein L19	1.80E-10
cluster369	4	60S ribosomal protein L2, mitochondrial	1.80E-13
cluster62	4	60S ribosomal protein L27a-2	3.00E-30
cluster471	4	60S ribosomal protein L5	2.20E-13
cluster115	4	60S ribosomal protein L6	3.80E-13
cluster422	4	60S ribosomal protein L7a	5.40E-32
cluster583	4	6-phosphogluconate dehydrogenase, decarboxylating 2, chloroplastic	2.40E-70
cluster658	4	ABC transporter B family member 10	3.30E-05
cluster700	4	ABC transporter B family member 4	9.70E-06
cluster231	4	ABC transporter C family member 2	3.30E-24
cluster132	4	Acetyl-CoA acetyltransferase, cytosolic 1	1.40E-58
cluster278	4	ADP-ribosylation factor 1	2.70E-21

cluster421	4 ADP-ribosylation factor-like protein 2 Adrenodoxin-like protein 2, mitochondrial	3.80E-13
cluster743	4 {ECO:0000303   PubMed:12714594}	1.20E-27
cluster575	4 Alpha,alpha-trehalose-phosphate synthase [UDP-forming] 6	5.00E-60
cluster566	4 Anaphase-promoting complex subunit 8	1.60E-42
cluster221	4 Armadillo repeat-containing kinesin-like protein 3	8.70E-35
cluster445	4 Asparagine--tRNA ligase, chloroplastic/mitochondrial ATP synthase subunit alpha, chloroplastic {ECO:0000255   HAMAP-	1.20E-05
cluster78	4 Rule:MF_01346}	1.70E-15
cluster544	4 ATPase 4, plasma membrane-type ATP-dependent (S)-NAD(P)H-hydrate dehydratase	9.10E-19
cluster261	4 {ECO:0000255   HAMAP-Rule:MF_03157}	2.60E-16
cluster742	4 ATP-dependent zinc metalloprotease FTSH 1, chloroplastic	1.30E-30
cluster349	4 ATP-dependent zinc metalloprotease FTSH 2, chloroplastic	7.20E-31
cluster744	4 ATP-dependent zinc metalloprotease FTSH 2, chloroplastic	4.10E-39
cluster393	4 ATP-dependent zinc metalloprotease FTSH, chloroplastic	3.00E-53
cluster105	4 B-box zinc finger protein 24 {ECO:0000303   PubMed:19920209}	8.10E-04
cluster418	4 CAAX prenyl protease 1 homolog	5.10E-64
cluster360	4 Calcium-binding protein CML19	1.10E-03
cluster65	4 Calcium-dependent protein kinase 15	5.50E-06
cluster10	6 Calcium-transporting ATPase 1, plasma membrane-type	6.60E-03
cluster367	4 CBL-interacting protein kinase 2	4.20E-38
cluster245	4 CBL-interacting protein kinase 32	6.00E-28
cluster229	4 CBL-interacting serine/threonine-protein kinase 16	1.10E-45
cluster739	4 CBL-interacting serine/threonine-protein kinase 8	1.10E-23

cluster520	4 Cell division control protein 2 homolog	3.30E-34
cluster612	4 Cell division control protein 48 homolog A	8.30E-48
cluster109	4 Cell division control protein 48 homolog D	1.00E-53
cluster180	4 Cell division control protein 48 homolog D	3.30E-52
cluster446	4 Cell division control protein 48 homolog D	1.40E-45
cluster535	4 Cell division control protein 48 homolog E	2.90E-46
cluster306	4 Chaperone protein ClpC4, chloroplastic	6.70E-58
cluster333	4 CHD3-type chromatin-remodeling factor PICKLE	1.40E-03
cluster189	4 Choline-phosphate cytidyltransferase 1	1.00E-55
cluster317	4 Cleavage and polyadenylation specificity factor subunit 2	2.00E-38
cluster332	4 Cleavage and polyadenylation specificity factor subunit 2	1.20E-22
cluster213	4 Cleavage and polyadenylation specificity factor subunit 3-l	8.40E-146
cluster260	4 Coatomer subunit beta-1	3.60E-16
cluster230	4 Coatomer subunit gamma-2	2.10E-55
cluster388	4 Condensin complex subunit 2	2.40E-03
cluster195	4 COP9 signalosome complex subunit 5b	2.20E-29
cluster117	4 Copper-transporting ATPase PAA1, chloroplastic	3.40E-05
cluster341	4 Crossover junction endonuclease MUS81	5.80E-14
cluster101	4 Cyclin-dependent kinase A-1	1.90E-07
cluster499	4 Cyclin-dependent kinase B1-1	2.50E-45
cluster467	4 Cyclin-dependent kinase B2-2	1.80E-37
cluster177	4 Cyclin-dependent kinase D-3	2.60E-65
cluster148	4 Cyclin-dependent kinase G1	7.40E-40
cluster493	4 Cyclin-dependent kinase G-1	7.00E-50

cluster240	4 Cysteine desulfurase 1, mitochondrial Cytoplasmic tRNA 2-thiolation protein 1 {ECO:0000255   HAMAP- Rule:MF_03053}	5.30E-144
cluster561	4 DEAD-box ATP-dependent RNA helicase 10	5.30E-85
cluster682	4 DEAD-box ATP-dependent RNA helicase 12	5.90E-47
cluster629	4 DEAD-box ATP-dependent RNA helicase 13	5.30E-48
cluster416	4 DEAD-box ATP-dependent RNA helicase 26	1.30E-07
cluster521	4 DEAD-box ATP-dependent RNA helicase 3, chloroplastic	9.60E-43
cluster507	4 DEAD-box ATP-dependent RNA helicase 42	4.20E-43
cluster232	4 DEAD-box ATP-dependent RNA helicase 52B	3.00E-04
cluster174	4 DEAD-box ATP-dependent RNA helicase 56	4.50E-75
cluster123	4 DEAD-box ATP-dependent RNA helicase 56	8.80E-55
cluster387	4 DEAD-box ATP-dependent RNA helicase ISE2, chloroplastic	2.80E-51
cluster562	4 DEAD-box ATP-dependent RNA helicase ISE2, chloroplastic	6.00E-13
cluster627	4 Deoxyhypusine hydroxylase-A {ECO:0000255   HAMAP- Rule:MF_03101}	8.20E-25
cluster731	4 Deoxyhypusine synthase	4.70E-16
cluster713	4 Developmentally-regulated G-protein 2	6.90E-67
cluster551	4 DNA ligase 1	1.50E-107
cluster660	4 DNA mismatch repair protein MLH3	2.70E-119
cluster528	4 DNA mismatch repair protein MSH6	2.20E-17
cluster480	4 DNA mismatch repair protein PMS1	1.00E-27
cluster258	4 DNA polymerase alpha catalytic subunit	4.50E-35
cluster285	4 DNA polymerase delta catalytic subunit	3.50E-07
cluster666	4 DNA polymerase epsilon catalytic subunit A	1.70E-222
cluster235		2.50E-232

cluster257	4	DNA repair helicase UVH6	1.90E-15
cluster281	4	DNA repair helicase UVH6	2.50E-186
cluster350	4	DNA repair helicase XPB1	1.70E-149
cluster525	4	DNA repair protein RAD50	1.60E-60
cluster77	4	DNA replication licensing factor MCM3 homolog 3	4.20E-32
cluster19	5	DNA replication licensing factor MCM7	2.80E-42
cluster33	5	DNA topoisomerase 2	0.00E+00
cluster347	4	DNA topoisomerase 3-beta	3.00E-83
cluster424	4	DNA-directed RNA polymerase II subunit 7	9.40E-27
cluster90	4	DNA-directed RNA polymerase subunit beta'	3.90E-23
cluster161	4	DNA-directed RNA polymerase subunit beta {ECO:0000255   HAMAP-Rule:MF_01321}	4.20E-39
cluster459	4	DNA-directed RNA polymerase subunit beta {ECO:0000255   HAMAP-Rule:MF_01321}	1.00E-34
cluster11	6	DNA-directed RNA polymerase subunit beta' {ECO:0000255   HAMAP-Rule:MF_01323}	2.80E-31
cluster681	4	DNA-directed RNA polymerase subunit beta' {ECO:0000255   HAMAP-Rule:MF_01323}	8.40E-35
cluster160	4	DNA-directed RNA polymerases IV and V subunit 3B	5.30E-29
cluster595	4	DNA-directed RNA polymerases IV and V subunit 5B	3.70E-15
cluster186	4	Double-strand break repair protein MRE11	1.50E-58
cluster456	4	Dynamin-2B	2.70E-16
cluster646	4	Dynamin-2B	3.70E-15
cluster104	4	Dynein assembly factor with WDR repeat domains 1	3.40E-10
cluster37	5	Dynein-1-beta heavy chain, flagellar inner arm I1 complex	1.60E-16
cluster458	4	E2F transcription factor-like E2FE	2.60E-05



cluster524	4	E3 ubiquitin-protein ligase UPL2	6.50E-41
cluster121	4	Elongation factor 1-alpha S	5.90E-31
cluster184	4	Elongation factor G-2, mitochondrial	2.80E-16
cluster539	4	Elongation factor Tu, chloroplastic	7.00E-33
cluster652	4	Elongator complex protein 3	5.70E-215
cluster536	4	Endoplasmic reticulum oxidoreductin-1	3.40E-10
cluster283	4	Enolase	1.30E-67
cluster207	4	Ethanolamine-phosphate cytidyltransferase	3.00E-51
cluster379	4	Eukaryotic peptide chain release factor subunit 1-1	1.50E-12
cluster320	4	Eukaryotic translation initiation factor 3 subunit I {ECO:0000255   HAMAP-Rule:MF_03008}	1.80E-05
cluster538	4	Eukaryotic translation initiation factor 6 {ECO:0000255   HAMAP- Rule:MF_03132}	2.70E-49
cluster712	4	Eukaryotic translation initiation factor NCBP	3.90E-05
cluster26	5	Exonuclease 1	1.00E-37
cluster531	4	Exonuclease 1	2.20E-12
cluster199	4	Farnesyl pyrophosphate synthase	1.60E-19
cluster135	4	Fimbrin-1 {ECO:0000303   PubMed:11123801}	5.90E-10
cluster67	4	FIP1 [III]-like protein {ECO:0000303   PubMed:16282318}	8.70E-08
cluster227	4	Flap endonuclease 1-A {ECO:0000255   HAMAP-Rule:MF_03140}	3.10E-03
cluster73	4	Flowering time control protein FY	1.90E-98
cluster277	4	Fructose-bisphosphate aldolase 1, chloroplastic	1.90E-56
cluster636	4	Gamma-glutamyltranspeptidase 3	3.30E-77
cluster639	4	GATA transcription factor 18	2.40E-06
cluster295	4	Glucose-6-phosphate 1-dehydrogenase, chloroplastic	1.20E-26

cluster209	4	Glucose-6-phosphate isomerase, cytosolic A	4.40E-73
cluster578	4	Glutamate--tRNA ligase, chloroplastic/mitochondrial Glutamine--fructose-6-phosphate aminotransferase [isomerizing]	6.70E-10
cluster106	4	{ECO:0000255   HAMAP-Rule:MF_00164}	5.20E-24
cluster415	4	Glutamine--tRNA ligase	2.10E-118
cluster634	4	Glycerol-3-phosphate dehydrogenase SDP6, mitochondrial	1.50E-112
cluster733	4	Glycylpeptide N-tetradecanoyltransferase 1	7.10E-38
cluster588	4	GTP-binding protein yptV1	4.50E-19
cluster632	4	Guanosine nucleotide diphosphate dissociation inhibitor At5g09550	1.10E-53
cluster476	4	Guanylate kinase 3, chloroplastic	3.00E-16
cluster623	4	H/ACA ribonucleoprotein complex subunit 2-like protein	1.30E-12
cluster378	4	H/ACA ribonucleoprotein complex subunit 4	4.20E-115
cluster134	4	Heat shock 70 kDa protein	2.40E-119
cluster518	4	Heat shock 70 kDa protein 18	1.20E-57
cluster301	4	Heat shock protein 82	3.80E-96
cluster351	4	Heat stress transcription factor A-9	2.60E-14
cluster380	4	Heterogeneous nuclear ribonucleoprotein 1	5.90E-09
cluster684	4	High mobility group B protein 9	3.90E-08
cluster673	4	Histidine--tRNA ligase	1.30E-81
cluster748	4	Histone acetyltransferase GCN5	9.80E-20
cluster81	4	Histone deacetylase 15	8.20E-23
cluster607	4	Histone deacetylase 8	6.20E-25
cluster395	4	Histone deacetylase HDT3	4.70E-03
cluster377	4	Histone H3.3	1.50E-31

cluster308	4 Histone H4	2.10E-14
cluster390	4 Histone-binding protein MSI1 homolog	1.20E-03
cluster552	4 Homeobox-leucine zipper protein HOX23	4.30E-04
cluster404	4 Homeobox-leucine zipper protein PROTODERMAL FACTOR 2	2.70E-06
cluster677	4 Importin subunit alpha-2	2.00E-28
cluster191	4 Iron-sulfur cluster assembly protein 2	2.40E-45
cluster63	4 K(+) efflux antiporter 1, chloroplastic	5.70E-04
cluster549	4 Kinesin-13A	3.00E-31
cluster120	4 Kinesin-5	2.00E-29
cluster420	4 Kinesin-like protein NACK1	3.50E-35
cluster449	4 Kinesin-like protein NACK2	3.40E-21
cluster717	4 La protein 1	1.60E-11
	Leucine-rich repeat receptor-like serine/threonine-protein kinase	
cluster107	4 BAM2	2.50E-04
cluster302	4 Long chain base biosynthesis protein 1	9.90E-35
cluster579	4 Long chain base biosynthesis protein 1c	1.30E-46
cluster389	4 Luminal-binding protein	1.10E-136
cluster201	4 Lysine--tRNA ligase	5.10E-130
cluster542	4 MADS-box transcription factor 50	1.90E-03
cluster130	4 Mechanosensitive ion channel protein 7	1.80E-19
cluster144	4 Mechanosensitive ion channel protein 7	2.50E-10
cluster400	4 Mechanosensitive ion channel protein 8	8.60E-23
cluster188	4 Mediator of RNA polymerase II transcription subunit 6	4.70E-14
cluster372	4 Meiotic recombination protein DMC1 homolog	8.30E-89

cluster490	4 Metal tolerance protein C2	4.50E-10
cluster567	4 Methionine aminopeptidase 2A {ECO:0000255   HAMAP- Rule:MF_03175}	8.70E-68
cluster358	4 MIP18 family protein At1g68310	1.70E-23
cluster642	4 Mitochondrial import inner membrane translocase subunit TIM50	1.40E-18
cluster767	4 Monothiol glutaredoxin-S1, mitochondrial	1.70E-12
cluster482	4 Myosin-9	1.60E-77
cluster401	4 NAD-dependent protein deacetylase SRT1	3.50E-08
cluster479	4 NEDD8-activating enzyme E1 catalytic subunit	1.30E-08
cluster563	4 NEDD8-activating enzyme E1 catalytic subunit	2.20E-17
cluster139	4 Notchless protein homolog	1.30E-05
cluster331	4 Notchless protein homolog	4.40E-07
cluster466	4 Nuclear poly(A) polymerase 3 {ECO:0000303   PubMed:18479511}	3.40E-49
cluster118	4 Nuclear transcription factor Y subunit B-9	5.70E-16
cluster491	4 Nuclear transcription factor Y subunit C-1	3.10E-04
cluster723	4 Nuclear transcription factor Y subunit C-7	1.60E-11
cluster696	4 Nucleolar GTP-binding protein 1	2.00E-113
cluster411	4 Nucleoside diphosphate kinase III, chloroplastic/mitochondrial	2.00E-34
cluster87	4 Peptidyl-prolyl cis-trans isomerase CYP65 {ECO:0000305}	2.80E-24
cluster154	4 Phosphatidylinositol 4-kinase alpha 2	1.10E-24
cluster391	4 Phosphatidylinositol 4-kinase beta 1	1.30E-53
cluster761	4 Phosphatidylinositol 4-kinase gamma 4	1.60E-05
cluster325	4 Phosphoacetylglucosamine mutase	3.10E-77
cluster127	4 Phosphoglycerate kinase, chloroplastic	2.90E-48

cluster23	5	Phosphoinositide phosphatase SAC5	5.30E-07
cluster196	4	Phospholipase D gamma 2	1.10E-04
cluster190	4	Phosphomannomutase	7.90E-63
cluster100	4	Phytochrome-associated serine/threonine-protein phosphatase 1	8.60E-93
cluster202	4	Pleiotropic drug resistance protein 5	1.70E-36
cluster489	4	Polyadenylate-binding protein 6	3.90E-03
cluster222	4	Polycomb group protein FERTILIZATION-INDEPENDENT ENDOSPERM	4.30E-04
cluster553	4	Polyubiquitin	2.40E-35
cluster152	4	Probable ADP-ribosylation factor GTPase-activating protein AGD5	5.00E-14
cluster410	4	Probable ADP-ribosylation factor GTPase-activating protein AGD9	3.20E-16
cluster251	4	Probable calcium-transporting ATPase 6, plasma membrane-type	3.80E-05
cluster693	4	Probable CCR4-associated factor 1 homolog 5	9.60E-24
cluster273	4	Probable DNA helicase MCM8	1.70E-62
cluster362	4	Probable DNA helicase MCM8	3.20E-59
cluster363	4	Probable DNA helicase MCM8	1.90E-58
cluster616	4	Probable DNA helicase MCM8	1.80E-53
cluster255	4	Probable DNA helicase MCM9	5.60E-63
cluster640	4	Probable DNA helicase MCM9	1.00E-67
cluster598	4	Probable DNA primase large subunit	3.60E-30
cluster732	4	Probable glutathione peroxidase 4	1.20E-22
cluster263	4	Probable LRR receptor-like serine/threonine-protein kinase At1g74360	6.70E-03
cluster586	4	Probable magnesium transporter NIPA8	7.90E-29
cluster674	4	Probable mannose-1-phosphate guanylyltransferase 3	6.60E-73

cluster203	4 Probable methionine--tRNA ligase Probable mitochondrial saccharopine dehydrogenase-like	6.80E-120
cluster112	4 oxidoreductase At5g39410	2.10E-35
cluster435	4 Probable mitochondrial-processing peptidase subunit beta	5.40E-10
cluster413	4 Probable MYST-like histone acetyltransferase 1	2.70E-80
cluster442	4 Probable phenylalanine--tRNA ligase alpha subunit	4.30E-123
cluster558	4 Probable protein S-acyltransferase 13	2.00E-19
cluster136	4 Probable sugar phosphate/phosphate translocator At1g06470	3.20E-34
cluster386	4 Probable threonine--tRNA ligase, cytoplasmic	2.00E-70
cluster488	4 Probable trehalase	2.00E-54
cluster200	4 Probable tRNA (guanine(26)-N(2))-dimethyltransferase 2	9.90E-34
cluster249	4 Probable V-type proton ATPase subunit d	5.60E-21
cluster465	4 Proteasome subunit alpha type-1-A	6.10E-22
cluster398	4 Proteasome subunit alpha type-3	8.80E-13
cluster371	4 Proteasome subunit alpha type-5	1.60E-14
cluster365	4 Proteasome subunit alpha type-6	1.80E-15
cluster444	4 Proteasome subunit alpha type-6	6.30E-19
cluster289	4 Proteasome subunit beta type-4	9.50E-36
cluster429	4 Proteasome subunit beta type-5-A	1.30E-61
cluster657	4 Proteasome subunit beta type-7-A	5.70E-49
cluster432	4 Protein CHROMATIN REMODELING 19 {ECO:0000303   PubMed:16547115}	3.10E-19
cluster237	4 Protein CHROMATIN REMODELING 8 {ECO:0000303   PubMed:16547115}	3.50E-17
cluster718	4 Protein gamma response 1	4.90E-04

cluster125	4 Protein HIRA	1.10E-05
cluster216	4 Protein PELOTA 2	3.80E-28
cluster450	4 Protein PEROXIN-4	3.50E-11
cluster150	4 Protein RER1C	7.00E-28
cluster613	4 Protein S-acyltransferase 24	1.40E-06
cluster484	4 Protein SGT1 homolog	4.80E-07
cluster140	4 Protein STICHEL	9.00E-21
cluster610	4 Protein SUPPRESSOR OF PHYA-105 1	3.10E-06
cluster500	4 Protein transport protein Sec61 subunit alpha	1.30E-119
cluster643	4 Proton pump-interactor 1	1.10E-04
cluster267	4 Pumilio homolog 15	1.10E-07
cluster375	4 Putative 60S ribosomal protein L18-1	3.00E-17
cluster655	4 Putative 60S ribosomal protein L37a-1	3.90E-17
cluster64	4 Putative cyclin-F1-1	2.80E-03
cluster683	4 Putative glycine--tRNA ligase, cytoplasmic	8.20E-92
cluster757	4 Putative leucine aminopeptidase 1	1.70E-45
cluster309	4 Putative nucleolar protein 5-3	1.70E-45
cluster593	4 Putative nucleolar protein 5-3	8.20E-54
cluster264	4 Putative rRNA 2'-O-methyltransferase fibrillar in 3	3.10E-71
cluster270	4 Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 2	1.50E-18
cluster474	4 Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 2	6.30E-21
cluster751	4 Putative SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 3-like 3	3.20E-23

cluster197	4	Putative WEB family protein At1g65010, chloroplastic	1.30E-05
cluster147	4	Putative zinc finger CCCH domain-containing protein 57	1.00E-09
cluster412	4	Pyruvate dehydrogenase E1 component subunit alpha	8.00E-34
cluster618	4	Rac-like GTP-binding protein ARAC10	1.40E-26
cluster769	4	Rac-like GTP-binding protein ARAC10	7.90E-34
cluster114	4	Rac-like GTP-binding protein ARAC7	7.80E-24
cluster86	4	Ras-related protein Rab7	1.40E-22
cluster555	4	Ras-related protein RABA4c	2.20E-43
cluster392	4	Ras-related protein RABA6b	4.70E-26
cluster452	4	Ras-related protein RABB1a	1.90E-35
cluster205	4	Ras-related protein RABC1	1.40E-03
cluster163	4	Ras-related protein RABH1c	4.50E-34
cluster266	4	Receptor-like protein kinase	4.70E-10
cluster210	4	Receptor-like protein kinase 5	5.40E-04
cluster336	4	Regulator of nonsense transcripts 1 homolog	2.10E-36
cluster540	4	Regulator of nonsense transcripts 1 homolog	8.90E-25
cluster710	4	Replication factor C subunit 1	4.90E-57
cluster526	4	Replication factor C subunit 2	1.80E-42
cluster247	4	Replication factor C subunit 5	5.10E-52
cluster434	4	Replication protein A 70 kDa DNA-binding subunit B	1.50E-58
cluster370	4	Ribonuclease H2 subunit A	6.50E-35
cluster31	5	Ribonuclease II, chloroplastic/mitochondrial	3.20E-20
cluster236	4	Ribulose-phosphate 3-epimerase, cytoplasmic isoform	1.80E-20
cluster513	4	RNA polymerase II C-terminal domain phosphatase-like 3	3.30E-10



cluster707	4 RNA pseudouridine synthase 4, mitochondrial	3.10E-08
cluster608	4 RuvB-like protein 1	2.10E-113
cluster436	4 SEC1 family transport protein SLY1	2.10E-54
cluster766	4 Serine/arginine-rich-splicing factor SR34	1.40E-03
cluster590	4 Serine/threonine protein phosphatase 2A 55 kDa regulatory subunit B beta isoform	1.70E-13
cluster38	5 Serine/threonine-protein kinase ATM	1.10E-41
cluster604	4 Serine/threonine-protein kinase Nek2	1.60E-21
cluster565	4 Serine/threonine-protein kinase Nek4	2.50E-03
cluster609	4 Serine/threonine-protein kinase UCNL {ECO:0000305}	1.50E-04
cluster153	4 Serine/threonine-protein phosphatase PP1 isozyme 1	9.60E-52
cluster475	4 Serine/threonine-protein phosphatase PP1 isozyme 1	5.50E-43
cluster274	4 Serine/threonine-protein phosphatase PP2A catalytic subunit	3.20E-71
cluster157	4 Serine--tRNA ligase	7.40E-98
cluster477	4 Signal recognition particle 54 kDa protein 2	1.50E-96
cluster29	5 Small RNA degrading nuclease 1	2.70E-16
cluster167	4 Soluble inorganic pyrophosphatase 1, chloroplastic	2.70E-48
cluster151	4 Structural maintenance of chromosomes protein 1	1.20E-13
cluster173	4 Structural maintenance of chromosomes protein 3	2.10E-26
cluster688	4 Structural maintenance of chromosomes protein 4	6.30E-35
cluster414	4 Structural maintenance of chromosomes protein 5	2.50E-83
cluster737	4 SUMO-activating enzyme subunit 2	3.50E-10
cluster727	4 Superoxide dismutase [Fe] 2, chloroplastic	3.60E-27
cluster172	4 TATA-box-binding protein	1.20E-74

cluster179	4 T-complex protein 1 subunit alpha	3.10E-20
cluster641	4 T-complex protein 1 subunit alpha	1.40E-47
cluster722	4 T-complex protein 1 subunit alpha	2.10E-46
cluster534	4 T-complex protein 1 subunit gamma	3.00E-56
cluster603	4 T-complex protein 1 subunit gamma	8.90E-42
cluster145	4 Thioredoxin reductase NTRC	1.40E-61
cluster447	4 Thioredoxin-like protein YLS8	2.60E-15
cluster330	4 Transcription elongation factor SPT4 homolog 1	5.00E-17
cluster76	4 Transcription factor MYB29	4.50E-24
cluster638	4 Transcription factor MYB98	1.70E-09
cluster706	4 Transcription initiation factor IIB-1	9.00E-45
cluster183	4 Transcription initiation factor TFIIID subunit 1	8.30E-05
cluster311	4 Transcription initiation factor TFIIID subunit 1	2.50E-07
cluster408	4 Transcription initiation factor TFIIID subunit 2	1.30E-35
cluster146	4 Transcription initiation factor TFIIID subunit 5	5.20E-08
cluster405	4 Transcription initiation factor TFIIID subunit 5	2.90E-09
cluster650	4 Transcription initiation factor TFIIID subunit 6b	4.60E-45
cluster596	4 Transketolase, chloroplastic	1.80E-76
cluster667	4 Translation factor GUF1 homolog, chloroplastic {ECO:0000255   HAMAP-Rule:MF_03138}	2.90E-11
cluster724	4 Translation initiation factor IF-2, chloroplastic	4.70E-13
cluster556	4 Triosephosphate isomerase, cytosolic	5.20E-32
cluster265	4 tRNA (guanine-N(7)-)-methyltransferase {ECO:0000255   HAMAP- Rule:MF_03055}	4.00E-40

cluster280	4	tRNA-splicing ligase RtcB homolog {ECO:0000255   HAMAP- Rule:MF_03144}	6.90E-113
cluster419	4	Tubulin alpha-6 chain	2.70E-172
cluster606	4	Tubulin beta-3 chain	9.20E-64
cluster423	4	Ubiquitin carboxyl-terminal hydrolase 16	4.30E-18
cluster303	4	Ubiquitin-conjugating enzyme E2 19	7.80E-21
cluster625	4	Ubiquitin-conjugating enzyme E2 35	2.60E-17
cluster570	4	Ubiquitin-conjugating enzyme E2 5	5.80E-19
cluster92	4	Ubiquitin-conjugating enzyme E2 7	8.50E-37
cluster143	4	UDP-N-acetylglucosamine diphosphorylase 2	1.80E-38
cluster577	4	UTP--glucose-1-phosphate uridylyltransferase	1.40E-108
cluster334	4	Valine--tRNA ligase	1.40E-191
cluster80	4	Vesicle-associated protein 1-2	2.20E-04
cluster271	4	V-type proton ATPase subunit a3	2.20E-91
cluster605	4	V-type proton ATPase subunit B 2	3.30E-171
cluster93	4	WEB family protein At5g16730, chloroplastic	8.00E-04
cluster703	4	Zinc finger CCCH domain-containing protein 21	2.20E-17
cluster747	4	Zinc finger CCCH domain-containing protein 31	1.20E-25
cluster66	4	Zinc finger CCCH domain-containing protein 4	2.50E-38

APÊNDICE 11: ANOTAÇÕES FUNCIONAIS DE GENES EXCLUSIVOS DE ILBN-2

Query	Predicted Name	Best OG	COG Category	EggNOG HMM Description
<b>ILBN2_0001p00060</b>	FG07018.1	OPNRK   2.7e-23   83.0	A	Small nuclear ribonucleoprotein
<b>ILBN2_0662p19380</b>	PAB1	OPFWK   1e-237   792.3	A	Binds the poly(A) tail of mRNA. Appears to be an important mediator of the multiple roles of the poly(A) tail in mRNA biogenesis, stability and translation. In the nucleus, involved in both mRNA cleavage and polyadenylation. Is also required for efficient mRNA export to the cytoplasm. Acts in concert with a poly(A)-specific nuclease (PAN) to affect poly(A) tail shortening, which may occur concomitantly with either nucleocytoplasmic mRNA transport or translational initiation. In the cytoplasm, stimulates translation initiation and regulates mRNA decay through translation termination-coupled poly(A) shortening, probably mediated by PAN (By similarity)
<b>ILBN2_0022p01660</b>	CFT1	OPFX4   1.1e-28   100.2	A	RNA-binding component of the cleavage and polyadenylation factor (CPF) complex, which plays a key role in polyadenylation-dependent pre-mRNA 3'-end formation and cooperates with cleavage factors including the CFIA complex and NAB4 CFIB. Involved in poly(A) site recognition. May be

<b>ILBN2_0098p06090</b>	ROK1	OPG11   7.1e-57   194.4	A	involved in coupling transcription termination and mRNA 3'-end formation
<b>ILBN2_0375p14380</b>	PGUG_03037	OPGAG   1e-43   150.0	A	ATP-dependent RNA helicase involved in 40S ribosomal subunit biogenesis. Required for the processing and cleavage of 35S pre-rRNA at sites A0, A1, and A2, leading to mature 18S rRNA
<b>ILBN2_0524p17220</b>	FGSG_06799	OPGRH   6e-255   850.4	A	helicase
<b>ILBN2_0701p19930</b>	FG10391.1	OPH2K   0.0   1360.4	A	5'-3' exoribonuclease
<b>ILBN2_0174p08960</b>	HAS1	OPH4U   4.6e-174   581.5	A	RNA helicase
<b>ILBN2_0803p21370</b>	HAS1	OPH4U   8.8e-48   164.1	A	ATP-dependent RNA helicase involved in 40S ribosomal subunit biogenesis. Required for the processing and cleavage of 35S pre-rRNA at sites A0, A1, and A2, leading to mature 18S rRNA
<b>ILBN2_0775p20980</b>	FG02781.1	OPHEF   1.5e-44   153.5	A	ATP-dependent RNA helicase involved in 40S ribosomal subunit biogenesis. Required for the processing and cleavage of 35S pre-rRNA at sites A0, A1, and A2, leading to mature 18S rRNA
<b>ILBN2_0017p01360</b>	PRP8	OPHSR   1.1e-188   630.5	A	helicase (ski2)
<b>ILBN2_0628p18940</b>	PRP8	OPHSR   1.1e-33   116.3	A	factor 8
<b>ILBN2_0819p21570</b>	PRP8	OPHSR   2.6e-25   88.5	A	factor 8
<b>ILBN2_0550p17630</b>	CWC24	OPJWC   1.1e-24   88.6	A	factor 8
<b>ILBN2_0028p01990</b>	CBC2	OPN15   1.3e-17   65.7	A	factor cwc24
<b>ILBN2_0413p15160</b>	CEF1	OPFBZ   2.4e-55   189.6	A, D	nuclear cap-binding protein subunit 2
<b>ILBN2_0813p21510</b>	ADA2	OPFFS   4.3e-68   231.3	B	factor cef-1
<b>ILBN2_0183p09320</b>	PGUG_01521	OPJWH   1.5e-22   81.5	B	SAGA complex subunit Ada2
<b>ILBN2_0243p11140</b>	PGUG_04200	OPH47   1.6e-193   646.5	B, D, L, T	PHD finger domain protein
				histone acetylase complex subunit Paf400

<b>ILBN2_0108p06510</b>	LIA1		OPJAQ   2.2e-81   274.4	C	Catalyzes the hydroxylation of the N(6)-(4-aminobutyl)- L-lysine intermediate to form hypusine, an essential post-translational modification only found in mature eIF-5A factor (By similarity)
<b>ILBN2_0257p11570</b>	CTK1		OPGKY   1.8e-25   90.8	D	CMGC CDK CRK7 protein kinase
<b>ILBN2_0052p03650</b>	CTK1		OPGKY   8.3e-21   75.4	D	CMGC CDK CRK7 protein kinase
<b>ILBN2_0277p12190</b>	CDC		OPGWY   1.5e-74   252.3	D	Cell division control protein
<b>ILBN2_0780p21040</b>	Exodeoxyribonuclease alpha	v	OPURM   2.9e-77   261.5	D	Pfam:DUF889
<b>ILBN2_0617p18780</b>	Exodeoxyribonuclease alpha	v	OPURM   3.6e-102   344.0	D	Pfam:DUF889
<b>ILBN2_0617p18790</b>			OQDSN   1.1e-103   348.7	D	helicase-like protein
<b>ILBN2_0017p01450</b>			OPMC5   2.9e-31   110.4	H	Pantetheine-phosphate adenylyltransferase
<b>ILBN2_0290p12440</b>	FG00688.1		OPGMX   1.1e-67   229.6	I	phosphoinositide phosphatase
<b>ILBN2_0815p21530</b>	FG09744.1		OPFYD   1.4e-142   477.8	J	tRNA (Cytosine-5-)-methyltransferase
<b>ILBN2_0347p13850</b>	DPH5		OPGKF   1.1e-89   301.6	J	S-adenosyl-L-methionine-dependent methyltransferase that catalyzes the trimethylation of the amino group of the modified target histidine residue in translation elongation factor 2 (EF-2), to form an intermediate called diphthine. The three successive methylation reactions represent the second step of diphthamide biosynthesis (By similarity)
<b>ILBN2_0745p20580</b>	ILS1		OPGN9   8.8e-95   319.5	J	Isoleucyl-tRNA synthetase
<b>ILBN2_0897p22450</b>	RPS3		OPGWV   7.8e-71   239.3	J	40s ribosomal protein s3
<b>ILBN2_0325p13380</b>	RLP24		OPHOY   5.5e-46   158.1	J	ribosome biogenesis protein RLP24
<b>ILBN2_0042p03150</b>	PGUG_04190		OPHNI   2.7e-30   105.5	J	protein kinase Gcn2
<b>ILBN2_0733p20400</b>	FG06850.1		OPI3F   1.1e-222   742.9	J	Valyl-trna synthetase
<b>ILBN2_0313p12980</b>	FG04476.1		OPPDN   9e-27   95.1	J	40S ribosomal protein S24

<b>ILBN2_0106p06450</b>	GAR1	OPPNM   1.5e-29   104.5	J	Required for ribosome biogenesis. Part of a complex which catalyzes pseudouridylation of rRNA. This involves the isomerization of uridine such that the ribose is subsequently attached to C5, instead of the normal N1. Pseudouridine ( psi ) residues may serve to stabilize the conformation of rRNAs
<b>ILBN2_0028p02000</b>	RPL34	OPPQN   1.5e-26   93.9	J	60S ribosomal protein L34
<b>ILBN2_0338p13620</b>	NHP2	OPQ50   1.6e-32   114.2	J	ribosome biogenesis protein Nhp2
<b>ILBN2_0777p21010</b>	FG07336.1	OQDMR   4e-32   112.6	J	peptidyl-tRNA hydrolase domain protein
<b>ILBN2_0428p15490</b>	FG05627.1	OPFMF   7.9e-125   418.8	K	dna-directed rna polymerase
<b>ILBN2_0708p20030</b>	PGUG_03709	OPG11   1.9e-90   305.2	K	DNA-directed RNA Polymerase
<b>ILBN2_0042p03120</b>	PGUG_03709	OPG11   1e-126   425.4	K	DNA-directed RNA Polymerase
<b>ILBN2_0542p17510</b>	FG09860.1	OPH1P   0.0   1399.7	K	DNA-directed RNA Polymerase
<b>ILBN2_0760p20790</b>	FG07306.1	OPH8Z   1.2e-81   276.3	K	SNF2 family ATP-dependent chromatin-remodeling factor snf21
<b>ILBN2_0867p22120</b>	ESF2	OPHQE   3.7e-55   188.5	K	Involved in the small subunit (SSU) processome assembly and function, and in the 18S rRNA synthesis. Required for the early cleavages at sites A0, A1 and A2 (By similarity)
<b>ILBN2_0128p07380</b>	PGUG_02470	OPHZ3   7.9e-21   74.2	K	cellular morphogenesis regulator DopA
<b>ILBN2_0220p10460</b>	RPC40	OPI11   4.8e-87   293.1	K	DNA-directed RNA polymerases I and III
<b>ILBN2_0277p12200</b>	SFP1	OPI5F   2.4e-16   61.1	K	Transcription factor
<b>ILBN2_0865p22100</b>	BDP1	OPIBH   4.4e-31   109.8	K	transcription factor (TFIIB)
<b>ILBN2_0642p19100</b>	MOT1	OPIIR   1.2e-146   491.4	K	factor mot1
<b>ILBN2_0250p11330</b>	MOT1	OPIIR   1.8e-13   49.9	K	factor mot1
<b>ILBN2_0232p10800</b>	CDC73	OPK36   1.4e-19   71.2	K	Pol II transcription elongation factor subunit Cdc73

<b>ILBN2_0550p17660</b>	TOA1	OPM7H   9e-26   92.6	K	transcription factor TFIIA complex subunit Toa1
<b>ILBN2_0452p15980</b>	NCB2	OPPQR   2.3e-41   142.2	K	transcription corepressor activity
<b>ILBN2_0417p15240</b>	RPC11	OPQ6F   5.8e-33   114.4	K	DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates (By similarity)
<b>ILBN2_0117p06870</b>	FG00303.1	OPQMI   9.7e-22   78.5	K	transcriptional elongation regulator Elc1 Elongin C
<b>ILBN2_0294p12530</b>	PGUG_02820	OPFQP   2.5e-52   178.9	L	replication factor c
<b>ILBN2_0059p03980</b>	TEL1	OPFW9   4.1e-82   277.5	L	Serine threonine protein kinase which activates checkpoint signaling upon genotoxic stresses such as ionizing radiation (IR), ultraviolet light (UV), or DNA replication stalling, thereby acting as a DNA damage sensor. Recognizes the substrate consensus sequence ST - Q. Phosphorylates histone H2A to form H2AS128ph (gamma-H2A) at sites of DNA damage, involved in the regulation of DNA damage response mechanism. Required for the control of telomere length and genome stability (By similarity)
<b>ILBN2_0789p21190</b>	Retrotransposon protein	OPGPZ   1.1e-28   100.7	L	to reverse transcriptase
<b>ILBN2_0782p21070</b>	Retrotransposon protein	OPGPZ   1.1e-44   153.8	L	to reverse transcriptase
<b>ILBN2_0518p17110</b>	Retrotransposon protein	OPGPZ   1.1e-60   206.8	L	to reverse transcriptase
<b>ILBN2_0799p21320</b>	Retrotransposon protein	OPGPZ   1.2e-101   342.5	L	to reverse transcriptase
<b>ILBN2_0843p21850</b>	Retrotransposon protein	OPGPZ   1.3e-109   368.9	L	to reverse transcriptase
<b>ILBN2_0679p19590</b>	Retrotransposon protein	OPGPZ   1.3e-35   123.6	L	to reverse transcriptase
<b>ILBN2_0889p22370</b>	Retrotransposon protein	OPGPZ   1.3e-49   170.1	L	to reverse transcriptase
<b>ILBN2_0876p22220</b>	Retrotransposon protein	OPGPZ   1.3e-83   282.8	L	to reverse transcriptase



<b>ILBN2_0757p20750</b>	Retrotransposon protein	OPGPZ   1.3e-96   325.8	L	to reverse transcriptase
<b>ILBN2_0625p18910</b>	Retrotransposon protein	OPGPZ   1.4e-32   113.7	L	to reverse transcriptase
<b>ILBN2_0900p22480</b>	Retrotransposon protein	OPGPZ   1.4e-42   146.8	L	to reverse transcriptase
<b>ILBN2_0804p21390</b>	Retrotransposon protein	OPGPZ   1.4e-51   176.6	L	to reverse transcriptase
<b>ILBN2_0704p19980</b>	Retrotransposon protein	OPGPZ   1.4e-81   276.0	L	to reverse transcriptase
<b>ILBN2_0714p20110</b>	Retrotransposon protein	OPGPZ   1.5e-175   587.5	L	to reverse transcriptase
<b>ILBN2_0731p20370</b>	Retrotransposon protein	OPGPZ   1.5e-20   73.7	L	to reverse transcriptase
<b>ILBN2_0265p11790</b>	Retrotransposon protein	OPGPZ   1.5e-38   133.4	L	to reverse transcriptase
<b>ILBN2_0615p18760</b>	Retrotransposon protein	OPGPZ   1.6e-58   199.6	L	to reverse transcriptase
<b>ILBN2_0890p22380</b>	Retrotransposon protein	OPGPZ   1.6e-66   226.1	L	to reverse transcriptase
<b>ILBN2_0835p21750</b>	Retrotransposon protein	OPGPZ   1.6e-71   242.7	L	to reverse transcriptase
<b>ILBN2_0644p19130</b>	Retrotransposon protein	OPGPZ   1.7e-236   789.4	L	to reverse transcriptase
<b>ILBN2_0898p22460</b>	Retrotransposon protein	OPGPZ   1e-25   90.9	L	to reverse transcriptase
<b>ILBN2_0853p21950</b>	Retrotransposon protein	OPGPZ   1e-29   104.1	L	to reverse transcriptase
<b>ILBN2_0753p20700</b>	Retrotransposon protein	OPGPZ   1e-40   140.6	L	to reverse transcriptase
<b>ILBN2_0879p22260</b>	Retrotransposon protein	OPGPZ   1e-47   163.8	L	to reverse transcriptase
<b>ILBN2_0805p21400</b>	Retrotransposon protein	OPGPZ   2.1e-227   759.3	L	to reverse transcriptase
<b>ILBN2_0680p19600</b>	Retrotransposon protein	OPGPZ   2.1e-44   152.8	L	to reverse transcriptase
<b>ILBN2_0730p20360</b>	Retrotransposon protein	OPGPZ   2.1e-72   245.6	L	to reverse transcriptase
<b>ILBN2_0771p20930</b>	Retrotransposon protein	OPGPZ   2.2e-60   205.8	L	to reverse transcriptase
<b>ILBN2_0892p22400</b>	Retrotransposon protein	OPGPZ   2.2e-69   235.6	L	to reverse transcriptase
<b>ILBN2_0327p13400</b>	Retrotransposon protein	OPGPZ   2.3e-251   838.7	L	to reverse transcriptase
<b>ILBN2_0868p22130</b>	Retrotransposon protein	OPGPZ   2.3e-39   136.1	L	to reverse transcriptase
<b>ILBN2_0703p19950</b>	Retrotransposon protein	OPGPZ   2.3e-46   159.3	L	to reverse transcriptase
<b>ILBN2_0769p20910</b>	Retrotransposon protein	OPGPZ   2.4e-105   354.8	L	to reverse transcriptase
<b>ILBN2_0833p21730</b>	Retrotransposon protein	OPGPZ   2.4e-201   672.9	L	to reverse transcriptase
<b>ILBN2_0753p20690</b>	Retrotransposon protein	OPGPZ   2.4e-27   96.3	L	to reverse transcriptase
<b>ILBN2_0702p19940</b>	Retrotransposon protein	OPGPZ   2.6e-61   208.9	L	to reverse transcriptase

<b>ILBN2_0246p11200</b>	Retrotransposon protein	OPGPZ   2.7e-45   155.8	L	to reverse transcriptase
<b>ILBN2_0875p22200</b>	Retrotransposon protein	OPGPZ   2.8e-63   215.4	L	to reverse transcriptase
<b>ILBN2_0656p19320</b>	Retrotransposon protein	OPGPZ   2.8e-68   231.9	L	to reverse transcriptase
<b>ILBN2_0866p22110</b>	Retrotransposon protein	OPGPZ   2.9e-99   334.6	L	to reverse transcriptase
<b>ILBN2_0811p21490</b>	Retrotransposon protein	OPGPZ   3.1e-37   129.0	L	to reverse transcriptase
<b>ILBN2_0719p20190</b>	Retrotransposon protein	OPGPZ   3.4e-30   105.7	L	to reverse transcriptase
<b>ILBN2_0895p22430</b>	Retrotransposon protein	OPGPZ   3.4e-68   231.7	L	to reverse transcriptase
<b>ILBN2_0861p22040</b>	Retrotransposon protein	OPGPZ   3.5e-30   105.7	L	to reverse transcriptase
<b>ILBN2_0818p21560</b>	Retrotransposon protein	OPGPZ   3.7e-41   142.1	L	to reverse transcriptase
<b>ILBN2_0246p11210</b>	Retrotransposon protein	OPGPZ   3.7e-44   152.0	L	to reverse transcriptase
<b>ILBN2_0877p22230</b>	Retrotransposon protein	OPGPZ   3.7e-60   205.0	L	to reverse transcriptase
<b>ILBN2_0882p22290</b>	Retrotransposon protein	OPGPZ   3e-36   125.8	L	to reverse transcriptase
<b>ILBN2_0558p17810</b>	Retrotransposon protein	OPGPZ   4.1e-186   622.4	L	to reverse transcriptase
<b>ILBN2_0800p21340</b>	Retrotransposon protein	OPGPZ   4.2e-37   128.6	L	to reverse transcriptase
<b>ILBN2_0782p21080</b>	Retrotransposon protein	OPGPZ   4.4e-129   433.4	L	to reverse transcriptase
<b>ILBN2_0857p21990</b>	Retrotransposon protein	OPGPZ   4.4e-24   85.5	L	to reverse transcriptase
<b>ILBN2_0878p22240</b>	Retrotransposon protein	OPGPZ   4.7e-66   224.6	L	to reverse transcriptase
<b>ILBN2_0820p21600</b>	Retrotransposon protein	OPGPZ   5.4e-69   234.3	L	to reverse transcriptase
<b>ILBN2_0776p21000</b>	Retrotransposon protein	OPGPZ   5.5e-85   287.3	L	to reverse transcriptase
<b>ILBN2_0265p11800</b>	Retrotransposon protein	OPGPZ   5.6e-39   134.8	L	to reverse transcriptase
<b>ILBN2_0634p19000</b>	Retrotransposon protein	OPGPZ   5.6e-69   234.3	L	to reverse transcriptase
<b>ILBN2_0879p22250</b>	Retrotransposon protein	OPGPZ   5.8e-49   167.9	L	to reverse transcriptase
<b>ILBN2_0886p22340</b>	Retrotransposon protein	OPGPZ   6.1e-32   111.5	L	to reverse transcriptase
<b>ILBN2_0771p20940</b>	Retrotransposon protein	OPGPZ   6.1e-53   181.1	L	to reverse transcriptase
<b>ILBN2_0820p21590</b>	Retrotransposon protein	OPGPZ   6.3e-43   147.9	L	to reverse transcriptase
<b>ILBN2_0873p22180</b>	Retrotransposon protein	OPGPZ   6.3e-59   200.9	L	to reverse transcriptase
<b>ILBN2_0684p19690</b>	Retrotransposon protein	OPGPZ   6.3e-82   277.2	L	to reverse transcriptase
<b>ILBN2_0862p22050</b>	Retrotransposon protein	OPGPZ   6.4e-66   224.1	L	to reverse transcriptase

<b>ILBN2_0810p21470</b>	Retrotransposon protein	OPGPZ   6.5e-45   154.5	L	to reverse transcriptase
<b>ILBN2_0800p21330</b>	Retrotransposon protein	OPGPZ   6.6e-74   250.6	L	to reverse transcriptase
<b>ILBN2_0579p18150</b>	Retrotransposon protein	OPGPZ   6.7e-42   144.5	L	to reverse transcriptase
<b>ILBN2_0863p22070</b>	Retrotransposon protein	OPGPZ   6e-27   95.0	L	to reverse transcriptase
<b>ILBN2_0730p20350</b>	Retrotransposon protein	OPGPZ   7.2e-19   68.2	L	to reverse transcriptase
<b>ILBN2_0439p15710</b>	Retrotransposon protein	OPGPZ   7.3e-95   320.0	L	to reverse transcriptase
<b>ILBN2_0842p21840</b>	Retrotransposon protein	OPGPZ   7.7e-31   107.9	L	to reverse transcriptase
<b>ILBN2_0893p22410</b>	Retrotransposon protein	OPGPZ   7.8e-27   94.6	L	to reverse transcriptase
<b>ILBN2_0265p11810</b>	Retrotransposon protein	OPGPZ   8.2e-20   71.3	L	to reverse transcriptase
<b>ILBN2_0608p18670</b>	Retrotransposon protein	OPGPZ   8.5e-107   359.6	L	to reverse transcriptase
<b>ILBN2_0891p22390</b>	Retrotransposon protein	OPGPZ   8e-94   316.6	L	to reverse transcriptase
<b>ILBN2_0821p21610</b>	Retrotransposon protein	OPGPZ   9.8e-51   173.8	L	to reverse transcriptase
<b>ILBN2_0888p22360</b>	Retrotransposon protein	OPGPZ   9e-34   117.6	L	to reverse transcriptase
<b>ILBN2_0885p22330</b>	Retrotransposon protein	OPGPZ   9e-53   180.5	L	to reverse transcriptase
<b>ILBN2_0511p17030</b>	POL2	OPGZD   0.0   1026.7	L	DNA polymerase
<b>ILBN2_0841p21820</b>	MSH6	OPH6E   2.2e-76   258.7	L	DNA mismatch repair protein Msh6
<b>ILBN2_0554p17720</b>	RVB2	OPHPG   6.5e-167   557.0	L	DNA helicase which participates in several chromatin remodeling complexes, including the SWR1 and the INO80 complexes. The SWR1 complex mediates the ATP-dependent exchange of histone H2A for the H2A variant HZT1 leading to transcriptional regulation of selected genes by chromatin remodeling. The INO80 complex remodels chromatin by shifting nucleosomes and is involved in DNA repair. Also involved in pre-rRNA processing (By similarity)
<b>ILBN2_0884p22320</b>		OPJX6   5.1e-32   111.3	L	protein Missing 5 prime End
<b>ILBN2_0605p18630</b>	transposon protein	OPX77   4.6e-38   132.1	L	transposon protein

<b>ILBN2_0555p17750</b>		OPUCJ   8.7e-26   91.0	M, P, T	Ankyrin repeat
<b>ILBN2_0176p09090</b>	PGUG_00775	OPFGD   2.4e-52   179.2	O	proteasome regulatory
<b>ILBN2_0773p20960</b>	RPN1	OPFH6   7.5e-244   812.8	O	26S proteasome regulatory subunit
<b>ILBN2_0284p12330</b>	FG07222.1	OPFX7   2.4e-122   411.1	O	Aaa family atpase
<b>ILBN2_0305p12790</b>	FG09689.1	OPFZC   1.3e-53   183.1	O	ubiquitin-activating enzyme E1
<b>ILBN2_0748p20630</b>	FG09689.1	OPFZC   1.5e-45   156.4	O	ubiquitin-activating enzyme E1
<b>ILBN2_0482p16590</b>	FG09689.1	OPFZC   2e-53   182.4	O	ubiquitin-activating enzyme E1
<b>ILBN2_0332p13480</b>	FG09689.1	OPFZC   3.2e-21   75.8	O	ubiquitin-activating enzyme E1
<b>ILBN2_0736p20460</b>	FGSG_11966	OPGAX   1.9e-32   113.3	O	ubiquitin conjugation factor E4
<b>ILBN2_0076p04860</b>	FG00564.1	OPGK8   2.9e-50   172.4	O	The proteasome is a multicatalytic proteinase complex which is characterized by its ability to cleave peptides with Arg, Phe, Tyr, Leu, and Glu adjacent to the leaving group at neutral or slightly basic pH. The proteasome has an ATP-dependent proteolytic activity (By similarity)
<b>ILBN2_0012p00920</b>	FG07110.1	OPGTA   2.2e-103   346.8	O	Geranylgeranyltransferase beta subunit
<b>ILBN2_0467p16310</b>	HRD1	OPGY0   1.8e-49   169.8	O	Ring finger protein
<b>ILBN2_0059p03970</b>	HULA	OPH9A   3.4e-38   131.6	O	E3 ubiquitin-protein ligase
<b>ILBN2_0464p16210</b>	GDI1	OPHJ6   4.9e-51   174.5	O	rab gdp-dissociation inhibitor
<b>ILBN2_0419p15280</b>	STI1	OPIT1   1.4e-64   220.0	O	Heat shock protein
<b>ILBN2_0177p09120</b>	PGUG_04025	OPKFU   1.2e-25   91.7	O	DnaJ domain protein
<b>ILBN2_0002p00170</b>	FG06532.1	OPMV8   6.6e-31   108.7	O	Peptidyl-prolyl cis-trans isomerase
<b>ILBN2_0493p16780</b>	UBC12	OPN5Z   5.7e-39   134.5	O	NEDD8-conjugating enzyme UBC12
<b>ILBN2_0092p05740</b>	FG06877.1	OPP17   5.8e-36   124.9	O	ubiquitin conjugating enzyme
<b>ILBN2_0420p15290</b>		OPZ43   7.1e-26   93.0	O	Zinc finger, C3HC4 type (RING finger)
<b>ILBN2_0448p15890</b>		OPZNE   1.5e-17   65.4	O	Zinc finger, C3HC4 type (RING finger)
<b>ILBN2_0509p17000</b>	FG04095.1	OPFD4   1.3e-139   468.3	P	Na H antiporter

<b>ILBN2_0471p16380</b>	NEO1	OPG6X   2.8e-41   142.3	P	P-type ATPase
<b>ILBN2_0088p05490</b>	FG06743.1	OPGA7   2.3e-07   29.6	P	phospholipid-transporting ATPase
<b>ILBN2_0483p16600</b>	FG07466.1	OPH07   1.2e-192   643.7	P	P-type ATPase
<b>ILBN2_0072p04600</b>	FG07466.1	OPH07   3e-255   851.0	P	P-type ATPase
<b>ILBN2_0069p04450</b>	PGUG_04776	OPGHX   1.6e-13   50.7	Q	(ABC) transporter
<b>ILBN2_0042p03140</b>	PGUG_04776	OPGHX   3.6e-85   287.9	Q	(ABC) transporter
<b>ILBN2_0566p17910</b>	FG10935.1	OPHN2   6.2e-70   237.0	Q	(ABC) transporter
<b>ILBN2_0685p19710</b>	MIDASIN	OPFRP   5.1e-40   137.0	S	Nuclear chaperone required for maturation and nuclear export of pre-60S ribosome subunits (By similarity)
<b>ILBN2_0280p12240</b>	FG07194.1	OPFXG   7.8e-34   118.3	S	(GTP)-binding protein
<b>ILBN2_0718p20170</b>	FG05698.1	OPGEI   9.1e-49   167.4	S	WD repeat protein
<b>ILBN2_0577p18120</b>	FG00732.1	OPGJP   2.2e-68   232.0	S	N-acetyltransferase 10
<b>ILBN2_0017p01420</b>	FG10152.1	OPI6A   3.3e-36   125.9	S	Unfolded protein response protein Orm1
<b>ILBN2_0081p05180</b>	PGUG_03061	OPJ6V   2.7e-31   110.1	S	Pyrimidine 5'-nucleotidase
<b>ILBN2_0135p07610</b>	PGUG_01675	OPNRQ   1e-37   130.8	S	Peptidyl-tRNA hydrolase
<b>ILBN2_0273p12060</b>	Retrotransposon protein	OQDNF   5.6e-27   94.7	S	Reverse transcriptase (RNA-dependent DNA polymerase)
<b>ILBN2_0795p21260</b>	PGUG_04710	OQEHI   2.7e-07   29.8	S	Lipid-binding protein
<b>ILBN2_0001p00030</b>	PKA2	OPGKH   4.3e-119   399.0	T	cAMP-dependent protein kinase catalytic subunit
<b>ILBN2_0118p06880</b>	FG05139.1	OPG8U   1.4e-151   505.5	T	(GTP)-binding protein
<b>ILBN2_0221p10510</b>	CNA1	OPGAU   1e-124   418.1	T	serine threonine-protein phosphatase
<b>ILBN2_0004p00330</b>	FG05406.1	OPGBR   8.4e-86   289.6	T	protein kinase NpkA
<b>ILBN2_0047p03410</b>		OPGS1   9.1e-48   163.3	T	Adenylate cyclase
<b>ILBN2_0177p09130</b>	PKC1	OPHX3   3.6e-119   400.4	T	protein kinase C
<b>ILBN2_0334p13520</b>	KIN3	OPIY3   2.8e-62   212.7	T	G2-specific protein kinase
<b>ILBN2_0717p20150</b>	FG06878.1	OPJ6N   3.5e-116   389.4	T	calcium calmodulin-dependent protein kinase

<b>ILBN2_0048p03500</b>	GLO2	OPJ9Y   2.3e-09   37.8	T	Hydroxyacylglutathione hydrolase
<b>ILBN2_0527p17250</b>		OPPPX   2.4e-59   202.5	T	Muskelin N-terminus
<b>ILBN2_0874p22190</b>		OPPPX   3e-57   195.6	T	Muskelin N-terminus
<b>ILBN2_0201p09850</b>		OPUAY   2.9e-21   76.7	T	Mitogen-activated protein kinase
<b>ILBN2_0220p10450</b>	TIM50	OPGG7   6.3e-32   112.2	U	Essential component of the TIM23 complex, a complex that mediates the translocation of transit peptide-containing proteins across the mitochondrial inner membrane. Required to direct preproteins in transit and direct them to the channel protein TIM23, and possibly facilitates transfer of the translocating proteins from the TOM complex to the TIM23 complex
<b>ILBN2_0569p17980</b>	FG01644.1	OPGQR   4.6e-47   162.2	U	TBC domain protein
<b>ILBN2_0109p06550</b>	FG06920.1	OPKEM   1e-38   133.8	U	ADP-ribosylation
<b>ILBN2_0459p16110</b>	FG02475.1	OPGTF   7.6e-25   88.7	Z	CTLH domain-containing protein
<b>ILBN2_0768p20900</b>	MYO2	OPI8B   1.1e-34   120.2	Z	class V myosin
<b>ILBN2_0412p15150</b>	MYO2	OPI8B   1.5e-39   136.3	Z	class V myosin
<b>ILBN2_0072p04620</b>	TUB2	OPIDE   1.5e-265   881.9	Z	Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha chain

APÊNDICE 12: TABELA DE CATEGORIAS COG

<b>PROCESSOS CELULARES E SINALIZAÇÃO</b>	[D] Controle do ciclo celular, divisão celular, partição de cromossomos
	[M] Biogênese de parede/membrana/envelope celular
	[N] Motilidade celular
	[O] Modificação pós-traducional, <i>turnover</i> de proteínas, e chaperonas
	[T] Mecanismos de transdução de sinal
	[U] Tráfego intracelular, secreção, e transporte de vesículas
	[V] Mecanismos de defesa
	[W] Estruturas extracelulares
	[Y] Estrutura nuclear
	[Z] Citoesqueleto
<b>PROCESSAMENTO E ARMAZENAMENTO DE INFORMAÇÃO</b>	[A] Processamento e modificação de RNA
	[B] Dinâmica e estrutura de cromatina
	[J] Tradução, estrutura ribossômica e biogênese
	[K] Transcrição
	[L] Replicação, recombinação e reparo
<b>METABOLISMO</b>	[C] Produção e conversão de energia
	[E] Transporte de aminoácido e metabolismo
	[F] Transporte de nucleotídeo e metabolismo
	[G] Transporte de carboidrato e metabolismo
	[H] Transporte de coenzima e metabolismo
	[I] Transporte de lipídeo e metabolismo
	[P] Transporte de ion orgânico e metabolismo
	[Q] Biossíntese, transporte, e catabolismo de metabólitos secundários
<b>POUCO CARACTERIZADAS</b>	[R] Somene função geral predita
	[S] Função desconhecida

APÊNDICE 13: ANOTAÇÕES DE GENES COMPARTILHADOS ENTRE ILBN-2 E BEOM-2 A PARTIR DE BLAST BIDIRECIONAL

Query	Predicted Name	Best OG	COG Category	EggNOG HMM Description
<b>ILBN2_0046p03370</b>		OPHKN   8e-21   74.8	A	MRNA cleavage and polyadenylation specificity factor complex subunit
<b>ILBN2_0071p04550</b>	FG01381.1	OPS3G   1.6e-25   89.7	A	Small nuclear ribonucleoprotein
<b>ILBN2_0087p05460</b>	FG01187.1	OPI89   1.4e-31   111.0	A	Splicing factor 3A subunit 2
<b>ILBN2_0225p10620</b>	FG05638.1	OPHFC   2.9e-69   235.2	A	Exchanges the guanine residue with 7-aminomethyl-7- deazaguanine in tRNAs with GU(N) anticodons (tRNA-Asp, -Asn, -His and -Tyr). After this exchange, a cyclopentendiol moiety is attached to the 7-aminomethyl group of 7-deazaguanine, resulting in the hypermodified nucleoside queuosine (Q) (7-(((4,5-cis- dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine) (By similarity)
<b>ILBN2_0013p01110</b>	HTA1	OPNTK   1.3e-15   58.6	B	Core component of nucleosome which plays a central role in DNA double strand break (DSB) repair.
<b>ILBN2_0059p03960</b>	Sep-01	OPH74   8.6e-27   94.9	B	Catalytic component of the COMPASS (Set1C) complex that specifically mono-, di- and trimethylates histone H3 to form H3K4me1 2 3, which subsequently plays a role in telomere length maintenance and transcription elongation regulation
<b>ILBN2_0781p21050</b>	HAT2	OPHNE   3.7e-89   300.6	B	histone acetyltransferase type b subunit 2
<b>ILBN2_0093p05890</b>	FG10837.1	OPIEK   8.9e-43   147.9	J	tRNA pseudouridine synthase



<b>ILBN2_0127p07350</b>	FG06893.1	OPG18   7e-58   195.7	J	40S ribosomal protein S18
<b>ILBN2_0134p07590</b>	RPS9	OPG10   4.7e-51   174.2	J	40S ribosomal protein S9
<b>ILBN2_0134p07600</b>	FG08895.1	OPHM8   5.2e-23   82.5	J	60s ribosomal protein
<b>ILBN2_0154p08390</b>	FRS1	OPGUM   3.9e-121   406.5	J	Phenylalanyl-tRNA synthetase, beta
<b>ILBN2_0195p09690</b>	FG10845.1	OPPNW   4.2e-36   125.1	J	60S ribosomal protein I26
<b>ILBN2_0224p10600</b>	RPS15	OPIWK   8.8e-56   188.6	J	40s ribosomal protein s15
<b>ILBN2_0237p10970</b>	RPL30	OPPNR   2.5e-38   133.4	J	60s ribosomal protein I30
<b>ILBN2_0532p17320</b>	RPL32	OPNPG   2.1e-25   90.3	J	60S ribosomal protein L32
<b>ILBN2_0556p17780</b>	RPL12	OPN4A   4.7e-60   203.5	J	60S ribosomal protein L12
<b>ILBN2_0557p17800</b>	RPS1	OPGZY   7.3e-79   265.7	J	40S ribosomal protein S1
<b>ILBN2_0636p19020</b>	FG02764.1	OPH3G   2.1e-33   116.9	J	GTP-binding protein YchF
<b>ILBN2_0034p02580</b>	TAF10	OPRA4   1.9e-32   114.3	K	Transcription initiation factor TFIID subunit
<b>ILBN2_0063p04150</b>	PGUG_02088	OPHE0   2.1e-33   117.4	K	transcription initiation factor
<b>ILBN2_0084p05320</b>	RPB10	OPRZ2   8.9e-29   100.3	K	DNA-directed RNA polymerase
<b>ILBN2_0176p09050</b>	RPC34	OPFKU   3.4e-49   169.1	K	DNA-directed RNA polymerase III subunit
<b>ILBN2_0195p09700</b>	RPB6	OPQRP   6e-41   141.7	K	DNA-directed RNA polymerase
<b>ILBN2_0260p11680</b>	FG07476.1	OPQMN   9.5e-26   92.5	K	Transcription elongation factor
<b>ILBN2_0379p14440</b>	Retrotransposon protein	OPGPZ   1e-105   356.0	L	to reverse transcriptase
<b>ILBN2_0536p17420</b>	FG06339.1	OPG27   4.9e-67   227.8	L	DNA primase
<b>ILBN2_0589p18320</b>	FG10482.1	OPG5A   3.3e-57   195.1	L	Polynucleotide
<b>ILBN2_0604p18600</b>	OGG1	OPIA2   2.3e-63   215.4	L	N-glycosylase DNA lyase
<b>ILBN2_0632p18980</b>	MEC1	OPGIW   1.6e-97   328.6	L	Serine threonine protein kinase which activates checkpoint signaling upon genotoxic stresses such as ionizing radiation (IR), ultraviolet light (UV), or DNA replication stalling, thereby acting as a DNA damage sensor. Recognizes

					the substrate consensus sequence ST-Q. Recruited to DNA lesions in order to initiate the DNA repair by homologous recombination. Phosphorylates histone H2A to form H2AS128ph (gamma-H2A) at sites of DNA damage, also involved in the regulation of DNA damage response mechanism. Required for cell growth and meiotic recombination (By similarity)
<b>ILBN2_0661p19370</b>	Retrotransposon protein	OPGPZ   5.6e-27   95.1	L		to reverse transcriptase
<b>ILBN2_0804p21380</b>	Retrotransposon protein	OPGPZ   7.6e-74   250.4	L		to reverse transcriptase
<b>ILBN2_0816p21540</b>	Retrotransposon protein	OPGPZ   2.3e-60   205.7	L		to reverse transcriptase
<b>ILBN2_0887p22350</b>	Retrotransposon protein	OPGPZ   6.8e-99   333.4	L		to reverse transcriptase
<b>ILBN2_0052p03660</b>	ESF1	OPGQE   4.9e-25   89.2	F		Pre-rRNA processing protein Esf1
<b>ILBN2_0013p01020</b>	FG04487.1	OPFZ3   1.4e-47   163.7	G		Glucose transporter
<b>ILBN2_0052p03640</b>	FG06055.1	OPHC1   2.5e-163   545.3	G		Phosphoglycerate mutase
<b>ILBN2_0176p09080</b>	RKI1	OPGR7   2.1e-38   133.4	G		Ribose 5-phosphate isomerase
<b>ILBN2_0896p22440</b>	GPD	OPF9K   2e-89   300.8	G		Glyceraldehyde-3-phosphate dehydrogenase
<b>ILBN2_0132p07480</b>	PGUG_01834	OPGF7   2e-21   77.4	H		tRNA dimethylallyltransferase
<b>ILBN2_0031p02220</b>	ACC1	OPH59   2.8e-229   765.2	I		Acetyl-CoA carboxylase
<b>ILBN2_0052p03630</b>		OPK30   2.7e-107   360.6	I		Hydroxymethylglutaryl-coenzyme A synthase N terminal
<b>ILBN2_0398p14880</b>		OPHMB   2.2e-63   215.7	I		O-acyltransferase
<b>ILBN2_0183p09330</b>	CDC25	OPI8H   1e-58   200.7	D		Inducer phosphatase
<b>ILBN2_0271p12020</b>	FG02610.1	OPPWC   1.5e-58   200.5	D		Meiosis protein MEI2
<b>ILBN2_0001p00020</b>		OPG6N   1.6e-290   968.0	M		Chitin synthase

<b>ILBN2_0221p10520</b>	ALG2	OPFGA   2.1e-62   212.4	M	Alpha-1,3-mannosyltransferase alg-2
<b>ILBN2_0013p01090</b>	UBI4	OPHU0   1e-112   378.2	O	Ubiquitin
<b>ILBN2_0118p06890</b>	RPN8	OPHXE   9.9e-72   242.2	O	Proteasome regulatory subunit
<b>ILBN2_0176p09100</b>	PGUG_05283	OPH2Q   3.3e-79   268.2	O	RING and UBP finger domain protein
<b>ILBN2_0193p09590</b>	RPN1	OPFH6   1.1e-92   312.7	O	26S proteasome regulatory subunit
<b>ILBN2_0202p09900</b>	FG09635.1	OPFK5   4.1e-19   69.8	O	Gpi-anchor transamidase
<b>ILBN2_0319p13230</b>	FG02460.1	OPFI2   2.5e-162   543.2	P	Cation-transporting atpase
<b>ILBN2_0032p02340</b>	PTC2	OPGQI   9.7e-30   105.0	T	Phosphatase 2C
<b>ILBN2_0039p02890</b>	CNB1	OPG8E   5.8e-18   66.0	T	Calcineurin subunit B
<b>ILBN2_0505p16960</b>	STT4	OPHQ4   1.6e-136   458.1	T	Phosphatidylinositol 4-kinase
<b>ILBN2_0176p09040</b>	FG05891.1	OPIQY   2.1e-37   129.9	U	Inositol polyphosphate phosphatase
<b>ILBN2_0774p20970</b>	MIDASIN	OPFRP   1.9e-10   38.9	S	Nuclear chaperone required for maturation and nuclear export of pre-60S ribosome subunits (By similarity)
<b>ILBN2_0176p09110</b>	FG00732.1	OPGJP   4e-65   221.2	S	N-acetyltransferase 10
<b>ILBN2_0213p10240</b>	PGUG_03578	OPFCW   3.5e-79   267.9	S	Pseudouridine synthase
<b>ILBN2_0232p10810</b>		OPZUM   4.6e-32   112.5	S	Hydrolase domain containing
<b>ILBN2_0237p10980</b>	DRE2	OPIMK   8.5e-20   73.1	S	Component of the cytosolic iron-sulfur (Fe S) protein assembly machinery. Required for the maturation of extramitochondrial Fe S proteins (By similarity). Has anti- apoptotic effects in the cell. Involved in negative control of H(2)O(2)-induced cell death, probably by tethering the pro- apoptotic factor tah18 in the cytoplasm in the absence of oxidative stress (By similarity)
<b>ILBN2_0447p15860</b>	FG06615.1	OPJ7H   2.3e-23   83.8	S	F-box and WD40 domain protein
<b>ILBN2_0740p20520</b>	FG08784.1	OPGDI   1.6e-52   180.1	S	Transport protein yif1

