



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (ZOOLOGIA)

IDALIO DO AMARANTO COSTA

SEQUENCIAMENTO E ANÁLISE COMPARATIVA DO DNA DE
***Cyphoderus similis* FOLSOM E *Cyphoderus innominatus* MILLS**

JOÃO PESSOA – PB

2019

IDALIO DO AMARANTO COSTA

SEQUENCIAMENTO E ANÁLISE COMPARATIVA DO DNA DE
***Cyphoderus similis* FOLSOM E *Cyphoderus innominatus* MILLS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas (Zoologia) da Universidade Federal da Paraíba, como requisito para obtenção do título de Mestre em Ciências Biológicas (Zoologia).

Dr. Douglas Zeppelini

Orientador

Dr. Guilherme C. Oliveira

Co-orientador

JOÃO PESSOA – PB

2019

Catálogo na publicação
Seção de Catálogo e Classificação

A485s Amaranto, Idalio.

SEQUENCIAMENTO E ANÁLISE COMPARATIVA DO DNA DE
Cyphoderus similis FOLSOM E Cyphoderus innominatus
MILLS / Idalio Amaranto. - João Pessoa, 2019.
99 f. : il.

Orientação: Douglas Zeppelini.

Coorientação: Guilherme C Oliveira.

Dissertação (Mestrado) - UFPB/CCEN.

1. Collembola. 2. Genômica. 3. Draft Genome. 4. Next
Generation Sequencing (NGS). 5. Filogenia Molecular. I.
Zeppelini, Douglas. II. Oliveira, Guilherme C. III.
Título.

UFPB/BC

IDALIO DO AMARANTO COSTA

**SEQUENCIAMENTO E ANÁLISE COMPARATIVA DO DNA DE
Cyphoderus similis FOLSOM E *Cyphoderus innominatus* MILLS**

Dissertação apresentada ao programa de Pós-graduação em Ciências Biológicas (Zoologia), da Universidade Federal da Paraíba, em cumprimento as exigências para obtenção do título de Mestre em Ciências Biológicas.

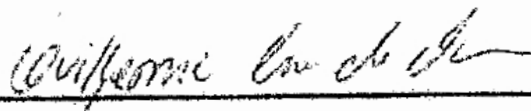
Aprovado em **31 de julho de 2019**

Banca Examinadora:



Dr. Douglas Zeppelini

Presidente - Orientador (Universidade Federal da Paraíba)



Dr. Guilherme C. Oliveira

Co-Orientador (Instituto Tecnológico Vale – Desenvolvimento Sustentável)



Dr. Bruno C. Bellini

Membro Titular (Universidade Federal do Rio Grande do Norte)



Dr. Eder S. Pires

Membro Titular (Instituto Tecnológico Vale – Desenvolvimento Sustentável)

Dr. Nikolas Gioia Cipola

Membro Suplente (Instituto Nacional de Pesquisas da Amazônia)

DEDICATÓRIA

À Deus, o criador de tudo e o conhecedor da sabedoria que envolve todos os mistérios do mundo.

À minha família, que me educou para o bem e para a justiça, torcendo pela minha felicidade.

Aos meus amigos, que me apoiaram e me alegraram nos momentos mais difíceis.

Aos meus professores, que me trouxeram o desejo de descobrir, a arte de questionar, e a persistência de seguir no caminho da ciência.

AGRADECIMENTOS

À Deus, senhor supremo de minha vida e minha mãe Santíssima por sempre me guiar pelos caminhos da vida e por sempre colocarem as pessoas certas no meu destino, proporcionando conquistas e realização de sonhos.

À Universidade Federal da Paraíba, campus João Pessoa e a todos os profissionais que fazem parte da rede UFPB.

Aos meus orientadores, Prof. Dr. Douglas Zeppelini e Prof. Dr. Guilherme C. Oliveira por toda a confiança, apoio, paciência e assistência durante toda a caminhada do Mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela bolsa de Pós-Graduação e o auxílio financeiro durante o período do mestrado.

À todos os professores do PPGCB - UFPB que participaram desta etapa de minha vida através dos ensinamentos e dedicação durante as aulas. Um agradecimento também ao Prof. Víctor Serrano-Solís por dividir seus conhecimentos durante minha participação nas disciplinas de bioinformática no PPGBCM – UFPB. Aos professores Wilson J. S. Júnior e seu grupo por me receber de forma muito atenciosa durante os cursos de bioinformática e análises de dados NGS no LABBEE – UFPE.

À todos os integrantes e amigos do Instituto Tecnológico Vale – Desenvolvimento Sustentável em Belém – PA, o qual tive o prazer de fazer parte desse ‘mundo de outro nível’ durante 2 meses de treinamento, dedicação e atenção. Em especial aos amigos e parceiros de trabalhos Renato R. M. Oliveira e Éder S. Pires, além de todo o grupo de colegas do ITV pelos ensinamentos, companhia, cafés, e algumas aventuras inesquecíveis em Belém.

À Roniere A. Brito e Robson A. Zampaulo pela colaboração na coleta e transporte dos colêmbolos para o laboratório.

À minha mãe Fátima e meu pai João, por me proporcionarem os maiores ensinamentos da vida, pela educação, dedicação e amor dado a mim até hoje. Aos meus irmãos, cunhadas, sobrinhos, tios e primos que contribuíram para que esta caminhada fosse proveitosa. Ao meu companheiro de vida Joelson Lino pela atenção, paciência, dedicação, apoio e força durante esta etapa de minha vida. Obrigado família, amo vocês!

Aos meus grandes amigos da vida que sempre estiveram comigo, uns desde o ensino fundamental e médio, e outros mais recentes, porém agradeço sempre pelo carinho, força e atenção. Aos meus grandes amigos da biologia que permaneceram comigo. Aos meus colegas de laboratório (LSCC - UEPB) pela companhia, colaboração e diversão nos momentos mais difíceis, em especial aos meus amigos Bruna, Nathan, Roni, João Victor, Aila, Estevan, Misael e Luiz que diariamente fazem o laboratório um ambiente completamente louco, porém divertido. Obrigado Amigos!

“Seja a mudança que você quer ver no mundo”

(Mahatma Gandhi)

RESUMO

O gênero *Cyphoderus* Nicolet 1842 é o maior na subfamília Paronellidae, com distribuição mundial. No Brasil ocorrem nove espécies e dentre elas temos *Cyphoderus similis* Folsom 1927 e *Cyphoderus innominatus* Mills 1938. Atualmente apenas quatro espécies da Classe Collembola têm seu genoma completo publicado enquanto apenas 14 espécies apresentam seu genoma mitocondrial acessível. Estudos envolvendo o sequenciamento de DNA ampliam o conhecimento da biologia do grupo a nível molecular, tornando-se ponto inicial para estudos genéticos, evolutivos, filogenéticos e de conservação. Portanto esta dissertação tem como objetivo principal sequenciar e comparar o DNA nuclear e mitocondrial de *C. similis* e *C. innominatus*, e como objetivos específicos temos: a descrição estrutural e funcional dos genomas sequenciados; a comparação dos genomas sequenciados com os demais genomas publicados; realizar um teste filogenético molecular utilizando os genes sequenciados. A espécie *C. similis* foi coletada aleatoriamente a partir de solos de cavernas de minério de ferro em Minas Gerais, enquanto que a espécie *C. innominatus* foi coletada em áreas verdes da Universidade Estadual da Paraíba, João Pessoa. Posteriormente os espécimes foram mantidos em cultivos, separados por espécie. No Instituto Tecnológico Vale – Desenvolvimento Sustentável (ITV-DS/PA) foram realizadas as etapas de extração e sequenciamento do DNA. Utilizou-se o kit *DNeasy Blood & Tissue* para extração e as tecnologias de sequenciamento Illumina *NextSeq* e *MiniSeq* com as técnicas *paired-end* e *mate-pair* combinadas. Temos como resultados o sequenciamento dos DNA mitocondriais com um comprimento total de 14.988 bp com 36 genes para *C. similis* e 14.811 pb com 37 genes, para *C. innominatus*, ambas apresentando conteúdo gênico completo e semelhante a outros genomas mitocondriais de Collembola. A filogenia baseada em sequências de DNA mitocondrial corrobora estudos anteriores, porém traz maior robustez para a filogenia molecular atual de Collembola. Os genomas nucleares contêm como comprimento total máximo 358 Mb com N50 de 3.3 Kb para *C. similis* e 370 Mb com N50 de 3.2 Kb para *C. innominatus*. A anotação gênica apresentou 278.865 genes putativos encontrados no genoma de *C. innominatus* com densidade de 3.8 genes para cada éxon encontrado. Os genomas nucleares apresentam-se ainda fragmentado e com baixo valor de N50 ao ser comparado com outros genomas disponíveis. Portanto, concluímos aqui os dois novos genomas mitocondriais completos, além dos primeiros *Drafts Genomes* nucleares para *Cyphoderus similis* e *Cyphoderus innominatus*, os primeiros genomas de Collembola para a região brasileira.

Palavras-chave: Collembola, Genômica, *Draft Genome*, *Next Generation Sequencing* (NGS), Filogenia Molecular.

ABSTRACT

The genus *Cyphoderus* Nicolet 1842 is the largest in the Paronellidae subfamily, with worldwide distribution. In Brazil there are nine species and among them there are *Cyphoderus similis* Folsom 1927 and *Cyphoderus innominatus* Mills 1938. Currently only four Collembola Class species have their complete genome published while only 14 species have their accessible mitochondrial genome. Studies involving DNA sequencing broaden knowledge of group biology at the molecular level, becoming a starting point for genetic, evolutionary, phylogenetic, and conservation studies. Therefore this dissertation has as main objective to sequence and compare the nuclear and mitochondrial DNA of *C. similis* and *C. innominatus*, and as specific objectives we have: the structural and functional description of the sequenced genomes; the comparison of sequenced genomes with other published genomes; perform a molecular phylogenetic test using the sequenced genes. The species *C. similis* was randomly collected from iron cave soils in Minas Gerais, while the species *C. innominatus* was collected from green areas of Paraíba State University, João Pessoa. Subsequently the specimens were kept in cultures, separated by species. At Vale Technological Institute - Sustainable Development (ITV-DS / PA), DNA extraction and sequencing steps were performed. The *DNeasy Blood & Tissue* extraction kit and Illumina *NextSeq* and *MiniSeq* sequencing technologies were used with *paired-end* and *mate-pair* techniques. The result is the sequencing of mitochondrial DNA with a total length of 14,988 bp with 36 genes for *C. similis* and 14,811 bp with 37 genes for *C. innominatus*, both presenting complete gene content and similar to other Collembola mitochondrial genomes. Phylogeny based on mitochondrial DNA sequences corroborates previous studies, but brings greater robustness to Collembola's current molecular phylogeny. Nuclear genomes contain a maximum total length of 358 Mb with N50 of 3.3 Kb for *C. similis* and 370 Mb with N50 of 3.2 Kb for *C. innominatus*. The gene annotation showed 278,865 putative genes found in the *C. innominatus* genome with a density of 3.8 genes for each exon found. Nuclear genomes are still fragmented and low in N50 when compared to other available genomes. Therefore, we conclude here the two new complete mitochondrial genomes, in addition to the first Nuclear *Drafts Genomes* for *Cyphoderus similis* and *Cyphoderus innominatus*, the first Collembola genomes for the Brazilian region.

Keywords: Collembola, Genomic, Draft Genome, Next Generation Sequencing (NGS), Molecular Phylogeny.

LISTA DE FIGURAS

Figura 1.	Esquema da morfologia básica de Collembola. Apontando suas apomorfias: Colóforo, Tenáculo e Fúrcula, e estruturas referentes ao <i>bauplan</i> geral de hexápodes: Cabeça (com 1 par de antenas), Tórax (com 3 pares de pernas) e Abdome.....	19
Figura 2.	Registros de <i>Cyphoderus similis</i> em comensalismo com formigueiros nas imagens A, B e D , com indivíduos foréticos na imagem C	23
Figura 3.	Esquema simplificado das terminologias e processos do sequenciamento <i>shotgun</i> . Primeiramente o DNA do organismo é fragmentado aleatoriamente; estes fragmentos podem ser representados por <i>single-reads</i> (~50-1000 pb) ou <i>pair-reads</i> (<i>paired-end</i> ou <i>mate-pair</i> com tamanhos variáveis); a montagem ocorre, de acordo com o tipo de fragmento gerado, por sobreposição de <i>reads</i> , formando <i>contigs</i> e <i>scaffolds</i> ; modelos de genes podem ser inferidos através de genomas evolutivamente próximos já anotados, ou por dados de transcriptoma do próprio organismo alvo. Formando, assim, o <i>Draft Genome</i> de um organismo.....	25
Figura 4.	Bibliotecas genômicas de DNA <i>paired-end</i> e <i>mate-pair</i> . Combinação de insertos curtos e longos proporciona uma ampliação da cobertura na montagem <i>de novo</i>	25
Figura 5.	Ferramentas genômicas aplicadas na detecção de variabilidade genética em espécies e suas possíveis limitações.....	28
Figura 6.	Ordem gênica mitocondrial de alguns hexápodes. Pode-se observar o tamanho e a localização dos genes referentes ao genoma de cinco artrópodes.....	29
Figura 7.	Espécies de Collembola com genoma completamente sequenciados e sua distribuição biogeográfica atual. (A) fotografias de: Cornwall, N.J. (Inglaterra - 2006), Van Duinen, J. (Holanda - 2014), McQuillan, B. (Nova Zelandia – 2014 e Nanz, S. (USA - 2008). (B) mapa de distribuição biogeográfica das quatro espécies consecutivas. Em vermelho temos a localização registrada.....	31
Figura 8.	Mapa de distribuição biogeográfica de <i>C. similis</i> e <i>C. innominatus</i> , respectivamente. Apresentando uma distribuição bem mais representativa para o território brasileiro.....	32

Figura 9.	Mapa do estado de Minas Gerais destacando a região metropolitana de Belo Horizonte. Ao lado, uma imagem via satélite da Serra da Moeda, região alvo para um estudo de diversidade de Collembola associados a cavernas de ferro, local da primeira coleta.....	34
Figura 10.	Mapa do estado da Paraíba destacando a região de João Pessoa. Ao lado, uma imagem via satélite do campus da Universidade Estadual da Paraíba, local da segunda coleta.....	35
Figura 11.	Cultivos laboratoriais de <i>Cyphoderus</i> . (A) cultivos inicialmente desenvolvidos para pequenas populações; (B) cultivos posteriormente desenvolvidos para grandes populações.....	36
Figura 12.	Esquema do Sequenciamento por Síntese (SBS). Tecnologia de sequenciamento NGS utilizada neste trabalho.....	39
Figura 13.	Amplificação por Ponte dos fragmentos de DNA indexados presentes na biblioteca genômica, com detalhes na explicação das etapas de formação dos <i>clusters</i> de DNA.....	41
Figura 14.	Intensidade luminosa emitida pelas chamadas de bases em diferentes <i>clusters</i> na superfície da célula de fluxo.....	42
Figura 15.	Representação do Sequenciador NextSeq® e sua célula de fluxo. (A) máquina de sequenciamento NGS NextSeq®500/550. (B) célula de fluxo específica para o sequenciador NextSeq®.....	43
Figura 16.	Representação do Sequenciador MiniSeq® e sua célula de fluxo. A) máquina de sequenciamento NGS MiniSeq. B). célula de fluxo específica para o sequenciador MiniSeq.....	43
Figura 17.	Ferramenta FastQC para tratamento de qualidade de dados de sequenciamento de alto rendimento, onde se destaca o <i>status</i> de todos os parâmetros avaliados.....	45
Figura 18.	Genoma mitocondrial de <i>Cyphoderus similis</i> . Em verde está representado o conjunto de CDS; em vermelho o conjunto de rRNA (rrnL – 16S; rrnS – 12S); em rosa o conjunto de tRNA (especializados para cada aminoácido); e em azul o conjunto de Origem de Replicação das fitas (OH).....	51

Figura 19.	Genoma mitocondrial de <i>Cyphoderus innominatus</i> . Em verde está representado o conjunto de CDS; em vermelho o conjunto de rRNA (rrnL – 16S; rrnS – 12S); em rosa o conjunto de tRNA (especializados para cada aminoácido); e em azul o conjunto de Origem de Replicação das fitas (OH).....	53
Figura 20.	Anotação dos genes presentes em <i>scaffolds</i> do genoma de <i>C. innominatus</i> . Sendo a ordem: <i>scaffold 1</i> (acima) com tamanho 31.655 bp; <i>scaffold 2</i> (central) com tamanho 31.690 bp; e o <i>scaffold 34</i> (abaixo) com tamanho 33.968 bp.....	56
Figura 21.	Rearranjo gênico do genoma mitocondrial encontrado em <i>Cyphoderus similis</i> (sequência mitocondrial superior) e em <i>Cyphoderus innominatus</i> (sequência mitocondrial inferior). As cores representam os rRNA (vermelho), Origem de Replicação (azul), CDS (verde) e tRNA (rosa). A direção das setas indica o sentido do gene.....	58
Figura 22.	Exemplo do alinhamento múltiplo das 13 CDS anotadas no genoma mitocondrial de <i>C. similis</i> e <i>C. innominatus</i>	58
Figura 23.	Matriz de identidade referente aos genomas mitocondriais de Collembola atualmente disponíveis no GenBank com adição dos mitogenomas de <i>Cyphoderus</i> sequenciados.....	61
Figura 24.	Matriz de identidade referente aos <i>Barcodes</i> de Collembola atualmente disponíveis no GenBank com adição dos <i>Barcodes</i> de <i>Cyphoderus</i> sequenciados.....	61
Figura 25.	Filogenia molecular de Collembola baseada em genomas mitocondriais completos por análise de Máxima Verossimilhança e <i>bootstrap</i> de 1000. Com destaque para o monofiletismo das espécies <i>Cyphoderus</i> (em laranja). Importante: o ramo “CYPHODERUS_SP.N” refere-se à <i>Cyphoderus innominatus</i>	65
Figura 26.	Filogenia molecular de Collembola baseada em genomas mitocondriais representando as famílias monofiléticas (Onychiuridae, Neanuridae, Isotomidae e Paronellidae) com valores de <i>bootstrap</i> de 100. Com destaque para o monofiletismo da família Paronellidae (em laranja) e o parafiletismo de três táxons (em azul).....	66

Figura 27.	Filogenia molecular de Collembola baseada em genomas mitocondriais representando o monofiletismo da Ordem Poduromorpha (<i>bootstrap</i> de 81) e o Polifiletismo recuperado para as Ordens Entomobryomorpha (em laranja com <i>bootstrap</i> de 100) e Symphypleona (em azul com <i>bootstrap</i> de 95).....	66
Figura 28	Gráfico extraído do FastQC referente à qualidade das sequências brutas (imagem à esquerda) e tratadas (imagem à direita) para o sequenciamento NextSeq P-E (imagens de cima) e NextSeq M-P (imagens de baixo) da amostra de DNA de <i>C. similis</i>	94
Figura 29	Gráfico extraído do FastQC referente à qualidade das sequências brutas (imagem à esquerda) e tratadas (imagem à direita) para o sequenciamento MiniSeq P-E (imagens de cima) e NextSeq M-P (imagens de baixo) da amostra de DNA de <i>C. innominatus</i>	94

LISTA DE TABELAS

Tabela 1.	Lista das espécies de Collembola alvo para análises comparativas baseado em sequências de DNA nuclear. Foram baixados no Genbank através de seus códigos de acesso para genomas completos disponibilizados. As Ordens de Collembola estão representadas por Entomobryomorpha (E) e Poduromorpha (P).....	48
Tabela 2.	Lista das espécies de Collembola alvo para as análises comparativas baseadas em sequências de DNA mitocondrial. Foram baixados no Genbank através de seus códigos de acesso para genomas mitocondriais disponibilizados. As Ordens de Collembola estão representadas por Entomobryomorpha (E), Poduromorpha (P) e Symphypleona (S).....	48
Tabela 3.	Estatísticas básicas da montagem <i>de novo</i> dos genomas mitocondriais de <i>Cyphoderus</i> . * Acesso disponibilizado após publicação dos genomas.....	50
Tabela 4.	Atributos estruturais referentes aos genes anotados no genoma mitocondrial de <i>Cyphoderus similis</i> . Referente a diferenciação das fitas de mtDNA temos: fita <i>foward</i> (F) e fita <i>reverse</i> (R).....	52
Tabela 5.	Tabela de uso e frequência de códons do mtDNA de <i>Cyphoderus similis</i> . Sendo a média de códons de 4996.....	52
Tabela 6.	Atributos estruturais referentes aos genes anotados no genoma mitocondrial de <i>Cyphoderus innominatus</i> . Referente a diferenciação das fitas de mtDNA temos: fita <i>foward</i> (F) e fita <i>reverse</i> (R).....	54
Tabela 7.	Tabela de uso e frequência de códons do mtDNA de <i>Cyphoderus innominatus</i> . Sendo a média de códons de 4937.....	54
Tabela 8.	Características estruturais referente ao genoma nuclear de <i>Cyphoderus similis</i> . Comparando as duas ferramentas de montagens de genomas nucleares.....	55
Tabela 9.	Características estruturais referente ao genoma nuclear de <i>Cyphoderus innominatus</i> . Comparando as duas ferramentas de montagens de genomas nucleares.....	55
Tabela 10.	Composição funcional do genoma nuclear de <i>Cyphoderus innominatus</i> anotado por referência do genoma de <i>Drosophila melanogaster</i>	56
Tabela 11.	Composição de aminoácidos contidos nos genomas mitocondriais sequenciados de <i>Cyphoderus</i> . Representados em porcentagem (%)......	59

Tabela 12.	Composição estrutural e funcional comparativa dos genomas mitocondriais de <i>Collembola</i> disponíveis até então no GenBank em relação aos genomas mitocondriais de <i>Cyphoderus</i> sequenciados.....	59
Tabela 13.	Análise estrutural comparativa entre os genomas nucleares de <i>Collembola</i> atualmente sequenciados e os dois novos genomas de <i>Cyphoderus</i> sequenciados. *dados referentes à montagem pelo SOAPdenovo2. **calculada através da razão entre total de bases sequenciadas e o total de bases montadas...	63
Tabela 14.	Análise funcional comparativa entre os genomas nucleares de <i>Collembola</i> atualmente sequenciados e o novo genoma de <i>Cyphoderus innominatus</i> devidamente anotado.....	62
Tabela 15	Resumo dos dados moleculares referentes ao pós-sequenciamento dos <i>Cyphoderus similis</i> e <i>Cyphoderus innominatus</i> . Comparando os dados brutos (B) e tratados (T). Valores de <i>phred-score</i> <30 é considerado ruim.....	93

SUMÁRIO

1. INTRODUÇÃO.....	18
1.1. INTRODUÇÃO A COLLEMBOLA COM ÊNFASE NA FAUNA BRASILEIRA DE <i>Cyphoderus</i> (Nicolet 1842).....	18
1.2. INTRODUÇÃO A GENÔMICA E AO SEQUENCIAMENTO DE DNA.....	22
1.3. APLICABILIDADE DE DADOS GENÔMICOS COM ÊNFASE NA GENÔMICA DE COLLEMBOLA.....	27
2. OBJETIVOS.....	33
2.1. OBJETIVO GERAL.....	33
2.2. OBJETIVOS ESPECÍFICOS.....	33
3. MATERIAIS E MÉTODOS.....	34
3.1. ÁREA E METODOLOGIA DE COLETA.....	34
3.2. CRIAÇÃO DE CULTIVOS E IDENTIFICAÇÃO DOS COLLEMBOLA COLETADOS.....	35
3.3. PROCEDIMENTOS LABORATORIAIS PARA SEQUENCIAMENTO DE DNA.....	37
3.3.1. Extração de DNA.....	37
3.3.2. Construção das bibliotecas de sequências de DNA.....	38
3.3.2.1. Biblioteca <i>Paired-End</i>	38
3.3.2.2. Biblioteca <i>Mate-Pair</i>	38
3.3.3. Sequenciamento de DNA.....	39
3.3.3.1. Sequenciamento NextSeq® 500.....	42
3.3.3.2. Sequenciamento MiniSeq®	43
3.4. ANÁLISES DOS DADOS DE SEQUENCIAMENTO DE DNA.....	44
3.4.1. Tratamento Pós-Sequenciamento dos dados moleculares.....	44
3.4.2. Montagem dos genomas.....	45
3.4.2.1. Montagem <i>de novo</i> dos Genomas Mitocondriais.....	45
3.4.2.2. Montagem <i>de novo</i> dos Genomas Nucleares.....	45
3.4.3. Anotação dos genomas.....	46
3.4.3.1. Anotação <i>ab initio</i> dos Genomas Mitocondriais.....	46
3.4.3.2. Anotação <i>ab initio</i> dos Genomas Nucleares.....	46
3.4.4. Análise estrutural dos genomas.....	47

3.4.5.	Análise comparativa dos genomas.....	48
3.4.6.	Análise filogenética dos genomas.....	49
4.	RESULTADO E DISCUSSÃO.....	50
4.1.	ANÁLISES ESTRUTURAIS E FUNCIONAIS DOS GENOMAS.....	50
4.1.1.	Análise Estrutural e Funcional dos Genomas Mitocondriais.....	50
4.1.1.1.	Genoma Mitocondrial de <i>Cyphoderus similis</i>	50
4.1.1.2.	Genoma Mitocondrial de <i>Cyphoderus innominatus</i>	52
4.1.2.	Análise Estrutural e Funcional dos Genomas Nucleares.....	55
4.2.	ANÁLISES COMPARATIVAS DOS GENOMAS.....	57
4.2.1.	Comparações entre os Genomas Mitocondriais.....	57
4.2.2.	Comparação entre Genomas Nucleares.....	62
4.3.	FILOGENIA MOLECULAR DE COLLEMBOLA APLICANDO DOIS NOVOS GENOMAS MITOCONDRIAIS DE <i>Cyphoderus</i>	64
5.	CONCLUSÃO.....	68
	REFERÊNCIAS.....	69
	APÊNDICE A.....	88
	APÊNDICE B.....	89
	APÊNDICE C.....	91
	APÊNDICE D.....	92

1. INTRODUÇÃO

1.1. INTRODUÇÃO A COLLEMBOLA COM ÊNFASE NA FAUNA BRASILEIRA DE *Cyphoderus* (Nicolet 1842)

Comunidades de animais edáficos constituem um grande reservatório de biodiversidade terrestre (DECAËNS *et al.*, 2006) das quais fazem parte de ambientes complexos cuja organização e composição não são facilmente compreendidos (ANDRÉ *et al.*, 1994, 2001; DANCE, 2008), seja por meio da diversidade microbiana (ROESCH *et al.*, 2007; FULTHORPE *et al.*, 2008) ou de organismos da micro/meso/megafauna do solo (DECAËNS *et al.*, 2008; STORK *et al.*, 2008). Nestas perspectivas de aplicabilidade, os invertebrados ganham destaque, tanto pela sua abundância quanto pela sua riqueza nestes ambientes. Um efeito significativo para a crescente importância no estudo de invertebrados de solo é de que estas comunidades relacionam-se à princípios ecológicos gerais, portanto invertebrados como minhocas, ácaros, e hexápodes (colêmbolos, besouros, gafanhotos, formigas, lagartas) são alvos de diferentes estudos em ecologia, monitoramento e conservação ambiental dentre os mais variados ambientes (HEMMER, 1990; HOPKIN, 1997; HOLMSTRUP *et al.*, 2002; DE DEYN *et al.*, 2003; NARDI *et al.*, 2003a; HEEMSBERGEN *et al.*, 2004; SETÄLÄ *et al.*, 2005; DOMES *et al.*, 2007; NAKAMORI *et al.*, 2010; GERLACH *et al.*, 2013).

Collembola (*springtails*) são os mais basais dos hexápodes entognatos (artrópodes com peças bucais encapsuladas), sendo a espécie *Rhyniella praecursor* (Hirst & Maulik 1926) o registro fóssil mais antigo datado do Devoniano identificado em 1919 (HIRST & MAULIK, 1926; KRISTENSEN, 1981; WHALLEY & JARZEMBOWSKI, 1981; GREENSLADE & WHALLEY, 1986). Conhecidos como os entognatos mais diversos e amplamente distribuídos, os colêmbolos habitam ecossistemas terrestres desde regiões tropicais a regiões polares (WISE, 1967; WISE *et al.*, 1971; MIRANDA-RANGEL & PALACIOS-VARGAS, 1992; GREENSLADE, 1995; FRATI *et al.*, 1997; HOPKIN, 1997; FRATI *et al.*, 2000b; DEHARVENG, 2004; ZEPPELINI & BELLINI, 2004; DEHARVENG *et al.*, 2008; ZEPPELINI *et al.*, 2009; ANSLAN, 2017).

Os colêmbolos possuem um número reduzido de segmentos abdominais (**Figura 1**) e são definidos por três características autapomórficas:

- a) O tubo ventral ou colóforo – um apêndice ventral responsável por funções como limpar e osmorregular a superfície do corpo, captação de água e fixação ao substrato, e está localizado no primeiro segmento abdominal;
- b) O retináculo ou tenáculo – apêndice ventral localizada no terceiro segmento abdominal;
- c) A fúrcula ou furca (do inglês *spring* – “mola”) - localizada ventralmente no quarto segmento abdominal, este apêndice permanece retido pelo retináculo e, ao ser estendida, possibilita que os colêmbolos realizem grandes saltos.

Vale ressaltar que o retináculo e a fúrcula podem ser secundariamente perdidos em linhagens internas (BRUSCA & BRUSCA, 2003; DELSUC *et al.*, 2003; CHEN *et al.*, 2019)

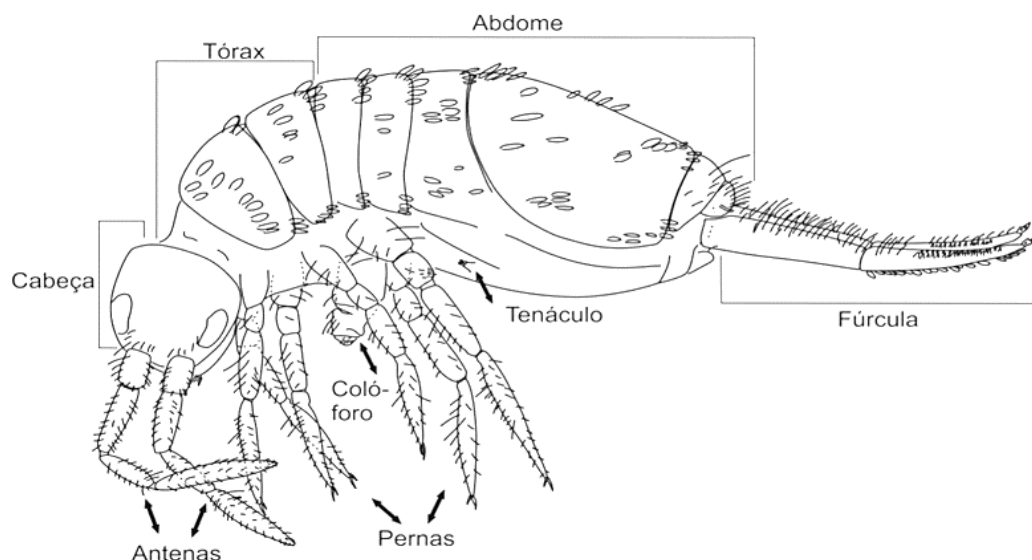


Figura 1. Esquema da morfologia básica de Collembola. Apontando suas apomorfias: Colóforo, Tenáculo e Fúrcula, e estruturas referentes ao *bauplan* geral de hexápodes: Cabeça (com 1 par de antenas), Tórax (com 3 pares de pernas) e Abdome. Fonte: BELLINI, 2016.

Habitantes típicos de ecossistemas terrestres, os colêmbolos fazem parte da mesofauna edáfica, podendo ser encontrados na superfície do solo como nas camadas da serapilheira (epiedáfico), ou nos horizontes mais profundos do solo (euedáfico) (BARDGETT & GRIFFITHS 1997; HOPKIN, 1997). Há registros de espécies de colêmbolos relacionados à ninhos de insetos sociais como cupins e formigas, e devido a esta circunstância de inquilinismo algumas espécies de Collembola foram classificadas como foréticas, por serem encontradas sob

partes do corpo de seus “anfitriões” insetos (CHRISTIANSEN, 1950, 1992; HERMANN *et al.*, 1970; KISTNER, 1982; HÖLDOBLER & WILSON, 1990; WAGNER *et al.*, 1997; KAUFMANN *et al.*, 2003; YÉO *et al.*, 2006; BAKR *et al.*, 2007; MOSER & BLOMQUIST, 2011; CASTAÑO-MENESES *et al.*, 2014;).

Alimentam-se de maneira não-seletiva, a partir de uma ampla variedade de resíduos de plantas, fungos ou animais (vivos ou mortos), sendo classificados como decompositores e, assim importantes para a cadeia trófica terrestre (PARKINSON *et al.*, 1979; HUNT *et al.*, 1987; BARDGETT & GRIFFITHS 1997; HOPKIN, 1997; RUSEK, 1998; FILSER, 2002; CASSAGNE *et al.*, 2003; CASTAÑO-MENESES *et al.*, 2004; HOSKINS *et al.*, 2015; POTAPOV *et al.*, 2016). Estes microartrópodes de solo também afetam o crescimento de plantas, uma vez que utilizam de micélios de fungos micorrízicos como alimento e dispersam seus esporos, sendo os fungos considerados como a dieta mais importante para Collembola (KOOISTRA, 1964; CHEN *et al.*, 1995; KLIRONOMOS & KENDRICK, 1995; THIMM & LARINK, 1995; GANGE, 2000; BAKONYI *et al.*, 2002; CASSAGNE *et al.*, 2003). Há também relatos de espécies de colêmbolos predando nematoides, relação ecológica benéfica relacionado ao controle populacional destes organismos no solo (GILMORE & GILMORE, 1970). Por outro lado, colêmbolos servem como presa para outros artrópodes de solo, como ácaros por exemplo (CHRISTIANSEN & BELLINGER, 1980, 1998; JOHNSON & WELLINGTON, 1980; USHER, 1985; BILDE *et al.*, 2000; LAWRENCE & WISE, 2000; KOPECKY *et al.*, 2016; NYFFELER & BIRKHOFFER, 2017; YIN *et al.*, 2017).

A reprodução do grupo é aparentemente primitiva, por se tratar da transferência indireta de espermatozoides através de estruturas delicadas chamadas de espermatóforos depositados no substrato por machos e capturados por fêmeas (CHRISTIANSEN, 1964; HOPKIN, 1997). Entretanto, muitas espécies são consideradas partenogenéticas, estratégia que pode sustentar uma população a partir de uma baixa densidade populacional, proporcionando uma adaptabilidade a flutuações populacionais (PETERSEN, 2002; GOLOSHCHAPOVA *et al.*, 2005; CHERNOVA *et al.*, 2010).

Algumas espécies de colêmbolos sustentam comunidades bacterianas abrigando-as em seu trato intestinal, ovários e células neurais, um exemplo é a infecção por *Wolbachia* (Bacteria, Rickettsiaceae) detectadas por microscopia eletrônica de transmissão, sendo esta coevolução apontada por alguns autores como benéfica para o hospedeiro no sentido de possibilitar a indução reprodutiva por partenogênese, ou seja o desenvolvimento de juvenis por meio de ovos não-fecundados (PALÉVODY, 1972; BORKOTT & INSAM, 1990; O'NEILL *et al.*, 1997;

WERREN, 1997; THIMM *et al.*, 1998; STOUTHAMER *et al.*, 1999; KOIVISTO & BRAIG, 2003; CZARNETZKI & TEBBE, 2004).

A presença/ausência de certas espécies de colêmbolos (como as espécies endêmicas) podem estar relacionado com alterações ambientais, como o pH, disponibilidade de alguns íons e/ou água, presença de componentes tóxicos (herbicidas, pesticidas ou até mesmo outros contaminantes), portanto podem ser consideradas como indicadores de qualidade de solo principalmente relacionados à impactos ambientais (BELLINGER *et al.*, 1996, 2008; FRAMPTON, 2000; CASSAGNE *et al.*, 2003, 2004; CHOI & MOORHEAD, 2006).

Mesmo estando entre os artrópodes dominantes em ambientes edáficos, há um número subestimado de espécies de colêmbolos descritas, contendo atualmente cerca de 9023 espécies distribuídas em 689 gêneros, 36 famílias e 4 Ordens (Poduromorpha, Entomobryomorpha, Symphypleona e Neelipleona) presentes em todo o mundo, com as famílias mais numerosas em espécies sendo Neanuridae (~1.463 espécies), Isotomidae (~1.406) e Entomobryidae (~1.014). (BELLINGER *et al.*, 1996-2019; DEHARVENG, 2004; ZHANG *et al.*, 2018b). Sendo a cor e a pigmentação as clássicas características utilizadas para identificação taxonômica (RUSEK, 2002), recentemente estudos com peças bucais e quetotaxia (número, natureza e posição de cerdas) representam um avanço tanto na taxonomia quanto na sistemática do grupo (DEHARVENG, 2004; ZEPPELINI *et al.*, 2018). Com o desenvolvimento de novas ferramentas e abordagens de identificação esse número de espécies possivelmente aumentará drasticamente, alcançando cerca de 50.000 espécies, número estimado por alguns autores (HOPKIN, 1997; DEHARVENG, 2004; CICCONARDI *et al.*, 2013), principalmente para regiões pouco amostradas, como é o caso da região tropical (CASSAGNAU, 1991; PALACIOS-VARGAS & ARBEA, 2009).

No Brasil, atualmente encontra-se um registro de 401 espécies de Collembola, sendo estas distribuídas em 113 gêneros e 21 famílias (BRITO *et al.*, 2017; ZEPPELINI *et al.*, 2019). Dentre elas está a família Paronellidae, representante da superfamília Entomobryiodea (responsável por cerca de 20% das espécies de colêmbolos), que inclui colêmbolos de corpo alongado, fúrcula geralmente bem desenvolvida subdividida nas subfamílias Paronellinae e Cyphoderinae (MITRA, 1993; BELLINGER *et al.*, 1996-2019; SOTO-ADAMES *et al.*, 2008, 2014; SOTO-ADAMES & BELLINI, 2015). Os Cyphoderinae (*sensu* Soto-Adames *et al.*, 2008) ou Cyphoderidae (*sensu* JANTARIT *et al.*, 2014) brasileiros agrupam espécies dos gêneros *Cyphoderodes* Silvestri 1910 (*Cyphoderodes xenopus* Börner 1913), *Cyphoderus* Nicolet 1841 (*C. agnotus* Börner 1906; *C. arlei* Cassagnau 1963; *C. caetetus* Zeppelini &

Oliveira 2016; *C. equidenticulati* Nunes & Bellini, 2018; *C. innominatus* Mills 1938; *C. javanus* Börner 1906; *C. mucrominimus* Oliveira, Alves & Zeppelini 2017; *C. mucrostrimenus* Oliveira, Alves e Zeppelini, 2017; *C. similis* Folsom 1927) e *Troglobius* Palacios-Vargas & Wilson, 1990 (*T. albertinoi* Cipola & Bellini, 2016; *T. brasiliensis* Palacios-Vargas & Zeppelini, 1995; e *T. ferroicus* Zeppelini, Silva & Palacios-Vargas, 2014) (BÖRNER, 1913; CASSAGNAU, 1963; CULIK *et al.* 2006; ABRANTES *et al.*, 2012; ZEPPELINI & OLIVEIRA, 2016; OLIVEIRA *et al.*, 2017).

A maioria das espécies de *Cyphoderus* são adaptadas para uma vida comensal (**Figura 2**) em ninhos de formigas e cupins, ou seja, são consideradas espécies mirmecófilas e termitófilas, respectivamente (DELAMARE-DEBOUTTEVILLE, 1948, CHRISTIANSEN, 1957; YOSHII, 1980, 1987, 1990, 1992). Além disso há registros de *Cyphoderus* foréticos a espécies de formigas, e registros de colêmbolos cifoderídeos habitando solos de cavernas. Ambientes cavernícolas mesmos pobres em matéria orgânica, quando comparado com ambientes da superfície, apresentam transportes ativos ou passivos de recursos orgânicos, sendo considerados bons ecossistemas para colêmbolos (CHRISTIANSEN, 1950; KISTNER, 1982; CAMASSA, 2004; ENDLWEBER *et al.*, 2009; MOSER & BLOMQUIST, 2011; CASTAÑO-MENESES *et al.*, 2014; JANTARIT *et al.*, 2014; ZHANG *et al.*, 2017).

1.2. INTRODUÇÃO A GENÔMICA E AO SEQUENCIAMENTO DE DNA

Em 1986 o pesquisador Thomas Roderick cunhou o termo genômica como sendo a parte da genética que estuda o sequenciamento, mapeamento e análises do genoma, podendo ser dividida atualmente em genômica estrutural, que remete-se ao estudo da estrutura e organização do genoma; e genômica funcional, que lida com a expressão e função dos genes (MCKUSICK, 1989; XU, 2012). A genômica já ficou limitada a organismos modelo, entretanto com o avanço das técnicas moleculares e diminuição de custo de sequenciamento os genomas de milhares de organismos não-modelo são sequenciados e os resultados aplicados a novas abordagens como metabolômica, proteômica e transcriptômica (FRATI *et al.*, 2000b; STEINER *et al.*, 2013; JOHNSON & KOEPFLI, 2014).



Figura 2. Registros de *Cyphoderus similis* em comensalismo com formigueiros nas imagens **A, B e D**, com indivíduos foréticos na imagem **C**. Fonte: Fotografias de Jon Hart (USA - 2010) e Graham Montgomery (USA - 2018) disponível em <www.bugguide.net>. Fotografias de Arnt Kvinnesland (Noruega - 2018) e Geir Drange (Noruega - 2011) disponível em <www.gbif.org>.

O DNA (*deoxyribonucleic acid*) é uma molécula relativamente estável e que, de acordo com as condições de armazenamento, congelada entre -80° a -20° Celsius por exemplo, pode sobreviver intacta por um longo intervalo de tempo sem apresentar perdas consideráveis em sua estrutura molecular (HIGUCHI *et al.*, 1984; PÄÄBO, 1989; LINDAHL, 1993; PHILLIPS & SIMON, 1995; HANRAHAN & JOHNSTON, 2011). A possibilidade de investigar o DNA de forma mais minuciosa e utilizá-lo como material de pesquisa concedeu-se a partir de rápidos avanços na tecnologia de sequenciamento de próxima geração (*New Generation Sequencing* - NGS) e de ferramentas de bioinformática durante a última década, quando houve a transição da era da genética clássica para a era da genômica, ampliando o poder das análises de DNA e melhorando a confiabilidade dos resultados (PRIMMER, 2009; ALLENDORF *et al.*, 2010; OUBORG *et al.*, 2010a; STEINER *et al.*, 2013).

Informações, principalmente sobre o tamanho de genomas são imprescindíveis tanto para estudos de comparação de genomas, quanto para estimar o gasto financeiro e determinar plataformas de sequenciamento adequadas (BENNETT & LEITCH, 2005; GREGORY, 2005a; 2005b). Estas informações sobre a dimensão de genomas geralmente são expostas através de números de pares de base, como valores de kilobase (Kb - 10^3 pb); megabase (Mb - 10^6 pb); ou gigabase (Gb - 10^9 pb) (GREGORY, 2002, 2005b; DOLEZEL *et al.*, 2003).

Os primeiros projetos de sequenciamento de genoma usam uma estratégia conhecida como *shotgun* (Sequenciamento Sanger – Primeira Geração), simplificada na **Figura 3** (SANGER *et al.*, 1977). Sobre esta técnica EKBLOM & WOLF (2014) explicam que a partir da extração o DNA é cortado em curtos fragmentos aleatórios (de 100–1000 pb), logo depois é sequenciado; estas tecnologias, principalmente, levam em consideração os tamanhos dos fragmentos, portanto o sequenciamento Sanger ficou amplamente relacionado à pequenos projetos de sequenciamento, enquanto os novos métodos de Sequenciamento NGS relacionados à projetos de sequenciamento maiores.

As técnicas de sequenciamento abrangem diferentes tipos de bibliotecas de DNA (**Figura 4**), estas bibliotecas representam o conjunto de fragmentos do DNA produzidos a partir de amplificações por PCR (Reação em cadeia da polimerase) da amostra inicial de DNA (ILLUMINA SEQUENCING, 2017). As bibliotecas genômicas podem ser:

- a) *Paired-End* (P-E): em que os fragmentos de DNA apresentam em torno de 200-500 pb (sendo chamadas de leituras curtas) e que as duas extremidades dos fragmentos são ligadas a adaptadores (etapa chamada de tagmentação), resultando em um sequenciamento de leituras curtas por ambas as extremidades, duplicando o total de *reads* geradas no fim do sequenciamento, apresentando assim uma maior capacidade na identificação de regiões de deleções ou inserções de nucleotídeos e permitindo maior precisão nos alinhamentos de bases;
- b) *Mate-Pair* (M-P): em que os fragmentos de DNA estão entre 2-5 kb (sendo chamadas de leituras longas), ocorre a tagmentação seguida por uma circularização e fragmentação, resultando em um sequenciamento de leituras longas por ambas as extremidades.

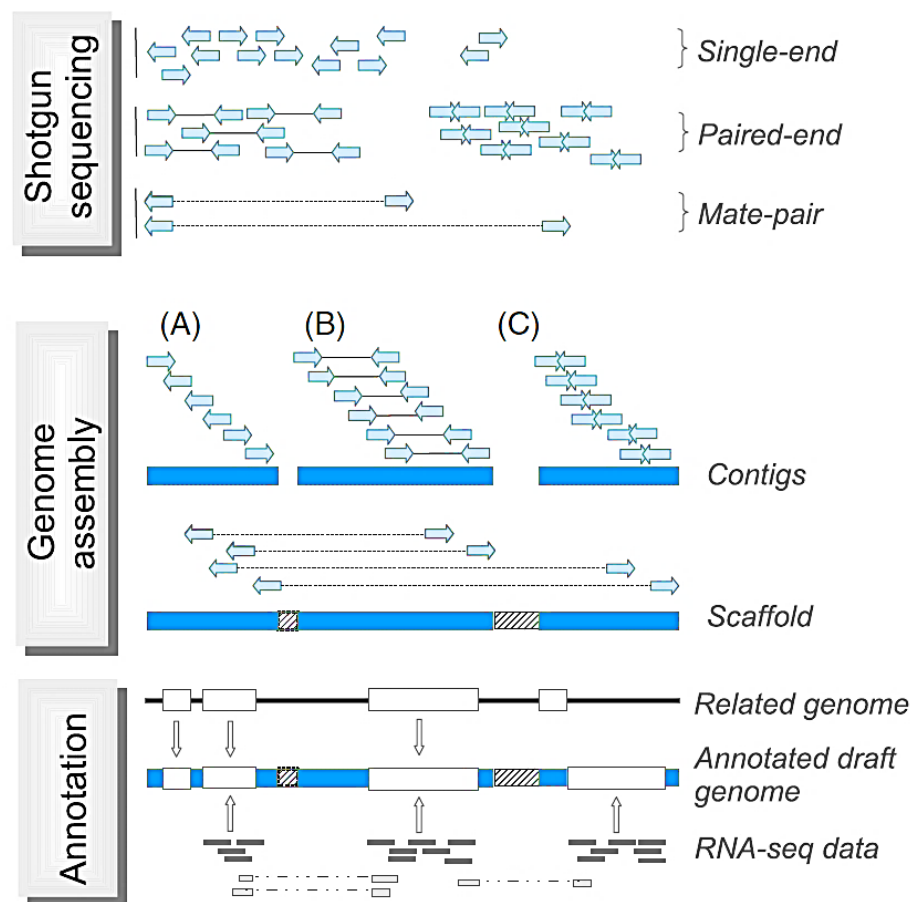


Figura 3. Esquema simplificado das terminologias e processos do sequenciamento *shotgun*. Primeiramente o DNA do organismo é fragmentado aleatoriamente; estes fragmentos podem ser representados por *single-reads* (~50-1000 pb) ou *pair-reads* (*paired-end* ou *mate-pair* com tamanhos variáveis); a montagem ocorre, de acordo com o tipo de fragmento gerado, por sobreposição de *reads*, formando *contigs* e *scaffolds*; modelos de genes podem ser inferidos através de genomas evolutivamente próximos já anotados, ou por dados de transcriptoma do próprio organismo alvo. Formando, assim, o *Draft Genome* de um organismo. Fonte: EKBLÖM & WOLF, 2014



Figura 4. Bibliotecas genômicas de DNA *paired-end* e *mate-pair*. Combinação de insertos curtos e longos proporciona uma ampliação da cobertura na montagem *De Novo*. Fonte: ILLUMINA SEQUENCING, 2017.

A tendência atual é de que os projetos de sequenciamento migrem do tradicional Sanger (sequências de ~1 kb) e Roche 454 (~800 pb) partindo para as tecnologias de sequenciamento NGS que trabalham com *reads* mais curtas, como o Illumina HiSeq (geralmente 150 pb) e SOLiD (geralmente 50 pb) (LI et al., 2010; WEI et al., 2012; EKBLOM & WOLF, 2014; VICENTE *et al.*, 2017). Uma gama de estudos recentes começou a combinar dados genômicos de diferentes comprimentos de *reads* e de plataformas de sequenciamento diferentes para ampliar a cobertura (proporcionando maiores *contigs* e *scaffolds*) e confiabilidade das montagens e anotações, e ainda para contrabalancear desvantagens de determinados métodos ou protocolos utilizados, principalmente para projetos de sequenciamento *de novo*, pois a facilidade de completar *gaps* na montagem de leitura curtas utilizando os fragmentos de longas leituras resulta em uma montagem mais eficaz (KOREN *et al.*, 2012; BRADNAM *et al.*, 2013).

O avanço nas tecnologias de sequenciamento foi proporcionado principalmente pelo avanço concomitante da área da bioinformática, na qual utiliza algoritmos computacionais para gerenciar e lidar com os dados moleculares, deixando mais prático agrupar as sequências geradas em trechos contínuos cada vez maiores, a partir de montagem por referência a algum modelo evolutivamente próximo, ou montagem sem referência pela técnica *de novo* (ILLUMINA SEQUENCING, 2017). Para uma boa montagem é necessária a sobreposição das *reads* suficientes para contabilizar uma cobertura elevada do genoma. Sendo assim, a escolha da plataforma de sequenciamento é uma das etapas primordiais em um projeto de sequenciamento (EKBLOM & WOLF, 2014).

Sequências de *Draft Genome* (Genoma Esboço) estão sendo produzidas em uma velocidade cada vez maior, onde bases de dados tradicionais estão fornecendo acesso a seus meta-dados, como o ENSEMBL (projeto científico em conjunto com o *European Bioinformatics Institute* e o *Wellcome Trust Sanger Institute*) e o NCBI (*National Center for Biotechnology Information*). O esboço de um genoma representa o sequenciamento geralmente de um único indivíduo, entretanto às vezes o genoma representa o sequenciamento de um “consenso” de amostras agrupadas (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2004; WHEELER *et al.*, 2008). Vale ressaltar que é particularmente impossível sequenciar e montar todos os nucleotídeos de um genoma, pois grande parte da sequência de DNA não são facilmente acessíveis, especialmente regiões da heterocromatina (região altamente condensada), ao redor de centrômeros, telômeros e as regiões repetitivas, mesmo nos mais conhecidos modelos de genomas (EKBLOM & WOLF, 2014; ELLEGREN, 2014). A aplicabilidade biológica de um genoma para ser totalmente decifrada precisa passar

por um processo de anotação genômica descrevendo as informações biologicamente relevantes do organismo, muitas vezes em forma de sequência de codificação de proteínas ou as várias classes de RNA (YANDELL & ENCE, 2012; PRIMMER *et al.*, 2013).

Anotações genômicas via de regra pretendem identificar os mais diferentes genes presentes no genoma do organismo, promovendo o início da interpretação biológica de um genoma (DENOEUDE *et al.*, 2007; ROZOWSKY *et al.*, 2007; HARROW *et al.*, 2009;). Os genes estão relacionados a regiões codificadoras de proteínas (CDS), onde geralmente inicia-se na extremidade 5' com um códon de iniciação (*Start codon*) e termina na extremidade 3' com um códon de parada (*Stop codon*), ou seja, CDS estão incluídas nas chamadas fases de leitura aberta (*open reading frame* - ORF) não incluindo os íntrons (se houver). Entretanto, nem todas as ORF's representam CDS válidas, mesmo estas sendo anotadas por ferramentas bioinformáticas, ou serem formadas tipicamente por aproximadamente 300 pb, e por apresentarem regiões ricas em Guanina e Citosina - GC (possíveis *start codon*) e rica em Adenina e Timina - AT (possíveis *stop codon*), estas CDS podem vir a ser identificadas ao acaso sendo chamadas de CDS putativa (CARNINCI *et al.*, 2005; CHENG *et al.*, 2005; CLAMP *et al.*, 2007; ENCODE PROJECT CONSORTIUM, 2007). Algumas CDS putativas referem-se a possíveis transposons (sequências de DNA com a habilidade de se mover de uma posição a outra do genoma), a genes ortólogos e parálogos (genes de espécies diferentes mas semelhantes entre si com ancestral comum; genes com função distintas e com ancestral comum presente em um mesmo organismo), a genes codificadores de proteínas propriamente ditos, ou genes órfãos (genes presentes apenas em algumas espécies) (CLAMP *et al.*, 2007).

1.3. APLICABILIDADE DE DADOS GENÔMICOS COM ÊNFASE NA GENÔMICA DE COLLEMBOLA

A estrutura genômica padrão de uma espécie reflete sua história evolutiva, contida exclusivamente em seu DNA, fornecendo informações detalhadas desde a biologia de indivíduos à padrões populacionais atuais ou passados, além de permitir novas descobertas aplicáveis em diversos campos da biologia, como a ecologia e biologia da conservação, a genética de populações, a biologia evolutiva e a sistemática filogenética, entre outras vertentes (HEDRICK, 1999; CRANDALL *et al.*, 2000; FRATI *et al.*, 2000b; FRASER & BERNATCHEZ, 2001; TIMMERMAN *et al.*, 2005; DE QUEIROZ, 2007; ALLENDORF *et al.*, 2010; THE 1000 GENOMES PROJECT CONSORTIUM, 2010; OUBORG *et al.*, 2010b;

MILLER *et al.*, 2012; KOBOLDT *et al.*, 2013; GOODWIN *et al.*, 2016; ZHANG *et al.*, 2018a). As informações moleculares vêm se tornando cada vez mais empregadas (**Figura 5**), seja como complemento para abordagens tradicionais ou até mesmo decisivas para muitos projetos de pesquisa, produzindo novas perspectivas para pesquisadores e seus mais variados grupos biológicos (CARAPELLI *et al.*, 1995; FRATI *et al.*, 2000a; SCHEFFER, 2000; CAMERON *et al.*, 2008; KING *et al.*, 2008; CHANG *et al.*, 2009).

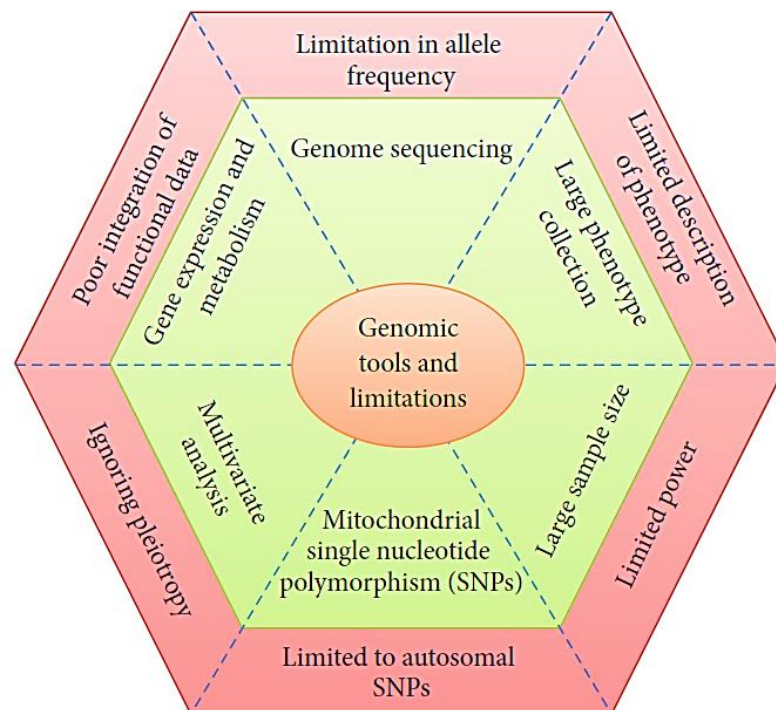


Figura 5. Ferramentas genômicas aplicadas na detecção de variabilidade genética em espécies e suas possíveis limitações. Fonte: KHAN *et al.*, 2016.

Abordagens comparativas de genomas possibilitam a identificação, a quantificação e a caracterização de sequências nucleotídicas ou protéicas presente nos genomas, descrevendo suas semelhanças (conservação) ou diferenças (divergência) (PARRA *et al.*, 2007). Fornecendo informações como a dimensão da montagem de novos genomas, a detecção de sequências perdidas ou de possíveis mutações nos genomas, a conservação de ordens gênicas entre grupos (EKBLOM & WOLF, 2014). A nível gênico, a genômica comparativa interpreta as diferenças relacionadas às suas características fenotípicas, ou seja, remetendo-se à adaptação das espécies envolvidas (FADDEEVA-VAKRHUSHEVA, 2017). Por exemplo, temos os genomas mitocondriais que vêm sendo alvo crescente para estudos comparativos (**Figura 6**) por

consistirem em um único éxon contínuo possuindo todo o conjunto gênico mitocondrial, particularmente empregados em análises filogenéticas, mecanismos de rearranjos genômicos e variabilidade de sequências (BERNT *et al.*, 2013). O DNA mitocondrial (mtDNA) trata-se de uma molécula circular que representa menos de 1% do DNA celular total dos mamíferos, por exemplo, e não difere muito em tamanho entre os mais diversos animais, em torno de 16.500 pb. Os mtDNA's de animais codificam precisamente 2 genes de RNA ribossomal (rRNA), sendo eles as subunidades 12S e 16S; 22 genes de RNA transportador (tRNA), abrangendo todos os aminoácidos necessários para produzir proteínas; 13 CDS (envolvidos no processo de fosforilação oxidativa); e em artrópodes, uma grande região rica em AT; todos de origem parental materna (ALBERTS *et al.*, 2010).

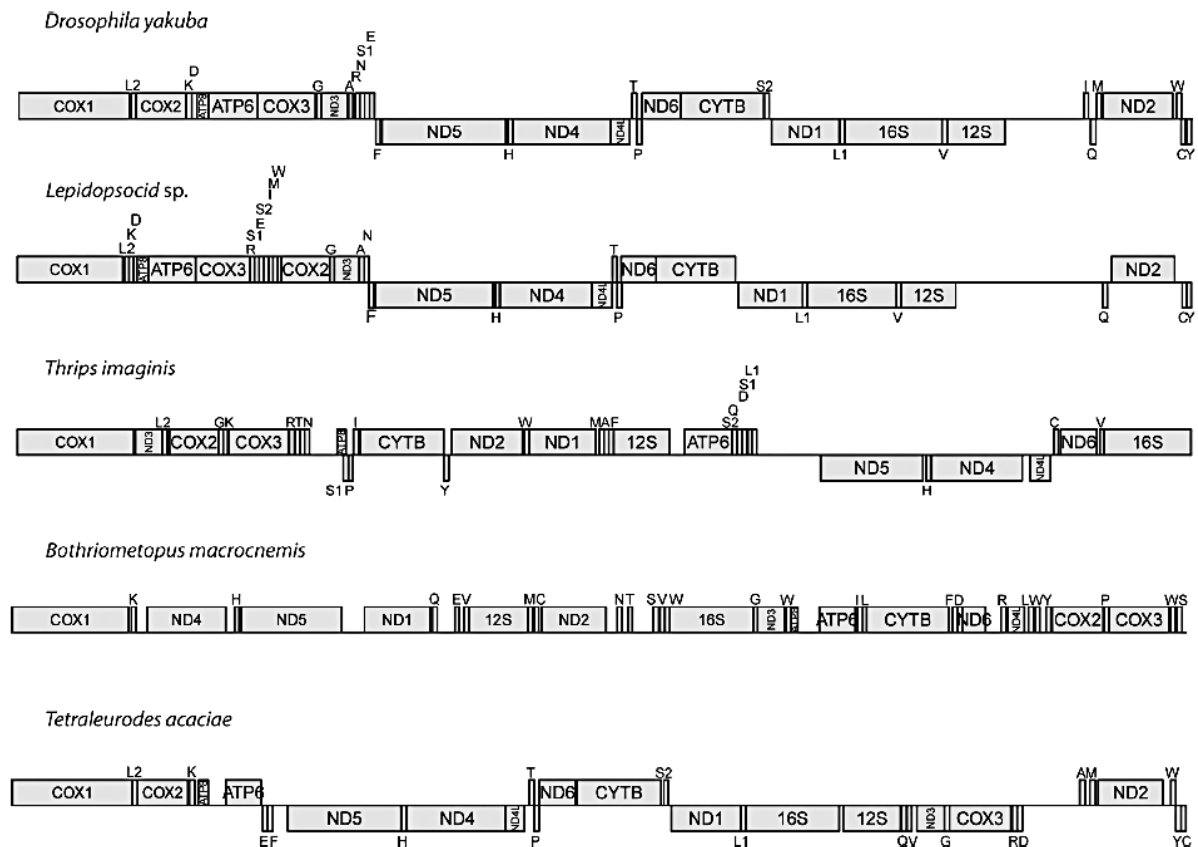


Figura 6. Ordem gênica mitocondrial de alguns hexápodes. Pode-se observar o tamanho e a localização dos genes referentes ao genoma de cinco artrópodes. Fonte: SIMON & HADRY, 2013.

O uso do DNA na taxonomia e filogenia tem se mostrado cada vez mais interessante e vantajoso, principalmente quando combinado com abordagens taxonômicas tradicionais e vem

sendo comumente utilizado para identificação e delimitação de espécies, útil principalmente na abordagens de espécies crípticas, que são aquelas separadas reprodutivamente porém mantêm um certo grau altamente semelhante morfológicamente entre si (WILL & RUBINOFF, 2004; DESALLE *et al.*, 2005; HEBERT & GREGORY, 2005; WILL *et al.*, 2005; RUBINOFF *et al.*, 2006; HAJIBABAEI *et al.*, 2007; ROUGERIE *et al.*, 2009; ANSLAN, 2017). Como é o caso do *DNA barcode* que se baseia em marcadores moleculares através de abordagens com marcadores únicos, geralmente o gene mitocondrial citocromo oxidase subunidade I – COI (HEBERT *et al.*, 2003; HAILER *et al.*, 2012; CARSTENS *et al.*, 2013; STEVENS & D’HAESE, 2017), ou abordagens *multilocus*, envolvendo tanto genes mitocondriais (citocromo b - Cob, citocromo oxidase subunidade II - COII, rRNA 16S), quanto genes nucleares (rRNA 28S, sub-região do espaçador intergênico nuclear - ITS2, Histona H3, fator de alongamento 1 alfa EF1- α) (LANGEL, 1993; PALUMETS, 1993; SELLIN, 1994; KNOWLES & CARSTENS, 2007; YANG & RANNALA, 2010; VON SALTZWEDEL *et al.*, 2017; MATEOS *et al.*, 2018; ZHANG *et al.*, 2018a,b).

Para Collembola os métodos moleculares baseados em marcadores de DNA vêm resolvendo problemas de delimitações de algumas espécies de *Orchesella* Templeton 1836, *Isotomurus* Börner 1903, *Deutonura* Cassagnau 1979, *Gomphiocephalus* Carpenter 1908, *Lepidocyrtus* Bourlet 1839, *Pseudosinella* Schäffer 1897, *Megalotorax* Willem 1900, *Entomobrya* Rondani 1861, *Parisetoma* Bagnall 1940, *Cryptopygus* Willem 1901, *Friesea* von Dalla Torre 1895, *Tomocerus* Nicolet 1842, *Pogonognathellus* Paclt 1944 e *Protaphorura* Absolon 1901 (CARAPELLI *et al.*, 1995b; FRATI *et al.*, 2000a; SOTO-ADAMES, 2002; STEVENS & HOGG, 2003; FELDERHOFF *et al.*, 2010; PORCO *et al.*, 2010; TORRICELLI *et al.*, 2010; SCHNEIDER & D’HAESE, 2013; ZHANG *et al.*, 2014a; KATZ *et al.*, 2015; CARAPELLI *et al.*, 2017; SUN *et al.*, 2017).

Poucos modelos genômicos são encontrados para os apterigotos, especificamente para a Classe Collembola, sendo as espécies *Folsomia candida* (Willem 1902) e *Orchesella cincta* (Linnæus 1758) as mais conhecidas e estudadas, fazendo com que as espécies não-modelo dividam os esforços de pesquisa em uma rede escassa de investigadores, abrangendo o interesse e a concentração das espécies modelos moleculares do grupo.

O sequenciamento genômico nuclear de Collembola conta, até o presente momento, com apenas 4 espécies investigadas (**Figura 7**):

- a) O genoma da *O. cincta* (Entomobryidae) – publicado em 2016 por VAKHRUSHEVA e colaboradores;
- b) O genoma da *F. candida* (Isotomidae) – publicado em junho de 2017 por FADDEEVA-VAKHRUSHEVA e colaboradores;
- c) O genoma da *Holacanthella duospinosa* Salmon 1942 (Neanuridae) – publicado em outubro de 2017 por WU e colaboradores;
- d) O genoma da *Sinella curviseta* Brook 1882 (Entomobryidae) – publicado em 2019 por ZHANG e colaboradores (b).

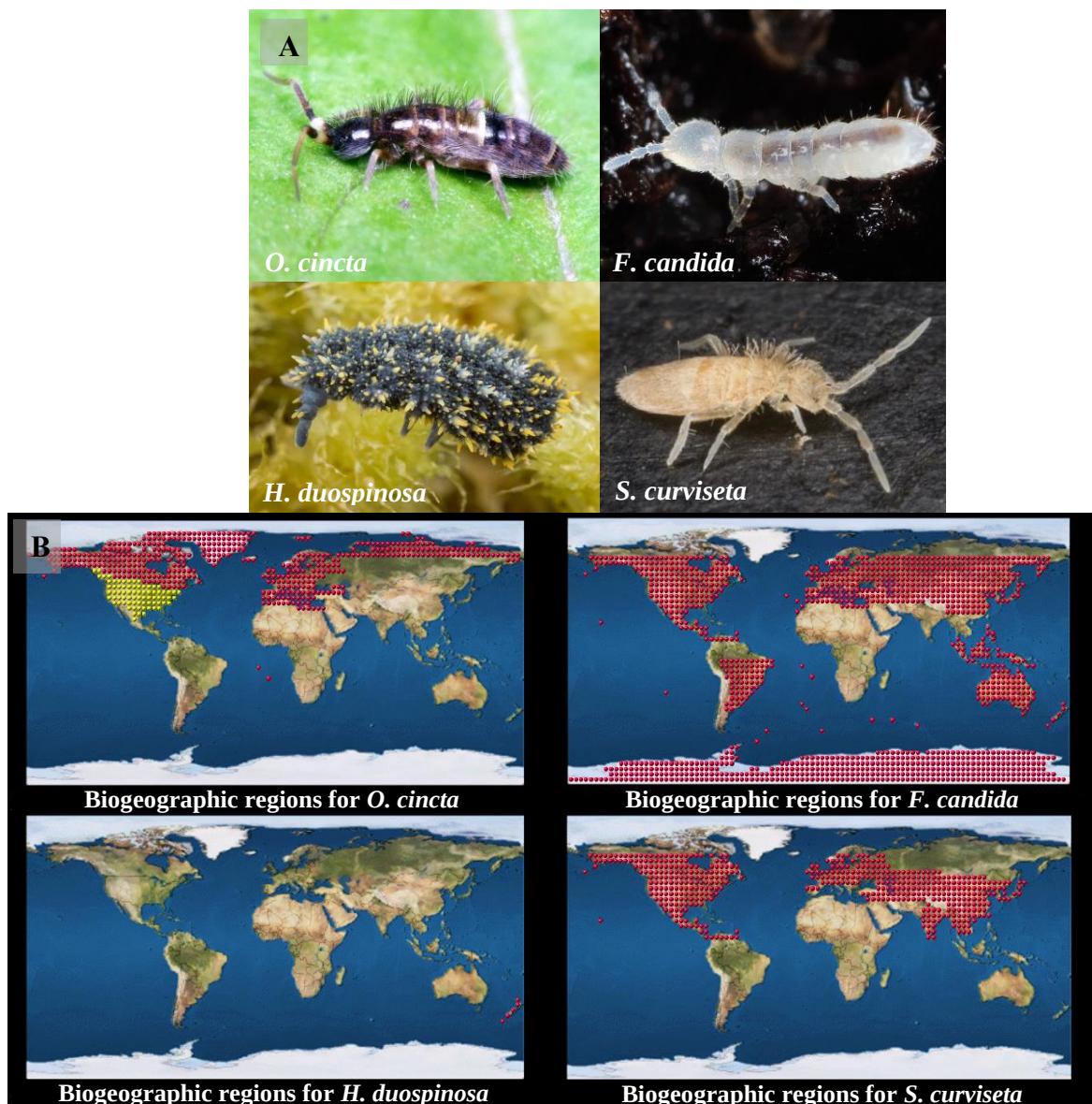


Figura 7. Espécies de Collembola com genoma completamente sequenciados e sua distribuição biogeográfica atual. **(A)** fotografias de: Cornwall, N.J. (Inglaterra - 2006), Van Duinen, J. (Holanda - 2014), McQuillan, B. (Nova Zelandia – 2014 e Nanz, S. (USA - 2008). **(B)** mapa de distribuição biogeográfica das quatro espécies consecutivas. Em vermelho temos a localização registrada. Fonte: Disponível em <www.collembola.org>.

Já o sequenciamento genômico mitocondrial de Collembola, contamos atualmente com 16 espécies investigadas, porém duas encontram-se recentemente sequenciadas, não possibilitando acesso a seus dados genômicos, sendo elas as espécies *Isotomurus maculatus* Schäffer 1896 e *Tullbergia bisetosa* Börner 1903 (JAGATAP *et al.*, 2019a,b), totalizando 14 genomas mitocondriais acessíveis.

De modo geral, as espécies *C. similis* e *C. innominatus* com uma distribuição mais ampla nas américas, em destaque para o Brasil (**Figura 8**), referem-se aos dois novos grupos de investigação e sequenciamento genômico que serão exploradas dentro da Classe Collembola.

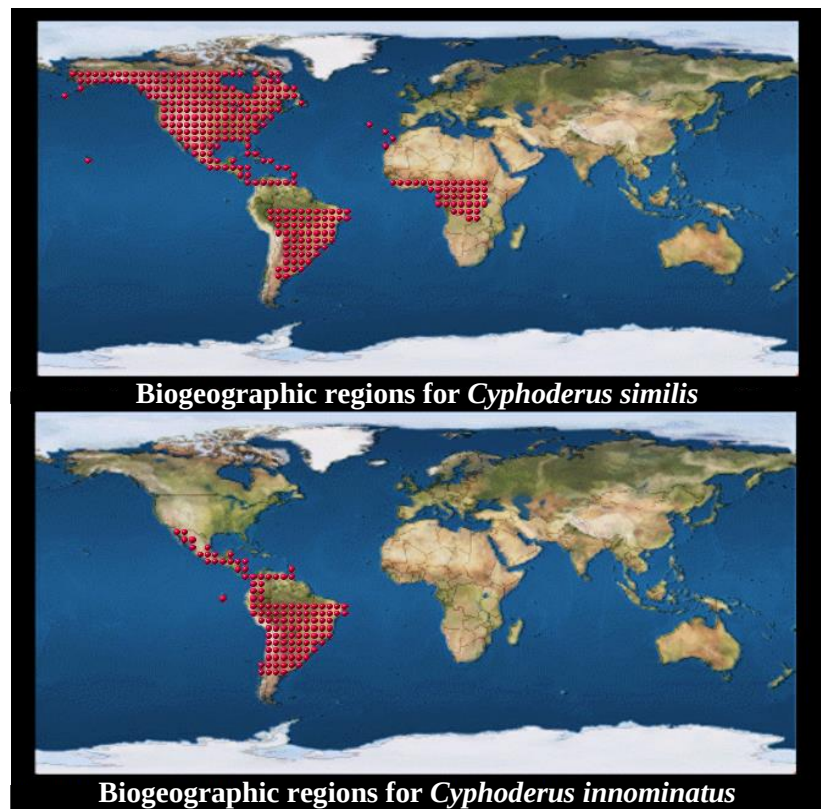


Figura 8. Mapa de distribuição biogeográfica de *C. similis* e *C. innominatus*, respectivamente. Apresentando uma distribuição bem mais representativa para o território brasileiro. Fonte: Disponível em < www.collembola.org >.

2. OBJETIVOS

2.1. OBJETIVO GERAL

- Sequenciar e comparar o genoma nuclear e mitocondrial de *Cyphoderus similis* e *Cyphoderus innominatus*.

2.2. OBJETIVOS ESPECÍFICOS

- Descrever as características estruturais e funcionais encontradas nos genomas das duas espécies de *Cyphoderus*;
- Comparar os genomas sequenciados de *Cyphoderus* com outros genomas nucleares e mitocondriais de Collembola disponíveis;
- Gerar uma nova hipótese evolutiva molecular para Collembola testando filogeneticamente os genes sequenciados de *Cyphoderus*.

3. MATERIAIS E MÉTODOS

3.1. ÁREA E METODOLOGIA DE COLETA

Foram realizadas duas coletas, a primeira ocorreu por meio da colaboração de Roniere A. Brito e Robson A. Zampaulo no período de Julho/Agosto de 2017 a partir de busca ativa aleatória com uso de aspiradores entomológicos em resíduos de formigueiros em zona hipógea (interna) e afótica de cavernas de minério de ferro localizadas na Serra da Moeda (**Figura 9**), nas seguintes coordenadas: 20°10'60.00''S, 43°58'19.00''O; enquanto que a segunda foi realizada no mesmo período em áreas verdes da Universidade Estadual da Paraíba (UEPB), campus V através de busca ativa aleatória com uso de aspiradores entomológicos (**Figura 10**), nas coordenadas: 7°09'31.21''S, 34°52'23.01''O.

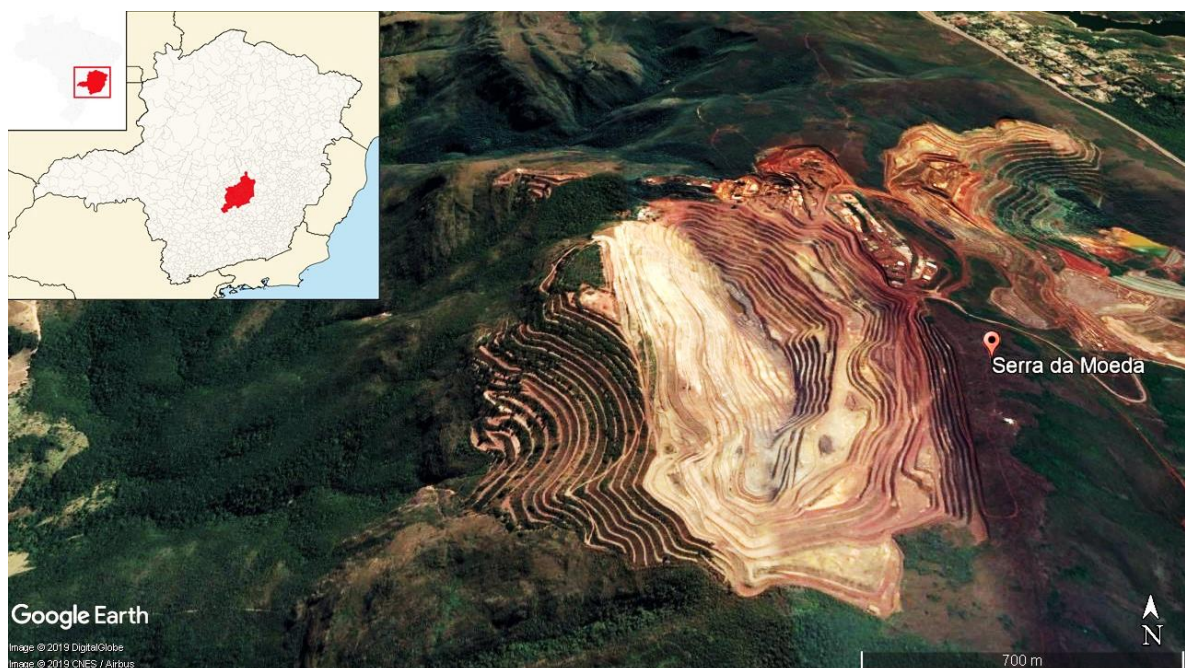


Figura 9. Mapa do estado de Minas Gerais destacando a região metropolitana de Belo Horizonte. Ao lado, uma imagem via satélite da Serra da Moeda, região alvo para um estudo de diversidade de Collembola associados a cavernas de ferro, local da primeira coleta. Fonte: Google Earth e colaboração de Raphael Lorenzeto de Abreu.



Figura 10. Mapa do estado da Paraíba destacando a região de João Pessoa. Ao lado, uma imagem via satélite do campus da Universidade Estadual da Paraíba, local da segunda coleta. Fonte: Google Earth e colaboração de Raphael Lorenzeto de Abreu.

Vale ressaltar que a escolha das áreas de coleta se deu meramente aleatória, sem haver interesse especiais em suas escolhas, com intenção apenas na obtenção de grandes quantidades de espécimes para elaboração de cultivos de *Cyphoderus* spp. A Serra da Moeda foi escolhida como um dos pontos de coleta por apresentar registro prévio de *Cyphoderus* (dados ainda não publicados da tese de doutorado de Roniere A. Brito) e pela praticidade e logística na obtenção dos espécimes através dos colaboradores mencionados anteriormente. Já o campus V da UEPB foi investigado, de forma independente, apenas como propósito de obtenção extra de espécimes para dar início aos testes de criação em cultivos laboratoriais.

3.2. CRIAÇÃO DE CULTIVOS E IDENTIFICAÇÃO DOS COLLEMBOLA COLETADOS

Em laboratório, as amostras coletadas foram inicialmente identificadas e separadas por espécie e inserida em cultivos (**Figura 11-A**), constituídos, a partir de recipientes plásticos com 4 cm de diâmetro, 5.55 cm de altura e 0.09 L de volume, adaptado de AMARANTO e colaboradores (2018), como uma estrutura eficaz para acomodar pequeno artrópodes em

experimentação laboratorial. Um substrato sólido especial foi implementado nos cultivos, baseado em ADAMS & SALMON (1972), a base de uma mistura de carvão vegetal e gesso de Paris ($\text{CaSO}_4 \frac{1}{2} \text{H}_2\text{O}$) na proporção de 1:9 e umedecido com água destilada (KROGH, 2009; FIEIRA, 2014). Os cultivos foram acondicionados à 22° Celsius e umidade do ar em aproximadamente 85%. Ao longo do desenvolvimento reprodutivo laboratorial de *Cyphoderus*, os cultivos foram ampliados para recipientes plásticos com volume de 600 ml preenchidos com húmus esterilizado (**Figura 11-B**).

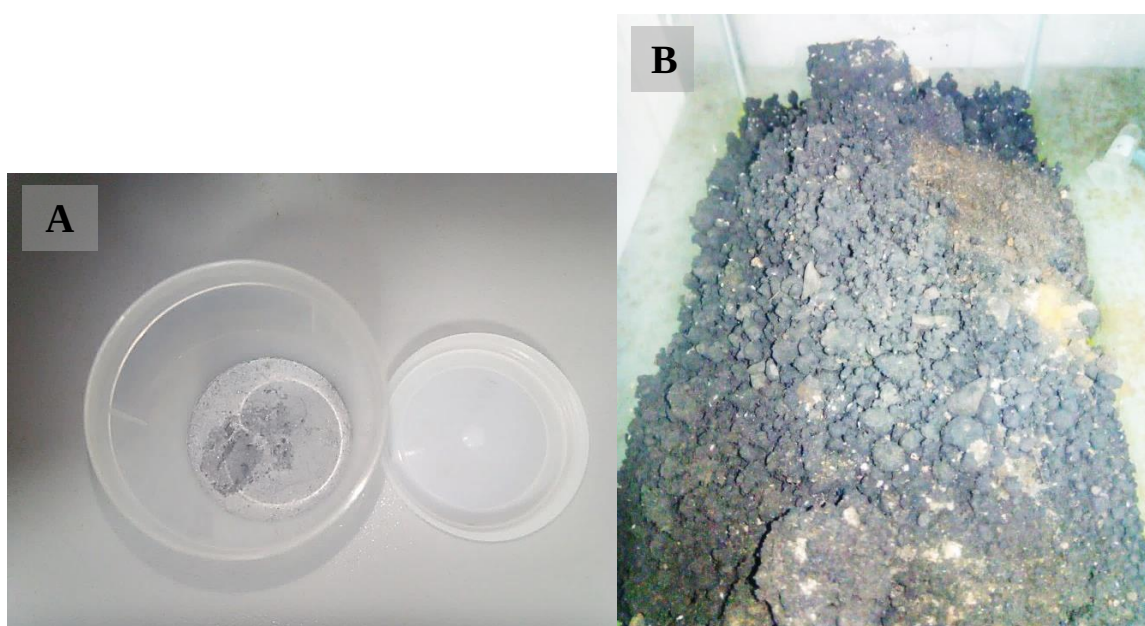


Figura 11. Cultivos laboratoriais de *Cyphoderus*. (A) cultivos inicialmente desenvolvidos para pequenas populações; (B) cultivos posteriormente desenvolvidos para grandes populações.

A dieta dos espécimes em cultivos se dava através de grãos de fermento seco (*Saccharomyces cerevisiae*) posto diretamente sobre o substrato (AL-SAFADI, 1988; THORNE *et al.*, 2011). Mesmo não havendo uma dieta sistematicamente balanceada a utilização de grãos de leveduras revela-se suficientemente adequada para espécies de colêmbolos mantidos em laboratório (CHRISTIANSEN, 1964; BELLINGER *et al.*, 2016). Diariamente os cultivos passavam por manutenção, através da observação direta da umidade do substrato, onde se necessário ocorria a adição de água destilada e/ou alimento *ad libitum*. Ao passo que a população de *Cyphoderus* crescia ocorria a divisão para novos cultivos inabitados, sempre respeitando a separação por espécie.

A identificação dos espécimes ocorreu a partir de sub amostragens do material coletado. Sob estereomicroscópio os espécimes passavam por diafanização (procedimento para diluir a musculatura e preservar o exoesqueleto) por KOH 5% e solução *Nesbitt*, logo depois foram montados em lâminas semipermanentes com líquido de Hoyer (ARLÉ & MENDONÇA, 1982; FARAJI & BAKKER, 2008). Após secagem em estufa, as lâminas foram identificadas em microscópio óptico a partir de morfologia e quetotaxia dos colêmbolos e com auxílio de chaves de identificação especializadas (OLIVEIRA et al., 2017). O registro da reprodução e identificação de *Cyphoderus* em laboratório encontra-se citado no **APÊNDICE A**, enquanto que o registro de polimorfismos em espécies de *Cyphoderus* é encontrado no **APÊNDICE B**.

3.3. PROCEDIMENTOS LABORATORIAIS PARA SEQUENCIAMENTO DE DNA

As etapas laboratoriais referentes ao sequenciamento de DNA ocorreram no Laboratório de Genômica do Instituto Tecnológico Vale Desenvolvimento Sustentável – ITV-DS, com orientação e supervisão dos doutores em microbiologia e especialistas em Genômica Eder S. Pires e Guilherme C. Oliveira.

3.3.1. Extração de DNA

Foram realizados testes pilotos com diferentes quantidades de espécimes de colêmbolos para padronizar o número de indivíduos por extração (CZARNETZKI & TEBBE, 2004; HOGG & HEBERT, 2004; NAKAMORI *et al.*, 2008; ROUGERIE *et al.*, 2009; ROBERTS & WEEKS, 2011; THORNE *et al.*, 2011; PORCO *et al.*, 2012; VON SALTZWEDEL *et al.*, 2017). De acordo com outros protocolos de extração de DNA de Collembola, utilizamos o KIT *DNeasy Blood & Tissue* da Qiagen® e um pool de 50 indivíduos (animais inteiros congelados a -80° Celsius) para cada extração (KUUSK & AGUSTÍ, 2008; NAKAMORI *et al.*, 2010; ZHANG *et al.*, 2014b; AOYAMA *et al.*, 2015; ZHANG *et al.*, 2015; VON SALTZWEDEL *et al.*, 2017). Foram realizadas duas extrações, uma para a população de *C. similis* e outra para a população de *C. innominatus*.

As extrações ocorreram através de maceração mecânica (beads magnéticas – minúsculas esferas metálicas) executada no Tissue Lyser, seguida por digestão química no Thermo Shaker com adição de Proteinase K, e finalizada por várias etapas de adição de *buffers* específicos do

Kit sob várias etapas de centrifugação. As amostras de DNA extraídas receberam as seguintes identificações: 10629 (amostra de *C. similis*) e 10660 (amostra de *C. innominatus*).

Após as extrações, as amostras passaram por processos de quantificação, referindo-se a uma fase de teste de qualidade. A quantificação foi realizada no Qubit 3.0 da Invitrogen®. A quantificação ocorreu por leitura de fluorescência das amostras, e sua unidade é o ng/µl. Valores da Extração de DNA de *Cyphoderus* é encontrado no **APÊNDICE C**.

3.3.2. Construção das bibliotecas de sequências de DNA

Foram construídas bibliotecas de DNA de dois comprimentos de fragmentos diferentes: a biblioteca *Paired-End* e *Mate-Pair*. Valores referentes as Bibliotecas de DNA de *Cyphoderus* é encontrado no **APÊNDICE C**.

As bibliotecas genômicas passam por um processo de indexação, que permite que um conjunto de fragmentos da biblioteca sejam agrupadas e sequenciadas juntas, sem que se misturem no sequenciador. Este processo constitui na adição de duas sequências únicas (index), já conhecidas. Nos protocolos seguidos as indexações são duplas (do tipo B), logo ocorre a adição de dois índice (i5 e i7) em cada fragmento de DNA.

3.3.2.1. Biblioteca *Paired-End*

O processo de construção da biblioteca de DNA *Paired-End* (P-E) foi o mesmo demonstrado no protocolo *SureSelect QXT Whole Genome Library Prep for Illumina Multiplexed Sequencing* da Agilent Technologies®.

Após a construção da biblioteca P-E foi realizado sua quantificação no Qubit 3.0 da Invitrogen® e no BioAnalyser da Agilent® através de eletroforese capilar.

3.3.2.2. Biblioteca *Mate-Pair*

O procedimento referente à construção da biblioteca de DNA *Mate-Pair* (M-P) foi o mesmo demonstrado no protocolo *Nextera® Mate Pair Library Prep Reference Guide* da Illumina®, com modificação ao utilizar o equipamento Biorruptor da Diagenode® no qual

ocorreu duas etapas de fragmentação por sonicação, uma vez que o tamanho dos fragmentos não estavam ideais.

Após a construção da biblioteca M-P houve a quantificação no Qubit 3.0 e avaliação final do tamanho dos fragmentos por eletroforese capilar no BioAnalyser.

3.3.3. Sequenciamento de DNA

Uma etapa preliminar ao sequenciamento é a Desnaturação e Diluição das bibliotecas de DNA. Procedimento realizado através do protocolo *NextSeq Denature and Dilute Libraries Guide* (Protocolo A) da Illumina®, comum para ambas as bibliotecas e tecnologias de sequenciamento.

Os sequenciamentos foram baseados no processo de Sequenciamento por Síntese (*Sequencing by synthesis - SBS*), amplamente adotada em sequenciamento NGS (**Figura 12**). Neste sequenciamento é incorporado um terminador (base nitrogenada) marcado com fluorescência, sendo visualizado assim que cada base é incorporada à sequência, emitindo um determinado comprimento de luz que será lido e interpretado ao longo do sequenciamento de DNA.

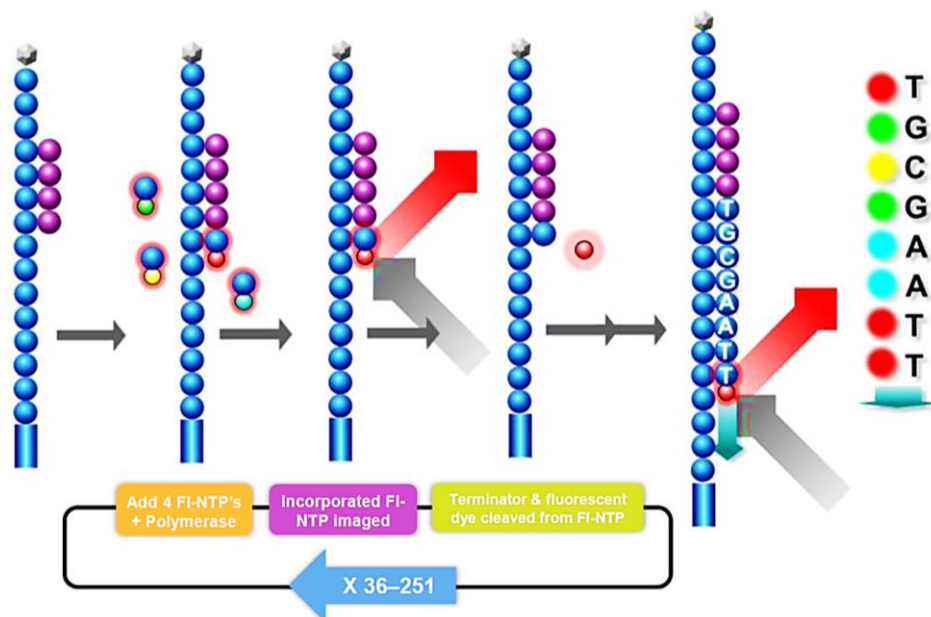


FIGURA 12. Esquema do Sequenciamento por Síntese (SBS). Tecnologia de sequenciamento NGS utilizada neste trabalho. Fonte: < Illumina_Sequencing_Overview_15045845_D.pdf >

Ao longo da etapa de sequenciamento, as amostras passam por um processo conhecido como Amplificação por Ponte (*Bridge Amplification*), no qual os fragmentos iniciais da biblioteca de DNA são captados e presos em pequenos oligonucleotídeos contidos nas células de fluxo (*flow cell*) do sequenciador, onde ocorre o sequenciamento propriamente dito. Por fim ocorrem os ciclos de síntese de DNA, onde, a cada ciclo de sequenciamento, a quantidade de moléculas ancoradas à *flow cell* cresce exponencialmente, formando agrupamentos de moléculas de DNA conhecidos como *clusters* (**Figura 13**).

Os *clusters* são visualizados por fluorescência, emitidas pelos adaptadores ligados a cada *reads* e que serão lidos a cada ciclo de sequenciamento, onde ao fim do sequenciamento essas *reads* são filtradas, como um tratamento de qualidade. A quantidade de ciclos do sequenciamento corresponde a duas etapas de 150 ciclos, no caso 150 x 2 ciclos. Porém, são realizados 151 x 2 ciclos, sendo esses 2 ciclos adicionais finais reservados para as etapas conhecidas por *phasing* e *prephasing*, pois a cada fim de ciclo as *reads* sintetizadas podem se estender ou não por 1 único nucleotídeo, sendo então necessária para o cálculo destes fenômenos.

A execução do sequenciamento é configurada e acompanhada por uma plataforma eletrônica integrada ao sequenciador, chamada de *BaseSpace*[®]. Assim que o sequenciamento avança os dados são enviados, em tempo real, para esta plataforma eletrônica.

Durante cada inserção de base no sequenciamento SBS, são extraídas as leituras de fluorescência, ou, intensidade luminosa emitida por cada uma das quatro bases do DNA (**Figura 14**), esta leitura ocorre a cada ciclo de sequenciamento através da ferramenta de Análise em Tempo-Real (*Real-Time Analysis*). Existe um índice de qualidade para estas chamadas de bases, conhecido como *Q-score*, onde valores altos deste índice, geralmente acima de 30, reflete uma maior qualidade na leitura de chamadas de bases.

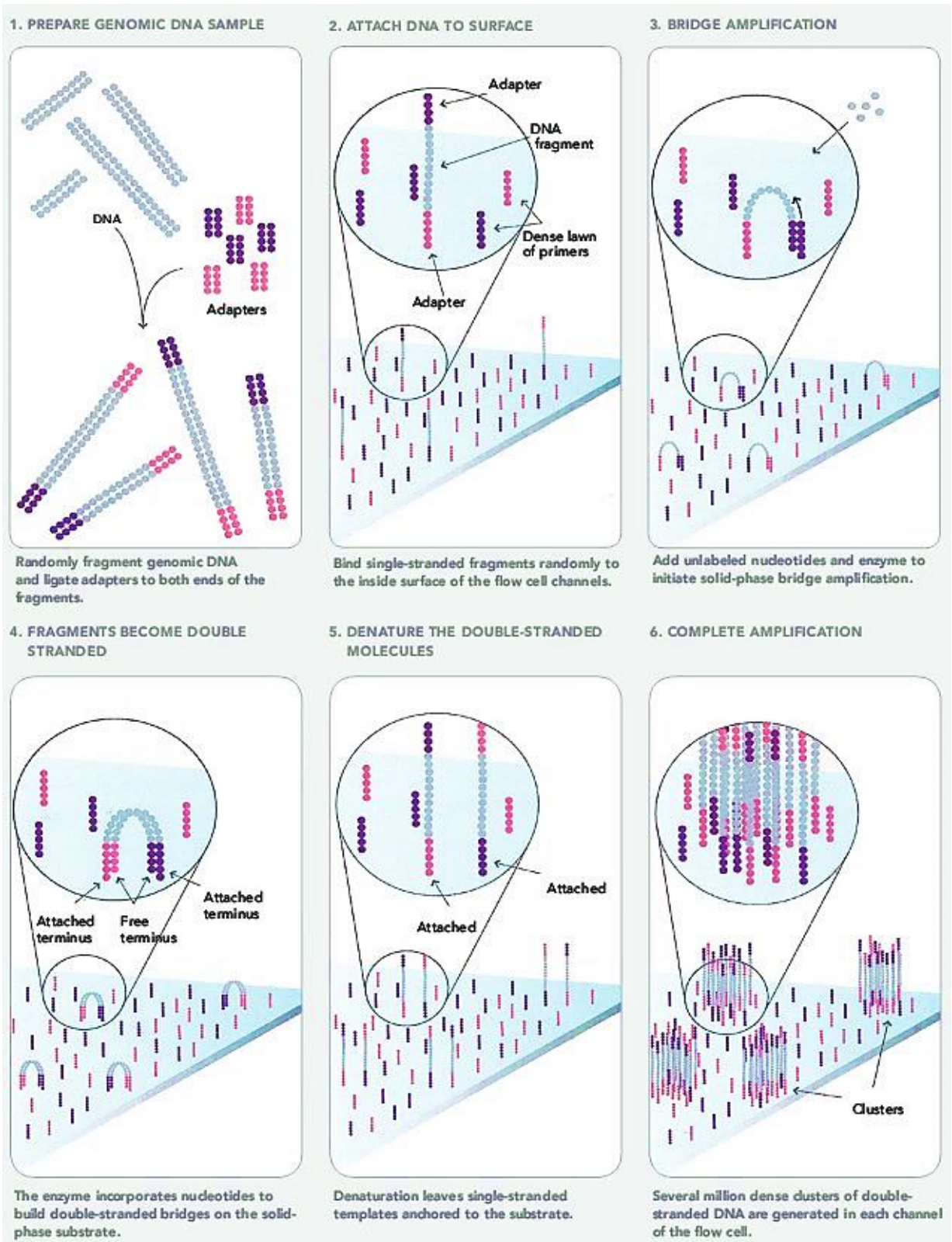


Figura 13. Amplificação por Ponte dos fragmentos de DNA indexados presentes na biblioteca genômica, com detalhes na explicação das etapas de formação dos *clusters* de DNA. Fonte: < <https://www.illumina.com/> >

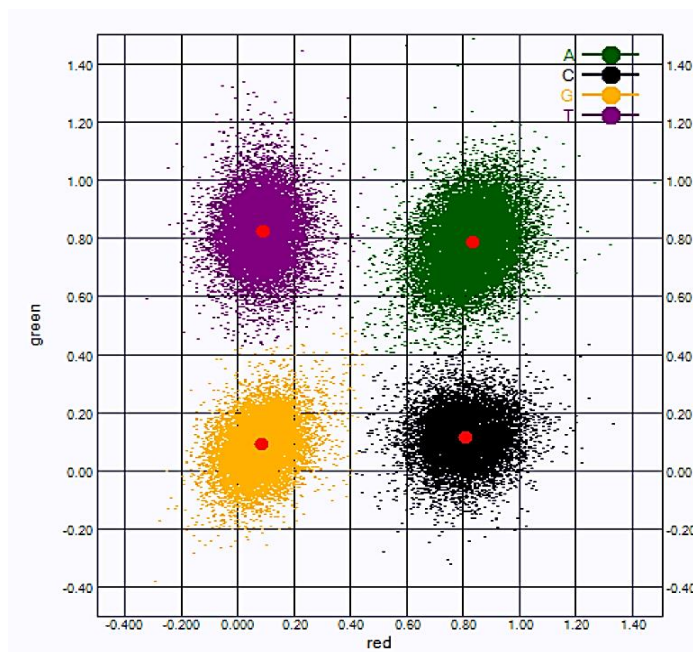


Figura 14. Intensidade luminosa emitida pelas chamadas de bases em diferentes *clusters* na superfície da célula de fluxo. Fonte: ILLUMINA SEQUENCING, 2016.

Foram empregados dois tipos de tecnologias de sequenciamento NGS da Illumina®, o sequenciamento NextSeq® e o sequenciamento MiniSeq®, referente às amostras de *C. similis* e *C. innominatus*, respectivamente. Tratam-se de sistemas de sequenciamento de alto rendimento com finalidade de sequenciar genomas completos, ambos consistem em uso de KIT's únicos, constituído por reagentes químicos e uma célula de fluxo, diferem apenas na quantidade de volume de dados acumulado e no tempo máximo de sequenciamento.

3.3.3.1. Sequenciamento NextSeq®500

O sequenciamento NextSeq® (**Figura 15**) apresenta uma carga máxima de rendimento um valor de 120 Gb (giga bases) de dados em até 30 horas de sequenciamento.

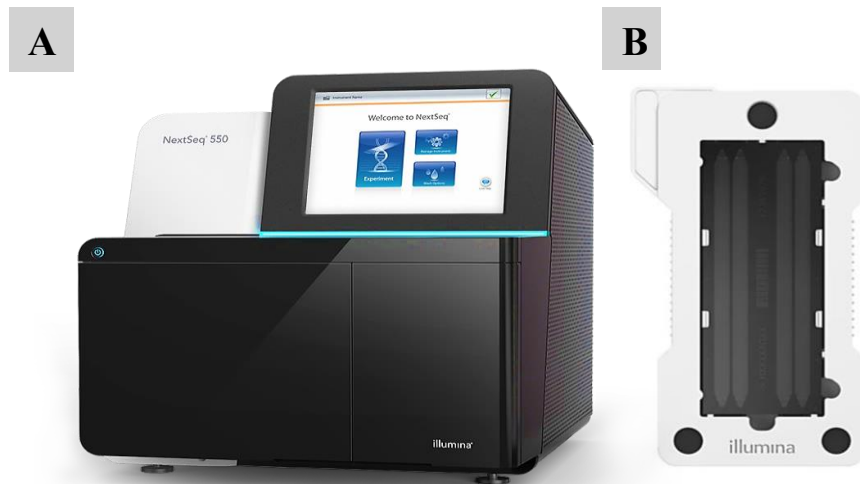


Figura 15. Representação do Sequenciador NextSeq® e sua célula de fluxo. **(A)** máquina de sequenciamento NGS NextSeq® 500/550. **(B)** célula de fluxo específica para o sequenciador NextSeq®. Fonte: Adaptador de < Illumina_Sequencing_Overview_15045845_D.pdf>

3.3.3.2. Sequenciamento MiniSeq®

O sequenciamento MiniSeq® (**Figura 16**) apresenta como capacidade máxima de rendimento um total de 7,5 Gb de dados em um máximo de 24 horas de sequenciamento.



Figura 16. Representação do Sequenciador MiniSeq® e sua célula de fluxo. **A)** máquina de sequenciamento NGS MiniSeq. **B).** célula de fluxo específica para o sequenciador MiniSeq. Fonte: Adaptado de < Illumina_Sequencing_Overview_15045845_D.pdf>

3.4. ANÁLISES DOS DADOS DE SEQUENCIAMENTO DE DNA

As etapas referentes aos itens 3.4.1, 3.4.2, 3.4.3. e 3.4.4 (referentes a procedimentos de tratamentos de qualidade pós-sequenciamento, montagem e anotação dos genomas, respectivamente) estão relacionadas à procedimentos de bioinformática, no qual ocorreu por supervisão e orientação do doutorando em bioinformática Renato R. M. Oliveira.

3.4.1. Tratamento Pós-Sequenciamento dos dados moleculares

Com a conclusão do sequenciamento, todo o volume de dados gerados é armazenado na plataforma do BaseSpace. Nesta plataforma inicia-se as primeiras etapas do tratamento das sequencias geradas. A primeira etapa chama-se Desmultiplexação (*demultiplexing*), que se refere a identificação/organização/remoção das sequências index únicas associadas a cada amostra, que foram adicionadas na etapa da construção das bibliotecas. Resultando em dados de sequenciamento desmultiplexados e convertidos em arquivos ‘FASTQ R1’ para a fita sintetizada *forward* e ‘FASTQ R2’ para a fita sintetizada *reverse* de cada amostra de DNA.

O controle de qualidade para dados de sequência de alto rendimento é averiguado através da ferramenta FastQC (**Figura 17**) que fornece um conjunto de análises pré-estabelecidas que são visualizadas por gráficos e estatísticas. Sendo os parâmetros mais importantes a serem avaliados:

- a) O *score* de qualidade de cada base;
- b) A quantidade de bases não identificadas (N);
- c) O conteúdo GC;
- d) A presença de sequencias de adaptadores index.

Sendo assim, foi utilizado a ferramenta PRINSEQ para remoção das sequencias de bases com baixa qualidade (*score* Phred < 30) identificadas no FastQC (ANDREWS, 2016; VICENTE *et al.*, 2017). Valores referentes ao tratamento pós-sequenciamento dos genomas de *Cyphoderus* é encontrado no **APÊNDICE D**.

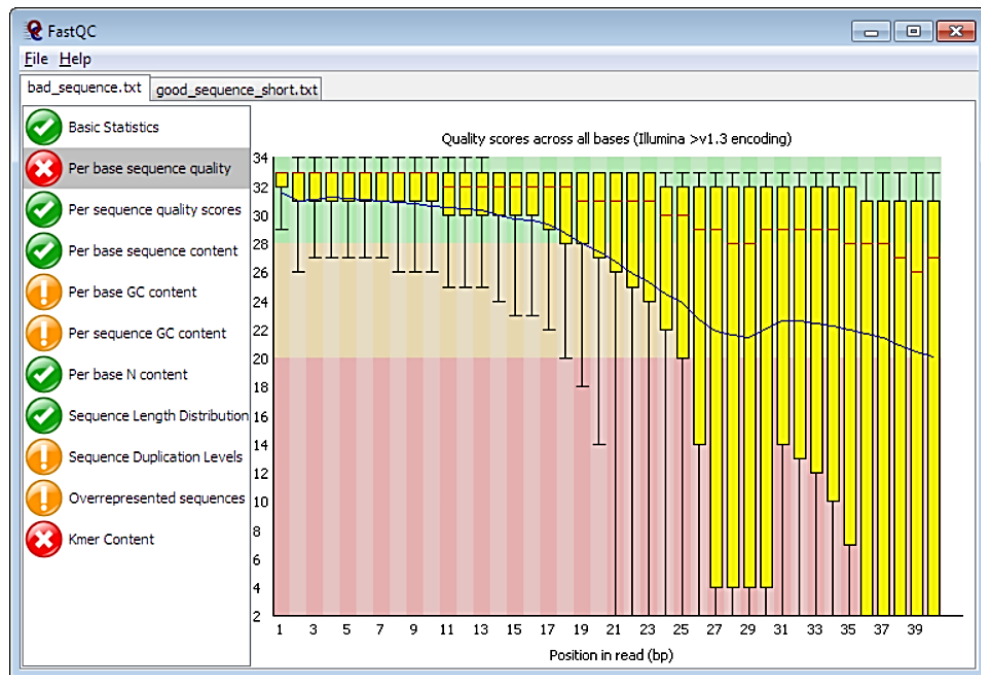


Figura 17. Ferramenta FastQC para tratamento de qualidade de dados de sequenciamento de alto rendimento, onde se destaca o *status* de todos os parâmetros avaliados. Fonte: <<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>

3.4.2. Montagem dos genomas

3.4.2.1. Montagem *de novo* dos Genomas Mitocondriais

A montagem dos genomas mitocondriais se deu a partir do *software* de código aberto NOVOPlasty v2.7.2. que se baseia em *seed-extend*, onde a ‘*seed*’ refere-se a uma sequência referência (geralmente um gene conservado ou até um genoma mitocondrial de organismos filogeneticamente próximo) que servirá apenas para extrair as leituras mitocondriais dentre todas leituras provenientes do sequenciamento. A *seed* utilizada foi o gene Citocromo Oxidase subunidade I (COI) completo da espécie *F. candida*. Com *k-mer* padrão de 38 (representando um conjunto de 38 nucleotídeos lidos em sequência em cada *read*).

3.4.2.2. Montagem *de novo* dos Genomas Nucleares

Foram utilizados dois montadores de sequências NGS, sendo eles:

- a) MEGAHIT - um montador otimizado para metagenomas ou genomas grandes e complexos via grafo sucinto De Bruijn. Foram estabelecidos *k-mer* de 21, 33, 55 e 77 (BOWE *et al.*, 2012; LI *et al.*, 2016) no qual foi montado através das sequências P-E e sequências *singletons* de M-P (sequências montadas sem estarem pareadas da biblioteca M-P);
- b) SOAPdenovo2 - um montador de leituras curtas projetado para montar grandes genomas eucariotos (plantas e animais) com *k-mer* de 41 (LUO *et al.*, 2012) no qual foi montado por sequências M-P e P-E.

Após a montagem realizada com SOAPdenovo2 foi necessário unir tanto os *contigs* quanto os *scaffolds* gerados com o auxílio da ferramenta SSPACE (BOETZER *et al.*, 2010), um script para estender *contigs* e *scaffolds* pré-montados, e foi preciso fechar *gaps* encontrados durante cada ciclo de montagem de *contigs* e *scaffolds* utilizando a ferramenta GapCloser.

3.4.3. Anotação dos genomas

3.4.3.1. Anotação *ab initio* dos Genomas Mitocondriais

A anotação *ab initio* dos genomas mitocondriais se deu pelo software MITOS, que fornece acesso a um pipeline automatizado baseado em estratégias de agregação de similaridade por pesquisas do BLAST (*Basic Local Alignment Search Tool*) em sequências de proteínas previamente anotadas e por modelos específicos de covariância na anotação de RNAt e RNAr.

O MITOS investiga um intervalo de até seis aminoácidos a montante e a jusante para cada predição de genes, evitando possíveis erros de anotação, sendo considerado como uma ferramenta adequada para anotações *ab initio* de alta qualidade para genomas mitocondriais, além de padronizar a nomenclatura e designação do limite de todos os *features* anotados no genoma (BERNT *et al.*, 2013).

3.4.3.2. Anotação *ab initio* dos Genomas Nucleares

Com o software AUGUSTUS foi realizada a predição gênica nas montagens realizadas (KELLER *et al.*, 2011). Foi realizado o upload das sequências montadas e escolhido o genoma de referência a servir como fonte de extração de informações gênicas, logo após a anotação

concluída é baixado o arquivo contendo os éxons, íntrons e arquivos com as sequências gênicas anotadas com suas estatísticas básicas. Foi usado como base de comparação o genoma de *D. melanogaster*.

3.4.4. Análise estrutural dos genomas

Através do software Artemis 17.0.1. foi realizada uma caracterização dos mitogenomas, onde foram avaliados:

- a) A composição nucleotídica total;
- b) A composição nucleotídica de cada *start* e *stop* códons das CDS's encontradas.

Com este programa também foi realizado a curadoria manual do genoma mitocondrial, através da avaliação de seus nucleotídeos e códons, ajustando trincas, averiguando os códons de iniciação e terminação. Com auxílio do banco de dados de proteínas Pfam (*Protein families database*) foi averiguado as sequencias de todos os conjuntos de CDS anotadas nos mitogenomas para, se necessário, correção de sequências na etapa de curadoria. Com os mitogenomas devidamente curados, foram realizadas suas submissões ao GenBank, um banco de dados público, constituído e distribuído pelo NCBI (Centro Nacional de Informações sobre Biotecnologia).

O software Geneious de análise, interpretação e aplicação de dados moleculares de DNA foi utilizado para geração das imagens referentes aos genomas que serão apresentados em breve.

Através da ferramenta de pré-processamento e informações de sequencias PRINSEQ foram extraídas as principais estatísticas estruturais do genoma nuclear como:

- a) Seu tamanho, medido em pares de base ou suas variações;
- b) A quantidade e tamanho de *contigs* ou *scaffolds*;
- c) A porcentagem nucleotídica para cada base do genoma;
- d) O valor GC, o qual contribui para uma inferência da normalidade do genoma para regiões ricas em genes (AÍSSANI & BERNARDI, 1991; SUMNER *et al.*, 1993);
- e) O valor de N50, que refere-se a qualidade da montagem no sentido de contiguidade, ou seja, representa o tamanho (pb) do *contig/scaffold* que contém 50% do tamanho do genoma, quando ordenado de forma crescente, isto infere que metade das sequencias do genoma está em *contigs/scaffolds* maiores ou iguais ao N50 (MILLER *et al.*, 2010);
- f) A quantidade de sequências ainda desconhecidas.

O software MEGAX foi utilizado para algumas outras análises estruturais e funcionais, como a frequência de uso de códons e composição de aminoácidos (baseado em TAMURA *et al.*, 2011).

3.4.5. Análise comparativa dos genomas

Os genomas foram comparados entre eles, para avaliar o grau de similaridade nucleotídica entre os *Cyphoderus*, além de seguir comparações com outros genomas de Collembola (mitocondriais e nucleares) disponíveis e com alto grau de cobertura. Segue na **Tabela 1** a lista das espécies de Collembola com genomas nucleares disponíveis para comparação, enquanto na **Tabela 2** a lista das espécies de Collembola com genomas mitocondriais disponíveis para comparação.

Tabela 1. Lista das espécies de Collembola alvo para análises comparativas baseado em sequências de DNA nuclear. Foram baixados no Genbank através de seus códigos de acesso para genomas completos disponibilizados. As Ordens de Collembola estão representadas por Entomobryomorpha (E) e Poduromorpha (P).

Ordem	Família	Espécie	Número de Acesso
E	Isotomidae	<i>Folsomia candida</i>	ASM221717v1
	Entomobryidae	<i>Sinella curviseta</i>	ASM411504v1
		<i>Orchesella cincta</i>	ASM171814v1
P	Neanuridae	<i>Holacanthella duospinosa</i>	ASM273828v1

Tabela 2. Lista das espécies de Collembola alvo para as análises comparativas baseadas em sequências de DNA mitocondrial. Foram baixados no Genbank através de seus códigos de acesso para genomas mitocondriais disponibilizados. As Ordens de Collembola estão representadas por Entomobryomorpha (E), Poduromorpha (P) e Symphypleona (S).

Ordem	Família	Espécie	Número de Acesso
E	Isotomidae	<i>Cryptopygus antarcticus</i>	EU016194.1
		<i>Cryptopygus terranovus</i>	NC_037610.1
		<i>Folsomia candida</i>	KU198392.1
		<i>Folsomotoma octooculata</i>	KC862316.1
	Entomobryidae	<i>Orchesella cincta</i>	KT985987.1
		<i>Orchesella villosa</i>	EU016195.1
P	Neanuridae	<i>Bilobella aurantiaca</i>	EU084034.1
		<i>Friesea grisea</i>	EU124719.1

	Hypogastruridae	<i>Gomphiocephalus hodgsoni</i>	AY191995.1
	Onychiuridae	<i>Onychiurus orientalis</i>	AY639938.1
		<i>Tetradontophora bielanensis</i>	AF272824.1
	Poduridae	<i>Podura aquatica</i>	AY639939.1
S	Bourletiellidae	<i>Bourletiella arvalis</i>	NC_039558.1
	Sminthuridae	<i>Sminthurus viridis</i>	EU016192.1

As comparações ocorrerão através dos softwares de alinhamento múltiplo MAFFT v7.388 e Clustal Omega onde os genomas serão comparados e construído uma matriz de identidade com base nas divergências de seus nucleotídeos, além de valores de similaridade para cada análise comparativa realizada, baseado em comparações por alinhamento realizada por FRATI e colaboradores (1997, 2000b) e NAKAMORI (2013).

3.4.6. Análise filogenética dos genomas

Foi realizado uma busca por genomas mitocondriais de Collembola no NCBI, o qual foram encontrados 14 genomas mitocondriais acessíveis, porém foi utilizado apenas 13 genomas mitocondriais devido a qualidade de suas sequências gênicas. Foi excluído da análise filogenética o genoma mitocondrial da espécie *Bourletiella arvalis*.

Como grupo externo foi escolhida a espécie *Daphnia magna* (número de acesso NC_026914.1) representante do grupo Crustacea, o qual seu genoma também foi buscado e baixado através do GenBank.

Após a obtenção dos genomas mitocondriais foi preciso concatenar a ordem gênica dos genomas para cada espécie e posteriormente alinhá-los utilizando o MAFFT v7.388, uma ferramenta ideal para alinhamentos de múltiplas sequências com arquivo de entrada no formato FASTA, baseado na metodologia de alinhamento de MATEOS e colaboradores (2018).

Para a reconstrução filogenética dos genomas mitocondriais foi utilizado o CIPRES com RAxML para uma análise de Máxima Verossimilhança (ML), com um total de 1000 replicações e modelo GTRGAMMA. A topologia foi editada e visualizada pelo FIGTREE v1.4.3. com *bootstrap* de 1000 e valores acima de 60% foram considerados significativos (baseado em CHEN *et al.*, 2014).

4. RESULTADOS E DISCUSSÃO

4.1. ANÁLISES ESTRUTURAIS E FUNCIONAIS DOS GENOMAS

A seguir são apresentadas as análises da estrutura e das anotações gênicas dos 2 genomas de *Cyphoderus* sequenciados, primeiramente referente aos genomas mitocondriais e posteriormente referente aos genomas nucleares.

4.1.1. Análise Estrutural e Funcional dos Genomas Mitocondriais

A **Tabela 3** mostra os valores referentes às características estruturais e funcionais (referente à anotação gênica) resumidas obtidas a partir dos sequenciamentos dos genomas mitocondriais de *C. similis* e *C. innominatus*.

Tabela 3. Estatísticas básicas da montagem *de novo* dos genomas mitocondriais de *Cyphoderus*. *Acesso disponibilizado após publicação dos genomas.

	<i>Cyphoderus similis</i>	<i>Cyphoderus innominatus</i>
Comprimento Total	14988 bp	14811 bp
%A	37.1%	37.6%
%T	34.9%	33.5%
%G	10.9%	10.9%
%C	17.1%	17.9%
CDS	13	13
tRNA	22	21
rRNA	2	2
Número de Acesso GB	MK510939*	MK510940*

4.1.1.1. Genoma Mitocondrial de *Cyphoderus similis*

Observa-se na **Figura 18** a representação gráfica do mitogenoma de *C. similis* (amostra ITV10629) e seu conjunto gênico.

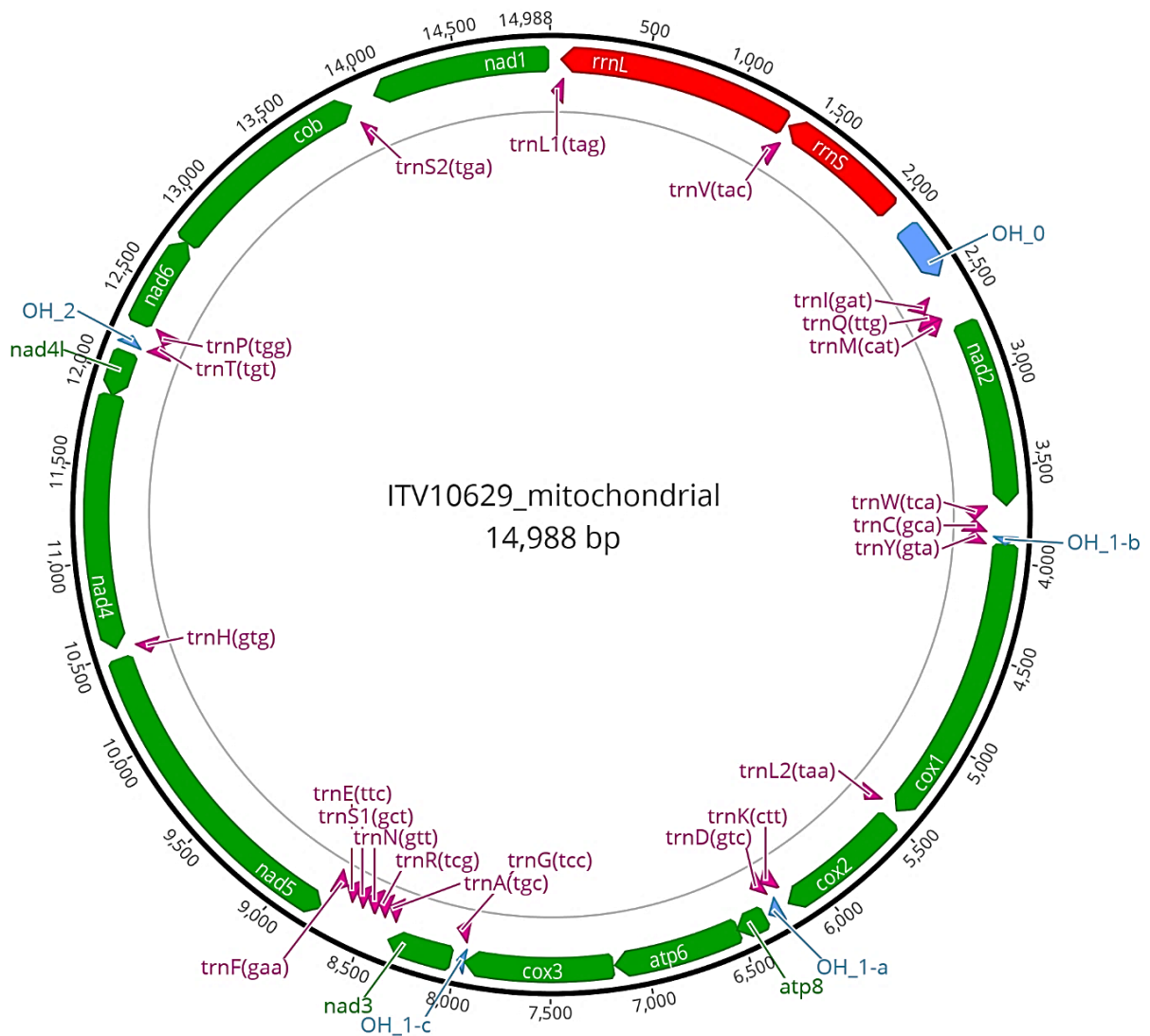


Figura 18. Genoma mitocondrial de *Cyphoderus similis*. Em verde está representado o conjunto de CDS; em vermelho o conjunto de rRNA (rrnL – 16S; rrnS – 12S); em rosa o conjunto de tRNA (especializados para cada aminoácido); e em azul o conjunto de Origem de Replicação das fitas (OH).

Os atributos dos genes anotados podem ser encontrados na **Tabela 4**, que traz a sua posição na molécula de DNA mitocondrial, o comprimento de cada uma das CDS's além de evidenciar o conjunto de *start* e *stop* códons que foram identificados no mitogenoma de *C. similis*.

Acerca da tradução do mtDNA de *C. similis*, observa-se a frequência de uso de códons para a espécie na **Tabela 5**.

Tabela 4. Atributos estruturais referentes aos genes anotados no genoma mitocondrial de *Cyphoderus similis*. Referente a diferenciação das fitas de mtDNA temos: fita *forward* (F) e fita *reverse* (R).

	Comprimento (bp)	Start-Stop Códon	Posição da CDS	Fita de mtDNA
NAD2	993	ATT - TAA	2568 / 3560	F
COI	1536	ATT - TAA	3807 / 5342	F
COII	696	ATG - TAA	5408 / 6103	F
ATP8	162	ATT - TAA	6303 / 6467	F
ATP6	681	ATG - TAA	6468 / 7148	F
COIII	789	ATG - TAA	7149 / 7937	F
NAD3	345	ATT - TAG	8036 / 8380	F
NAD5	1715	ATT - TTA	8772 / 10486	R
NAD4	1344	ATG - TAA	10550 / 11893	R
NAD4I	237	ATG - TAA	11894 / 12130	R
NAD6	495	ATA - TAA	12303 / 12797	F
COB	1134	ATG - TAA	12798 / 13931	F
NAD1	933	ATG - TAG	14002 / 14934	R

Tabela 5. Tabela de uso e frequência de códons do mtDNA de *Cyphoderus similis*. Sendo a média de códons de 4996.

UUU(F)	265.0 (1.40)	UCU(S)	110.0 (1.81)	UAU(Y)	241.0 (1.49)	UGU(C)	52.0 (1.17)
UUC(F)	114.0 (0.60)	UCC(S)	49.0 (0.80)	UAC(Y)	82.0 (0.51)	UGC(C)	37.0 (0.83)
UUA(L)	235.0 (2.40)	UCA(S)	79.0 (1.30)	UAA(*)	256.0 (1.61)	UGA(W)	55.0 (1.16)
UUG(L)	61.0 (0.62)	UCG(S)	18.0 (0.30)	UAG(*)	63.0 (0.39)	UGG(W)	40.0 (0.84)
CUU(L)	103.0(1.05)	CCU(P)	52.0 (0.99)	CAU(H)	72.0 (1.13)	CGU(R)	17.0 (1.01)
CUC(L)	49.0(0.50)	CCC(P)	67.0 (1.27)	CAC(H)	55.0 (0.87)	CGC(R)	19.0 (1.13)
CUA(L)	110.0(1.12)	CCA(P)	77.0 (1.46)	CAA(Q)	89.0 (1.34)	CGA(R)	21.0 (1.25)
CUG(L)	30.0 (0.31)	CCG(P)	15.0 (0.28)	CAG(Q)	44.0 (0.66)	CGG(R)	10.0 (0.60)
AUU(I)	248.0 (1.47)	ACU(T)	91.0 (1.45)	AAU(N)	277.0 (1.36)	AGU(S)	70.0 (1.15)
AUC(I)	89.0 (0.53)	ACC(T)	62.0 (0.99)	AAC(N)	131.0 (0.64)	AGC(S)	62.0 (1.02)
AUA(M)	202.0 (1.65)	ACA(T)	81.0 (1.29)	AAA(K)	351.0 (1.67)	AGA(S)	55.0 (0.90)
AUG(M)	43.0 (0.35)	ACG(T)	17.0 (0.27)	AAG(K)	69.0 (0.33)	AGG(S)	44.0 (0.72)
GUU(V)	58.0 (1.57)	GCU(A)	40.0 (1.37)	GAU(D)	48.0 (1.48)	GGU(G)	32.0 (1.16)
GUC(V)	17.0 (0.46)	GCC(A)	32.0 (1.09)	GAC(D)	17.0 (0.52)	GGC(G)	16.0 (0.58)
GUA(V)	61.0 (1.65)	GCA(A)	39.0 (1.33)	GAA(E)	50.0 (1.30)	GGA(G)	30.0 (1.09)
GUG(V)	12.0 (0.32)	GCG(A)	6.0 (0.21)	GAG(E)	27.0 (0.70)	GGG(G)	32.0 (1.16)

4.1.1.2. Genoma Mitocondrial de *Cyphoderus innominatus*

Na **Figura 19** visualiza-se graficamente o mitogenoma de *C. innominatus* (amostra ITV10660) e seu conjunto de genes anotados.

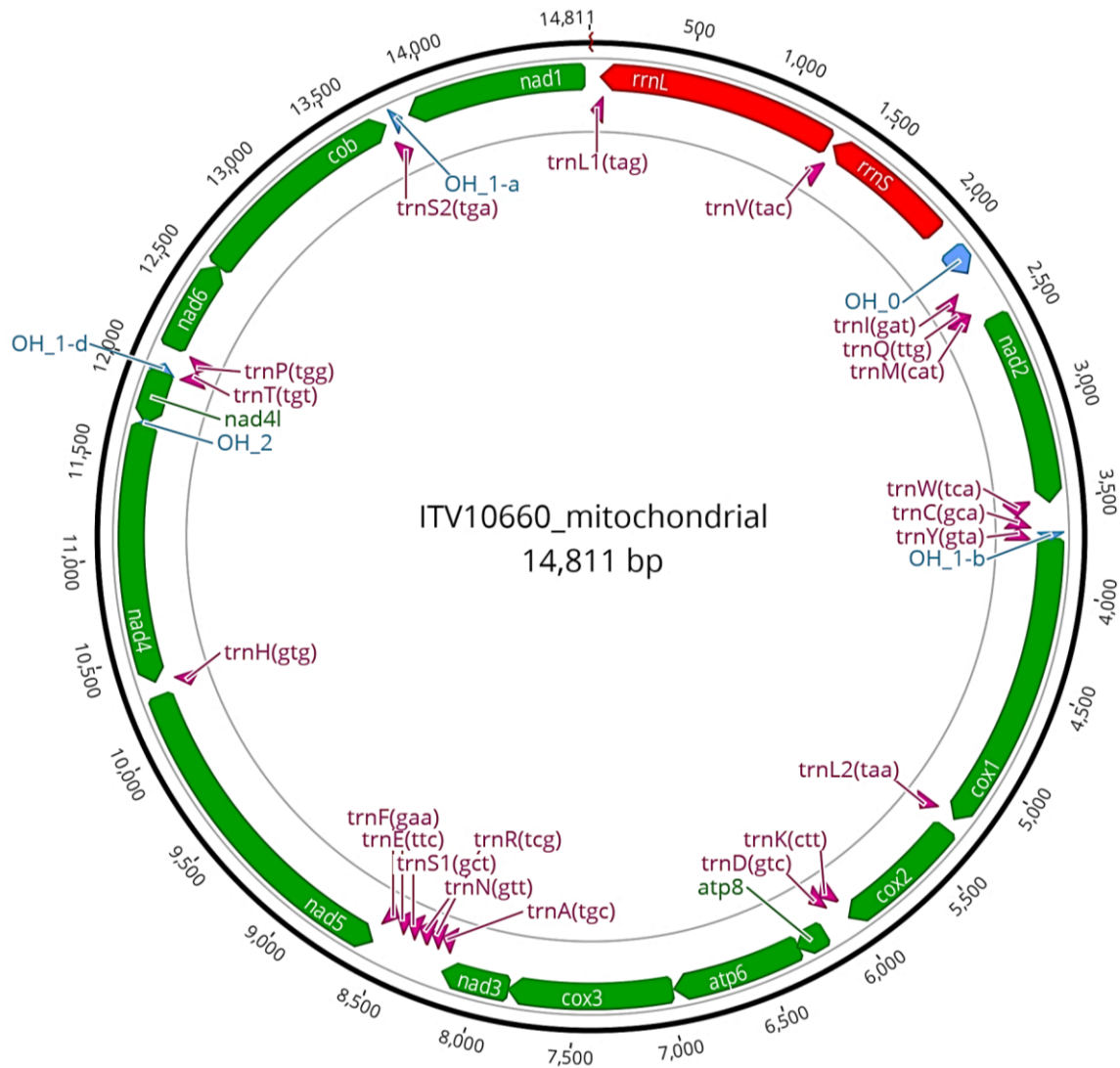


Figura 19. Genoma mitocondrial de *Cyphoderus innominatus*. Em verde está representado o conjunto de CDS; em vermelho o conjunto de rRNA (rrnL – 16S; rrnS – 12S); em rosa o conjunto de tRNA (especializados para cada aminoácido); e em azul o conjunto de Origem de Replicação das fitas (OH).

Encontra-se na **Tabela 6** os atributos gênicos anotados no mitogenoma de *C. innominatus*, com informações de posição na molécula de mtDNA, de comprimento e seus respectivos *start* e *stop* códons.

Relacionado ao conjunto nucleotídico traduzido, temos a frequência de uso de códons para a espécie *C. innominatus* na **Tabela 7**. Sendo também construída a partir do código genético padrão para invertebrados.

Tabela 6. Atributos estruturais referentes aos genes anotados no genoma mitocondrial de *Cyphoderus innominatus*. Referente a diferenciação das fitas de mtDNA temos: fita *forward* (F) e fita *reverse* (R).

	Comprimento (bp)	Start-Stop códon	Posição da CDS	Fita de mtDNA
NAD2	996	ATT - TAA	2530 / 3525	F
COI	1545	ATT - TAA	3716 / 5260	F
COII	681	ATG - TAG	5328 / 6008	F
ATP8	159	ATA - TAA	6153 / 6311	F
ATP6	681	ATG - TAA	6305 / 6985	F
COIII	849	ATG - TAA	6988 / 7836	F
NAD3	345	ATT - TAG	7841 / 8185	F
NAD5	1701	ATT - TAA	4532 / 6232	R
NAD4	1341	ATG - TAA	3118 / 4458	R
NAD4I	276	ATT - TAA	2835 / 3110	R
NAD6	486	ATA - TAA	12111 / 12596	F
COB	1134	ATG - TAA	12596 / 13729	F
NAD1	927	ATT - TAG	25 / 951	R

Tabela 7. Tabela de uso e frequência de códons do mtDNA de *Cyphoderus innominatus*. Sendo a média de códons de 4937.

UUU(F)	223.0 (1.51)	UCU(S)	99.0 (1.39)	UAU(Y)	243.0 (1.52)	UGU(C)	55.0 (1.22)
UUC(F)	72.0 (0.49)	UCC(S)	52.0 (0.73)	UAC(Y)	77.0 (0.48)	UGC(C)	35.0 (0.78)
UUA(L)	195.0 (2.26)	UCA(S)	91.0 (1.28)	UAA(*)	273.0 (1.70)	UGA(W)	66.0 (1.33)
UUG(L)	48.0 (0.56)	UCG(S)	23.0 (0.32)	UAG(*)	49.0 (0.30)	UGG(W)	33.0 (0.67)
CUU(L)	113.0 (1.31)	CCU(P)	65.0 (1.07)	CAU(H)	87.0 (1.23)	CGU(R)	13.0 (0.71)
CUC(L)	56.0 (0.65)	CCC(P)	92.0 (1.51)	CAC(H)	55.0 (0.77)	CGC(R)	23.0 (1.26)
CUA(L)	79.0 (0.92)	CCA(P)	61.0 (1.00)	CAA(Q)	94.0 (1.58)	CGA(R)	15.0 (0.82)
CUG(L)	26.0 (0.30)	CCG(P)	26.0 (0.43)	CAG(Q)	25.0 (0.42)	CGG(R)	22.0 (1.21)
AUU(I)	256.0 (1.46)	ACU(T)	80.0 (1.20)	AAU(N)	284.0 (1.42)	AGU(S)	70.0 (0.99)
AUC(I)	94.0 (0.54)	ACC(T)	75.0 (1.12)	AAC(N)	117.0 (0.58)	AGC(S)	83.0 (1.17)
AUA(M)	239.0 (1.70)	ACA(T)	87.0 (1.30)	AAA(K)	321.0 (1.70)	AGA(S)	81.0 (1.14)
AUG(M)	42.0 (0.30)	ACG(T)	25.0 (0.37)	AAG(K)	57.0 (0.30)	AGG(S)	69.0 (0.97)
GUU(V)	58.0 (1.62)	GCU(A)	32.0 (1.25)	GAU(D)	45.0 (1.34)	GGU(G)	22.0 (1.04)
GUC(V)	17.0 (0.48)	GCC(A)	34.0 (1.33)	GAC(D)	22.0 (0.66)	GGC(G)	21.0 (0.99)
GUA(V)	57.0 (1.59)	GCA(A)	33.0 (1.29)	GAA(E)	56.0 (1.51)	GGA(G)	25.0 (1.18)
GUG(V)	11.0 (0.31)	GCG(A)	3.0 (0.12)	GAG(E)	18.0 (0.49)	GGG(G)	17.0 (0.80)

4.1.2. Análise Estrutural e Funcional dos Genomas Nucleares

Na **Tabela 8** contém os resultados relacionados à estrutura genômica nuclear obtidos à partir do sequenciamento de DNA de *C. similis*.

Tabela 8. Características estruturais referente ao genoma nuclear de *Cyphoderus similis*. Comparando as duas ferramentas de montagens de genomas nucleares.

	MEGAHIT	SOAPdenovo2
Nº Scaffolds	219.658	249.620
Total Bases	303.394.992	357.912.514
Comprimento Mínimo (bp)	340	362
Comprimento Máximo (bp)	60.885	38.814
Comprimento Médio (bp)	1.381,22	1.433,83
N50 (bp)	2.628	3.293
Conteúdo GC	35,8%	33,16%
Sequências 'N'	6,38%	45,07%

Agora, na **Tabela 9** encontra-se os resultados do sequenciamento nuclear de *C. innominatus*.

Tabela 9. Características estruturais referente ao genoma nuclear de *Cyphoderus innominatus*. Comparando as duas ferramentas de montagens de genomas nucleares.

	MEGAHIT	SOAPdenovo2
Nº Scaffolds	249.982	249.310
Total Bases	338.881.231	370.224.676
Comprimento Mínimo (bp)	571	495
Comprimento Máximo (bp)	33.969	27.801
Comprimento Médio (bp)	1.355,62	1.485,0
N50 (bp)	1.518	3.216
Conteúdo GC	35,81%	32,40%
Sequências 'N'	6,39%	45,65%

Com isso, foi determinada a montagem pela ferramenta SOAPdenovo2 como o genoma menos fragmentado e com grandes *scaffolds* apresentados, representando 50% do tamanho total do genoma (maiores N50), em comparação com os genomas montados pela ferramenta MEGAHIT.

Acerca da estrutura funcional, é visto na **Tabela 10** um resumo sobre a composição de genes, éxons e íntros, além da densidade de genes em cada sequência funcional encontrada no genoma nuclear de *C. innominatus*. Comparação baseada na estrutura funcional do genoma da mosca-da-fruta (*D. melanogaster*), um dos principais modelos utilizado na comparação de genes.

Tabela 10. Composição funcional do genoma nuclear de *Cyphoderus innominatus* anotado por referência do genoma de *Drosophila melanogaster*.

Atributos	Valores
Nº Genes	278.865
Nº Éxons	73.812
Nº Íntros	50.624
Densidade de Genes	3.8

Observa-se na **Figura 20** a localização e o tamanho dos genes presentes em três dos *scaffolds* montados do genoma de *C. innominatus*.

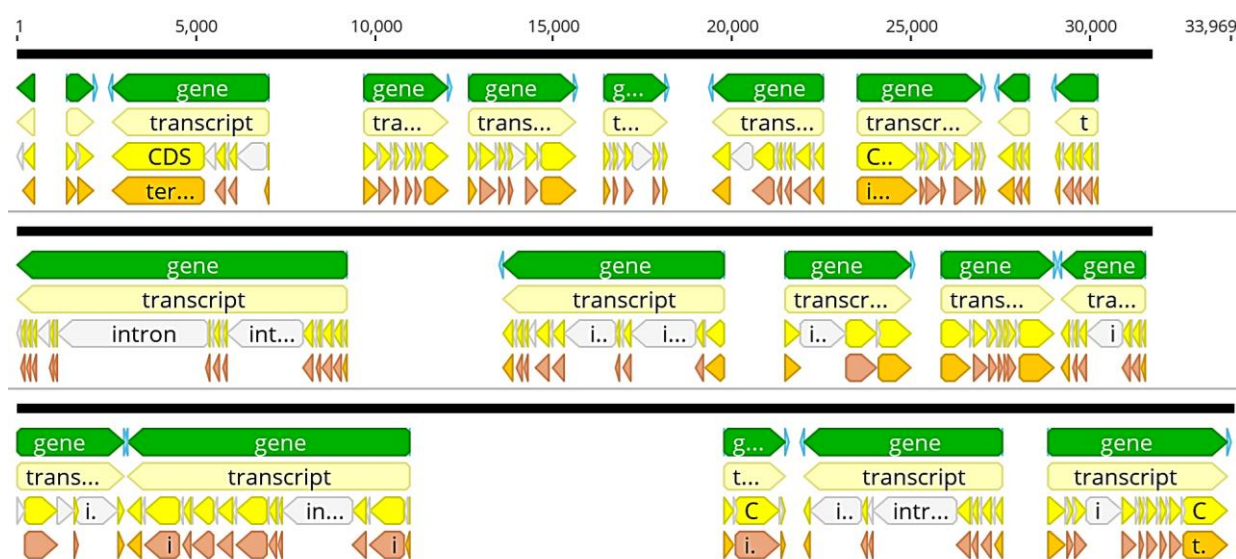


Figura 20. Anotação dos genes presentes em *scaffolds* do genoma de *C. innominatus*. Sendo a ordem: *scaffold 1* (acima) com tamanho 31.655 bp; *scaffold 2* (central) com tamanho 31.690 bp; e o *scaffold 34* (abaixo) com tamanho 33.968 bp

O sequenciamento genômico vem produzindo descobertas aplicáveis a muitas linhas de pesquisas, ainda por cima ao se tratar de genomas *de novo* no qual acrescentam novos ramos à árvore da vida, da ótica molecular. Uma vez que a biologia moderna vem sendo aplicada com

tendência apenas em organismos modelo, o progresso dos sequenciamentos de nova geração e métodos computacionais aplicados à bioinformática vem beneficiando e proporcionando progressos à exploração de organismos não-modelo, de certa forma, negligenciados, porém com potencial científico ainda desconhecido (VAN STRAALLEN *et al.*, 2008).

4.2. ANÁLISES COMPARATIVAS DOS GENOMAS

A seguir são apresentadas as análises de comparações realizadas entre os genomas sequenciados de *Cyphoderus* (inicialmente entre os genomas mitocondriais e, posteriormente entre os genomas nucleares) e entre outros genomas de Collembola publicados e com acessos disponíveis em bancos de dados online.

4.2.1. Comparações entre os Genomas Mitocondriais

Relacionando os dois mitogenomas de *Cyphoderus* sequenciados, observa-se muitas semelhanças, principalmente referentes a ordem dos conjuntos de genes e suas posições bastante similares na dupla fita de DNA mitocondrial, como pode ser visto na **Figura 21** e na **Figura 22** que representa um recorte do alinhamento múltiplo entre todas as 13 CDS anotadas. Já a composição de aminoácidos presentes em seus mitogenomas está presente na **Tabela 11**.

Ao comparar os mitogenomas sequenciados com os outros 14 genomas mitocondriais de Collembola disponibilizados no GenBank (levando em consideração tanto os genomas completos quanto os parcialmente completos) observa-se algumas semelhanças estruturais e funcionais na composição dos mtDNA de Collembola (**Tabela 12**).

Encontra-se na **Figura 23** os dados relacionados à matriz de identidade nucleotídica baseada no alinhamento múltiplo comparativo dos genomas mitocondriais inteiros analisados para Collembola. Já referente ao gene COI (citocromo oxidase I), por ser apresentado como um importante marcador molecular para delimitação e identificação molecular de espécies, encontra-se na **Figura 24** a matriz de similaridade referente a comparação do *Barcodes* genético das espécies de Collembola e os *Barcodes* sequenciados de *Cyphoderus*.

Tabela 11. Composição de aminoácidos contidos nos genomas mitocondriais sequenciados de *Cyphoderus*. Representados em porcentagem (%).

	<i>Cyphoderus similis</i>	<i>Cyphoderus innominatus</i>
Ala	2.50	2.21
Cys	1.90	1.95
Asp	1.39	1.45
Glu	1.65	1.60
Phe	8.10	6.39
Gly	2.35	1.84
His	2.72	3.08
Ile	7.21	7.58
Lys	8.98	8.19
Leu	12.57	11.20
Met	5.24	6.09
Asn	8.72	8.69
Pro	4.51	5.29
Gln	2.84	2.58
Arg	1.43	1.58
Ser	10.41	12.31
Thr	5.37	5.79
Val	3.16	3.10
Trp	2.03	2.15
Tyr	6.91	6.93
Total	4677	4615

Tabela 12. Composição estrutural e funcional comparativa dos genomas mitocondriais de Collembola disponíveis até então no GenBank em relação aos genomas mitocondriais de *Cyphoderus* sequenciados.

Espécie	Comprimento Total (bp)	CDS	tRNA	rRNA
<i>Cyphoderus similis</i>	14.988	13	22	2
<i>Cyphoderus innominatus</i>	14.811	13	21	2
<i>Cryptopygus antarcticus</i>	15.297	13	23	2
<i>Cryptopygus terranovus</i>	15.352	13	22	2
<i>Folsomia candida</i>	15.147	13	22	2
<i>Folsomotoma octooculata</i>	15.338	13	22	2
<i>Orchesella cincta</i>	15.728	13	22	2
<i>Orchesella villosa</i>	14.924	13	22	2
<i>Bilobella aurantiaca</i>	16.312	13	24	2
<i>Friesea grisea</i>	15.442	13	22	2
<i>Gomphiocephalus hodgsoni</i>	15.075	13	22	2
<i>Onychiurus orientalis</i>	12.984	13	20	1
<i>Tetradontophora bielensis</i>	15.455	13	22	2
<i>Podura aquatica</i>	13.809	13	20	1

<i>Bourletiella arvalis</i>	14.794	13	22	2
<i>Sminthurus viridis</i>	14.817	13	22	2

Ambos os genomas mitocondriais sequenciados incluem 13 genes codificadores de proteínas, 2 genes de RNA ribossômico, e 22 genes de RNA transportador (sendo 21 para *C. innominatus*) representando assim o conjunto gênico completo proposto por BOORE (1999) a partir do arranjo ancestral de genomas de Insecta.

Para Collembola, há registros de comparações de variância genética dentro e entre linhagens referentes ao marcador molecular Citocromo Oxidase subunidade I (COI), amplamente utilizado como ferramenta de delimitação taxonômica a nível de espécie (*Barcode*). Como é o caso de linhagens europeias de *Parisotoma notabilis* com baixa variância dentro de linhagens (<3%) e alta variância entre linhagens (21%), podendo então relacionar estas altas variâncias como preditor para linhagens crípticas com evolução independente e baixa diferenciação morfológica (VON SALTZWEDEL *et al.*, 2017). Sequências do gene COI também se mostrou útil na discriminação de linhagens do gênero *Ceratophysella* (NAKAMORI, 2013; PORCO *et al.*, 2012) e de linhagens europeias do gênero *Deutonura* (ROUGERIE *et al.*, 2006; PORCO *et al.*, 2010). Estudos mostraram variabilidade do COI também para *Isotoma riparia* comparada com *Isotoma viridis* (0,2 – 4,3%), no qual o autor relata que valores abaixo de 3% se encontra dentro dos limites intraespecíficos de divergência para este marcador molecular (HEBERT *et al.*, 2003; ANSLAN & TEDERSOO, 2015). Inclusive estudos apontam que este marcador apresenta uma divergência da sequência de 8 – 13,5% para espécies de Collembola em comparação com outros animais (HOGG & HEBERT, 2004; PORCO *et al.*, 2014). Assim, o uso do *Barcode* recém sequenciado para *Cyphoderus* mostrou-se eficaz na separação das espécies de *C. similis* e *C. innominatus*, uma vez que apresentaram maior similaridade a nível de gênero em comparação com outras espécies de Collembola, além de apresentarem variabilidade de sequência acima do limite descrito para delimitação de espécies (aproximadamente 20% de divergência).

1: Orchesella_cincta	100.00	43.96	42.67	43.84	44.24	44.53	43.21	43.35	42.30	43.98	43.14	42.55	45.19	44.51	44.74	45.26
2: Bilobella_aurantiaca	43.96	100.00	61.10	64.62	60.94	61.29	59.96	61.05	58.81	64.01	61.59	61.40	60.88	61.59	60.32	61.15
3: Podura_aquatica	42.67	61.10	100.00	63.06	61.20	61.05	60.36	61.99	60.21	64.35	62.46	60.74	62.06	61.83	61.90	62.38
4: Friesea_grisea	43.84	64.62	63.06	100.00	64.30	63.98	62.23	63.08	60.57	67.18	64.33	63.32	63.80	63.73	63.31	63.75
5: CYPHODERUS_INNOMINATUS	44.24	60.94	61.20	64.30	100.00	77.21	64.02	68.54	62.63	65.64	65.32	62.59	67.26	67.12	67.24	67.47
6: CYPHODERUS_SIMILIS	44.53	61.29	61.05	63.98	77.21	100.00	63.42	68.08	62.21	65.53	64.83	62.34	66.51	66.82	66.55	67.21
7: Bourletiella_arvalis	43.21	59.96	60.36	62.23	64.02	63.42	100.00	66.66	67.09	63.76	63.17	61.55	64.64	64.41	63.86	63.93
8: Orchesella_villosa	43.35	61.05	61.99	63.08	68.54	68.08	66.66	100.00	73.57	65.63	65.01	62.90	66.50	66.72	66.35	66.91
9: Sminthurus_viridis	42.30	58.81	60.21	60.57	62.63	62.21	67.09	73.57	100.00	62.65	61.85	60.57	62.48	62.85	61.95	62.78
10: Gomphiocephalus_hodgsoni	43.98	64.01	64.35	67.18	65.64	65.53	63.76	65.63	62.65	100.00	66.32	64.88	65.87	65.99	65.92	65.98
11: Tetrodontophora_bielanensis	43.14	61.59	62.46	64.33	65.32	64.83	63.17	65.01	61.85	66.32	100.00	69.51	64.80	65.95	65.60	65.95
12: Onychiurus_orientalis	42.55	61.40	60.74	63.32	62.59	62.34	61.55	62.90	60.57	64.88	69.51	100.00	63.63	64.04	63.22	64.01
13: Folsomia_candida	45.19	60.88	62.06	63.80	67.26	66.51	64.64	66.50	62.48	65.87	64.80	63.63	100.00	69.79	70.87	71.45
14: Folsomotoma_octooculata	44.51	61.59	61.83	63.73	67.12	66.82	64.41	66.72	62.85	65.99	65.95	64.04	69.79	100.00	70.13	71.46
15: Cryptopygus_antarcticus	44.74	60.32	61.90	63.31	67.24	66.55	63.86	66.35	61.95	65.92	65.60	63.22	70.87	70.13	100.00	74.21
16: Cryptopygus_terranovus	45.26	61.15	62.38	63.75	67.47	67.21	63.93	66.91	62.78	65.98	65.95	64.01	71.45	71.46	74.21	100.00

Figura 23. Matriz de identidade referente aos genomas mitocondriais de Collembola atualmente disponíveis no GenBank com adição dos mitogenomas de *Cyphoderus* sequenciados.

1: Orchesella_cincta	100.00	39.71	40.19	40.34	40.69	40.47	41.66	39.78	41.44	40.75	38.55	39.30	40.34	41.74	40.54	40.40
2: Sminthurus_viridis	39.71	100.00	75.64	73.62	75.07	77.49	74.54	73.50	74.05	75.29	72.80	71.73	74.17	72.67	73.71	74.84
3: Bourletiella_arvalis	40.19	75.64	100.00	76.22	76.81	76.55	76.88	76.03	77.47	78.18	72.37	73.02	75.57	77.01	74.40	76.55
4: CYPHODERUS_INNOMINATUS	40.34	73.62	76.22	100.00	81.64	78.40	77.41	76.69	77.51	78.10	72.73	74.05	75.86	75.93	73.65	76.66
5: CYPHODERUS_SIMILIS	40.69	75.07	76.81	81.64	100.00	78.29	76.30	77.41	76.47	77.51	73.71	74.28	77.43	76.45	74.89	76.66
6: Orchesella_villosa	40.47	77.49	76.55	78.40	78.29	100.00	79.14	76.34	77.84	79.01	73.15	75.08	76.94	77.14	75.18	77.18
7: Folsomotoma_octooculata	41.66	74.54	76.88	77.41	76.30	79.14	100.00	78.39	80.51	80.57	74.89	72.66	77.17	78.02	74.43	77.12
8: Folsomia_candida	39.78	73.50	76.03	76.69	77.41	76.34	78.39	100.00	80.18	80.64	71.82	73.36	75.28	76.17	73.44	75.10
9: Cryptopygus_antarcticus	41.44	74.05	77.47	77.51	76.47	77.84	80.51	80.18	100.00	83.25	73.61	74.38	75.64	77.73	75.77	76.34
10: Cryptopygus_terranovus	40.75	75.29	78.18	78.10	77.51	79.01	80.57	80.64	83.25	100.00	74.27	75.36	76.88	77.01	75.18	77.44
11: Bilobella_aurantiaca	38.55	72.80	72.37	72.73	73.71	73.15	74.89	71.82	73.61	74.27	100.00	73.71	75.54	76.32	72.41	73.87
12: Podura_aquatica	39.30	71.73	73.02	74.05	74.28	75.08	72.66	73.36	74.38	75.36	73.71	100.00	76.78	75.65	72.85	75.03
13: Friesea_grisea	40.34	74.17	75.57	75.86	77.43	76.94	77.17	75.28	75.64	76.88	75.54	76.78	100.00	80.43	76.19	77.47
14: Gomphiocephalus_hodgsoni	41.74	72.67	77.01	75.93	76.45	77.14	78.02	76.17	77.73	77.01	76.32	75.65	80.43	100.00	76.63	78.51
15: Onychiurus_orientalis	40.54	73.71	74.40	73.65	74.89	75.18	74.43	73.44	75.77	75.18	72.41	72.85	76.19	76.63	100.00	79.03
16: Tetrodontophora_bielanensis	40.40	74.84	76.55	76.66	76.66	77.18	77.12	75.10	76.34	77.44	73.87	75.03	77.47	78.51	79.03	100.00

Figura 24. Matriz de identidade referente aos *Barcodes* de Collembola atualmente disponíveis no GenBank com adição dos *Barcodes* de *Cyphoderus* sequenciados.

4.2.2. Comparação entre Genomas Nucleares

Tendo em vista os quatro genomas nucleares de Collembola sequenciados e disponíveis para análise, a **Tabela 13** representa uma avaliação comparativa a nível estrutural dos genomas nucleares de Collembola em relação aos dois novos genomas de *Cyphoderus* sequenciados.

De acordo com a anotação disponível para o genoma de *C. innominatus* segue abaixo uma análise comparativa dos genomas nucleares até então sequenciados e anotados para Collembola (**Tabela 14**).

Tabela 14. Análise funcional comparativa entre os genomas nucleares de Collembola atualmente sequenciados e o novo genoma de *Cyphoderus innominatus* devidamente anotado.

	<i>C. innominatus</i>	<i>F. candida</i>	<i>O. cincta</i>	<i>H. duospinosa</i>	<i>S. curviseta</i>
Genes	278.865	28.734	20.249	9.911	23.943
Exóns	73.812	197.859	118.474	79.659	133.951
Íntrons	50.624	524,921	336.337	242.640	381.850

Os genomas nucleares sequenciados para *Cyphoderus* apresentaram um nível de montagem semelhante aos outros genomas publicados (nível de *scaffold*), porém se observa uma alta fragmentação em ambas as montagens (seja MEGAHIT ou SOAPdenovo2), um baixo N50, além de uma baixa cobertura calculada quando comparados com os outros genomas nucleares de Collembola disponíveis. As tecnologias de sequenciamento utilizadas (NextSeq e MiniSeq) representam abordagens novas para montagens *de novo* de genomas de Collembola até então publicados, portanto a questão de trabalhar com *short reads* possa ter favorecido a alta fragmentação, questão a ser solucionada com o sequenciador de terceira geração (PacBio) que apresenta fragmentação do DNA por *long reads*, o que poderá satisfazer uma montagem final menos fragmentada e com maior cobertura para os genomas. Entretanto, consideramos os genomas nucleares aqui apresentados como a nível de *Draft Genomes*, ou seja, representam os primeiros esboços de um genoma parcialmente completo, apresentando um nível de anotação a ser considerável, mesmo por se apresentar como possíveis genes anotados, ainda necessita de mais algumas etapas para validação definitiva.

Tabela 13. Análise estrutural comparativa entre os genomas nucleares de Collembola atualmente sequenciados e os dois novos genomas de *Cyphoderus* sequenciados. *dados referentes à montagem pelo SOAPdenovo2. **calculada através da razão entre total de bases sequenciadas e o total de bases montadas.

	<i>C.</i> <i>similis</i> *	<i>C.</i> <i>innominatus</i> *	<i>F.</i> <i>candida</i>	<i>O.</i> <i>cincta</i>	<i>H.</i> <i>duospinosa</i>	<i>S.</i> <i>curviseta</i>
Total de Bases	357.912.514	370.224.676	221.702.752	286.766.519	327.566.771	381.458.724
Nº de Sequências	249.620	249.310	162	9.402	62.430	599
Comprimento Médio (bp)	1.433,83	1.485,0	1.368.535,51	30.500,59	5.246,94	636.825,92
N50 (bp)	3.293	3.216	6.519.406	65.879	310.226	3.284.409
Conteúdo GC	33,16%	32,40%	37.52%	36.81%	33.35%	37.51%
Número Sequências 'N'	45,07%	45,65%	28.40%	2.67%	45.35%	0.17%
Nível da Montagem	<i>Scaffold</i>	<i>Scaffold</i>	<i>Scaffold</i>	<i>Scaffold</i>	<i>Scaffold</i>	<i>Contig</i>
Montador	SOAPdenovo2	SOAPdenovo2	Falco	DBG2 OLC	SOAPdenovo	Canu
Sequenciador	Illumina NextSeq	Illumina MiniSeq/NextSeq	PacBio	PacBio/Illumina HiSeq	Illumina HiSeq	PacBio
Cobertura	6.19 x**	8.02 x**	58 x	40 x	180 x	110 x

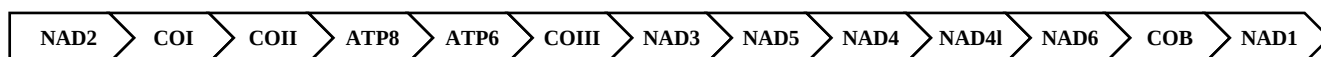
Abordagens comparativas estão se tornando muito aplicáveis no fornecimento de informações relevantes sobre a cobertura da montagem e da qualidade do genoma, principalmente no que se refere a *Draft Genomes*. Ao se tratar de uma análise de genomas proveniente de um sequenciamento com baixa cobertura e bibliotecas de inserções curtas (*short reads*) haverá portanto um desempenho de montagem muito abaixo do esperado para grandes genomas, pelo motivo de que *short reads* aumentam a complexidade dos algoritmos de montagem, proporcionando, assim, baixos N50 e alta proporção de fragmentos sem contiguidade, resultando em análises moleculares sem confiabilidade biológica (ZHANG *et al.*, 2019a; MILLER *et al.*, 2010).

O desafio de desvendar a biologia a partir de óticas diferentes só se torna possível ao passo que pesquisas descritivas vêm se realizando, como o sequenciamento de novos genomas e suas respectivas divulgações. Sendo assim, os primeiros *Draft Genomes* de *Cyphoderus* seguirão sendo aperfeiçoados em trabalhos futuros, visando, *a priori*, a diminuição de sua fragmentação com uma anotação gênica mais robusta, e, posteriormente, atingindo investigações mais profundas, como por exemplo aplicações em pesquisas de ecologia, biodiversidade e taxonomia de *Cyphoderus*.

4.3. FILOGENIA MOLECULAR DE COLLEMBOLA APLICANDO DOIS NOVOS GENOMAS MITOCONDRIAIS DE *Cyphoderus*

A proposta de topologia molecular que reflita a relação de parentesco filogenético para a Classe Collembola foi construída com base nos mtDNA de todo um conjunto de espécies que apresentem dados mitogenômicos completos disponíveis. A partir desse conjunto de mitogenomas disponíveis foi adicionado à análise filogenética os dois novos genomas montados referentes às espécies *Cyphoderus similis* e *C. innominatus*, primeiros exemplares do gênero *Cyphoderus* a fazer parte de uma filogenia molecular geral para a Classe.

Entretanto, algumas espécies que apresentam mitogenoma parciais foram excluídas da análise por apresentar genes fragmentados, foi preciso selecionar e concatenar apenas os genes que se apresentaram com melhor estrutura e qualidade para o alinhamento genômico. Com isso foi selecionada a maior quantidade de táxon possível e definido a ordem gênica seguinte:



Pode ser visto na **Figura 25** o resultado da inferência filogenética molecular a partir de dados mitogenômicos atuais de Collembola. Com *bootstrap* dos ramos variou de 60% a 100% e apresenta a *Daphnia magna* (NC_026914) como *outgroup*. A filogenia refletiu o monofiletismo da Classe Collembola, monofiletismo de algumas famílias representadas (**Figura 26**) e monofiletismo para os representantes da ordem Poduromorpha (**Figura 27**). As espécies *C. similis* e *C. innominatus*, representadas pelos genomas sequenciados, apresentaram-se como monofiléticas. Entretanto, a topologia apresenta algumas incoerências como a espécie *Sminthurus viridis*, a única representante da Ordem Symphypleona como táxon interno do grupo dos Entomobryidae, no qual pode ser um reflexo da baixa amostragem para a Classe Collembola e a falta de muitos táxons não representados como por exemplo grande parte da Ordem Symphypleona e a total falta de representantes da Ordem Neelipleona.

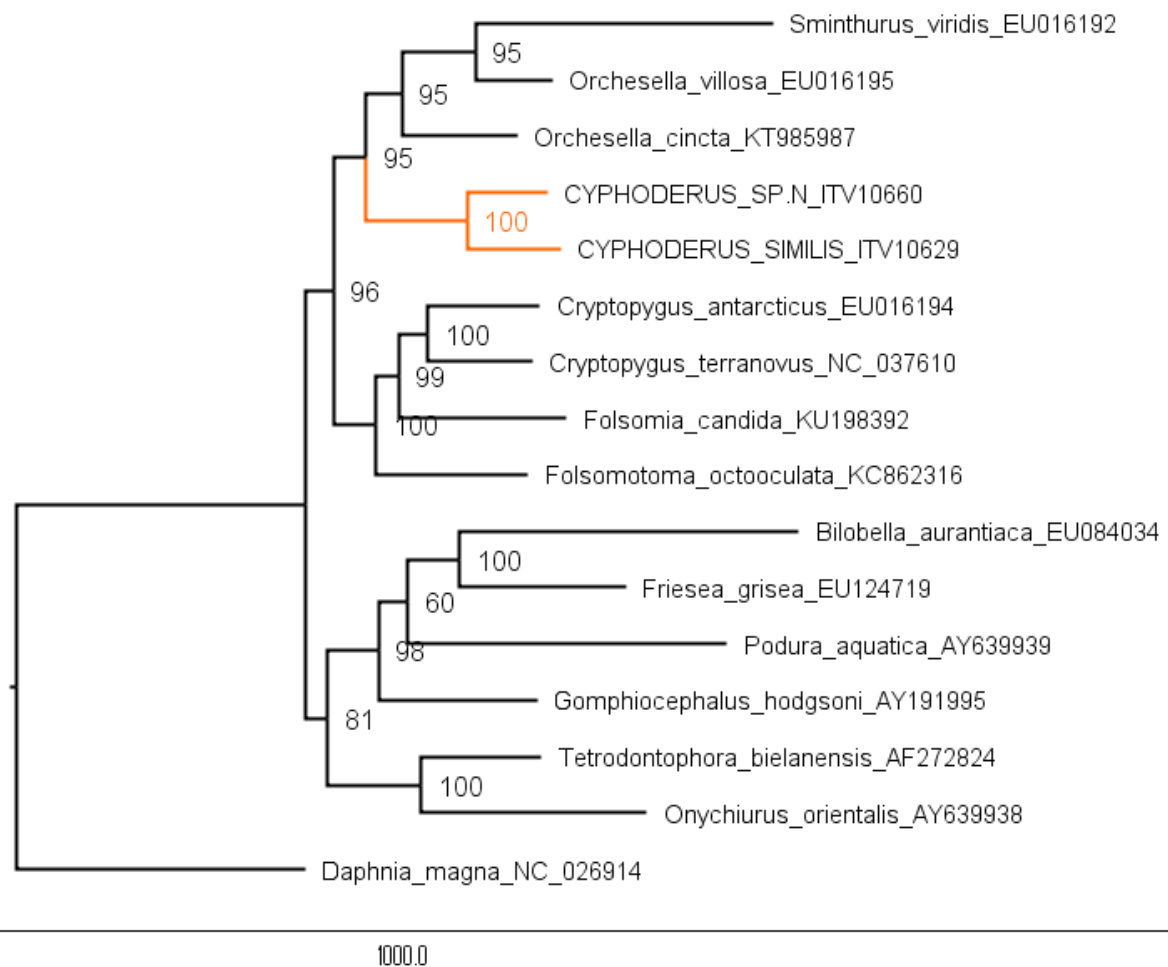


Figura 25. Filogenia molecular de Collembola baseada em genomas mitocondriais completos por análise de Máxima Verossimilhança e *bootstrap* de 1000. Com destaque para o monofiletismo das espécies *Cyphoderus* (em laranja). Importante: o ramo “CYPHODERUS_SP.N” refere-se à *Cyphoderus innominatus*.

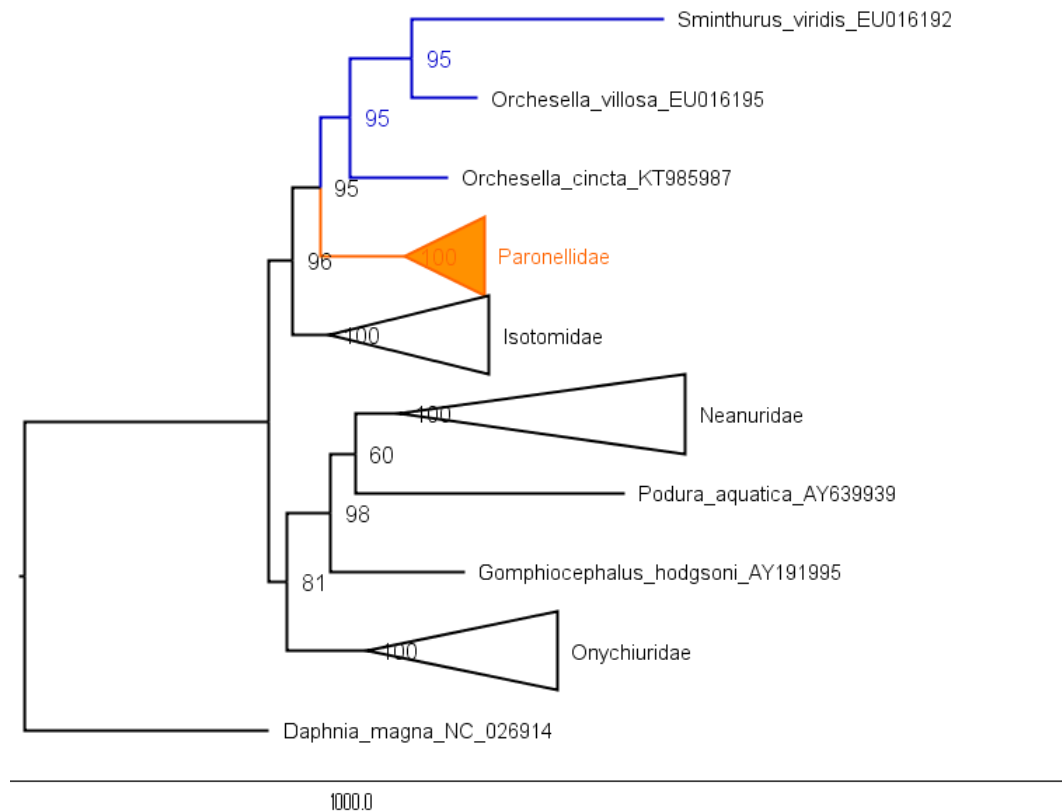


Figura 26. Filogenia molecular de Collembola baseada em genomas mitocondriais representando as famílias monofiléticas (Onychiuridae, Neanuridae, Isotomidae e Paronellidae) com valores de *bootstrap* de 100. Com destaque para o monofiletismo da família Paronellidae (em laranja) e o parafiletismo de três táxons (em azul).

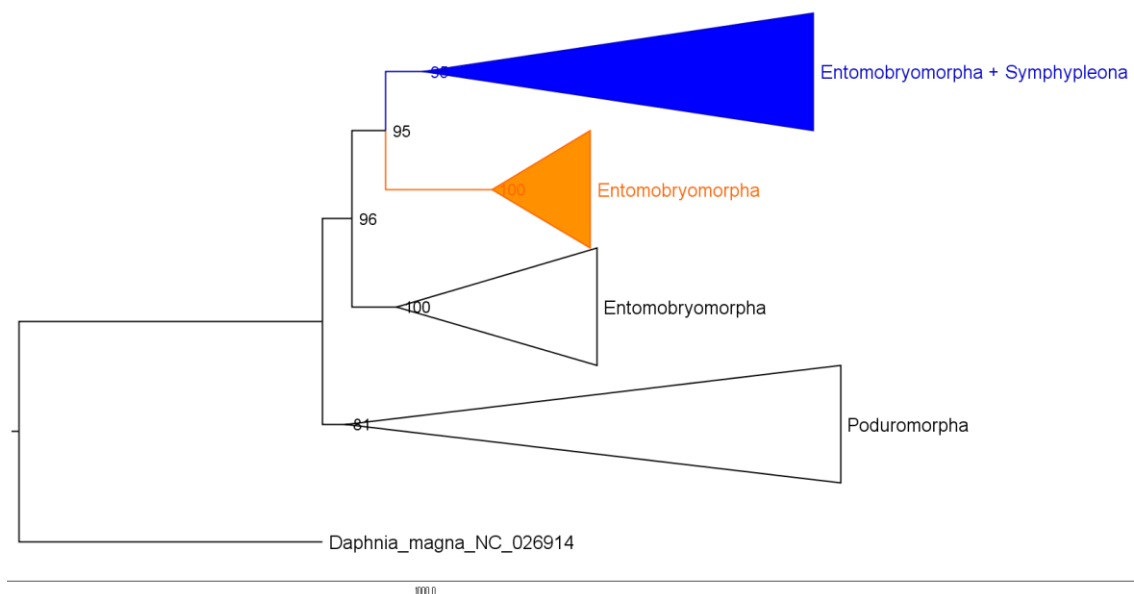


Figura 27. Filogenia molecular de Collembola baseada em genomas mitocondriais representando o monofiletismo da Ordem Poduromorpha (*bootstrap* de 81) e o Polifiletismo recuperado para as Ordens Entomobryomorpha (em laranja com *bootstrap* de 100) e Symphypleona (em azul com *bootstrap* de 95).

Semelhante às análises filogenéticas moleculares de CAMERON e colaboradores (2004) a análise realizada também apoia a Classe Collembola como monofilética, evidenciando o gênero *Daphnia* como bom grupo externo para a reconstrução filogenética do grupo. Além de corroborar à árvore filogenética de CARAPELLI e colaboradores (2014), inferida através de sequências de nucleotídeos referente às posições I e II de 13 mitogenomas de Collembola. Porém, a atual amostragem molecular de colêmbolos não representa de forma eficaz a Classe Collembola como um todo, com baixa cobertura, mesmo representando importantes famílias (Entomobryidae, Sminthuridae, Hypogastruridae, Isotomidae, por exemplo), há ausência de dados para muitas outras famílias, inclusive ainda não encontra-se dados moleculares mitogenômicos para a Ordem Neelipleona (CARAPELLI *et al.*, 2014).

Entretanto estudos relatam que por mais que o DNA esteja mais próximo de uma realidade evolutiva, no sentido de estar diretamente associado a mutações e seleção natural, ele ainda vem sendo altamente ruidoso no que se refere ao alto grau de saturação de substituições em seus códons, mais precisamente de terceiro códon. Assim, os métodos básicos de alinhamento múltiplo vêm se tornando mais necessários nesse processo de uso do genoma mitocondrial na da história filogenética (CAMERON *et al.*, 2004). Inclusive o uso do genoma completo para análises filogenéticas também se torna alvo de debates, NARDI e colaboradores (2003a,b) propõe o uso apenas do COI, COII, COIII e COB como suficiente e confiável.

Apesar da crescente disponibilidade de genomas mitocondriais de Collembola os dados existentes até então são insuficientes para resolver problemas como: determinar o grupo-irmão de Collembola, ou resolver relações internas da Classe. Além destes conjuntos de dados moleculares ainda se parecer vulnerável à seleção de táxons, de grupos externos, e até mesmo de genes, sendo assim, o limite e a aplicabilidade desses dados em questões filogenéticas ainda está em desenvolvimento (CAMERON *et al.*, 2004). No entanto, acreditamos que nossos dados representam o primeiro passo para uma futura reconstrução filogenética de *Cyphoderus*.

5. CONCLUSÃO

Ao final desta dissertação elencam-se os seguintes objetivos alcançados:

- Os primeiros *Draft Genomes* de colêmbolos brasileiros, sendo eles referentes ao sequenciamento e montagem *de novo* com anotações *ab initio* das espécies *C. similis* e *C. innominatus* representando os primeiros genomas para gênero *Cyphoderus*;
- A descrição dos primeiros genomas mitocondriais de *Cyphoderus* sequenciados, apresentando 100% de cobertura, apresentando seus conteúdos gênicos completos, sendo um total de 14.988 bp com 37 produtos gênicos e 14.811 bp com 36 produtos gênicos para as espécies *C. similis* e *C. innominatus*, respectivamente;
- Construção de uma filogenia molecular de Collembola utilizando pela primeira vez genomas mitocondriais sequenciados de *Cyphoderus* e um conjunto de 15 genomas mitocondriais completos de Collembola disponibilizados em banco de dados genômicos;
- Primeiras montagens de dois genomas nucleares de *Cyphoderus* com comprimento aproximado de 358 Mb com N50 máximo de 3,3 Kb para *C. similis* e comprimento aproximado de 370 Mb com N50 máximo de 3,3 Kb para *C. innominatus*;
- Primeira anotação genômica nuclear para *Cyphoderus*, o qual foi anotado um total de 278.865 genes putativos, com densidade de 3.8 genes por éxons sequenciados, referentes à espécie *C. innominatus*.

REFERÊNCIAS

- ABRANTES, E.A.; BELLINI, B.C.; BERNARDO, A.N.; FERNANDES, L.H.; MENDONÇA, M.C.; OLIVEIRA, E.P.; QUEIROZ, G.C.; SAUTTER, K.D.; SILVEIRA, T.C; ZEPPELINI, D. Errata Corrigenda and update for the “Synthesis of Brazilian Collembola: an update to the species list.” ABRANTES et al. (2010). **Zootaxa**, 2388, 1–22. **Zootaxa**, 3168, 1–21. 2012.
- ADAMS, E.C.G. & SALMON, J.T. The mouthparts and feeding methods of *Brachystomella parvula* (Collembola: Brachystomellidae). **Trans. R. ent. Soc. Lond.** 124 (3), 269-286 pp. 1972.
- AÏSSANI, B. & BERNARDI, G. CpG islands, genes and isochores in the genomes of vertebrates. **Gene**, 106 (2): 185-195, 1991.
- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. Biologia Molecular da Célula. **Artmed**, 5ªed. Porto Alegre, 2010.
- ALLENDORF, F.W.; HOHENLOHE, P.A.; LUIKART, G. Genomics and the future of conservation genetics. **Nature Reviews Genetics** 11:697–709. 2010.
- AL-SAFADI, M.M. Observation on the behaviour of *Cryptopygus thermophilus* Axelson (Collembola: Isotomidae). **Rev. Écol. Biol. Sol.** 25 (3), p.333-34. 1988.
- AMARANTO, I.; SALIMON, C.I.; ZEPPELINI, D. Relação entre massa corporal e taxa respiratória em Arthropoda de Floresta Atlântica (Paraíba, Brasil). **Pesquisa e Ensino em Ciências Exatas e da Natureza**, 2(2): 119–129. 2018.
- ANDRÉ, H.M.; NOTI, M.I.; LEBRUN, P. The soil fauna: the other last biotic frontier. **Biodiversity and Conservation**, v.3, p.45-56, 1994.
- ANDREWS, S. FastQC: A Quality Control Tool for High Throughput Sequence Data; 1. Available online <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>> 2016.
- ANSLAN, S. Molecular identification of Collembola and their fungal associates (thesis). **University of Tartu Press**, p7-123, 2017.
- ANSLAN, S. & TEDERSOO, L. Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. **European Journal of Soil Biology** 69: 1e7. 2015.
- AOYAMA, H.; SAITOH, S.; FUJII, S.; NAGAHAMA, H.; SHINZATO, N.; KANEKO, N.; NAKAMORI, T. A rapid method of non-destructive DNA extraction from individual springtails (Collembola). **Appl Entomol Zool**, 50:419–425. 2015.
- ARLÉ, R. & MENDONÇA, C. Estudo preliminar das espécies de *Dicranocentrus* Schött, 1893, ocorrentes no Parque Nacional da Tijuca, Rio de Janeiro. **Revista Brasileira de Biologia** 42: 41-49. 1982.
- AUBRET, F. & SHINE, R. Genetic assimilation and the postcolonization erosion of phenotypic plasticity in island tiger snakes. **Current Biology** 19: 1932–1936. 2009.
- BAKONYI, G.; POSTA, K.; KISS, I.; FABIAN, M.; NAGY, P.; NOSEK, J.N. Density-dependent regulation of arbuscular mycorrhiza by Collembola. **Soil Biol. Biochem.** 34, 661–664. 2002.

- BAKR, R.F.; FADL, H.H.; BADAWEY, R.M.; SHARAF, M.R. Myrmecophile insects associated with some ant species (Hymenoptera- Formicidae) in Egypt. Proc. 2nd **Intl. Conf. Entomol. Soc.** Egypt 1: 207-235. 2007.
- BARDGETT, R.D. & GRIFFITHS, B.S. Ecology and biology of soil protozoa, nematodes and microarthropods. In **Modern Soil Microbiology**. van Elsas, J.D., Trevors, J.T., and Wellington, E.M.H. (eds). New York: Marcel Dekker Inc., pp. 129–163, 1997.
- BELLINGER, P.F.; CHRISTIANSEN, K.A.; JANSSENS, F. Checklist of the Collembola of the World. <<http://www.collembola.org>> 1996-2008.
- BELLINGER, P.F.; CHRISTIANSEN, K.A.; JANSSENS, F. Checklist of the Collembola of the World, <<http://www.collembola.org>> 1996-2019.
- BELLINI, B. C. Colêmbolos: uma riqueza microscópica no Semiárido. 1.ed.-São Paulo: **Métis Produção Editorial**, 43-55 p. 2016.
- BENNETT, M.D. & LEITCH, I.J. Genome size evolution in plants. In Gregory, T.R. (ed.), The Evolution of the Genome. **Elsevier**, San Diego, CA, pp. 89–162. 2005.
- BERNT, M.; DONATH, A.; JÜHLING, F.; EXTERNBRINK, F.; FLORENTZ, C.; FRITZSCH, G.; PÜTZ, J.; MIDDENDORF, M.; STADLER, P.F. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. **Molecular Phylogenetics and Evolution**, 69(2):313-9. 2013.
- BILDE, T.; AXELSEN, J.A.; TOFT, S. The value of Collembola from agricultural soils as food for a generalist predator. **J. Appl. Ecol.** 37, 672–683. 2000.
- BOETZER, M., HENKE, C. V., JANSEN, H. J., BUTLER, D., AND PIROVANO, W. Scaffolding pre-assembled contigs using SSPACE. **Bioinformatics**. 27, 578–579. 2010.
- BOORE, J.L. Animal mitochondrial genomes. **Nucleic Acids Res**, 27(8):1767-1780. 1999.
- BORKOTT, H. & INSAM, H. Symbiosis with bacteria enhances the use of chitin by the springtail, *Folsomia candida* (Collembola). **Biol Fertil Soils** 9: 126–129, 1990.
- BÖRNER, C. Neue Cyphoderinen. **Zoologischer Anzeiger**, 41, 274–284. 1913.
- BOWE, A.; ONODERA, T.; SADAKANE, K. & SHIBUYA, T. Succinct de Bruijn Graphs, in Algorithms in Bioinformatics, B. Raphael and J. Tang, Editors, **Springer Berlin Heidelberg**. p. 225-235. 2012.
- BRADSHAW, A.D. Evolutionary significance of phenotypic plasticity in plants. **Advances in Genetics** 13: 115–155. 1965.
- BRAENDLE, C. & FLATT, T. A role for genetic accommodation in evolution? **BioEssays** 28:868–873. 2006.
- BRADNAM, K., FASS, J.; ALEXANDROV, A.; BARANAY, P.; BECHNER, M.; BIROL, I.; BOISVERT, S.; CHAPMAN, J.A.; CHAPUIS, G.; CHIKHI, R.; CHITSAZ, H.; CHOU, W.-C.; CORBEIL, J.; DEL FABBRO, C.; DOCKING, T.R.; DURBIN, R.; EARL, D.; EMRICH, S.; FEDOTOV, P.; FONSECA, N.A.; GANAPATHY, G.; GIBBS, R.A.; GNERRE, S.; GODZARIDIS, É.; GOLDSTEIN, S.; HAIMEL, M.; HALL, G.; HAUSSLER, D.; HIATT, J.B.; HO, I.Y.; HOWARD, J.; HUNT, M.; JACKMAN, S.D.; JAFFE, D.B.; JARVIS, E.D.; JIANG, H.; KAZAKOV, S.; KERSEY, P.J.; KITZMAN, J.O.; KNIGHT, J.R.; KOREN, S.; LAM, T.-W.; LAVENIER, D.; LAVIOLETTE, F.; LI, Y.; LI, Z.; LIU, B.; LIU, Y.; LUO, R.; MACCALLUM, I.; MACMANES, M.D.; MAILLET, N.; MELNIKOV, S.; NAQUIN, D.;

NING, Z.; OTTO, T.D.; PATEN, B.; PAULO, O.S.; PHILLIPPY, A.M.; PINA-MARTINS, F.; PLACE, M.; PRZYBYLSKI, D.; QIN, X.; QU, C.; RIBEIRO, F.J.; RICHARDS, S.; ROKHSAR, D.S.; RUBY, J.G.; SCALABRIN, S.; SCHATZ, M.C.; SCHWARTZ, D.C.; SERGUSHICHEV, A.; SHARPE, T.; SHAW, T.I.; SHENDURE, J.; SHI, Y.; SIMPSON, J.T.; SONG, H.; TSAREV, F.; VEZZI, F.; VICEDOMINI, R.; VIEIRA, B.M.; WANG, J.; WORLEY, K.C.; YIN, S.; YIU, S.-M.; YUAN, J.; ZHANG, G.; ZHANG, H.; ZHOU, S.; KORE, I.F. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. **GigaScience** 2:10. 2013.

BRITO, R.A.; LIMA, E.C.A.; FERREIRA, A.S.; OLIVEIRA, J.V.L.C.; AMORIM, T.P.L., MACHADO, T.G.; LOPES, B.C.H.; COSTA, I.A; ALVES, J.L.S.; BRITO, N.P.; ZEPPELINI, D. A Catalogue of Species of Collembola (Hexapoda: Ellipura) Deposited in Coleção de Referência de Fauna de Solo of the Universidade Estadual da Paraíba, Brazil. **Florida Entomologist**. Volume 100, No. 1, 2017.

BRUSCA, R.C. & BRUSCA, G. J. Invertebrates. 2nd edition. **Sinauer Associations, Sunderland**, Mass, 936 pp, 2003.

BURKHARDT, U. & FILSER, J. Molecular evidence for a fourth species within *the Isotoma viridis* group (Insecta, Collembola). **Zoologica Scripta** 34: 177–185. 2005.

CAMASSA, M.M. Food resources. In: Gunn J. (Ed.), Encyclopaedia of caves and karst science. **Fitzroy Dearborn**, New York, p. 755-761. 2004.

CAMERON, E.K.; BAYNE, E.M.; COLTMAN, D.W. Genetic structure of invasive earthworms *Dendrobaena octaedra* in the boreal forest of Alberta: insights into introduction mechanisms. **Molecular Ecology**, v.17, p.1189-1197, 2008.

CAMERON, S.L.; MILLER, K.B.; D’HAESE, C.A.; WHITING, M.F.; BARKER, S.C. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea *sensu lato* (Arthropoda). **Cladistics**, 20, 534–557, 2004.

CARAPELLI, A.; CONVEY, P.; NARDI, F.; FRATI, F. The mitochondrial genome of the antarctic springtail *Folsomotoma octooculata* (Hexapoda; Collembola), and an update on the phylogeny of collembolan lineages based on mitogenomic data. **Entomologia**; 2:190. 2014.

CARAPELLI, A.; FANCIULLI, P.P.; FRATI, F.; DALLAI, R. The use of genetic markers for the diagnosis of sibling species in the genus *Isotomurus* (Insecta, Collembola). **Bolletino di Zoologia**, v.62, p.71-76, 1995a.

CARAPELLI, A.; FRATI, F.; FANCIULLI, P.P.; DALLAI, R. Genetic differentiation of six sympatric species of *Isotomurus* (Collembola, Isotomidae) – is there any difference in their microhabitat preference? **European Journal of Soil Biology**, 31, 87–99. 1995b.

CARAPELLI, A.; LEO, C.; FRATI, F. High levels of genetic structuring in the Antarctic springtail *Cryptopygus terranovus*. **Antarctic Science**, 29, 311–323. 2017.

CARNINCI, P.; KASUKAWA, T.; KATAYAMA, S.; GOUGH, J.; FRITH, M.C.; MAEDA, N.; OYAMA, R.; RAVASI, T.; LENHARD, B.; WELLS, C.; KODZIUS, R.; SHIMOKAWA, K.; BAJIC, V.B.; BRENNER, S.E.; BATALOV, S.; FORREST, A.R.; ZAVOLAN, M.; DAVIS, M.J.; WILMING, L.G.; AIDINIS, V.; ALLEN, J.E.; AMBESI-IMPIOMBATO, A.; APWEILER, R.; ATURALIYA, R.N.; BAILEY, T.L.; BANSAL, M.; BAXTER, L.; BEISEL, K.W.; BERSANO, T.; BONO, H.; CHALK, A.M.; CHIU, K.P.; CHOUDHARY, V.; CHRISTOFFELS, A.; CLUTTERBUCK, D.R.; CROWE, M.L.; DALLA, E.; DALRYMPLE, B.P.; DE BONO, B.; DELLA GATTA, G.; DI BERNARDO, D.; DOWN, T.; ENGSTROM,

P.; FAGIOLINI, M.; FAULKNER, G.; FLETCHER, C.F.; FUKUSHIMA, T.; FURUNO, M.; FUTAKI, S.; GARIBOLDI, M.; GEORGII-HEMMING, P.; GINGERAS, T.R.; GOJOBORI, T.; GREEN, R.E.; GUSTINCICH, S.; HARBERS, M.; HAYASHI, Y.; HENSCH, T.K.; HIROKAWA, N.; HILL, D.; HUMINIECKI, L.; IACONO, M.; IKEO, K.; IWAMA, A.; ISHIKAWA, T.; JAKT, M.; KANAPIN, A.; KATOH, M.; KAWASAWA, Y.; KELSO, J.; KITAMURA, H.; KITANO, H.; KOLLIAS, G.; KRISHNAN, S.P.; KRUGER, A.; KUMMERFELD, S.K.; KUROCHKIN, I.V.; LAREAU, L.F.; LAZAREVIC, D.; LIPOVICH, L.; LIU, J.; LIUNI, S.; MCWILLIAM, S.; MADAN BABU, M.; MADERA, M.; MARCHIONNI, L.; MATSUDA, H.; MATSUZAWA, S.; MIKI, H.; MIGNONE, F.; MIYAKE, S.; MORRIS, K.; MOTTAGUI-TABAR, S.; MULDER, N.; NAKANO, N.; NAKAUCHI, H.; NG, P.; NILSSON, R.; NISHIGUCHI, S.; NISHIKAWA, S.; NORI, F.; OHARA, O.; OKAZAKI, Y.; ORLANDO, V.; PANG, K.C.; PAVAN, W.J.; PAVESI, G.; PESOLE, G.; PETROVSKY, N.; PIAZZA, S.; REED, J.; REID, J.F.; RING, B.Z.; RINGWALD, M.; ROST, B.; RUAN, Y.; SALZBERG, S.L.; SANDELIN, A.; SCHNEIDER, C.; SCHÖNBACH, C.; SEKIGUCHI, K.; SEMPLÉ, C.A.; SENO, S.; SESSA, L.; SHENG, Y.; SHIBATA, Y.; SHIMADA, H.; SHIMADA, K.; SILVA, D.; SINCLAIR, B.; SPERLING, S.; STUPKA, E.; SUGIURA, K.; SULTANA, R.; TAKENAKA, Y.; TAKI, K.; TAMMOJA, K.; TAN, S.L.; TANG, S.; TAYLOR, M.S.; TEGNER, J.; TEICHMANN, S.A.; UEDA, H.R.; VAN NIMWEGEN, E.; VERARDO, R.; WEI, C.L.; YAGI, K.; YAMANISHI, H.; ZABAROVSKY, E.; ZHU, S.; ZIMMER, A.; HIDE, W.; BULT, C.; GRIMMOND, S.M.; TEASDALE, R.D.; LIU, E.T.; BRUSIC, V.; QUACKENBUSH, J.; WAHLESTEDT, C.; MATTICK, J.S.; HUME, D.A.; KAI, C.; SASAKI, D.; TOMARU, Y.; FUKUDA, S.; KANAMORI-KATAYAMA, M.; SUZUKI, M.; AOKI, J.; ARAKAWA, T.; IIDA, J.; IMAMURA, K.; ITOH, M.; KATO, T.; KAWAJI, H.; KAWAGASHIRA, N.; KAWASHIMA, T.; KOJIMA, M.; KONDO, S.; KONNO, H.; NAKANO, K.; NINOMIYA, N.; NISHIO, T.; OKADA, M.; PLESSY, C.; SHIBATA, K.; SHIRAKI, T.; SUZUKI, S.; TAGAMI, M.; WAKI, K.; WATAHIKI, A.; OKAMURA-OHO, Y.; SUZUKI, H.; KAWAI, J.; HAYASHIZAKI, Y.; FANTOM CONSORTIUM; RIKEN GENOME EXPLORATION RESEARCH GROUP AND GENOME SCIENCE GROUP (GENOME NETWORK PROJECT CORE GROUP). The transcriptional landscape of the mammalian genome. **Science** 309:1559–1563. 2005.

CARSTENS, B.C.; PELLETIER, T.A.; REID, N.M.; SALTER, J.D. How to fail at species delimitation. **Molecular Ecology**, 22, 4369–4383. 2013.

CASSAGNE, N.; BAL-SERIN, M. C.; GERS, C.; GAUQUELIN, T. Changes in humus properties and collembolan communities following the replanting of beech forests with spruce. **Pedobiologia** 48: 267–276. 2004.

CASSAGNE, N.; GERS, C.; GAUQUELIN, T. Relationships between Collembola, soil chemistry and humus types in forest stands. **Biology and Fertility of Soils** 37: 355–361, 2003.

CASSAGNAU, P. Collembolles d'Amérique du Sud, II. Orchesellini, Paronellinae, Cyphoderinae. **Biologie de la Amérique Australe**, 2, 127–148. 1963.

CASSAGNAU, P. Les collembolles Neaurinae de l'Himalaya 2: Paranurini et Paleonurini paucitubercules. **Travaux du Laboratoire d'Ecobiologie des Arthropodes Edaphiques Toulouse** 6, 1–20. 1991.

CASTAÑO-MENESES, G.; PALACIOS-VARGAS, J.G.; CUTZ-POOL, L.Q. Feeding habits of Collembola and their ecological niche. **Anales del Instituto de Biología**, Universidad Nacional Autónoma de México, Serie Zoología 75: 135–142, 2004.

- CASTAÑO-MENESES, G.; PALACIOS-VARGAS, J.G.; DELABIE, J.H.C.; SANTOS, R.d.J.; MARIANO, C.S.F. Springtails (Collembola) From Nests Of Ponerinae (Hymenoptera: Formicidae) Ants In Brazilian Cacao Plantations. **Florida Entomologist** 97(4), 2014.
- CHANG, C.H.; ROUGERIE, R.; CHEN, J.H. Identifying earthworms through DNA barcodes: pitfalls and promise. **Pedobiologia**, v.52, p.171-180, 2009.
- CHEN, B.; SNIDER, R.J.; SNIDER, R.M. Food preference and effects of food type on the life history of some soil Collembola. **Pedobiologia** 39: 496–505. 1995.
- CHEN, C-G.; CHEN, T.; HUA, B-Z.; WAN, T-R. Structure and functions of the ventral tube of the clover springtail *Sminthurus viridis* (Collembola: Sminthuridae). **Scientific Reports**. 9:897, 1-6, 2019.
- CHEN, W-J.; KOCH, M.; MALLATT, J.M.; LUAN, Y-X. Comparative Analysis of Mitochondrial Genomes in Diplura (Hexapoda, Arthropoda): Taxon Sampling Is Crucial for Phylogenetic Inferences. **Genome Biol. Evol.** 6(1):105–120. 2014.
- CHENG, J.; KAPRANOV, P.; DRENKOW, J.; DIKE, S.; BRUBAKER, S.; PATEL, S.; LONG, J.; STERN, D.; TAMMANA, H.; HELT, G.; SEMENTCHENKO, V.; PICCOLBONI, A.; BEKIRANOV, S.; BAILEY, D.K.; GANESH, M.; GHOSH, S.; BELL, I.; GERHARD, D.S.; GINGERAS, T.R. .Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. **Science**, 308:1149–1154. 2005.
- CHERNOVA, N.M.; POTAPOV, M.B.; SAVENKOVA, Y.Y.; BOKOVA, A.Y. Ecological significance of parthenogenesis in Collembola. **Entomol. Rev.** 90 (1), 23–38. 2010.
- CHOI, W.I. & MOORHEAD, D.L. A modeling study of soil temperature and moisture effects on population dynamics of *Paronychiurus kimi* (Collembola: Onychiuridae). **Biology and Fertility of Soils** 43: 69–75. 2006.
- CHRISTIANSEN, K. Massachusetts records of *Cyphoderus assimilis* Börner (Collembola). **Psyche** 57: 94. 1950.
- CHRISTIANSEN, K. The collembolan of Lebanon and western Syria part II families Cyphoderidae and Oncopoduridae. **Psyche** 64: 119–133. 1957.
- CHRISTIANSEN, K. Bionomics of Collembola. **Ann Ver Entomol** 9:147–178. 1964.
- CHRISTIANSEN, K. A. Springtails. **Kansas School Naturalist** 39: 3-16. 1992.
- CHRISTIANSEN, K. & BELLINGER, P. The Collembola of North America. North of Rio Grande. **Grinnell College**, Grinnell, Iowa. 1322 p. 1980.
- CICCONARDI, F.; FANCIULLI, P.P.; EMERSON, B.C. Collembola, the biological species concept and the underestimation of global species richness. **Molecular Ecology** 22: 5382–5396, 2013.
- CLAMP, M.; FRY, B.; KAMAL, M.; XIE, X.; CUFF, J.; LIN, M.F.; KELLIS, M.; LINDBLAD-TOH, K.; LANDER, E.S. Distinguishing protein-coding and noncoding genes in the human genome. **PNAS**, vol. 104, no. 49: 19428–19433. 2007.
- CRANDALL, K.A.; BININDA-EMONDS, O.R.P.; MACE, G.M.; WAYNE, R.K. Considering evolutionary processes in conservation biology. **Trends in Ecology & Evolution** 15:290–295. 2000.

- CULIK, M.P.; MARTINS, D.S.; VENTURA, J.A. Collembola (Arthropoda: Hexapoda) communities in the soil of papaya orchards managed with conventional and integrated production in Espírito Santo, Brazil. **Biota Neotropica**, 6 (n2). Available from: <<http://www.biotaneotropica.org.br/v6n3/en/abstract?short-communication+bn02006032006>> 2006.
- CZARNETZKI, A. B. & TEBBE, C. C. Detection and phylogenetic analysis of *Wolbachia* in Collembola. **Environmental Microbiology**, 6 (1), 35–44, 2004.
- DANCE, A. Soil ecology: what lies beneath. **Nature**, v.455, p.724-725, 2008.
- DECAËNS, T.; JIMÉNEZ, J.J.; GIOIA, C.; MEASEY, G.J.; LAVELLE, P. The values of soil animals for conservation biology. **European Journal of Soil Biology**, v.42, p.23-38, 2006.
- DECAËNS, T.; LAVELLE, P.; JIMÉNEZ, J.J. Priorities for conservation of soil animals. **CAB Reviews**, v.3, p.1-18, 2008.
- DE DEYN, G.B.; RAAIJMAKERS, C.E.; ZOOMER, H.R.; BERG, M.P.; DE RUITER, P.C.; VERHOEF, H.A.; BEZEMER, T.M.; VAN DER PUTTEN, W.H. Soil invertebrate fauna enhances grassland succession and diversity, **Nature** 422: 711–713. 2003.
- DEHARVENG, L. Recent advances in Collembola systematics. **Pedobiologia** 48: 415–433, 2004
- DEHARVENG, L.; D'HAESE, C.A.; BEDOS, A. Global diversity of springtails (Collembola; Hexapoda) in freshwater. **Hydrobiologia** 595: 329–338, 2008.
- DELAMARE-DEBOUTTEVILLE, C. Recherches sur les Collembolles termitophiles et myrmécophiles (Ecologie, Ethologie, Systématique). **Archives de Zoologie Expérimentale et Générale** 85: 261–425. 1948.
- DELSUC, F.; PHILLIPS, M.J.; PENNY, D. Comment on "hexapod origins: monophyletic or paraphyletic?". **Science**, 301:1482d, 2003.
- DENOEUDE, F.; KAPRANOV, P.; UCLA, C.; FRANKISH, A.; CASTELO, R.; DRENKOW, J.; LAGARDE, J.; ALIOTO, T.; MANZANO, C.; CHRAST, J.; DIKE, S.; WYSS, C.; HENRICHSEN, C.N.; HOLROYD, N.; DICKSON, M.C.; TAYLOR, R.; HANCE, Z.; FOISSAC, S.; MYERS, R.M.; ROGERS, J.; HUBBARD, T.; HARROW, J.; GUIGÓ, R.; GINGERAS, T.R.; ANTONARAKIS, S.E.; REYMOND, A. Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. **Genome Res**, 17:746-759. 2007.
- DE QUEIROZ, K. Species concepts and species delimitation. **Systematic Biology**, 56, 879–886. 2007.
- DESALLE, R.; EGAN, M.G.; SIDDALL, M. The unholy trinity: taxonomy, species delimitation and DNA barcoding. **Philos. Trans. R. Soc. B** 360, 1905–1916. 2005.
- DOLEZEL, J.; BARTOS, J.; VOGLMAYR, H.; GREILHUBER, J. Nuclear DNA content and genome size of trout and human. **Cytometry**, 51A, 127–128. 2003.
- DOMES, K.; NORTON, R.A.; MARAUN, M.; SCHEU, S. Reevolution of sexuality breaks Dollo's law, **Proc. Natl. Acad. Sci. U.S.A.** 104: 7139–7144. 2007.
- EHRENREICH, I.M. & PFENNIG, D.W. Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. **Annals of Botany**, 117(5):769-79. 2016.

- EKBLOM, R. & WOLF, J.B.W. A field guide to whole-genome sequencing, assembly and annotation. **Evolutionary Applications**, 7(9):1026-42. 2014.
- ELLEGREN, H. Genome sequencing and population genomics in nonmodel organisms. **Trends in Ecology & Evolution** 29:51–63. 2014.
- ENCODE PROJECT CONSORTIUM. An integrated encyclopedia of DNA elements in the human genome. **Nature**, 447:799–816. 2007.
- ENDLWEBER, K.; RUESS, L.; SCHEU, S. Collembola switch diet in presence of plant roots thereby functioning as herbivores. **Soil Biology & Biochemistry**, 41: 1151- 1154. 2009.
- FADDEEVA-VAKRHUSHEVA, A. Hacking the genomes of soil arthropods (thesis). Department of Ecological Science. **VU - University Amsterdam**, p. 9-22. URL <<https://research.vu.nl/en/publications/hacking-the-genomes-of-soil-arthropods> > 2017.
- FADDEEVA-VAKHRUSHEVA, A.; KRAAIJEVELD, K.; DERKS, M.F.L.; ANVAR, S.Y.; AGAMENNONE, V.; SURING, W.; KAMPFRAATH, A.A.; ELLERS, J.; NGOC, G.L.; VAN GESTEL, C.A.M.; MARIËN J.; SMIT S.; VAN STRAALLEN, N.M.; ROELOFS, D. Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*. **BMC Genomics**, 18:493, 2017.
- FARAJI, F. & BAKKER, F. A modified method for clearing, staining and mounting plant-inhabiting mites. **Eur. J. Entomol.** 105: 793–795, 2008.
- FELDERHOFF, K.L.; BERNARD, E.C.; MOULTON, J.K. Survey of *Pogonognathellus* Börner (Collembola: Tomoceridae) in the southern Appalachians based on morphological and molecular data. **Annals of the Entomological Society of America**, 103, 472–491. 2010.
- FILSER, J. The role of Collembola in carbon and nitrogen cycling in soil. In Proceedings of the Xth international Colloquium on Apterygota, Ceské Budejovice 2000: Apterygota at the Beginning of the Third Millennium. **Pedobiologia**, 46(3): 234–245. 2002.
- FITZPATRICK, B.M. Underappreciated consequences of phenotypic plasticity for ecological speciation. **International Journal of Ecology** 2012: 256017. 2012.
- FRAMPTON, G. K. Recovery responses of soil surface Collembola after spatial and temporal changes in long-term regimes of pesticide use. **Pedobiologia** 44: 489–501. 2000.
- FRASER, D.J. & BERNATCHEZ, L. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. **Molecular Ecology** 10:2741–2752. 2001.
- FRATI, F.; DELL'AMPIO, E.; CASASANTA, S.; CARAPELLI, A.; PAOLO FANCIULLI, P. Large amounts of genetic divergence among Italian species of the genus *Orchesella* (Insecta, Collembola) and the relationships of two new species. **Molecular Phylogenetics and Evolution**, v.17, p.456-461, 2000a.
- FRATI, F.; FANCIULLI, P.P.; CARAPELLI, A.; DELL'AMPIO, E.; NARDI, F.; SPINSANTI, G.; DALLAI, R. DNA Sequence Analysis to Study the Evolution of Antarctic Collembola. **Ital. J. Zool.** 1: 133-139, 2000b.
- FRATI, F.; FANCIULLI, P.P.; CARAPELLI, A.; DALLAI, R. The Collembola of northern Victoria Land (Antarctica): distribution and ecological remarks. **Pedobiologia**, 41: 50-55, 1997.
- FULTHORPE, R.R.; ROESCH, L.F.W.; RIVA, A.; TRIPLETT, E.W. Distantly sampled soils carry few species in common. **The ISME Journal**, v.2, p.901-910, 2008.

- GANGE, A. Arbuscular mycorrhizal fungi, Collembola and plant growth. **Trends Ecol. Evol.** 15, 369–372. 2000.
- GERLACH, J.; SAMWAYS, M.; PRYKE, J. Terrestrial invertebrates as bioindicators: an overview of available taxonomic groups. **Journal of Insect Conservation** 17, 831–850. 2013.
- GILBERT, S.F. & EPEL, D. Ecological developmental biology: integrating epigenetics, medicine, and evolution. **Sunderland, MA**: Sinauer. 2009.
- GILMORE, S. & GILMORE, K. Collembola predation on nematodes. *Search*: **Agriculture** 1: 1–12. 1970.
- GOLOSHCHAPOVA, N.P.; POTAPOV, M.B.; CHERNOVA, N.M. Sexual behavior in threespecies of Isotomidae (Collembola). **Zool. Zh.** 84 (1), 63–69. 2005.
- GOODWIN, S.; MCPHERSON, J.D.; MCCOMBIE, W.R. Coming of age: ten years of next generation sequencing technologies. **Nat Rev Genet** 17, 333–351, 2016.
- GREENSLADE, P. Collembola from Scotia Arc and Antarctic Peninsula including descriptions of two new species and notes on biogeography. **Polkie Pismo Entomol.**, 64: 305–319, 1995.
- GREENSLADE, P. & WHALLEY, P.E.S. The systematic position of *Rhyniella praecursor* Hirst & Maulik (Collembola). The earliest known hexapod. **International Seminar on Apterygota**, University of Siena, Italy. 1986.
- GREGORY, T.R. Genome size and developmental complexity. **Genetica** 115(1):131–146. 2002.
- GREGORY, T.R. Synergy between sequence and size in large-scale genomics. **Nature Rev. Genet.**, 6, 699–708. 2005a.
- GREGORY, T.R. Genome size evolution in animals. In Gregory, T.R. (ed.), *The Evolution of the Genome*. **Elsevier**, San Diego, CA, pp. 3–87. 2005b.
- HAILER, F.; KUTSCHERA, V.E.; HALLSTRÖM, B.M.; KLASSERT, D.; FAIN, S.R.; LEONARD, J.A.; ARNASON, U.; JANKE, A. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. **Science**, 336, 344–347. 2012.
- HAJIBABAEI, M.; JANZEN, D.H.; BURNS, J.M.; HALLWACHS, W.; HEBERT, P.D.N. DNA barcodes distinguish species of tropical Lepidoptera. **Proc. Natl. Acad. Sci. USA** 103, 968–971. 2007.
- HALLIGAN, D.L. & KEIGHTLEY, P.D. Spontaneous mutation accumulation studies in evolutionary genetics. **Annual Review of Ecology, Evolution, and Systematics** 40: 151–172. 2009.
- HANRAHAN, S.J. & JOHNSTON, J.S. New genome size estimates of 134 species of arthropods. **Chromosome Res**, 19:809–823. 2011.
- HARROW, J.; NAGY, A.; REYMOND, A.; ALIOTO, T.; PATTHY, L.; STYLIANOS & ANTONARAKIS; GUIGÓ, R. Identifying protein-coding genes in genomic sequences. **Genome Biology**, 10:201. 2009.
- HEBERT, P.D.N. & GREGORY, T.R. The Promise of DNA Barcoding for Taxonomy. **Syst. Biol.** 54(5), 852–859. 2005.

- HEBERT, P.D.N.; CYWINSKA, A.; BALL, S.L.; DEWAARD, J.R. Biological identifications through DNA barcodes. **Proceedings of the Royal Society of London. Series B. Biological Sciences** 270, 313–321. 2003.
- HEDRICK, P.W. Perspective: highly variable loci and their interpretation in evolution and conservation. **Evolution** 53:313–318. 1999.
- HEEMSBERGEN, D.A.; BERG, M.P.; LOREAU, M.; VAN HAL, J.R.; FABER, J.H.; VERHOEF, H.A. Biodiversity effects on soil processes explained by interspecific functional dissimilarity, **Science** 306: 1019–1020. 2004.
- HEMMER, W. Karyotype differentiation and chromosomal variability in springtails (Collembola, Insecta). **Biol Fert Soils**, 9:119-125. 1990.
- HERMANN, H.R.; BLUM, M.S.; HUNT, A.N. Myrmecophilous arthropods associated with imported Pre ant, *Solenopsis saevissima* (Hymenoptera: Formicidae). **LA Acad. Sci.** 33: 13-18. 1970.
- HIGUCHI, R.; BOWMAN, B.; FREIBERGER, M.; RYDER, O.A.; WILSON, A.C. DNA sequences from the quagga, an extinct member of the horse family. **Nature** (Lond.) 312: 282-284. 1984.
- HIRST, S. & MAULIK, S. On some Arthropod Remains from the Rhynie Chert (Old Red Sandstone). 63 (2): 69–71. 1926.
- HOGG, I.D. & HEBERT, P.D.N. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. **Can. J. Zool.** 82: 749–754. 2004.
- HÖLLDOBLER, B. & WILSON, E.O. The ants. **Harvard University Press**, Cambridge, 732 pp. 1990.
- HOLMSTRUP, M.; BAYLEY, M.; RAMLØV, H. Supercool or dehydrate? An experimental analysis of overwintering strategies in small permeable arctic invertebrates, **Proc. Natl. Acad. Sci. U.S.A.** 99: 5716–5720. 2002.
- HOPKIN, S.P. Biology of the springtails (Insecta: Collembola): **Oxford University Press**. 1997.
- HOSKINS, J.L.; JANION-SCHEEPERS, C.; CHOWN, S.L.; DUFFY, G.A. Growth and reproduction of laboratory-reared neanurid Collembola using a novel slime mould diet. **Scientific Reports** 5: 11957. 2015.
- HUNT, H.W.; COLEMAN, D.C.; INGHAM, R.E.; ELLIOT, E.T.; MOORE, J.C.; ROSE, S.L.; REID, C.P.P.; MORLEY, C.R. The detrital food web in a shortgrass prairie. **Biol Fert Soils** 3: 17–68. 1987.
- ILLUMINA SEQUENCING. NextSeq® 500 System Guide. <<https://www.well.ox.ac.uk/ogc/wp-content/uploads/2019/02/nextseq-500-system-guide-15046563-04.pdf>> 2016.
- ILLUMINA SEQUENCING. An introduction to Next-Generation Sequencing Technology. <www.illumina.com/technology/next-generation-sequencing.html> Pub. No. 770-2012-008-B. 2017.
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Finishing the euchromatic sequence of the human genome. **Nature** 431:931–945. 2004.

- JAGATAP, H.; MONSANTO, D.M.; VAN VUUREN, B.J.; JANION-SCHEEPERS, C.; SEKAR, S.; TESKE, P.R. & EMAMI-KHOYI, A. The complete mitogenome of the springtail *Tullbergia bisetosa*: a subterranean springtails from the sub-Antarctic region, **Mitochondrial DNA Part B**, 4:1, 1594-1596, 2019a.
- JAGATAP, H.; MONSANTO, D.M.; VAN VUUREN, B.J.; PARBHU, S.P.; DINOI, A.; JANION-SCHEEPERS, C.; SEKAR, S.; TESKE, P.R. & EMAMI-KHOYI, A. The complete mitogenome of *Isotomurus maculatus*: a widespread species that is invading the sub-Antarctic region, **Mitochondrial DNA Part B**, 4:1, 1706-1708, 2019b.
- JANTARIT, S., SATASOOK, C.; DEHARVENG, L. *Cyphoderus* (Cyphoderidae) as a major component of collembolan cave fauna in Thailand, with description of two new species. **ZooKeys** 368: 1–21. 2014.
- JOHNSON, D.L. & WELLINGTON, W.G. Predation of *Apochthonius minimus* (Pseudoscorpionida: Chthoniidae) on *Folsomia candida* (Collembola: Isotomidae) I. Predation rate and size-selection. **Res. Popul. Ecol.** 22(2): 339–352. 1980.
- JOHNSON, W.E. & KOEPFLI, K. The role of genomics in conservation and reproductive sciences. **Advances in Experimental Medicine and Biology**, vol. 753, pp. 71–96, 2014.
- KASSEN, R. & BATAILLON, T.M. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. **Nature Genetics** 38: 484–488. 2006.
- KATZ, A.D.; GIORDANO, R.; SOTO-ADAMES, F.N. Operational criteria for cryptic species delimitation when evidence is limited, as exemplified by North American *Entomobrya* (Collembola: Entomobryidae). **Zoological Journal of the Linnean Society** 173, 818–840. 2015.
- KAUFMANN, E.; MALSCH, A.; K. F.; ERLE, M.; MASCHWITZ, U. Compound nesting of *Strumigenys* sp. (Myrmicinae) and *Diacamma* sp. (Ponerinae), and other nesting symbioses of myrmicinae and ponerinae ants in Southeast Asia. **Insect. Soc.** 50: 88-97. 2003.
- KELLER, O.; KOLLMAR, M.; STANKE, M. & WAACK, S. A novel hybrid gene prediction method employing protein multiple sequence alignments. **Bioinformatics**, 27, 757–763. 2011.
- KHAN, S.; NABI, G.; ULLAH, M.W.; YOUSAF, M.; MANAN, S.; SIDDIQUE, R.; HOU, H. Overview on the Role of Advance Genomics in Conservation Biology of Endangered Species. **International Journal of Genomics**, 2016, Article ID 3460416, 8 p, 2016.
- KING, R.A.; TIBBLE, A.L.; SYMONDSON, W.O.C. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. **Molecular Ecology**, v.17, p.4684-4698, 2008.
- KISTNER, D. H. The social insects' bestiary, pp. 1-244. In Herman, H.R. (ed.) **The social Insects**, Vol. 3. Academic Press. 1982.
- KLIRONOMOS, J.N. & KENDRICK, B. Relationships among microarthropods, fungi, and their environment. **Plant Soil** 170: 183–197. 1995.
- KNOWLES, L.L. & CARSTENS, B.C. Delimiting species without monophyletic gene trees. **Systematic Biology**, 56, 887–895. 2007.
- KOBOLDT, D.C.; STEINBERG, K.M.; LARSON, D.E.; WILSON, R.K.; MARDIS, E.R. The next-generation sequencing revolution and its impact on genomics. **Cell**, 155, 27–38, 2013.

- KOIVISTO, R.K.K. & BRAIG, H.R. Microorganisms and parthenogenesis. **Biol J Linnean Soc** 79: 43–58, 2003.
- KOOISTRA, G. Some data concerning the presence and behavior of springtails (Collembola) on Grass and white clover. **Netherlands Journal of Agricultural Science** 70: 136–141, 1964.
- KOPECKY, O.; NOVAK, K.; VOJAR, J.; SUSTA, F. Food composition of alpine newt (*Ichthyosaura alpestris*) in the post-hibernation terrestrial life stage. **North-Western Journal of Zoology** 12: 299 – 303. 2016.
- KOREN, S.; SCHATZ, M.C.; WALENZ, B.P.; MARTIN, J.; HOWARD, J.T.; GANAPATHY, G.; WANG, Z.; RASKO, D.A.; MCCOMBIE, W.R.; JARVIS, E.D.; PHILLIPPY, A.M. Hybrid error correction and de novo assembly of single-molecule sequencing reads. **Nature Biotechnology**; 30:693–700. 2012.
- KRISTENSEN, N.P. Phylogeny of insect orders. **Annual Review of Entomology** 26: 135–157, 1981.
- KROGH, P.H. Toxicity testing with the collembolans *Folsomia fimetaria* and *Folsomia candida* and the results of a ringtest. **Environmental Project. Miljøprojekt**, 66.p. ISBN: 978-87-7052-882-5. 2009.
- KUUSK, A.K. & AGUSTÍ, N. Group-specific primers for DNA-based detection of springtails (Hexapoda: Collembola) within predator gut contents. **Molecular Ecology Resources**, 8, 678–681. 2008.
- LANDE, R. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. **Journal of Evolutionary Biology** 22: 1435–1446. 2009.
- LANGEL, Ü. Galanin and galanin antagonists (thesis). **University of Tartu Press**, 97 p, 1993.
- LAWNICZAK, M.K.N.; HOLLOWAY, A.K.; BEGUN, D.J.; JONES, C.D. Genomic analysis of the relationship between gene expression. variation and DNA polymorphism in *Drosophila simulans*. **Genome Biology**, 9:R125. 2008.
- LAWRENCE, K.L. & WISE, D.H. Spider predation on forest-floor Collembola and evidence for indirect effects on decomposition. **Pedobiologia**, 44(1): 33–39. 2000.
- LI, R.; FAN, W.; TIAN, G.; ZHU, H.; HE, L.; CAI, J.; HUANG, Q.; CAI, Q.; LI, BO.; BAI, Y.; ZHANG, Z.; ZHANG, Y.; WANG, W.; LI, J.; WEI, F.; LI, H.; JIAN, M.; LI, J.; ZHANG, Z.; NIELSEN, R.; LI, D.; GU, W.; YANG, Z.; XUAN, Z.; RYDER, O.A.; LEUNG, F.C-C.; ZHOU, Y.; CAO, J.; SUN, X.; FU, Y.; FANG, X.; GUO, X.; WANG, B.; HOU, R.; SHEN, F.; UM, B.; NI, P.; LIN, R.; QIAN, W.; WANG, G.; YU, C.; NIE, W.; WANG, J.; WU, Z.; LIANG, H.; MIN, J.; WU, Q.; CHENG, S.; RUAN, J.; WANG, M.; SHI, Z.; WEN, M.; LIU, B.; REN, X.; ZHENG, H.; DONG, D.; COOK, K.; SHAN, G.; ZHANG, H.; KOSIOL, C.; XIE, X.; LU, Z.; ZHENG, H.; LI, Y.; STEINER, C.C.; LAM, T.T-Y.; LIN, S.; ZHANG, Q.; LI, G.; TIAN, J.; GONG, T.; LIU, H.; ZHANG, D.; FANG, L.; YE, C.; ZHANG, J.; HU, W.; XU, A.; REN, Y.; ZHANG, G.; BRUFORD, M.W.; LI, Q.; MA, L.; GUO, Y.; AN, N.; HU, Y.; ZHENG, Y.; SHI, Y.; LI, Z.; LIU, Q.; CHEN, Y.; ZHAO, J.; QU, N.; ZHAO, S.; TIAN, F.; WANG, X.; WANG, H.; XU, L.; LIU, X.; VINAR, T.; WANG, Y.; LAM, T-W.; YIU, S-M.; LIU, S.; ZHANG, H.; LI, D.; HUANG, Y.; WANG, X.; YANG, G.; JIANG, Z.; WANG, J.; QIN, N.; LI, L.; LI, J.; BOLUND, L.; KRISTIANSEN, K.; WONG, G.K-S.; OLSON, M.; ZHANG, X.; LI, S.; YANG, H.; WANG, J.; WANG, J. The sequence and de novo assembly of the giant panda genome. **Nature** 463:311–317. 2010.

LI, D.; LUO, R.; LIU, C-M.; LEUNG, C-M.; TING, H-F.; SADAKANE, K.; YAMASHITA, H.; LAM, T-W. MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler driven by Advanced Methodologies and Community Practices, **Methods**, 2016.

LINDAHL, T. Instability and decay of the primary structure of DNA. **Nature** (Lond.) 362: 709-715. 1993.

LUO, R.; LIU, B.; XIE, Y.; LI, Z.; HUANG, W.; YUAN, J.; HE, G.; CHEN, Y.; PAN, Q.; LIU, Y.; TANG, J.; WU, G.; ZHANG, H.; SHI, Y.; LIU, Y.; YU, C.; WANG, B.; LU, Y.; HAN, C.; CHEUNG, D.W.; YIU, S-M.; PENG, S.; XIAOQIAN, Z.; LIU, G.; LIAO, X.; LI, Y.; YANG, H.; WANG, J.; LAM, T-W. & WANG, J. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. **GigaScience**, 1:18, 2012.

MATEOS, E.; ESCUER, P.; BUSMACHI, G.; RIUTORT, M.; ÁLVAREZ-PRESAS, M.; Untangling *Lepidocyrtus* (Collembola, Entomobryidae): new molecular data shed light on the relationships of the European groups. **Invertebrate Systematics**, 32, 639–651. 2018.

MCKUSICK, V.A. HUGO news. The human genome organisation: history, purposes, and membership. **Genomics**, vol. 5, no. 2, pp. 385–387, 1989.

MILLER, J.R.; KOREN, S. & SUTTON, G. Assembly algorithms for next-generation sequencing data. **Genomics**. v. 95 (6): 315-327. 2010.

MILLER, M.R.; BRUNELLI, J.P.; WHEELER, P.A.; LIU, S.; REXROAD, C.E.; PALT, Y.; DOE, C.Q.; THORGAARD, G.H. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. **Molecular Ecology**, vol. 21, no. 2, pp. 237–249, 2012.

MIRANDA-RANGEL, A. & PALACIOS-VARGAS J.G. Estudio Comparativo de las comunidades de colémbolos edáficos de bosque de Abies religiosa y cultivo de haba (*Vicia faba*). **Agrociencia, Serie Protección Vegetal**, 3: 1–14, 1992.

MITRA, S.K. Chaetotaxy, phylogeny and biogeography of Paronellinae (Collembola: Entomobryidae). **Records of the Zoological Survey of India, Occasional Papers** 154: 1-100. 1993.

MOCZEK, A.P.; SULTAN, S.E.; FOSTER, S.; LEDÓN-RETTIG, C.; DWORKIN, I.; NIJHOUT, H.F.; ABOUHEIF, E.; PFENNIG, D.W. The role of developmental plasticity in evolutionary innovation. **Proceedings of the Royal Society of London Series B: Biological Sciences** 278: 2705–2713. 2011.

MOSER, J.C. & BLOMQUIST, S.R. Phoretic Arthropods of the Red Imported Fire Ant in Central Louisiana. **Annals Of The Entomological Society Of America**, Vol. 104, no. 5, 2011.

NAKAMORI, T. A new species of *Ceratophysella* (Collembola: Hypogastruridae) from Japan, with notes on its DNA barcode and a key to Japanese species in the genus. **Zootaxa** 3641 (4): 371–378, 2013.

NAKAMORI, T.; FUJIMORI, A.; KINOSHITA, K.; BAN-NAI, T.; KUBOTA, Y.; YOSHIDA, S. Application of HiCEP to Screening of Radiation Stress-Responsive Genes in the Soil Microarthropod *Folsomia candida* (Collembola). **Environ. Sci. Technol**, 42, 6997–7002, 2008.

NAKAMORI, T.; FUJIMORI, A.; KINOSHITA, K.; BAN-NAI, T.; KUBOTA, Y.; YOSHIDA, S. mRNA expression of a cadmium-responsive gene is a sensitive biomarker of

- cadmium exposure in the soil collembolan *Folsomia candida*. **Environmental Pollution** 158: 1689–1695, 2010.
- NARDI, F.; SPINSANTI, G.; BOORE, J.L.; CARAPELLI, A.; DALLAI, R.; FRATI, F. Hexapod origins: monophyletic or paraphyletic? **Science**, 299, 1887–1889, 2003a.
- NARDI, F., SPINSANTI, G., BOORE, J.L., CARAPELLI, A., DALLAI, R., FRATI, F. Response to comment on “Hexapod Origins: Monophyletic or Paraphyletic?” **Science**, 301, 1482. 2003b.
- NIJHOUT, H.F. Development and evolution of adaptive polyphenisms. **Evolution and Development** 5: 9–18. 2003.
- NUNES, R.C. & BELLINI, B.C. Three new species of Entomobryoidea (Collembola: Entomobryomorpha) from Brazilian Caatinga-Cerrado transition, with identification keys to Brazilian *Cyphoderus*, *Pseudosinella* and *Trogolaphysa* species. **Zootaxa** 4420 (1): 071–096. 2018.
- NYFFELER, M. & BIRKHOFFER, K. An estimated 400 – 800 million tons of prey are annually killed by the global spider community. **Die Naturwissenschaften** 104: 30. 2017.
- OLIVEIRA, J.V.L.C.; ALVES, J.L.S.; ZEPPELINI, D. Two new *Cyphoderus* (Collembola: Paronellidae) of “tridenticulati” and “bidenticulati” groups from Brazilian Amazon. **Zootaxa**, 4350 (1), 47–60. 2017.
- O’NEILL, S.L.; HOFFMANN, A.A.; WERREN, J.H. Influential Passengers. Inherited Microorganisms and Arthropod Reproduction. Oxford, UK: **Oxford University Press**, 1997.
- OUBORG, N.J.; ANGELONI, F.; VERGEER, P. An essay on the necessity and feasibility of conservation genomics. **Conservation Genetics**, vol. 11, no. 2, pp. 643–653, 2010a.
- OUBORG, N.J.; PERTOLDI, C.; LOESCHCKE, V.; BIJLSMA, R.; HEDRICK, P.W. Conservation genetics in transition to conservation genomics. **Trends in Genetics** 26:177–187. 2010b.
- PÄÄBO, S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. **Proc. Natl. Acad. Sci. U.S.A.** 86: 1939–1943. 1989.
- PALACIOS-VARGAS, J.G., & ARBEA, J.I. Neotropical species of *Proisotoma* (Collembola: Isotomidae), with description of two new cave species from America. **Revista Mexicana De Biodiversidad** 80, 445–453. 2009.
- PALÉVODY, C. Microorganismes intracellulaires dans l’ovaire du Collemboule, *Isotomide Folsomia candida*. **C R Acad Sc Paris Series D** 275: 401–404, 1972.
- PALUMETS, J. Analysis of phytomass partition in Norway spruce (thesis). **University of Tartu Press**, 47 p, 1993.
- PARRA, G.; BRADNAM, K.; KORF, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. **Bioinformatics**, 23:1061–1067. 2007.
- PARKINSON, D.; VISSER, S.; WHITTAKER, J.B. Effects of collembolan grazing on fungal colonization of leaf litter. **Soil Biol. Biochem.** 11, 529–535. 1979.
- PETERSEN, H. General aspects of collembolan ecology at the turn of the millen-nium. **Pedobiologia** 46, 246–260. 2002.

- PFENNIG, D.W. & EHRENREICH, I.M. Towards a gene regulatory network perspective on phenotypic plasticity, genetic accommodation and genetic assimilation. **Molecular Ecology** 23: 4438–4440. 2014.
- PFENNIG, D.W. & MCGEE, M. Resource polyphenism increases species richness: a test of the hypothesis. **Philosophical Transactions of the Royal Society of London Series B: Biological Sciences** 365: 577–591. 2010.
- PHILLIPS, A.J. & SIMON, C. Simple, Efficient, and Nondestructive DNA Extraction Protocol for Arthropods. **Annals Of The Entomological Society Of America**: Vol. 88, no. 3. 1995.
- PIGLIUCCI, M. & MURREN, C.J. Genetic assimilation and a possible evolutionary paradox: can macroevolution sometimes be so fast as to pass us by? **Evolution** 57: 1455–1464. 2003.
- PORCO, D.; BEDOS, A.; DEHARVENG, L. Description and DNA barcoding assessment of the new species *Deutonura gibbosa* (Collembola: Neanuridae: Neanurinae), a common springtail of Alps and Jura. **Zootaxa** 2639, 59–68. 2010.
- PORCO, D.; BEDOS, A.; GREENSLADE, P.; JANION, C.; SKARZYNSKI, D.; STEVENS, M.I., JANSEN VAN VUUREN, B.; DEHARVENG, L. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. **Invertebrate Systematics**, 26, 470–477, 2012.
- PORCO, D.; SKARZYNSKI, D.; DECAENS, T.; HEBERT, P.D.N.; DEHARVENG, L. Barcoding the Collembola of Churchill: a molecular taxonomic reassessment of species diversity in a sub-Arctic area, **Mol. Ecol. Resour.** 14: 249-261. 2014.
- POTAPOV, A.A.; SEMENINA, E.E.; KOROTKEVICH, A.Y.; KUZNETSOVA, N.A.; TIUNOV, A.V. Connecting taxonomy and ecology: Trophic niches of collembolans as related to taxonomic identity and life forms. **Soil Biology and Biochemistry** 101: 20 – 31. 2016.
- PRICE, T.D.; QVARNSTROM, A.; IRWIN, D.E. The role of phenotypic plasticity in driving genetic evolution. **Proceedings of the Royal Society of London Series B: Biological Sciences** 270: 1433–1440. 2003.
- PRIMMER, C.R. From conservation genetics to conservation genomics. **Annals of the New York Academy of Sciences** 1162:357–368. 2009.
- PRIMMER, C.R.; PAPAKOSTAS, S.; LEDER, E.H.; DAVIS, M.J.; RAGAN, M.A. Annotated genes and nonannotated genomes: cross-species use of Gene Ontology in ecology and evolution research. **Molecular Ecology**. 22:3216–3241. 2013.
- ROBERTS, J.M.K. & WEEKS, A.R. Genetic structure and long-distance dispersal in populations of the wingless pest springtail, *Sminthurus viridis* (Collembola: Sminthuridae). **Genet. Res., Camb**, 93, pp. 1–12. 2011.
- ROESCH, L.F.W.; FULTHORPE, R.R.; RIVA, A.; CASELLA, G.; HADWIN, A.K.M.; KENT, A.D.; DAROUB, S.H.; CAMARGO, F.A.O.; FARMERIE, W.G.; TRIPLETT, E.W. Pyrosequencing enumerates and contrasts soil microbial diversity. **The ISME Journal**, v.1, p.283-290, 2007.
- ROUGERIE, R.; DECAËNS, T.; DEHARVENG, L.; PORCO, D.; JAMES, S.W.; CHANG, C-H.; RICHARD, B.; POTAPOV, M.; SUHARDJONO, Y.; HEBERT, P.D.N. DNA barcodes for soil animal taxonomy. **Pesquisa Agropecuária Brasileira** 44(8), 789–802. 2009.

- ROUGERIE, R., DEHARVENG, L., PORCO, D. & BEDOS, A. Geographical barriers and ecological gradients: geographical distribution of the subspecies and morphotypes of *Deutonura deficiens* Deharveng, 1979 (Collembola: Neanuridae). **Pedobiologia**, 50, 83–93. 2006.
- ROZOWSKY, J.S.; NEWBURGER, D.; SAYWARD, F.; WU, J.; JORDAN, G.; KORBEL, J.O.; NAGALAKSHMI, U.; YANG, J.; ZHENG, D.; GUIGÓ, R.; GINGERAS, T.R.; WEISSMAN, S.; MILLER, P.; SNYDER, M.; GERSTEIN, M.B. The DART classification of unannotated transcription within the ENCODE regions: associating transcription with known and novel loci. **Genome Res**, 17:732–745. 2007.
- RUBINOFF, D.; CAMERON, S.; WILL, K. A genomic perspective on the shortcomings of mitochondrial DNA for ‘barcoding’ identification. **J. Heredity** 97, 581–594. 2006.
- RUSEK, J. Biodiversity of Collembola and their functional role in the ecosystem. **Biodivers. Conserv.** 7, 1207–1219. 1998.
- RUSEK, J. Taxonomy of Collembola at the beginning of the new millennium. **Pedobiologia** 46, 215–224. 2002.
- SANGER, F.; NICKLEN, S.; COULSON, A.R. DNA sequencing with chain-terminating inhibitors. **PNAS**;74(12):5463–5467. 1977.
- SCHEFFER, S.J. Molecular evidence of cryptic species within the *Liriomyza huidobrensis* (Diptera: Agromyzidae). **Journal of Economic Entomology**, v.93, p.1146–1151, 2000.
- SCHNEIDER, C. & D’HAESE, C.A. Morphological and molecular insights on *Megalothorax*: the largest Neelipleona genus revisited (Collembola). **Invertebr. Syst.** 27(3), 317–364. 2013.
- SELLIN, A. Variation in hydraulic architecture of *Picea abies* (L.) Karst. trees grown under different environmental conditions (thesis). **University of Tartu Press**, 119 p, 1994.
- SETÄLÄ, H.; BERG, M.P.; JONES, T.H. Trophic structure and functional redundancy in soil communities, in: BARDGETT, R.D.; USHER, M.B.; HOPKINS, D.W. (Eds.), **Biological Diversity and Function in Soil**, Cambridge University Press, Cambridge, pp. 236–249. 2005.
- SIMON, S. & HADRYIS, H. A comparative analysis of complete mitochondrial genomes among Hexapoda. **Molecular Phylogenetics and Evolution**. 2013.
- SOTO-ADAMES, F.N. Molecular phylogeny of the Puerto Rican *Lepidocyrtus* and *Pseudosinella* (Hexapoda: Collembola), a validation of Yoshii’s “color pattern species”. **Molecular phylogenetics and evolution** 25: 27–42. 2002.
- SOTO-ADAMES, F.N.; BARRA, J.A.; CHRISTIANSEN, K.; JORDANA, R. Suprageneric classification of Collembola Entomobryomorpha. **Annals Entomological Society America** 101: 501–513. 2008.
- SOTO-ADAMES, F.N.; JORDANA, R.; BAQUERO, E. Comparative analysis of the dorsal chaetotaxy of *Troglopedetes*, *Trogolaphysa* and *Campylothorax* supports the synonymization of tribes Paronellini and Troglopedetini (Collembola: Paronellidae). **Journal of Insect Science** 14: 278–29. 2014.
- SOTO-ADAMES, F.N. & BELLINI, B.C. Dorsal chaetotaxy of neotropical species supports a basal position for the genus *Lepidonella* among scaled Paronellidae (Collembola, Entomobryodea). *Florida Entomologist*, Volume 98, No. 1. 2015.

- STEARNS, S.C. The evolutionary significance of phenotypic plasticity. **Bioscience** 39: 436–445. 1989.
- STEINER, C.C.; PUTNAM, A.S.; HOECK, P.E.A.; RYDER, O.A. Conservation genomics of threatened animal species. **Annual Review of Animal Biosciences**, vol. 1, pp. 261–281, 2013.
- STEVENS, M.I. & D’HAESE, C.A. Morphologically tortured: Taxonomic placement of an Antarctic springtail (Collembola: Isotomidae) misguided by morphology and ecology. **Zoologica Scripta**, 46, 180–187. 2017.
- STEVENS, M. I. & HOGG, I. D. Long-term isolation and recent range expansion revealed for the endemic springtail *Gomphiocephalus hodgsoni* from southern Victoria Land, Antarctica. **Molecular Ecology**, 12, 2357–2369, 2003.
- STORK, N.E.; GRIMBACHER, P.S.; STOREY, R.; OBERPRIELER, R.G.; REID, C.; SLIPINSKI, S.A. What determines whether a species of insect is described? Evidence from a study of tropical forest beetles. **Insect Conservation and Diversity**, v.1, p.114–119, 2008.
- STOUTHAMER, R.; BREEUWER, J.A.J.; HURST, G.D.D. *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. **Ann Rev Microbiol** 53: 71–102, 1999.
- SULTAN, S.E. Development in context: the timely emergence of eco-devo. **Trends in Ecology & Evolution** 22: 575–582. 2007.
- SUMNER, A.T., DE LA TORRE, J. & STUPPIA, L. The distribution of genes on chromosomes: A cytological approach. **J Mol Evol**: 37: 117, 1993.
- SUN, X.; ZHANG, F.; DING, Y.; DAVIES, D.W.; LI, Y.; WU, D. Delimiting species of *Protaphorura* (Collembola: Onychiuridae): Integrative evidence based on morphology, DNA sequences and geography. **Scientific Reports**, 7, 8261. 2017.
- TAMURA, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Mol. Biol. Evol.** 28, 2731–2739. 2011.
- THE 1000 GENOMES PROJECT CONSORTIUM. A map of human genome variation from population-scale sequencing. **Nature** 467, 1061–1073, 2010.
- THIMM, T.; HOFFMANN, A.; BORKOTT, H.; MUNCH, J.C.; TEBBE, C.C. The gut of the soil microarthropod *Folsomia candida* (Collembola) is a frequently changeable but selective habitat and a vector for microorganisms. **Appl Environ Microbiol** 64: 2660–2669, 1998.
- THIMM, T. & LARINK, O. Grazing preferences of some Collembola for endomycorrhizal fungi. **Biol. Fertil. Soils** 19, 266–268. 1995.
- THORNE, M.A.S.; WORLAND, M.R.; FERET, R.; DEERY, M.J.; LILLEY, K.S.; CLARK, M.S. Proteomics of cryoprotective dehydration in *Megaphorura arctica* Tullberg 1876 (Onychiuridae: Collembola). **Insect Molecular Biology**, 20(3), 303–310. 2011.
- TIMMERMANS, M.J.T.N.; ELLERS, J.; MARIËN, J.; ERHOEFS, C.V.; FERWERDA, E.B.; VAN STRAALLEN, N.M. Genetic structure in *Orchesella cincta* (Collembola): strong subdivision of European populations inferred from mtDNA and AFLP markers. **Molecular Ecology**, 14, 2017–2024, 2005.
- TORRICELLI, G.; CARAPELLI, A.; CONVEY, P.; NARDI, F.; BOORE, J. L.; FRATI, F. High divergence across the whole mitochondrial genome in the “pan-Antarctic” springtail *Friezea grisea*: Evidence for cryptic species? **Gene**, 449, 30–40, 2010.

- USHER, M.B. Population and community dynamics in the soil ecosystem. In Ecological interactions in soil: plants, microbes and animals. *Edited by* FITTER, A.H.; ATKINSON, D.; READ, D.J.; USHER, M.B. **Blackwell, Palo Alto**, pp. 243–265. 1985.
- VAN STRAALLEN, N.M.; TIMMERMAN, M.J.T.N.; ROELOFS, D.; BERG, M.P. Apterygota in the spotlights of ecology, evolution and genomics. **European Journal of soil biology** 44: 452– 457. 2008.
- VICENTE, V.A.; WEISS, V.A.; BOMBASSARO, A.; MORENO, L.F.; COSTA, F.F.; RAITTZ, R.T.; LEÃO, A.C.; GOMES, R.R.; BOCCA, A.L.; FORNARI, G.; DE CASTRO, R.J.A.; SUN, J.; FAORO, H.; TADRA-SFEIR, M.Z.; BAURA, V.; BALSANELLI, E.; ALMEIDA, S.R.; DOS SANTOS, S.S.; TEIXEIRA, M.D.M.; SOARES FELIPE, M.S.; DO NASCIMENTO, M.M.F.; PEDROSA, F.O.; STEFFENS, M.B.; ATTILI-ANGELIS, D.; NAJAFZADEH, M.J.; QUEIROZ-TELLES, F.; SOUZA, E.M.; DE HOOG, S. Comparative Genomics of Sibling Species of *Fonsecaea* Associated with Human Chromoblastomycosis. **Front. Microbiol.** 8:1924, 2017.
- VON SALTZWEDEL, H.; SCHEU, S.; SCHAEFER, I. Genetic structure and distribution of *Parisotoma notabilis* (Collembola) in Europe: Cryptic diversity, split of lineages and colonization patterns. **PLoS ONE**, 12(2): e0170909, 2017.
- WAGNER, D.; BROWN, M.J.F.; GORDON, D.M. Harvester ants nests, soil biota and soil chemistry. **Oecologia** 112: 232-236. 1997.
- WEI, F.; HU, Y.; ZHU, L.; BRUFORD, M.W.; ZHAN, X.; ZHANG, L. Black and white and read all over: the past, present and future of giant panda genetics. **Molecular Ecology** 21:5660–5674. 2012.
- WERREN, J.H. Biology of *Wolbachia*. **Ann Rev Entomol** 42: 587–609, 1997.
- WHALLEY, P. & JARZEMBOWSKI, E.A. A new assessment of *Rhyniella*, the earliest known insect. From the Devonian of Rhynie, Scotland. **Nature**, Lond. 291:317. 1981.
- WHEELER, D.A.; SRINIVASAN, M.; EGHOLM, M.; SHEN, Y.; CHEN, L.; MCGUIRE, A.; HE, W.; CHEN, Y.J.; MAKHIJANI, V.; ROTH, G.T.; GOMES, X.; TARTARO, K.; NIAZI, F.; TURCOTTE, C.L.; IRZYK, G.P.; LUPSKI, J.R.; CHINAULT, C.; SONG, X.Z.; LIU, Y.; YUAN, Y.; NAZARETH, L.; QIN, X.; MUZNY, D.M.; MARGULIES, M.; WEINSTOCK, G.M.; GIBBS, R.A.; ROTHBERG, J.M. The complete genome of an individual by massively parallel DNA sequencing. **Nature** 452:872–876. 2008.
- WILL, K.W. & RUBINOFF, D. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. **Cladistics** 20, 47–55. 2004.
- WILL, K.W.; MISHLER, B.D.; WHEELER, Q.D. The perils of DNA barcoding and the need for integrative taxonomy. **Syst. Biol.** 54, 844–851. 2005.
- WISE, K. A. J. Collembola (Springtails). **Antarc. Res. Ser.**, 10: 123-147, 1967.
- WISE, K. A. J.; FEARON, C. E.; WILKES, O. R. Entomological investigations in Antarctica, 1962-63 Season. **Pacific Insects**, 6: 541-570, 1971.
- WU, C.; JORDAN, M.D.; NEWCOMB, R.D.; GEMMELL, N.J.; BANK, S.; MEUSEMANN, K.; DEARDEN, P.K.; DUNCAN, E.J.; GROSSER, S.; RUTHERFORD, K.; GARDNER, P.P.; CROWHURST, R.N.; STEINWENDER, B.; TOOMAN, L.K.; STEVENS, M.I.; BUCKLEY, T.R. Analysis of the genome of the New Zealand giant collembolan (*Holacanthella duospinosa*) sheds light on hexapod Evolution. **BMC Genomics**, 18:795, 2017.

- XU, Y. Molecular plant breeding. **Molecular Breeding Tool: Omics and Arrays**, p. 68, 2012.
- YANDELL, M. & ENCE, D. A beginner's guide to eukaryotic genome annotation. **Nature Reviews Genetics** 13:329–342. 2012.
- YANG, Z. & RANNALA, B. Bayesian species delimitation using multilocus sequence data. **Proceedings of the National Academy of Sciences USA**, 107, 9264–9926. 2010.
- YÉO, K.; MOLET, M.; PEETERS, C. When David and Goliath share a home: compound nesting of *Pyramica* and *Platythyrea* ants. **Insect. Soc.** 53: 435–438. 2006.
- YIN, Z.W.; CAI, C.Y.; HUANG, D.Y.; LI, L.Z. Specialized adaptations for springtail predation in Mesozoic beetles. **Scientific Reports** 7: 98. 2017.
- YOSHII, R. Cyphoderid Collembola of Sabah. **Contributions of the Biological Laboratory of Kyoto University** 26: 1–16. 1980.
- YOSHII, R. Notes on some Cyphoderid Collembola of the tropical Asia. **Contributions of the Biological Laboratory of Kyoto University** 27: 121–136. 1987.
- YOSHII, R. Miscellaneous notes on the Collembola of Macaronesia. **Contributions of the Biological Laboratory of Kyoto University** 27: 535–540. 1990.
- YOSHII, R. Interim report of the taxonomic researches toward the Collembolan family Cyphoderidae. **Contributions of the Biological Laboratory of Kyoto University** 28: 99–118. 1992.
- ZEPPELINI, D.F. & BELLINI B.C. Introdução ao estudo dos Collembola. João Pessoa, Paraíba: **Editora Universitária**, Universidade Federal da Paraíba. 82 p, 2004.
- ZEPPELINI, D. & OLIVEIRA, J.V.L.C. Chaetotaxy of Neotropical *Cyphoderus caetetus* sp. nov. with comments on the taxonomic position of Cyphoderinae within Paronellidae (Collembola, Entomobryoidea). **Zootaxa** 4098 (3): 560–570. 2016.
- ZEPPELINI, D.; BELLINI, B.C.; DUARTE, A.J.C.; HERNANDEZ, M.I.M. Collembola as bioindicators of restoration in mined sand dunes of northeastern Brazil. **Biodiversity and Conservation** 18: 1161–1170, 2009.
- ZEPPELINI, D.; LOPES, B.C.H.; LIMA, E.C.A. A New Species of *Szeptyckitheca* (Collembola, Symphypleona, Sphyrothecinae) from Brazil. **Neotrop Entomol**, 2018.
- ZEPPELINI, D.; QUEIROZ, G.C.; BELLINI, B.C. Collembola in **Catálogo Taxonômico da Fauna do Brasil**. PNUD. Disponível em: <<http://fauna.jbrj.gov.br/fauna/faunadobrasil/379>>. 2019.
- ZETTEL, J. & ZETTEL, U. Seasonal and reproductional polymorphism in *Ceratophysella sigillata* (Uzel) (Collembola: Hypogastruridae). **Acta Zool. Fenn.** 195, 154–156. 1994.
- ZHANG, F.; YU, D.; LUO, Y.; HO, S.Y.W.; WANG, B.; ZHU, C. Cryptic diversity, diversification and vicariance in two species complexes of *Tomocerus* (Collembola, Tomoceridae) from China. **Zoologica Scripta** 43: 393–404. 2014a.
- ZHANG, F.; CHEN, Z.; DONG, R-R.; DEHARVENG, L.; STEVENS, M.I.; HUANG, Y-.; ZHU, C-D. Molecular phylogeny reveals independent origins of body scales in Entomobryidae (Hexapoda: Collembola). **Molecular Phylogenetics and Evolution**, 70: 231–239. 2014b.

ZHANG, F.; SUN, D-D.; YU, D-Y.; WANG, B-X. Molecular phylogeny supports S-chaetae as a key character better than jumping organs and body scales in classification of Entomobryoidea (Collembola). *Scientific Reports*, 5:12471. 2015.

ZHANG, F.; MA, Y.; GREENSLADE, P. New Australian Paronellidae (Collembola) reveal anomalies in existing tribal diagnoses. *Invertebrate Systematics*, 31, 375–393. 2017.

ZHANG, F.; JANTARIT, S.; NILSAI, A.; STEVENS, M.I.; DING, Y.; SATASOOK, C. Species delimitation in the morphologically conserved *Coecobrya* (Collembola: Entomobryidae): A case study integrating morphology and molecular traits to advance current taxonomy. *Zoologica Scripta*;1–15. 2018a.

ZHANG, F.; YU, D.; STEVENS, M.I.; DING, Y. Colouration, chaetotaxy and molecular data provide species-level resolution in a species complex of *Dicranocentrus* (Collembola: Entomobryidae). *Invertebrate Systematics*, 32, 1298–1315. 2018b.

ZHANG, F.; DING, Y.; ZHU, C-D.; ZHOU, X.; ORR, M. C.; SCHEU, S. LUAN, Y-X. Phylogenomics from low-coverage whole-genome sequencing. *Methods Ecol Evol*; 1–11. 2019a.

ZHANG, F.; DING, Y.; ZHOU, Q-S.; WU, J.; LUO, A.; ZHU, C-D. A High-quality Draft Genome Assembly of *Sinella curviseta*: A SoilModel Organism (Collembola). *Genome Biol. Evol.* 11(2):521–530. 2019b.

APÊNDICE A

REPRODUÇÃO E IDENTIFICAÇÃO DOS CULTIVOS LABORATORIAIS DE *Cyphoderus*

Registra-se nesta dissertação a primeira reprodução de colêmbolos coletados em cavernas de minério de ferro, MG e áreas verdes urbanas de João Pessoa, PB. Ao longo das gerações cultivadas, através da confecção de lâminas semipermanentes e uso de chaves específicas foi identificada a espécie *Cyphoderus similis* a partir dos cultivos de espécimes de MG e a espécie *Cyphoderus innominatus* a partir dos cultivos de espécimes da PB. Torna-se registrado também os métodos aplicados à criação de duas novas espécies de Collembola em laboratório, destacando sua praticidade de manutenção de cultivos e boa capacidade de reprodução *ex situ*.

Bem como citado em literatura, encontramos algumas variações morfológicas nas populações mantidas em cultivos, sendo a mais frequente relacionada a variação no número de cerdas (pluriquetose) na região dorsal do dens (região da fúrcula). Mencionado por OLIVEIRA e colaboradores (2017) sobre o polimorfismo acentuado das espécies *C. javanus* e *C. similis* de região Neotropical e sendo reforçado por NUNES & BELLINI (2018) a necessidade de revisão taxonômica e novos recursos taxonômicos integrativos mais claros para o grupo. Foi considerado, então, esta variação morfológica dos *Cyphoderus* cultivados em laboratório como uma resposta à alta endogamia, inúmeras gerações mantidas a partir de uma variabilidade genética fixa, ou seja, podendo ser correspondente ao alto número de alelos da população inicial coletada, discutido no **Apêndice B**.

APÊNDICE B

RELATO DE POLIMORFISMO EM *Cyphoderus*

Todos os organismos têm a capacidade de alterar algum aspecto em seu fenótipo em resposta a mudanças ambientais, seja em sua fisiologia, seu comportamento, sua morfologia ou até a nível de expressão gênica (NIJHOUT, 2003; SULTAN, 2007; GILBERT & EPEL, 2009). Fenótipos diferentes dentro de uma mesma população pode ser resultado da variação na expressão gênica causada por polimorfismo na sequência codificadoras de proteínas no DNA, evidenciando a conexão que o fenótipo apresenta com alterações observadas no genótipo (LAWNICZAK *et al.*, 2008).

A morfologia de um Collembola adulto está sujeita a interferências tanto de fatores ambientais quanto de fatores genéticos intrínsecos da espécie, podendo com isso resultar em diferentes morfotipos surgidos em curtos intervalos de tempo durante seu ciclo de vida (CHRISTIANSEN, 1957; ZETTEL & ZETTEL, 1994).

Dentro do gênero *Cyphoderus* a espécie *C. similis* apresenta um histórico de dificuldades morfológicas em distinguir-se de *C. assimilis*, *C. subsimilis* e *C. simulans* devido a caracteres com alta variação interpretativa e pouca capacidade de aceita-los como características reais que para fins taxonômicos, como por exemplo a ausência da túnica, o dente mediano não pareado, a *tenente hair* clavada e o comprimento relativo do mucro (CHRISTIANSEN, 1957). Desde os mais antigos registros que descrevem a sistemática de *Cyphoderus* baseando-se tradicionalmente na estrutura do mucro, no número e arranjo de escamas nos dentes e da estrutura da garra (DELAMARE-DEBOUTTEVILLE, 1948), sendo que a maioria dos caracteres foram inicialmente descritos a partir de espécimes raros, tornando-se difícil reconhecer o valor taxonômico verdadeiro, uma vez que foram determinados na ausência de quaisquer valores estatísticos ou distribuição geográfica (CASSAGNAU, 1963). Além disso, CASSAGNAU (1963) reconhece também a plasticidade morfológica de espécies de cifoderídeos, destacando a importância de não subestimar evoluções regressivas especificamente do mucro, podendo levar a alguns polimorfismos dentro de uma população típica, no caso, levando a interpretar o termo “típica” como sendo uma população anteriormente homogênea para aquela característica, ainda assim reconheceu o polimorfismo referente a estrutura do mucro de *C. albinus*, *C. innominatus*, *C. javanus* e *C. similis* observando o tamanho, a posição e a presença de espinhos do mucro destas espécies.

Um fenômeno amplamente citado como sendo a rota que explica a diversificação evolutiva proporcionada pela plasticidade fenotípica é o processo evolutivo conhecido como assimilação genética. Em que este mecanismo é causado por uma mudança ambiental que desencadeia a expressão de uma nova variante fenotípica hereditária, em que este fenótipo variante que é reconhecido, inicialmente, como raro apresenta um aumento em sua frequência dentro da população, assim ao longo das gerações esta nova variante fenotípica poderá ser fixada constituindo-se como uma nova variação dentro da população (BRADSHAW, 1965; STEARNS, 1989; PRICE *et al.*, 2003; BRAENDLE & FLATT, 2006; PFENNIG & MCGEE, 2010; FITZPATRICK, 2012; EHRENREICH & PFENNIG, 2016). As alterações ambientais frequentemente tendem a afetar diretamente um número muito maior de indivíduos simultaneamente, diferente de processo de mutação genética que inicialmente afeta um único indivíduo e seus descendentes direto, portanto a assimilação genética tem sido considerada como um potencial mecanismo de origem de novos traços fenotípicos (PIGLIUCCI & MURREN, 2003; KASSEN & BATAILLON, 2006; AUBRET & SHINE, 2009; HALLIGAN & KEIGHTLEY, 2009; LANDE, 2009; MOCZEK *et al.*, 2011).

Assim, avanços em técnicas de sequenciamento de DNA proporciona uma ferramenta adicional nos meios até então utilizados para delimitar espécies, principalmente quando se refere a espécies crípticas de Collembola (FRATI *et al.*, 2000a; SOTO-ADAMES, 2002; BURKHARDT & FILSER, 2005; ZHANG *et al.*, 2014a; KATZ *et al.*, 2015). Consecutivamente, a partir de novas tecnologias de sequenciamento e facilitação na geração de genomas e transcriptomas de alta qualidade, no qual abrangem organismos modelos ou não, estes dados genéticos também tem a capacidade de servir como ferramentas de identificação de mecanismos referentes ao processo de assimilação genética (PFENNIG & EHRENREICH, 2014; EHRENREICH & PFENNIG, 2016).

APÊNDICE C

EXTRAÇÃO E QUANTIFICAÇÃO DO DNA DE *Cyphoderus*

Registra-se nesta dissertação a primeira extração de DNA de Collembola realizado no ITV-DS com utilização de 50 indivíduos em cada população de *Cyphoderus*, com valores de 30Hz por 1 minuto de maceração mecânica.

Foi alcançado valores de 44.1 ng/μl (*C. similis*) e 16.0 ng/μl (*C. innominatus*). Portanto, devido ao baixo valor de concentração de DNA de *C. innominatus*, foi aconselhado a tecnologia de sequenciamento MiniSeq para a amostra.

A construção das bibliotecas genômicas resultaram em 8.6 ng/μl com tamanho médio dos fragmentos de 512 pb para *C. similis*, enquanto para *C. innominatus* foi encontrado valores de 23.6 ng/μl com tamanho médio dos fragmentos de 469 pb, no qual esperava-se valores de concentração da biblioteca de até 2 ng/μl e tamanho médio dos fragmentos de 400-600 pb como valores ideais.

APÊNDICE D

ANÁLISE ESTRUTURAL PÓS-SEQUENCIAMENTO DOS GENOMAS DE *Cyphoderus*

A **Tabela 15** mostra os valores de *output* de cada sequenciamento para os *Cyphoderus*, trazendo um comparativo entre os dados brutos não-tratados e os dados pós-tratamentos de qualidade, informando o comprimento total sequenciado, o comprimento das *reads*, e seus valores de qualidade (*phred*), no qual podemos observar a necessidade de tratamentos dos dados pós sequenciamento genômico.

Abaixo observam-se os gráficos referentes à qualidade das sequências (**Figura 28 e 29**), onde as regiões alaranjada e avermelhada apresentam valores de *phred* considerados ruins. Estes gráficos são exemplos de como se deu o tratamento dos dados pós-sequenciamento através do FastQC (dados brutos à esquerda, dados tratados à direita). Para os sequenciamentos foram realizados o tratamento de qualidade das sequências e eliminando sequencias curtas (<50 pb de comprimento mínimo) e com valores de *phred* baixos (<30).

O sequenciamento com bibliotecas M-P resultou também em um conjunto *singletons* de *reads*, referentes às sequências tratadas não-pareadas. Os *singletons* podem representar importantes sequências para uso de montagem para algumas ferramentas de montagem. Um total de 18.014.630 *reads* para *C. similis* e 11.408.174 *reads* para *C. innominatus*.

Tabela 15. Resumo dos dados moleculares referentes ao pós-sequenciamento dos *Cyphoderus similis* e *Cyphoderus innominatus*. Comparando os dados brutos (B) e tratados (T). Valores de *phred-score* <30 é considerado ruim.

Espécie	Sequenciamento	Tipo dos Dados	Comprimento (bp)	Comprimento das reads	Phred-score
<i>C. similis</i>	NextSeq <i>P-E</i>	B	46.842.269	35-151	< 30
		T	39.799.096	50-151	> 30
	NextSeq <i>M-P</i>	B	32.365.835	35-151	< 30
		T	11.362.287	70-151	> 30
<i>C. innomunatus</i>	MiniSeq <i>P-E</i>	B	38.851.815	48-151	< 30
		T	34.745.970	50-151	>30
	NextSeq <i>M-P</i>	B	21.112.749	35-151	<30
		T	7.482.507	70-151	>30

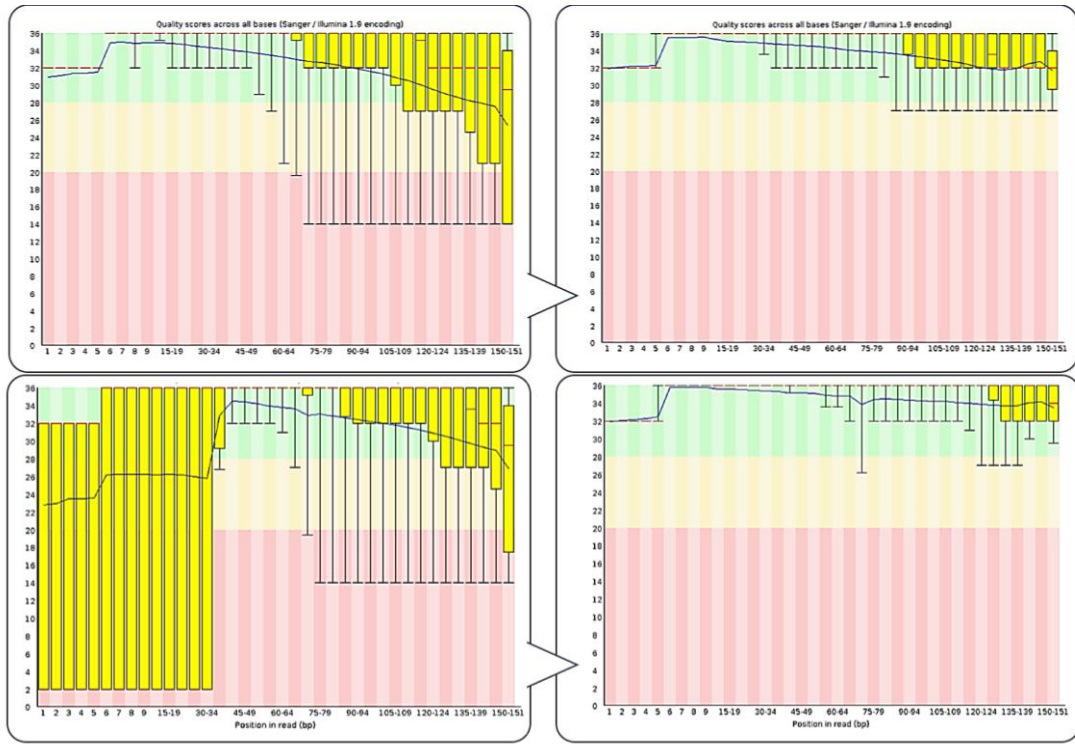


Figura 28. Gráfico extraído do FastQC referente à qualidade das seqüências brutas (imagem à esquerda) e tratadas (imagem à direita) para o sequenciamento NextSeq P-E (imagens de cima) e NextSeq M-P (imagens de baixo) da amostra de DNA de *C. similis*.

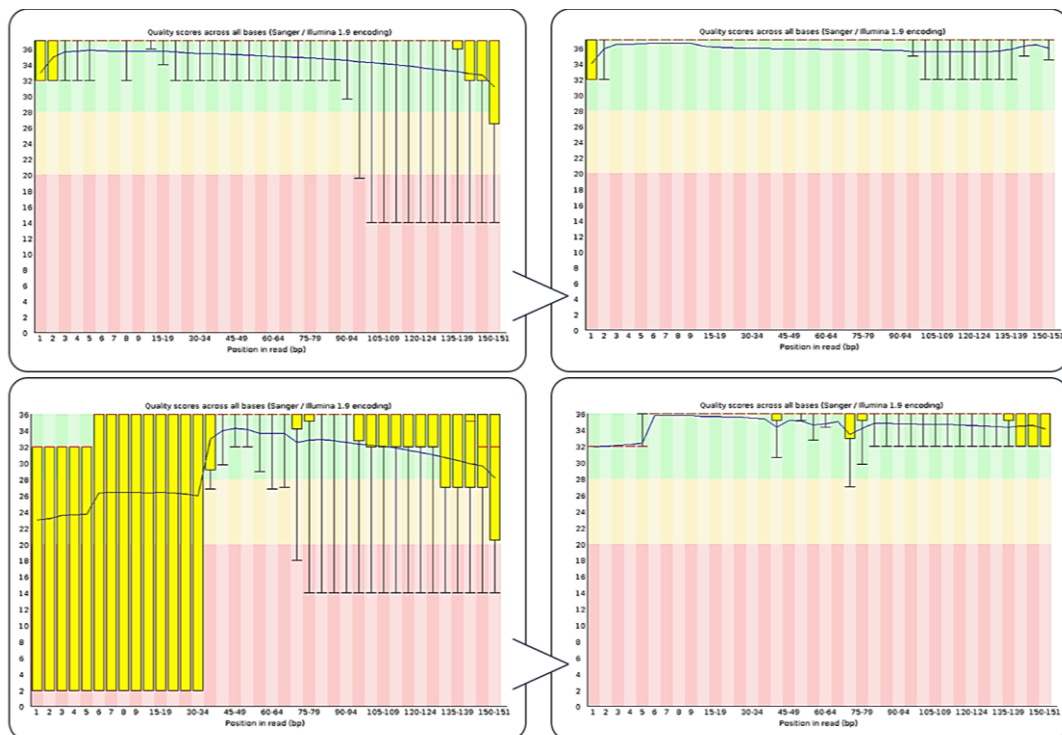


Figura 29. Gráfico extraído do FastQC referente à qualidade das seqüências brutas (imagem à esquerda) e tratadas (imagem à direita) para o sequenciamento MiniSeq P-E (imagens de cima) e NextSeq M-P (imagens de baixo) da amostra de DNA de *C. innominatus*.