

Gabriel Alves Vieira

**FILOGENÔMICA MITOCONDRIAL SEM
CUSTOS: UMA PROVA DE CONCEITO
COM FORMIGAS DA SUBFAMÍLIA
PSEUDOMYRMECINAE
(HYMENOPTERA : FORMICIDAE)**

Brasil

Março de 2019

Gabriel Alves Vieira

**FILOGENÔMICA MITOCONDRIAL SEM CUSTOS:
UMA PROVA DE CONCEITO COM FORMIGAS
DA SUBFAMÍLIA PSEUDOMYRMECINAE
(HYMENOPTERA : FORMICIDAE)**

Dissertação apresentada ao Programa de
Pós-Graduação em Química Biológica,
Instituto de Bioquímica Médica Leopoldo
de Meis, Universidade Federal do Rio de
Janeiro, como parte dos requisitos para
obtenção do título de Mestre em Química
Biológica.

Ministério da Educação

Universidade Federal do Rio de Janeiro

Instituto de Bioquímica Médica Leopoldo de Meis

Orientador: Dr. Francisco Prosdocimi

Brasil

Março de 2019

Gabriel Alves Vieira

FILOGENÔMICA MITOCONDRIAL SEM CUSTOS: UMA PROVA DE CONCEITO COM FORMIGAS DA SUBFAMÍLIA PSEUDOMYRMECINAE (HYMENOPTERA : FORMICIDAE)/ Gabriel Alves Vieira. – Brasil, Março de 2019-
83p. : il. (algumas color.) ; 30 cm.

Orientador: Dr. Francisco Prosdocimi

Tese (Doutorado) – Ministério da Educação

Universidade Federal do Rio de Janeiro

Instituto de Bioquímica Médica Leopoldo de Meis , Março de 2019.

1. Palavra-chave1. 2. Palavra-chave2. 2. Palavra-chave3. I. Orientador. II. Universidade xxx. III. Faculdade de xxx. IV. Título

FILOGENÔMICA MITOCONDRIAL SEM CUSTOS: UMA PROVA DE CONCEITO
COM FORMIGAS DA SUBFAMÍLIA PSEUDOMYRMECINAE (HYMENOPTERA :
FORMICIDAE) Gabriel Alves Vieira

Dissertação apresentada ao Programa de Pós-Graduação em Química
Biológica, Instituto de Bioquímica Médica Leopoldo de Meis,
Universidade Federal do Rio de Janeiro, como parte dos requisitos para
obtenção do título de Mestre em Química Biológica.

Aprovado em: 26/03/2019

BANCA EXAMINADORA

Dr. Francisco Prosdocimi

Prof. Adjunto do Instituto de Bioquímica Médica Leopoldo de Meis da
Universidade Federal do Rio de Janeiro – UFRJ

Dr^a. Carla Ribeiro Polycarpo

Prof^a. Associada do Instituto de Bioquímica Médica Leopoldo de Meis
da Universidade Federal do Rio de Janeiro – UFRJ

Dr^a Ana Carolina Martins Junqueira

Prof^a. Adjunta do Instituto de Bioquímica Médica Leopoldo de Meis da
Universidade Federal do Rio de Janeiro – UFRJ

Dr. Marcus Fernandes de Oliveira

Prof. Adjunto do Instituto de Bioquímica Médica Leopoldo de Meis da
Universidade Federal do Rio de Janeiro – UFRJ

Suplente externo: Dr. Marcelo Weksler

Prof. Titular do Programa de Pós-Graduação em Zoologia do Museu
Nacional da Universidade Federal do Rio de Janeiro - UFRJ

Revisor: Dr. Fernando Lucas Palhano Soares

Prof. Adjunto do Instituto de Bioquímica Médica Leopoldo de Meis da
Universidade Federal do Rio de Janeiro – UFRJ

Agradecimentos

Aos meus pais que, apesar da enorme vontade de me ter ao lado deles, demonstraram seu amor ao darem apoio incondicional a um filho que, tal qual Brás Cubas, foi dominado por uma idéia fixa: ver e aprender mais sobre bioinformática e o mundo.

À Agatinha, minha mafagafa e companheira de todos os momentos, com a qual constituí família e cresci como pessoa. Abro um sorriso cada vez que encontro uma “AGATA” em alguma mitocôndria e me lembro de como sou sortudo. Completamente desprovido de dons artísticos para te escrever músicas ou poemas, o melhor que consigo fazer é dedicar todas essas sequências, junto com meu coração, a ti.

Aos meus filhos: Seth, que vive roubando meu cobertor e pulando no meu colo sem ser convidado; e Bílquis, que só quer saber de me afofar com suas garrinhas afiadas e me trazer proteína na forma de lagartos (vivos, mortos ou em uma estranha mescla dos dois estados). Amo muito ambos e sem eles não seria plenamente feliz.

Ao Exmo. Prof. Dr. Francisco Prosdocimi, por ter lido um email (redigido com pressa e provavelmente contendo múltiplos erros ortográficos) no final de 2016 de um cara meio maluco e afobado que queria estudar bioinformática. Mais do que isso, agradeço por ele ter depositado confiança no dito cujo e aceitado orientá-lo.

A todos os meus colegas do LAMPADA/Laboratório de Genômica e Biodiversidade, em especial à Ana e Deise. Sem a ajuda e apoio de vocês, meu trabalho seria muito menos divertido /empolgante e eu certamente teria traçado arestas menos satisfatórias na tentativa de resolver esse grafo de dois anos que foi meu mestrado.

A todos aqueles que geraram e disponibilizaram os *datasets* utilizados aqui. Se eu não tivesse tido o privilégio de subir nos ombros desses gigantes, esse trabalho não existiria. Embora nós tenhamos explorado perguntas, metodologias e evidências diferentes ao nos aventurar por esses dados, acredito que comungamos de um

objetivo comum: promover o avanço de uma ciência tão aberta e colaborativa quanto possível.

À CAPES e demais agências de fomento, por darem suporte a esse trabalho

“I like the hand we’ve been dealt.”
(Uncharted 4: A Thief’s End - 2016)

Resumo

O advento do Sequenciamento de Nova Geração reduziu os custos de sequenciamento e aumentou o número de projetos genômicos para uma enorme gama de organismos, gerando uma quantidade sem precedentes de conjuntos de dados genômicos publicamente disponíveis. Muitas vezes, apenas uma pequena fração da relevância dos dados produzidos pelos pesquisadores é contemplada em seus trabalhos. Este fato permite que os dados gerados sejam reciclados em outros projetos ao redor do mundo. A montagem de mitogenomas completos é frequentemente negligenciada, embora seja útil para entender as relações evolutivas entre táxons, especialmente aqueles que apresentam baixa amostragem de mtDNA ao nível de gêneros e famílias. Esse é exatamente o caso das formigas (Hymenoptera: Formicidae) e, mais especificamente, da subfamília Pseudomyrmecinae, um grupo de formigas arbóreas com vários casos de coevolução convergente mas sem qualquer sequência mitocondrial completa disponível. Nesta dissertação reunimos, anotamos e realizamos análises genômicas comparativas de 14 novos genomas mitocondriais completos de espécies de Pseudomyrmecinae, usando apenas os conjuntos de dados públicos disponíveis no Sequence Read Archive (SRA). Utilizamos todos os mitogenomas completos disponíveis para formigas para estudar a conservação da ordem gênica e também para gerar duas árvores filogenéticas usando (i) conjunto concatenado de 13 genes mitocondriais e (ii) as seqüências mitocondriais completas. Embora as topologias das árvores tenham divergido sutilmente umas das outras (e de estudos anteriores), nossos resultados confirmam várias relações conhecidas e geram novas evidências para a classificação de grupos irmão dentro de Pseudomyrmecinae. Também realizamos uma análise de sentença para a família Formicidae e identificamos possíveis sítios nos quais inserções nucleotídicas ocorreram em mitogenomas de formigas do gênero Pseudomyrmex. Usando uma abordagem de mineração de dados/bioinformática, a dissertação atual aumentou o número de genomas mitocondriais completos disponíveis para formigas de 15 para 29, demonstrando o potencial único dos bancos de dados públicos para estudos mitogenômicos. As amplas aplicações de mitogenomas na pesquisa e a presença de dados mitocondriais em diferentes tipos de dados públicos tornam a abordagem da “mitogenômica no-budget” ideal para estudos moleculares abrangentes, especialmente para taxóons subamostrados.

Palavras-chave: Pseudomyrmecinae, Mitogenômica, Mineração de dados, Bioinformática, Filogenômica, Biologia evolutiva de formigas, Sequenciamento de nova geração, Dados públicos.

Abstract

The advent of Next Generation Sequencing has reduced sequencing costs and increased genomic projects from a huge amount of organismal taxa, generating an unprecedented amount of genomic datasets publicly available. Often, only a tiny fraction of outstanding relevance of the genome data produced by researchers is used in their works. This fact allows the data generated to be recycled in further projects worldwide. The assembly of complete mitogenomes is frequently overlooked though it is useful to understand evolutionary relationships among taxa, especially those presenting poor mtDNA sampling at the level of genera and families. This is exactly the case for ants (Hymenoptera:Formicidae) and more specifically for the subfamily Pseudomyrmecinae, a group of arboreal ants with several cases of convergent coevolution without any complete mitochondrial sequence available. In this dissertation, we assembled, annotated and performed comparative genomics analyses of 14 new complete mitochondrial genomes from Pseudomyrmecinae species relying solely on public datasets available from the Sequence Read Archive (SRA). We used all complete mitogenomes available for ants to study the gene order conservation and also to generate two phylogenetic trees using both (i) concatenated set of 13 mitochondrial genes and (ii) the whole mitochondrial sequences. Even though the tree topologies diverged subtly from each other (and from previous studies), our results confirm several known relationships and generate new evidences for sister clade classification inside Pseudomyrmecinae clade. We also performed a synteny analysis for Formicidae and identified possible sites in which nucleotidic insertions happened in mitogenomes of pseudomyrmecine ants. Using a data mining/bioinformatics approach, the current dissertation increased the number of complete mitochondrial genomes available for ants from 15 to 29, demonstrating the unique potential of public databases for mitogenomics studies. The wide applications of mitogenomes in research and presence of mitochondrial data in different public dataset types makes the “no budget mitogenomics” approach ideal for comprehensive molecular studies, especially for subsampled taxa.

Keywords: Pseudomyrmecinae, Mitogenomics, Data mining, Bioinformatics, Phy-

logenomics, Ant evolutionary biology, Next Generation Sequencing, Public data.

Lista de ilustrações

Figura 1 – Principais etapas do sequenciamento Illumina	21
Figura 2 – Preparação de bibliotecas <i>paired-end</i> and <i>mate-pair</i>	24
Figura 3 – Conversão de um único spot do formato sra para fastq	28
Figura 4 – Uso de dados pareados para o fechamento de gaps	29
Figura 5 – Anotação de montagens mitocondriais e separação de spots . . .	38
Figura 6 – Esquema representativo das montagens denovo e por referência	39
Figura 7 – Grafos de Bruijn aplicados à montagem de genomas	40

Lista de tabelas

Lista de Abreviaturas e Siglas

3'OH:	Hidroxila presente no carbono 3' de um nucleotídeo
A:	Amina (base púrica)
ATP6:	ATP sintase F0 subunidade 6
ATP8:	ATP sintase F0 subunidade 8
BLAST:	Basic Local Alignment Search Tool
BLASTp:	Protein-protein BLAST
BRIG:	Blast Ring Image Generator
BS:	Suporte de Bootstrap
C:	Citosina (base pirimídica)
COX1:	Citocromo oxidase 1
COX2:	Citocromo oxidase 2
DDBJ:	DNA Data Bank of Japan
DNA:	Ácido desoxirribonucleico
EBI:	European Bioinformatics Institute
EMBL:	European Molecular Biology Laboratory
G:	Guanina (base púrica)
Gpb:	Giga pares de base (1000000000 bp)
GTR+G+I:	Modelo de substituição nucleotídica “General Time Reversible + Gamma distributed + Invariant sites”
INSDC:	International Nucleotide Sequence Database Collaboration

kpb:	Kilo pares de base (1000 pb)
ML:	Máxima Verossimilhança
MMG:	Metagenômica mitocondrial
Mpb:	Mega pares de base (1000000 pb)
mRNA:	RNA mensageiro
mtDNA:	DNA mitocondrial
MYA:	Milhões de anos atrás
N:	Base desconhecida
NCBI:	National Center for Biotechnology Information
NGS:	Sequenciamento de Nova Geração
ORF:	Fase aberta de leitura
pb:	Pares de base
PCG:	Gene codificador de proteína
PCR:	Reação em cadeia da polimerase
RNA:	Ácido ribonucleico
RNA-Seq:	Sequenciamento de mRNA
rRNA:	RNA ribossomal
rrnS:	RNA ribossomal 16S
SBS:	Sequenciamento por síntese
SRA:	Sequence Read Archive
T:	Timina (base pirimídica)
TPA:	Banco de anotação terceirizada do Genbank

tRNA:	RNA transportador
trn-X:	RNA transportador relativo ao aminoácido X
UCE:	Elementos ultra-conservados
WGS:	Sequenciamento de genoma completo

Sumário

1	INTRODUÇÃO	19
1.1	Bases moleculares do Sequenciamento de Nova Geração Illumina	20
1.2	O formato sra e o conceito de “spot”	25
1.3	Montagem e anotação de genomas	27
1.4	Mitogenomas: evolução, aplicações e ubiquidade em datasets	31
1.5	Formigas: relevância e disponibilidade de dados moleculares	34
1.6	A subfamília Pseudomyrmecinae: taxonomia, ecologia e evolução	35
2	OBJETIVOS	41
2.1	Objetivos gerais	41
2.2	Objetivos gerais	41
2.3	Objetivos específicos	41
3	METODOLOGIA	42
3.1	Aquisição de dados	42
3.2	Montagem e anotação do genoma mitocondrial	42
3.3	Análises filogenômicas	44
4	RESULTADOS	46
4.1	Montagem e anotação dos mitogenomas de Pseudomyrmecinae	46
5	DISCUSSÃO	53
I	PREPARAÇÃO DA PESQUISA	64
6	CONTEÚDOS ESPECÍFICOS DO MODELO DE TRABALHO ACADÊMICO	65

6.1	Quadros	65
II	REFERENCIAIS TEÓRICOS	66
7	LOREM IPSUM DOLOR SIT AMET	67
7.1	Aliquam vestibulum fringilla lorem	67
III	RESULTADOS	69
8	LECTUS LOBORTIS CONDIMENTUM	70
8.1	Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae	70
9	NAM SED TELLUS SIT AMET LECTUS URNA ULLAMCORPER TRISTIQUE INTERDUM ELEMENTUM	71
9.1	Pellentesque sit amet pede ac sem eleifend consetetuer . . .	71
10	CONCLUSÃO	72
	REFERÊNCIAS	73
	APÊNDICES	77
	APÊNDICE A – QUISQUE LIBERO JUSTO	78
	APÊNDICE B – NULLAM ELEMENTUM URNA VEL IMPERDIET SODALES ELIT IPSUM PHARETRA LIGULA AC PRETIUM ANTE JUSTO A NULLA CURABITUR TRISTIQUE ARCU EU METUS	79

ANEXOS	80
ANEXO A – MORBI ULTRICES RUTRUM LOREM.	81
ANEXO B – CRAS NON URNA SED FEUGIAT CUM SO- CIIS NATOQUE PENATIBUS ET MAGNIS DIS PARTURIENT MONTES NASCETUR RI- DICULUS MUS	82
ANEXO C – FUSCE FACILISIS LACINIA DUI	83

1 Introdução

Mais de uma década após o advento do sequenciamento de nova geração (NGS, do inglês “Next Generation Sequencing”) (MARGULIES *et al.*, 2005), é evidente que essa tecnologia incrível e madura promoveu um aumento sem precedentes na geração de dados genômicos e uma importante redução dos custos de sequenciamento (MARDIS, 2008; GOODWIN; MCPHERSON; MCCOMBIE, 2016). A fim de reunir e democratizar o acesso a dados genômicos, a International Nucleotide Sequence Database Collaboration (INSDC, <http://www.insdc.org/>) foi estabelecida em 1987. Esse esforço contínuo compreende três centros internacionais: (i) o National Center for Biotechnology Information (NCBI), (ii) European Bioinformatics Institute (EBI) e (iii) DNA Data Bank of Japan (DDBJ) (KARSCH-MIZRACHI; TAKAGI; COCHRANE, 2018; COCHRANE *et al.*, 2006). Como parte dessa notável iniciativa, o banco de dados Sequence Read Archive (SRA) foi criado para hospedar reads e metadados de sequências brutas geradas por projetos NGS (KODAMA; SHUMWAY; LEINONEN, 2012). Disponibilizar dados brutos de sequenciamento é fundamental para a reprodutibilidade experimental (STODDEN; SEILER; MA, 2018), um pilar do esforço científico. Além disso, o SRA tem sido utilizado recorrentemente para apoiar novas pesquisas, tais como: a avaliação de polimorfismos e deleções de base única (BORDBARI *et al.*, 2017), o teste de novos programas de bioinformática (SIMPSON *et al.*, 2009; LANGMEAD; SALZBERG, 2012; BOLGER; LOHSE; USADEL, 2014), para avaliar os impactos de alguns procedimentos comuns sobre os dados, como o trimming (Del Fabbro *et al.*, 2013), dentre outros estudos (KAYAL *et al.*, 2015; BERNSTEIN; DOAN; DEWEY, 2017; LINARD *et al.*, 2018) .

A disponibilidade de dados públicos está crescendo em conjunto com os potenciais usos de tais bancos de dados para a comunidade científica. Em um período de 2 anos (agosto-2015 a agosto-2017), 3000 trilhões de pares de bases (pb) foram adicionados ao SRA, promovendo um crescimento de 233 % do repositório (Karsch-Mizrachi, Takagi Cochrane, 2017)¹. No entanto, o potencial desses dados

¹ Essa versão da dissertação foi desenvolvida apenas para experimentar as funcionalidades do

está longe de ser totalmente explorado, uma vez que os bancos de dados públicos apresentam recursos que poderiam ser usados para abordar diversos tipos de questões biológicas previamente inexploradas. Neste trabalho, nos concentramos em obter genomas mitocondriais completos usando conjuntos de dados (datasets) genômicos disponíveis publicamente.

1.1 Bases moleculares do Sequenciamento de Nova Geração Illumina

Este trabalho se valeu exclusivamente de conjuntos de dados genômicos paired-end gerados pela plataforma Illumina. Segundo o [site da companhia](#), o processo de sequenciamento gerado por essa tecnologia pode ser dividido em 4 etapas principais:

Preparação de biblioteca (Figura 1-A): O DNA do organismo de interesse deve ser clivado aleatoriamente (por processos como sonicação ou uso de nucleases inespecíficas) (Knierim et al., 2011). Após isso, fragmentos são selecionados por tamanho através de gel, se obtendo assim sequências com menor variação de tamanho. A preparação das bibliotecas consiste na modificação e amplificação desses fragmentos para que o sequenciamento seja possível. Em um primeiro passo, oligonucleotídeos específicos (chamados adaptadores) se ligam às duas extremidades do fragmento. Na preparação de uma biblioteca, geralmente são usados dois adaptadores: uma para cada extremidade do fragmento (5' e 3'). A sequência dos adaptadores varia de acordo com o modelo de sequenciador utilizado e contém tanto regiões de ligação a primers, quanto sequências identificadoras (índices) e uma região imobilizadora, que consiste de oligonucleotídeos necessários para a ligação da sequência à flowcell (mais informações na próxima etapa). O tamanho da sequência proveniente do organismo, a qual se encontra entre os adaptadores, é chamado de insert size. Depois da adição dos adaptadores, os fragmentos podem ser

¹LaTeX. Assim sendo, algumas referências que estão no texto - as que não apresentam hyperlink - não serão encontradas na [seção 10](#). O mesmo é válido para a itálização de nomes científicos e termos estrangeiros. Para a versão final dessa dissertação, acesse a cópia presente em meu [github](#) ou a disponível no [portal do Departamento de Bioquímica Médica da UFRJ](#) (ano de defesa: 2019)

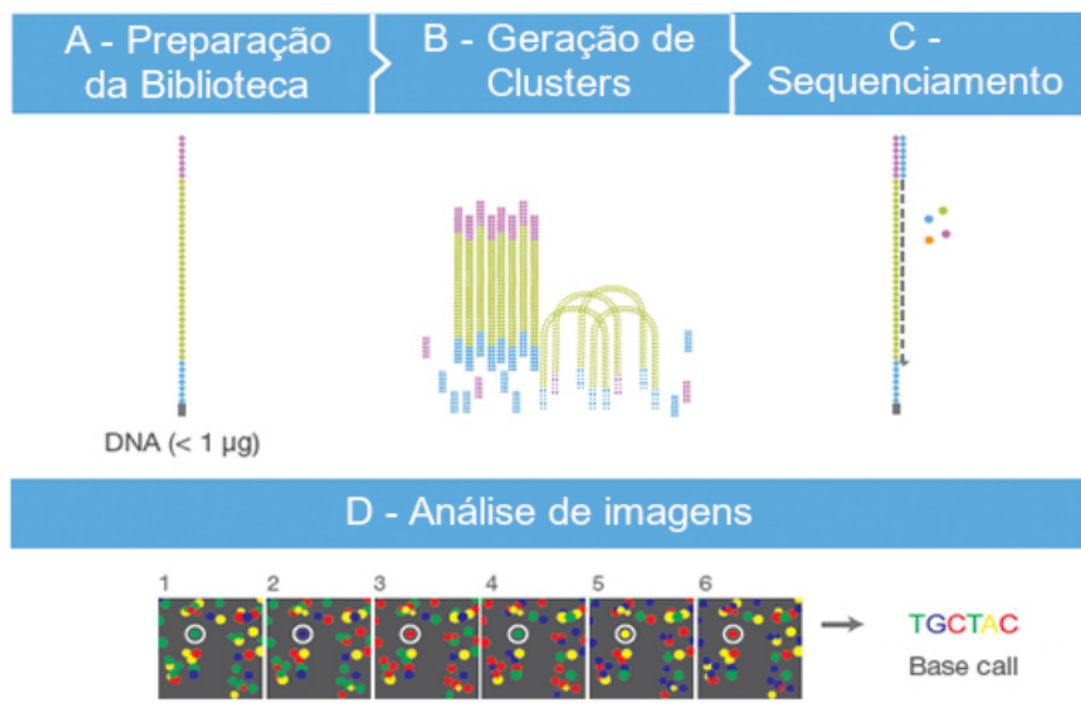


Figura 1 – Quatro principais etapas do sequenciamento Illumina.

A. O DNA obtido a partir do organismo de interesse (seja ele DNA genômico ou cDNA, no caso de RNA-Seq) é modificado de forma a gerar uma biblioteca. **B.** Os fragmentos dessa biblioteca são então ligados à flowcell e amplificados, gerando aglomerações da mesma sequência (clusters). **C.** Uma nova sequência é sintetizada pela adição de bases complementares às sequências dos clusters. Cada uma das quatro bases emite um sinal luminoso de coloração específica que é capturado em imagens. **D.** Essas imagens são analisadas de forma a determinar qual base foi adicionada em cada ciclo de sequenciamento. Adaptado de: <<https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html>>

amplificados pela Reação em Cadeia da Polimerase (PCR, do inglês “Polymerase Chain Reaction”). Essa etapa de amplificação (também chamada de “enriquecimento de biblioteca”) nem sempre é realizada, mas é importante por tornar possível a realização de sequenciamentos mesmo a partir de baixas concentrações de DNA.

Ligação e amplificação dos fragmentos na flowcell (Figura 1-B):

A biblioteca gerada é então dispersa pela flowcell, que consiste em uma placa

de vidro com um número definido de canais (lanes) densamente populados com oligonucleotídeos. Os adaptadores apresentam sequências que são complementares àquelas presentes nesses canais, hibridizando-se a elas e fixando à flowcell os fragmentos de DNA que serão sequenciados. Essa hibridização se dá de forma aleatória e ocorre com apenas uma das extremidades do fragmento de DNA original, de forma que, em um cenário ideal, é esperado que as sequências se liguem e fiquem posicionadas a uma distância considerável uma das outras, muito embora isso dependa de outros fatores, como a concentração da biblioteca.

Como a próxima etapa do sequenciamento depende da captação de sinal luminoso, há a necessidade de que cada sequência da flowcell gere cópias que estejam próximas a ela para a amplificação desse sinal. Primeiramente, são adicionadas polimerases que gerarão o reverso complementar das sequências de DNA originais. Então, a fita original é lavada e permanecem apenas seus complementares na placa, que passam pelo processo de “amplificação em ponte” (Mayer et al., 2011). Nesse processo, o adaptador do topo do fragmento se hibridiza a uma sequência complementar na superfície da flowcell. Isso fornece uma extremidade livre que permite a amplificação de uma nova cópia de DNA, a princípio resultando em uma dupla fita que é posteriormente desnaturada e dá origem a duas sequências complementares. Em um novo ciclo, os adaptadores dessas sequências novamente se hibridizam aleatoriamente com aqueles presentes na placa, gerando de forma clonal outras sequências. Repetido múltiplas vezes, esse processo acaba por amplificar várias cópias de um fragmento a ser sequenciado, que ficam próximos e consequentemente emitem um sinal luminoso mais forte, perceptível pelo sequenciador. A essas aglomerações de cópias de uma mesma sequência se dá o nome de cluster.

A princípio, os fragmentos em questão podem ser tanto da sequência senso (forward) quanto da antissenso (reverse). Entretanto, para evitar a captação de sinais luminosos conflitantes na próxima etapa, cada etapa de sequenciamento deve ser feita exclusivamente com sequências senso ou antissenso. Isso proporciona dois tipos principais de sequenciamento: (i) single-end, no qual as sequências antissenso são removidas e apenas as sequências senso são sequenciadas; e (ii) paired-end, onde após o sequenciamento das sequências senso, ocorre uma nova série de amplificações em ponte e o processo inverso é realizado, removendo os fragmentos forward e

mantendo os reverse. Com isso, ambas as fitas são sequenciadas e, conseqüentemente, as duas extremidades do fragmento. Conhecendo as extremidades da sequência e o tamanho dos fragmentos gerados, podemos estimar a distância entre as reads. A obtenção dessa informação é a maior vantagem do sequenciamento paired sobre o single-end, já que saber a distância entre sequências nos permite, dentre outros, fechar gaps reais e estabelecer a ordem e distância entre dois ou mais contigs durante a montagem. O uso dessa informação nas montagens frequentemente aumenta o tamanho dos contigs/scaffolds obtidos e conseqüentemente gera uma representação mais fidedigna do genoma estudado.

Em trabalhos de sequenciamento de genoma nuclear é comum se usar dados advindos de bibliotecas de diferentes tamanhos médios para se obter uma montagem de maior qualidade (RUBIN et al., 2016; WIRTHLIN et al., 2018). Vale ressaltar que o tamanho de fragmento ou insert size aceito pelo sequenciamento paired-end é limitado (200-800 pb) e para se conseguir ordenar e unir contigs muito afastados geralmente se faz necessário o uso de um terceiro tipo de sequenciamento: o mate-pair, que consegue obter sequências separadas por uma distância muito maior (2000-5000 pb). A única diferença do sequenciamento mate-pair para o paired-end está na preparação de sua biblioteca (Figura 2): enquanto o paired-end usa os fragmentos gerados na fragmentação diretamente na preparação da biblioteca, no sequenciamento mate-pair moléculas de biotina são ligadas covalentemente às extremidades de DNA (biotinilação), o que leva à circularização dos fragmentos. Posteriormente, os fragmentos são clivados na região de circularização, que é então enriquecida. Assim, as extremidades sequenciadas correspondem às sequências que estão separadas por distâncias muito maiores do que aquelas encontradas em paired-end.

Sequenciamento (Figura 1-C): O processo é baseado no “Sequenciamento por Síntese” (Sequencing-by-Synthesis ou SBS) (Mardis et al. 2008), metodologia na qual um primer se liga ao adaptador da extremidade livre na flowcell e uma DNA polimerase sintetiza uma nova fita pela incorporação de nucleotídeos marcados com fluorescência. Os quatro tipos de nucleotídeos (A, G, C, T) são adicionados à flowcell simultaneamente e cada um apresenta fluoróforos que emitem um sinal luminoso de comprimento de onda específico. Em outras palavras,

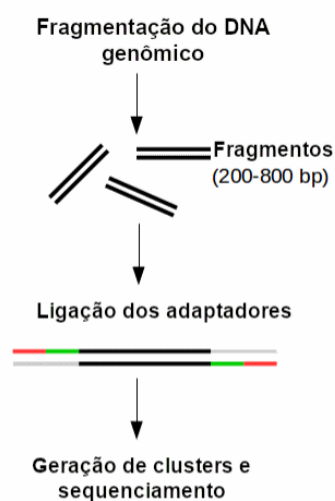
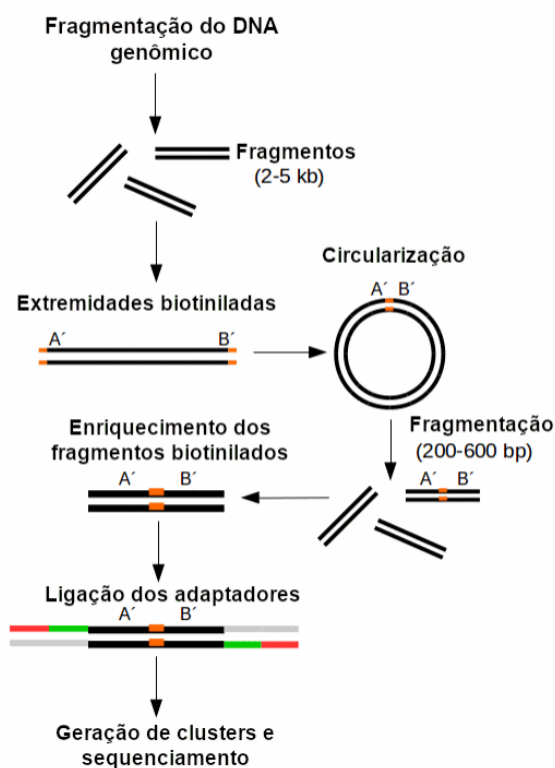
Sequenciamento Paired-End**Sequenciamento Mate-Pair**

Figura 2 – Comparação entre o processo de preparação de bibliotecas *paired-end* e *mate-pair*

Adaptado de: <https://www.ecseq.com/support/ngs/what-is-mate-pair-sequencing-useful-for>

cada nucleotídeo emite luz de uma coloração característica, que é registrada pelo sequenciador e usada posteriormente na identificação da base incorporada. Esses nucleotídeos também são bloqueados de forma reversível em sua extremidade 3'OH, o que permite que apenas um nucleotídeo seja adicionado e identificado por ciclo de síntese. Essa incorporação de um único nucleotídeo por ciclo é necessária para, por exemplo, a determinação correta do comprimento de regiões homopoliméricas.

Após a captação do sinal luminoso, o bloqueador 3'OH é clivado e um novo nucleotídeo é incorporado. Esse processo ocorre simultaneamente para milhões ou

mesmo bilhões de clusters e se repete “n” vezes até gerar uma read de tamanho “n” por cluster (esse valor pode ir de 50 a 300, dependendo do sequenciador utilizado). A paralelização massiva dessa metodologia, apesar de gerar reads consideravelmente curtas, permite a geração de uma grande quantidade de dados por corrida de sequenciamento. Isso implica uma diminuição do custo de sequenciamento por nucleotídeo quando comparada à tecnologia de sequenciamento Sanger (Mardis, 2008; van Dijk et al., 2014; Goodwin, McPherson McCombie, 2016).

Análise de imagens (Figura 1-D): Cada vez que uma base é adicionada durante o sequenciamento, o sequenciador captura imagens com as fluorescências emitidas por uma grande quantidade de clusters. Essas imagens serão analisadas pelo software do sequenciador, que identifica a posição dos sinais luminosos emitidos por meio de coordenadas X e Y das imagens. As posições emissoras de fluorescência (correspondentes aos clusters) são chamadas de spots. O software do sequenciador analisa essa imagem e, ao avaliar o comprimento de onda e a intensidade do sinal luminoso obtido, realiza a identificação da base (base call) para cada nucleotídeo adicionado a um determinado spot. Ao final desse processo obtemos as sequências de DNA com base na leitura dessas imagens, que por conta disso são chamadas de reads ou leituras. Em sequenciamentos paired-end, as duas extremidades de um mesmo fragmento de DNA pertencerão a um mesmo spot.

O algoritmo utilizado pode apresentar maior ou menor grau de confiança em sua identificação, que é convertido em um valor de qualidade associado a cada base. Essa informação é relevante em diversos procedimentos. Por exemplo, regiões de baixa confiabilidade podem ser identificadas e removidas para que etapas subsequentes (como a montagem) sejam realizada apenas com o “gold standard” dos dados gerados, diminuindo a presença de ruído nas análises e consequentemente propiciando resultados com maior suporte.

1.2 O formato sra e o conceito de “spot”

O Sequence Read Archive disponibiliza seus dados primariamente nos formatos fastq e sra. Os arquivos sra geralmente precisam passar por uma conversão para fastq para serem usados na maior parte dos programas de bioinformática.

O programa mais comumente utilizado nessa conversão é o fastq-dump, parte do pacote de programas SRAToolkit, distribuído pelo NCBI. Os arquivos sra apresentam duas grandes vantagens sobre os fastq para trabalhos em larga escala: i) ocupam consideravelmente menos espaço do disco rígido, sendo mais eficientes para o armazenamento de cópias locais, especialmente em se tratando de datasets massivos; e ii) o principal programa utilizado para a conversão de sra para fastq (fastq-dump) permite a manipulação de diversos parâmetros para gerar arquivos fastq modificados sem a necessidade de os baixar novamente. Várias opções podem ser utilizadas para se obter exatamente o tipo de dado desejado: tamanho mínimo das reads (-M), qualidade das reads (-R), remoção de sequências adaptadoras (-W), formatação do cabeçalho das reads (-readids e -define-seq, importantes para que o fastq seja compatível com alguns programas), dentre outros.

Os arquivos sra são divididos em spots, não em reads. Além disso, há tanto uma opção do fastq-dump para imprimir um número fixo de spots (-X) quanto um que permite separar esses spots e armazenar as reads resultantes em um único arquivo (-split-spot) ou em dois arquivos diferentes (-split-files). Embora as documentações oficiais do NCBI não deixem explícito o que é um spot para esses arquivos, a associação com o conceito introduzido no sequenciamento Illumina é muito provável (<https://www.biostars.org/p/12047/>). Conforme supracitado, em um spot são sequenciadas tanto a read forward quanto a reverse relativas às extremidades de um fragmento de DNA. Assim sendo, o mais provável é que o conceito de “spot” em um arquivo sra paired-end seja próximo de algo como “toda a informação gerada por um spot em um sequenciamento”, o que corresponde a dizer que um spot é uma concatenação das reads forward e reverse. Isso é corroborado pelo fato de que, ao se separar os spots, o tamanho das sequências geradas caem pela metade e apresentam os característicos cabeçalhos terminando em “1” ou “2” para indicar os pares de sequências advindas do mesmo fragmento. O dataset paired-end SRR5852657, para o qual mitocôndrias de camundongo (*Mus musculus*) foram isoladas e sequenciadas (<https://www.ncbi.nlm.nih.gov/sra/?term=SRR5852657>), foi utilizado na avaliação da diferença observada entre reads geradas com relação à separação de spots (Figura 3). Esse mesmo dataset foi utilizado para gerar arquivos fastq e montar o genoma mitocondrial do camundongo usando uma referência da

mesma espécie (KY018919.1). A anotação dos mitogenomas obtidos com e sem a separação de spots da [Figura 5](#) evidencia que o uso de spots não partidos acarreta mais erros de montagem.

As sequências concatenadas em um spot, por não corresponderem àquelas encontradas no organismo de estudo, não são recomendadas para realizar montagens. Com isso, concluímos que entender o conceito de spots e utilizar a opção “split-spot” ou “split-files” do fastq-dump é fundamental para se trabalhar com dados públicos resultantes de sequenciamento paired-end.

1.3 Montagem e anotação de genomas

Como o sequenciamento gera sequências curtas, a montagem dessas reads é necessária para se obter sequências maiores ou mesmo genomas completos. O processo de montagem consiste na junção de reads contíguas guiada pela sobreposição dos nucleotídeos presentes em suas extremidades, o que permite gerar sequências maiores (chamadas de contigs). Também há os scaffolds, que consistem em contigs adjacentes cujo afastamento consegue ser estimada pela distância entre as reads geradas por sequenciamentos paired-end ou mate-pair. Como as bases que ligam os contigs não são conhecidas, elas são substituídas por “N”s, que podem ser lidos como “bases desconhecidas” ([página 29](#)).

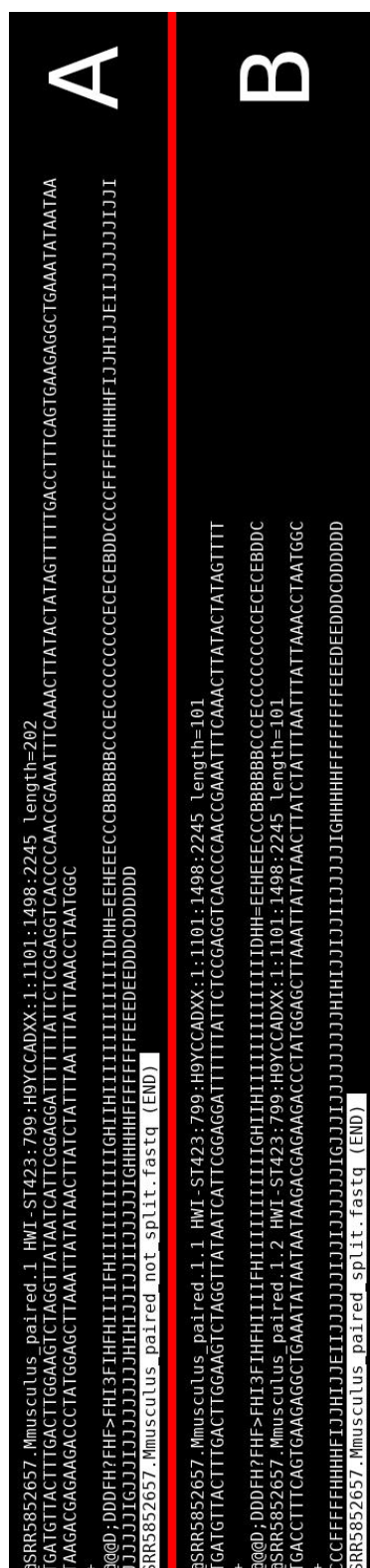


Figura 3 – Conversão de um único spot do formato sra para fastq usando o software fastq-dump

O dataset mitocondrial SRR5852657 foi utilizado para se obter um único spot com cabeçalho adaptado para dados paired-end (opções “-X 1” e “-read-ids”). Em **A**, nenhuma outra opção foi especificada, e o spot corresponde a uma única sequência de 202 nucleotídeos, que corresponde à concatenação das duas reads de 101 pb geradas em **B** (onde a opção “-split-spot” foi utilizada). As reads de **B** também se diferenciam pelo cabeçalho (.1 e .2), indicando que elas vieram do mesmo fragmento sequenciado, ao contrário de **A**.

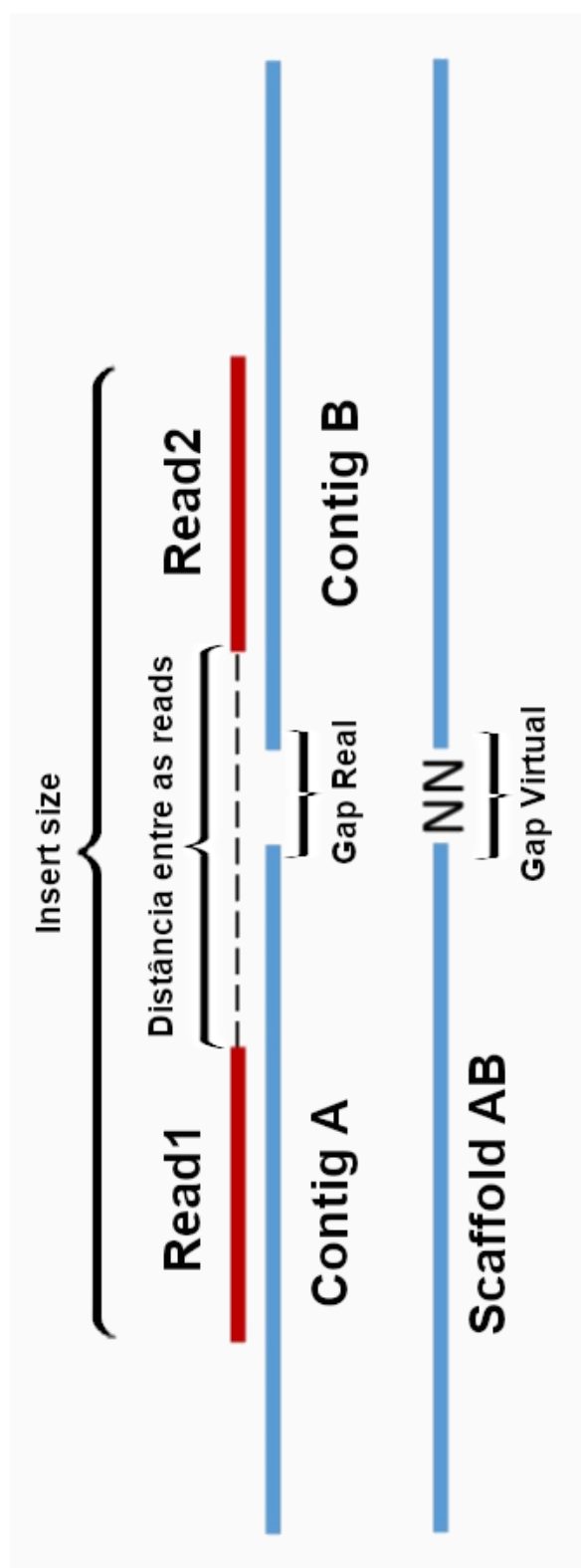


Figura 4 – Uso de dados pareados para o fechamento de gaps

Em dados pareados (paired-end ou mate-pair) é possível estimar a distância entre as reads, já que tanto o tamanho do fragmento sequenciado (Insert size) quanto o das sequência em si são conhecidas. Logo, quando as reads pareadas são mapeadas em contigs diferentes, podemos calcular a distância entre eles e uni-los em uma única sequência (scaffold). No processo de fusão, o que antes era um gap real (no qual tanto o tamanho quanto a sequência entre os contigs é desconhecida) se torna um gap virtual (onde se conhece a extensão do gap, que é preenchida por N's). Adaptado de: <http://madsalbertsen.github.io/multi-metagenome/docs/step10.html>

Há dois tipos principais de montagem (Lesk, 2014):

- (i) a montagem de novo, na qual somente a informação de sobreposição/distância entre as reads é utilizada na montagem de uma sequência (chamada de sequência consenso - [Figura 6-A](#))
- (ii) a montagem por referência, na qual informação exterior (na forma de uma sequência já existente, chamada de referência ou backbone) é utilizada na montagem. As reads são mapeadas a essa referência que idealmente deve pertencer a um organismo evolutivamente próximo do objeto de estudo, e a sobreposição entre as sequências é então utilizada para se obter o consenso ([Figura 6-B](#)).

Entretanto, geralmente priorizamos realizar montagens de novo pois elas não incluem vieses associados a referências, que levam a erros de montagem quando, por exemplo, a referência apresenta ordem gênica (sintenia) diferente daquela encontrada no genoma que está sendo montado. Além disso, ao se trabalhar com organismos não-modelo, referências próximas ao objeto de estudo comumente não estão disponíveis.

Múltiplos algoritmos foram desenvolvidos para a montagem de genomas, mas os mais amplamente utilizados são aqueles que se baseiam nos Grafos de Bruijn (Compeau et al., 2011; Lesk, 2014). Nas implementações desse algoritmo, uma subsequências das reads de tamanho k (chamadas de k -mers) são os vértices e arestas direcionadas são traçadas entre os k -mers nos quais há uma sobreposição de $k-1$ nucleotídeos, visitando cada nó apenas uma vez ([Figura 7](#)). Ao ser resolvido, o grafo nos dá um caminho entre os nós que indica a sucessão de sequências contíguas sobrepostas, correspondente à sequência montada. O valor de k -mer é geralmente estabelecido por aquele que irá montar o genoma, e a escolha do k -mer mais adequado é pautado pelo balanço entre especificidade e sensibilidade. Por exemplo, ao se escolher k -mers longos, se torna mais fácil estabelecer a direção correta do grafo em uma região repetitiva de DNA ou na ocorrência de erros de sequenciamento (alta especificidade), mas o número de reads sobrepostas acaba sendo muito reduzido (baixa sensibilidade), o que pode levar a uma montagem muito

fragmentada. Por outro lado, k-mers curtos geralmente não conseguem resolver corretamente regiões repetitivas do genoma, mas conseguem recrutar um número consideravelmente maior de reads para a montagem.

Múltiplas sequências podem se sobrepor em uma mesma região, e o número de reads que se sobrepõem confirmando uma determinada base é denominado cobertura (Verli, 2014). Esse conceito pode ser estendido para o cálculo da cobertura média do genoma inteiro que consiste na divisão do total de bases sobrepostas (ou seja, número de reads multiplicado pelo tamanho das mesmas) pelo total de bases da montagem. De forma geral, a cobertura está associada à qualidade da montagem, já que regiões de baixa cobertura apresentam menos dados confirmando sua sequência e consequentemente estão mais sujeitas a estarem incorretas.

A montagem, por melhor que seja, contempla apenas a sequência de nucleotídeos encontrada no objeto de estudo, o que geralmente é insuficiente. Para complementar essa informação e realizar análises adicionais que revelem novos aspectos sobre a biologia do organismo estudado, elementos relevantes dessa sequência nucleotídica (como genes codificadores de proteínas, tRNAs, rRNAs e afins) devem ser identificados e nomeados na montagem. A esse processo de adição de informação funcional às sequências montadas se dá o nome de anotação (Verli, 2014), a qual geralmente é feita em duas etapas: primeiramente, é realizada a anotação automática, que se vale de programas que, geralmente por meio do alinhamento de sequências contra bancos de dados de genes conhecidos (Wyman et al., 2004; Bernt et al., 2013), identificam a localização de vários ou mesmo de todos os genes presentes na sequência fornecida. Entretanto, muitas vezes o estabelecimento dos limites gênicos não é tão preciso, fazendo-se necessária uma segunda etapa de curadoria manual para confirmar ou refinar os resultados obtidos pela anotação automática.

1.4 Mitogenomas: evolução, aplicações e ubiquidade em datasets

Em animais, a mitocôndria é uma organela de origem materna - salvo raras exceções, como em algumas espécies de bivalves (Zouros et al., 1994; Theologidis et

al., 2008) - na qual ocorre o processo de fosforilação oxidativa, indispensável para a obtenção primária de energia em eucariotos. Essa organela está envolvida não só no metabolismo energético (BRAND, 1997), como também na apoptose (Wang, 2001), diferenciação celular (WANET et al., 2015) e várias doenças (CHAN, 2006). Apesar da gênese dessa organela ser tópico de discussão ainda hoje, classicamente sua origem comumente é explicada por meio da teoria endossimbiótica, a qual dita que a mitocôndria evoluiu a partir de uma bactéria de vida livre fagocitada e incorporada por uma célula hospedeira (GRAY, 2017). Algumas características mitocondriais corroboram essa hipótese, como a presença de dupla membrana e ribossomos similares aos encontrados em procariotos, mas nenhuma delas é tão contundente quanto a presença de DNA nessas organelas (GRAY; BURGER; LANG, 1999; KUTSCHERA; NIKLAS, 2005). O genoma mitocondrial (também chamado de mitogenoma) também apresenta similaridades com um genoma procarioto típico: possui arranjo circular e, no caso do mitogenoma animal, cada fita é transcrita como um único mRNA policistrônico que é então clivado, dando origem a todos os elementos funcionais desse mitogenoma (BOORE, 1999).

O conteúdo gênico mitocondrial varia consideravelmente em diferentes linhagens de eucariotos, o que acredita-se ocorrer em grande parte devido à transferência de vários genes mitocondriais para o núcleo durante a evolução da organela (ADAMS; PALMER, 2003). Em mitogenomas mitocondrias animais tipicamente são encontrados 38 genes (ou features, como são chamados no processo de anotação): 13 genes codificadores de proteínas (Protein Coding Genes ou PCG's); 22 RNAs transportadores e 2 RNAs ribossomais (WOLSTENHOLME, 1992; BOORE, 1999).

Devido aos seus tamanhos pequenos (aproximadamente 16 kpb em animais), alto grau de conservação no conteúdo e ausência de íntrons, os mitogenomas são os cromossomos mais comumente seqüenciados, em especial para metazoários (SMITH, 2016). Os genomas mitocondriais são mal amostrados para muitos táxons e, portanto, nosso conhecimento atual sobre a biologia evolutiva de muitos clados poderia ser incrementado pelo uso de dados públicos. Sendo primariamente de herança materna e não recombinantes, tais seqüências são frequentemente usadas para estudar biologia evolutiva (FINSTERMEIER et al., 2013; KRZEMIŃSKA et al., 2018), genética populacional (Pečnerová et al., 2017; Kılınç et al., 2018),

filogeografia (Chang et al., 2017; Fields et al., 2018), sistemática (Lin et al., 2017; Crainey et al., 2018) e conservação (Moritz, 1994; Rubinoff, 2006; Rosel et al., 2017) de vários clados (Avise, 1994), sendo especialmente convenientes para estudos com táxons subamostrados (Gotzek, Clarke Shoemaker, 2010; Duan, Peng Qian, 2016) e organismos não-modelo (Prosdocimi et al., 2012; Tilak et al., 2014; Plese et al., 2018).

Experimentos de sequenciamento de genoma completo (Whole Genome Sequencing or WGS) e projetos de sequenciamento parcial do genoma normalmente produzem reads mitocondriais suficientes para permitir a montagem de mitogenomas completos (Prosdocimi et al., 2012; Smith, 2015). Esses pequenos genomas organelares podem ser frequentemente montados em alta cobertura devido ao alto número de cópias dessa organela presentes nas células (Smith, 2015). Além disso, estudos anteriores indicam que é possível recuperar sequências mitocondriais completas e/ou quase completas a partir de dados de RNA-Seq (Tian Smith, 2016; Rauch et al., 2017, Plese et al., 2018) e dados de estratégias de amplificação que se valem do enriquecimento de regiões específicas do genoma para um sequenciamento direcionado a alguma região de interesse, como o exoma (Picardi Pesole, 2012 ; Guo et al., 2013; Samuels et al., 2013) ou os elementos ultra-conservados (Ultra Conserved Elements ou UCE) (Do Amaral et al., 2015; MILLER et al., 2016). Apesar do enriquecimento das sequências de interesse, outras regiões são amplificadas aleatoriamente, incluindo as mitocondriais. Logo, o uso de datasets de sequenciamento direcionado para obter mitogenomas completos só é possível devido ao fato dessa metodologia não ser completamente eficiente na amplificação dos sítios desejados. Outra estratégia que já foi realizada com sucesso é montagem de numerosos mitogenomas completos e/ou grandes contigs mitocondriais a partir do sequenciamento de amostras contendo várias espécies (Timmermans et al., 2015; Linard et al., 2018). Essa abordagem é denominada 'mito-metagenômica' (Tang et al., 2014) ou 'metagenômica mitocondrial' (MMG) (Crampton-Platt et al., 2015). Outros trabalhos utilizaram com sucesso dados públicos para montar sequências mitocondriais (Diroma et al., 2014; Kayal et al., 2015; Linard et al., 2018) demonstrando o potencial dos bancos de dados públicos para estudos mitogenômicos. Entretanto, ainda existem várias espécies sem mitogenoma completo disponível

que apresentam um grande número de dados genômicos na base de dados do SRA.

1.5 Formigas: relevância e disponibilidade de dados moleculares

Um exemplo de amostragem deficiente de mitogenomas ocorre na família Formicidae (clado correspondente às formigas). Apesar de ser um grupo onipresente, ecologicamente dominante e hiper-diversificado (Hölldobler Wilson, 1990; Bolton, 1994), com mais de 13.000 espécies descritas (Bolton, 2012), registros de sequências mitocondriais completas são restritos a apenas 15 espécies de formigas no GenBank. Os insetos pertencentes a essa família não só constituem uma parcela significativa da biomassa nos ambientes onde ocorrem - Hölldobler Wilson (1990) estimam que eles podem chegar a constituir mais de 10 % da biomassa faunal -, como também são considerados “engenheiros ecossistêmicos” que influenciam as propriedades físicas, químicas e biológicas do solo (Jones et al, 1994; Folgarait, 1998; Lobry de Bruin, 1999). Além disso, também impactam interações multitróficas (Sanders Veen, 2011) e contribuem para a estabilidade ambiental e subsequente manutenção da prestação de serviços ecossistêmicos como a regulação climática, captura de carbono e purificação de água e ar (Sanford et al., 2009).

As formigas também apresentam grande potencial como bioindicadores em áreas de recuperação ambiental, visto que: (i) elas são abundantes e ubíquas, ocorrendo tanto em habitats intactos quanto em áreas perturbadas (Majer, 1983; Hölldobler Wilson, 1990; Hoffman et al, 2000); (ii) sua coleta é relativamente simples (Majer, 1983; Lopes Vasconcelos, 2008); e (iii) são muito sensíveis a variações ambientais, apresentando respostas claras a essas (Majer, 1983; Hoffman et al., 2000). Estudos usando bioindicadores para analisar distúrbios ambientais comumente são realizados através da comparação da riqueza de espécies em regiões perturbadas e não-perturbadas (Pearson Carroll, 1998; Hoffman et al., 2000). Portanto, a eficácia do uso desses organismos como bioindicadores está diretamente associada à precisão da identificação das espécies. Por último, é necessário ressaltar que as formigas também podem participar ativamente do processo de regeneração de áreas degradadas (Lobry de Bruin, 1999; Gallegos et al., 2014).

Várias formigas têm sua notoriedade e importância indissociavelmente ligadas às suas interações com plantas, em especial àquelas que detêm status de pragas agrícolas. De fato, há espécies que são pragas de culturas economicamente importantes, como as formigas cortadeiras (popularmente conhecidas como saúvas), que cortam folhas para cultivar fungos (que constituem a fonte de alimento dessas formigas) e geram prejuízos enormes em várias monoculturas, em especial as de eucalipto (DellaLucia et al., 2014). Entretanto, mesmo as poucas espécies de saúvas que causam dano econômico também impactam positivamente seu ambiente - reduzindo o tempo de ciclagem de nutrientes e aumentando a aeração do solo, por exemplo - o que coloca em xeque sua classificação como “praga” (Fowler et al., 1989; Jones, 1994). Também há espécies de formigas que são utilizadas como agentes de controle biológico no manejo de pragas (Way Khoo, 1992; Philpott Armbrrecht, 2006) e a maioria das relações entre formigas e plantas são na verdade mutualísticas (Cannicci et al., 2008), não predatórias. Logo, informações sobre esse grupo é relevante no estudo e preservação desses e de seus ecossistemas.

Evidência molecular, em especial dados de sequenciamento de UCE, já foi usada para estudar a filogenia das formigas nos últimos anos (Blaimer et al., 2015; Ward Branstetter, 2017; Branstetter et al., 2017). Entretanto, são raras as tentativas que visam recuperar sequências mitocondriais com base nesses dados gerados para outros propósitos (Ströher et al., 2017) e usar essas informações para entender melhor as relações evolutivas para o clado. Estudos com esse escopo também adicionam evidências moleculares úteis à identificação de espécies, o que no caso das formigas pode potencializar seu uso como bioindicadoras.

1.6 A subfamília Pseudomyrmecinae: taxonomia, ecologia e evolução

Um grupo particular de formigas que sofre de má amostragem de mitogenômica é a subfamília Pseudomyrmecinae, que contém três gêneros: (i) Pseudomyrmex, encontrado no Novo Mundo e que possui ≈ 137 espécies, a maioria das quais pode ser classificada em um dos dez grupos de espécies, delimitados com base em caracteres morfológicos (Ward 1989, 1993, 1999, 2017); (ii) Tetraponera, com ≈ 93 espécies

e de distribuição paleotropical; e (iii) *Myrcidris*, gênero sul-americano que possui apenas uma espécie descrita, *Myrcidris epicharis* (Ward Downie, 2005; Bolton, 2012; Ward, 2017).

De acordo com Janzen (1966) e Ward (1991), existem dois grupos ecológicos conhecidos de Pseudomyrmecinae:

Grupo 1 Composto de espécies arbóreas generalistas. Nidificam em galhos mortos de vários tipos de plantas e são geralmente passivas em relação a objetos externos

Grupo 2 Formigas especializadas em colonização de plantas. habitantes obrigatórias de cavidades ocas em tecidos vivos de plantas. Essas cavidades são estruturas vegetais especializadas (chamadas de domatia) que provêm abrigo e proteção às formigas, que são freqüentemente agressivas em relação a outros insetos ou plantas.

As formigas do Grupo 2 fornecem proteção contra herbivoria e competição para sua planta hospedeira em um caso típico de mutualismo (JANZEN, 1966; WARD, 1991). Estudos anteriores usando dados morfológicos e moleculares (Ward, 1991; Ward Downie, 2005) sugerem que esse tipo de mutualismo das espécies do Grupo 2 evoluiu independentemente pelo menos 12 vezes na subfamília Pseudomyrmecinae. Por exemplo, o trabalho de Chomicki e colaboradores (2015) chama a atenção para o fato de formigas do gênero *Pseudomyrmex* desenvolverem comportamentos similares por convergência, apesar de evoluírem com diferentes hospedeiros vegetais. Esse mesmo trabalho elenca as plantas comumente associadas a essas formigas: (i) leguminosas (Fabaceae) dos gêneros *Vachellia* (anteriormente parte do gênero *Acacia*, cujo mutualismo com as *Pseudomyrmex* levou essas formigas a serem conhecidas como “acacia-ants”), *Platymiscium* e *Tachigali*; e (ii) as poligonáceas dos gêneros *Triplaris* e *Ruprechtia*.

Casos de evolução convergente são freqüentemente caracterizados usando abordagens filogenéticas (Ward Branstetter, 2017). Análises evolutivas de sequências mitocondriais geralmente permitem um melhor entendimento sobre a história de clados superiores ao nível de ordem (Mao, Gibson Dowton, 2015) e família (Miya et al., 2003; Kayal et al. 2015). A mitogenômica já foi usada para resolver relações

evolutivas em clados superiores de insetos (subfilo Hexapoda) (Mao, Gibson Dowton, 2015; Bourguignon et al., 2016). Assim sendo, a subfamília Pseudomyrmecinae é uma ótima candidata a ser estudada usando filogenômica mitocondrial, já que é um clado superior de Formicidae e apresenta diversos casos de coevolução.

Diversos estudos moleculares foram realizados no gênero *Pseudomyrmex*. Estes geralmente abordam questões coevolutivas, como o impacto de associações mutualistas na taxa de evolução do genoma (Rubin Moreau, 2016), ou a caracterização de associações entre formigas e plantas através do estudo de relações filogenéticas e biogeografia (Chomicki, Ward Renner, 2015; Ward Branstetter, 2017). No entanto, análises completas de mitogenomas nunca foram realizadas para Pseudomyrmecinae devido à ausência de sequências mitocondrais completas para o clado. Na atual abordagem de “mitogenômica *no-budget*” (definida aqui como o uso de dados públicos de NGS para montar grandes sequências mitocondriais indisponíveis em bancos de dados), usamos dados genômicos publicamente disponíveis gerados por outros trabalhos (Tabela 1) para montar e analisar a sequência mitocondrial completa para 14 representantes da subfamília Pseudomyrmecinae: 12 espécies do gênero *Pseudomyrmex* e duas espécies de *Tetraponera*. O tempo de divergência estimado entre *Pseudomyrmex* e *Tetraponera* é estimado em ≈ 95.8 MYA (Million Years Ago), de acordo com Gómez-Acevedo et al. (2010), apoiando um evento de vicariância durante a separação da América do Sul da África. Assim sendo, este trabalho também testa a eficiência da abordagem mitogenômica em resolver relações filogenéticas entre clados que divergiram em um passado relativamente distante.

Apresentamos as primeiras sequências mitocondriais completas para esta subfamília e realizamos análises evolutivas delas em conjunto com todos os outros mitogenomas de Formicidae disponíveis, tentando entender melhor as relações de grupos-irmãos dentro deste clado altamente diverso. A dissertação atual apresenta novos genomas mitocondriais para espécies de formigas que cobrem cinco dos 10 grupos de espécies de *Pseudomyrmex* e quase duplica o número de genomas mitocondriais disponíveis para formigas, aumentando este número de 15 para 29.

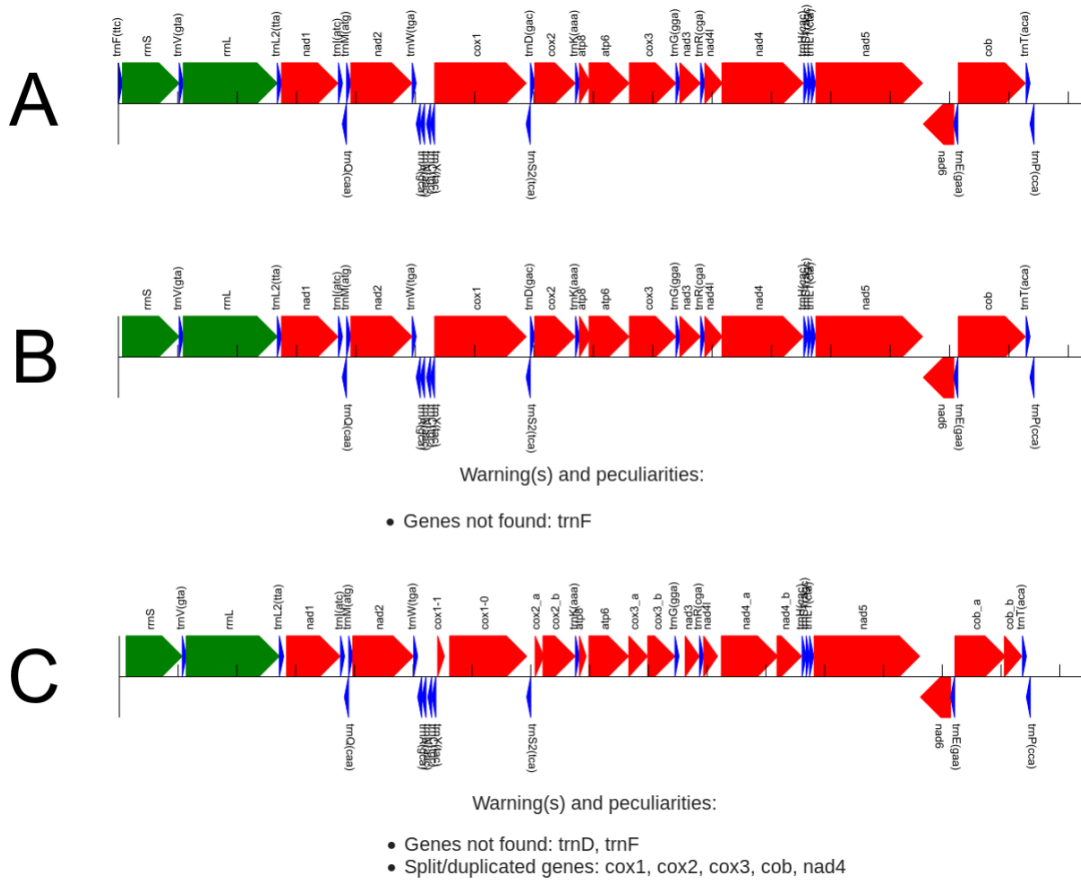
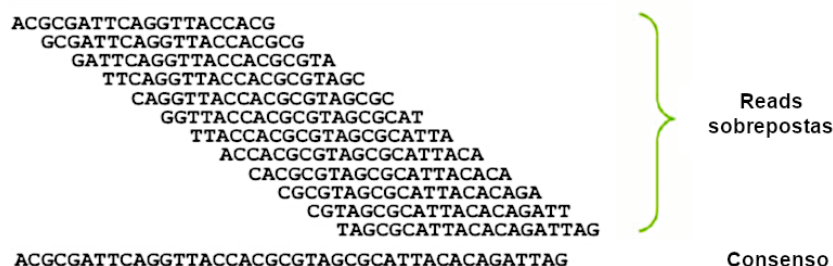


Figura 5 – Anotação de montagens mitocondriais com e sem a separação de spots

Por meio do software MIRA (Chevreux et al., 1999), realizamos montagens mitocondriais por referência para averiguar o impacto da separação dos spots nas montagens. A referência utilizada e os mitogenomas gerados foram anotados utilizando o MITOS Web Server (BERNT et al., 2013). Podemos observar que a montagem **A** (obtida através do Genbank – Acession number [KY018919.1](#)) não apresenta nenhum problema em sua anotação. As duas outras montagens foram realizadas utilizando a sequência de **A** como backbone. Na montagem **B** (obtida usando spots partidos), o tRNA da fenilalanina (trn-F) não pôde ser identificado. Por último, temos que a anotação da montagem **C** (spots inteiros) apresenta mais problemas: além da ausência de trn-D e trn-F, há vários genes duplicados. De maneira geral, a anotação de **B** é muito mais similar à da referência, o que sugere que a separação de spots é importante para a obtenção de sequências de maior qualidade.

A) MONTAGEM DE NOVO



B) MONTAGEM POR REFERÊNCIA

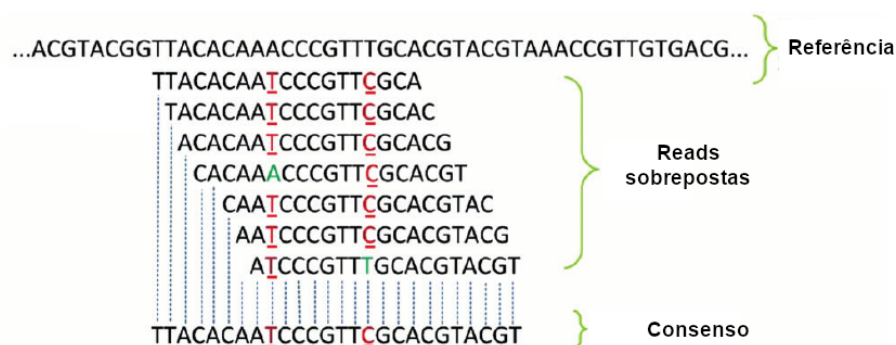


Figura 6 – Esquema representativo das montagens denovo e por referência

A. A montagem de novo se vale exclusivamente da informação contida no sequenciamento. Adaptado de <<https://contig.files.wordpress.com/2010/02/alignment1.jpg>>. **B.** A montagem por referência utiliza informação externa na forma de uma sequência previamente sequenciada. Adaptado de Wajid Serpedin (2014). Ambas geram a sequência final com base no consenso das sobreposições.

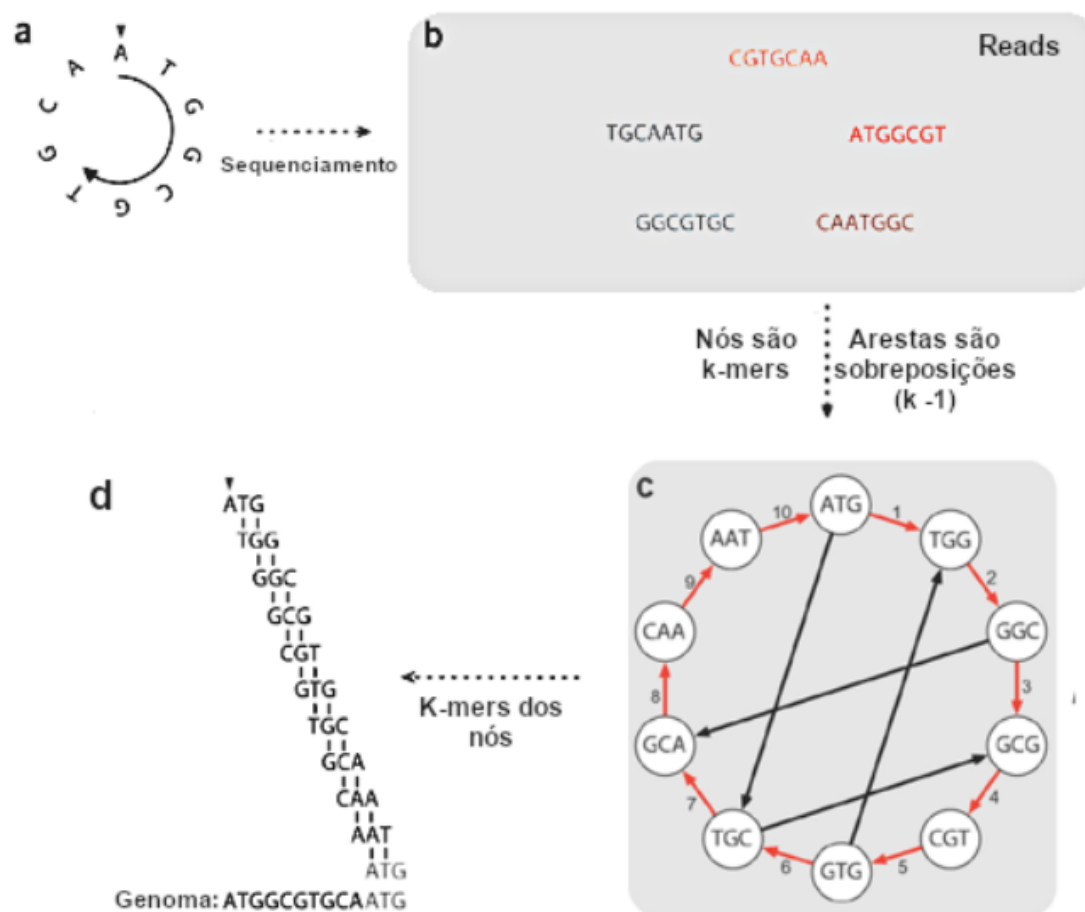


Figura 7 – Grafos de Bruijn aplicados à montagem de genomas

A. Um pequeno genomas circular é sequenciado. **B.** As reads são geradas (cada uma com sete nucleotídeos). **C.** Com $k=3$, um grafo de Bruijn é construído, onde os nós são sequências de três nucleotídeos (3-mers) e as arestas são traçadas ao se encontrar sobreposições de dois ($k-1$) nucleotídeos entre as sequências. Repare que o baixo valor de k-mer utilizado facilita a sobreposição entre vários nós, o que faz com que múltiplos caminhos sejam possíveis nesse grafo (arestas pretas) e dificultam a montagem. **D.** Ao final, o caminho que visita cada nó apenas uma vez (arestas vermelhas) nos dá a sequência montada. Adaptado de Compeau et al. (2011).

2 Objetivos

2.1 Objetivos gerais

2.2 Objetivos gerais

- Obter genomas mitocondriais completos usando dados públicos presentes no Sequence Read Archive
- Estudar a subfamília de formigas Pseudomyrmecinae por meio da análise de sequências mitocondriais

2.3 Objetivos específicos

- Montar, anotar e disponibilizar mitogenomas da subfamília Pseudomyrmecinae
- Realizar estudos de genômica comparativa que englobem os mitogenomas completos de formigas presentes no Genbank e as sequências geradas nessa dissertação
- Construir árvores filogenômicas para estudar a evolução da família Formicidae, com ênfase na subfamília Pseudomyrmecinae

3 Metodologia

3.1 Aquisição de dados

Quatorze datasets paired-end obtidos por sequenciamento Illumina foram baixados do EMBL Nucleotide Archive (<<https://www.ebi.ac.uk/ena>>) no formato de arquivo SRA (consulte a Tabela 1). Esses conjuntos de dados contêm tanto reads mitocondriais quanto nucleares que foram convertidas para FASTQ usando o software fastq-dump (com parâmetros `-readids` e `-split-files`) que integra o pacote SRAToolkit.2.8.2.

3.2 Montagem e anotação do genoma mitocondrial

Cada dataset completo foi usado como entrada para a montagem de novo usando NOVOPlasty2.6.3 (Dierckxsens et al., 2016) com os valores padrão dos parâmetros. Já que NOVOPlasty foi nosso montador principal e Dierckxsens, Mardulyn e Smits (2016) recomendam o uso de dados não trimados nesse software, optamos por usar os datasets brutos como entrada para a montagem. A única exceção foi o dataset de *Tetraponera rufonigra*, que teve de ser ajustado com o software Trimmomatic v.0.36 (Bolger, Lohse & Usadel, 2014) para produzir sequências com o mesmo comprimento em pares de base. Este controle dos dados foi realizado ao ajustar o parâmetro MINLEN do Trimmomatic para 125, que é o comprimento das maiores reads encontradas no dataset. Com isso, descartamos sequências menores e mantemos apenas aquelas de tamanho máximo para serem usadas como entrada na montagem inicial. Montagens NOVOPlasty precisam de uma sequência (denominada seed ou semente) que é utilizada para identificar uma read mitocondrial do dataset, a qual por sua vez será usada para iniciar a montagem (Dierckxsens et al., 2016). As seeds foram selecionadas utilizando sequências de COX1 (Citocromo Oxidase I) da mesma espécie (quando disponíveis) ou utilizando regiões de COX1 de espécies proximoamente relacionadas. As montagens de mitogenoma preliminares realizadas pelo NOVOPlasty foram utilizadas como

referência para uma segunda etapa de montagem utilizando o software MIRA v.4.0.2 com parâmetros padrão (Chevreux et al., 1999). NOVOPlasty não gera um arquivo de alinhamento mostrando as sequências mapeadas à montagem, então MIRA foi utilizado para mapear reads brutas à montagem preliminar e permitir a análise de cobertura da sequência mitocondrial consenso nos próximos passos. Quando a primeira montagem não gerou o mitogenoma completo, nós usamos o MITObim v.1.9 (Hahn et al., 2013) sem alterar seus parâmetros. Este programa realiza montagens MIRA sucessivas para estender o(s) contig(s) mitocondrial(ais) e fechar pequenas lacunas da montagem, gerando a versão final e circularizada do genoma mitocondrial.

O software Tablet versão 1.17.08.17 (Milne et al., 2012) foi usado com valores paramétricos padrão para verificar a cobertura de reads e a circularização dos mitogenomas completos. O processo de anotação automática foi realizado usando MITOSWebServer (Bernt et al., 2013) sem a alteração de parâmetros. Em seguida foi realizada uma etapa de curadoria manual com o software Artemis v.17.0.1 (Carver et al., 2012) usando a tabela de código genético número cinco (correspondente à mitocôndria dos invertebrados) para identificar os limites das fases abertas de leitura (Open Reading Frames ou ORF's).

Já que o genoma mitocondrial dá origem a um grande mRNA policistrônico que então é clivado (Boore, 1999), sobreposições gênicas poderiam incorrer na formação de proteínas, rRNAs ou tRNAs não funcionais. É então razoável pensarmos que mitogenomas nos quais as features não se sobrepõem são energeticamente mais econômicas para a célula e, conseqüentemente, foram selecionados ao longo da evolução. Partindo dessa premissa, tentamos ao máximo evitar sobreposições durante a anotação das sequências mitocondriais. Assim sendo, os limites gênicos dos tRNAs e rRNAs foram mantidos de acordo com os resultados do MITOS Web Server, salvo quando encontrada sobreposição entre duas features (gene codificador de proteína - PCG, tRNA ou rRNA) na mesma fita. Nesse caso, os nucleotídeos sobrepostos foram retirados de uma das features para que a sobreposição fosse completamente removida. O D-loop não foi explicitamente anotado, já que seu caráter hipervariável e de baixa complexidade (Moritz, 1994; Vanecek et al., 2004; Zhang & Hewitt, 1997) torna difícil estabelecer limites precisos para essa região,

em especial quando não se tem uma referência próxima. Entretanto, com base na análise comparativa de sintenia em Formicidae e no fato do D-loop ser geralmente a maior região intergênica do genoma mitocondrial (Liu et al., 2015; Zhang et al., 2016; Huang et al., 2017), identificamos que essa região variável se encontra entre o *rrnS* e *trn-M*.

Para os PCGs, em vários casos se fez necessário expandir a anotação fornecida pelo MITOS Web Server de forma a englobar a maior ORF que não apresente sobreposição com outras features na mesma fita. Então, essa ORF foi utilizada como entrada na versão online do BLASTp (Altschul et al., 1997), sendo alinhada contra o banco de sequências pertencentes a família Formicidae do GenBank. As informações obtidas por esse alinhamento contra sequências mitocondriais de outras formigas nos permitiu considerar a conservação de sequências entre as espécies e determinar o tamanho mais provável da proteína, refinando a anotação. Seguindo esse procedimento, nós alcançamos uma decisão racional, com base em genômica comparativa, sobre os limites gênicos. O conteúdo de AT para (i) o genoma mitocondrial completo; e (ii) a região intergênica que contém o D-loop foram calculados usando o programa online OligoCalc (Kibbe, 2007) com os valores padrão para seus parâmetros.

3.3 Análises filogenômicas

As relações filogenéticas de Formicidae foram reconstruídas usando (i) os 14 mitogenomas completos por nós produzidos juntamente com (ii) todos os outros 15 genomas mitocondriais completos atualmente disponíveis para o clado; e (iii) dois mitogenomas de abelhas (família Apidae) utilizados como grupos externos. Duas árvores filogenéticas foram construídas usando (i) toda a sequência mitocondrial e (ii) o conjunto de genes concatenados de todos os 13 genes codificadores de proteínas (PCGs). Para o primeiro, editamos manualmente as sequências para iniciar no gene COX1 quando necessário e alinhamos os mitogenomas inteiros usando o software ClustalW v.2.1 usando os parâmetros padrão (Thompson, Gibson & Higgins, 2003). Para o segundo, alinhamos e concatenamos os nucleotídeos de todos os PCGs usando o programa Phylomito (<<https://github.com/igorrcoستا/phylomito>>) sem

alteração de seus parâmetros. Modeltest (Posada & Crandall, 1998) foi executado através do software MEGA7 (Kumar, Stecher & Tamura, 2016) para os dois conjuntos de dados e identificou o modelo GTR + G + I como o modelo de substituição de nucleotídeos que melhor explica a variação das sequências. As sequências alinhadas foram usadas como entrada para uma análise de Máxima Verossimilhança (Maximum Likelihood ou ML) usando o MEGA7. A reamostragem foi realizada por bootstrap usando 1000 réplicas. O software BRIG (Blast Ring Image Generator) v.0.95 (Alikhan et al. 2011) foi utilizado com valores paramétricos padrão para comparar e visualizar todos os mitogenomas de Pseudomyrmecinae produzidos aqui.

4 Resultados

4.1 Montagem e anotação dos mitogenomas de *Pseudomyrmecinae*

Os 14 conjuntos de dados genômicos usados para montar o mitogenoma completo das formigas pertencentes à subfamília *Pseudomyrmecinae* foram baixados do banco de dados SRA (Tabela 1). Dois tipos de datasets diferentes foram usados: (i) Sequenciamento do Genoma Completo (WGS), que frequentemente continha uma quantidade maior de dados de sequenciamento, totalizando 212,7 Giga pares de base (Gpb) para seis espécies (de acordo com a informação fornecida pelo SRA); uma média de 35,45 Gpb por espécie (Rubin & Moreau, 2016); e (ii) experimentos de UCE, para os quais realizamos o download de 5,94 Gpb para oito espécies; uma média de 742,5 Mpb por espécie (Branstetter et al., 2017; Ward & Branstetter, 2017).

O dataset completo baixado para cada espécie foi usado como entrada para uma montagem de novo usando o montador NOVOPlasty. Após essa primeira etapa de montagem do genoma, usamos um subconjunto contendo dois ou quatro milhões de reads de sequenciamento como entrada para uma segunda etapa de montagem do genoma usando o software MIRA. Este procedimento foi realizado para mapear as reads na montagem preliminar e melhorar a qualidade do mitogenoma. Para alguns mitogenomas, o MIRA não conseguiu produzir o genoma mitocondrial completo e circularizado. Nesse caso, uma terceira etapa de montagem foi necessária, na qual o maior contig gerado pelo MIRA foi usado como backbone para concluir a montagem usando o MITObim (Tabela 2). Esta metodologia foi capaz de montar a mitocôndria completa de todas as espécies de *Pseudomyrmecinae* analisadas, com exceção da *T. aethiops*, para a qual tivemos que usar o dataset completo como entrada para MIRA e MITObim ao invés de filtrar o subconjunto de reads na segunda etapa. O uso de múltiplas estratégias para montar as sequências mitocondriais completas era esperado, já que dados de NGS são variáveis entre

diferente espécies e corridas de sequenciamento. Além disso, os datasets usados aqui vieram de trabalhos com abordagens experimentais diferentes, o que provavelmente potencializou a variabilidade de conjuntos de dados já muito díspares entre si. Os 14 genomas mitocondriais construídos aqui foram verificados quanto à circularidade e confirmaram apresentar, como esperado para metazoários, 13 genes codificadores de proteínas, 22 tRNAs, dois rRNAs e uma região de controle (Wolstenholme, 1992; Boore, 1999). A anotação do genoma para todos os mitogenomas completos é apresentada na Tabela S1. Todos os genomas mitocondriais produzidos aqui foram submetidos ao GenBank sob o banco de dados de anotação terceirizada (Third Party Annotation ou TPA) (COCHRANE et al., 2006) que forneceu números de acesso para cada genoma mitocondrial, permitindo a visualização e download das sequências (Tabela 1).

De acordo com as estimativas fornecidas pelo software TABLET, a cobertura média de reads para os mitogenomas variou entre 85x e 292x para as sequências mitocondriais nas quais um subconjunto dos dados foi usado. Para *T. aethiops*, a cobertura foi maior, dado que o dataset inteiro foi utilizado (712x). Observou-se uma distribuição uniforme da cobertura ao longo dos mitogenomas (Figura S1), exceto em casos nos quais segmentos ricos em AT da região de controle apresentaram baixa cobertura, geralmente próximos a sequências poli-T.

Species name	Bioproject	Experiment	Biosample	SRA Run number	FTP link used for dataset download	Data
<i>Pseudomyrmex concolor</i>	PRJNA268384	SRX831102	SAMN03275516	SRR1742927	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR174/007/SRR1742927	WGS
<i>Pseudomyrmex dendroicus</i>	PRJNA268384	SRX831097	SAMN03275515	SRR1742922	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR174/002/SRR1742922	WGS
<i>Pseudomyrmex elongatus</i>	PRJNA268384	SRX831106	SAMN03275518	SRR1742975	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR174/005/SRR1742975	WGS
<i>Pseudomyrmex feralis</i>	PRJNA357470	SRX2424867	SAMN06141944	SRR5112519	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR511/009/SRR5112519	UCE
<i>Pseudomyrmex ferrugineus</i>	PRJNA357470	SRX2424886	SAMN06141956	SRR5112538	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR511/008/SRR5112538	UCE
<i>Pseudomyrmex flavicornis</i>	PRJNA268384	SRX831107	SAMN03275519	SRR1742976	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR174/006/SRR1742976	WGS
<i>Pseudomyrmex gracilis</i>	PRJNA268384	SRX831110	SAMN03219222	SRR1742979	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR174/009/SRR1742979	WGS
<i>Pseudomyrmex janzeni</i>	PRJNA357470	SRX2424860	SAMN06141954	SRR5112512	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR511/002/SRR5112512	UCE
<i>Pseudomyrmex pallidus</i>	PRJNA268384	SRX831105	SAMN03275517	SRR1742947	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR174/007/SRR1742947	WGS
<i>Pseudomyrmex particeps</i>	PRJNA357470	SRX2424875	SAMN06141966	SRR5112527	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR511/007/SRR5112527	UCE
<i>Pseudomyrmex peperi</i>	PRJNA357470	SRX2424871	SAMN06141946	SRR5112523	ftp://ftp.sra.ebi.ac.uk/vol1/srr/SRR511/003/SRR5112523	UCE

Tabela 1: Informação acerca dos 14 datasets genômicos baixados do Sequence Read Archive para a montagem de mitogenomas completos de formigas da subfamília Pseudomyrmecinae

Pseudomyrmecinae Species	Species group	Mitogenome TPA accession number	NOVOPlasty seed	MITObim third assembly round needed	Mitogenome Coverage	Low coverage Region	Mitogenome Size (pb)	AT content: togenome (%)	AT content: D-loop region (%)
P. concolor	P. viidus	BK010475	KU985552.1	No	193.2x	No	15906	75	91
P. dendroicus	P. viidus	BK010473	KP271186.1	Yes	123.9x	No	17362	81	94
P. pallidus	P. pallidus	BK010383	KU985552.1	No	91.9x	No	17117	74	84
P. elongatus	P. oculatus	BK010474	KP271181.1	No	115.4x	No	17304	78	93
P. gracilis	P. gracilis	BK010472	FJ436821.1	No	165.5x	13761-13928	15704	77	93
P. feralis	P. ferrugineus	BK010379	FJ436819.1	No	128.0x	No	18835	78	92
P. ferrugineus	P. ferrugineus	BK010380	FJ436819.1	Yes	87.0x	No	18480	77	90
P. janzeni	P. ferrugineus	BK010382	FJ436819.1	No	125.8x	15848-15867	18380	77	89
P. particeps	P. ferrugineus	BK010384	FJ436819.1	No	126.8x	15799-15820	18524	80	90
P. peperi	P. ferrugineus	BK010385	FJ436819.1	Yes	87.4x	16006-16023	18709	78	91
P. veneficus	P. ferrugineus	BK010386	FJ436819.1	No	155.4x	15889-15928	18410	79	91
T. rufonigra	NE	BK010387	KX398231.1	No	292.2x	13889-13982	15907	74	91

Tabela 2: Informação sobre a montagem dos genomas mitocondriais das 14 espécies de formigas da subfamília Pseudomyrmecinae

4.2 Variação do tamanho de genomas mitocondriais e sítios de inserção no gênero *Pseudomyrmex* Mitogenomas de *Pseudomyrmex* mostraram variação significativa de tamanho, indo de 15704 a 18835 pb (Tabela 2). Observamos três faixas distintas de tamanho de mitogenoma para o gênero. O tamanho do genoma mitocondrial variou de: (i) menos de 16 kpb em *P. gracilis* e *P. concolor*; (ii) entre 17kpb e 18kpb em *P. pallidus* e *P. dendroicus*; e (iii) maior que 18kpb em outras espécies, pertencentes ao grupo de espécies *P. ferrugineus*. Uma análise de genômica comparativa usando o software BRIG identificou quatro regiões variáveis como segmentos putativos de inserção (Figura 8). Após a anotação do genoma, identificamos que essas supostas inserções possivelmente estão localizadas entre (i) COX2 e trn-K; (ii) ATP8 e ATP6; (iii) trn-N e trn-F; e (iv) trn-W e COX1.

4.3 Ordem gênica em mitogenomas de formigas Apesar da amostragem limitada de genomas mitocondriais completos disponível para formigas, cinco rearranjos de sintenia ligeiramente diferentes (Figura 9) foram observados na família Formicidae. Todos os mitogenomas de Pseudomyrmecinae e Dolichoderinae analisados mostraram uma única sintenia, conservada para todas as suas espécies e compartilhada pela maioria das espécies de Formicinae. Essa conservação de arranjo gênico foi levado em conta para determinar a região na qual o D-loop se encontra nas espécies de Pseudomyrmecinae. Também observamos que os clados Formicinae e Myrmicinae apresentam um arranjo modal de sintenia sugerindo uma possível ordem gênica ancestral para cada grupo. Uma única espécie de Formicinae (*Camponotus atrox*) apresenta inversões entre trn-M, I e Q que diferem de outros mitogenomas desta subfamília, possivelmente representando uma variação derivada. A subfamília Myrmicinae também apresenta dois outros rearranjos únicos restritos a uma única espécie cada, sugerindo sintenias derivadas: (i) *P. punctatus* tem uma inversão entre trn-K e D; e (ii) *W. auropunctata* apresenta uma inversão entre trn-V e D-loop e trnY na fita oposta quando comparada com as outras.

Figura 8: Análise de genômica comparativa de todas as 14 formigas da subfamília Pseudomyrmecinae. Comparação por BLAST de todos os genomas mitocondriais de Pseudomyrmecinae contra uma referência (*Pseudomyrmex janzeni*) gerada pelo Blast Ring Image Generator (BRIG). As lacunas presentes nos anéis correspondem a regiões com menos de 50 % de identidade com a sequência de

referência. A maioria das características mitocondriais é conservada dentro do clado, embora ATP8 e alguns tRNAs (trn-S, trn-E e trn-T) tenham apresentado maior variabilidade. Quatro regiões (identificadas por setas) apresentam variações de tamanho de nucleotídeos e são encontradas entre (i) COX2 e trn-K; (ii) ATP8 e ATP6; (iii) trn-N e trn-F e; (iv) trn-W e COX1.

Figura 9: Cinco sintenias observadas em mitogenomas completos da família Formicidae disponíveis no Genbank

Os dois arranjos gênicos modais estão representados dentro do retângulo horizontal e foram observados em 26 das 29 espécies analisadas: todas as Pseudomyrmecinae (14 espécies); todas as Dolichoderinae (duas espécies: *L. pallens* e *L. humile*); três das quatro Formicinae (*F. fusca*, *F. selysi* e *P. dives*) e em sete das nove Myrmicinae (*A. texana*; *C. obscurior*; *M. scabrinodis*; *S. richteri*; *S. geminata*; *S. invicta*; *V. emeryi*). Nós sugerimos que essas sintenias podem representar arranjos ancestrais para esses clados. As sintenias fora do retângulo horizontal correspondem às ordens gênicas cuja ocorrência é limitada a uma única espécie. Os retângulos verticais e linhas indicam regiões nas quais mudanças de sentença ocorreram e tanto o asterisco (*) quanto a linha vertical no trn-Y de *W. auropunctata* indicam que essa é a única feature em uma mitocôndria de formiga que mudou sua fita codificante ao longo de sua evolução.

4.4 Análises filogenéticas de Formicidae usando dados mitogenômicos

Para avaliar a filogenia do grupo, duas árvores de Máxima Verossimilhança foram produzidas usando dados de entrada ligeiramente diferentes: (i) as sequências alinhadas e concatenadas para todos os 13 PCGs mitocondriais (Figura 10); e (ii) os genomas mitocondriais completos (Figura 11). Analisamos todas as espécies de formigas que apresentam mitogenomas completos disponíveis no Genbank (Gotzek et al., 2010; Hasegawa et al., 2011; Berman et al., 2014; Babbucci et al., 2014; Kim et al., 2015; Duan et al., 2016; Liu et al., 2016; Yang et al., 2016) e duas abelhas da família Apidae como outgroups (Crozier & Crozier, 1993; Cha et al., 2007) (ver números de acesso e referências para todas as sequências na Tabela S2). As árvores reconstruídas a partir de dados mitocondriais corroboraram a maioria das relações filogenéticas conhecidas para formigas com vários clados observados como monofiléticos com alta confiança (bootstrap = 100). Ambas

as árvores apresentaram resultados semelhantes, embora diferenças possam ser observadas em vários nós quanto à topologia da árvore e/ou suporte estatístico. A principal diferença observada é que a árvore de genes concatenados exibiu todas as subfamílias como monofiléticas, enquanto Myrmicinae foi recuperada como parafilética na árvore baseada em mitogenomas completos.

Figura 10: Árvore filogenômica de concatenação gênica para todos os mitogenomas completos de Formicidae disponíveis no Genbank. A árvore foi construída usando as sequências nucleotídicas alinhadas e concatenadas para todos os 13 genes mitocondriais codificadores de proteínas. Modeltest identificou o GTR + G + I como o modelo de substituição mais adequado e a filogenia foi reconstruída por Maximum Likelihood usando o software MEGA7, com 1000 replicatas geradas pelo método de bootstrap. Abelhas da família Apidae foram utilizadas como grupo externo. Grupos de espécies do gênero *Pseudomyrmex* são evidenciados e espécies de *Pseudomyrmecinae* que apresentam características mutualistas são indicadas pela presença de um asterisco “*”.

Figura 11: Árvore filogenética usando a sequência mitocondrial completa de todos os mitogenomas de formigas disponíveis no Genbank. “GTR + G + I” foi escolhido como modelo de substituição, conforme sugerido pelo software Modeltest. A árvore foi construída usando MEGA7 pelo método de Maximum Likelihood com 1000 replicatas de bootstrap. Mitogenomas de abelhas foram usados como grupos externos. Grupos de espécies de *Pseudomyrmex* e espécies mutualistas de *Pseudomyrmecinae* são evidenciados.

5 Discussão

Neste estudo, usamos dados públicos para montar, anotar, comparar e fornecer análises evolutivas de 14 sequências completas de genoma mitocondrial da subfamília Pseudomyrmecinae e outros 15 mitogenomas de formigas baixados do GenBank.

5.1 Cobertura uniforme do mitogenoma e viés de AT

Mesmo que segmentos do genoma mitocondrial possam ser copiados para o núcleo formando NuMTs (Sequências Nucleares Mitocondriais), a cobertura genômica obtida para as montagens frequentemente apresentou distribuições uniformes (Figura S1), mesmo para *Pseudomyrmex gracilis* em que os NuMTs foram previamente identificados (Rubin & Moreau, 2016). A montagem correta dos genomas mitocondriais foi possível porque o número de reads mitocondriais é provavelmente muito maior do que o número de sequências provenientes de NuMTs.

A baixa cobertura em segmentos com uma tendência a AT pronunciada deve ser esperada pois as regiões ricas em AT são conhecidas por ter amplificação reduzida em protocolos de preparação de bibliotecas Illumina (Dohm et al., 2008; Aird et al., 2011; Oyola et al., 2012). Nas espécies analisadas, a região intergênica na qual o D-loop se encontra está localizada entre o RNA ribossomal 12S (rrns) e o RNA transportador da metionina (trn-M) e varia em tamanho de 527 pb (*P. concolor*) a 697 pb (*P. elongatus*) (Tabela S1). Vários estudos demonstraram que formigas apresentam viés de AT em seus mitogenomas, em especial na região de controle, que pode exceder 90 % (Berman et al., 2014; Liu et al., 2016). Nossos dados corroboram isso, já que a menor porcentagem de AT dentre as 14 espécies de Pseudomyrmecinae analisadas corresponde a 74 % e 12 dessas espécies apresentam valor de conteúdo de AT igual ou superior a 90 % na região intergênica que contém o D-loop (Tabela 1). Além disso, a região de controle já se mostrou particularmente difícil de sequenciar em himenópteros (Castro & Dowton, 2005; Dowton et al., 2009; Rodovalho et al., 2014). Tendo isso em vista, o fato de que segmentos de baixa cobertura em nossas montagens sempre ocorrerem no D-loop provavelmente está

associado ao viés de AT pronunciado dessa região.

O viés de AT dos mitogenomas de formigas, associado às limitações do sequenciamento para regiões ricas nesses nucleotídeos podem dificultar a obtenção de mitogenomas completos de formigas. Essa dificuldade provavelmente é parte do motivo pelo qual há tão poucos genomas mitocondriais completos disponíveis para esse grupo apesar da ampla disponibilidade de dados públicos. Ao mesmo tempo, devemos considerar que o advento de novas ferramentas podem tornar a montagem de mitogenomas mais acessível. Por exemplo, novos montadores como o NOVOPlasty superam programas clássicos (Dierckxsens, Mardulyn & Smits, 2016; Plese et al., 2018) e facilitam a produção de mitogenomas completos. Assim sendo, os incessantes avanços técnicos no ramo da bioinformática prenunciam perspectivas favoráveis para o fechamento de lacunas filogenéticas em Formicidae, especialmente se os dados públicos disponíveis para o clado forem utilizados para esse fim.

5.2 Mitogenômica Comparativa: tamanho do mitogenoma e análises de sintenia

Além da identificação de quatro sítios putativos de inserção que podem explicar as diferenças observadas no tamanho do mitogenoma (apontado pelas setas na Figura 8), também observamos que todos os sete mitogenomas incluídos no grupo de espécies *P. ferrugineus* têm aproximadamente o mesmo tamanho de sequência em pb, sugerindo que esse grupo é monofilético. Por outro lado, existe uma diferença significativa entre o tamanho do mitogenoma de *P. concolor* (15906 pb) e *P. dendroicus* (17362 pb), ambas pertencentes ao grupo *P. viidus*. Isto corrobora trabalhos anteriores que apontam este grupo de espécies como parafilético (Ward, 1989; Ward & Downie, 2005).

Há uma correlação positiva entre as múltiplas sintenias encontradas nos clados Myrmicinae e Formicinae e a notável biodiversidade observada para estas duas subfamílias: Myrmicinae, que apresentou o maior número de rearranjos gênicos (três), é a maior subfamília de formigas em termos de riqueza de espécies, com mais de 6.600 espécies descritas, quase metade de toda a biodiversidade documentada para formigas; e Formicidae, que apresentou duas sintenias distintas, é a segunda mais biodiversa, com mais de 3.100 espécies. As outras subfamílias de Formicidae

analisadas neste estudo (ambas com um único arranjo gênico) são menos diversas: Dolichoderinae exibe 713 espécies enquanto Pseudomyrmecinae apresenta 231 espécies documentadas (Bolton, 2012). Como o arranjo de genes ancestrais para Formicinae é idêntico ao observado em Pseudomyrmecinae e Dolichoderinae, a análise de sintenia indica que Formicinae está mais próximo filogeneticamente a este grupo do que a Myrmicinae. Um número maior de mitogenomas associado a uma cobertura taxonômica mais ampla melhorarão a avaliação da correlação entre a ordem dos genes mitocondriais e a biodiversidade da subfamília, permitindo um melhor entendimento da evolução mitocondrial da sintenia em Formicidae.

5.3 Relações filogenômicas de Formicidae inferidas usando dados de mitogenoma

As árvores filogenômicas geradas forneceram topologias ligeiramente diferentes devido a informação adicional presente na análise do mitogenoma completo. Enquanto a árvore de concatenação gênica utiliza apenas a informação contida nos genes codificadores de proteínas, a árvore construída com base na sequência mitocondrial completa usa, além das PCG's, informação proveniente das regiões intergênicas e dos tRNAs, rRNAs e D-loop para a inferência filogenética. Ademais, o DNA mitocondrial apresenta uma taxa de substituição relativamente elevada em regiões não codificantes (Vanecek, Vorel & Sip, 2004; DeSalle, 2017) e a adição dessa variabilidade às análises também justifica as diferenças topológicas observadas.

Em geral, nas árvores filogenômicas geradas para todas as formigas apresentando mitogenoma completo, a filogenia da subfamília Pseudomyrmecinae foi fortemente recuperada como monofilética, e as posições filogenéticas da maioria dos clados foram bem resolvidas. A monofilia para a subfamília Pseudomyrmecinae e também para os gêneros Pseudomyrmex e Tetraponera foi recuperada com 100 % de suporte de bootstrap (BS) em ambas as árvores. O gênero Pseudomyrmex apresentou poucos nós não suportados, mas Tetraponera foi completamente resolvida em ambas as árvores (BS = 100). Com relação aos grupos de espécies em Pseudomyrmex, em ambas as árvores o status monofilético do grupo *P. flavicornis* e o estado parafilético do grupo *P. viidus* confirmam (i) observações prévias baseadas exclusivamente em morfologia (Ward, 1989), (ii) filogenias usando caracteres morfológicos em conjunto com marcadores nucleares (Ward & Downie,

2005), e (iii) nossas próprias observações em relação ao tamanho do mitogenoma. Embora a divisão morfológica em grupos de espécies não tenha sido formalizada ou regulada nomenclaturalmente (Ward, 2017), o trabalho usando uma abordagem híbrida morfológica/molecular de Ward & Downie, 2005 mostra que apenas dois dos nove grupos definidos na época eram parafiléticos: *P. pallens* e *P. viidus*. A corroboração de estudos morfológicos pela análise de dados mitocondriais confirma a relevância do uso de caracteres morfológicos na determinação das relações entre clados. Ao mesmo tempo, nossos resultados reforçam que evidências moleculares podem esclarecer e complementar tais estudos, refinando e melhorando o suporte geral das filogenias reconstruídas. Neste trabalho, geramos sequências mitocondriais completas para formigas classificadas em cinco dos 10 grupos descritos para espécies de *Pseudomyrmex*, cobrindo pelo menos metade da diversidade genética do gênero e adicionando uma nova fonte de evidência molecular para estudos posteriores sobre o clado.

Ambas as árvores sugerem fortemente que os mutualismos de formigas são parafiléticos em *Pseudomyrmecinae* (por favor, verifique os asteriscos presentes nas Figuras 10 e 11), adicionando também evidências às suposições prévias de comportamento generalista como um traço basal do gênero *Pseudomyrmex* (Ward & Branstetter, 2017). Isso sugere que a relação de co-evolução entre plantas e essas formigas se desenvolveu mais tarde (e independentemente) várias vezes no clado. Espécies mutualistas são mais comuns no grupo de espécies *P. ferrugineus*, reforçando a hipótese do mutualismo ser uma característica derivada. No gênero *Pseudomyrmex*, o grupo *P. ferrugineus* possivelmente apresenta duas linhagens independentes de formigas mutualísticas (já que *P. feralis* é frequentemente considerada como exibindo comportamento generalista; BS = 50), enquanto outras duas linhagens mutualistas independentes podem ser observadas dentro do gênero ao se analisar o posicionamento filogenético de *P. concolor* e *P. dendroicus* nas árvores. Considerando o gênero *Tetraponera*, *T. aethiops* e *T. rufonigra* são espécies intimamente relacionadas e apenas *T. aethiops* apresenta comportamento mutualístico (Ward & Downie, 2005), mostrando-nos que a diferenciação de traços ecológicos pode ser observado mesmo entre espécies aparentadas. Infelizmente não há estudos que estimem o tempo de divergência para espécies do gênero *Tetraponera*, mas essa

diferenciação ecológica observada em espécies próximas (para as quais é esperado um passado evolutivo comum relativamente recente) pode ser indicativa de que a evolução de traços mutualistas em *Pseudomyrmecinae* pode ocorrer em períodos de tempo consideravelmente curtos. Considerando o número limitado de espécies amostradas aqui, fomos capazes de identificar 5 das 12 vezes em que associações mutualísticas desenvolvidas independentemente foram relatadas no clado (Ward, 1991; Ward & Downie, 2005). Com uma melhor cobertura taxonômica, esse número pode ser aumentado e novas análises realizadas, gerando resultados mais robustos e elucidando cada vez mais esses eventos coevolutivos.

Há diversas relações filogenéticas bem resolvidas para várias espécies de *Pseudomyrmecinae* (como *P. peperi*, *P. veneficus*, *P. particeps*, *P. gracilis*, *T. aethiops* e *T. rufonigra*) que corroboram tanto os resultados de Ward & Downie (2005) quanto a árvore de Máxima Verossimilhança gerada usando dados de elementos ultra-conservados de Ward & Branstetter (2017). A relação do grupo irmão entre *P. dendroicus* e *P. elongatus* também é bem suportada (BS = 100 na árvore de mitocôndria completa; e BS = 99 na árvore de concatenação gênica), indo ao encontro de um trabalho recente utilizando scaffolds de WGS concatenados como entrada para a construção de árvores de ML (Rubin & Moreau, 2016).

No entanto, diferenças sutis foram observadas entre nossos resultados e as relações filogenéticas inferidas com base em elementos ultra-conservados de Ward & Branstetter, 2017. Usando dados de UCE, *P. janzeni* foi observado como grupo irmão de *P. ferrugineus*. Neste trabalho, a árvore, usando a sequência mitogenômica completa, recapturou essa mesma relação com um valor replicado de bootstrap de 77. Por outro lado, na árvore de genes concatenados, além da relação de grupo irmão ter sido observada entre *P. janzeni* e *P. flavicornis*, ela mostrou um suporte inferior (BS = 47). No geral, essa relação pareceu ser melhor reconstruída pela análise da sequência mitocondrial completa, que corrobora as análises de UCE. Dentro de *Pseudomyrmecinae*, observamos duas espécies cujas posições filogenéticas não foram bem resolvidas pelas análises mitocondriais atuais e, portanto, sua relação pode ser vista como inconclusiva: (i) *P. feralis* em ambas as árvores filogenéticas; e (ii) *P. pallidus* na árvore de genes concatenados. Essas posições apresentam um valor de suporte de bootstrap de 50.

Ambas as árvores apresentaram a subfamília Dolichoderinae como monofilética, embora este resultado não tenha sido recuperado em todas as replicatas. Dolichoderinae é uma subfamília altamente diversificada e contém mais de 700 espécies, mas foi aqui representada por apenas duas espécies. Assim, acreditamos que uma maior cobertura de espécies irá melhorar a robustez das análises filogenéticas para o clado.

Trabalhos anteriores com caracteres morfológicos e/ou genes nucleares apresentam evidências de relação de grupo irmão entre Pseudomyrmecinae e Myrmeciinae (Ward & Downie, 2005; Brady et al., 2006). Nós esperaríamos que Myrmeciinae fosse o grupo irmão de Pseudomyrmecinae de acordo com os dados mitocondriais, mas como genomas mitocondriais completos não estão disponíveis para a subfamília Myrmeciinae, nós não pudemos testar essa hipótese. A ausência de mitogenomas para essa e outras subfamílias podem estar associadas ao fato de que sua biodiversidade e importância econômica não são tão expressivas quando comparadas à das subfamílias de formigas mais estudadas. Por exemplo, Myrmeciinae não apresenta nenhuma espécie de pronunciada relevância econômica e conta com apenas 94 espécies descritas, enquanto Myrmicinae engloba as saúvas (tribo Attini), notoriamente conhecidas como importantes pragas agrícolas e possui mais de 6600 espécies (Bolton, 2012).

Na ausência de Myrmeciinae, espera-se que Dolichoderinae seja o grupo mais próximo de Pseudomyrmecinae em nossas árvores. Isso foi confirmado em ambas as árvores, nas quais as duas subfamílias aparecem como grupos irmãos entre si, corroborando filogenias moleculares de larga escala usando poucos genes nucleares (Brady et al., 2006) e dados de UCE (Branstetter et al., 2017). A sintenia compartilhada entre todas as espécies de Pseudomyrmecinae e Dolichoderinae também suporta a relação do grupo irmão observada. Nossos resultados sugerem que Myrmicinae é o táxon mais próximo de um clado contendo Pseudomyrmecinae e Dolichoderinae, enquanto Formicinae foi observado como um grupo mais basal na família Formicidae. Esta posição basal de Formicinae é altamente suportada na árvore de concatenação de genes, mas não na árvore usando mitogenomas completos, ao contrário do que é mostrado por outros trabalhos usando dados nucleares que apontam para uma relação de grupo irmão entre Myrmicinae e Formicinae (Brady

et al., 2006; Branstetter et al., 2017).

A monofilia da subfamília Formicinae e todos os seus nós mostram suporte máximo em ambas as árvores (BS = 100). Nossos resultados também corroboram o caráter monofilético do gênero *Formica* e apresentam os gêneros *Camponotus* e *Polyrhachis* como intimamente relacionados entre si, conforme observado no trabalho de Blaimer e colaboradores (2015), que utilizaram locus de UCE para inferência filogenética. O único problema com relação a essa subfamília diz respeito à posição mal suportada de Formicinae em relação às outras subfamílias. Os dados de mitogenoma forneceram com sucesso relações filogenéticas robustas, mesmo para *Camponotus atrox*, uma espécie que mostrou sintenia única, mas teve sua posição bem resolvida em ambas as inferências, inclusive na árvore mitocondrial completa, que pode estar propensa a ser afetada por alterações de sintenia. Esta questão confirma a robustez das sequências mitocondriais para inferir filogenias de formigas.

No geral, os resultados mais controversos obtidos aqui estão relacionados à posição da subfamília Myrmicinae. Para esse clado, a árvore de concatenação de genes foi capaz de indicar monofilia (BS = 74), mas dados de mitogenoma total produziram parafilia. Neste último caso, as espécies *Atta texana*, *Myrmica scabrinodis* e *Pristomyrmex punctatus* divergiram antes das outras formigas. Por outro lado, ambas as árvores recapturaram com sucesso o caráter monofilético do gênero *Solenopsis* e as relações entre suas espécies (*S. geminata* como grupo irmão do clado consistindo de *S. invicta* e *S. richteri*) com 100 % de suporte de bootstrap. A relação do grupo irmão entre *Solenopsis* spp. e *Vollenhovia emeryi* também é recuperada. Estes resultados corroboram aqueles obtidos pelo uso de sequências de aminoácidos concatenados de todos os PCGs mitocondriais para inferência de árvores (Duan et al., 2016). Entretanto, nossa avaliação da posição de *V. emeryi* foi melhor suportada (BS = 90 na árvore de concatenação gênica e BS = 99 na árvore mitocondrial completa) do que a deste trabalho anterior (BS = 75). Considerando que Duan e colaboradores (2016) utilizaram uma abordagem semelhante à nossa (concatenação gênica seguida por construção de árvore usando Máxima Verossimilhança), podemos concluir que esses resultados indicam que os dados nucleotídicos apresentam informações mais confiáveis para a inferência

filogenômica desses clados do que os dados aminoacídicos. Isto é consistente com pesquisa anterior na qual a inferência filogenética utilizando nucleotídeos obteve resultados melhor suportados do que as análises ao nível de aminoácidos ou codons (Holder, Zwickl & Dessimoz, 2008). Além disso, os valores de bootstrap obtidos através de dados nucleotídicos já foram relatados como geralmente maiores do que aqueles provenientes de seus correspondentes em aminoácidos (Regier et al., 2010). Essas observações são ao menos parcialmente explicadas pelas diferenças na quantidade de sinal filogenético considerados por esses dois métodos. Sinal adicional presente em sequências de nucleotídeos é perdido na tradução para aminoácidos. Isso é particularmente importante em se tratando de aminoácidos hexacodônicos como a serina, que é codificada tanto por TCN quanto por AGY (Regier et al., 2010; Zwick, Regier & Zwickl, 2012).

As relações filogenéticas de outras espécies da subfamília Myrmicinae na nossa árvore de concatenação gênica não estão bem resolvidas, como a posição de *Myrmica scabrinodis* (BS = 52), *Wasmannia auropunctata* (BS = 43) e *Pristomyrmex punctatus* (BS = 16). No entanto, a posição dessas espécies na árvore de aminoácidos de Duan e colaboradores (2016) também é inconclusiva e difere daqui, agrupando *W. auropunctata* e *M. scabrinodis* em uma relação suportada apenas em 35 % das replicatas de bootstrap. Este clado é colocado como grupo irmão de *Solenopsis* spp. e *V. emeryi* com suporte ainda menor (BS = 21) e *P. punctatus* assume uma posição mais basal na árvore em 46 % das repetições. No entanto, *Atta laevigata* aparece na base de todas as Myrmecinae com suporte de bootstrap de 100 % na árvore de aminoácidos. Como o mitogenoma de *A. laevigata* disponível não está completo, ele não foi usado como entrada para a concatenação de nucleotídeos das PCGs aqui realizada, ao contrário de sua congênica *Atta texana*, cujo mitogenoma completo foi analisado aqui. *A. texana* também aparece na base da subfamília Myrmicinae, mas sob uma relação de grupo irmão com *M. scabrinodis*, mesmo que com baixa resolução (BS = 52). Este clado é irmão de todas as outras espécies de Myrmecinae (BS = 74). Por fim, a posição de *Cardiocondyla obscurior* também não foi bem suportada (BS = 43), mas como esse é um mitogenoma recentemente publicado, não foi utilizado no trabalho de Duan e colaboradores.

Em ambos os trabalhos, as análises mitogenômicas não foram totalmente capazes de resolver importantes nós do ramo das Myrmicinae e vários fatores podem estar associados a esses resultados insatisfatórios. É necessário destacar que Myrmicinae é a subfamília mais biodiversa (Bolton, 2012) e é conhecida por apresentar vários grupos monofiléticos duvidosos (Brady et al., 2006; Ward, 2011; Ward et al., 2015). Essa diversidade é evidenciada pelo fato de que, apesar de apenas nove mitogenomas estarem disponíveis para o grupo, três arranjos diferentes de genes mitocondriais podem ser observados, sugerindo uma alta taxa de evolução mitocondrial nessa subfamília.

Além disso, houveram divergências no ramo das Myrmicinae em estudos filogenéticos moleculares anteriores que tentaram estudar a família Formicidae como um todo (Brady et al., 2006; Moreau et al., 2006). Por outro lado, Ward et al. (2015) se foca no estudo dessa subfamília ao reconstruir uma filogenia em grande escala usando 11 marcadores nucleares de 251 espécies amostradas em todas as 25 tribos de Myrmicinae, a maioria delas parafiléticas. Utilizando uma grande quantidade de dados que cobre uma extensa parcela da diversidade de espécies dessa subfamília, eles conseguiram propor uma nova classificação de Myrmicinae composta exclusivamente por tribos monofiléticas, o que também reduziu o número de gêneros parafiléticos.

Assim, a natureza hiperdiversa deste clado, associada à subamostragem ou mesmo ausência de mitogenomas para vários táxons da subfamília e uma possível alta taxa de evolução do genoma mitocondrial são fatores que podem ter contribuído para os resultados inconclusivos das análises mitocondriais. Além disso, apesar de algumas relações não terem sido elucidadas pelo uso exclusivo da filogenômica mitocondrial, a informação fornecida pelo mitogenoma é classicamente considerada como útil no estudo das relações evolutivas para diversos táxons, seja confirmando (Prosdocimi et al., 2012; Finstermeier et al. al., 2013) ou refutando hipóteses filogenéticas anteriores (Kayal et al., 2015; Uliano-Silva et al., 2016). Assim, ainda recomendamos o uso de dados mitocondriais, de preferência ao lado de outros marcadores (por exemplo, genes nucleares), para aumentar o sinal filogenético e recapturar filogenias mais robustas. Entretanto, graças à taxa de substituição do mtDNA, árvores geradas a partir de dados mitocondriais apresentam

uma maior probabilidade de resolver ramos curtos corretamente (DeSalle, 2017). Portanto, também acreditamos que o uso de dados mitocondriais para inferência filogenômica, mesmo sem outros marcadores, produzirá resultados mais satisfatórios se trabalharmos no sentido de mitigar o problema da escassez de mitogenomas disponíveis para esse clado e melhorarmos a cobertura mitocondrial de seus táxons. Essa afirmativa não só é válida para a subfamília Myrmicinae, como também para a família Formicidae como um todo e para qualquer outro grupo com escassez de mitogenomas conhecidos e necessidade de elucidação sobre suas relações filogenéticas. Nesse sentido, os resultados aqui apresentados são extremamente relevantes para mostrar que as informações já disponíveis em bancos de dados públicos devem ser usadas para obter genomas mitocondriais completos e fomentar novas pesquisas que gerarão conhecimento sem incorrer em custos adicionais de sequenciamento.

5.4 Mitogenômica no-budget: análises integradas entre datasets e potencial para estudos de larga-escala

Os resultados aqui apresentados confirmam que os dados de UCE e WGS publicamente disponíveis podem ser usados para montar genomas mitocondriais completos com alta cobertura (Tabela 2), o que pode ser explicado pelo alto número de cópias de reads mitocondriais que pode alcançar algo entre 0,25 % a 0,5 % do número total de bases geradas (Prosdocimi et al., 2012) e chegar a 2 % do total de reads mapeando ao mtDNA (Ekblom, Smeds & Ellegren, 2014). Também confirmamos o potencial dos dados de UCE como uma alternativa de baixo custo para sequenciar mitogenomas completos com alta cobertura, conforme descrito por Raposo do Amaral et al. (2015). Dados de mitogenoma são usados em várias análises e sequências mitocondriais são encontradas em vários tipos de datasets, que geralmente fornecem informação suficiente para montar toda a sequência mitocondrial. Essa versatilidade e onipresença de sequências mitocondriais deve ser usada em favor dos estudos de biodiversidade, especialmente considerando que os datasets públicos estão disponíveis para um número cada vez maior de espécies.

O potencial dessas sequências na elucidação de filogenias não deve ser menos-prezado, especialmente se considerarmos que existem diferentes tipos de conjuntos de dados disponíveis para diferentes espécies (WGS, RNA-Seq, enriquecimento de UCE, dentre outros). Esses diferentes recursos dificultam a obtenção de árvores

filogenéticas/filogenômicas que integrem esses diferentes dados públicos, já que muitas vezes as análises dependem da ortologia das sequências comparadas (Kuzniar et al., 2008). Assim, o uso de diferentes tipos de dados para montar os mitogenomas completos ou quase completos para espécies com dados publicamente disponíveis apresenta uma solução para este problema com o genoma mitocondrial agindo como uma “sequência normalizadora” que permite a comparação de diferentes conjuntos de dados. Por exemplo, neste trabalho algumas espécies tinham apenas dados de UCE disponíveis publicamente, enquanto outros apresentavam datasets padrão de WGS. No entanto, anotando e analisando o mitogenoma completo para essas espécies, conseguimos ampliar nosso escopo e estudar todas elas juntas. Assim, sugerimos que o uso de mitogenomas obtidos a partir de dados públicos tem o potencial de se tornar uma importante fonte de informação filogenética. Além disso, o estudo das sequências mitocondriais pode ser uma das rotas mais rápidas para a obtenção de árvores abrangentes de alta qualidade para táxons hiperdiversos, como os insetos. Progresso se tem feito nesse sentido, como pode ser visto no trabalho recente de Linard et al. (2018), onde a mineração de dados do Genbank e montagem usando datasets metagenômicos forneceram contigs mitocondriais (> 3kpb) para quase 16.000 espécies de coleópteros. Essa enorme quantidade de dados mitogenômicos foi usada para gerar a maior árvore filogenética já vista para o clado.

Estudos que tentam montar mitogenomas completos usando dados públicos ainda são escassos, ao passo que o tamanho e a amplitude dos bancos de dados públicos estão em crescimento, juntamente com seu potencial para responder questões filogenéticas, dentre tantas outras. A mitogenômica no-budget é uma oportunidade sem precedentes de reconstruir e analisar filogenias em larga escala para vários grupos em diferentes níveis taxonômicos, o que por sua vez pode subsidiar estudos evolutivos e de biologia da conservação e incrementar nosso conhecimento sobre espécies não-modelo e sua diversidade.

Parte I

Preparação da pesquisa

6 Conteúdos específicos do modelo de trabalho acadêmico

6.1 Quadros

Este modelo vem com o ambiente **quadro** e impressão de Lista de quadros configurados por padrão. Verifique um exemplo de utilização:

Quadro 1 – Exemplo de quadro

Pessoa	Idade	Peso	Altura
Marcos	26	68	178
Ivone	22	57	162
...
Sueli	40	65	153

Fonte: Autor.

Este parágrafo apresenta como referenciar o quadro no texto, requisito obrigatório da ABNT. Primeira opção, utilizando **autoref**: Ver o [Quadro 1](#). Segunda opção, utilizando **ref**: Ver o Quadro [1](#).

Parte II

Referenciais teóricos

7 Lorem ipsum dolor sit amet

7.1 Aliquam vestibulum fringilla lorem

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar

lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Parte III

Resultados

8 Lectus lobortis condimentum

8.1 Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

9 Nam sed tellus sit amet lectus urna ullamcorper tristique interdum elemen- tum

9.1 Pellentesque sit amet pede ac sem eleifend consectetuer

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

10 Conclusão

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.

Sed eleifend, eros sit amet faucibus elementum, urna sapien consectetur mauris, quis egestas leo justo non risus. Morbi non felis ac libero vulputate fringilla. Mauris libero eros, lacinia non, sodales quis, dapibus porttitor, pede. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi dapibus mauris condimentum nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Etiam sit amet erat. Nulla varius. Etiam tincidunt dui vitae turpis. Donec leo. Morbi vulputate convallis est. Integer aliquet. Pellentesque aliquet sodales urna.

Referências

ADAMS, K. L.; PALMER, J. D. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution*, v. 29, n. 3, p. 380–395, 2003. ISSN 10557903. Citado na página 32.

BERNSTEIN, M. N.; DOAN, A.; DEWEY, C. N. MetaSRA: Normalized human sample-specific metadata for the Sequence Read Archive. *Bioinformatics*, v. 33, n. 18, p. 2914–2923, 2017. ISSN 14602059. Citado na página 19.

BERNT, M. et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, Elsevier Inc., v. 69, n. 2, p. 313–319, 2013. ISSN 10557903. Disponível em: <<http://dx.doi.org/10.1016/j.ympev.2012.08.023>>. Citado na página 38.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114–2120, 2014. ISSN 14602059. Citado na página 19.

BOORE, J. L. Animal mitochondrial genomes. *Nucleic Acids Research*, v. 27, n. 8, p. 1767–1780, 1999. ISSN 03051048. Citado na página 32.

BORDBARI, M. H. et al. Deletion of 2.7 kb near HOXD3 in an Arabian horse with occipitoatlantoaxial malformation. *Animal Genetics*, v. 48, n. 3, p. 287–294, 2017. ISSN 13652052. Citado na página 19.

BRAND, M. D. Regulation analysis of energy metabolism. *The Journal of experimental biology*, v. 200, n. Pt 2, p. 193–202, 1997. ISSN 0022-0949. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9050227>>. Citado na página 32.

CHAN, D. C. Mitochondria: Dynamic Organelles in Disease, Aging, and Development. *Cell*, v. 125, n. 7, p. 1241–1252, 2006. ISSN 00928674. Citado na página 32.

COCHRANE, G. et al. Evidence Standards in Experimental and Inferential INSDC Third Party Annotation Data. *OMICS A Journal of Integrative Biology*, v. 10, n. 2, 2006. Citado 2 vezes nas páginas 19 e 47.

Del Fabbro, C. et al. An extensive evaluation of read trimming effects on illumina NGS data analysis. *PLoS ONE*, 2013. ISSN 19326203. Citado na página 19.

DIERCKXSENS, N.; MARDULYN, P.; SMITS, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, v. 45, n. 4, p. gkw955, 2016. ISSN 0305-1048. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw955>>. Citado na página 42.

Do Amaral, F. R. et al. Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS ONE*, v. 10, n. 9, p. 1–9, 2015. ISSN 19326203. Citado na página 33.

FINSTERMEIER, K. et al. A Mitogenomic Phylogeny of Living Primates. *PLoS ONE*, 2013. ISSN 19326203. Citado na página 32.

GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, Nature Publishing Group, v. 17, n. 6, p. 333–351, 2016. ISSN 14710064. Disponível em: <<http://dx.doi.org/10.1038/nrg.2016.49>>. Citado na página 19.

GRAY, M. W. Lynn Margulis and the endosymbiont hypothesis: 50 years later. *Molecular Biology of the Cell*, v. 28, n. 10, p. 1285–1287, 2017. ISSN 1059-1524. Disponível em: <<http://www.molbiolcell.org/lookup/doi/10.1091/mbc.E16-07-0509>>. Citado na página 32.

GRAY, M. W.; BURGER, G.; LANG, B. F. *Mitochondrial evolution*. 1999. Citado na página 32.

JANZEN, D. H. Coevolution of Mutualism Between Ants and Acacias in Central America. *Evolution*, 1966. ISSN 00143820. Citado na página 36.

KARSCH-MIZRACHI, I.; TAKAGI, T.; COCHRANE, G. The international nucleotide sequence database collaboration. *Nucleic Acids Research*, v. 46, n. D1, p. D48–D51, 2018. ISSN 13624962. Citado na página 19.

KAYAL, E. et al. Phylogenetic analysis of higher-level relationships within Hydrozoa (Cnidaria: Hydrozoa) using mitochondrial genome data and insight into their mitochondrial transcription. *PeerJ*, v. 3, p. e1403, 2015. ISSN 2167-8359. Disponível em: <<https://peerj.com/articles/1403>>. Citado na página 19.

KODAMA, Y.; SHUMWAY, M.; LEINONEN, R. The sequence read archive: Explosive growth of sequencing data. *Nucleic Acids Research*, v. 40, n. D1, p. 2011–2013, 2012. ISSN 03051048. Citado na página 19.

KRZEMIŃSKA, U. et al. Population mitogenomics provides insights into evolutionary history, source of invasions and diversifying selection in the House Crow (*Corvus splendens*). *Heredity*, 2018. ISSN 13652540. Citado na página 32.

KUTSCHERA, U.; NIKLAS, K. J. Endosymbiosis, cell evolution, and speciation. *Theory in Biosciences*, v. 124, n. 1, p. 1–24, 2005. ISSN 14317613. Citado na página 32.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, v. 9, n. 4, p. 357–359, 2012. ISSN 15487091. Citado na página 19.

LINARD, B. et al. The contribution of mitochondrial metagenomics to largescale data mining and phylogenetic analysis of Coleoptera. *bioRxiv*, n. March, p. 280792, 2018. Disponível em: <<https://www.biorxiv.org/content/early/2018/03/12/280792.figures-only>>. Citado na página 19.

MARDIS, E. R. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, v. 24, n. 3, p. 133–141, 2008. ISSN 01689525. Citado na página 19.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005. ISSN 00280836. Citado na página 19.

MILLER, M. J. et al. Complete mitochondrial genomes of the New World jacanas: *Jacana spinosa* and *Jacana jacana*. *Mitochondrial DNA*, v. 27, n. 1, p. 764–765, 2016. ISSN 19401744. Citado na página 33.

RUBIN, B. E. R. et al. Comparative genomics reveals convergent rates of evolution in ant–plant mutualisms. *Nature Communications*, Nature Publishing Group, v. 7, p. 12679, 2016. ISSN 2041-1723. Disponível em: <<http://www.nature.com/doifinder/10.1038/ncomms12679>>. Citado na página 23.

SIMPSON, J. T. et al. ABySS : A parallel assembler for short read sequence data. *ABySS : A parallel assembler for short read sequence data*. p. 1117–1123, 2009. Citado na página 19.

SMITH, D. R. Goodbye genome paper, hello genome report: the increasing popularity of 'genome announcements' and their impact on science. *Briefings in Functional Genomics*, v. 16, n. 3, p. 156–162, 2016. Disponível em: <www.arrogantgenome.com>. Citado na página 32.

STODDEN, V.; SEILER, J.; MA, Z. An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, v. 115, n. 11, p. 2584–2589, 2018. ISSN 0027-8424. Disponível em: <<http://www.pnas.org/lookup/doi/10.1073/pnas.1708290115>>. Citado na página 19.

WANET, A. et al. Connecting Mitochondria, Metabolism, and Stem Cell Fate. *Stem Cells and Development*, v. 24, n. 17, p. 1957–1971, 2015. ISSN 1547-3287. Disponível em: <<http://online.liebertpub.com/doi/10.1089/scd.2015.0117>>. Citado na página 32.

WARD, P. S. *Phylogenetic analysis of pseudomyrmecine ants associated with domatia-bearing plants*. 1991. Citado na página 36.

WIRTHLIN, M. et al. Parrot Genomes and the Evolution of Heightened Longevity and Cognition. *Current Biology*, p. 1–8, 2018. ISSN 09609822. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0960982218314179>>. Citado na página 23.

WOLSTENHOLME, D. R. Animal Mitochondrial DNA: Structure and Evolution. *International Review of Cytology*, v. 141, n. C, p. 173–216, 1992. ISSN 00747696. Citado na página 32.

Apêndices

APÊNDICE A – Quisque libero justo

Quisque facilisis auctor sapien. Pellentesque gravida hendrerit lectus. Mauris rutrum sodales sapien. Fusce hendrerit sem vel lorem. Integer pellentesque massa vel augue. Integer elit tortor, feugiat quis, sagittis et, ornare non, lacus. Vestibulum posuere pellentesque eros. Quisque venenatis ipsum dictum nulla. Aliquam quis quam non metus eleifend interdum. Nam eget sapien ac mauris malesuada adipiscing. Etiam eleifend neque sed quam. Nulla facilisi. Proin a ligula. Sed id dui eu nibh egestas tincidunt. Suspendisse arcu.

APÊNDICE B – Nullam elementum urna vel imperdiet sodales elit ipsum pharetra ligula ac pretium ante justo a nulla curabitur tristique arcu eu metus

Nunc velit. Nullam elit sapien, eleifend eu, commodo nec, semper sit amet, elit. Nulla lectus risus, condimentum ut, laoreet eget, viverra nec, odio. Proin lobortis. Curabitur dictum arcu vel wisi. Cras id nulla venenatis tortor congue ultrices. Pellentesque eget pede. Sed eleifend sagittis elit. Nam sed tellus sit amet lectus ullamcorper tristique. Mauris enim sem, tristique eu, accumsan at, scelerisque vulputate, neque. Quisque lacus. Donec et ipsum sit amet elit nonummy aliquet. Sed viverra nisl at sem. Nam diam. Mauris ut dolor. Curabitur ornare tortor cursus velit.

Morbi tincidunt posuere arcu. Cras venenatis est vitae dolor. Vivamus scelerisque semper mi. Donec ipsum arcu, consequat scelerisque, viverra id, dictum at, metus. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut pede sem, tempus ut, porttitor bibendum, molestie eu, elit. Suspendisse potenti. Sed id lectus sit amet purus faucibus vehicula. Praesent sed sem non dui pharetra interdum. Nam viverra ultrices magna.

Aenean laoreet aliquam orci. Nunc interdum elementum urna. Quisque erat. Nullam tempor neque. Maecenas velit nibh, scelerisque a, consequat ut, viverra in, enim. Duis magna. Donec odio neque, tristique et, tincidunt eu, rhoncus ac, nunc. Mauris malesuada malesuada elit. Etiam lacus mauris, pretium vel, blandit in, ultricies id, libero. Phasellus bibendum erat ut diam. In congue imperdiet lectus.

Anexos

ANEXO A – Morbi ultrices rutrum lorem.

Sed mattis, erat sit amet gravida malesuada, elit augue egestas diam, tempus scelerisque nunc nisl vitae libero. Sed consequat feugiat massa. Nunc porta, eros in eleifend varius, erat leo rutrum dui, non convallis lectus orci ut nibh. Sed lorem massa, nonummy quis, egestas id, condimentum at, nisl. Maecenas at nibh. Aliquam et augue at nunc pellentesque ullamcorper. Duis nisl nibh, laoreet suscipit, convallis ut, rutrum id, enim. Phasellus odio. Nulla nulla elit, molestie non, scelerisque at, vestibulum eu, nulla. Ut odio nisl, facilisis id, mollis et, scelerisque nec, enim. Aenean sem leo, pellentesque sit amet, scelerisque sit amet, vehicula pellentesque, sapien.

ANEXO B – Cras non urna sed feugiat cum sociis natoque penatibus et magnis dis parturient montes nascetur ridiculus mus

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

ANEXO C – Fusce facilisis lacinia dui

Phasellus id magna. Duis malesuada interdum arcu. Integer metus. Morbi pulvinar pellentesque mi. Suspendisse sed est eu magna molestie egestas. Quisque mi lorem, pulvinar eget, egestas quis, luctus at, ante. Proin auctor vehicula purus. Fusce ac nisl aliquam ante hendrerit pellentesque. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Morbi wisi. Etiam arcu mauris, facilisis sed, eleifend non, nonummy ut, pede. Cras ut lacus tempor metus mollis placerat. Vivamus eu tortor vel metus interdum malesuada.