

Universidade Federal de Pernambuco  
Centro de Biociências  
Programa de Pós-Graduação em Genética

**ABIGAIL MARCELINO DOS SANTOS SILVA**

**AVALIAÇÃO DE MÉTODOS PROBABILÍSTICOS DE INFERÊNCIA  
FILOGENÉTICA NA INVESTIGAÇÃO DE COMPLEXOS DE ESPÉCIES  
CRÍPTICAS: ESTUDO DE CASO EM FLEBOTOMÍNEOS DE INTERESSE  
EPIDEMIOLÓGICO**

**Recife  
2015**

**Abigail Marcelino dos Santos Silva**

**Avaliação de métodos probabilísticos de inferência filogenética na  
investigação de complexos de espécies crípticas: estudo de caso em  
flebotomíneos de interesse epidemiológico**

Dissertação apresentada ao Programa de Pós-Graduação em Genética, da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Mestre em Genética.

Orientador: Prof. Dr. Valdir de Queiroz Balbino

Recife

2015

Catálogo na fonte  
Elaine Barroso  
CRB 1728

**Silva, Abigail Marcelino dos Santos**

**Avaliação de métodos probabilísticos de inferência filogenética na investigação de complexos de espécies crípticas: estudo de caso em flebotomíneos de interesse epidemiológico. / Abigail Marcelino dos Santos Silva - Recife: O Autor, 2015.**

**185 folhas: il., fig., tab.**

**Orientador: Valdir de Queiroz Balbino.**

**Dissertação (mestrado) – Universidade Federal de Pernambuco.**

**Centro de Biociências. Genética, 2015.**

**Inclui referências e apêndices**

**1. Análise cladística 2. Leishmaniose 3. Mosquito como transmissor de doenças I. Balbino, Valdir de Queiroz (orient.) II. Título**

**578.012**

**CDD (22.ed.)**

**UFPE/CB-2017-387**

**Abigail Marcelino dos Santos Silva**

**Avaliação de métodos probabilísticos de inferência filogenética na  
investigação de complexos de espécies crípticas: estudo de caso em  
flebotomíneos de interesse epidemiológico**

Dissertação apresentada ao Programa de Pós-Graduação em Genética, da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Mestre em Genética.

**Aprovado em 15/09/2015**

**BANCA EXAMINADORA**

---

Dr. Valdir de Queiroz Balbino  
Universidade Federal de Pernambuco

---

Dr. Antonio Mauro Rezende  
Fundação Oswaldo Cruz de Pernambuco

---

Dr. Martin Alejandro Montes  
Universidade Federal Rural de Pernambuco

---

Dr. Carlos Alberto Santiago Figueirêdo Júnior  
Universidade Federal de Pernambuco

## **AGRADECIMENTOS**

Ao professor Valdir de Queiroz Balbino, por ter arriscado e me dado seu voto de confiança. Por ter aberto as portas para a minha vida científica que é bastante importante na minha vida. E ainda mais: por ter me feito enxergar o papel de um verdadeiro cientista. Por ter acompanhado cada passo e com sabedoria ter permitido a desenvoltura e a aplicabilidade de algo que gosto muito de fazer. Sou muito grata por essa oportunidade de trabalhar ao seu lado.

Ao Centro de Ciências Biológicas, ao Departamento de Genética e ao Programa de Pós-Graduação em Genética, pela oportunidade de uma realização na minha vida e de convivência com discentes e docentes que tem me levado a importantes reflexões e apontado caminhos de conhecimento os quais sozinha talvez não seria possível descobrir ou alcançá-los. Sinto-me honrada de ter feito parte de vós um dia.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo fornecimento da bolsa de mestrado.

Aos meus pais, Aluizio Francisco da Silva e Josefa Marcelino dos Santos Silva. Mesmo muitas vezes sendo difícil me compreenderem, têm respeitado minhas decisões e também não tem poupado seus esforços para me ajudarem da maneira que podem. Agradeço pelo ombro amigo que nos meus momentos melancólicos me deu suporte e por, junto a mim, ansiarem minhas conquistas como se fossem suas próprias. Isso foi bastante importante para mim.

A minha irmã Priscila Marcelino dos Santos Silva que, apesar de não estarmos juntas na nossa vida profissional como premeditado pelos nossos pais e achávamos que iria ser - tomamos rumos diferentes (foi melhor assim!), tem feito

parte de cada momento meu. Nossas conversas e os momentos de diversões preenchem bastante meus dias. Desejo muito sucesso para você.

Aos meus demais parentes, apesar de muitas vezes eu sentir vocês tão distantes de mim. Só pelo fato de saber que torcem por meu sucesso e de me falar em nossos breves encontros que se sentem felizes com minhas realizações, faz-me superar vossas ausências, faz-me sorrir ao lembrar-se disso. Isso também foi bastante importante.

Ao meu instrumento musical, que permite pousar minhas mãos sobre ele e assim me dá liberdade para usar e abusar da imaginação e sonhar livremente - isso faz minha alma bastante leve... Afinal não vivo apenas de ciência.

Aos meus companheiros do Laboratório, os quais compartilharam comigo um caminho cheio de desafios. Sobretudo àqueles que participaram mais diretamente deste trabalho com suas experiências na área de entomologia (as quais não possuo) e foram indispensáveis para tornar possível a realização deste trabalho. E aos demais, por compartilharem comigo vossas experiências e deixarem nosso dia-a-dia sempre animado, tornando agradável o nosso ambiente de trabalho - o que é bastante importante para nossas realizações por mais que tenhamos paixão pelo que fazemos. Vocês fazem esses dias no laboratório serem lembrados com mais ênfase.

Aos meus amigos (cujos nomes não estou citando para não correr o risco de esquecer alguém), com os quais tenho compartilhado muitos momentos. Por me fazerem recordar do que há de bom na vida, de bons acontecimentos e acreditar que não vale a pena lembrar-se das dificuldades que passamos e que finalmente também já ultrapassamos, de modo que as veja como uma experiência para

enxergar o valor das coisas que muitas vezes achamos tão fúteis, simples e insignificantes.

## RESUMO

A falta de padronizações sistemáticas deixa questões em aberto na identificação das espécies crípticas dos complexos *Lutzomyia longipalpis* e *L. umbratilis*, que incluem vetores dos agentes etiológicos das leishmanioses no Brasil. As metodologias de identificação taxonômica têm contribuído para um maior conhecimento da diversidade biológica. Porém o *status* taxonômico destes dois grupos de flebotomíneos ainda permanece incerto no Brasil. A filogenética molecular é uma ferramenta essencial para investigação de espécies crípticas. Para inferência filogenética com dados de sequências de nucleotídeos existem métodos estatísticos de inferência, inclusive métodos baseados em probabilidade. Estão disponíveis em duas abordagens: Máxima Verossimilhança (MV) e Bayesiana (BE). Neste trabalho foi avaliada a adequação destas abordagens na investigação das referidas espécies crípticas utilizando um marcador mitocondrial (citocromo oxidase I – COI) e um nuclear (gene *period* – *per*). Para o complexo *L. umbratilis*, o método BE apresentou árvores filogenéticas com maiores probabilidades, ou seja, com melhores resultados em relação às árvores inferidas por MV, principalmente com a utilização do marcador COI. Já para o complexo de espécies *L. longipalpis*, o método BE apresentou a melhor árvore filogenética em relação ao método MV e apenas utilizando o marcador nuclear *period*. Tendo alcançado melhores resultados com o marcador mitocondrial COI, para *L. umbratilis*, e *period*, para *L. longipalpis*, e com a abordagem BE, conclui-se que esta abordagem juntamente com os respectivos marcadores teve maior adequação do que a abordagem MV.

**Palavras-chave:** espécies crípticas; *Lutzomyia longipalpis*; *Lutzomyia umbratilis*; filogenética; métodos probabilísticos.

## ABSTRACT

The lack of systematic standardization leaves open questions in the identification of cryptic species complex of *Lutzomyia longipalpis* and *L. umbratilis*, including vectors of etiological agents of leishmaniasis in Brazil. The taxonomic identification methodologies have contributed to a better understanding of biological diversity. But the taxonomic status of these two sandflies groups remains uncertain in Brazil. Molecular phylogenetics is an essential tool for research cryptic species. For phylogenetic inference nucleotide sequence data are statistical inference methods, including methods based on probability. two approaches are available: Maximum Likelihood (ML) and Bayesian (BE). This work evaluated the adequacy of these approaches in the investigation of these cryptic species using a mitochondrial marker (cytochrome oxidase I - COI) and nuclear (gene period - per). For the complex *L. umbratilis* the method presented phylogenetic trees BE with higher probability, that is, with better results in respect of trees inferred by MV, especially with the use of the COI marker. As for the species complex *L. longipalpis*, the BE method presented the best phylogenetic tree in relation to the MV method only using the nuclear period marker. Having achieved better results with the mitochondrial marker COI to *L. umbratilis*, and period to *L. longipalpis*, and with BE approach, it is concluded that this approach together with their markers had better match than the MV approach.

**Keywords:** cryptic species; *Lutzomyia longipalpis*; *Lutzomyia umbratilis*; phylogenetic; probabilistic methods.

## LISTA DE ILUSTRAÇÕES

- Figura 1 - Fluxograma do processo de decisão na descrição de uma espécie. Fonte original: Doyen e Slobodchikoff (1974), modificada por Winston (1999). 34
- Figura 2 - Ilustração do nó  $v$ , seus descendentes  $u_1$  e  $u_2$ , os caracteres  $x_i$  e os caracteres restritos por  $x_i^v$  para a sub-árvore enraizada por  $v$ . 48
- Figura 3 - Polarização única alcançada em todas as árvores filogenéticas de *L. umbratilis* resultantes que estavam entre as polarizações esperadas. 84
- Figura 4 - Polarização única alcançada na única árvore, de todas as árvores filogenéticas de *L. longipalpis* resultantes, que estava entre as polarizações esperadas. 85

## LISTA DE TABELAS

Tabela 1 - Categorização das proporções das 72 sequências do marcador <i>per</i> e COI referentes a <i>L. umbratilis</i> quanto ao gênero e área geográfica.	68
Tabela 2 - Categorização das proporções das 73 sequências do marcador <i>per</i> e COI referentes a <i>L. longipalpis</i> quanto ao gênero e área geográfica.	69
Tabela 3 – Grupos externos utilizados para cada marcador ( <i>per</i> e COI) referentes a <i>L. umbratilis</i> na primeira busca de grupos externos mais adequados.	70
Tabela 4 – Grupos externos utilizados para cada marcador ( <i>per</i> e COI) referentes a <i>L. longipalpis</i> na primeira busca de grupos externos mais adequado.	70
Tabela 5 – Grupos externos utilizados para cada marcador ( <i>per</i> e COI) referentes a <i>L. umbratilis</i> (e <i>L. longipalpis</i> ) na segunda busca de grupos externos mais adequado.	71
Tabela 6 – Grupos externos utilizados para cada marcador ( <i>per</i> e COI) referentes a <i>L. umbratilis</i> (e <i>L. longipalpis</i> ) na terceira busca de grupos externos mais adequado.	72
Tabela 7 - Oito possíveis combinações de táxons de grupo externo na primeira busca de grupo externo referentes a <i>L. umbratilis</i> utilizando marcador <i>per</i> .	73
Tabela 8 - Possíveis combinações de grupos externos na primeira busca de grupo externos referentes a <i>L. umbratilis</i> utilizando marcador COI.	74

Tabela 9 - Oito possíveis combinações de táxons de grupo externo na primeira busca de grupo externo referentes à <i>L. longipalpis</i> utilizando marcador per.	75
Tabela 10 - Dezesesseis possíveis combinações de táxons de grupo externo na primeira busca de grupo externo referentes a <i>L. longipalpis</i> utilizando marcador COI.	76
Tabela 11 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos <i>fasta</i> de <i>L. umbratilis</i> na primeira busca de grupo externo.	78
Tabela 12 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos <i>fasta</i> de <i>L. umbratilis</i> na segunda busca de grupo externo.	79
Tabela 13 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos <i>fasta</i> de <i>L. umbratilis</i> na terceira busca de grupo externo.	80
Tabela 14 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos <i>fasta</i> de <i>L. longipalpis</i> na primeira busca de grupo externo.	81
Tabela 15 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos <i>fasta</i> de <i>L. longipalpis</i> na segunda busca de grupo externo.	82
Tabela 16 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos <i>fasta</i> de <i>L. longipalpis</i> na terceira busca de grupo externo.	83

Tabela 17 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da primeira busca de grupo externo para <i>L. umbratilis</i> .	86
Tabela 18 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da segunda busca de grupo externo para <i>L. umbratilis</i> .	88
Tabela 19 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da terceira busca de grupo externo para <i>L. umbratilis</i> .	89
Tabela 20 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da primeira busca de grupo externo para <i>L. longipalpis</i> .	91
Tabela 21 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da segunda busca de grupo externo para <i>L. longipalpis</i> .	93
Tabela 22 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da terceira busca de grupo externo para <i>L. longipalpis</i> .	94

Tabela 23 - Análise da primeira busca de grupo externo para complexo <i>L. umbratilis</i> utilizando marcador COI.	96
Tabela 24 - Análise da primeira busca de grupo externo para complexo <i>L. umbratilis</i> utilizando marcador <i>per.</i>	99
Tabela 25 - Análise da terceira busca de grupo externo para complexo <i>L. umbratilis</i> utilizando marcador COI.	102
Tabela 26 - Análise da terceira busca de grupo externo para complexo <i>L. umbratilis</i> utilizando marcador <i>per.</i>	104
Tabela 27 - Análise da primeira busca de grupo externo para complexo <i>L. longipalpis</i> utilizando marcador COI.	106
Tabela 28 - Análise da primeira busca de grupo externo para complexo <i>L. longipalpis</i> utilizando marcador <i>per.</i>	107
Tabela 29 - Análise da terceira busca de grupo externo para complexo <i>L. longipalpis</i> utilizando marcador COI.	109
Tabela 30 - Análise da terceira busca de grupo externo para complexo <i>L. longipalpis</i> utilizando marcador <i>per.</i>	110
Tabela 31 - Modelos evolutivos das sequências concatenadas de <i>L. umbratilis</i> e valores de suporte de confiança dos dois principais clados com importância biológica apresentadas nas árvores filogenéticas inferidas utilizando esses dados concatenados.	112
Tabela 32 - Modelos evolutivos das sequências concatenadas de <i>L. longipalpis</i> e valores de suporte de confiança dos dois principais clados com importância biológica apresentadas nas árvores filogenéticas inferidas utilizando esses dados concatenados.	112

Tabela 33 - Valores de suporte de confiança dos dois principais clados com importância biológica para cada uma das árvores inferidas por dados concatenados ou isolados do complexo <i>L. umbratilis</i> .	113
Tabela 34 - Valores <i>Split</i> , nodal e de clados congruentes entre duas árvores de <i>L. umbratilis</i> cada das quais inferida por um método diferente (BE e MV) e utilizando mesmo marcador (ou mesma concatenação de marcadores) e grupo externo (quando tinham).	115
Tabela 35 - Valores <i>Split</i> , nodal e de clados congruentes entre duas árvores de <i>L. umbratilis</i> cada das quais inferida utilizando marcadores diferentes (COI e <i>per</i> ) e ambas inferidas pelo mesmo método (ambas inferidas por MV ou ambas por BE).	117
Tabela 36 - Valores <i>Split</i> , nodal e de clados congruentes entre duas árvores de <i>L. umbratilis</i> cada das quais inferidas por MV utilizando marcador COI e por BE utilizando marcador <i>per</i> , e também entre duas árvores inferidas por MV utilizando marcador <i>per</i> e por BE utilizando marcador COI.	118
Tabela 37 - Valores de suporte de confiança dos dois principais clados com importância biológica para cada uma das árvores inferidas com dados concatenados e isolados do complexo <i>L. longipalpis</i> .	119
Tabela 38 - Valores <i>Split</i> , nodal e de clados congruentes entre duas árvores de <i>L. longipalpis</i> cada das quais inferida por um método diferente (BE e MV) e utilizando mesmo marcador (ou mesma concatenação de marcadores) e grupo externo (quando tinham).	122

Tabela 39 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. longipalpis* cada das quais inferida utilizando marcadores diferentes (COI e *per*) e ambas inferidas pelo mesmo método (ambas inferidas por MV ou ambas por BE). 123

Tabela 40 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. longipalpis* cada das quais inferidas por MV utilizando marcador COI e por BE utilizando marcador *per*, e também entre duas árvores inferidas por MV utilizando marcador *per* e por BE utilizando marcador COI. 124

## LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

BE	Bayesiana – relativo à Bayes
BIC	<i>Bayesian Information Criterion</i> (em português, Critério de informação Bayesiana)
CAR	População do município de Caririaçu (Ceará, Brasil)
COI	Gene Citocromo c oxidase I
D	Dados de sequencia de nucleotídeos
FMAN	Fêmeas do município de Manacapuru (Amazonas, Brasil)
FREC	Fêmeas do município de Recife (Pernambuco, Brasil)
FRIP	Fêmeas do município de Rio Preto da Eva (Amazonas, Brasil)
LABBE	Laboratório de Bioinformática de Biologia Evolutiva/UFPE
Log	Logaritmo
MAN	População do município de Manacapuru (Amazonas, Brasil)
MCMC	Cadeia de Markov Monte Carlo
MMAN	Machos do município de Manacapuru (Amazonas, Brasil)
MREC	Machos do município de Recife (Pernambuco, Brasil)
MRIP	Machos do município de Rio Preto da Eva (Amazonas, Brasil)
MV, ML	Máxima verossimilhança (do inglês, <i>maximum likelihood</i> )
NCBI	<i>National Center for Biotechnology Information</i> (em português, Centro Nacional de Informações sobre Biotecnologia)
Per	Gene <i>Period</i>
REC	População do município de Recife (Pernambuco, Brasil)
RIP	População do município de Rio Preto da Eva (Amazonas, Brasil)

- SOB* População do município de Sobral (Ceará, Brasil)
- TOPD/FMST* *Topological Distance program / From Multiple To Single program*  
(em português, Programa de distância topológica / programa  
Múltiplo a Único)
- Θ Parâmetros da árvore filogenética: topologia de árvore, comprimento  
do ramo e modelo evolutivo.

## Sumário

<b>1. INTRODUÇÃO</b>	22
<b>2. REVISÃO DA LITERATURA</b>	25
2.1. ESPÉCIES CRÍPTICAS	25
2.2. <i>LUTZOMYIA UMBRATILIS</i> : UM COMPLEXO DE ESPÉCIES CRÍPTICAS NO TERRITÓRIO BRASILEIRO	25
2.3. <i>LUTZOMYIA LONGIPALPIS</i> : UM SEGUNDO COMPLEXO DE ESPÉCIES CRÍPTICAS NO TERRITÓRIO BRASILEIRO	27
2.4. IMPORTÂNCIA DA INVESTIGAÇÃO DE ESPÉCIES CRÍPTICAS	28
2.5. CLASSIFICAÇÃO DE ESPÉCIES	29
2.6. ESPÉCIE DENTRO DA ABORDAGEM TAXONÔMICA CLADÍSTICA	33
<b>2.6.1 A taxonomia dos flebotomíneos</b>	38
2.7. FILOGENÉTICA MOLECULAR	39
<b>2.7.1. Inferência e inferência estatística</b>	39
<b>2.7.2. Inferência filogenética</b>	40
2.7.2.1. Dados moleculares para investigação de espécies	41
2.7.2.2. Utilização de dados moleculares concatenados via particionamento	42
2.7.2.3. Métodos de inferência filogenética	43
2.7.2.4. Métodos de inferência filogenética baseados na análise de caracteres	44
2.8. MÉTODOS PROBABILÍSTICOS DE INFERÊNCIA FILOGENÉTICA	47

<b>2.8.1. Por abordagem de máxima verossimilhança</b>	47
<b>2.8.2. Por abordagem Bayesiana</b>	50
2.9. APLICAÇÃO DA ABORDAGEM MÁXIMA VEROSSIMILHANÇA EM INVESTIGAÇÕES DE POSSÍVEIS ESPÉCIES CRÍPTICAS NO BRASIL	54
2.10. APLICAÇÃO DA ABORDAGEM BAYESIANA EM INVESTIGAÇÕES DE POSSÍVEIS ESPÉCIES CRÍPTICAS NO BRASIL	55
2.11. AVALIAÇÃO METODOLÓGICA DE MÉTODOS DE INFERÊNCIA FILOGENÉTICA	55
<b>3. OBJETIVOS</b>	57
3.1. OBJETIVO GERAL	57
3.2. OBJETIVOS ESPECÍFICOS	57
<b>4. MATERIAL E MÉTODOS</b>	59
4.1. COLETA E ORGANIZAÇÃO DOS DADOS DE SEQUÊNCIAS DE NUCLEOTÍDEOS NUCLEAR E MITOCONDRIAL DE <i>L. UMBRATILIS</i> E <i>L. LONGIPALPIS</i>	59
4.2. OBTENÇÃO DE ÁRVORES FILOGENÉTICAS A PARTIR DOS DADOS DE SEQUENCIAS DOS MARCADORES <i>COI</i> OU <i>PER</i> DE <i>L. UMBRATILIS</i> E <i>L. LONGIPALPIS</i>	60
<b>4.2.1. Definição de grupos externos para a inferência de árvores filogenéticas</b>	60
<b>4.2.2. Alinhamento das sequências de nucleotídeos</b>	62
<b>4.2.3. Escolha do modelo evolutivo probabilístico</b>	63
<b>4.2.4. Obtenção das árvores filogenéticas por métodos probabilísticos</b>	63
4.3. OBTENÇÃO DE ÁRVORE FILOGENÉTICA A PARTIR DE DADOS	64

DE SEQUÊNCIAS CONCATENADAS DOS MARCADORES *COI* E *PER*  
DE *L. UMBRATILIS* E *L. LONGIPALPIS*

4.4. COMPARAÇÃO DO DESEMPENHO DOS MÉTODOS DE 66  
INFERÊNCIA A PARTIR DE ANÁLISE DAS ÁRVORES INFERIDAS POR  
DADOS CONCATENADOS E POR DADOS DE CADA MARCADOR  
ISOLADO

**5. RESULTADOS** 68

5.1. DADOS DE SEQUÊNCIAS DE *L. UMBRATILIS* E *L. LONGIPALPIS* 68  
ORGANIZADOS NOS ARQUIVOS FASTA

5.2. DEFINIÇÃO DE GRUPOS EXTERNOS PARA INFERÊNCIA DE 69  
ÁRVORES FILOGENÉTICAS DOS COMPLEXOS *L. UMBRATILIS* E *L.*  
*LONGIPALPIS*

5.3. ARQUIVOS COM SEQUÊNCIAS ALINHADAS 73

5.4. MODELOS EVOLUTIVOS PROBABILÍSTICOS 77

5.5. ÁRVORES FILOGENÉTICAS DE *L. UMBRATILIS* E *L.* 84  
*LONGIPALPIS*

5.6. RESULTADO DA ANÁLISE DAS ÁRVORES INFERIDAS A 95  
PARTIR DE DADOS DE CADA MARCADOR ISOLADO OBSERVANDO  
VALORES DE SUPORTE DE CADA POLARIDADE

5.7. DADOS CONCATENADOS DE SEQUÊNCIAS DOS 111  
MARCADORES *COI* E *PER* DE *L. UMBRATILIS* E *L. LONGIPALPIS* E  
ÁRVORES FILOGENÉTICAS OBTIDAS A PARTIR DESTES DADOS

5.8. ANÁLISES DE ÁRVORES INFERIDAS POR DADOS 113  
CONCATENADOS *VERSUS* AS POR DADOS DE CADA MARCADOR  
ISOLADO OBSERVANDO OS VALORES DE SUPORTE DE CADA

POLARIDADE	
5.9. ANÁLISES DE ÁRVORES INFERIDAS A PARTIR DE DADOS CONCATENADOS E DE DADOS DE CADA MARCADOR ISOLADO UTILIZANDO O TESTE DE CONGRUÊNCIA ENTRE TOPOLOGIAS	116
<b>6. DISCUSSÃO</b>	127
<b>7. CONCLUSÕES</b>	133
<b>8. REFERÊNCIAS BIBLIOGRÁFICAS</b>	134
<b>APÊNDICE A - FLUXOGRAMA DA METODOLOGIA UTILIZADA NESTE TRABALHO PARA CADA UM DOS COMPLEXOS DE ESPÉCIES CRÍPTICAS</b>	146
<b>APÊNDICE B: FLUXOGRAMA DA BUSCA DE GRUPO EXTERNO PARA CADA UM DOS COMPLEXOS DE ESPÉCIES</b>	149
<b>APÊNDICE C - ÁREAS E COORDENADAS GEOGRÁFICAS E ASPECTOS BIOGEOGRÁFICOS CORRESPONDENTES ÀS SEQUÊNCIAS DAS ESPÉCIES CRÍPTICAS <i>L. UMBRATILIS</i> E <i>L. LONGIPALPIS</i></b>	150
<b>APÊNDICE D - ESTRUTURAS DE UMA ÁRVORE FILOGENÉTICA</b>	152
<b>APÊNDICE E - CLADOS IMPORTANTES NO SENTIDO BIOLÓGICO - POLARIZAÇÃO ESPERADA COM AS ÁRVORES ENRAIZADAS VIA GRUPO EXTERNO</b>	153
<b>APÊNDICE F - MELHORES ÁRVORES DO COMPLEXO <i>L. UMBRATILIS</i></b>	158
<b>APÊNDICE G - MELHORES ÁRVORES DO COMPLEXO <i>L. LONGIPALPIS</i></b>	182

## 1. INTRODUÇÃO

O reconhecimento das espécies crípticas tem sido um desafio constante na taxonomia. A falta de análises ou de padronizações sistemáticas deixa muitas questões em aberto, como por exemplo, se as espécies crípticas são mais comuns em particulares habitats, latitudes ou grupos taxonômicos. A investigação das espécies crípticas tem profunda implicação para a teoria da evolução, biogeografia, planejamento de conservação e epidemiologia.

Muitas espécies com importância epidemiológica reconhecida possuem o status taxonômico de complexos de espécies crípticas irmãs (e.g. insetos vetores que transmitem a Doença de Chagas, malária e leishmanioses visceral e tegumentar). As espécies de um complexo podem diferir quanto a determinadas características biológicas, interferindo na disseminação e no controle de patologias. Os dípteros flebotomíneos *Lutzomyia longipalpis* e *Lutzomyia umbratilis*, importantes vetores de agentes etiológicos das leishmanioses no Brasil, possuem status taxonômico controverso e podem ser considerados como exemplos de complexos de espécies. Dentre as fontes de evidências que suportam essa possibilidade, destacam-se aquelas advindas do uso de marcadores moleculares, que vêm sendo frequentemente utilizados no reconhecimento de espécies irmãs simpátricas e alopátricas.

As evidências fornecidas pelos marcadores moleculares, que podem ser representadas através de ferramentas de filogenética molecular, facilitam o reconhecimento das espécies com status taxonômicos controverso, contribuindo para a identificação adequada de cada uma delas. Atualmente os marcadores moleculares, principalmente os marcadores de DNA (mitocondriais e nucleares), são

as fontes de dados mais utilizadas na filogenética molecular. Cada marcador pode apresentar padrões evolutivos diferentes do padrão de evolução do organismo estudado. Associados aos padrões de evolução intrínsecos de cada marcador, os métodos de filogenética molecular disponíveis até o momento possuem diferentes graus de eficiência (adequação) devido à natureza de cada método (quantitativos, baseados em matrizes de distâncias; e àqueles baseados na análise de caracteres – parcimônia e probabilidades).

Apesar da existência desses vários métodos de inferência filogenética e de marcadores genéticos empregados no reconhecimento de espécies crípticas terem contribuído para um maior conhecimento da diversidade biológica nos últimos anos, o status taxonômico dos flebotomíneos *Lutzomyia longipalpis sensu lato* e *L. umbratilis sensu lato* ainda permanece incerto. Uma das razões disto pode corresponder à falta de conhecimento de quais métodos e marcadores se adaptam mais na investigação especificamente de cada complexo de espécies crípticas.

Devido à diversidade de classes de marcadores de DNA e de métodos de inferência filogenética, o presente estudo propõe a sistematização da investigação de espécies crípticas, avaliando a adequação de diferentes métodos de inferência filogenética e marcadores genéticos, utilizando como modelo de estudo os complexos de espécies dos flebotomíneos *Lutzomyia longipalpis* e *L. umbratilis*. Os métodos de inferência filogenética incluídos neste estudo foram os probabilísticos e os marcadores genéticos foram citocromo oxidase I (*COI*) e gene *period* (*per*), uma vez que os métodos baseados na análise de caracteres são os mais eficientes para a detecção de espécies crípticas, sendo que os mais utilizados são os métodos probabilísticos de inferência filogenética (Máxima Verossimilhança e Inferência Bayesiana) e os marcadores mitocondrial citocromo oxidase I (*COI*) e nuclear gene

*period (per)* são os mais utilizados para investigação dos complexos *L. umbratilis* e *L. longipalpis*.

Espera-se que, futuramente, a partir deste estudo, possa-se definir uma alternativa metodológica eficiente para a investigação de espécies crípticas destes organismos.

## 2. REVISÃO DA LITERATURA

### 2.1. ESPÉCIES CRÍPTICAS

O sistema de classificação criado por Linneus (1735) que prioriza as características morfológicas ainda é o principal sistema de classificação das espécies, embora se saiba que uma grande quantidade da diversidade natural não é reconhecida por este tipo de abordagem (Bickford *et al.*, 2007). Espécies crípticas são morfológicamente indistinguíveis e, entretanto, possuem entre si diferenças genéticas, comportamentais e ecológicas.

O conceito de espécies crípticas mais correntemente utilizado é o de Bickford *et al.* (2006), que define espécies crípticas como sendo conjuntos de duas ou mais espécies que são classificadas como uma única espécie nominal em função da semelhança morfológica entre elas. Às mesmas podem ser atribuídas a expressão “espécies irmãs” (Sáez & Lozano, 2005). No aspecto etimológico, entretanto, a expressão “espécies irmãs” implica uma relação taxonômica mais próxima entre as espécies consideradas do que àquele referente às espécies crípticas (Knowlton, 1986).

### 2.2. *LUTZOMYIA UMBRATILIS*: UM COMPLEXO DE ESPÉCIES CRÍPTICAS NO TERRITÓRIO BRASILEIRO

*Lutzomyia umbratilis* (Diptera: *Psychodidae*), principal vetor da *Leishmania guyanensis*, agente etiológico da Leishmaniose Tegumentar Americana (LTA), pode ser encontrado na região amazônica (Young *et al.*, 1994) e no Nordeste, ou mais

especificamente, no Estado de Pernambuco (Balbino *et al.*, 2001). Um estudo utilizando genes mitocondriais sugere a presença de espécies crípticas (Scarpassa & Alencar, 2012). Em *L. umbratilis*, há diferenças biológicas entre populações: variações na susceptibilidade deste vetor à infecção por *Leishmania* (Arias & Freitas, 1978) e diferenças bionômicas significativas (*e.g.* ciclo de vida, fecundidade, fertilidade, grau de emergência e longevidade dos adultos entre as populações) (Justiniano *et al.*, 2004). Até o momento, as possíveis espécies pertencentes a este complexo ainda não foram descritas.

Ainda há um reduzido número de trabalhos de genética de populações de *L. umbratilis*, porém estes mostram evidências de espécies crípticas. Quanto aos aspectos ecológicos e ao isolamento reprodutivo entre as espécies do complexo, ainda não foram encontrados trabalhos. Quanto à morfologia, não foram conduzidos trabalhos conclusivos de morfometria, limitando-se a descrições de diferenças no comprimento das cerdas no quarto estado larval e no tamanho dos espinhos de armação do átrio genital das fêmeas (Scarpassa & Alencar, 2012). Também não foram encontrados trabalhos voltados à caracterização dos sons copulatórios e de feromônios, mas quanto às isoenzimas não foram identificadas diferenças significativas entre as populações (Scarpassa & Alencar, 2012). Quanto a genética molecular de populações, variações em genes mitocondriais mostraram indícios da existência de espécies crípticas, porém não foram encontrados trabalhos com genes autossômicos (*e.g. period* e *cacophony*), à semelhança do que se observa em *L. longipalpis* (Bauzer *et al.*, 2002-b).

### 2.3. *LUTZOMYIA LONGIPALPIS*: UM SEGUNDO COMPLEXO DE ESPÉCIES CRÍPTICAS NO TERRITÓRIO BRASILEIRO

*Lutzomyia longipalpis* s.l. (Diptera: Psychodidae) é o principal vetor da *Leishmania infantum chagasi* nas Américas, agente causador da Leishmaniose Visceral Americana (LVA). O status taxonômico deste organismo tem sido questionado desde 1969, quando Mangabeira (1969) descreveu divergências morfológicas entre machos de diferentes localidades do Brasil, sugerindo a possibilidade da existência de um complexo de espécies. Este autor propôs que os dois fenótipos identificados (1S – machos contendo apenas manchas pálidas no quarto tergito abdominal; e 2S, machos contendo um par de manchas pálidas nas terceira e quarta tergitos abdominais) poderiam representar duas espécies distintas.

Apesar dos vários estudos realizados desde a publicação das observações de Mangabeira (1969), ainda não se reconheceu no Brasil nenhuma espécie deste complexo por ainda não haver um consenso acerca do status taxonômico das populações brasileiras (Maigon *et al.*, 2003; Bauzer *et al.*, 2007). Em populações da região Nordeste, por exemplo, existem evidências que apontam a favor da uniformidade taxonômica de *L. longipalpis* (e.g. de Queiroz Balbino *et al.*, 2006), também há evidências de que muitas populações brasileiras correspondem a espécies crípticas. Neste caso, são consideradas como espécies crípticas porém pertencentes a um complexo de espécies ainda incipiente (Bauzer *et al.*, 2007), não sendo assim possível individualizar as espécies.

No território brasileiro, existem evidências da ocorrência de espécies crípticas do complexo *Lutzomyia longipalpis* já descritas na literatura: diferenças ecológicas (Mangabeira, 1969); isolamento reprodutivo entre populações alopátricas

e simpátricas (Ward *et al.*, 1983); populações reprodutivamente isoladas produzem feromônios diferentes (Ward *et al.*, 1988). Quanto à morfologia, não há relatos de diferenças significativas na morfometria (Azevedo *et al.*, 2000), apesar de terem sido apontadas algumas diferenças tênues nas espermatecas das fêmeas (Mangabeira, 1969) e nos padrões das pintas dorsais nos machos (Mangabeira, 1969).

Em relação aos machos, o consenso inicial era de que as diferenças nos padrões de manchas não estariam correlacionadas com o tipo de feromônio secretado (Ward *et al.*, 1988). Quanto aos sons copulatórios, foram percebidas diferenças notáveis entre espécimes de populações simpátricas e alopátricas (Souza *et al.*, 2002). Quanto às diferenças bioquímicas, não houve diferença significativa nos padrões de diversidade genética de locos isozímicos (Mukhopadhyay *et al.*, 1997; Mukhopadhyay *et al.*, 1998a; Mukhopadhyay *et al.*, 1998b; Mutebi *et al.*, 1999), porém existem diferenças nos feromônios (Hamilton *et al.*, 2004; Hamilton *et al.*, 2005), inclusive correlacionadas com divergências genéticas do gene *period* (Souza *et al.*, 2004). Quanto à genética molecular de populações, variações dos genes mitocondriais (Hodgkinson *et al.*, 2003) e de perfis RAPD-PCR (de Queiroz Balbino *et al.*, 2006) não permitiram concluir que se trata de espécies crípticas, enquanto que as variações de genes autossômicos: microssatélites (Maigon *et al.*, 2003), *period* gene (Bauzer *et al.*, 2002-a), (Bauzer *et al.*, 2002-b) e do gene *cacophony* (Bottecchia *et al.*, 2004) foram indicativas da existência de espécies crípticas.

#### 2.4. IMPORTÂNCIA DA INVESTIGAÇÃO DE ESPÉCIES CRÍPTICAS

O reconhecimento das espécies crípticas é indispensável para a quantificação da diversidade biológica. A correta identificação das espécies crípticas

é importante para a compreensão de seu possível papel ecológico e de outras diferenças biológicas, apresentando um papel crucial para planejamentos na biologia da conservação, no controle de pragas, na pesca e no controle de doenças, além de ser um passo crucial para a identificação de espécies transmissoras de patógenos (Velzen *et al.*, 2012).

Segundo Marcondes (1998), os membros de um complexo de espécies hematófagas podem diferir quanto às preferências alimentares ou quanto à capacidade vetorial, como foi descrito para *Anopheles gambiae*, vetor de *Plasmodium falciparum*, agente etiológico da malária no continente africano (White, 1974). O fluxo gênico entre espécies ou entre diferentes unidades evolutivas podem ter relevantes consequências epidemiológicas, uma vez que pode facilitar a transferência interespecífica de genes ou de alelos epidemiologicamente importantes e assim alterar os padrões de transmissão das doenças (Mazzoni *et al.*, 2008)

Espécies são um elemento fundamental da biologia, ecologia e conservação, de modo que sua delimitação correta é essencial (Petit & Excoffier, 2009), sendo assim indispensável a investigação de espécies para identificação correta de cada uma delas, inclusive a identificação das espécies crípticas (van Velzen *et al.*, 2012).

## 2.5. CLASSIFICAÇÃO DE ESPÉCIES

A área da biologia que trata da classificação e/ou do reconhecimento das espécies é a taxonomia. O sistema largamente utilizado remonta ao século XVIII, quando o botânico sueco Carl Linnaeus desenvolveu uma maneira de nomear e classificar os seres vivos. Ele usou um sistema hierárquico para organizar a natureza

inclusive os seres vivos e não como uma simples lista: sendo uma das categorias correspondentes a menor unidade – a espécie (Knapp, 2010).

Na biologia moderna (antes tinha a teoria essencialismo baseada na lógica de Aristóteles, também chamada de tipologia que é uma escola do feneticismo) há duas linhagens – fenéticas e filogenéticas - e três escolas:

**1. Fenética:** na biologia moderna, fenética corresponde à “taxonomia numérica” apenas, escola defendida principalmente por Sneath e Sokal, 1973, e não mais ao essencialismo. Neste método não é necessário saber de evolução. Seus princípios são basicamente:

- ✓ Dados utilizados para classificação: grande número de caracteres de todas as partes do corpo dos organismos analisados e de todo o ciclo vital;
- ✓ Todos os caracteres tem a mesma importância na classificação;
  - ✓ Exemplo de caracteres fenéticos: morfológicos; fisiológicos; ecológicos; etológicos; anatômicos; citológicos;
- ✓ Similaridade total é a soma de todas as similaridades de todos os caracteres;
- ✓ Está livre de inferência genealógica;
- ✓ Para os feneticistas, é impossível levar em conta as classificações que expressam a filogenia, uma vez que há pouco conhecimento acerca de detalhes relativos à história evolutiva da maioria dos seres organismos.
- ✓ Excluem a classificação de informações filogenéticas.
- ✓ Enfoque empírico e operacional: empírico porque as decisões taxonômicas provêm da experiência, a partir da qual se define um critério observável que

pode ser aplicado; posteriormente, por técnicas de análise de agrupamentos ou da análise de ordenação.

**2. Filogenética:** levam-se em conta as relações evolutivas entre os organismos. Classificam as espécies de acordo com a recenticidade que compartilham com um ancestral comum.

✓ **Cladismo (sistemática filogenética) (Henning, 1968)** – os critérios básicos são:

- ✓ Cada grupo ou táxon a se formar deve ser monofilético (todos os seus membros compartilham entre si um ancestral comum mais recente do que com qualquer outro membro de algum outro grupo de igual categoria na hierarquia filogenética);
- ✓ Nos caracteres utilizados, pode-se determinar o estado primitivo (estado presente no ancestral comum mais recente), que se pode obter pelo estudo do grupo;
- ✓ Estabelecem-se as sequências das ramificações da árvore genealógica e as posições são relativas ao tempo dessas ramificações (com a ajuda dos caracteres elegidos ou fósseis);
- ✓ Os grupos formados se baseiam na posse de estados evolucionários em comum, no monofiletismo e na proximidade temporal;
- ✓ Rejeitam-se grupos parafiléticos e polifiléticos;
- ✓ Pode-se enquadrar a hierarquia filogenética na Hierarquia de Linneus, embora algumas vezes possuam mais categorias do que a de Linneus.

**3. Evolucionismo:** esta escola não pretende chegar a uma classificação que expresse a filogenia, mas que seja conseqüente com ela. Combinam critérios fenéticos com informações genealógicas:

- ✓ Filogenia;
- ✓ Incluem não apenas grupos monofiléticos mais também os parafiléticos (que são definidos feneticamente);
- ✓ Homogeneidade em cada táxon formado, a partir de similaridades.

Geralmente a classificação resultante das duas correntes da taxonomia, fenética e filogenética, coincidem. Há alguns casos, contudo, em que organismos feneticamente muito semelhantes têm um ancestral mais distante entre si do que com aquele que tem mais diferenças fenéticas, seja porque alguns dos descendentes de um ancestral comum enfrentaram maior quantidade de mudanças ou possível coevolução (Amorim, 2002). Então se vê a importância de, até mesmo no nível taxonômico de espécie, já serem consideradas as relações evolutivas (o que não ocorre na fenética), inclusive nos casos de espécies crípticas e politípicas. Isso é importante não só para delimitá-las, assim como também na busca de possíveis explicações de processos que modelaram essa história evolutiva (Amorim, 2002).

A taxonomia fenética, além de ignorar as relações evolutivas e classificar as espécies por semelhanças aparentes, sofre ainda de um problema de subjetividade: subjetividade de qual método estatístico de análise de matrizes de distância a ser utilizado (incoerência constante entre os taxonomistas); os diferentes métodos de cálculo de distância das matrizes de distância, uma vez que as diferentes medidas

de distância resultavam em hierarquias diferentes, e novamente uma grande incoerência na escolha do método por diferentes taxonomistas (Williams & Ebach, 2009).

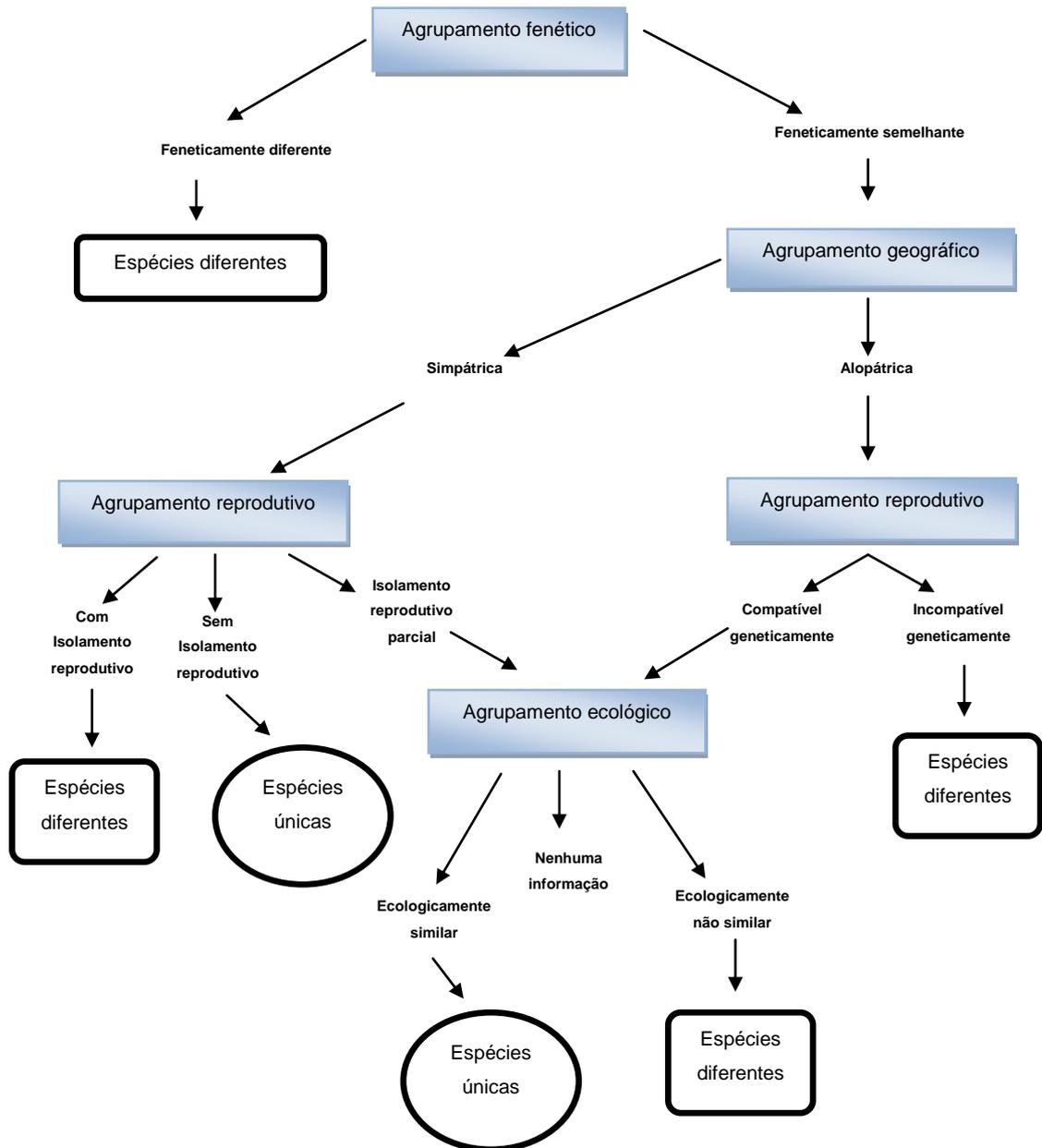
A taxonomia baseada nos conceitos da filogenética cladística está cada vez mais comum. Ela é conhecida como taxonomia cladística: os táxons são distribuídos em uma árvore evolucionária. A taxonomia cladística tem vantagens sobre a fenética, pois além de levar em conta os aspectos evolutivos, é mais objetiva; e sendo mais objetiva, tem sido o método de escolha na biologia contemporânea. Esta objetividade é decorrente da existência da hierarquia filogenética independente da técnica utilizada para descobri-la e é única para todos os seres vivos. Quando há discordância entre diferentes técnicas basta apelar para uma referência externa (Williams & Ebach, 2009).

A classificação biológica é hierárquica porque a evolução produziu um padrão hierárquico em rede divergente de semelhança entre os seres vivos. Sem dúvida que, desde as ideias evolutivas de Darwin, a natureza hierárquica das classificações biológicas tem feito parte das evidências da evolução (Coyne, 2009).

## 2.6. ESPÉCIE DENTRO DA ABORDAGEM TAXONÔMICA CLADÍSTICA

Doyen e Slobodchikoff (1974) (*apud* Wheeler & Meier, 2000) criaram um diagrama operacional de procedimentos lógicos para um estudioso descrever espécies, segundo a figura abaixo (modificada por Winston, 1999. Ver também em <http://iczn.org/content/how-can-i-describe-new-species>).

Figura 1 - Fluxograma do processo de decisão na descrição de uma espécie. Fonte original: Doyen e Slobodchikoff (1974), modificada por Winston (1999).



O primeiro passo, na maioria dos casos, é o reconhecimento de grupos fenéticos com base em diferenças morfológica ou físicas. Grupos feneticamente semelhantes ainda podem ser reconhecidos como espécies diferentes com base em distinções a partir de níveis geográficos, reprodutivos e ecológicos, já que são

barreiras que separam duas espécies irmãs no processo de cladogênese (de Queiroz, 1998; Aleixo, 2007).

Esse fluxograma reflete os conceitos biológico, ecológico (ecoespécies) e fenético (fenoespécies) de espécies. Esses são os mais utilizados pelos biólogos, já que em geral os membros de uma espécie diferem de outras espécies genética e ecologicamente, em comportamento e em morfologia (Wilkins, 2009). Abaixo está a definição de cada um desses conceitos:

- 1. Espécies biológicas** (Mayr, 1942; de Queiroz, 2007), usualmente chamadas de espécies genéticas: população mendeliana de organismos de reprodução sexuada, o cruzamento natural da população é isolado de outros grupos. Depende de mecanismos de isolamento reprodutivo (RIMs, do inglês *reproductive isolating mechanisms*) endógenos.
- 2. Ecoespécies (Ecótipos):** Uma linhagem (ou conjunto de linhagens intimamente relacionadas), que ocupa uma zona adaptativa minimamente diferente da de qualquer outra linhagem em seu alcance e que evolui separadamente a partir de todas as linhagens fora do seu alcance (Simpson, 1961; de Queiroz, 2007).
- 3. Fenoespécies:** Um conjunto de caracteres estatisticamente covariantes, um conceito de semelhança de família em que a posse da maior parte dos caracteres é necessária para inclusão em uma espécie, mas não todos. Uma classe de organismos que partilha mais de um conjunto de caracteres (Beckner, 1959; de Queiroz, 2007).

A busca inicial pelos grupos fenéticos é justificada pela seguinte suposição: é provável que organismos adaptados ao mesmo nicho sejam feneticamente

parecidos por compartilharem o mesmo conjunto de caracteres fenéticos para explorar os recursos ecológicos. Os ancestrais dos atuais membros da espécie se inter cruzam gerando membros com semelhanças genéticas e provavelmente fenéticas (Ridley, 2006).

O problema é que nem todas as espécies formam uma unidade fenética óbvia, como as espécies crípticas, que embora feneticamente muito semelhante muitas vezes possuem isolamento reprodutivo, ficando longe do conceito biológico de espécies, e também espécies polítípicas (são feneticamente muitos diferentes, porém não possuem isolamento reprodutivo uns com os outros) (Ridley, 2006).

Devido a essas causas, o conceito filogenético de espécies é também muito usado. Este conceito nasceu quando o uso do conceito biológico era considerado a única alternativa disponível para uma definição evolutiva de espécie (Cracraft, 1983). Ele estabelece que espécies devem ser delimitadas primeiramente com base em filogenias, ou seja, em hipóteses de ancestralidade (Cracraft, 1983; Cracraft *et al.*, 2004).

Remsen (2005) postula que, no conceito filogenético de espécie, independentemente do nível crítico de diagnose adotado (100%, 95% ou 75%), para que duas populações sejam consideradas espécies distintas, as definições sempre serão arbitrárias e em grande parte influenciadas pelo regime de amostragem do estudo taxonômico em questão.

Já *de* Queiroz (1998) propôs uma unificação conceitual da definição de espécie com pontos fundamentais de todos os conceitos de espécie baseados na teoria da evolução. Ele afirma que tanto o conceito biológico como o filogenético de espécies constituem unicamente "critérios" distintos e alternativos, e não "conceitos", para a definição de espécie dentro de um conceito unificado: Conceito Filético Geral

de Espécie. Segundo este conceito, as espécies são produtos de um processo temporal de cladogênese. E esse tempo contém as seguintes etapas sequências: a) separação de populações irmãs A e B por barreiras geográficas, ecológicas ou comportamentais (fluxo gênico interrompido); b) as duas populações irmãs já fixaram diferenças em vários caracteres, sendo plenamente diagnosticadas uma da outra, porém é provável que as populações A e B ainda mantenham a compatibilidade reprodutiva em função da sua relação filogenética ainda mais próxima; c) as diferenças iniciais acumuladas entre as populações A e B foram ampliadas para um conjunto ainda maior de caracteres, porém a compatibilidade reprodutiva ainda pode ser mantida; e d) tantas diferenças se acumularam entre as populações A e B, que elas desenvolveram uma barreira reprodutiva (ver em Aleixo, 2007).

Nos tempos “b” e “c” já é possível diagnosticar as populações A e B, ou seja, já são consideradas entidades distintas por ambos os conceitos, biológico e filogenético. A diferença é que no conceito biológico são denominadas de subespécies, enquanto que no conceito filogenético são denominadas de espécies. A denominação de subespécie é justificada pelo fato de que ainda podem ser capazes de entrecruzar e produzir descendentes férteis. No tempo “d”, os dois conceitos consideram as populações A e B como espécies distintas, pois A e B acumularam grande quantidade de diferenças, concretizando o isolamento reprodutivo, sendo praticamente impossível o fluxo gênico entre elas (Aleixo, 2007).

O processo de cladogênese é contínuo. Ao longo das gerações, entre os tempos “a” e “d”, a antiga população A, por exemplo, foi subdividida em várias subpopulações diagnosticáveis entre si como espécies distintas pelo conceito filogenético e como subespécies pelo conceito biológico. Novas populações vão surgindo enquanto outras são extintas. Como o processo de cladogênese é contínuo,

a definição destes limites pelos critérios de espécies, conceitos biológico e filogenético, é arbitrária, independentemente do critério adotado (Aleixo, 2007).

### **2.6.1. A taxonomia dos flebotomíneos**

Atualmente, no mundo, estão descritas mais de 900 espécies de flebotomíneos (Diptera, Psychodidae, Phlebotominae), das quais pelo menos 480 ocorrem no novo mundo (Galati, 2003). A sistemática desta subfamília tem sido baseada em critérios práticos, porém conservadores, que dividem os flebotomíneos em cinco gêneros (Lewis *et al.*, 1977), dos quais dois (*Phlebotomus* Rondani, 1840 e *Sergentomyia* França & Parrot, 1920) ocorrem no Velho Mundo e três (*Brumptomyia* França & Parrot, 1921; *Warileya* Hertig, 1948 e *Lutzomyia* França, 1924) estão representados no Novo Mundo.

Uma proposta para a sistemática dos Phlebotominae foi apresentada por Galati (1995; 2003), que utilizou o método cladístico para estudar os flebotomíneos americanos. Ou seja, é baseada em estudos filogenéticos e, conseqüentemente, em uma taxonomia mais refinada do grupo (Shimabukuro *et al.*, 2011). Apesar de a autora ter se utilizado de exclusivamente de caracteres morfológicos e/ou morfométricos para classificar as espécies, Galati reclassificou os flebotomíneos do Novo Mundo em 22 gêneros (Shimabukuro *et al.*, 2011). Independente das classificações, os estudos taxonômicos de flebotomíneos devem levar em conta a possibilidade da existência de complexos de espécies; a urgência de revisões taxonômicas, quer seja em nível de gênero como a realizada para *Psychodopygus* (Mangabeira, 1941), ou de um grupo de espécies, como o proposto para a série *shannoni* do gênero *Psathyromyia* (Barretto, 1962).

## 2.7. FILOGENÉTICA MOLECULAR

Como nenhum ser humano presenciou toda a história evolutiva de uma espécie ou de um grupo de espécies e os fósseis dos ancestrais são geralmente muito escassos, não é possível descrever a verdadeira história evolutiva. Mas dados cotidianos como caracteres fenéticos e caracteres de genes (genotípicos como sequência de nucleotídeos) podem oferecer evidências a partir das quais se pode então inferir a história evolutiva. E essa história evolutiva de um grupo de indivíduos (espécies ou agrupamentos taxonômicos superiores de organismos) inferida é o que é chamada filogenia (Felsenstein, 2004).

A filogenética molecular, ou sistemática molecular, pode ser definida como sendo o uso das estruturas moleculares (e.g. DNA e proteínas), para se ganhar informação sobre as relações evolutivas de um organismo (Hillis & Moritz, 1996). O resultado de uma análise filogenética molecular é expresso numa árvore filogenética (Page & Holmes, 1998). A árvore filogenética é, portanto, a principal forma de se representar a filogenia de um grupo (outra forma seria através de uma matriz numérica cujos números representariam o quanto cada uma das espécies do grupo analisado está próxima da outra) (Felsenstein, 2004).

### 2.7.1. Inferência e inferência estatística

Faz-se uma inferência quando se tem uma proposição que não é reconhecida diretamente como verdadeira, mas isto é feito através da relação dela com outras proposições que são verdadeiras (Hacking, 2011). Inferência estatística é um ramo da estatística cujo objetivo é fazer afirmações a partir de um conjunto de

valores representativo (amostra) sobre um universo. Tal tipo de afirmação deve sempre vir acompanhado de uma medida de precisão sobre sua veracidade. Para realizar este trabalho o estatístico coleta informações de dois tipos, experimentais (as amostras) e àquelas que obtêm na literatura. As duas principais escolas de inferência são a inferência frequentista (ou clássica) e a inferência Bayesiana (Evans & Rosenthal, 2010).

A inferência estatística é geralmente distinta da estatística descritiva. A descrição estatística pode ser vista como a simples apresentação dos fatos, nos quais o modelo de decisões feito pelo analista tem pouca influência. É natural que análises estatísticas avancem, indo da descrição para a inferência de padrões. Essa última tarefa depende do modelo usado e/ou criado pelo analista dos dados (Evans & Rosenthal, 2010).

### **2.7.2. Inferência filogenética**

Há vários tipos de dados a partir dos quais pode se inferir a filogenia. Mas os dados moleculares e, mais especificamente, dados de sequência de nucleotídeos têm uma significativa vantagem sobre os demais dados por serem mais exatos e reprodutíveis, por terem menor faixa de erro experimental e serem de fácil obtenção (Moret *et al.*, 2005). Graças a eles, tem sido possível esclarecer o mais correto posicionamento de espécies cuja morfologia torna difícil uma decisão sobre sua melhor classificação, como acontece nas espécies crípticas (*e.g.* complexos de espécies *L. longipalpis* e *L. umbratilis*). O sistema de classificação biológica permanece praticamente inalterado em sua estrutura desde o início da classificação de Linnaeus. Mas a maneira como as relações entre estes táxons são investigados

mudou drasticamente nas últimas décadas. Agora, é comum que os biólogos elaborem a classificação com base nos resultados de análises filogenéticas usando dados de sequência de DNA. A filogenética molecular tem impacto direto sobre as classificações taxonômicas, mesmo que não seja parte da taxonomia (Padial *et al.*, 2010).

Cada nucleotídeo em uma sequência de DNA é definido como um caractere enquanto os demais dados (*e.g.* dados morfológicos) devem primeiro ser categorizados em caracteres sendo assim sujeitos a problemas de interpretação e discretização, entre outros (Moret *et al.*, 2005). Para fazer inferência filogenética a partir de dados moleculares podem ser utilizados vários métodos com base estatística, chamados métodos de inferência filogenética, a partir de evidências de sequência de nucleotídeos ou aminoácidos.

#### 2.7.2.1. Dados moleculares para investigação de espécies

Atualmente, a identificação de flebotomíneos é essencialmente baseada em caracteres morfológicos (Young & Duncan, 1994; Galati, 2010). No entanto, a identificação morfológica requer habilidades e conhecimentos específicos. Além disso, a semelhança morfológica significativa entre algumas espécies pode causar dificuldades durante o processo de identificação (Scarpassa & Alencar 2012).

O DNA é uma ferramenta cada vez mais útil para estimar a diversidade de plantas, animais e insetos, inclusive de flebotomíneos, e para garantir a identificação rápida e precisa das espécies. Sequências de genes com alta taxa de evolução, como a sequência parcial da subunidade do gene mitocondrial citocromo oxidase I (COI) é uma das regiões mais utilizadas inclusive para identificação de

flebotomíneos (Contreras Gutiérrez *et al.*, 2014). Essas regiões são chamadas de sequências *barcoding* (sequências código de barra) (Contreras Gutiérrez *et al.*, 2014). Além de regiões de sequências mitocondriais, como COI, regiões de sequências nucleares também são utilizadas, como por exemplo, regiões do gene espaçador interno transcrito 2 (ITS2) (Depaquit *et al.*, 2000) e regiões do gene *period* (Bauzer *et al.*, 2002).

#### 2.7.2.2. Utilização de dados moleculares concatenados via particionamento

Além de se utilizar dados de um único gene ou haplótipo para inferência filogenética pode-se utilizar um conjunto deles simultaneamente para inferir uma única árvore representativa deste conjunto. A base para particionamento de dados é um grupo de genes ou sítios com características evolutivas semelhantes na mesma partição. Todos os sites na mesma partição são descritos usando o mesmo modelo evolutivo enquanto que diferentes partições usam diferentes modelos (Yang, 1996; Shapiro *et al.*, 2006).

As estratégias atuais para o particionamento de dados incluem particionamento de genes de acordo com a sua taxa relativa de substituição (Nishihara *et al.*, 2007) e separando as três posições de códons em genes codificantes em partições diferentes (Yang, 1996). O teste de razão de verossimilhança também tem sido utilizado para decidir se dois genes deve estar na mesma ou em diferentes partições (Leigh *et al.*, 2008).

### 2.7.2.3. Métodos de inferência filogenética

Cada conjunto de espécies tem apenas uma filogenia, ou seja, alguns dos caracteres considerados são enganosos. Um conjunto de espécies não podem ter múltiplas filogenias. Pode-se fazer a analogia com uma família da espécie humana que não pode ter mais de uma genealogia. Se houver duas hipóteses de genealogia, espera-se que pelo menos uma delas deve estar equivocada (Henning, 1966).

A técnica de cladística de inferência filogenética, ou reconstrução filogenética, tenta resolver este problema basicamente da seguinte forma (Henning, 1966):

- ✓ Distingue caracteres fontes e interpretações confiáveis e inconfiáveis;
- ✓ Os caracteres concordantes devem ser concordantes na filogenia, enquanto que os inconfiáveis devem ser descartados;
- ✓ As etapas para tal distinção consistem em: a) distinguir homologias das homoplasias, b) distinguir homologias derivadas (apomorfia) das homologias ancestrais (pleisiomorfia);
- ✓ Caractere homólogo é um caractere compartilhado por duas ou mais espécies e que estava presente no ancestral comum a elas;
- ✓ Caractere homoplásico é um caractere compartilhado entre duas ou mais espécies e que não estava presente no ancestral comum a elas. Podem surgir por acaso ou através da evolução convergente (paralelismo);
- ✓ Os caracteres de escolha para inferência filogenética são os homólogos derivados;
- ✓ Critérios gerais para que os caracteres sejam classificados como homólogos: mesma estrutura fundamental geralmente possui a mesma

relação com os caracteres circundantes, mesmo desenvolvimento em grupos diferentes;

- ✓ Homologia ancestral (plesiomorfia ou simplesiomorfia, para um grupo de indivíduos de uma ou mais espécie) é o estado de um caractere que estava presente no ancestral comum a todo o conjunto de espécies estudado;
- ✓ Homologia derivada (apomorfia ou sinapormofia, para um grupo de indivíduos de uma ou mais espécie) é o estado de um caractere que evoluiu dentro do conjunto de espécies estudado, após o ancestral comum;
- ✓ Diferentes homologias derivadas devem sugerir filogenias compatíveis. Para isso as homologias e as polaridades (analisar qual dos estados de um caractere é o estado ancestral e qual é o estado derivado) dos caracteres devem ser corretamente identificadas;
- ✓ Esta polaridade pode ser inferida por um grupo externo ou grupo fósseis.

#### 2.7.2.4. Métodos de inferência filogenética baseados na análise de caracteres

Esses métodos correspondem à parcimônia e aos dois métodos probabilísticos. Eles possuem abordagem essencialmente cladística, ao contrário dos métodos de inferência filogenética quantitativos, que são baseados em distâncias (*Yang & Rannala, 2012*).

##### A) MÉTODOS BASEADOS EM PARCIMÔNIA

São métodos não paramétricos e baseados na análise de caracteres. Estes métodos minimizam o total de números de mudanças de caracteres, que podem ser

ponderados para refletir a evidência estatística. Nestes métodos assume-se que cada característica evolui independente – assim, cada característica representa uma contribuição independente para o total. A fim de avaliar esta contribuição, os métodos de parcimônia reconstróem sequências ancestrais nos nós internos. Porém, em contraste com a distância e verossimilhança, métodos por parcimônia nem sempre são estatisticamente consistentes (Felsenstein, 1978). Esse método se baseia em caracteres, ou seja, compara simultaneamente todas as sequências no alinhamento, considerando-se um caractere (um sítio no alinhamento) por vez para calcular uma pontuação para cada topologia e não tem um modelo de substituição de base explícito (Yang & Rannala, 2012). A(s) melhor(es) árvore(s) é(são) aquela(s) com menor número de mudanças ao longo dos ramos e os testes de confiança para a(s) árvore(s) resultante(s) geralmente é feito pelo *bootstrap* (Yang & Rannala, 2012).

Algumas características favorece o uso desse método: fácil compreensão, as análises matemáticas são simples. Essas características facilitam a elaboração de algoritmos. Apesar dessas vantagens, esse método não permite uso de modelo evolutivo, não é eficiente na correção de possíveis substituições múltiplas de nucleotídeos levando a atração de ramos longos.

## B) MÉTODOS BASEADOS EM PROBABILIDADE

São métodos paramétricos e que tem sido muito utilizado para a análise de divergência de espécies crípticas. Isso se deve à utilização da probabilidade como uma ferramenta para se medir a incerteza na inferência filogenética. O processo evolutivo é fundamentado em modelos probabilísticos de evolução aumentando a

possibilidade de se aproximar à realidade biológica. Este método também se baseia em caracteres e faz uso dos modelos de substituição de base de nucleotídeos (Yang & Rannala, 2012). Os métodos estão disponíveis em duas formas: uma por abordagem por verossimilhança, pela qual a(s) melhor(e)s árvore(s) é(são) aquela(s) com maior valor de verossimilhança e o teste de confiança da árvore resultante é feito geralmente por *bootstrap*; a outra, por abordagem Bayesiana, pela qual a(s) melhor(e)s árvore(s) é (são) aquela(s) com maior probabilidade *a posteriori* e o teste de confiança da árvore resultante corresponde a própria probabilidade *a posteriori* (Bryant *et al.*, 2005; Yang, 2005).

Algumas das vantagens dos métodos de máxima verossimilhança e de inferência Bayesiana são: todas as suposições de modelo evolutivo são explícitas, podendo então ser avaliadas e melhoradas (Bryant *et al.*, 2005; Yang, 2005). Essas suposições são chamadas de modelos evolutivos ou mais especificamente modelos evolutivos probabilísticos.

Dentre as desvantagens estão: a alta exigência computacional tanto do cálculo de verossimilhança (apenas a abordagem verossimilhança) como da busca de topologias de árvore e propriedades estatísticas potencialmente pobres quando o modelo evolutivo é mal especificado (Bryant *et al.*, 2005; Yang, 2005).

Os métodos probabilísticos estão sendo muito utilizados para análise de divergência de espécies e espécies crípticas. Isso se deve as suas vantagens sobre os demais métodos. Primeiro a utilização da probabilidade como uma ferramenta de medir incerteza na inferência filogenética. A disponibilidade de um rico repertório de modelos evolutivos sofisticados é uma de suas principais vantagens em relação à máxima parcimônia, já que nem sempre a evolução é parcimoniosa, apesar de

terem essências cladísticas. Por serem baseados em caracteres, são mais eficientes do que os métodos de distância, inclusive para a detecção de espécies crípticas, apesar de ambos utilizarem modelos evolutivos. Diante disto, conclui-se que os métodos probabilísticos têm vantagens sobre os métodos de distância ou de parcimônia se o objetivo é compreender o processo de evolução das sequências (Yang & Rannala, 2012).

## 2.8. MÉTODOS PROBABILÍSTICOS DE INFERÊNCIA FILOGENÉTICA

(Este tópico está baseado em Bryant *et al.*, 2005 e Yang, 2005)

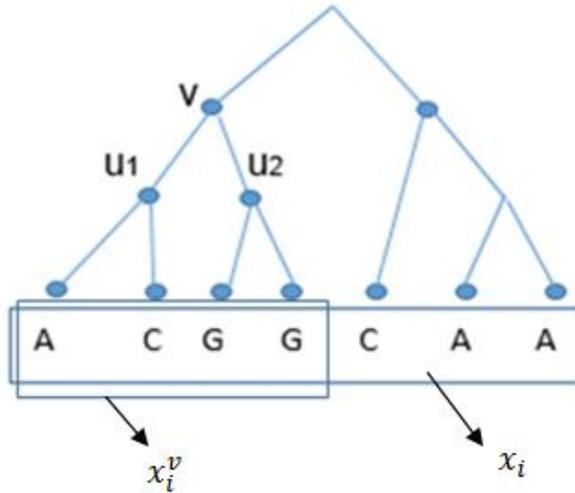
### 2.8.1. Por abordagem de máxima verossimilhança

A ideia chave por trás da verossimilhança é escolher os valores dos parâmetros (topologia da árvore, comprimento dos ramos e modelo evolutivo de substituição de bases) que maximizam a probabilidade de se observar os dados que dispomos. Os valores serão aceitos, considerando que a melhor verossimilhança é a máxima verossimilhança (Felsenstein, 1981). O algoritmo (Felsenstein, 1981) utilizado na filogenética está fundamentado na função  $L(\theta | D) = P(D | \theta)$  ou  $L(\theta) = P(D | \theta)$  que captura como é possível, plausível, os dados estimados alcançados ser os dados observados no cotidiano ( $D$ ) dado os valores dos parâmetros ( $\theta$ ). A verossimilhança de um alinhamento é calculada multiplicando-se a verossimilhança de cada um dos  $n$  sítios  $X_i$ ,  $L(\theta) = P[X_i | \theta]$ :

$$L(\theta) = \prod_{i=1}^n P[X_i | \theta] \quad (1.1)$$

Tendo a topologia abaixo,

Figura 2 - Ilustração do nó  $v$ , seus descendentes  $u_1$  e  $u_2$ , os caracteres  $x_i$  e os caracteres restritos por  $x_i^v$  para a sub-árvore enraizada por  $v$ .



sendo  $v$  um nó interno da árvore,  $x \in E$ , onde  $E = \{A, C, G, T\}$ , a verossimilhança condicional parcial é:

$$L_i^v(x) = P[X_i^v | \theta, \hat{X}_i(v) = x] \quad (1.2)$$

sendo  $X_i^v$  a restrição do caractere  $X_i$  aos descendentes do nó  $v$ ,  $\hat{X}_i(v)$  é o estado do ancestral para o sítio  $i$  ao nó  $v$ ,  $L_i^v(x)$  é a verossimilhança para o sítio  $i$  por sub-árvore adjacente ao nó  $v$ , condicional ao estado  $x$  de  $v$ .

Já que existem quatro estados, deve-se haver o somatório das probabilidades considerando cada possível estado de  $v$ . Então a verossimilhança fica:

$$P[X_i | \theta] = \sum_{x \in E} P[\hat{X}_i(v_0) = x] L_i^{v_0}(x) \quad (1.3)$$

onde  $v_0$  é o nó da árvore. Sendo  $P[\tilde{X}(v_0) = x]$  igual à probabilidade de  $x$  sobre uma distribuição estacionária,  $\pi_x$ .

A função  $L_i^v(x)$  satisfaz a recorrência:

$$L_i^v(x) = \left( \sum_{y \in E} P_{xy}(t_1) L_i^{u_1}(y) \right) \left( \sum_{y \in E} P_{xy}(t_2) L_i^{u_2}(y) \right) \quad (1.4)$$

para todos nós internos  $v$ , onde  $u_1$  e  $u_2$  são os nós descendentes de  $v$  enquanto  $t_1$  e  $t_2$  são os tamanhos de ramos conectados a  $v$ . Esta fórmula indica a independência dos processos das duas sub árvores que seguem o nó  $v$ . Para as pontas  $l$  da árvore a verossimilhança é:

$$L_i^l(x) = \begin{cases} 1, & \text{se } X_i(l) = x \\ 0, & \text{outros estados} \end{cases} \quad (1.5)$$

Computacionalmente é mais comum ser calculado o logaritmo da verossimilhança do que a verossimilhança propriamente dita, de forma que a primeira fórmula apresentada se apresenta como um somatório e não como multiplicação.

### 2.8.2. Por abordagem Bayesiana

Os métodos Bayesianos calculam a probabilidade *a posteriori* dos dados observados terem sido produzidos por várias árvores *a priori*. A melhor estimativa da filogenia é a árvore com a maior probabilidade *a posteriori*. A ideia central da inferência Bayesiana é que os parâmetros podem assumir distribuições binomiais, Poisson e normal. Antes os dados são observados – e assim os parâmetros tem uma distribuição *a priori*. Isto é combinado com a verossimilhança (probabilidade dos dados, dado cada valor dos parâmetros), para dar uma distribuição *a posteriori* (Rannala & Yang 1996), através do teorema de Bayes. Os algoritmos para inferência Bayesiana se fundamentam basicamente na seguinte função do teorema de Bayes:

$$f(\theta|D) = \frac{f(\theta)f(D|\theta)}{f(D)} = \frac{f(\theta)f(D|\theta)}{\int f(\theta)f(D|\theta) d\theta} \quad (1.6)$$

O Teorema de Bayes foi desenvolvido por Thomas Bayes na primeira metade do século XVIII. Porém os seus primeiros empregos na inferência filogenética foram feitos por Rannala & Yang (1996), Mau & Newton (1997) e Li et al. (2000).

A probabilidade dos dados,  $f(D)$  é uma normalização constante para marcar integração  $f(\theta/D)$  para um. Nesta equação mostra que a distribuição *a posteriori*  $f(\theta/D)$  é proporcional a distribuição *a priori*  $f(\theta)$  vezes a verossimilhança  $f(\theta/D)$ . Ou seja, a informação *a posteriori* é a soma das informações *a priori* e das informações da amostra.

Considere:  $D$  - dados de sequência de nucleotídeo;  $\tau_i$ - iésima topologia da árvore;  $i=1, 2, 3, \dots, N(s)$ , onde  $N(s)$  é o número total de topologias para  $s$  espécies; geralmente assume-se que  $f(\tau_i)=1/N(s)$  uniforme *a priori*;  $b_i$  – comprimentos de ramos sobre  $\tau_i$ , com probabilidade *a priori*  $f(b_i)$ .

Então a probabilidade *a posteriori* de uma árvore  $\tau_i$  é:

$$P(\tau_i|D) = \frac{\iint f(\theta)f(b_i|\theta)f(\tau_i|\theta)f(D|\tau_i, b_i, \theta) db_i d\theta}{\sum_{j=1}^{N(s)} \iint f(\theta)f(b_j|\theta)f(\tau_j|\theta)f(D|\tau_j, b_j, \theta) db_j d\theta} \quad (1.7)$$

O denominador, correspondente à probabilidade marginal dos dados  $f(D)$ , deve envolver o somatório sobre todas possíveis topologias de árvore e para cada topologia  $\tau_i$ , integração (envolve uma integral) sobre todos os comprimentos de ramos  $b_i$  e parâmetros  $\theta$ , uma tarefa praticamente impossível, exceto para árvores muito pequenas. Mas uma estratégia para sair desta dificuldade computacional é a abordagem estocástica de busca de topologias através do método Cadeia de Markov Monte Carlo (MCMC, do inglês *Markov chain Monte Carlo*), que é um algoritmo de amostragem aplicado em cálculos de integrais-multidimensionais.

Na máxima verossimilhança também tem a abordagem estocástica para escolha de árvore, sendo que não é MCMC, pois os cálculos de máxima verossimilhança em si não envolve integral. Mas segue a mesma lógica – de forma heurística procurar as topologias possíveis dando saltos pelo espaço amostral de topologias através de razão entre topologia atual e topologia após salto. Quem rege

esses saltos no MCMC é a razão proposta como será visto no próximo parágrafo. A busca de topologia utilizada é heurística na forma *hill-climbing* (mantendo a nova árvore se é melhor do que a anterior) como NNI (*Nearest neighbor interchange*), SPR (*Subtree pruning and regrafting*) e TBR (*Tree bisection and reconnection*) mas que são moldadas com a forma dos saltos regido pela razão proposta.

Mas a MCMC possui uma estratégia que possibilita o cálculo Bayesiano, inclusive na filogenética. Um algoritmo que utiliza o método MCMC é o de Metropolis *et al.* (1953). A ideia é gerar um processo de Markov, onde um estado pode ser aceito ou rejeitado como auxílio de um processo de iteração. Os estados da cadeia são os parâmetros  $\theta$  e a distribuição estacionária do estado é  $\pi(\theta) = f(\theta|\mathbf{D})$ , que é a distribuição a posteriori de  $\theta$ .

Suponha que o atual estado da cadeia de Markov é  $\theta$ . O algoritmo propõe um novo estado  $\theta^*$  através de uma densidade proposta (do inglês, *proposal*)  $q(\theta^*|\theta)$ . O estado candidato  $\theta^*$  é aceito com probabilidade:

$$\alpha = \min \left( 1, \frac{\pi(\theta^*)q(\theta|\theta^*)}{\pi(\theta)q(\theta^*|\theta)} \right) \quad (1.8)$$

Quando o algoritmo para o método MCMC é usado para aproximar à distribuição a posteriori dos parâmetros  $\theta$ , têm se  $\pi(\theta) = f(\theta|\mathbf{D}) = f(\theta)f(\mathbf{D}|\theta)/f(\mathbf{D})$ , de modo que:

$$\frac{\pi(\theta^*)}{\pi(\theta)} = \frac{\frac{f(\theta^*)f(D|\theta^*)}{f(D)}}{\frac{f(\theta)f(D|\theta)}{f(D)}} = \frac{f(\theta^*)f(D|\theta^*)}{f(\theta)f(D|\theta)} \quad (1.9)$$

Logo a constante de normalização  $f(D)$  é cancelada. Conclui-se que a probabilidade de aceitação é:

$$= \min\left(1, \frac{f(\theta^*)}{f(\theta)} \times \frac{f(D|\theta^*)}{f(D|\theta)} \times \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)}\right) =$$

$$= \min\left(1, \text{razão a priori} \times \text{razão verossimilhança} \times \text{razão proposta}(\text{proposal})\right) \quad (1.10)$$

Vários melhoramentos do MCMC têm sido propostos nos recentes anos os quais também têm sido introduzidos na inferência filogenética, incluindo *Gibbs sampler*, *Metropolis-coupled* MCMC (utilizado no programa Mr. Bayes), *Time-reversible jump* MCMC, *multiple-try* Metropolis e *population-based* MCMC.

A máxima verossimilhança usa uma função explícita que pode ser minimizada para se obter a explicação "mais provável" para os dados (infelizmente resultados evolutivos nem sempre são o resultado do processo o mais provável) (Hall, 2001). Uma abordagem que está rapidamente ganhando terreno utiliza metodologia Bayesiana para a construção de árvores (Lewis & Swofford, 2001).

Ambas abordagens usam a função de verossimilhança e, portanto, compartilham muitas propriedades estatísticas, tais como consistência e eficiência. No entanto, máxima verossimilhança e inferência Bayesiana representam filosofias

opostas de inferência estatística. Dependendo do tipo de dado analisado os recursos adicionais (probabilidade *a priori*) da Bayesiana podem ser visto como força ou fraqueza (Yang & Rannala, 2012).

## 2.9. APLICAÇÃO DA ABORDAGEM MÁXIMA VEROSSIMILHANÇA EM INVESTIGAÇÕES DE POSSÍVEIS ESPÉCIES CRÍPTICAS NO BRASIL

Os métodos da Máxima Verossimilhança já foram utilizados em estudos envolvendo complexo de espécies: delimitação de espécies crípticas por descrição taxonômica de mosquitos *Anopheles (Nyssorhynchus) albitarsis* (Diptera: *Culicidae*) (Lehret *et al.*, 2005 – utilizando marcador mitocondrial e RAPD); Wilkerson *et al.*, 2005 – utilizando marcadores mitocondriais e nucleares), elucidação de caráter ambíguo do complexo de espécies crípticas *Leptynia attenuata* (Bicho pau) utilizando um segmento do gene *citocromo oxidase subunidade 2* (COX-2), cariótipo e aloenzimas (Passamonti *et al.*, 2004), análise filogeográfica de possíveis espécies crípticas em um complexo de espécies de sapo gigante espinhoso (Dicroglossidae: *Paa spinosa*) a partir de dados de sequência mtDNA (genes 12S rRNA e 16S rRNA ) (Ye *et al.*, 2013) e análise de diversidade de espécies sugestivas de serem crípticas do gênero *Lutzomyia* a partir de sequência CO1 (gene DNA *barcoding*) (Azpurua *et al.*, 2010), porém não se encontrou utilização do método de Máxima Verossimilhança para inferência filogenética em análise de complexo de espécies *L. umbratilis*.

## 2.10. APLICAÇÃO DA ABORDAGEM BAYESIANA EM INVESTIGAÇÕES DE POSSÍVEIS ESPÉCIES CRÍPTICAS NO BRASIL

Os métodos Bayesianos também têm sido utilizados na avaliação de complexo de espécies: análise de diversidade de espécies sugestivas de serem crípticas do gênero *Lutzomyia* (Azpurua *et al.*, 2010); análise filogeográfica de possíveis espécies crípticas do sapo gigante espinhoso (Dicroglossidae: *Paa spinosa*) a partir de dados de sequência mtDNA (genes 12S rRNA e 16S rRNA) (Ye *et al.*, 2013), delimitação de espécies crípticas por descrição taxonômica dos mosquitos *Anopheles (Nyssorhynchus) albitarsis* (Diptera: *Culicidae*) através de PCR de DNA polimórfico e análise gene mitocondrial COI (Lehret *et al.*, 2005) e sugestão de espécies crípticas e incipientes de *L. umbratilis* através do gene COI (Scarpassa & Alencar, 2012).

## 2.11. AVALIAÇÃO METODOLÓGICA DE MÉTODOS DE INFERÊNCIA FILOGENÉTICA

Já foram feitos vários estudos do funcionamento dos métodos de inferência filogenética. Geralmente são utilizados alguns critérios para julgar os métodos de reconstrução árvore (Yang & Rannala, 2012):

- **Consistência:** gera árvore confiável com determinado número de dados. Um método inconsistente, com frequência produz árvore errada mesmo se for dado quantidades infinitas de dados. Métodos baseados em modelo (isto é, matriz de distância, máxima verossimilhança e inferência Bayesiana) são consistentes se o modelo assumido está correto. A parcimônia pode ser inconsistente sob alguns modelos evolutivos (Felsenstein, 1978).

- **Eficiência:** em filogenética pode ser medida através da probabilidade de recuperação de árvore correta ou sub-árvore dado o número de sítios. Simulações de computador geralmente têm encontrado maior eficiência na máxima verossimilhança do que na parcimônia ou distância (especialmente *neighbour joining*) na recuperação da árvore correta (Felsenstein, 2004).
- **Velocidade computacional:** tempo de execução do método. Métodos de distância são muito rápidos. Enquanto métodos que buscam a melhor árvore sob um critério, como a evolução máxima, máxima parcimônia e máxima verossimilhança, são mais lentos. A máxima verossimilhança e Inferência Bayesiana são tipicamente mais lento do que a parcimônia.

Nos dois primeiros pontos os métodos probabilísticos se superam, porém apresentam uma velocidade computacional menor. Mas mesmo assim as vantagens dos métodos probabilísticos superam os demais métodos em suas vantagens.

Diferentes métodos de inferência filogenética podem construir diferentes árvores. Essas diferenças podem ser avaliadas a partir da comparação entre si da topologia das árvores filogenéticas resultantes de cada método (Puigbò *et al.*, 2007). Existem testes já descritos que medem o grau de congruência entre topologias como “split” e “nodal” implementados no programa TOPD/FMTS (Puigbò *et al.*, 2007).

### 3. OBJETIVOS

#### 3.1. OBJETIVO GERAL

- Avaliar desempenho dos algoritmos de máxima verossimilhança (MV) e Bayesiana (BE) e de dois marcadores genéticos como ferramentas auxiliares no reconhecimento de espécies crípticas dos complexos *Lutzomyia longipalpis* e *L. umbratilis*, visando sistematizar a metodologia de investigação filogenética molecular de complexo de espécies destes flebotomíneos.

#### 3.2. OBJETIVOS ESPECÍFICOS

- Elaborar grupos de dados local contendo as sequências de nucleotídeos do gene mitocondrial citocromo oxidase 1 (COI) e do nuclear *period* (*per*) oriundas de espécimes de populações representativas dos complexos *Lutzomyia longipalpis* e *L. umbratilis*.
- Obtenção de árvores filogenéticas para as espécies dos complexos estudados, inferindo-as com os métodos de máxima verossimilhança e de inferência Bayesiana a partir desses grupos de dados. Foram utilizados dados de sequencias de um único gene de cada vez como também dados de sequencias de ambos genes concatenados.
- Comparar o desempenho de ambos os métodos de inferência filogenética relacionando-os aos dois marcadores genéticos: citocromo oxidase I (mitocondrial) e *period* (nuclear).

- A partir dos resultados da comparação ver qual método e combinado com que marcador teve melhor adequação para investigação das espécies crípticas *L. umbratilis* e *L. longipalpis*.

## 4. MATERIAL E MÉTODOS

### 4.1. COLETA E ORGANIZAÇÃO DOS DADOS DE SEQUÊNCIAS DE NUCLEOTÍDEOS NUCLEAR E MITOCONDRIAL DE *L. UMBRATILIS* E *L. LONGIPALPIS*

As sequências de segmentos dos genes *period* (nuclear) e COI (mitocondrial) de *L. longipalpis* e *L. umbratilis* foram obtidas através de busca pelas palavras-chave “*Lutzomyia longipalpis AND period gene*”, “*Lutzomyia longipalpis AND COI gene*”, “*Lutzomyia umbratilis AND period gene*” e “*Lutzomyia umbratilis AND COI gene*” no Genbank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), como também obtidas no Laboratório de Bioinformática e Biologia Evolutiva do Departamento de Genética da UFPE (LABBE/UFPE), onde foi desenvolvido este trabalho. Dentre as tantas sequências recuperadas, foram escolhidas aquelas cujas regiões geográficas (ver Anexos 3 e 5) de origem estão associadas a possíveis processos de especiação para as espécies estudadas. As sequências obtidas foram organizadas segundo o tipo de gene e da espécie em arquivos no formato *fasta*.

Para alcançar os objetivos deste trabalho, o principal passo foi inferir a filogenia das sequências dos táxons. Para isso foi necessário executar procedimentos preliminares indispensáveis para a inferência: coleta de sequências de espécies candidatas a grupo externo; alinhamento múltiplo das sequências; e escolha do modelo evolutivo.

## 4.2. OBTENÇÃO DE ÁRVORES FILOGENÉTICAS A PARTIR DOS DADOS DE SEQUENCIAS DOS MARCADORES COI OU *PER* DE *L. UMBRATILIS* E *L. LONGIPALPIS*

### 4.2.1. Definição de grupos externos para a inferência de árvores filogenéticas

Dependendo da história evolutiva inferida, a árvore filogenética que a mais bem representa pode separar os táxons em grupos diferentes. Para se obter ainda mais informação acerca do grau de proximidade entre os grupos, a obtenção de uma árvore enraizada é solução ideal. Aí é que reside a importância do grupo externo (também chamado de *outgroup*), o qual permitirá a formação do enraizamento da árvore. Por isso a escolha do grupo externo deve ser muito cautelosa. Deve-se conhecer muito bem a história evolutiva de unidades taxonômicas, relacionadas às unidades taxonômicas analisadas, mais antigas.

Pesquisou-se na literatura a história evolutiva da família (Psychodidae) da qual o gênero *Lutzomyia* faz parte (Curler & Moulton, 2012), a fim de buscar espécies que poderiam ser mais propriamente utilizadas como grupos externos para cada um dos complexos de espécies estudados. Houve três buscas que se aplicam tanto para *L. umbratilis* como para *L. longipalpis* (ver também fluxograma no Anexo 2):

- **Primeira busca – busca de grupo externo (nível de espécie):** As sequências COI e *period* das espécies foram coletadas do banco de dados POPSET do Genbank, utilizando palavras chaves compostas de: “nome\_da\_espécie *period* “; e “nome\_da\_espécie *COI* “,

- **Segunda busca – busca de grupo externo (nível de subgênero) com pré-inferência filogenética a partir de método de distancia:** pensando na possibilidade de se concatenar as sequências dos dois marcadores para inferir árvores filogenéticas mais informativas e diante da não correspondência de grupo externo em nível de espécie:
  - Foram feitas buscas de sequências dos dois marcadores (nível de subgênero, por já ser bem estabelecido na literatura)
  - Em seguida, foram testadas as combinações possíveis de subgênero como grupos externos, construindo árvores rápidas pelo método de distância UPGMA através do programa MEGA v6 (Tamura *et al.*, 2013).
- **Terceira busca – Busca de grupo externo (nível de subgênero):** depois de se observar os resultados das duas buscas anteriores, uma espécie representante para cada subgênero coincidentes foi escolhida para ambos marcadores.

Vale salientar que esta estratégia de busca de grupos externos não segue nenhum relato na literatura e representa, portanto, decisões tomadas de acordo como resultado de cada uma das interações. A lógica seguida foi a de se começar com uma busca geral, seguida de buscas mais específicas que dependeriam dos resultados de cada uma das etapas anteriores. A busca foi encerrada quando foi possível atingir árvores filogenéticas possíveis de serem submetidas as avaliações de comparação de métodos de inferência filogenética (citados adiante). Estas árvores, para os dois marcadores, além de serem aceitáveis no ponto de vista biológico (Anexo 5), deveriam possuir táxons correspondentes, inclusive os táxons do grupo externo.

#### 4.2.2. Alinhamento das sequências de nucleotídeos

Antes de se fazer qualquer análise evolutiva, é necessário que as sequências estejam organizadas quanto às suas similaridades. Um alinhamento de sequências é uma forma de se organizar sequências primárias de nucleotídeos de regiões similares através das relações funcionais, estruturais ou evolutivas entre elas. Sequências alinhadas de nucleotídeos ou de resíduos de aminoácidos são representadas tipicamente como linhas de uma matriz. Espaçamentos (*gaps*) podem ser inseridos entre os resíduos para que caracteres semelhantes sejam alinhados em colunas sucessivas.

Se duas sequências em um alinhamento descenderem de um ancestral comum, discordâncias (*mismatches*) podem ser interpretadas como mutações pontuais e os espaços (*gaps*) como inserções e/ou deleções introduzidas em uma ou ambas as sequências desde quando estas divergiram no tempo evolutivo. A abordagem computacional para os alinhamentos podem ser local ou global, quanto ao tipo, e ótima ou heurística, quanto à precisão do alinhamento.

Neste trabalho foi utilizada a abordagem global-heurística através do programa computacional Clustal v. 1.83.1 (Chenna *et al.*, 2003), implementada no programa MEGA v. 6.0 (Tamura *et al.*, 2011).

Após cada uma das três buscas de grupos externos foi feito o alinhamento das sequências. As sequências alinhadas estavam em arquivos no formato *fasta*. Cada um desses arquivos continha tanto sequências dos flebotomíneos (inclusive as dos grupos externos tentativos). Cada arquivo referente ao marcador *period*, por exemplo, difere dos demais pela ausência ou presença de um grupo externo, e estes entre si pela diferentes combinações de táxons de grupo externo.

O ideal seria se as possíveis espécies utilizadas como grupos externos para os complexos estudados fossem as mesmas para os dois marcadores. Esta situação facilitaria as etapas de comparação de dados e de topologias de árvores relacionados aos diferentes marcadores já que espécies de referência seriam as mesmas. Porém os bancos de dados não contem até o presente momento sequências dos dois marcadores para cada uma das espécies consideradas.

#### **4.2.3. Escolha do modelo evolutivo probabilístico**

A partir das sequencias alinhadas faz-se a escolha do modelo evolutivo probabilísticos de sequencia, para então fazer a inferência filogenética pelos métodos probabilísticos. A escolha do modelo evolutivo foi feita usando o programa jModelTest v. 2.1.6. (Posada, 2008), com operação de busca de topologia de árvore BEST e método de avaliação do modelo BIC (*Bayesian Information Criterion*; ou Critério de Informação Bayesiana).

#### **4.2.4. Obtenção das árvores filogenéticas por métodos probabilísticos**

As árvores filogenéticas foram inferidas com duas abordagens: Máxima Verossimilhança (MV) e Bayesiana (BE). Todos os arquivos obtidos a partir das buscas 1 e 3 dos grupos externos e os arquivos contendo a melhor combinação de grupos externos na inferência pelo método de distância na busca 2 de grupos externos, foram submetidos a essas duas abordagens:

- **Máxima Verossimilhança:** através do programa PHYML v. 3.0 (Guindon *et al.*, 2010), com teste de suporte de ramos feitos através do método *bootstrap* com 100 a 1000 replicatas. Os arquivos com sequências alinhadas foram

convertidos do formato *fasta* para o *Phylip* (ver <http://sequenceconversion.bugaco.com/converter/biology/sequences/>).

- **Bayesiana:** realizada com o programa Mr. Bayes (Ronquist & Huelsenbeck, 2003) com 20.000 a 4.000.000 gerações, tendo a obtenção de cada árvore filogenética do espaço amostral de cada corrida a cada 500 gerações e tendo 25% das primeiras amostras da cadeia fria (*cold chain*) descartadas, ou seja, submetidas ao *burn-in*. Todos os arquivos de sequências alinhadas foram antes convertidos desta vez do formato *Fasta* para o formato *Nexus* (ver <http://sequenceconversion.bugaco.com/converter/biology/sequences/>).

A visualização destas árvores foi feita a partir do *software* Figtree (Rambaut, 2012). Foi observada a polarização de cada uma delas. As árvores filogenéticas resultantes que se polarizaram como umas das formas descritas no Anexo 5, ou seja, seguindo a lógica biológica de possíveis processo de especiação terão suportes dos grupos formados descritos nos resultados. Elas serão consideradas como árvores com boa polarização. Já aquelas que não apresentaram boa polarização foram desconsideradas na comparação dos métodos filogenéticos probabilísticos.

#### 4.3. OBTENÇÃO DE ÁRVORE FILOGENÉTICA A PARTIR DE DADOS DE SEQUÊNCIAS CONCATENADAS DOS MARCADORES COI E *PER* DE *L. UMBRATILIS* E *L. LONGIPALPIS*

Além de avaliar os métodos probabilísticos observando as árvores filogenéticas inferidas por eles a partir das sequências de cada marcador

isoladamente, também se pretendeu obter árvores inferidas a partir de sequências concatenadas dos dois marcadores. Antes desta etapa, verificou-se se havia congruência dos conjuntos de sequências.

Criou-se um arquivo de cada marcador para concatenar as sequências e em seguida submetê-las ao teste de homogeneidade de partição, para analisar se haveria realmente possibilidade de se inferir uma árvore filogenética. O arquivo escolhido foi aquele cujas árvores inferidas pelos métodos de Máxima verossimilhança e Bayesiano (ver valores de *bootstrap*, de números de gerações, conversão de arquivos, visualização de árvores no tópico 4.2.4 – foram os mesmos utilizados para inferência de árvore filogenética a partir de um único gene, ou seja, sequências não concatenadas) foram aceitáveis no contexto biológico (ver figuras 3 e 4) e continham sequências de grupos externos correspondentes.

Em um conjunto de dados filogenéticos dividido em partições menores ou constituídos de partições diferentes (correspondentes a genes diferentes, por exemplo), cada um pode originar árvores com diferentes topologias. Isto porque algumas (ou todas) árvores são incorretas e as partições compartilham a mesma história, ou ainda as árvores estão corretas e as partições diferentes têm experimentado histórias evolutivas distintas. A distinção entre essas duas opções requer testes estatísticos para determinar se as diferenças na topologia podem se dever apenas por acaso.

Foram realizados testes ILD (*Incongruence Length Difference*) (Farris *et al.*, 1994), implementado como o teste de homogeneidade de partição (HOMPART) no PAUP (Swofford, 2003), para verificar a congruência entre as matrizes combinadas das sequências alinhadas de *period* e COI. Apenas valores de  $p < 0,05$  foram

considerados como evidência significativa de incongruência entre os dados. O teste foi conduzido a partir de cem (100) replicações, com adição randômica de uma árvore por passo. Nessas combinações ou concatenações foram consideradas como uma única sequência para cada um dos táxons.

O teste de partição na verdade indica o caminho certo, apesar de não levar em conta o modelo evolutivo. Embora o intuito principal desse trabalho tenha sido comparar o método e não a capacidade do marcador, este teste é indispensável, uma vez que ele aponta *a priori* se vale a pena inferir árvores filogenéticas com dados concatenados (Hipp *et al.*, 2004).

Após o teste de partição, as sequências foram submetidas ao programa JModelTest para escolha do modelo evolutivo e logo em seguida a inferência filogenética MV e BE.

#### 4.4. COMPARAÇÃO DO DESEMPENHO DOS MÉTODOS DE INFERÊNCIA A PARTIR DE ANÁLISE DAS ÁRVORES INFERIDAS POR DADOS CONCATENADOS E POR DADOS DE CADA MARCADOR ISOLADO

Foram utilizados vários critérios para verificar as árvores. Esta avaliação dos melhores métodos foi nos seguintes pontos:

- O valor de suporte de cada polaridade das árvores aceitáveis, também a partir da observação individual de cada árvore. As árvores inferidas por dados de apenas um marcador isolado foram analisadas antes de concatenar os dados (a razão disto está no segundo parágrafo do tópico 4.3). E as árvores inferidas por dados de sequências concatenadas de ambos marcadores (COI e *per*)

foram avaliadas e também comparadas com árvores inferidas por sequências de cada marcador isolado.

- Teste de congruência entre as árvores. A partir dos métodos *split* (mais congruentes são as árvores quando os valores são próximos de zero. A escala de valores possíveis varia de zero a um) e nodal (mais congruentes serão as árvores quando o valor for mais próximo de um. A escala de valores possíveis vai de 1 a  $+\infty$ ), a partir do *software* TOPD/FMTS (Puigbò *et al.*, 2007). É evidente que métodos diferentes produzem muitas vezes árvores diferentes, mesmo utilizando os mesmos dados. Logo é bastante importante ter uma forma de medir objetivamente quão semelhante estas árvores são e a partir de proporção de clados coincidentes obtidos através do *software* Densitree (Bouckaert, 2010).

## 5. RESULTADOS

### 5.1. DADOS DE SEQUÊNCIAS DE *L. UMBRATILIS* E *L. LONGIPALPIS* ORGANIZADOS NOS ARQUIVOS FASTA

As sequências de cada um dos marcadores foram organizadas em arquivos diferentes para *L. umbratilis*: um dos arquivos conteve 72 sequências do gene *per* de *L. umbratilis*, sendo 24 de machos e 12 de fêmeas de Manacapuru (MAN – Amazonas, Brasil); 15 de machos e sete de fêmeas de Rio Preto da Eva (RIP – Amazonas, Brasil) e 14 de fêmeas de Recife (REC – Pernambuco, Brasil); o segundo arquivo, por sua vez, conteve 72 sequências de um segmento do gene COI de *L. umbratilis*, dos mesmos indivíduos da descrição anterior (sumarizado na Tabela 1).

Tabela 1 - Categorização das proporções das 72 sequências do marcador *per* e COI referentes a *L. umbratilis* quanto ao gênero e área geográfica.

Gene	Populações amostradas						Total
	Manacapuru / AM		Ribeirão Preto da Eva / AM		Recife / PE		
	Machos	Fêmeas	Machos	Fêmeas	Machos	Fêmeas	
<i>Period</i>	24	12	15	7	-	14	72
COI	24	12	15	7	-	14	72
<b>TOTAL</b>	<b>48</b>	<b>24</b>	<b>30</b>	<b>14</b>	<b>-</b>	<b>28</b>	<b>144</b>

Do mesmo modo, para *L. longipalpis*: um arquivo conteve 73 sequências do gene *per* de *L. longipalpis*, sendo 10 de machos 1S de Sobral (SOB – Ceará, Brasil), 26 de machos 1S de Caririaçu (CAR– Ceará, Brasil), oito de machos 2S de Sobral (SOB) e 29 de machos 2S de Caririaçu (CAR); o outro arquivo conteve 73 sequências do gene COI dos mesmos indivíduos anteriormente descritos (sumarizado na Tabela 2).

Tabela 2 - Categorização das proporções das 73 sequências do marcador *per* e COI referentes a *L. longipalpis* quanto ao gênero e área geográfica.

Populações amostradas					
Gene	Sobral / CE		Caririaçu / CE		Total
	Machos 1s	Machos 2S	Machos 1s	Machos 2S	
<i>Period</i>	10	8	26	29	73
COI	10	8	26	29	73
<b>TOTAL</b>	<b>20</b>	<b>16</b>	<b>52</b>	<b>58</b>	<b>146</b>

## 5.2. DEFINIÇÃO DE GRUPOS EXTERNOS PARA INFERÊNCIA DE ÁRVORES FILOGENÉTICAS DOS COMPLEXOS *L. UMBRATILIS* E *L. LONGIPALPIS*

- **Primeira busca - busca de grupo externo (em nível de espécie):** As espécies cujas sequências foram escolhidas estão sumarizadas na Tabelas 3 (*L. umbratilis*) e 4 (*L. longipalpis*).

Tabela 3 – Grupos externos utilizados para cada marcador (*per* e COI) referentes a *L. umbratilis* na primeira busca de grupos externos mais adequados.

<b>Marcador</b>	<b>Espécie</b>	<b>Referência no POPSET (NCBI)</b>
<b><i>Period</i></b>	<i>Phlebotomus duboscqi</i>	gi 22035171 gb AY071913.1
	<i>Lutzomyia intermedia</i>	gi 62861658 gb AY927155.1
	<i>Lutzomyia whitmani</i>	gi 62861516 gb AY927084.1
<b>COI</b>	<i>Phlebotomus argentipes</i>	gi 393660114 gb JX105038.1
	<i>Phlebotomus papatasi</i>	gi 393660112 gb JX105037.1
	<i>Phlebotomus sergenti</i>	gi 635543995 gb KJ481078.1
	<i>Lutzomyia longipalpis</i>	gi 403181582 gb JN897369.1

Tabela 4 – Grupos externos utilizados para cada marcador (*per* e COI) referentes a *L. longipalpis* na primeira busca de grupos externos mais adequado.

<b>Marcador</b>	<b>Espécie</b>	<b>Referência no POPSET (NCBI)</b>
<b><i>Period</i></b>	<i>Phlebotomus duboscqi</i>	gi 22035171 gb AY071913.1
	<i>Lutzomyia intermedia</i>	gi 62861658 gb AY927155.1
	<i>Lutzomyia whitmani</i>	gi 62861516 gb AY927084.1
<b>COI</b>	<i>Phlebotomus argentipes</i>	gi 393660114 gb JX105038.1
	<i>Phlebotomus papatasi</i>	gi 393660112 gb JX105037.1
	<i>Phlebotomus sergenti</i>	gi 635543995 gb KJ481078.1
	<i>Lutzomyia umbratilis</i>	Obtida no LABBE

- **Segunda busca - busca de grupo externo (em nível de subgênero) com pré-inferência filogenética a partir de método de distancia:** O subgênero e a respectiva espécie a cujo indivíduo pertence estão listados a seguir (Tabela 5):

Tabela 5 – Grupos externos utilizados para cada marcador (*per* e COI) referentes a *L. umbratilis* (e *L. longipalpis*) na segunda busca de grupos externos mais adequado.

Subgêneros	Marcador <i>period</i>		Marcador COI	
	Espécie(s)	Referência no POPSET (NCBI)	Espécie(s)	Referência no POPSET (NCBI)
<b><i>Nyssomia</i></b>	<i>L. umbratilis</i>	gi 22035167 gb AY071911.1	<i>L. antunesi</i>	gi 402715578 gb JN845549.1
	<i>L. intermedia</i>	gi 22035173 gb AY071914.1	<i>L. umbratilis</i>	gi 543174749 gb KF467562.1
	<i>L. whitmani</i>	gi 22035161 gb AY071908.1		
<b><i>Lutzomyia</i></b>	<i>L. dispar</i>	gi 22035169 gb AY071912.1	<i>L. gomezi</i>	gi 546605230 gb KC921253.1
	<i>L. renei</i>	gi 19849578 gb AY071907.1		
<b><i>Phlebotomus</i></b>	<i>P. duboscqi</i>	gi 22035171 gb AY071913.1	<i>P. sergenti</i>	gi 507472879 gb KC755398.1
<b><i>Migonei</i></b>	<i>L. migonei</i>	gi 22035163 gb AY071909.1	-	-
	<i>L. evandroi</i>	gi 22035165 gb AY071910.1		

<b><i>Psathyromyia</i></b>	-	-	<i>L. aragaoi</i>	gi 546605186 gb KC 921231.1
<b><i>Bichromomyia</i></b>	-	-	<i>L. olmeca</i>	gi 291066714 gb GU 001743.1
<b><i>Psychodopygus</i></b>	-	-	<i>L. panamensis</i>	gi 291066728 gb GU 001750.1

Nesta busca foram consideradas apenas as espécies cujos subgêneros eram coincidentes para ambos os marcadores (e.g. *Nyssomyia*, *Lutzomyia* e *Phlebotomus*).

- **Terceira busca- Busca de grupo externo (em nível de subgênero):** os subgêneros e as respectivas espécies escolhidas estão descritos na Tabela 6.

Tabela 6 – Grupos externos utilizados para cada marcador (*per* e COI) referentes a *L. umbratilis* (e *L. longipalpis*) na terceira busca de grupos externos mais adequado.

<b>Marcador</b>	<b>Subgênero</b>	<b>Espécie</b>
<b>COI</b>	<i>Nyssomyia</i> (N)	<i>Lutzomyia anduzei</i> (A)
	<i>Lutzomyia</i> (L)	<i>Lutzomyia gomezi</i> (G)
	<i>Phlebotomus</i> (P)	<i>Phlebotomus sergenti</i> (S)
<b>Period</b>	<i>Phlebotomus</i> (P)	<i>Phlebotomus dubosqui</i> (D)
	<i>Lutzomyia</i> (L)	<i>Lutzomyia renei</i> (R)
	<i>Nyssomyia</i> (N)	<i>Lutzomyia whitmani</i> (W)

### 5.3. ARQUIVOS COM SEQUÊNCIAS ALINHADAS

Foram executados no total, para cada espécie, 110 alinhamentos, correspondentes aos 24 arquivos *fasta* gerados a partir da primeira busca de grupos externos, 72 arquivos *fasta* (todos com grupos externos e esses correspondentes apenas aos subgêneros coincidentes, ou seja, *Nyssomyia*, *Lutzomyia* e *Phlebotomus*) a partir da segunda busca e 14 a partir da terceira busca de grupos externos.

Nas Tabelas 7, 8, 9 e 10 a seguir é descrita a lógica das diferentes combinações de sequências contidas em cada arquivo a partir da primeira busca de grupo externo para *L. umbratilis* (Tabelas 7 e 8) e *L. longipalpis* (Tabelas 9 e 10), como exemplo:

Tabela 7 - Oito possíveis combinações de táxons de grupo externo na primeira busca de grupo externo referentes a *L. umbratilis* utilizando marcador *per*.

Arquivos	Sequências do complexo de espécie	+	Táxons das sequências dos grupos externos
1			-
2			<i>P. duboscqi</i>
3			<i>L. intermedia</i>
4	72 sequências de <i>L.</i>		<i>L. whitmani</i>
5	<i>umbratilis</i>	+	<i>P. duboscqi</i> e <i>L. intermedia</i>
6			<i>P. duboscqi</i> e <i>L. whitmani</i>
7			<i>L. intermedia</i> e <i>L. whitmani</i>
8			<i>P. duboscqi</i> , <i>L. intermedia</i>

Tabela 8 - Possíveis combinações de grupos externos na primeira busca de grupo externos referentes a *L. umbratilis* utilizando marcador COI.

Arquivos	Sequências do complexo de espécie	+	Táxons das sequências dos grupos externos
1			-
2			<i>P. argentipes</i>
3			<i>P. papatasi</i>
4			<i>P. sergenti</i>
5			<i>L. longipalpis</i>
6			<i>P. argentipes</i> e <i>P. papatasi</i>
7			<i>P. argentipes</i> e <i>P. sergenti</i>
8	72 sequências de <i>L.</i>		<i>P. argentipes</i> e <i>L. longipalpis</i>
9	<i>umbratilis</i>	+	<i>P. papatasi</i> e <i>P. sergenti</i>
10			<i>P. papatasi</i> e <i>L. longipalpis</i>
11			<i>P. sergenti</i> e <i>L. longipalpis</i>
12			<i>P. argentipes</i> , <i>P. papatasi</i> e <i>P. sergenti</i>
13			<i>P. argentipes</i> , <i>P. papatasi</i> e <i>L. longipalpis</i>
14			<i>P. argentipes</i> , <i>P. sergenti</i> e <i>L. longipalpis</i>

15	<i>P. papatasi</i> , <i>P. sergenti</i> e <i>L. Longipalpis</i>
16	<i>P. argentipes</i> , <i>P. papatasi</i> , <i>P. sergenti</i> e <i>L. longipalpis</i>

Tabela 9 - Oito possíveis combinações de táxons de grupo externo na primeira busca de grupo externo referentes à *L. longipalpis* utilizando marcador *per*.

Arquivos	Sequências do complexo de espécie	+	Táxons das sequências dos grupos externos
1			-
2			<i>P. duboscqi</i>
3			<i>L. intermedia</i>
4			<i>L. whitmani</i>
5	73 sequências de <i>L. longipalpis</i>	+	<i>P. duboscqi</i> e <i>L. Intermedia</i>
6			<i>P. duboscqi</i> e <i>L. Whitmani</i>
7			<i>L. intermedia</i> e <i>L. whitmani</i>
8			<i>P. duboscqi</i> , <i>L. intermedia</i> e <i>L. whitmani</i>

Tabela 10 - Dezesesseis possíveis combinações de táxons de grupo externo na primeira busca de grupo externo referentes a *L. longipalpis* utilizando marcador COI.

Arquivos	Sequências do complexo de espécie	+	Táxons das sequências dos grupos externos
1			-
2			<i>P. argentipes</i>
3			<i>P. papatasi</i>
4			<i>P. sergenti</i>
5			<i>L. umbratilis</i> ,
6			<i>P. argentipes</i> e <i>P. papatasi</i>
7			<i>P. argentipes</i> e <i>P. sergenti</i>
8			<i>P. argentipes</i> e <i>L. umbratilis</i>
9	73 sequências de <i>L.</i>		<i>P. papatasi</i> e <i>P. sergenti</i>
10	<i>longipalpis</i>	+	<i>P. papatasi</i> e <i>L. umbratilis</i>
11			<i>P. sergenti</i> e <i>L. umbratilis</i>
12			<i>P. argentipes</i> , <i>P. papatasi</i> e <i>P. sergenti</i>
13			<i>P. argentipes</i> , <i>P. papatasi</i> e <i>L. umbratilis</i>
14			<i>P. argentipes</i> , <i>P. sergenti</i> e <i>L. umbratilis</i>
15			<i>P. papatasi</i> , <i>P. sergenti</i> e <i>L. umbratilis</i>
16			<i>P. argentipes</i> , <i>P. papatasi</i> , <i>P. sergenti</i> e <i>L. umbratilis</i>

A mesma lógica foi usada nas outras duas buscas de grupo externo nas montagens dos arquivos de sequência alinhadas.

#### 5.4. MODELOS EVOLUTIVOS PROBABILÍSTICOS

Abaixo estão apresentados os modelos evolutivos selecionados para cada alinhamento da primeira, segunda e terceira busca de grupos externos para *L. umbratilis* (Tabelas 11, 12 e 13, respectivamente) e *L. longipalpis* (tabelas 14, 15 e 16, respectivamente):

Tabela 11 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos *fasta* de *L. umbratilis* na primeira busca de *grupo* externo.

<b>Marcador</b>	<b>Grupo externo de cada arquivo <i>fasta</i></b>	<b>Modelo evolutivo</b>	<b>Marcador</b>	<b>Grupo externo de cada arquivo <i>fasta</i></b>	<b>Modelo evolutivo</b>
<b>COI</b>	Sem grupo externo	HKY+I	<b><i>Period</i></b>	Sem grupo externo	TPM2UF+I
	<i>P. argentipes</i> (A)	HKY+G		<i>P. dubosqui</i> (D)	TPM2UF+G
	<i>P. papatasi</i> (P)	HKY+G		<i>L. intermedia</i> (I)	TPM2UF+I
	<i>P. sergenti</i> (S)	HKY+G		<i>L. whitmani</i> (W)	TPM2UF+G
	<i>L. longipalpis</i> (L)	HKY+G		D+I	TPM3UF+G
	A+P	TPM3UF+G		D+W	TPM3UF+G
	A+S	TPM2UF+G		I+W	HKY+G
	P+S	TPM2UF+G		D+I+W	TPM3UF+G
	A+L	HKY+G			
	P+L	TPM1UF+G			
	A+P+S	TPM1UF+G			
	A+P+L	TIM1+G			
	A+S+L	TPM1UF+G			
	P+S+L	TVM+G			
	A+P+S+L	TPM1UF+G			
	S+L	TPM2UF+G			

Tabela 12 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos *fasta* de *L. umbratilis* na segunda busca de grupo externo.

<b>Marcador</b>	<b>Grupo externo (subgênero) de cada arquivo <i>fasta</i></b>	<b>Espécies correspondentes</b>	<b>Modelo evolutivo</b>
<b>COI</b>	<i>Phlebotomus</i> + <i>Lutzomyia</i>	<i>P. sergenti</i> + <i>L. gomezi</i>	TIM2+I
<b>Per</b>	<i>Phlebotomus</i> + <i>Lutzomyia</i>	<i>P. dubosqci</i> + <i>L. dispar</i>	TPM2uf+G

Tabela 13 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos *fasta* de *L. umbratilis* na terceira busca de grupo externo.

<b>Grupo externo</b>			
<b>Marcador</b>	<b>(subgênero) de cada arquivo <i>fasta</i></b>	<b>Espécies correspondentes</b>	<b>Modelo evolutivo</b>
<b>COI</b>	<i>Nyssomyia</i> (N)	<i>Lutzomyia anduzei</i> (A)	TPM3uf+I
	<i>Lutzomyia</i> (L)	<i>Lutzomyia gomezi</i> (G)	HKY+G
	<i>Phlebotomus</i> (P)	<i>Phlebotomus sergenti</i> (S)	TPM2uf+G
	L+N	G + A	HKY+G
	P+L	S + G	TPM2uf+G
	N+P	A + S	HKY+G
	N+P+L	A+ G + S	TPM2uf+G
	<b>Per</b>	<i>Phlebotomus</i> (P)	<i>Phlebotomus duboscqi</i> (D)
<i>Lutzomyia</i> (L)		<i>Lutzomyia renei</i> (R)	TPM1uf+G
<i>Nyssomyia</i> (N)		<i>Lutzomyia whitmani</i>	TPM2UF+G
		(W)	
P+L		D + R	TPM3uf+G
P+N		D + W	TPM3uf+G
L+N		R + W	TPM3uf+G
P+L+N		D + R + W	TPM3+G

Tabela 14 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos *fasta* de *L. longipalpis* na primeira busca de grupo externo.

<b>Marcador</b>	<b>Grupo externo de cada arquivo <i>fasta</i></b>	<b>Modelo evolutivo</b>	<b>Marcador</b>	<b>Grupo externo de cada arquivo <i>fasta</i></b>	<b>Modelo evolutivo</b>
<b>COI</b>	Sem grupo externo	HKY+G	<b><i>Period</i></b>	Sem grupo externo	K80+I+G
	<i>P. argentipe</i> (A)	HKY+G		<i>P. duboscqi</i> (D)	HKY+I+G
	<i>P. papatasi</i> (P)	HKY+G		<i>L. intermedia</i> (I)	HKY+I+G
	<i>P. sergenti</i> (S)	HKY+G		<i>L. whitmani</i> (W)	HKY+I+G
	<i>L. umbratilis</i> (U)	TrN+I		<i>D+I</i>	HKY+I+G
	<i>A+P</i>	TPM1UF+G		<i>D+W</i>	HKY+I+G
	<i>A+S</i>	TPM1UF+G		<i>I+W</i>	HKY+I+G
	<i>A+U</i>	TPM1UF+G		<i>D+I+W</i>	HKY+I+G
	<i>P+S</i>	TPM1UF+G			
	<i>P+U</i>	TPM1UF+G			
	<i>S+U</i>	TPM1UF+G			
	<i>A+P+S</i>	TPM1UF+G			
	<i>A+P+L</i>	TPM1UF+G			
	<i>A+S+L</i>	TPM1UF+G			
	<i>P+S+L</i>	TPM1UF+G			
	<i>A+P+S+L</i>	TPM1UF+G			

Tabela 15 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos *fasta* de *L. longipalpis* na segunda busca de grupo externo.

Marcador	Grupo externo		Modelo Evolutivo
	(subgênero) de cada arquivo <i>fasta</i>	Espécies correspondentes	
COI	<i>Phlebotomus</i> +	<i>P. segenti</i> +	TPM1uf+G
	<i>Nyssomyia</i>	( <i>L. umbratilis</i> e <i>L. antunesi</i> )	
Per	<i>Phlebotomus</i> +	<i>P. dubosqci</i> +	HKY+I+G
	<i>Nyssomyia</i>	( <i>L. whitmani</i> e <i>L. intermedia</i> )	

Tabela 16 - Modelos evolutivos selecionados para as sequências alinhadas contidas em cada um dos arquivos *fasta* de *L. longipalpis* na terceira busca de grupo externo.

<b>Marcador</b>	<b>Grupo externo (subgênero) de cada arquivo <i>fasta</i></b>	<b>Espécies correspondentes</b>	<b>Modelo evolutivo</b>
<b>COI</b>	<i>Nyssomyia (N)</i>	<i>L. anduzei (A)</i>	TrN+I
	<i>Lutzomyia (L)</i>	<i>L. gomezi (G)</i>	TrN+G
	<i>Phlebotomus (P)</i>	<i>P. sergenti (S)</i>	HKY+G
	<i>L+N</i>	<i>G + A</i>	TPM1uf+G
	<i>P+L</i>	<i>S + G</i>	TPM1uf+G
	<i>N+P</i>	<i>A + S</i>	TPM1uf+G
	<i>N+L+P</i>	<i>A+ G + S</i>	TPM1uf+G
<b>Per</b>	<i>Phlebotomus (P)</i>	<i>P. duboscqi (D)</i>	HKY+I+G
	<i>Lutzomyia (L)</i>	<i>L. renei (R)</i>	K80+I+G
	<i>Nyssomyia (N)</i>	<i>L. whitmani (W)</i>	HKY+I+G
	<i>P+L</i>	<i>D + R</i>	HKY+I+G
	<i>P+N</i>	<i>D + W</i>	TrN+G
	<i>L+N</i>	<i>R + W</i>	HKY+I+G
	<i>P+L+N</i>	<i>D + R + W</i>	HKY+I+G

### 5.5. ÁRVORES FILOGENÉTICAS DE *L. UMBRATILIS* E *L. LONGIPALPIS*

As árvores filogenéticas resultantes que se polarizaram dentro da lógica de entomologistas desses flebotomíneos (ver Anexo 5) apresentaram um padrão único. As relacionadas a *L. umbratilis* se apresentaram polarizadas em dois grupos: um com a população de Manacapuru (MAN) e o outro com as populações de Rio Preto da Eva (RIP) e Recife (REC). E a relacionada a *L. longipalpis* apresentou-se polarizada em dois grupos: um formado pelos indivíduos com fenótipo 1S (populações 1S de Caririçu e de Sobral) e o outro formado por indivíduos com fenótipo 2S (populações 2S de Caririçu e de Sobral), conforme descrito nas Figuras 3 e 4:

Figura 3 - Polarização única alcançada em todas as árvores filogenéticas de *L. umbratilis* resultantes que estavam entre as polarizações esperadas.

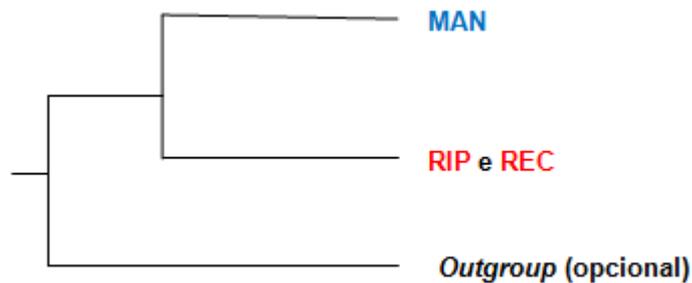
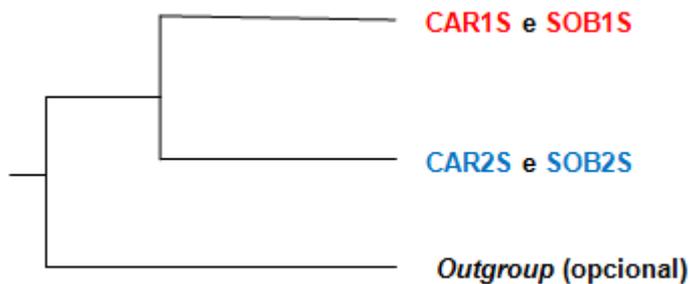


Figura 4 - Polarização única alcançada na única árvore, de todas as árvores filogenéticas de *L. longipalpis* resultantes, que estava entre as polarizações esperadas.



Os valores de suporte para cada uma das árvores obtidas na primeira, segunda e terceira busca de grupo externo para *L. umbratilis* estão descritos nas Tabelas 17, 18 e 19 e para *L. longipalpis*, nas Tabelas 20, 21 e 22.

Tabela 17 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da primeira busca de grupo externo para *L. umbratilis* (ver figura 3). Os valores em vermelho correspondem ao suporte de confiança do clado contendo os táxons correspondentes à população de Recife (REC) e de Rio Preto da Eva (RIC) e os valores em azul ao do clado contendo táxons da população de Manacapuru (MAN). Valores sublinhados: maiores valores de suporte de ambos cladogramas, ou seja, os melhores valores (acima de 80%). E presença de hífen: corresponde a obtenção de árvore filogenética sem polarização esperada (ver figura 3).

<b>Marcador</b>	<b>Grupo externo de cada arquivo <i>fasta</i></b>	<b>MV (<i>bootstrap</i> - 100 replicatas)</b>		<b>Bayes (Probabilidade a posteriori)</b>	
<b>COI</b>	Sem grupo externo	<u>100</u>	<u>100</u>	<u>1</u>	<u>1</u>
	<i>P. argentipes</i> (A)	-	-	-	
	<i>P. papatasi</i> (P)	-	-	0,2289	0,9154
	<i>P. sergenti</i> (S)	-	-	0,9983	0,5555
	<i>L. longipalpis</i> (L)	-	-	0,3249	0,7897
	A+P	-	-	-	
	A+S	-	-	-	
	P+S	-	-	-	
	A+L	-	-	-	
	P+L	-	-	-	
	A+P+S	-	-	-	
	A+P+L	-	-	-	
	A+S+L	61	68	<u>0,9352</u>	<u>0,9785</u>
	P+S+L	-	-	<u>0,8615</u>	<u>0,9356</u>

		A+P+S+L	-	-	-	
		S+L	-	-	<u>0,9024</u>	<u>0,9424</u>
		Total:		2	Total:	7
<b>Period</b>	Sem grupo externo	<u>100</u>	<u>100</u>		<u>1</u>	<u>1</u>
	<i>P. dubosqui (D)</i>	-	-		-	-
	<i>L. intermedia (I)</i>	-	-		-	-
	<i>L. whitmani (W)</i>	90	49		<u>1</u>	<u>0,9724</u>
	<i>D+I</i>	69	23		0,9986	0,4571
	<i>D+W</i>	64	48		0,3659	0,3576
	<i>I+W</i>	93	32		<u>0,9998</u>	<u>0,9509</u>
	<i>D+I+W</i>	68	51		0,8052	0,6397
		Total:		6	Total:	6

Tabela 18 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da segunda busca de grupo externo para *L. umbratilis* (ver figura 3). Os valores em vermelho correspondem ao suporte de confiança do clado contendo os táxons correspondentes à população de Recife (REC) e de Rio Preto da Eva (RIC) e os valores em azul ao do clado contendo táxons da população de Manacapuru (MAN). Valores sublinhados: maiores valores de suporte de ambos cladogramas, ou seja, os melhores valores (acima de 80%). E presença de hífen: corresponde a obtenção de árvore filogenética sem polarização esperada (ver figura 3).

Marcador	Grupo externo		Bayes (Probabilidade a posteriori)			
	(subgênero) de cada arquivo <i>fasta</i>	Espécies correspondentes	MV ( <i>bootstrap</i> - 1000 replicatas)			
COI	<i>Phlebotomus</i> +	<i>P. sergenti</i> +	<u>852</u>	<u>855</u>	<u>1</u>	<u>0,9999</u>
	<i>Lutzomyia</i>	<i>L. gomezi</i>				
		Total	1	Total	1	
Per	<i>Phlebotomus</i> +	<i>P. dubosqci</i> +	-	-	-	-
	<i>Lutzomyia</i>	<i>L. dispar</i>				
		Total	0	Total	0	

Tabela 19 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da terceira busca de grupo externo para *L. umbratilis* (ver figura 3). Os valores em vermelho correspondem ao suporte de confiança do clado contendo os táxons correspondentes à população de Recife (REC) e de Rio Preto da Eva (RIC) e os valores em azul ao do clado contendo táxons da população de Manacapuru (MAN). Valores sublinhados: maiores valores de suporte de ambos clados, ou seja, os melhores valores (acima de 80%). E presença de hífen: corresponde a obtenção de árvore filogenética sem polarização esperada (ver figura 3).

Marcador	Grupo externo		MV ( <i>bootstrap</i>		Bayes	
	(subgênero) de cada arquivo <i>fasta</i>	Espécies Correspondentes	- 1000 replicatas)		(Probabilidade <i>a posteriori</i> )	
<b>COI</b>	<i>Nyssomyia</i> (N)	<i>Lutzomyia anduzei</i>	<u>981</u>	<u>981</u>	0,7040	0,9969
		(A)				
	<i>Lutzomyia</i> (L)	<i>Lutzomyia gomezi</i>	712	698	<u>0,9902</u>	<u>0,9989</u>
		(G)				
	<i>Phlebotomus</i> (P)	<i>Phlebotomus sergenti</i> (S)	851	371	<u>0,9998</u>	<u>0,9613</u>
		L+N	G + A	<u>957</u>	<u>964</u>	<u>1</u>
	P+L	S + G	909	684	<u>1</u>	<u>0,9984</u>
	N+P	A + S	793	591	<u>0,9979</u>	<u>0,9299</u>
N+P+L	A+ G + S	<u>889</u>	<u>906</u>	<u>0,9968</u>	<u>0,9999</u>	
		Total	7	Total	7	
<b>Per</b>	<i>Phlebotomus</i> (P)	<i>Phlebotomus</i>	-	-	0,6255	0,5222
		<i>duboscqi</i> (D)				

<i>Lutzomyia</i> (L)	<i>Lutzomyia renei</i> (R)	-	-	-	-
	<i>Lutzomyia whitmani</i>				
<i>Nyssomyia</i> (N)	(W)	520	888	<u>0,9724</u>	<u>1</u>
P+L	D + R	391	832	-	-
P+N	D + W	585	498	<u>0,9843</u>	<u>0,9898</u>
L+N	R + W	562	770	<u>0,7772</u>	<u>0,9734</u>
P+L+N	D + R + W	431	807	<u>0,6689</u>	<u>0,9937</u>
	Total	5		Total	5

Tabela 20 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da primeira busca de grupo externo para *L. longipalpis* (ver figura 4). Os valores em vermelho correspondem ao suporte de confiança do clado contendo os táxons correspondentes ao fenótipo 1S (da cidade Caririáçu – CAR1S e de Sobral - SOB1S) e os valores em azul ao do clado contendo táxons com fenótipo 2S (da cidade de Caririáçu – CAR2S e de Sobral - SOB2S). Valores sublinhados: maiores valores de suporte de ambos clados, ou seja, os melhores valores (acima de 80%). E presença de hífen: corresponde a obtenção de árvore filogenética sem polarização esperada (ver figura 4).

Marcador	Grupo externo de cada arquivo <i>fasta</i>	MV ( <i>bootstrap</i> - 100 replicatas)				Bayes (Probabilidade <i>a posteriori</i> )	
COI	Sem grupo externo	-	-	-	-	-	
	<i>P. argentipe</i> (A)	-	-	-	-	-	
	<i>P. papatasi</i> (P)	-	-	-	-	-	
	<i>P. sergenti</i> (S)	-	-	-	-	-	
	<i>L. umbratilis</i> (U)	-	-	-	-	-	
	A+P	-	-	-	-	-	
	A+S	-	-	-	-	-	
	A+U	-	-	-	-	-	
	P+S	-	-	-	-	-	
	P+U	-	-	-	-	-	
	S+U	-	-	-	-	-	
	A+P+S	-	-	-	-	-	
A+P+L	-	-	-	-	-		

	<i>A+S+L</i>	-	-	-	-
	<i>P+S+L</i>	-	-	-	-
	<i>A+P+S+L</i>	-	-	-	-
	Total		0	Total	0
<b>Period</b>	Sem grupo externo	-	-	-	-
	<i>P. duboscqi (D)</i>	-	-	0,5301	0,4855
	<i>L. intermedia (I)</i>	-	-	-	-
	<i>L. whitmani (w)</i>	-	-	-	-
	<i>D+I</i>	-	-	-	-
	<i>D+W</i>	-	-	-	-
	<i>I+W</i>	-	-	-	-
	<i>D+I+W</i>	-	-	-	-
	Total		0	Total	1

Tabela 21 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da segunda busca de grupo externo para *L. longipalpis* (ver figura 4). Os valores em vermelho correspondem ao suporte de confiança do clado contendo os táxons correspondentes ao fenótipo 1S (da cidade Caririaçu – CAR1S e de Sobral - SOB1S) e os valores em azul ao do clado contendo táxons com fenótipo 2S (da cidade de Caririaçu – CAR2S e de Sobral - SOB2S). Valores sublinhados: maiores valores de suporte de ambos clados, ou seja, os melhores valores (acima de 80%). E presença de hífen: corresponde a obtenção de árvore filogenética sem polarização esperada (ver figura 4).

Marcador	Grupo externo		MV (bootstrap - 100 replicatas)		Bayes (Probabilidade a posteriori)	
	(subgênero) de cada arquivo <i>fasta</i>	Espécies correspondentes				
<b>COI</b>	<i>Phlebotomus</i> + <i>Nyssomyia</i>	<i>P. segenti</i> +				
		( <i>L. umbratilis</i> e <i>L. antunesi</i> )	-	-	-	-
		Total	0	Total		
<b>Per</b>	<i>Phlebotomus</i> + <i>Nyssomyia</i>	<i>P. dubosqci</i> +				
		( <i>L. whitmani</i> e <i>L. intermedia</i> )	-	-	-	-
		Total	0	Total		

Tabela 22 - Valores de suporte de teste de confiança (para MV e BE) dos dois grupos esperados nas árvores filogenéticas obtidas por MV e BE a partir dos alinhamentos da terceira busca de grupo externo para *L. longipalpis* (ver figura 4). Os valores em vermelho correspondem ao suporte de confiança do clado contendo os táxons correspondentes ao fenótipo 1S (da cidade Caririaçu – CAR1S e de Sobral - SOB1S) e os valores em azul ao do clado contendo táxons com fenótipo 2S (da cidade de Caririaçu – CAR2S e de Sobral - SOB2S). Valores sublinhados: maiores valores de suporte de ambos clados, ou seja, os melhores valores (acima de 80%). E presença de hífen: corresponde a obtenção de árvore filogenética sem polarização esperada (ver figura 4).

Marcador	Grupo externo		MV		Bayes (Probabilidade a posteriori)	
	(subgênero) de cada arquivo <i>fasta</i>	Espécies correspondentes	(bootstrap - 1000 replicatas)			
<b>COI</b>	<i>Nyssomyia</i> (N)	<i>L. anduzei</i> (A)	-	-	-	-
	<i>Lutzomyia</i> (L)	<i>L. gomezi</i> (G)	-	-	-	-
	<i>Phlebotomus</i> (P)	<i>P. sergenti</i> (S)	-	-	-	-
	L+N	G+A	-	-	-	-
	P+L	S+G	-	-	-	-
	N+P	A+S	-	-	-	-
	N+L+P	A+G+S	-	-	-	-
	Total		0	Total	0	
<b>Per</b>	<i>Phlebotomus</i> (P)	<i>P. duboscqi</i> (D)	-	-	0,4855	0,5301
	<i>Lutzomyia</i> (L)	<i>L. renei</i> (R)	-	-	-	-
	<i>Nyssomyia</i> (N)	<i>L. whitmani</i> (W)	-	-	-	-
	P+L	D+R	-	-	-	-

$P+N$	$D+W$	-	-	-	-
$L+N$	$R+W$	-	-	-	-
$P+L+N$	$D+R+W$	-	-	-	-
Total		0	Total	1	

## 5.6. RESULTADO DA ANÁLISE DAS ÁRVORES INFERIDAS A PARTIR DE DADOS DE CADA MARCADOR ISOLADO OBSERVANDO VALORES DE SUPORTE DE CADA POLARIDADE

Para as árvores de *L. umbratilis* foram alcançadas as seguintes observações (ver tópico 5.5 para entender a polarização esperada da árvore filogenética):

### a) Primeira busca de grupo externo

Utilizando o marcador COI (Tabela 23):

Tabela 23 - Análise da primeira busca de grupo externo para complexo *L. umbratilis* utilizando marcador COI.

Métodos de inferência	Proporções de árvores polarizadas como o esperado		Grupo externo	Valores de		Características comuns	
				suporte acima de 80%	Valores de suporte		
MV (16 árvores)	Proporção sem grupo externo		-	Ambos	100	100	
	Proporção com grupo externo		A+S+L	Nenhum ou 1	61	68	Grupo externo com mais de 1 taxon; ambos suportes não são altos; presença de táxons S+L; presença grupo externo com ancestral mais recente.
BE (16 árvores)	Proporção sem grupo externo			Ambos	1	1	
	Proporção com grupo externo		A+S+L	Ambos	0,9352	0,9785	Grupo externo com + de 1 taxon; Ambos
			S+L	Ambos	0,9024	0,9424	suportes altos; presença de
			P+S+L	Ambos	0,8615	0,9356	

---

				S+L; sempre presença grupo externo com ancestral recente.
<i>P</i>	Nenhum ou 1	0,2289	0,9154	Grupo externo com apenas 1 táxon; Ambos suportes não são altos; presença grupo externo com ancestral mais recente ou não.
<i>S</i>	Nenhum ou 1	0,9983	0,5555	
<i>L</i>	Nenhum ou 1	0,3249	0,7897	

---

Utilizando o marcador COI todas as árvores com grupo externo polarizadas em dois grupos esperados e com suportes maiores do que 80% possuem os táxons S e L no grupo externo. Ambos os métodos de inferência conseguiram boa polarização sem grupo externo ou utilizando grupo externo A+S+L.

O método de inferência Bayesiana mostrou-se com mais vantagem frente ao método MV, a saber:

- Maior quantidade de árvores polarizadas como esperado no aspecto biológico (Figura 3) - seis dentre as quinze árvores com grupo externo enquanto que por MV, apenas uma das quinze árvores com grupo externo;
- Três das seis árvores possuem suportes superiores a 80%, enquanto que a árvore gerada por MV não possui ambos os suportes superiores a 80%.

Para o marcador *per*, veja a Tabela 24:

Tabela 24 - Análise da primeira busca de grupo externo para complexo *L. umbratilis* utilizando marcador *per*.

Métodos de inferência	Proporções de árvores polarizadas como o esperado		Grupo externo	Valores de suporte acima de 80%	Valores de suporte		Características comuns	
MV (8 árvores)	Proporção	1/1	-	Ambos	100	100		
	sem grupo externo							
	Proporção	5/7	<i>W</i>	Nenhum ou 1	90	49	Grupo externo	
	com grupo externo			<i>D+I</i>	Nenhum ou 1	69	23	com 1 ou mais táxons; ambos
				<i>D+W</i>	Nenhum ou 1	64	48	suportes não são
			<i>I+W</i>	Nenhum ou 1	93	32	altos(os casos	
		<i>D+I+W</i>	Nenhum ou 1	68	51	em que um deles foi acima de 80%: <i>W</i> , <i>W+I</i> ); presença grupo externo com ancestral mais recente e antigo.		
BE (8 árvores)	Proporção	1/1	-	Ambos	1	1		
	sem grupo externo							
	Proporção	5/7	<i>W</i>	Ambos	1	0,9724	Grupo externo com 1 ou 2	
com grupo externo			<i>I+W</i>	Ambos	0,9998	0,9509	táxons; ambos suportes altos;	

---

				presença de w; apenas com grupo externo com acentrais recentes.
$D+W$	Nenhum ou 1	0,3659	0,3576	Grupo externo
$D+I$	Nenhum ou 1	0,9986	0,4571	com 2 táxons; ambos suportes não são altos;
$D+I+W$	Nenhum ou 1	0,8052	0,6397	presença de D; presença de um táxon grupo externo com ancestral antigo.

---

Para o marcador *period*, a maioria das árvores com grupo externo polarizadas em dois grupos esperados possui o táxon W no grupo externo. Ambos os métodos de inferência conseguiram boa polarização sem grupo externo ou utilizando os grupos externos W, I+W, D+I, D+W, D+W+I. O método de inferência Bayesiana mostrou-se com mais vantagem frente ao método MV:

- Apesar de mesma quantidade de árvores polarizadas como esperado no aspecto biológico (Figura 3) para ambos os métodos - cinco dentre sete árvores com grupo externo;
- Duas dessas cinco árvores possuem ambos suportes maior do que 80%, ou seja, um suporte alto, enquanto que nenhuma das cinco árvores por MV possuem ambos suportes maior do que 80%.

Na primeira busca de grupo externo utilizando marcador COI foi obtido um número maior de árvores com polarização esperada e com ambos os suportes altos em relação ao marcador *period*. E todas as árvores com essas características simultaneamente foram obtidas apenas pela inferência BE.

#### **b) Segunda busca de grupo externo**

Na segunda busca, apesar dos grupos externos pertencentes aos subgêneros *Phlebotomus* e *Lutzomyia* terem sido suficientes para polarização e suporte acima de 80% pelo método de distância tanto utilizando o marcador COI (grupo externo: *P. sergenti* e *L. gomezi*) como o marcador *period*(grupo externo: *P. dubosqi* e *L. dispar*), as árvores do complexo *L. umbratilis* inferidas pelo método MV e BE utilizando esses mesmos grupo externo (com similaridades em nível de subgênero) e marcadores não foram concordantes com a inferência pelo método de distância. Apesar de que utilizando o marcador COI foi obtida árvore com alto valor de suporte (acima de 80%) para os dois clados esperados (Figura 3), tanto pelo método MV (85,2% e 85,5%) como BE (100% e 99,99%), o mesmo não ocorreu utilizando o marcador *period*. Utilizando este marcador, a árvore não foi polarizada como esperada, tanto a inferida por MV como por BE. Utilizando marcador COI e método de inferência Bayesiana foi obtido o melhor resultado na segunda busca do grupo externo.

#### **c) Terceira busca de grupo externo**

Utilizando o marcador COI (Tabela 25):

Tabela 25 - Análise da terceira busca de grupo externo para complexo *L. umbratilis* utilizando marcador COI.

Métodos de inferência	Proporções de árvores polarizadas como o esperado		Grupo externo	Valores de suporte acima de 80%		Valores de suporte	Características comuns	
	Proporção	7/7		Ambos	981			981
MV (7 árvores)	Proporção com grupo externo	7/7	<i>N</i>	Ambos	981	981	Grupo externo com 1 ou mais táxons; ambos suportes são altos; presença de <i>N</i> ; presença grupo externo com ancestral mais recente.	
			<i>L+N</i>	Ambos	957	964		
			<i>N+P+L</i>	Ambos	889	906		
			<i>P</i>	Nenhum ou 1	851	371		Grupo externo com 1 ou mais táxons; ambos suportes não são altos; presença grupo externo com ancestral mais antigo
			<i>L</i>	Nenhum ou 1	712	698		
	<i>P+L</i>	Nenhum ou 1	909	684				
	<i>N+P</i>	Nenhum ou 1	793	591				
BE (7 árvores)	Proporção com grupo externo	7/7	<i>L+N</i>	Ambos	1	1	Grupo externo com 1 ou mais táxons; Ambos suportes altos; Sempre presença outgroup	
			<i>P+L</i>	Ambos	1	0,9984		
			<i>N+P+L</i>	Ambos	0,9968	0,9999		
			<i>L</i>	Ambos	0,9902	0,9989		

<i>P</i>	<b>Ambos</b>	0,9998	0,9613	com ancestral antigo e médio.
<i>N+P</i>	<b>Ambos</b>	0,9979	0,9299	
<i>N</i>	Nenhum ou 1	0,704	0,9969	Grupo externo com apenas 1 taxon; Ambos suportes não são altos; presença grupo externo com ancestral mais recente sozinho.

Utilizando o marcador COI (na terceira busca de grupo externo), ambos os métodos de inferência conseguiram boa polarização utilizando grupos externos L+N, N+P+L. O método de inferência Bayesiana mostrou-se com mais vantagem frente ao método MV:

- Apesar de mesma quantidade de árvores polarizadas como esperado no aspecto biológico (ver figura 3) para ambos os métodos - sete dentre sete árvores com grupo externo;
- Seis dessas seis árvores possuem ambos os suportes acima de 80%, enquanto que em apenas três das sete árvores geradas por MV possuem ambos os suportes maior do que 80%.

Utilizando o marcador *period* (Tabela 26):

Tabela 26 - Análise da terceira busca de grupo externo para complexo *L. umbratilis* utilizando marcador *per*.

Métodos de inferência	Proporções de árvores polarizadas como o esperado			Grupo externo	Valores de suporte acima de 80%	Valores de suporte		Características comuns
	Proporção							
MV (7 árvores)	Proporção com grupo externo	5/7	<i>N</i>	Nenhum ou 1	520	888	Grupo externo com 1 ou mais taxons; Ambos suportes não são altos; presença grupo externo com ancestral mais recente e antigo	
			<i>P+L</i>	Nenhum ou 1	391	832		
			<i>P+N</i>	Nenhum ou 1	585	498		
			<i>L+N</i>	Nenhum ou 1	562	770		
			<i>P+L+N</i>	Nenhum ou 1	431	807		
BE (7 árvores)	Proporção com grupo externo	5/7	<i>N</i>	Ambos	0,9724	1	Grupo externo com 1 ou 2 taxons; ambos suportes altos; N; apenas com grupo externo com acentrais recentes.	
			<i>P+N</i>	Ambos	0,9843	0,9898		
			<i>P</i>	Nenhum ou 1	0,6255	0,5222		
			<i>L+N</i>	Nenhum ou 1	0,7772	0,9734		
			<i>P+L+N</i>	Nenhum ou 1	0,6689	0,9937		

---

altos; Grupo  
externo com  
ancestral antigo  
e recente.

---

Utilizando o marcador *per*, ambos os métodos de inferência conseguiram boa polarização utilizando os grupos externos N, P+N, L+N, P+L+N. O método de inferência Bayesiana mostrou-se com mais vantagem frente ao método MV:

- Apesar de mesma quantidade de árvores polarizadas como esperado no aspecto biológico (ver figura 3) para ambos os métodos - cinco dentre sete árvores com grupos externos;
- Duas dessas cinco árvores exibiram suportes maiores do que 80%, enquanto que nenhuma das cinco árvores geradas por MV exibiu suportes maiores do que 80%.

Nesta terceira busca de grupo externo, utilizando marcador COI foi obtido um número maior de árvores com polarização esperada e com ambos os suportes altos em relação ao marcador *per*. Apesar de nesta terceira busca ter obtido árvores inferidas por MV com ambos os suportes altos (para MV não foi alcançado essas duas características simultaneamente nem na primeira e nem na segunda busca tanto utilizando COI como *per*), BE superou em quantidade e em nível de suporte de árvores com polaridades esperadas.

Na terceira busca pela combinação da utilização de marcador COI e da inferência BE foram alcançadas as melhores árvores. Todas as melhores árvores para *L. umbratilis*, quanto à polarização e quanto ao suporte (maior do que 80%) estão no anexo 6.

Para as árvores de *L. longipalpis* foram alcançadas as seguintes observações:

**a) Primeira busca de grupo externo**

Utilizando o marcador COI, veja tabela 27:

Tabela 27 - Análise da primeira busca de grupo externo para complexo *L. longipalpis* utilizando marcador COI.

Métodos de inferência	Proporções de árvores polarizadas como o esperado		Grupo externo	Valores de suporte		Valores de suporte	Características comuns
				suporte acima de 80%			
MV (16 árvores)	Proporção	0/1		-	-	-	-
	sem grupo externo						
	Proporção	0/15	-	-	-	-	-
	com grupo externo						
BE (16 árvores)	Proporção	0/1		-	-	-	-
	sem grupo externo						
	Proporção	0/15	-	-	-	-	-
	com grupo externo						

Utilizando o marcador COI tanto com o método de inferência MV como BE não foi possível obter árvores polarizadas em dois grupos esperados, conforme figura 4.

Utilizando o marcador *per*, veja tabela 28:

Tabela 28 - Análise da primeira busca de grupo externo para complexo *L. longipalpis* utilizando marcador *per*.

Métodos de inferência	Proporções de árvores polarizadas como o esperado	Grupo externo	Valores de suporte acima de 80%	Valores de suporte	Características comuns
<b>MV</b>	<b>Proporção</b>				
<b>(8 árvores)</b>	<b>sem grupo externo</b>	0/1	-	-	-
	<b>Proporção com grupo externo</b>	0/7	-	-	-
<b>BE</b>	<b>Proporção</b>				
<b>(8 árvores)</b>	<b>sem grupo externo</b>	0/1	-	-	-
	<b>Proporção com grupo externo</b>	1/7	W	Nenhum ou 1	0,5301 0,4855

Utilizando o marcador *per* apenas pela inferência BE foi possível obter uma árvore polarizada em dois grupos esperados, embora com suporte muito baixo para ambos clados.

Na primeira busca de grupo externo utilizando o marcador COI não foram alcançadas árvores com polarização esperada e com ambos os suportes altos. Apenas utilizando o marcador *per* foi possível alcançar uma árvore polarizada como

o esperado via inferência BE. Apesar de a combinação do uso do marcador *per* e inferência BE ter surtido melhor resultado, a árvore não possui simultaneamente ambas as características necessárias para considerarmos árvores boas: polarização esperada e suporte de ambos clados esperados maior do que 80%.

### **b) Segunda busca de grupo externo**

Apesar de os grupos externos pertencentes aos subgêneros *Phlebotomus* e *Nyssomyia* terem sido suficientes para polarização e suporte acima de 80% pelo método de distância tanto utilizando o marcador COI (grupo externo: *P. segenti* + *L. umbratilis* e *L. antunesi*) como o marcador *per* (grupo externo: *P. dubosqci* + *L. whitmani* e *L. intermedia*), as árvores do complexo *L. longipalpis* inferidas pelo método MV e BE utilizando esses mesmos grupo externo (com similaridades em nível de subgênero) e marcadores não foram concordantes com a inferência por método de distância. Utilizando o marcador COI ou *per*, a árvore não foi polarizada como esperada, tanto a inferindo por MV como por BE. Assim, na segunda busca de grupo externo não houve boas árvores.

### **c) Terceira busca de grupo externo**

Utilizando o marcador COI, veja tabela 29:

Tabela 29 - Análise da terceira busca de grupo externo para complexo *L. longipalpis* utilizando marcador COI.

Métodos de inferência	Proporções de árvores polarizadas como o esperado	Grupo externo	Valores			
			de suporte acima de 80%	Valores de suporte	Características comuns	
MV (7 árvores)	Proporção com grupo externo	0/7	-	-	-	-
	Proporção com grupo externo	0/7	-	-	-	-

Utilizando o marcador COI tanto com o método de inferência MV como BE não foi possível obter árvores polarizadas em dois grupos esperados, conforme figura 4.

Utilizando o marcador *per*, veja tabela 30:

Tabela 30 - Análise da terceira busca de grupo externo para complexo *L. longipalpis* utilizando marcador *per*.

Métodos de inferência	Proporções de árvores polarizadas como o esperado	Grupo externo	Valores de suporte acima de 80%	Valores de suporte	Características comuns
MV (7 árvores)	Proporção com grupo externo 0/7	-	-	-	-
BE (7 árvores)	Proporção com grupo externo 1/7	W	Nenhum ou 1	0,5301	0,4855

Utilizando o marcador *per* apenas pela inferência BE foi possível obter uma árvore polarizada em dois grupos esperados, embora com suporte muito baixo para ambos clados. Essa árvore corresponde a única árvore polarizada de modo esperado da primeira busca de grupo externo.

Nesta terceira busca de grupo externo utilizando o marcador COI não foram alcançadas árvores com polarização esperada e com ambos os suportes altos. Apenas utilizando o marcador *per* foi possível alcançar uma árvore polarizada como o esperado via inferência BE. Apesar de a combinação do uso do marcador *per* e inferência BE ter surtido melhor resultado, a árvore não possui simultaneamente

ambas as características necessárias para considerarmos árvores boas: polarização esperada e suporte de ambos clados esperados maior do que 80%.

A melhor árvore alcançada para *L. longipalpis* está no anexo 7.

#### 5.7. DADOS CONCATENADOS DE SEQUÊNCIAS DOS MARCADORES COI E PER DE *L. UMBRATILIS* E *L. LONGIPALPIS* E ÁRVORES FILOGENÉTICAS OBTIDAS A PARTIR DESTES DADOS

Ao todo foram criados quatro arquivos com sequências concatenadas: sequências de *L. umbratilis* sem grupo externo; sequências de *L. umbratilis* com grupo externo (subgêneros *Nyssomyia*, *Lutzomyia* e *Phlebotomus*); sequências de *L. longipalpis* sem grupo externo; e sequências de *L. longipalpis* com grupo externo (subgêneros *Phlebotomus*).

As sequências dos marcadores *period* e COI de *L. umbratilis* mostraram-se congruentes entre si na análise do teste de partição ( $p=0,9900$ , independente da utilização das sequências dos grupos externos). Dados de sequências do marcador nuclear *per* e do marcador mitocondrial COI de *L. longipalpis* concatenados mostraram-se congruentes entre si considerando a ausência das sequências grupo completo, embora com *p-value* muito baixo  $p=0,08$ . Porém considerando a presença do grupo externo, os dados concatenados não se mostraram congruentes ( $p=0,04$ ).

Nas tabelas 31 e 32 estão os modelos evolutivos escolhidos e os valores de suporte de cada grupo formado para *L. umbratilis* e *L. longipalpis* respectivamente nas árvores filogenéticas obtidas (segundo figuras 3 e 4, respectivamente). Essas árvores estão nos anexos 6 e 7 respectivamente.

Tabela 31 - Modelos evolutivos das sequências concatenadas de *L. umbratilis* e valores de suporte de confiança dos dois principais clados com importância biológica apresentadas nas árvores filogenéticas inferidas utilizando esses dados concatenados. Números em azul - referentes ao clado com táxons de Manacapuru (MAN) e em vermelho referentes ao clado com táxons de Recife (REC) e Rio Preto da Eva (RIC).

Macador	Modelo evolutivo	MV ( <i>bootstrap</i> em proporção decimal)		BE (probabilidade a posteriori)	
<b>Concatenados</b>					
sem grupo externo	HKY+I+G	1	1	1	1
<b>Concatenados</b>					
com grupo externo (N+L+P)	TVM+G	0,943	1	1	0,9998

Tabela 32 – Modelos evolutivos das sequências concatenadas de *L. longipalpis* e valores de suporte de confiança dos dois principais clados com importância biológica apresentadas nas árvores filogenéticas inferidas utilizando esses dados concatenados. Números em vermelho referentes ao clado com táxons 1S (da cidade de Caririáçu – CAR1S- Sobral - SOB1S) e em azul - referentes ao clado com táxons 2S (da cidade de Caririáçu – CAR2S- Sobral - SOB2S).

Macador	Modelo evolutivo	MV ( <i>bootstrap</i> em proporção decimal)		BE (probabilidade a posteriori)	
Concatenados sem grupo externo	TrN+I+G	0,5700	0,5700	0,9943	0,9943
Concatenados com grupo externo (grupo externo: P)	HKY+I+G	-	-	0,4316	0,5820

5.8. ANÁLISES DE ÁRVORES INFERIDAS POR DADOS CONCATENADOS VERSUS AS POR DADOS DE CADA MARCADOR ISOLADO OBSERVANDO OS VALORES DE SUPORTE DE CADA POLARIDADE

a) *L. umbratilis*

Na Tabela 33 são apresentados os valores de suportes dos dois clados com importância biológica (Figura 3) tanto na árvore inferida por cada método utilizando dados de sequências dos marcadores COI e *period* isolados e também concatenados (sem e com grupo externo):

Tabela 33 - Valores de suporte de confiança dos dois principais clados com importância biológica para cada uma das árvores inferidas por dados concatenados ou isolados do complexo *L. umbratilis*. Essas árvores envolvem as mesmas sequências do grupo interno e grupos externos tanto isolados como também concatenados (COI e *per*). Números em vermelho referentes ao clado com táxons de Recife (REC) e Rio Preto da Eva (RIC) e em azul - referentes ao clado com táxons de Manacapuru (MAN).

Macador	MV ( <i>bootstrap</i> em proporção decimal)		BE (probabilidade <i>a posteriori</i> )	
COI (Grupo externo: N+L+P)	0,906	0,889	0,9999	0,9968
<i>Period</i> (Grupo externo: N+L+P)	0,807	0,431	0,9937	0,6689
Concatenados sem grupo externo	1	1	1	1
Concatenados com grupo externo (N+L+P)	1	0,943	0,9998	1

Apenas lembrando: o grupo externo das árvores às quais se refere à Tabela 33 para COI corresponde ao grupo externo para *per* em nível de subgênero -

*Phlebotomus* (*P. sergentie*, *P. duboscqi*, respectivamente), *Lutzomyia* (*L. gomezie*, *L. renei*, respectivamente) e *Nyssomyia* (*L. anduzei*, *L. whitmani*, respectivamente).

Ao observar a Tabela 33 percebe-se que tanto utilizando os dados de marcador COI e *per* isolados como concatenados os maiores valores de suporte foram alcançados pela inferência Bayesiana.

Os suportes das árvores inferidas utilizando o marcador COI (88,9% e 90,6%, inferida por MV; 99,68% e 99,99%, inferida por BE) tem o suporte maior do que utilizando *per* (43,1% e 80,7%, inferida por MV; 66,89% e 99,37%, inferida por BE), inclusive ambas acima de 80%, ou seja, alto suporte.

Utilizando dados concatenados, com (94,3% e 100%, inferida por MV; 100 e 99,98% inferida por BE) ou sem grupo externo (100% e 100%, inferida tanto por MV como por BE), foram obtidos valores de suporte ainda maiores do que utilizando apenas dados COI.

#### **b) *L. longipalpis***

Na Tabela 34 estão os valores de suportes dos dois clados com importância biológica (ver figura 4) tanto na árvore inferida por cada método utilizando dados de se sequências dos marcadores COI e *per* isolados, como também concatenados (sem e com grupo externo):

Tabela 34 - Valores de suporte de confiança dos dois principais clados com importância biológica para cada uma das árvores inferidas com dados concatenados e isolados do complexo *L. longipalpis*. Essas árvores envolvem as mesmas sequências grupo interno e grupos externos tanto isolados como também concatenados (COI e *per*). Números em vermelho referentes ao clado com táxons 1S (da cidade de Caririaçu – CAR1S- Sobral - SOB1S) e em azul - referentes ao clado com táxons 2S (da cidade de Caririaçu – CAR2S- Sobral - SOB2S).

Macador	MV ( <i>bootstrap</i> em proporção decimal)		BE (probabilidade a posteriori)	
<b>COI (Grupo externo: P)</b>	-	-	-	-
<b><i>Period</i> (Grupo externo: P)</b>	-	-	0,4539	0,5136
<b>Concatenados sem grupo externo</b>	0,5700	0,5700	0,9943	0,9943
<b>Concatenados com grupo externo* (Grupo externo: P)</b>	-	-	0,4316	0,5820

\*Embora essas sequências concatenadas tenham apresentado  $p < 0,05$ , ou seja,  $p = 0,04$  no teste de homogeneidade de partição, a partir delas também foram inferidas árvores filogenéticas.

Apenas lembrando: o grupo externo para COI correspondem ao grupo externo para *per* em nível de subgênero - *Phlebotomus* (*P. duboscqi* e *P. sergenti*, respectivamente).

Ao observar a Tabela 34 observa-se que utilizando os dados de marcador COI isolado não foi possível inferir árvores polarizadas como o esperado e consequentemente não tem valor de suportenem por MV e nem por BE. Enquanto que utilizando marcador *per* apenas a árvore inferida por BE obteve polarização esperada, embora com suportes baixos (51,36% e 45,39%). Utilizando dados

concatenados dos marcadores COI e *per* com ou sem grupo externo, as árvores inferidas por BE foram melhores. E a árvore inferida por BE utilizando marcadores concatenados sem grupo externo apresentou melhor suporte (além disso, acima de 80%).

Utilizando dados concatenados, com (43,16% e 58,20% inferida por BE, enquanto que por MV não polarizou como esperado) ou sem grupo externo (57% e 57%, inferida por MV; 99,43% e 99,43%, inferida tanto por BE), foram obtidos valores de suporte ainda maiores do que utilizando apenas dados de sequências *per* combinados com inferência BE.

## 5.9. ANÁLISES DE ÁRVORES INFERIDAS A PARTIR DE DADOS CONCATENADOS E DE DADOS DE CADA MARCADOR ISOLADO UTILIZANDO O TESTE DE CONGRUÊNCIA ENTRE TOPOLOGIAS

### **a) *L. umbratilis***

Na Tabela 35 estão os valores referentes aos testes de congruência de topologia. As pontuações *Split* e nodal (métodos de comparação de topologia de árvores filogenética) e a quantidade de clados congruentes entre duas árvores inferidas por MV e BE, ambas utilizando o mesmo marcador e, quando incluído, o mesmo grupo externo.

Tabela 35 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. umbratilis* cada das quais inferida por um método diferente (BE e MV) e utilizando mesmo marcador (ou mesma concatenação de marcadores) e grupo externo (quando tinham).

Marcador/método de inferência	Árvore por MV x árvore por BE				
	<i>Split</i> (escala: 0 a 1) Quanto mais próximo de 0, maior é a congruência entre as árvores	Nodal (escala: 1 a +∞) Quanto mais próximo de 1, maior é a congruência entre as árvores	Árvore por MV	Árvore por BE	Coincidentes
COI					91
(Grupo externo: N+L+P)	0,8056	6,8104	148	149	(MV: 61,49% e BE:61,07%)
<i>Per</i>					97
(Grupo externo: N+L+P)	0,7222	9,0257	149	149	(MV e BE: 65,10%)
Concatenados sem grupo externo	0,5797	5,3022	143	143	103 (MV e BE: 72,03%)
Concatenados com grupo externo (N+L+P)	0,5278	4,6921	149	149	111 (MV e BE: 74,50%)

Na Tabela 36 estão as pontuações *Split* e nodal e a quantidade de clados congruentes entre duas árvores inferidas utilizando COI e *per*, ambas inferidas pelo mesmo método (ambas por MV ou ambas por BE):

Tabela 36 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. umbratilis* cada das quais inferida utilizando marcadores diferentes (COI e *per*) e ambas inferidas pelo mesmo método (ambas inferidas por MV ou ambas por BE).

Método de inferência/ marcador	Árvore COI (Grupo externo: N+L+P) x árvore <i>per</i> (Grupo externo: N+L+P)				
	<i>Split</i>	Nodal	Quantidade aproximada de clados:		
			Árvore COI	Árvore <i>per</i>	Coincidentes
<b>MV</b>	0,9306	10,5833	149	149	82 (COI e <i>per</i> 55,03%)
<b>BE</b>	0,9444	4,5874	149	149	81 (COI e <i>per</i> 54,36%)

E, por fim, na Tabela 37 estão as pontuações *Split* e nodal e a quantidade de clados congruentes entre duas árvores inferidas por MV utilizando marcador COI e por BE utilizando marcador *per*, e também entre duas árvores inferidas por MV utilizando marcador *per* e por BE utilizando marcador COI:

Tabela 37 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. umbratilis* cada das quais inferidas por MV utilizando marcador COI e por BE utilizando marcador *per*, e também entre duas árvores inferidas por MV utilizando marcador *per* e por BE utilizando marcador COI.

Marcador (cada com método de inferência diferente)	Quantidade aproximada de clados:				
	<i>Split</i>	Nodal	Árvore por	Árvore por	Coincidentes
			MV	BE	
COI (por MV) x					82
<i>per</i> (por BE)	0,9306	7,6282	148	149	(COI+MV 55,41% e <i>per</i> +BE 55,03%)
(ambos os grupo externo: N+L+P)					
<i>Per</i> ( por MV) x					81
COI (por BE)	0,9445	11,0345	149	149	( <i>per</i> +MV e COI+BE 54,36%)
(Grupo externo: N+L+P).					

Apenas lembrando que o grupo externo para COI correspondem ao grupo externo para *per* em nível de subgênero – *Phlebotomus* (*P. duboscqi* e *P. sergenti*, respectivamente), *Lutzomyia* (*L. gomezi* e *L. renei*, respectivamente) e *Nyssomyia* (*L. anduzei* e *L. whitmani*, respectivamente). Alguns pontos importantes podem-se citar em relação às comparações de topologias realizadas:

- ❖ **Comparação de topologia de árvore inferida por MV com a de árvore inferida por BE (ver Tabela 35):** comparações entre as árvores inferidas utilizando o mesmo marcador - ambas as árvores por COI, ambas por *per* ou ambas por COI e *per* concatenados com grupo externo e sem grupo externo -

deram baixo grau de incongruência. Todas tiveram em média 146 clados. Os valores de pontuação *split* de cada uma das quatro comparações variaram entre 0,50 e 0,85 e de pontuação nodal, de 4,5 e 9,5. A porcentagem de clados coincidentes entre cada duas árvores foi entre 60% e 75%, respectivamente.

❖ **Comparação de topologia de árvore inferida utilizando marcador COI com a de árvore inferida utilizando marcador *period* (Tabela 36):** comparações entre as árvores inferidas, porém cada utilizando marcador diferente - uma por COI e a outra por *per* - deram grau de incongruência considerável (mais alto do que ao apresentado na comparação anterior). Todas tiveram em média cento e quarenta e nove (149) clados. Os valores de pontuação *split* de cada uma das duas (2) comparações foram entre 0,93 e 0,95 e de pontuação nodal (quanto mais próximo de 1 mais congruentes são duas árvores) em cada comparação foi entre 4,5 e 11,0. A porcentagem de clados coincidentes entre cada duas árvores foi entre 54% e 56%.

❖ **Comparação de topologia de árvore inferida utilizando marcador COI (por MV) com a de árvore inferida utilizando marcador *per* (por BE) (ver Tabela 37):** ambas árvores tiveram grau de incongruência considerável (muito semelhantes à segunda comparação de topologia). As duas árvores tiveram em média cento e quarenta e oito (148) clados. O valor de pontuação *split* dessa comparação foi de 0,9306 e de pontuação nodal (quanto mais próximo de 1 mais congruentes são duas árvores) foi de 7,6282. A porcentagem de clados coincidentes entre estas duas árvores foi em torno de 55%.

❖ **Comparação de topologia de árvore inferida utilizando marcador COI (por BE) com a de árvore inferida utilizando marcador *per* (por MV) (ver também Tabela 37):** ambas as árvores tiveram grau de incongruência considerável (muito semelhantes à segunda comparação). As duas árvores tiveram em média cento e quarenta e nove (149) clados. O valor de pontuação *split* dessa comparação foi de 0,9445 e de pontuação nodal (quanto mais próximo de 1 mais congruentes são duas árvores) foi de 11,0345. A porcentagem de clados coincidentes entre estas duas árvores foi em torno de 54%.

**b) *L. longipalpis***

Na Tabela 38 estão valores do teste de congruência de topologia. As pontuações *Split* e nodal (métodos de comparação de topologia de árvores filogenética) e a quantidade de clados congruentes entre duas árvores inferidas por MV e BE, ambas utilizando o mesmo marcador e, quando tinham, o mesmo grupo externo:

Tabela 38 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. longipalpis* cada das quais inferida por um método diferente (BE e MV) e utilizando mesmo marcador (ou mesma concatenação de marcadores) e grupo externo (quando tinham).

Método de inferência /Marcador	Árvore por MV x árvore por BE				
	<i>Split</i> (escala: 0 a 1)	Nodal (escala: 1 a +∞)	Quantidade aproximada de clados:		
	Quanto mais próximo de 0, mais congruência entre as árvores	Quanto mais próximo de 1, mais congruência entre as árvores	Árvore por MV	Árvore por BE	Coincidentes
COI (Grupo externo: P)	0,7183	3,2273	147	147	96 (MV e BE 65,31%)
<i>Per</i> (grupo externo: P)	0,6697	3,0131	147	147	101 (MV e BE 68,71%)
Concatenados sem grupo externo	0,6714	4,5509	145	145	98 (MV e BE 67,59%)
Concatenados com grupo externo (grupo externo: P)	0,7324	4,2882	147	147	95 (MV e BE 64,63%)

Na Tabela 39 estão as pontuações *Split* e nodal e a quantidade de clados congruentes entre duas árvores inferidas utilizando COI e *per*, ambas inferidas pelo mesmo método (ambas por MV ou ambas por BE):

Tabela 39 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. longipalpis* cada das quais inferida utilizando marcadores diferentes (COI e *per*) e ambas inferidas pelo mesmo método (ambas inferidas por MV ou ambas por BE).

Método de inferência/ marcador	Árvore COI*(Grupo externo: P) x árvore <i>per</i> (Grupo externo: P)				
	<i>Split</i>	Nodal	Quantidade aproximada de clados:		
			Árvore COI	Árvore <i>per</i>	Coincidentes
					76
MV	1	8,73386	147	147	(MV e BE 51,70%)
					76
BE	1	8,2819	147	147	(MV e BE 51,70%)

\*Mesmos não obtendo árvores aceitáveis segundos aspectos biológicos aceitáveis, elas foram submetidas a esses testes, propósito de curiosidade.

E por fim na Tabela 40 estão as pontuações *Split* e nodal e a quantidade de clados congruentes entre duas árvores inferidas por MV utilizando marcador COI e por BE utilizando marcador *per*, e também entre duas árvores inferidas por MV utilizando marcador *per* e por BE utilizando marcador COI:

Tabela 40 - Valores *Split*, nodal e de clados congruentes entre duas árvores de *L. longipalpis* cada das quais inferidas por MV utilizando marcador COI e por BE utilizando marcador *per*, e também entre duas árvores inferidas por MV utilizando marcador *per* e por BE utilizando marcador COI.

Marcador  (cada com método de inferência diferente)	<i>Split</i>	Nodal	Quantidade aproximada de clados:		
			Árvore por MV	Árvore por BE	Coincidentes
COI (por MV) x <i>per</i> (por BE) (Ambos, grupo externo: P).	1	8,9036	147	147	76 (MV e BE 51,70%)
<i>Per</i> (por MV) x COI (por BE) (Ambos, grupo externo: P).	1	8,0742	147	147	76 (MV e BE 51,70%)

Mesmo essas sequências concatenadas apresentando valor  $p < 0,05$  no teste de homogeneidade de partição, a partir delas foram inferidas árvores filogenéticas.

Apenas lembrando: o grupo externo para COI correspondem aos grupos externos para *per* em nível de subgênero - *Phlebotomus* (*P. sergentie*, *P. duboscqi*, respectivamente). Analisando esses valores de comparações de topologias, perceberam-se algumas características importantes:

- ❖ **Comparação de topologia de árvore inferida por MV com a de árvore inferida por BE (ver Tabela 38):** comparações entre ambas as árvores inferidas cada uma por cada um dos métodos, porém utilizando o mesmo

marcador - ambas árvores por COI, ambas por *per* ou ambas por COI e *per* concatenados com e sem grupo externo - deram baixo grau de incongruência. Todas tiveram em média cento e quarenta e seis (146) clados. Os valores de pontuação *split* de cada uma das quatro (4) comparações foram entre 0,66 e 0,74 e de pontuação nodal (quanto mais próximo de 1 mais congruentes são duas árvores) em cada comparação foi entre 3,0 e 4,6. A porcentagem de clados coincidentes entre cada duas árvores foi entre 64% e 69%.

❖ **Comparação de topologia de árvore inferida utilizando marcador COI com a de árvore inferida utilizando marcador *per* (ver Tabela 39):** comparações entre ambas as árvores inferidas cada uma pelo mesmo método - ambas por MV ou ambas por BE - , porém cada utilizando marcador diferente - uma por COI e outra por *per* - deram grau de incongruência considerável (mais alto do que ao apresentado na comparação anterior). Todas tiveram em média cento e quarenta e nove (147) clados. Os valores de pontuação *split* de cada uma das duas (2) comparações foram 1,0 e de pontuação nodal (quanto mais próximo de 1 mais congruentes, são duas árvores) em cada comparação foi entre 8,0 e 9,0. A porcentagem de clados coincidentes entre cada duas árvores foi em média 51,7%.

❖ **Comparação de topologia de árvore inferida utilizando marcador COI (por MV) com a de árvore inferida utilizando marcador *per* (por BE) (ver Tabela 40):** ambas as árvores tiveram grau de incongruência considerável (muito semelhantes à segunda comparação). As duas árvores tiveram em média

cento e quarenta e sete (147) clados. O valor de pontuação *split* dessa comparação foi de 1,0 e de pontuação nodal (quanto mais próximo de 1 mais congruentes são duas árvores) foi de 8,9036. A porcentagem de clados coincidentes entre estas duas árvores foi em média 51,7%.

❖ **Comparação de topologia de árvore inferida utilizando marcador COI (por BE) com a de árvore inferida utilizando marcador per (por MV) (ver também Tabela 40):** ambas as árvores tiveram grau de incongruência considerável (muito semelhantes à segunda comparação). As duas árvores tiveram em média cento e quarenta e sete (147) clados. O valor de pontuação *split* dessa comparação foi de 1,0 e de pontuação nodal (quanto mais próximo de 1 mais congruentes são duas árvores) foi de 8,0742. A porcentagem de clados coincidentes entre estas duas árvores foi em média 51,7%.

## 6. DISCUSSÃO

Em todas as três buscas de grupo externo para *L. umbratilis*, a combinação da utilização do marcador mitocondrial COI com a inferência filogenética BE resultou nas melhores árvores filogenéticas, além de maior quantidade de árvores com suporte alto de ambos clados esperados, apresentando simultaneamente polarização esperada e suportes altos, ou seja, maiores do que 80%. Os genes que codificam DNA mitocondrial ( $_{mt}$ DNA) foram amplamente utilizado em estudos populacionais, taxonomia molecular e relações filogenéticas de muitos organismos (Avice, 1994). A alta velocidade de mudanças no processo de evolução e a herança materna no geral deste marcador tornam-o interessante para estudos evolutivos de grupos que divergiram recentemente. Nesses grupos apenas os sítios de nucleotídeos de evolução mais rápida acumulam substituições (Scarpassa & Alencar, 2012). O marcador COI tem sido utilizado frequentemente em estrutura filogeográfica e em estudos de genética de população de insectos (Simon *et al.*, 1994; Pramual *et al.*, 2005; Scarpassa *et al.*, 2008).

Dentre as três buscas de grupo externo para *L. longipalpis*, a combinação da utilização do marcador *per* com a inferência filogenética BE resultou na melhor árvore filogenética apresentando polarização esperada. Embora o suporte não tenha sido alto, ou seja, maiores do que 80%. Pouco se sabe sobre a genética molecular deste grupo de flebotomíneos (Esseghir *et al.*, 1997; Pronto *et al.*, 1997, 1998; Ishikawa *et al.*, 1999). No entanto, estudos de filogenética molecular de outros grupos de artrópodes sugerem que o gene COI pode ser informativo para esclarecer a taxonomia de *Lutzomyia* no nível de espécie (Navajas *et al.*, 1998; Roderick e Gillespie, 1998). Apesar disso, poucos estudos de flebotomíneos utilizam o gene COI (Arrivillaga *et al.*, 2002). Inclusive em vários estudos de genética molecular de

populações, variações dos genes mitocondriais (Hodgkinson *et al.*, 2003) não permitiram concluir que se *L. longipalpis* se trata de espécies crípticas enquanto que as variações nos genes cromossômicos (nucleares), como o *period gene* (Bauzer *et al.*, 2002-a), (Bauzer *et al.*, 2002-b) foram indicativas de espécies crípticas. Talvez isso se deve a ausência de divergência. Houve divergência entre as sequências, porém não foi suficiente para indicar espécies crípticas (Hodgkinson *et al.*, 2003) e muito menos pra discriminar estas espécies crípticas.

Já foi visto em estudos que apesar de ambos os métodos de inferência (MV e BE) serem probabilísticos, não todas e sim a maioria das árvores consenso de cada um desses métodos possuem topologias idênticas (Leaché, 2010). Também foi visto mais forte correlação entre valores de suportes da inferência MV e BE em comparação com correlação de suportes por métodos MV e parcimônia (Leaché, 2010). Mas isso não significa que não haja diferenças entre os suportes de clados obtidos pelo método MV e BE. Leaché (2010) diz que entre tantas causas dessas diferenças estão: erro estocástico, problemas de inicialização do método MV, ou combinações destas causas.

Observou-se também a utilização de sequências concatenadas dos marcadores COI e *per*. A partir das sequências concatenadas de *L. umbratilis* foram inferidas árvores com ambos os clados esperados com suportes ainda mais elevados do que se utilizando COI isoladamente, inferindo tanto pelo método MV como BE. E o uso isolado do marcador COI sempre supera o marcador *per* tanto em quantidade de árvores com polaridades esperadas como no valor de suporte dos clados esperados, e principalmente quando combinado com o método BE.

As árvores inferidas a partir das sequências concatenadas de *L. longipalpis* apresentaram ambos os clados esperados com suportes mais elevados do que se

utilizando *per* isoladamente, inferido tanto pelo método MV (embora com suporte abaixo de 80% tanto em árvore com como sem grupo externo) como BE (tanto com grupo externo, com suportes de ambos clados esperados abaixo de 80%; como sem grupo externo, esta com suportes maiores do que 80%). E o uso isolado do marcador *per* supera o marcador COI na possibilidade de polarizar a árvore de forma esperada quando combinado com o método de inferência BE.

Concatenar dados é uma boa alternativa para alcançar árvore filogenética de espécie e elevar o valor de suporte dos clados (Leaché, 2010), desde que os dados exibem forte congruência pois caso contrário pode ocorrer em uma sub-estimativa da árvore filogenética de espécies (Wiens, 1998).

Neste trabalho, segundo a análise via teste de homogeneidade de partição (pelo *software* PAUP) os dados de sequências dos marcadores nuclear *per* e do marcador mitocondrial COI concatenados mostraram-se congruentes entre si ( $p=0,9900>0,05$ ) para *L. umbratilis*, tanto a análise do teste de partição considerando a presença das sequências grupo externo como a análise com a ausência das sequências grupo externo).

Já para *L. longipalpis*, os dados de sequências do marcador nuclear *per* e do marcador mitocondrial COI concatenados mostraram-se congruentes entre si na análise do teste de partição considerando a ausência das sequências grupo externo, embora com p-value muito baixo ( $p=0,08$ ). Porém considerando a presença do grupo externo, os dados concatenados não se mostraram congruentes ( $p=0,04<0,05$ ). Porém sabe-se que combinar dados que exibem forte incongruência é questionável, e pode resultar em árvore espécie subestimada (Wiens, 1998).

Quanto à comparação de topologias, para *L. umbratilis* foi observado que se comparando duas árvores com os mesmos táxons, cada uma inferida por métodos

diferentes (uma por MV e outra por BE), porém utilizando o mesmo marcador (COI ou *per*) foi observada baixa incongruência. E para *L. longipalpis* foram observados resultados semelhantes.

Em um estudo concluiu-se que apesar de ambos os métodos de inferência (MV e BE) serem probabilísticos a maioria das árvores consenso de cada uma dos métodos possuem topologias idênticas (Leaché, 2010). Ou seja, não necessariamente todas as árvores possuem topologias idênticas. E ainda menor quando se utiliza os dados de sequências concatenadas dos dois marcadores. Comparando-se a topologia de duas árvores cada uma a partir de marcadores diferente (uma inferida por dados do marcador COI e outra pelos de *per*) foi observada maior incongruência tanto quando ambas foram inferidas pelo mesmo método de inferência (ambas por MV ou ambas por BE) ou por métodos diferentes (uma por MV e outra por BE). Isso se deve a histórias evolutivas de genes nucleares e de genes mitocondriais frequentemente serem distintas entre si (Leaché, 2010). Apesar de comparação de topologia ser das alternativas de avaliar o uso de diferentes métodos de inferência ou diferentes tipos de dados para inferir árvores filgenéticas de uma determinado grupo de táxons, neste trabalho essa comparação não teve grande contribuição quanto a conclusão geral deste trabalho pois como se trata de grupos de populações de uma espécie que corresponde a uma complexo de espécies crípticas, nem todos os clados serão relevantes, a não ser aqueles que indicam tal hipótese. Ou seja, a homogeneidade da árvore não interessa tanto e sim o poder de discriminar as diferentes populações como diferentes espécies. Talvez se cada um dos táxons de cada árvore correspondesse a uma espécie diferente, a comparação de topologia teria uma contribuição significativa quanto a homogeneidade dos clados, já que todos os clados formados teriam importância

praticamente similares. Os métodos de comparação de topologia utilizados neste trabalho apesar de serem indicados para comparar diferentes árvores que distinguem entre si quanto o método de inferência (Puigbò *et al.*, 2007), eles compara a topologia da árvore por inteiro analisam todos os clados de igual modo, além de não considerar os valores de suporte dos clados (Puigbò *et al.*, 2007) muito menos de um clado específico.

A aplicação de métodos baseados em modelos como os de máxima verossimilhança (ML) gasta um período de tempo muito extenso, por fazer a busca da árvore através do espaço de árvores, sendo difícil de aplicar-se tal método quando temos muitos táxons e/ou avaliação de confiabilidade através de métodos padrão *bootstrap* não paramétrico (Felsenstein, 1985). Apesar de testes de razão de verossimilhança ter uma base estatística sólida, tipicamente assume que a filogenia é conhecida sem erro, ou seja, a incerteza é ignorada, assim como os métodos de distância e de parcimônia. Os métodos Bayesianos prometem tratabilidade computacional com grandes conjuntos de dados e modelos evolutivos complexos (Huelsenbeck *et al.* 2001) e as topologias e comprimentos dos ramos não são tratadas como parâmetros, como em métodos ML (Felsenstein, 1981), mas como variáveis aleatórias em uma distribuição. Eles podem fazer uso dos mesmos modelos de substituição de DNA usados para análises de verossimilhança.

Em questões práticas o método Bayesiano tem uma vantagem sobre a máxima verossimilhança permite facilidade de interpretação dos resultados já que os valores de suporte de clado são dados em forma de probabilidade e a facilidade de incorporação de informação *a priori* do modelo. Além disso, a probabilidade *a posteriori* de uma árvore é simplesmente a probabilidade de que a árvore é correta, ou seja, também representa tendo em conta os dados e modelo (Yang & Rannala,

2012). Porém os conceitos desse intervalo de confiança tem uma interpretação imaginária que escapa de muitos que utilizam a análise Bayesiana e se as informações *a priori* forem mal especificadas podem ter influências inesperadas sobre a probabilidade *a posteriori*. Em compensação o teste de confiança do *bootstrap* utilizado no método de máxima verossimilhança é mais fácil interpretar. A inferência BE necessita de 80% menos tempo computacional em comparação a inferência MV (Leaché, 2010).

Apesar de ser tão lógica, a utilização de espécies reconhecidas como ancestrais anteriores ao *L. umbratilis* como grupo externo, muitas vezes não se consegue com sucesso com todas as opções possíveis talvez devido o tipo de homologies entre eles para o gene considerado, devido também a questão de filogenia da espécie nem sempre é possível ser representada com a filogenia por um único marcador molecular, entre outras razões.

## 7. CONCLUSÕES

Para o complexo *L. umbratilis*, o método de inferência filogenética BE apresentou árvores filogenéticas com maiores probabilidades, ou seja, com melhores resultados em relação às árvores inferidas por MV, principalmente com a utilização do marcador mitocondrial COI. Já para o complexo de espécies *L. longipalpis*, o método BE apresentou a melhor árvore filogenética em relação ao método MV e apenas utilizando o marcador nuclear *period*. Tendo alcançado melhores resultados com o marcador mitocondrial COI, para *L. umbratilis*, e *period*, para *L. longipalpis*, e com a abordagem BE, conclui-se que esta abordagem juntamente com os respectivos marcadores teve maior adequação do que a abordagem MV.

Uma alternativa de sistema padronizado da metodologia investigação de complexos de espécies via filogenética molecular para o complexo *L. umbratilis* é a combinação da utilização do marcador mitocondrial COI com a inferência Bayesiana. Enquanto que para o complexo *L. longipalpis* uma alternativa seria a combinação da utilização do marcador nuclear *per* com a inferência Bayesiana.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

- Aleixo, A. Conceitos de espécie e o eterno conflito entre continuidade e operacionalidade: uma proposta de normatização de critérios para o reconhecimento de espécies pelo Comitê Brasileiro de Registros Ornitológicos. *Revista Brasileira de Ornitologia* 15 (2) 297-310, 2007.
- Amorim, D. S. Fundamentos de Sistemática Filogenética. Editora Holos, ano 2002.
- Arias, J. R.; Freitas, R. A. Sobre os vectores da leishmaniose cutânea na Amazônia central do Brasil. 2: Incidência de flagelados em flebotomíneos selváticos. *Acta Amazônica* 8: 387–396, 1978.
- Arrivillaga, J. C.; Norris, D.E.; Feliciangeli, M. D.; Lanzaro, G. C. Phylogeography of the neotropical sand fly *Lutzomyia longipalpis* inferred from mitochondrial DNA sequences. *Inf Genet Evol* 2: 83–95, 2002.
- Avise, J. C. *Molecular Markers, Natural History and Evolution* Chapman & Hall, New York, ano 1994.
- Azevedo, A. C.; Monteiro, F. A.; Cabello, P. H.; Souza, N. A.; Rosa-Freitas, M. G.; Rangel, E. F. Studies on populations of *Lutzomyia longipalpis* (Lutz & Neiva, 1912) (Diptera: Psychodidae: Phlebotominae) in Brazil. *Mem Inst Oswaldo Cruz* 95: 305-322, ano 2000.
- Azpuruá, J., De La Cruz, D., Valderama, A., Windsor, D. *Lutzomyia* sand fly diversity and rates of infection by *Wolbachia* and an exotic *Leishmania* species on Barro Colorado Island, Panama. *PLoS Negl. Trop. Dis.* 4:e627, 2010.
- Balbino, V. Q.; Marcondes, C. B.; Alexander, B.; Luna, L. K. S.; Lucena, M. M. M.; Mendes, A. C. S.; Andrade, P. P. First report of *Lutzomyia* (*Nyssomyia*) *umbratilis* Ward & Frahia, 1977 outside of Amazonian region, in Recife, State of Pernambuco, Brazil (Diptera: *Psychodidae*: *Phlebotominae*). *Mem Inst Oswaldo Cruz* 96: 315–317, 2001. <http://dx.doi.org/10.1590/S0074-02762001000300005>
- Barreto, M. P. Novos subgêneros de *Lutzomyia* França, 1924 (Diptera. *Psychodidae*. Subfamilia *Phlebotominae*). *Be11. It18t. Meti. Trop.*, S. Paulo, 4: 91-100, 1962.
- Bauzer, L. G.; Souza, N. A.; Ward, R. D.; Kyriacou, C. P.; Peixoto, A. A. The period gene and genetic differentiation between three Brazilian populations of *Lutzomyia longipalpis*. *In-sect MolBiol* 11: 315-323, 2002a.
- Bauzer, L. G.; Gesto, J. S.; Souza, N. A.; Ward, R. D.; Hamilton, J. G.; Kyriacou, C. P.; Peixoto, A. A. Molecular divergence in the period gene between two putative

- sympatric species of the *Lutzomyia longipalpis* complex. *Mol Biol Evol* 19:1624-1627, 2002b.
- Bauzer, L. G. S. R.; Souza, N. A.; Maingon, R. D. C.; Peixoto, A. A. *Lutzomyia longipalpis* in Brazil: a complex or a single species? A mini-review. *Mem Inst Oswaldo Cruz, Rio de Janeiro*, Vol. 102(1): 1-12, 2007.
- Beckner, M. *The biological way of thought*. Columbia University Press, New York, ano 1959.
- Bickford, D.; Lohman, D. J.; Sodhi, N. S.; Ng, P. K. L.; Meier, R.; Winker, K.; Ingram, K. K.; Das, I. Cryptic species as a window on diversity and conservation. *TRENDS in Ecology and Evolution* Vol.22 No.3, p. 148-55, 2006.
- Bickford, D.; Lohman, D. J.; Sodhi, N. S.; et al. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22, 148–155, 2007.
- Bottecchia, M.; Oliveira, S. G.; Bauzer, L. G.; Souza, N. A.; Ward, R. D.; Garner, K. J.; Kyriacou C. P.; Peixoto, A. A. Genetic divergence in the cacophony IVS6 intron among five Brazilian populations of *Lutzomyia longipalpis*. *J MolEvol* 58: 754-761, 2004.
- Bouckaert, R. R. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* 15; 26(10):1372-3, 2010. doi: 10.1093/bioinformatics/btq110.
- Bryant, D.; Galtier, N.; Poursat, M. A. (2005) Likelihood calculation in molecular phylogenetics. In *Mathematics of evolution & phylogeny* (ed. Olivier Gascuel). Oxford University Press, New York. Ano 2005. Cap.2 (p. 33 – 62).
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J.; Higgins, D. G.; Thompson, J. D. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, 31, 3497–3500, 2003.
- Contreras Gutiérrez, M. A.; Vivero, R. J.; Vélez, I. D.; Porter, C. H.; Uribe, S. (2014) DNA Barcoding for the Identification of Sand Fly Species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS ONE* 9(1): e85496, 2014. doi:10.1371/journal.pone.0085496
- Coyne, J. A. *Why Evolution Is True?* Oxford University Press. Oxford, New York. Ano 2009.
- Cracraft, J. Species concepts and speciation analysis. *Current Ornithol.* 1: 159-187, 1983.

- Cracraft, J.; Barker, F. K.; Braun, M.; Harshman, J.; Dyke, G. J.; Feinstein, J.; Stanley, S.; Cibois, A.; Schikler, P.; Beresford, P.; Garcia-Moreno, J.; Sorenson, M. D.; Yuri, T.; Mindell, D. (2004) Phylogenetic relationships among modern birds (Neornithes). Toward an avian tree of life, Ed: Cracraft J and Donoghue MJ (eds.) Assembling the tree of life. Oxford. Ano 2004, p. 468-489.
- Curler, G. R and Moulton, J. K. Phylogeny of psychodid subfamilies (Diptera: *Psychodidae*) inferred from nuclear DNA sequences with a review of morphological evidence for relationships. *Systematic Entomology* 37, 603–616, 2012.
- Depaquit, J.; Ferté, H.; Léger, N.; Killick-Kendrick, R.; Rioux, J. A.; et al. Molecular systematic of the Phlebotomine sandflies of the subgenus *Paraphlebotomus* (Diptera, *Psychodidae*, *Phlebotomus*) based on ITS2 rDNA sequences. Hypotheses of dispersion and speciation. *Insect Mol Biol* 9: 293–300, 2000.
- de Queiroz Balbino, V.; Coutinho-Abreu, I. V.; Sonoda, I. V.; Melo, M. A.; de Andrade P. P.; de Castro, J. A.; Rebelo, J. M.; Carvalho, S. M.; Ramalho-Ortigao, M. Genetic structure of natural populations of the sand fly *Lutzomyia longipalpis* (Diptera: *Psychodidae*) from the Brazilian northeastern region. *Acta Trop* 98: 15-24, 2006.
- de Queiroz, K. The general lineage concept of species, species criteria, and the process of speciation: A conceptual unification and terminological recommendations in *Endless forms: Species and speciation* (Howard DJ and Berlocher SH eds.). Oxford University Press, New York. Ano 1998.
- de Queiroz, K. Species Concepts and Species Delimitation. *Syst. Biol.* 56(6):879–886, 2007. DOI: 10.1080/10635150701701083
- Dobzhansky, T. (1970) *Genetics of the evolutionary process*. Columbia university press, New York. Ano 1970.
- Doyen, J. T. and Slobodchikoff, C. N. An operational approach to species classification. *Systematic Zoology* 23: 239–247, 1974.
- Esseghir, S.; Ready, P. D.; Killick-Kendrick, R.; Ben Ismail, R. Mitochondrial haplotypes and phylogeography of *Phlebotomus* vectors of *Leishmania major* Insect. *Insect Mol. Biol.* 6, 211–225, 1997.

- Evans, M. J.; Roshental, J. S. Probability and Statistics: The Science of Uncertainty. University of Toronto. 2 ed. W. H. Freeman and Company. New York. Ano 2010.
- Farris, J. D.; Kallersj, M.; Kluge, A. G. and Bult, C. Testing significance of incongruence. *Cladistics* 10:315–319, 1994.
- Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, 27, 401-410, 1978.
- Felsenstein, J. Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *J Mol Evol* 17:368-376, 1981.
- Felsenstein, J. Inferring phylogenies. Sinauer Associates Incorporated, Sunderland, Massachusetts. ISBN 0-87893-177-5. Ano 2004.
- Fitch, W. M.; Margoliash, E. Construction of phylogenetic trees. *Science*, 155, 279-284, 1967.
- França, C.; Parrot, L. Introduction à l'étude systématique des Diptères du genre *Phlebotomus*. *Bul Soc Path Exot* 12: 695-708, 1920.
- França, C.; Parrot, L. Essai de classification des Phlébotomes. *Archs Insts Pasteur Afr N* 1: 279-284, 1921.
- Galati, E. A. B. Morfologia e taxonomia: classificação de Phlebotominae. In: Rangel, E.F. & Lainson, R. (Orgs.). *Flebotomíneos do Brasil*. Editora Fiocruz, Rio de Janeiro. Ano 2003, p. 23-51.
- Galati, E. A. B. Phlebotominae (Diptera, *Psychodidae*) Classificação, Morfologia, Terminologia e Identificação de Adultos [Phlebotominae (Diptera, Psychodidae) classification, morphology, terminology and adult identification]. Apostila Disciplina HEP 5752 Bioecologia e Identificação de Phlebotominae Vol. I. São Paulo: Brasil. Press, Departamento de Epidemiologia Faculdade de Saúde Pública Universidade de São Paulo. Ano 2010. 126 p.
- Galati, E. A. B. Phylogenetic systematics of Phlebotominae (Diptera: *Psychodidae*) with emphasis on American groups. *Boletín de la Dirección de Malariología y Ambiental*, 35(Supl. 1):133-142, 1995.
- Guindon, S.; Dufayard, J. F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3):307-21, 2010.

- Hacking, I. An Introduction to Probability and Inductive Logic. Cambridge University Press. Ano 2011.
- Hall, B. G. Phylogenetic trees Made Easy. Sinauer Assoc., Sunderland, MA, pp. 87-89. Ano 2001.
- Hamilton, J. G.; Brazil, R. P.; Maingon, R. A fourth chemotype of *Lutzomyia longipalpis* (Diptera: Psychodidae) from Jaibas, Minas Gerais State, Brazil. J Med Entomol 41: 1021-1026, 2004.
- Hamilton, J. G. C.; Maingon, R.; Alexander, B.; Ward, R.; Brazil, R. P. Analysis of the sex pheromone extract of individual male *Lutzomyia longipalpis* sandflies from six regions in Brazil. Med Vet Entomol 19: 480-8, 2005.
- Henning, W. Phylogenetic systematics. Urbana usw. (Univ. Illinois Press). Ano 1966.
- Henning, W. (1968). Elementos de uma sistemática filogenética, Eudeba, Buenos Aires. Ano 1968.
- Hertig, M. A new genus of blood sucking Psychodids from Peru (Diptera: *Psychodidae*). Ann Entomol Soc Am 41: 8-16, 1948
- Hey, J. Genes Categories & Species. Oxford University Press. Ano 2001.
- Hillis, D. M. and Moritz, C. (1996) Molecular systematics. 2nd ed. Sinauer Associates Incorporated. ISBN 0-87893-282-8. Ano 1996.
- Hipp, A. L., Hall, J. C.; Sytsma, K. J. Congruence *Versus* Phylogenetic Accuracy: Revisiting the Incongruence Length Difference Test. Syst. Biol. 53(1):81–89, 2004.
- Hodgkinson, V. H.; Birungi, J.; Quintana, M.; Dietze, R.; Munstermann, L. E. Mitochondrial cytochrome b variation in populations of the visceral leishmaniasis vector *Lutzomyia longipalpis* across eastern Brazil. Am J Trop Med Hyg 69:386-392, 2003.
- Huelsenbeck, J. P.; Ronquist, F.; Nielsen, R.; Bollback, J. P. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314, 2001. DOI: 10.1126/science.1065889.
- Ishikawa, E. A. Y.; Ready, P. D.; Souza, A. A.; Day, J. J.; Rangel, E. F.; Davies, C. R.; Shaw, I. J. A mitochondrial DNA phylogeny indicates close relationships between populations of *Lutzomyia whitmani* (Diptera: *Psychodidae*, *Phlebotominae*) from the rain forest regions of amazonian and northeast Brazil. Memorias Instituto Oswaldo Cruz Rio de Janeiro 94, 339–345, 1999.

- Justiniano, S. C. B., Chagas, A. C.; Pessoa, F. A. C.; Queiroz, R. G. Comparative biology of two populations of *Lutzomyia umbratilis* (Diptera: *Psychodidae*) of central Amazonia, Brazil, under laboratory conditions. *Braz J Biol* 64: 227–235, 2004.
- Knapp, S. What's in a name? A history of taxonomy. Ano 2010. Ver em website <http://www.nhm.ac.uk/nature-online/science-of-natural-history/taxonomy-systematics/history-taxonomy/index.html> (Ver também: What is taxonomy? Natural History Museum London. Em website <http://www.nhm.ac.uk/nature-online/science-of-natural-history/taxonomy-systematics/what-is-taxonomy/index.html>)
- Knowlton, N. Cryptic and Sibling Species among the Decapod Crustacea. *Journal of Crustacean Biology* Vol. 6, No. 3, pp. 356-363, 1986. <http://www.jstor.org/discover/10.2307/1548175?uid=2&uid=4&sid=21102566662027>
- Leaché, A. D. Species trees for spiny lizards (*Genus Sceloporus*): Identifying points of concordance and conflict between nuclear and mitochondrial data. *Molecular Phylogenetics and Evolution* 54: 162–171, 2010.
- Lehr, M. A.; Kilpatrick, C. W.; Wilkerson, R. C.; Conn, J. E. Cryptic Species in the *Anopheles (Nyssorhynchus) albitarsis* (Diptera: *Culicidae*) Complex: Incongruence between Random Amplified Polymorphic DNA-Polymerase Chain Reaction Identification and Analysis of Mitochondrial DNA COI Gene Sequences. *Ann Entomol Soc Am*; 98(6): 908–917, 2005.
- Leigh, J. W.; Susko, E.; Baumgartner, M and Roger, A. J. Testing congruence in phylogenomic analysis. *Syst. Biol.* 57, 104–115, 2008.
- Lewis, D. J.; Young, D. G.; Fairchild, G. B and Minter, D. M. Proposals for a stable classification of phlebotomine sandflies. *Systematic Entomology*, 2:319-332, 1977.
- Lewis, P. O and Swofford, D. L. Back to the future: Bayesian inference arrives in phylogenetics. *Trends Ecol. Evol.* 16: 600-601, 2001
- Li, S.; Pearl, D. K.; Doss, H. Phylogenetic tree construction using Markov chain Monte Carlo. *Journal of the American Statistical Association*; Jun 2000; 95, 450; *Platinum Full Text Periodicals* pg. 493, 2000.

- Linnaeus, C. Systema naturae, sive regna tria naturae systematice proposita per classes, ordines, genera, & species. Haak, Leiden. Ano 1735.
- Lutz, A. and Neiva, A. Contribuição para o conhecimento das espécies do gênero *Phlebotomus* existentes no Brasil. Mem. Inst. O. Cruz, 4:84-95, 1912.
- Maingon, R. D.; Ward, R. D.; Hamilton, J. G.; Noyes, H. A.; Souza, N.; Kemp, S. J.; Watts, P. C. Genetic identification of two sibling species of *Lutzomyia longipalpis* (Diptera: *Psychodidae*) that produce distinct male sex pheromones in Sobral, Ceará State, Brazil. Mol Ecol 12: 1879-1894, 2003.
- Maingon, R. D.; Ward, R. D.; Hamilton, J. G.; Noyes, H. A.; Souza, N.; Kemp, S. J.; Watts, P. C. Genetic identification of two sibling species of *Lutzomyia longipalpis* (Diptera: *Psychodidae*) that produce distinct male sex pheromones in Sobral, Ceará State, Brazil. Mol Ecol. 12(7):1879-94, 2003.
- Mangabeira, O. Sobre a sistemática e biologia dos flebotomos do Ceará. Rev Bras Mal Doen Trop 21: 3-26, 1969.
- Marcondes, C. B. Carta ao editor: Would taxonomy be insignificant to the vector's ecology? Revista da Sociedade Brasileira de Medicina Tropical 31:491-493, 1998.
- Mau, B. and Newton, M. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. Journal of Computational and Graphical Statistics 6:122-131, 1997.
- Mayr, E. Systematics and the origin of species. New York: Columbia University Press. Ano 1942.
- Mayr, E. Biological classification: toward a synthesis of opposing methodologies, Science, 214, 510, 1981.
- Mazzoni, C. J.; Araki, A. S.; Ferreira, G. E. M.; Azevedo, R. V. D. M.; Barbujani, G.; et al. Multilocus analysis of introgression between two sand fly vectors of leishmaniasis. BMC Evolutionary Biology 8: 141, 2008.
- Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. The Journal of Chemical Physics 21, 1087, 1953. doi: 10.1063/1.1699114
- Moret, B. M. E.; Tang, J.; Warnow, T. Reconstructing phylogenies from gene-content and gene-order data. In Mathematics of evolution & phylogeny (ed. Olivier

- Gascuel). Oxford University Press, New York. Ano 2005. Cap. 12 (p. 321 – 352), p. 321 – 329.
- Mukhopadhyay, J.; Ghosh, K.; Azevedo, A. C.; Rangel, E. F.; Munstermann, L. E. Genetic polymorphism of morphological and biochemical characters in a Natal, Brazil, population of *Lutzomyia longipalpis* (Diptera: Psychodidae). J Am Mosq Control Assoc 14: 277-282, 1998a.
- Mukhopadhyay, J.; Ghosh, K.; Rangel, E. F.; Munstermann, L. E. Genetic variability in biochemical characters of Brazilian field populations of the Leishmania vector, *Lutzomyia longipalpis* (Diptera: Psychodidae). Am J Trop Med Hyg 59: 893-901, 1998b.
- Mukhopadhyay, J.; Rangel, E.; Ghosh, K.; Munstermann, L. Patterns of genetic variability in colonized strains of *Lutzomyia longipalpis* (Diptera: Psychodidae) and its consequences. Am J Trop Med Hyg 57: 216-221, 1997.
- Mutebi, J. P.; Alexander, B.; Sherlock, I.; Wellington, J.; Souza, A. A.; Shaw, J.; Rangel, E. F.; Lanzaro, G. C. Breeding structure of the sand fly *Lutzomyia longipalpis* (Lutz & Neiva) in Brazil. Am J Trop Med Hyg 61: 149-157, 1999.
- Navajas, M.; Gutierrez, M.; Lagnel, J.; Boursot, P. Mitochondrial cytochrome oxidase I in a Tetranychidae mite: a comparison between molecular phylogeny and change of morphological and life history traits. Bull. Entomol. Res. 86, 407–417, 1996.
- Nishihara, H.; Okada, N. and Hasegawa, M. Rooting the Eutherian tree — the power and pit falls of phylogenomics. Genome Biol. 8, R199, 2007.
- Padial, J. M.; Miralles, A.; Riva, I. D.; Vences, M. The integrative future of taxonomy. Frontiers in Zoology, 7:16. Ano 2010. (Ver em <http://www.frontiersinzoology.com/content/7/1/16>)
- Page, R. D. M. and Holmes, E. C. Molecular evolution: a phylogenetic approach. Blackwell Science, Oxford. ISBN 0-86542-889-1. Ano 1998.
- Passamonti, M.; Mantovani, B.; Scali, V. Phylogeny and karyotype evolution of the Iberian *Leptynia attenuata* species complex (Insecta Phasmatodea). Molecular Phylogenetics and Evolution Volume 30, Issue 1, Pages 87–96, 2004. [http://dx.doi.org/10.1016/S1055-7903\(03\)00156-8](http://dx.doi.org/10.1016/S1055-7903(03)00156-8)
- Petit, R. J. and Excoffier, L. Gene flow and species delimitation. Trends Ecol Evol 24: 386–393, 2009.

- Posada, D. jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25: 1253–1256, 2008.
- Pramual, P.; Kuvangkadilok, C.; Baimai, V.; Walton, C. Phylogeography of the black fly *Simulium tani* (Diptera: *Simuliidae*) from Thailand as inferred from mtDNA sequences. *Mol Ecol* 14: 3989–4001. 39, 2005.
- Puigbò, P.; Garcia-Vallvé, S. and McInerney, J. O. TOPD/FMTS: a new software to compare phylogenetic trees. *Bioinformatics Applications Note*. Vol. 23 no. 12, pages 1556 –1558, 2007. doi:10.1093/bioinformatics/btm135
- Rambaut, A. FigTree v1.4. Institute of Evolutionary Biology, University of Edinburgh, 2006-2012.
- Rannala, B.; Yang, Z. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol.*;43(3):304-11, 1996.
- Ready, P. D.; Day, J. C.; Souza, A. A.; Rangel, E. F.; Davies, C. R. Mitochondrial DNA characterization of populations of *Lutzomyia whitmani* (Diptera: Psychodidae) incriminated in the peridomestic and selvatic transmission of leishmania species in Brazil. *Bull. Entomol. Res.* 87, 187–195, 1997.
- Ready, P. D.; de Souza, A. A.; Macario Rebelo, J. M.; Day, J. C.; Silveira, F. T.; Cambell-Ledrum, D.; Davies, C. R.; Costa, J. M. L. Phylogenetic species and domesticity of *Lutzomyia whitmani* at the south-east boundary of Amazonian Brazil. *Transact. R. Soc. Trop. Med. Hyg.* 92, 159–160, 1998.
- Remsen, J. V. Pattern, process, and rigor meet classification. *Auk* 122: 403-413, 2005.
- Ridley, M. *Evolução*. Ed.3. Porto alegre: Artmed. Ano 2006. Cap. 15.
- Roderick, G. K.; Gillespie, R. G. Speciation and phylogeography of Hawaiian terrestrial arthropods. *Mol. Ecol.* 7, 519–532, 1998.
- Rondani, C. *Sopra una Specie di Insetto Dittero*. Memoria Prima per Servire alla Ditterologia Italiana n°1, Donati, Parma. Ano 1840. 16 pp.
- Ronquist, F.; Huelsenbeck, J. P. Mr. Bayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574, 2003.
- Sáez, A. G.; Lozano, E. Body doubles. *Nature* 433, 111, 2005
- Scarpassa, V. M.; Cardoza, T. B.; Cardoso-Júnior, R. P. Population genetics and phylogeography of *Aedes aegypti* (Diptera: *Culicidae*) from Brazil. *Am J Trop Med Hyg* 78: 895–903, 2008.

- Scarpassa, V.M.; Alencar, R. B. *Lutzomyia umbratilis*, the Main Vector of *Leishmania guyanensis*, Represents a Novel Species Complex? PLoS ONE: Speciation in *Lutzomyia umbratilis*. Volume 7, Issue 5, e37341, May 2012. doi:10.1371/journal.pone.0037341
- Shapiro, B.; Rambaut, A. and Drummond, A. J. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol. Biol. Evol. 23, 7–9, 2006.
- Shimabukuro, P. H. F.; Tolezano, J. E.; Galati, E. A. B. Chave de identificação ilustrada dos Phlebotominae (Diptera, Psychodidae) do Estado de São Paulo, Brasil. Papéis Avulsos de Zoologia, Volume 51(27):399-441, 2011.
- Simon, C.; Frati, F.; Beckenbach, A.; Crespi, B.; Liu, H.; et al. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87: 651–701. 38, 1994.
- Simpson, G. G. Principles of Animal taxonomy, Columbia university press, New York. Ano 1961.
- Sneath, F. H. A, Sokal, R. R. Numerical taxonomy. The principles and practice of numerical classification, Freeman, San Francisco, Ca. Ano 1973.
- Souza, N. A.; Vigoder, F. M.; Araki, A. S.; Ward, R. D.; Kyriacou, C. P.; Peixoto, A. A. Analysis of the copulatory courtship songs of *Lutzomyia longipalpis* in six populations from Brazil. JMed Entomol 41: 906-913, 2004.
- Souza, N. A.; Ward, R. D.; Hamilton, J. G.; Kyriacou, C. P.; Peixoto, A. A. Copulation songs in three siblings of *Lutzomyia longipalpis* (Diptera: *Psychodidae*). Trans R Soc Trop MedHyg 96: 102-103, 2002.
- Swofford, D. L. PAUP\*:Phylogenetic analysis using parsimony (\*and other methods), vesion 4.0. Sinauer Associate, Sunterland, Massachusetts. Ano 2003.
- Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A. and Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30(12):2725–2729, 2013. doi:10.1093/molbev/mst197
- Velzen, R. V.; Weitschek, E.; Felici, G.; Bakker, F. T. DNA Barcoding of Recently Diverged Species: Relative Performance of Matching Methods. PLoS ONE. Volume 7, Issue 1, e30490, p. 1-12, 2012. doi:10.1371/journal.pone.0030490

- Ward, R. D., Phillips, A.; Burnet, B.; Marcondes, C. B. The *Lutzomyia longipalpis* complex: reproduction and distribution. In MW Service, Biosystematics of Haematophagous Insects, Oxford University Press, Oxford, p. 258-269, 1998.
- Ward, R. D.; Ribeiro, A. L.; Ready, P. R.; Murtagh, A. (Reproductive isolation between different forms of *Lutzomyia longipalpis* (Lutz & Neiva), (Diptera: *Psychodidae*), the vector of *leishmania donovani chagasi* Cunha & Chagas and its significance to kala-azar distribution in South America. Mem Inst Oswaldo Cruz 78: 269-280, 1983.
- Wheeler, Q. D. and Meier, R. (Eds.). Species Concepts and Phylogenetic Theory: A Debate. Acta Biotheoretica 2002, Volume 50, Issue 2, pp 137-140, 2000.
- White, G. B. (1974) *Anopheles gambiae* complex and disease transmission in Africa. Transactions of the Royal Society of Tropical Medicine and Hygiene 68: 278-301, 1974.
- Wiens, J. J. Combining data sets with different phylogenetic histories. Syst. Biol. 47, 568–581, 1998.
- Wilkerson, R. C.; Foster, P. G.; Li, C.; Sallum, M. A. M. Molecular Phylogeny of Neotropical *Anopheles (Nyssorhynchus) Albitarsis* Species Complex (Diptera: *Culicidae*). Ann Entomol Soc Am.; 98(6): 918–925, 2005.
- Wilkins, J. S. Defining species: a sourcebook from antiquity to today. American University Studies.V. Philosophy. Ano 2009.
- Williams, D. M.; Ebach, M. C. What, Exactly, is Cladistics? Re-writing teh History of Systematics and Biogeography. ActaBiotheor, 57, 2009.
- Winston, J. Describing Species. New York: Columbia University press. Ano 1999.
- Yang, Z. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587–596, 1996.
- Yang, Z. (2005) Bayesian inference in molecular phylogenetics. In Mathematics of evolution & phylogeny (ed. Olivier Gascuel), Oxford Univesity Press, New York. Ano 2005. Cap.3 (p. 63 – 90).
- Yang, Z. and Rannala, B. Molecular phylogenetics: principles and practice. Nature Reviews. Genetics. V. 13. P. 303-314, 2012. doi:10.1038/nrg3186
- Ye, S.; Huang, H.; Zheng, R.; Zhang, J.; Yang, G.; et al. Phylogeographic Analyses Strongly Suggest Cryptic Speciation in the Giant Spiny Frog (Dicroglossidae:

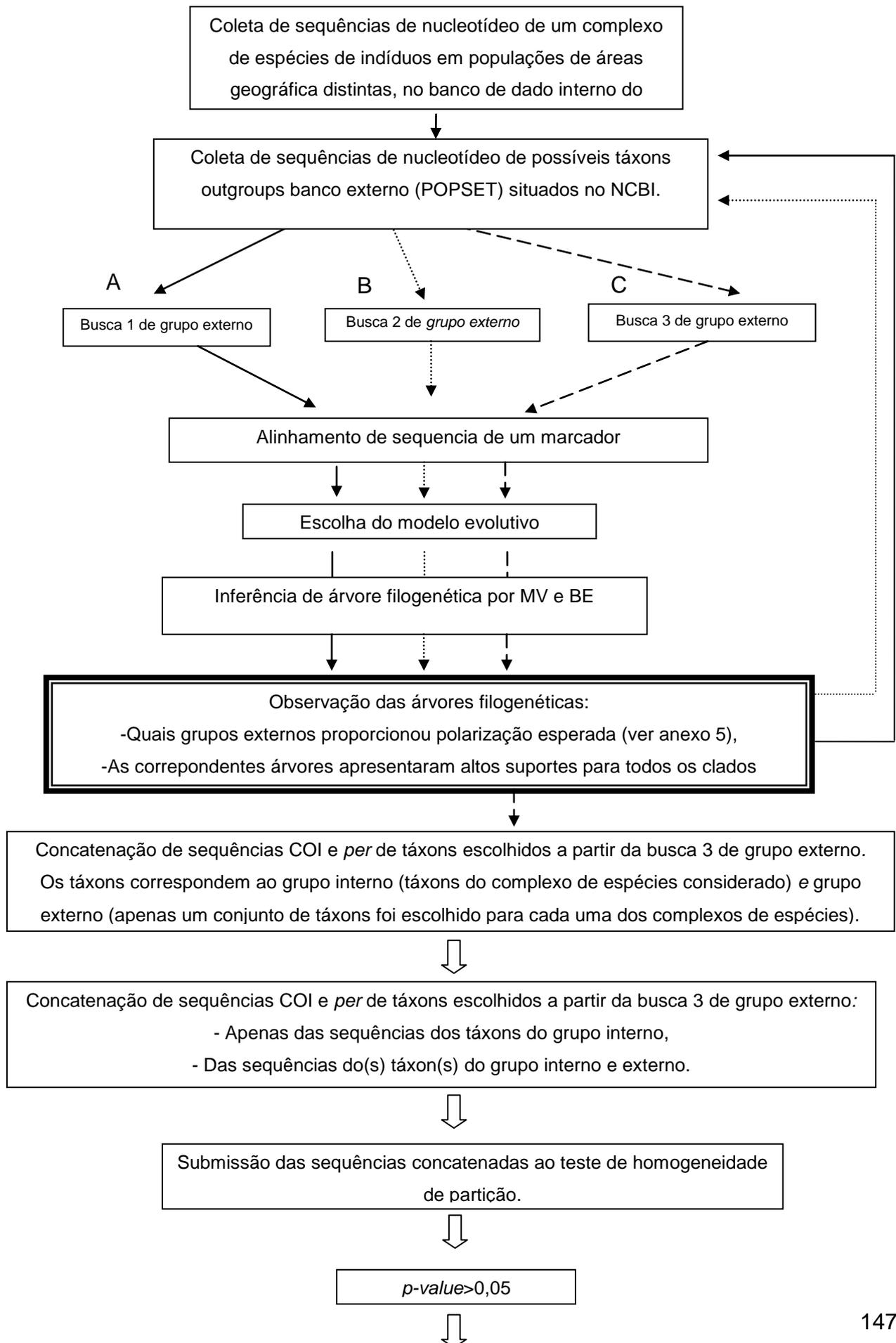
*Paaspinosa*) and Interspecies Hybridization in Paa. PLoS ONE 8(7): e70403, 2013. doi:10.1371/journal.pone.0070403

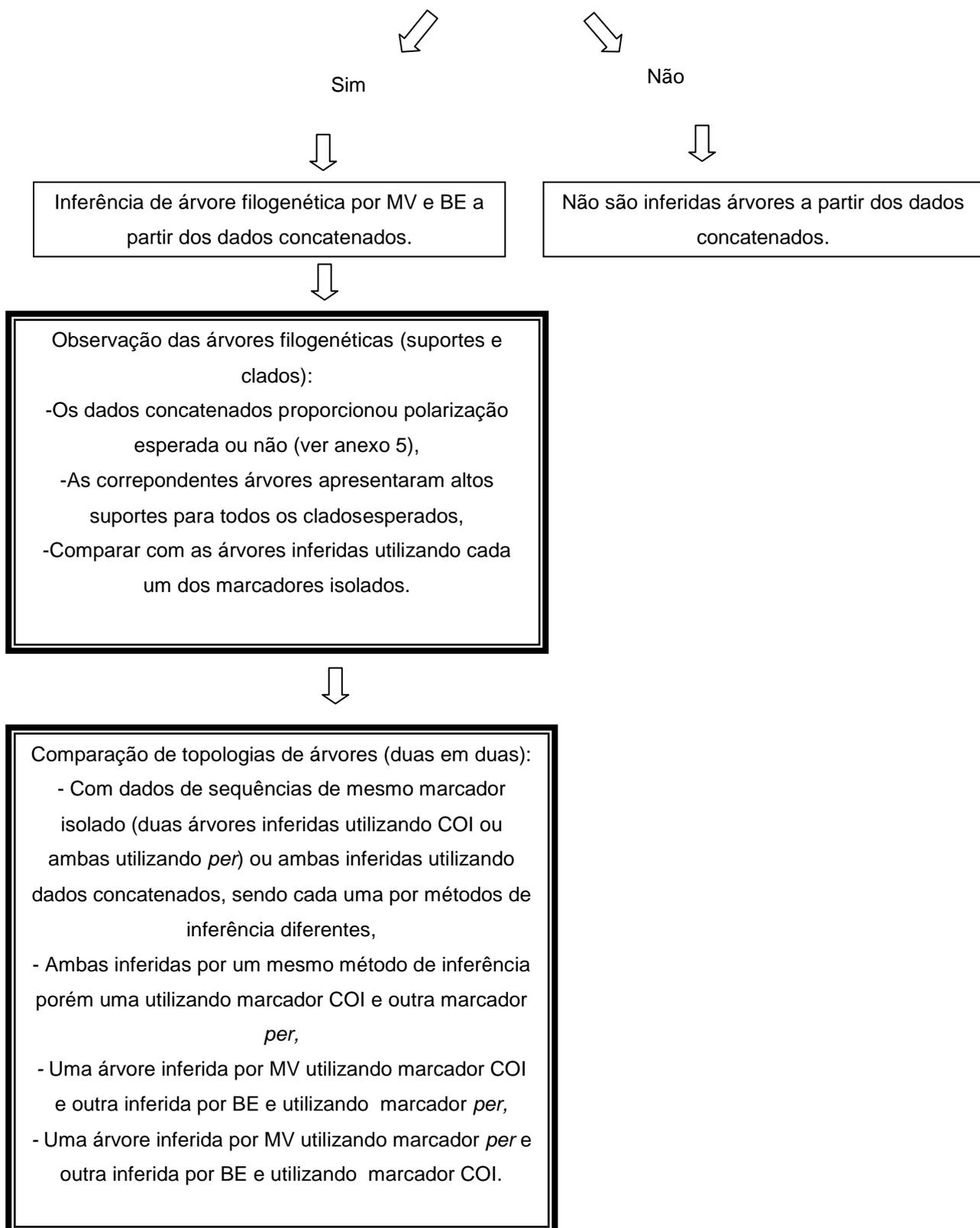
Young, D. G. and Duncan, N. A. Guide to the identification and geographic distribution of *Lutzomyia* sandflyies in Mexico, the West Indies, Central and South America (Diptera: *Psychodidade*). Mem Am Entomollnst 54: 1–881, 1994.

## **APÊNDICE A - FLUXOGRAMA DA METODOLOGIA UTILIZADA NESTE TRABALHO PARA CADA UM DOS COMPLEXOS DE ESPÉCIES CRÍPTICAS**

Espécies crípticas modelo de flebotomíneos: *L. umbratilis*.e *L. longipalpis*.

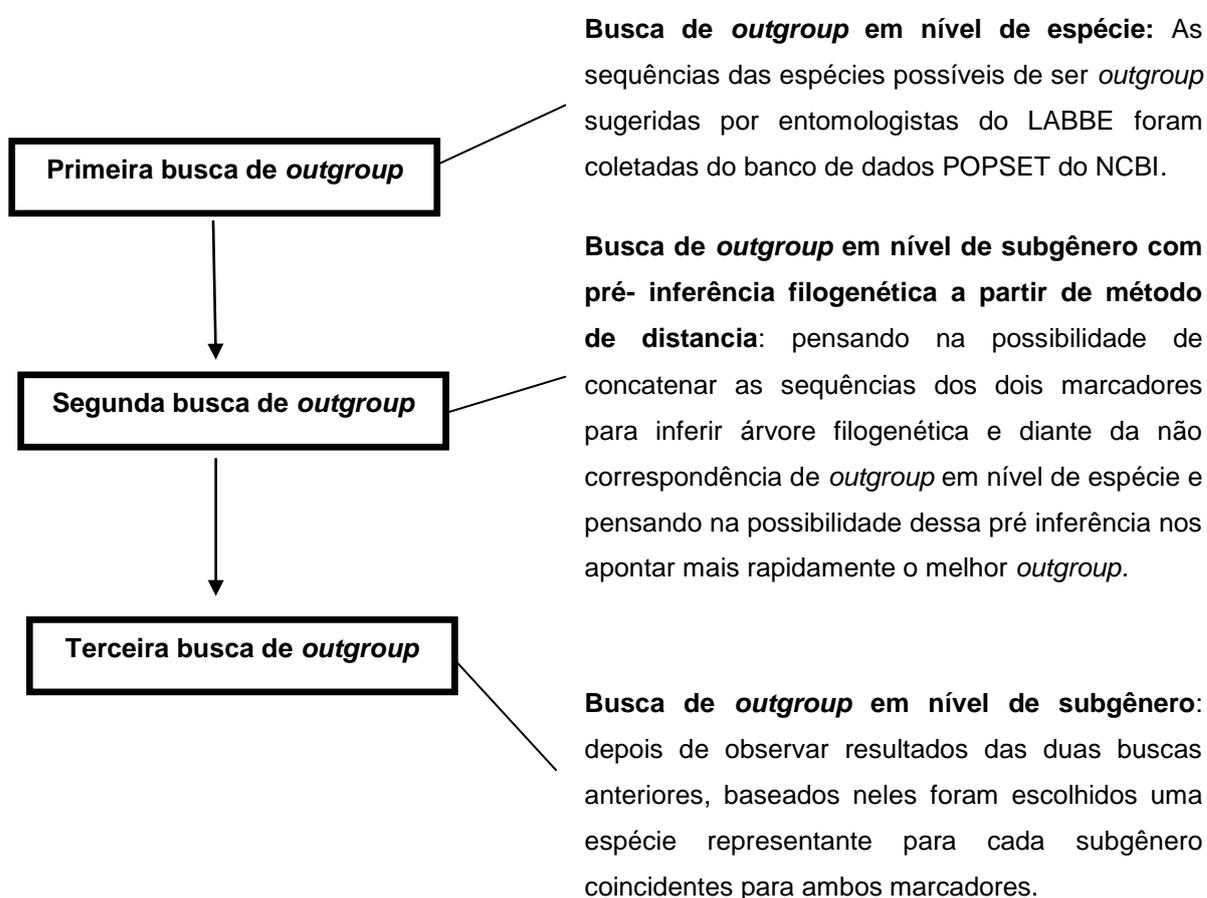
Essas espécies foram utilizadas apenas por questão de conveniência, e não porque essas espécies são modelos padrões de espécies crípticas. O acesso mais fácil aos especialistas dessas espécies crípticas, ou seja, dos entomologistas, aos quais foi preciso fazer perguntas específicas frequentemente e também a alguns dados de sequências de nucleotídeos (referentes à esses complexos) indisponíveis nos bancos de dados do NCBI porém presentes no banco interno do laboratório onde foi desenvolvido esse trabalho, Laboratório de Bioinformática e Biologia Evolutiva (LABBE). Veja abaixo um fluxograma dos passos seguidos na metodologia deste trabalho:





Os três quadros que têm as bordas mais espessas correspondem às etapas nas quais foram obtidos os resultados. Observe que antes do primeiro desses três quadros há 3 tipos de setas, cada uma correspondente caminhos sequências. Ao sair do segundo quadro geral do fluxograma, inicia-se o caminho "A" percorrendo todas as setas sólidas. Este caminho é finalizado no segundo quadro geral. Ao chegar neste passo novamente, o próximo passo é percorrer todo o caminho "B", representado pelas setas pontilhadas, o qual também será finalizado no segundo quadro geral. O próximo passo desta vez é percorrer o caminho "C" representado pelas setas tracejadas. Este por sua vez será finalizado no quadro logo após o primeiro quadro com bordas mais espessas.

## APÊNDICE B: FLUXOGRAMA DA BUSCA DE GRUPO EXTERNO PARA CADA UM DOS COMPLEXOS DE ESPÉCIES



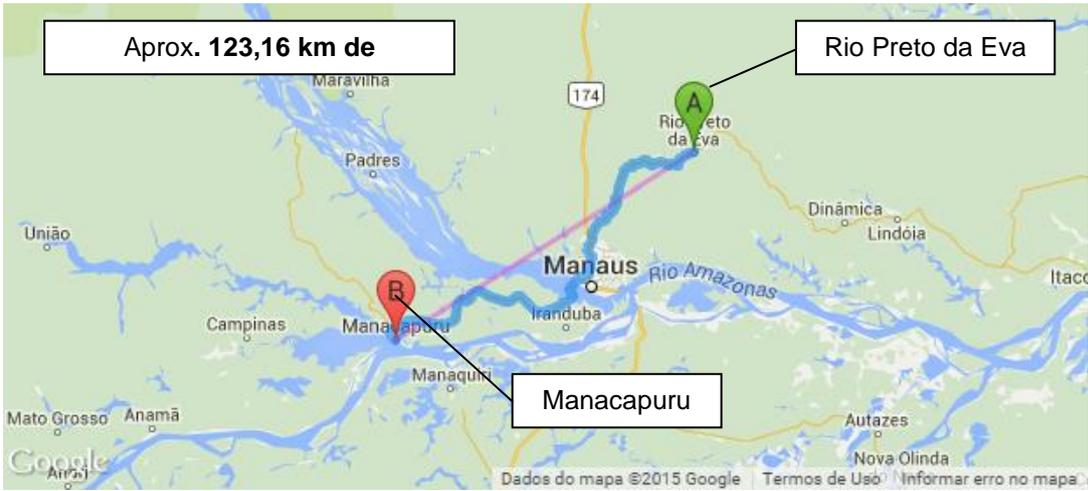
**APÊNDICE C - ÁREAS E COORDENADAS GEOGRÁFICAS E ASPECTOS  
BIOGEOGRÁFICOS CORRESPONDENTES ÀS SEQUÊNCIAS DAS ESPÉCIES  
CRÍPTICAS *L. UMBRATILIS* E *L. LONGIPALPIS***

As sequencias do grupo interno (ou seja, correspondentes aos táxons do considerado complexo de espécies crípticas) de áreas diferentes do território brasileiro, dentre de um mesmo estado ou não. Essas diferentes áreas são ditantes umas das outras e quando não, entre elas há alguma barreira, como por exemplo, o Rio Amazonas entre Rio Preto da Eva e Manacapuru.

**a) *L. umbratilis***

- Município Rio Preto da Eva - localizado ao norte do Rio Amazonas (2°50'50"Sul / 59°56'28"Oeste), no estado da Amazônia, Região Norte do Brasil,
- Município Manacapuru - localizado ao sul do Rio Amazonas (3°12'41"Sul / 60°26'20" Oeste), localizado no estado da Amazônia, Região Norte do Brasil,
- Reserva ecológica da Mata Atlântica no bairro de Dois Irmãos - (8°03'14"Sul / 34°52'52" Oeste) pertencente ao município de Recife, Estado de Pernambuco, Região Nordeste do Brasil.

Veja mapas abaixo (Fonte: *Google maps*):



**b) *L. longipalpis***

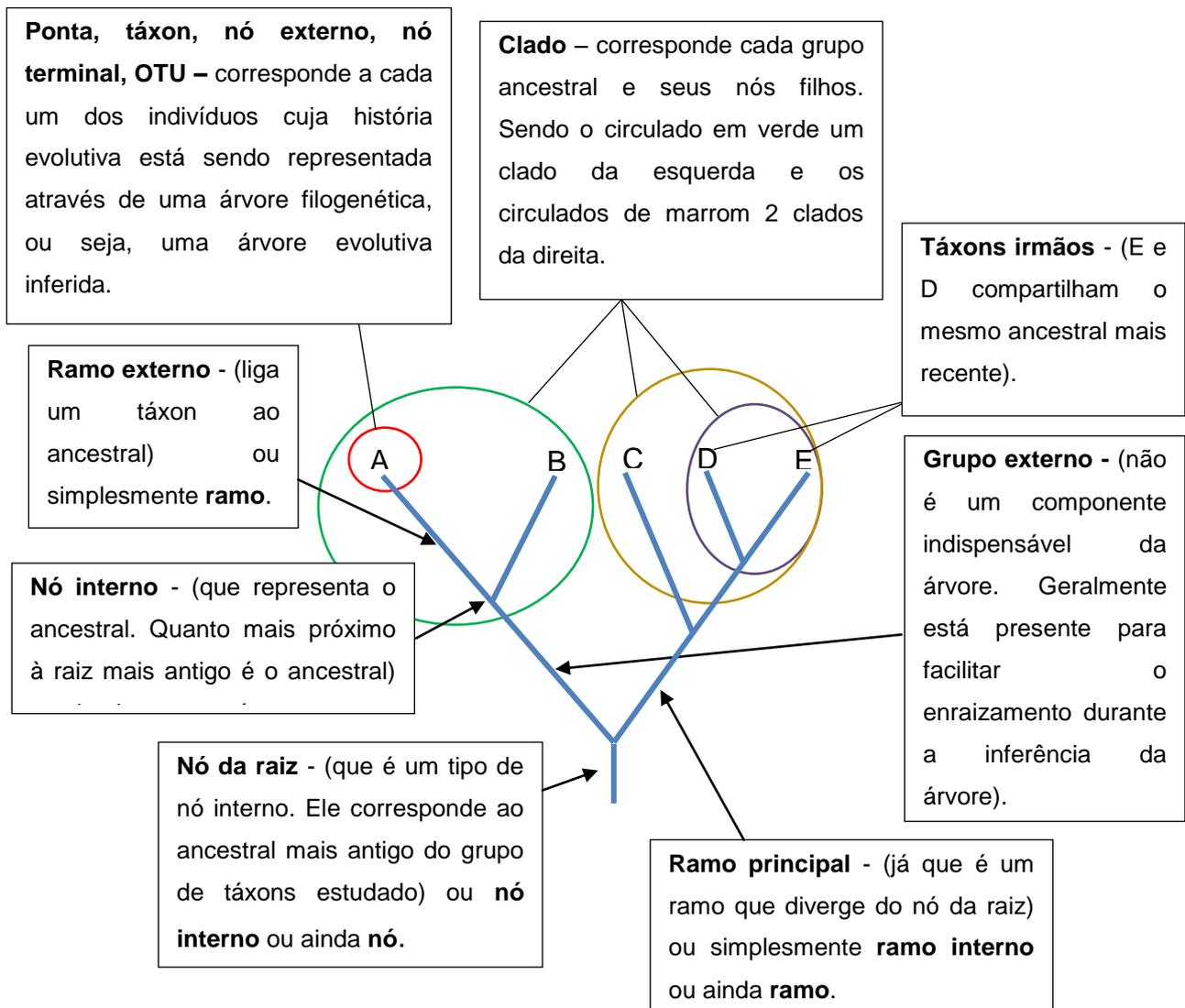
- Município Sobral - (3°41'15" Sul / 40°21'5" Oeste) - localizado no norte estado do Ceará, Região Nordeste do Brasil,
- Município Caririçu - (07°02'31" Sul / 39°17'02" Oeste) localizado no sul do estado do Ceará, Região Nordeste do Brasil.

Veja mapa abaixo (Fonte: *Google Maps*):



**APÊNDICE D - ESTRUTURAS DE UMA ÁRVORE FILOGENÉTICA**

É uma forma gráfica de se representar a filogenia: forma de árvore (um tipo de grafo). Os componentes da árvore filogenética são:

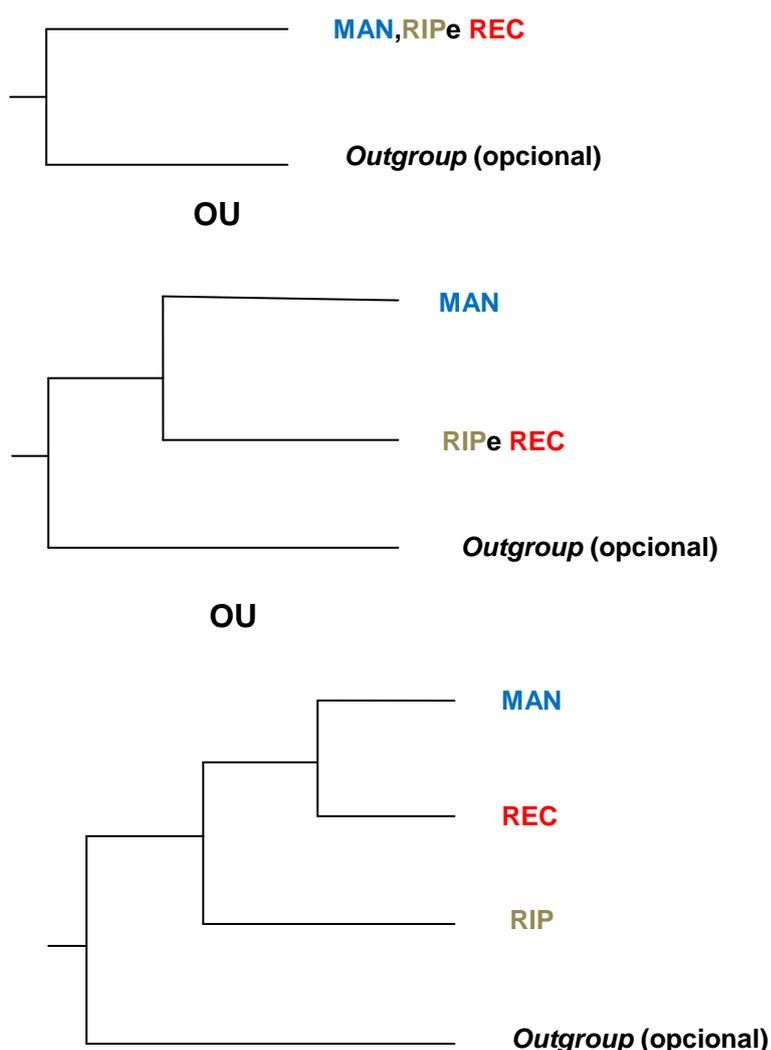


## APÊNDICE E - CLADOS IMPORTANTES NO SENTIDO BIOLÓGICO - POLARIZAÇÃO ESPERADA COM AS ÁRVORES ENRAIZADAS VIA GRUPO EXTERNO

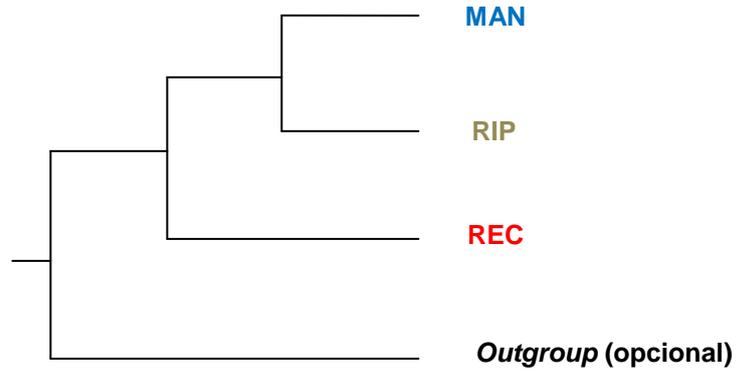
### a) *L. umbratilis*: um a três cladros esperados

Os táxons utilizados neste trabalho correspondem a indivíduos pertencentes a três populações que estão em duas áreas geográficas diferentes como visto no anexo 3 (Manacapuru - Amazonas; Rio Preto da Eva - Amazonas e Recife - Pernambuco). Sendo *L. umbratilis* um complexo de espécies crípticas no Brasil, tem

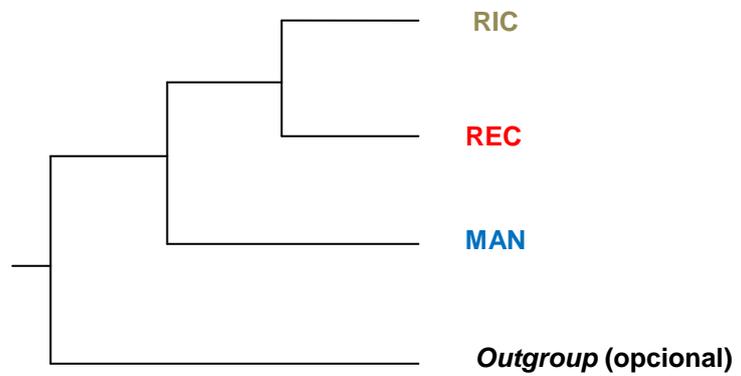
possibilidade de que tais populações dessas regiões possam corresponder à apenas 1 espécie ou à 2 ou 3 espécies distintas. As árvores que conseguiram distinguir grupos bem distintos neste trabalho sempre polarizaram em 2 grupos, provavelmente correspondentes a duas espécies distintas: grupo 1- população de Manacapuru (MAN) e grupo 2 - população de Rio Preto da Eva (RIP) e Recife (REC). Essas são as características das árvores referidas neste trabalho como "polarizadas de modo esperado" ou "melhores no aspecto biológico" ou ainda "com clados principais esperados" para *L. umbratilis*. Veja na figura abaixo um modelo gráfico geral das possibilidades de 1, 2 ou 3 espécies:



OU



OU

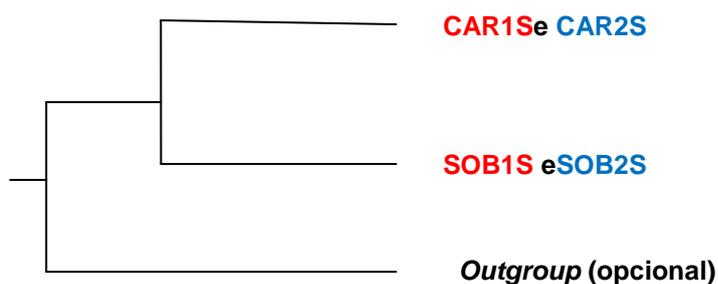


b) *L. longipalpis*: um a quatro clados esperados

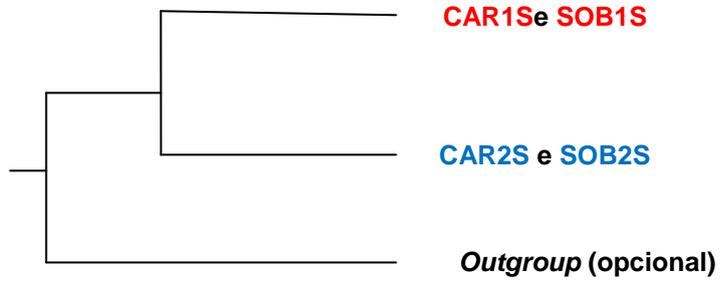
Os táxons utilizados neste trabalho correspondem a indivíduos pertencentes à quatro populações que estão em áreas geográficas diferentes como visto no anexo 3: duas populações foram do município de Caririçu - Ceará, sendo uma com fenótipo 1S (CAR1S) e outra com fenótipo 2S (CAR2S) e as outras duas foram do município de Sobral - Ceará, sendo uma com fenótipo 1S (SOB1S) e a outra com fenótipo 2S (SOB2S). Sendo *L. longipalpis* um complexo de espécies crípticas no Brasil, tem possibilidade de que tais populações dessas regiões possam corresponder à apenas 1 espécie ou à 2 ou 3 ou 4 espécies distintas. As árvores que conseguiram distinguir grupos bem distintos neste trabalho sempre polarizaram em 2 grupos, provavelmente correspondentes a duas espécies distintas: grupo 1- populações com fenótipo 1S(CAR1S e SOB1S) e grupo 2 - populações com fenótipo 2S (CAR2S e SOB2S). Essas são as características da árvore referida neste trabalho como "polarizadas de modo esperado" ou "melhores no aspecto biológico" ou ainda "com clados principais esperados" para *L. longipalpis*. Veja na figura abaixo um modelo gráfico geral das possibilidades de 1, 2, 3 ou 4 espécies:



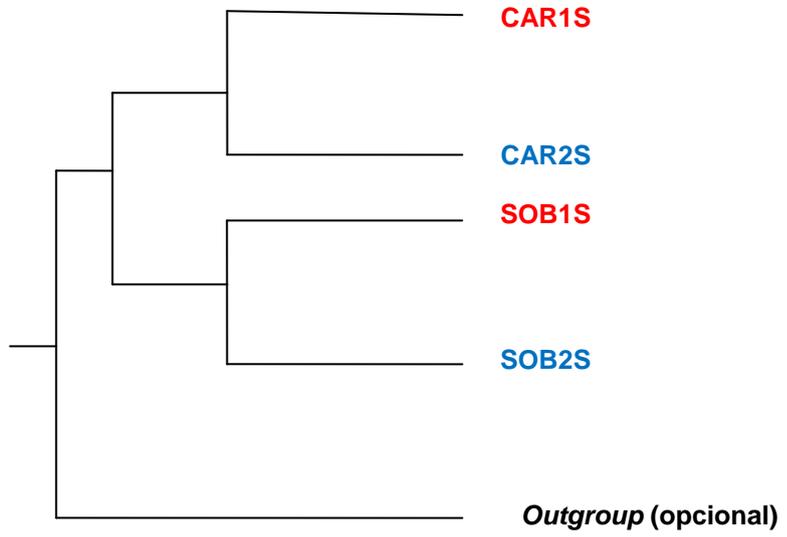
OU



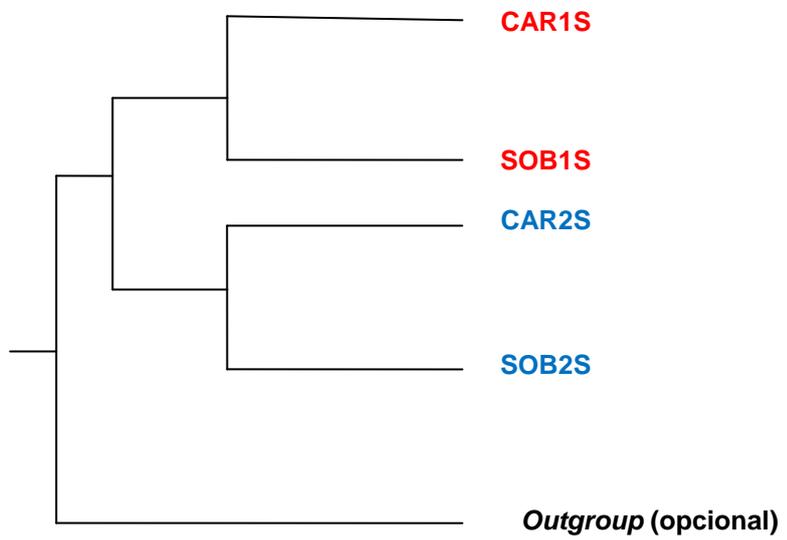
**OU**



**OU**



**OU**



## APÊNDICE F - MELHORES ÁRVORES DO COMPLEXO *L. UMBRATILIS*

Neste anexo estão as árvores que apresentaram melhores resultados quanto a polarização e quanto ao suporte dos dois clados principais formados (cujos táxons estão nas cores azul e vermelha).

Antes de apresentar as figuras vale dar algumas três informações importantes:

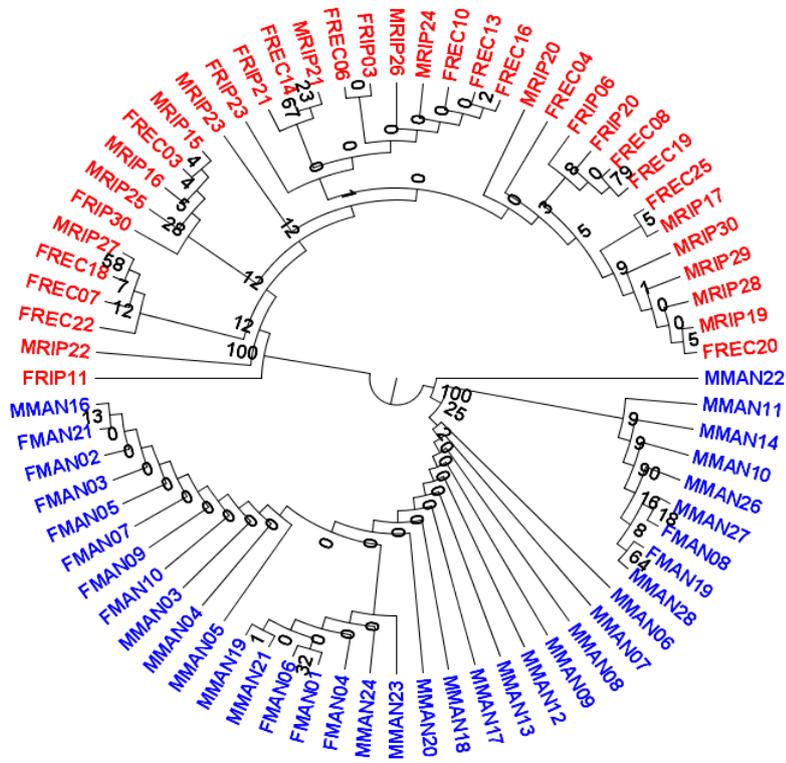
- Em árvores de *L. umbratilis*: Grupo de táxons vermelhos - Táxons dos municípios de Rio Preto da Eva (FRIP - fêmeas, MRIP - machos) e Recife (FREC - fêmeas e MREC - machos) e Grupo de táxons azuis - Táxons do município de Manacapuru (MAN), sendo as fêmeas representadas por FMAN e os machos por MMAN;

- Em árvores de *L. longipalpis*: Grupo de táxons vermelhos - Táxons com fenótipo 1S do município de Caririçu (CAR) e de Sobral (SOB) e Grupo de táxons azuis - Táxons com fenótipo 2S do município de Caririçu (CAR) e de Sobral (SOB).

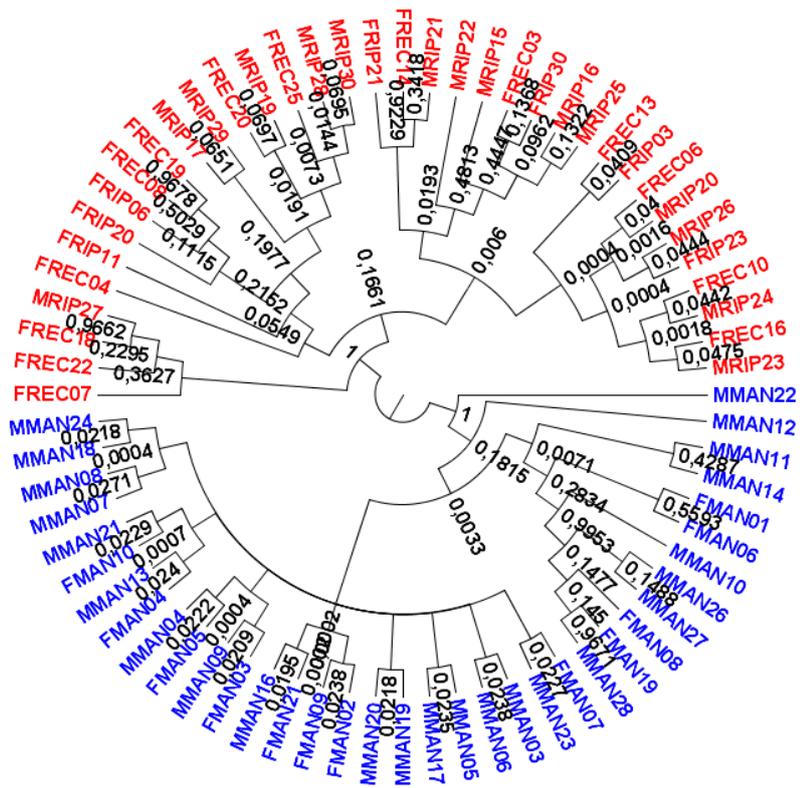
### **a) Primeira busca de grupo externo para *L. umbratilis*:**

- **Utilizando marcador COI:**

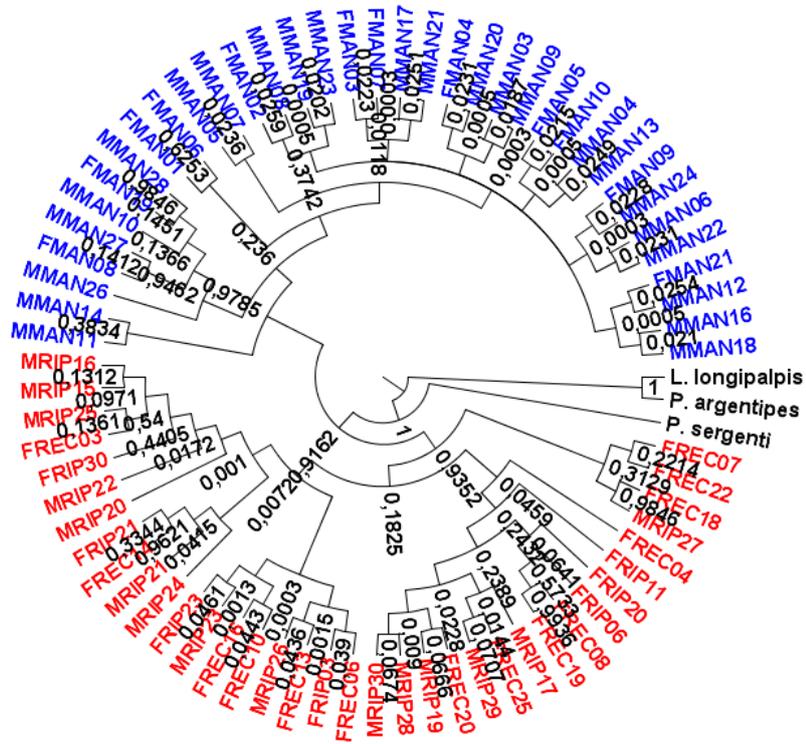
1. Sem grupo externo e inferida por MV



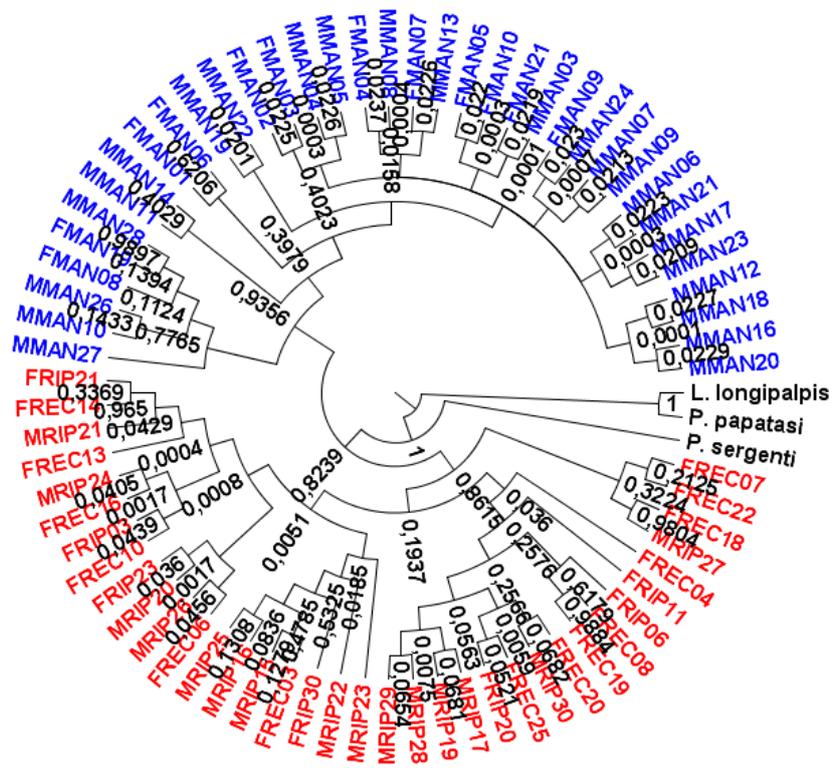
2. Sem grupo externo e inferida por BE



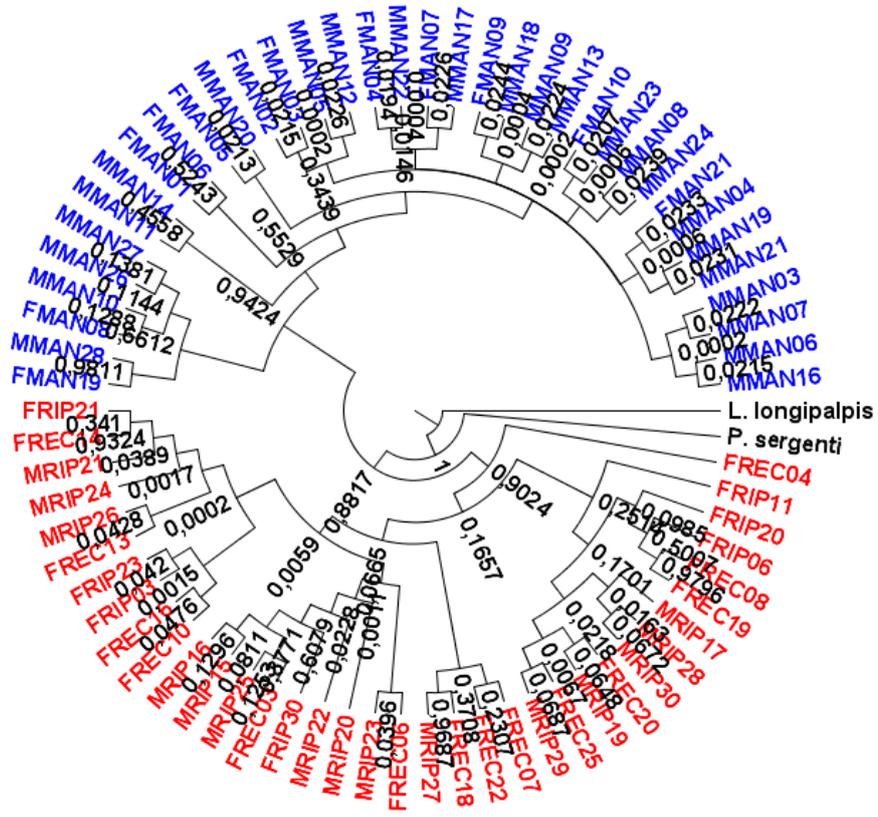
3. Grupo externo A+S+L e inferida BE



4. Grupo externo P+S+L e inferido por BE

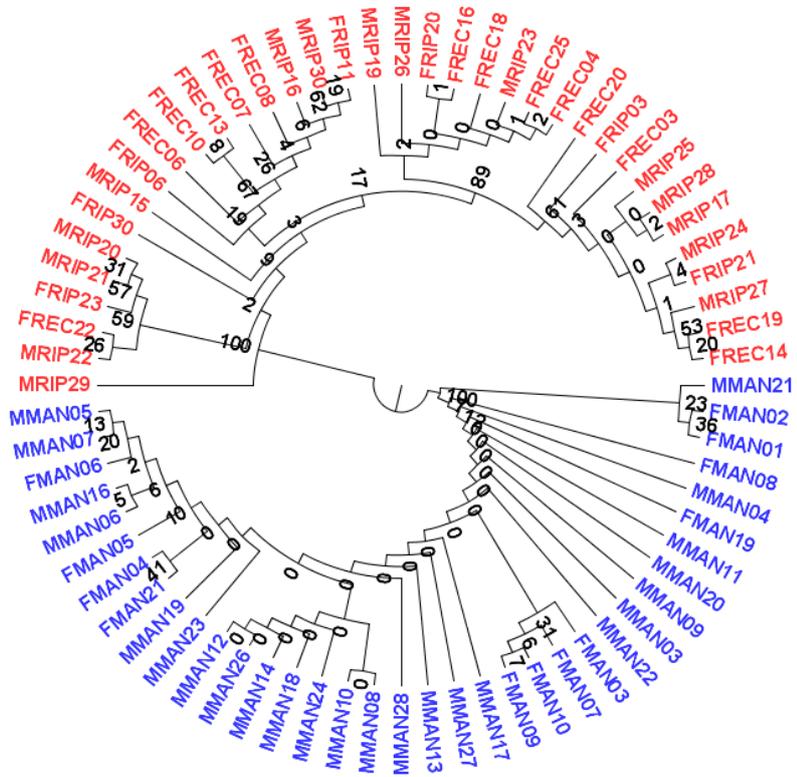


5. Grupo externo S+L e inferido por BE

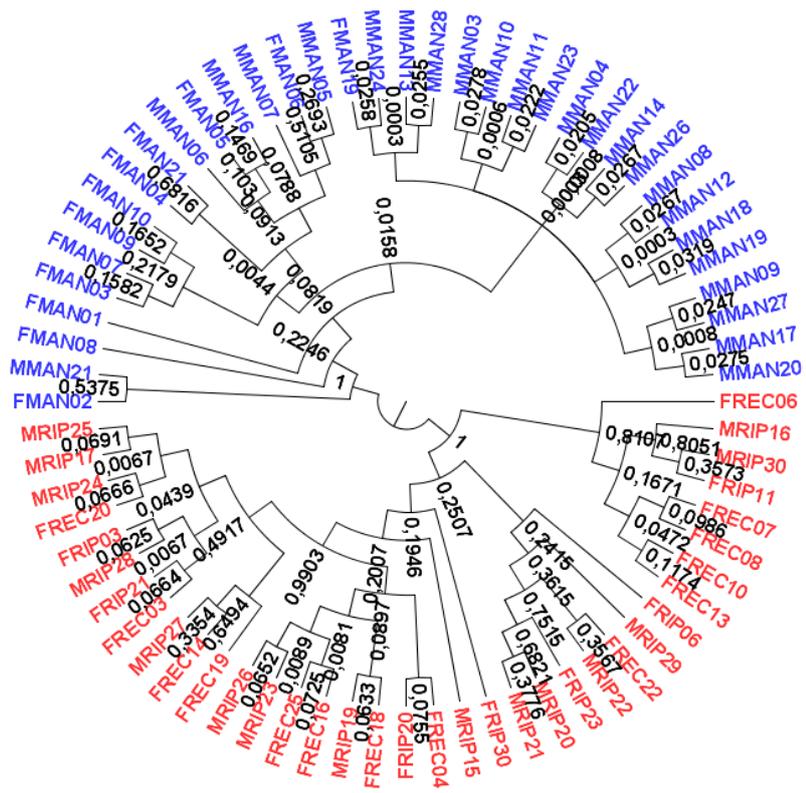


- Utilizando marcador *per*:

1. Sem grupo externo e inferida por MV



2. Sem grupo externo e inferida por BE



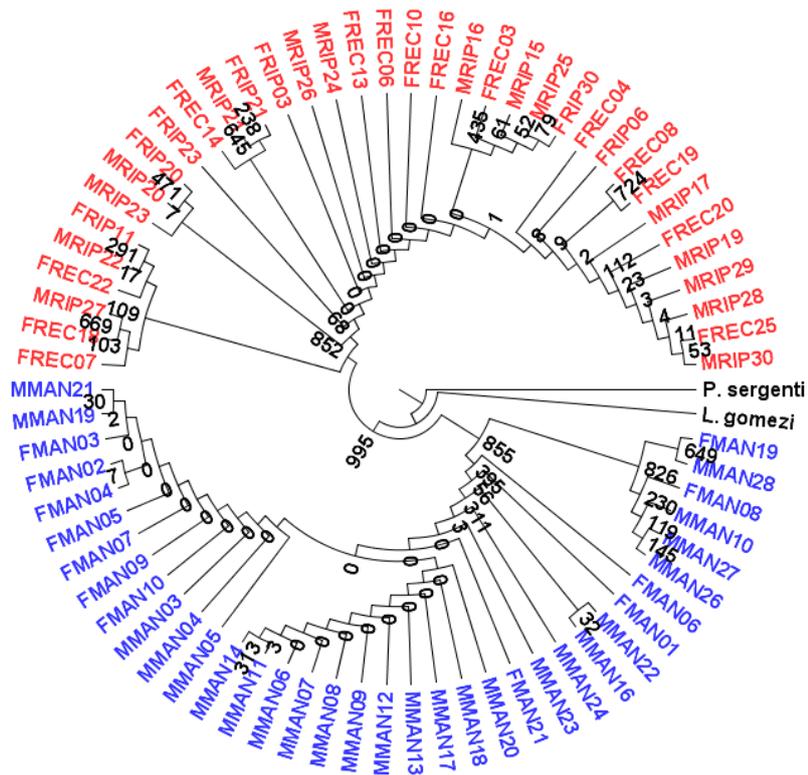




b) Segunda busca de grupo externo para *L. umbratilis*:

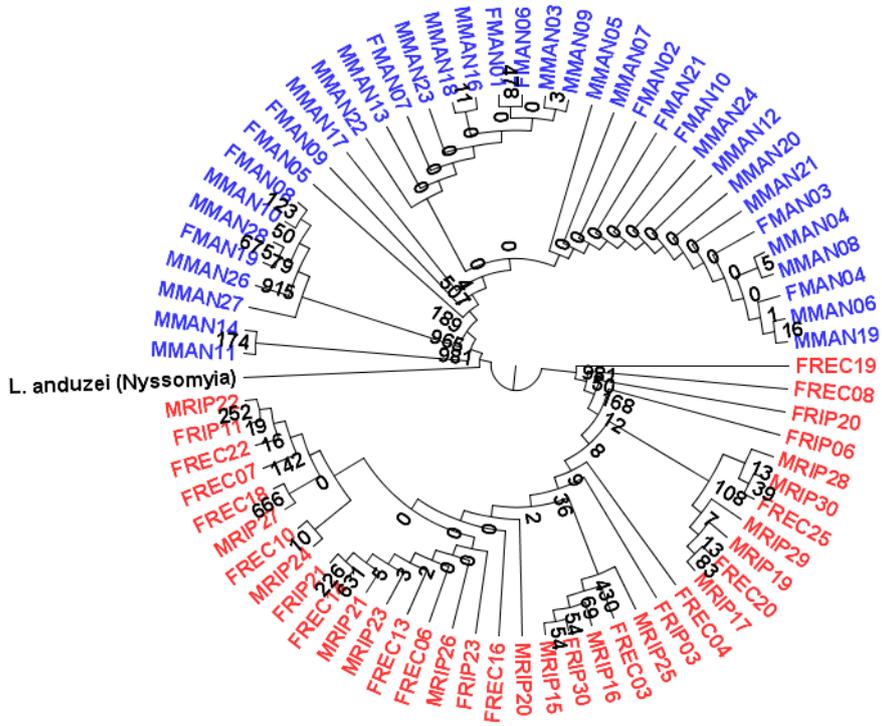
- Utilizando marcador COI:

1. Grupo externo S+G e inferida por MV

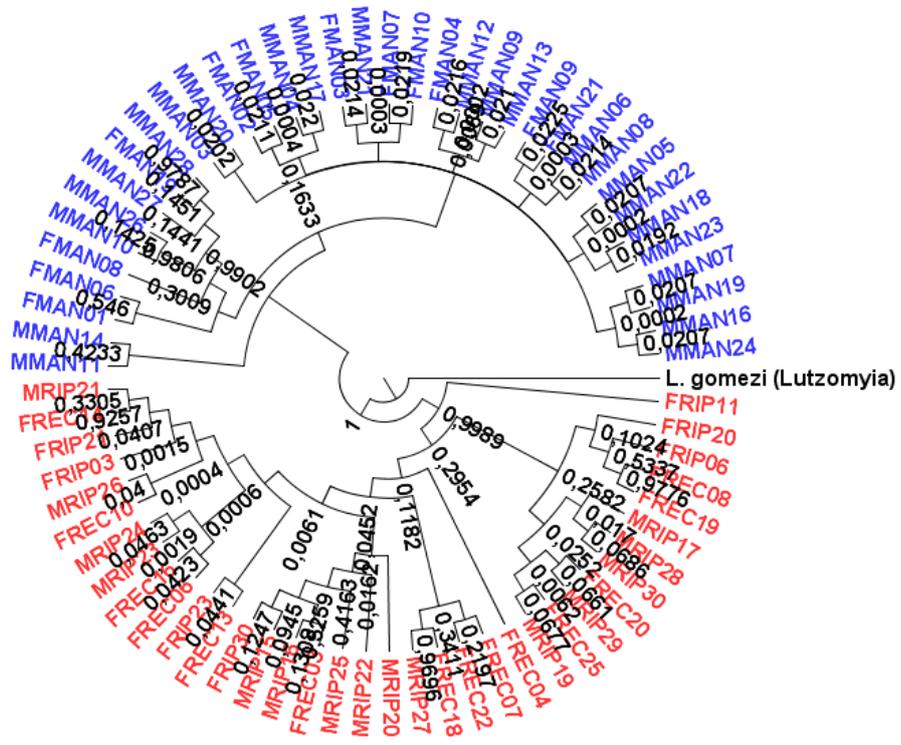




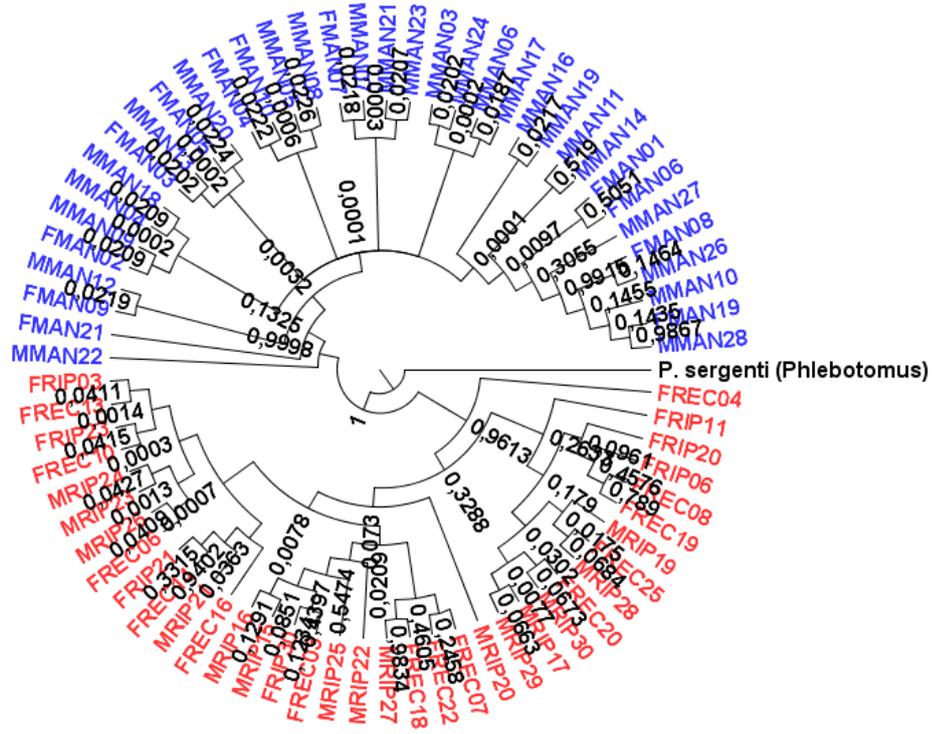
1. Grupo externo A e inferida por MV



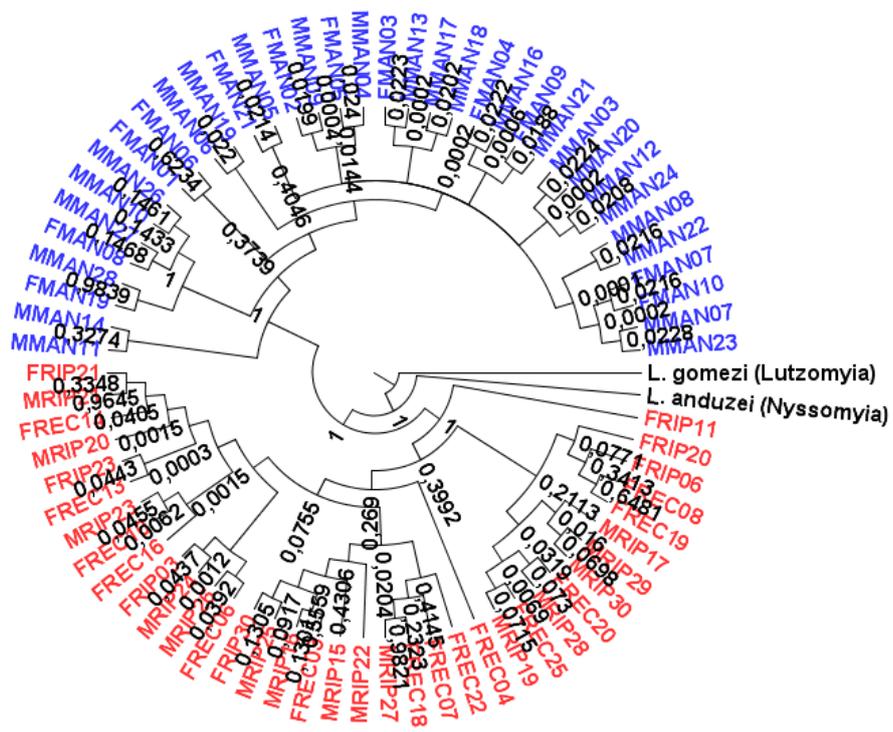
2. Grupo externo G e inferida por BE



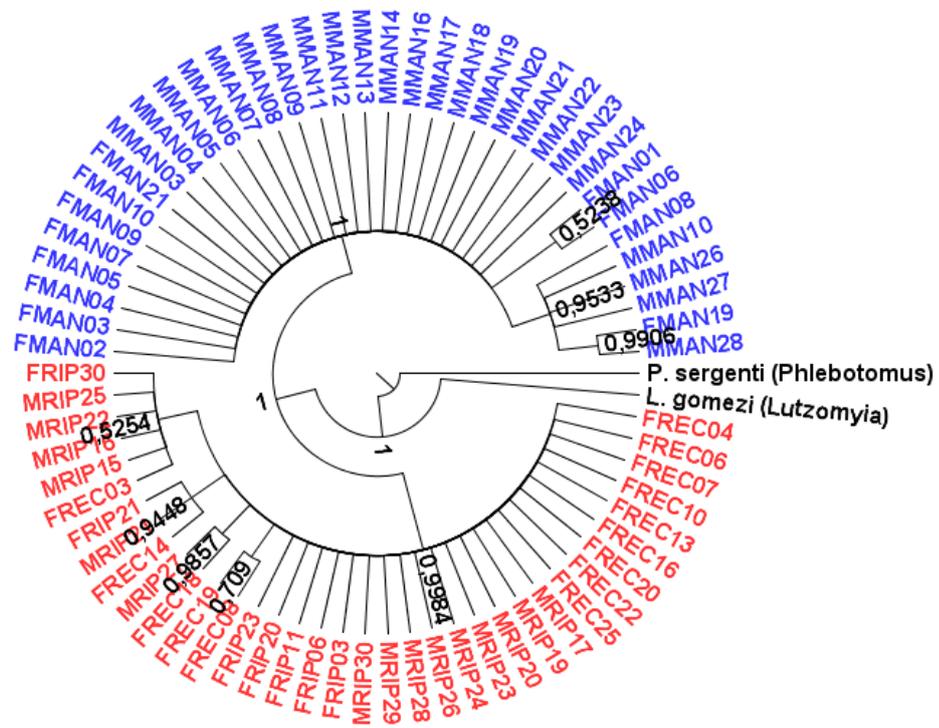
3. Grupo externo S e inferida por BE



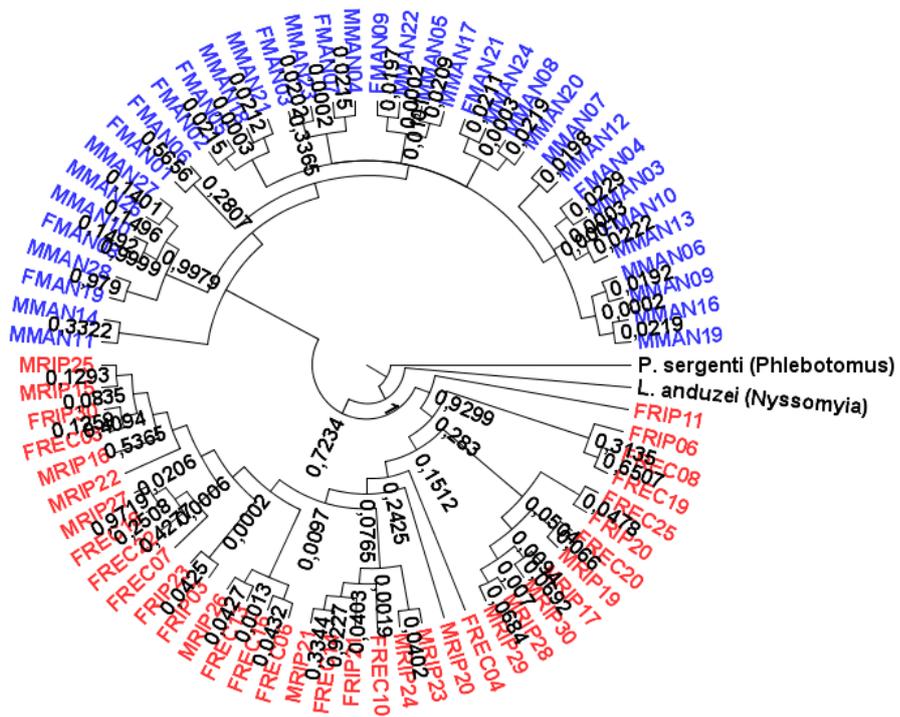
4. Grupo externo G+A e inferida por BE



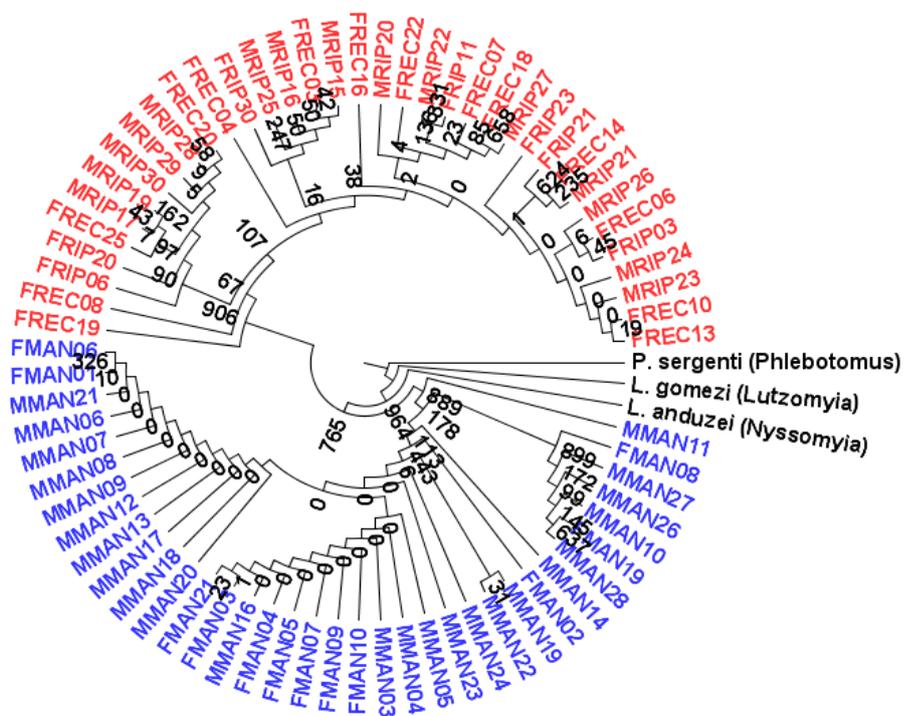
5. Grupo externo S+G e inferida por BE



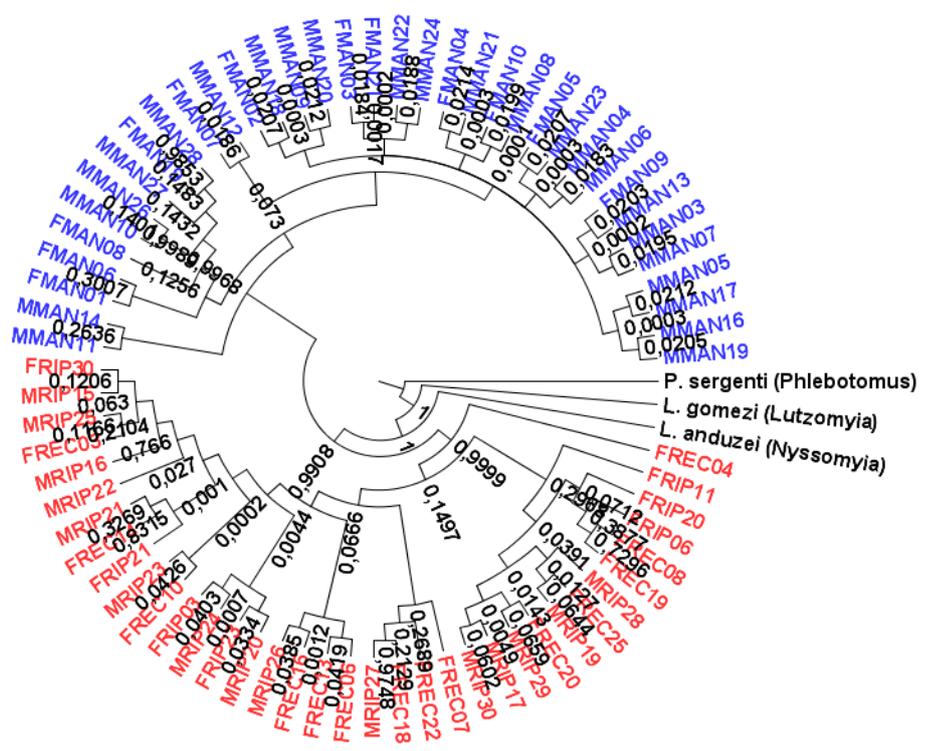
6. Grupo externo A+S e inferida por BE



7. Grupo externo A+G+S e inferida por MV

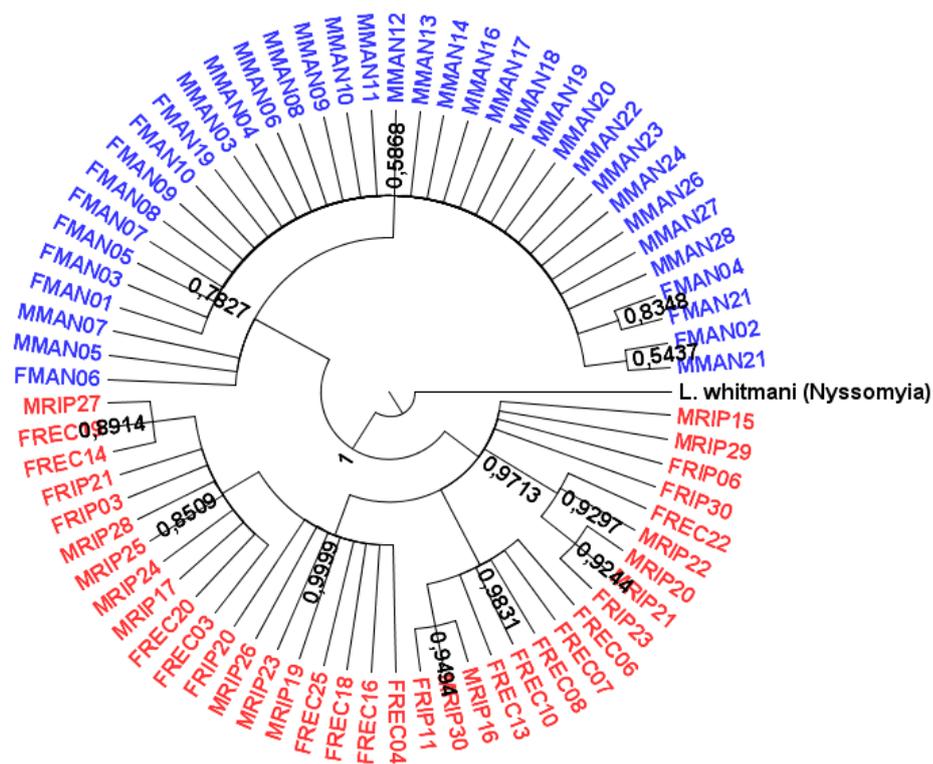


8. Grupo externo A+G+S e inferida por BE

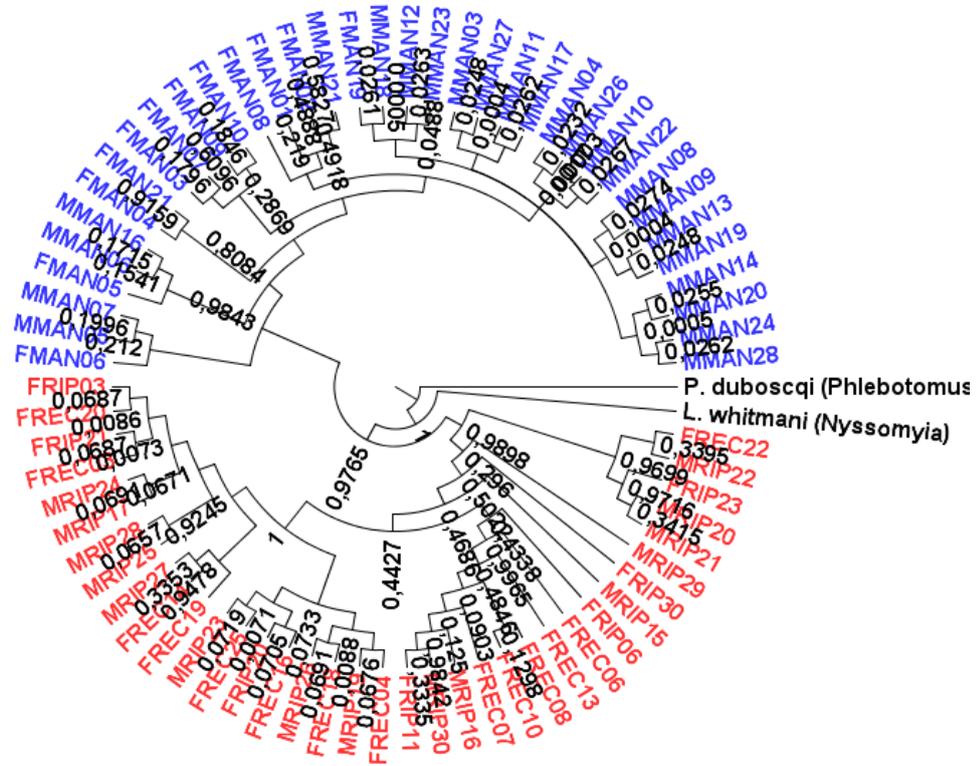


d) Utilizando marcador *per*:

1. Grupo externo W e inferida por BE

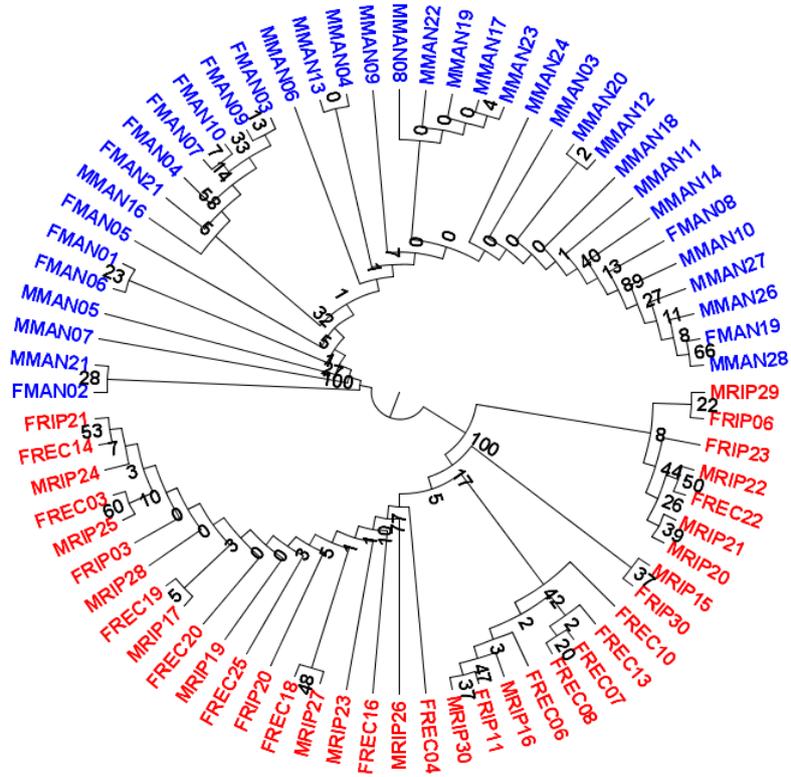


2. Grupo externo W+D e inferida por BE

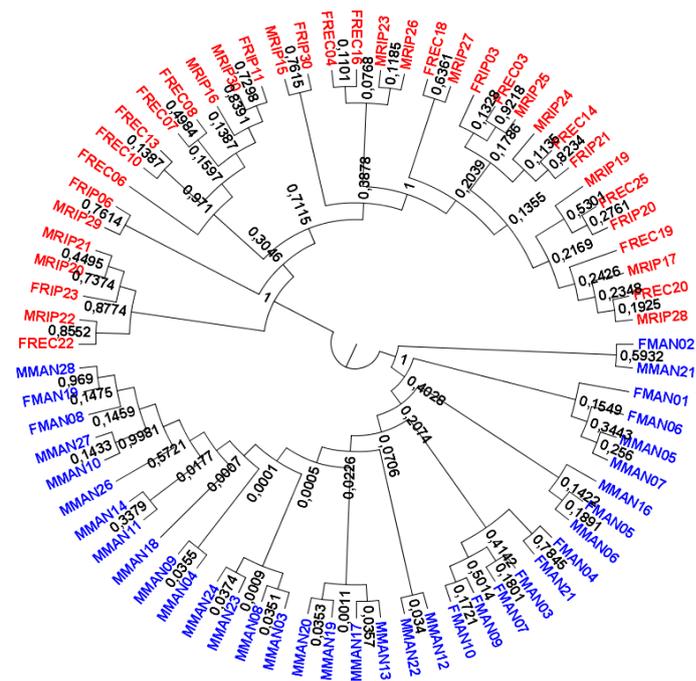


d) Utilizando dados concatenados:

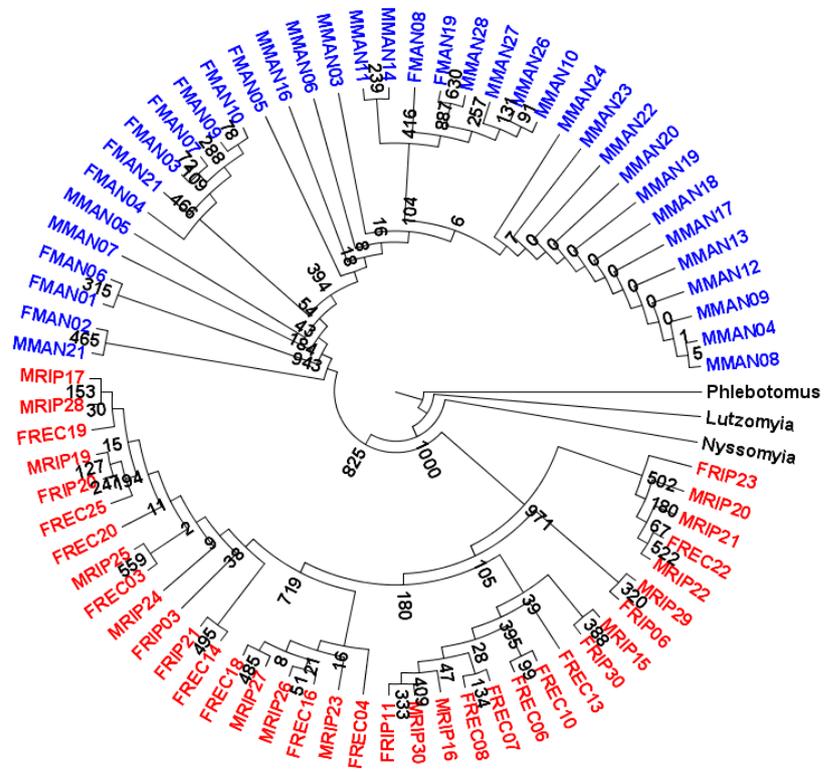
1. Sem grupo externo e inferida por MV (*bootstrap* - 100)



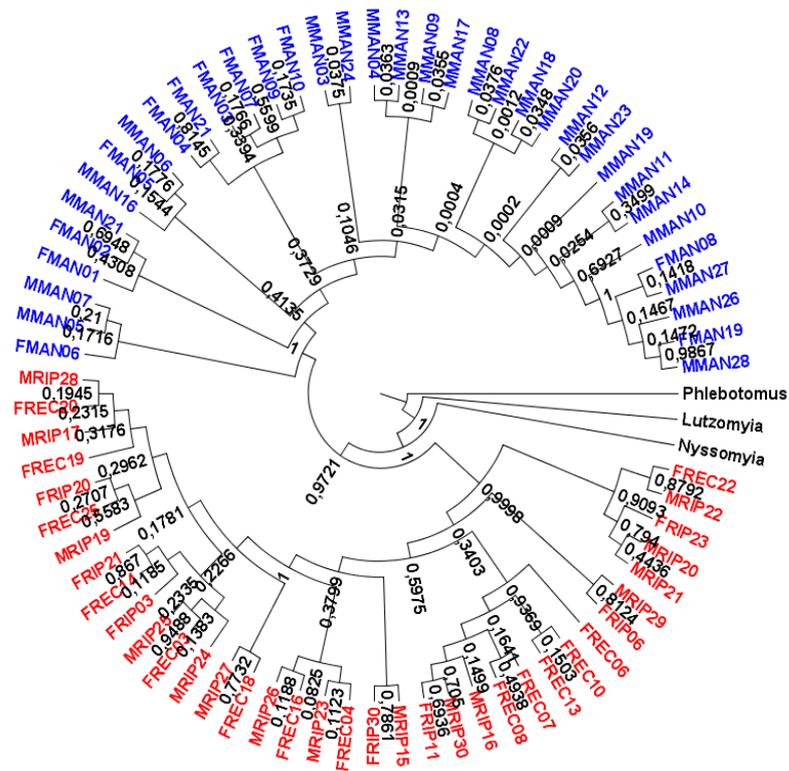
2. Sem grupo externo e inferida por BE



3. Com grupo externo e inferida por MV



4. Com grupo externo e inferida por BE

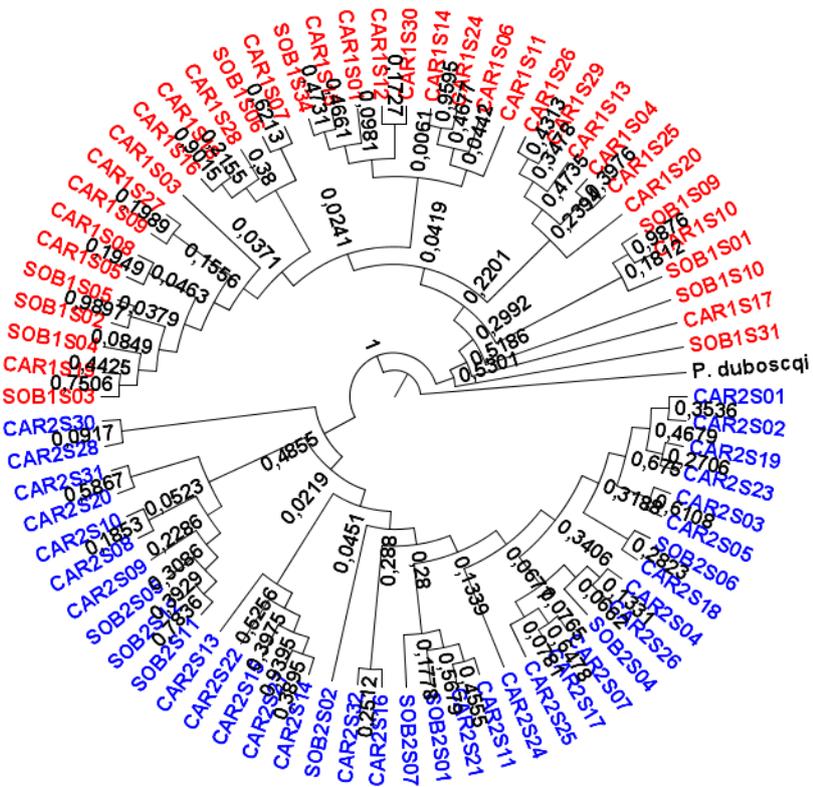


## APÊNDICE G - MELHORES ÁRVORES DO COMPLEXO *L. LONGIPALPIS*

### a) Primeira e terceira busca de grupo externo para *L. longipalpis*:

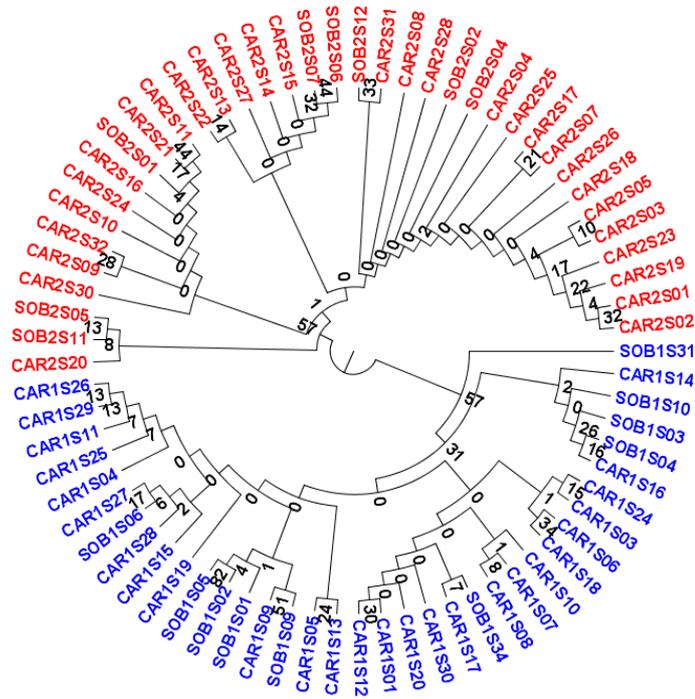
- Utilizando marcador *per*:

#### 1. Grupo externo D e inferida por BE

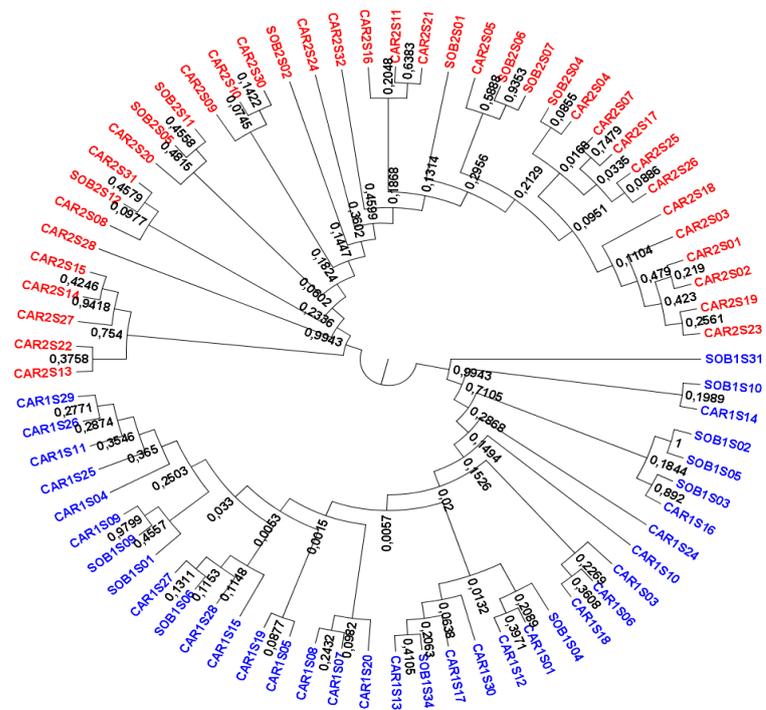


### b) Utilizando dados concatenados:

1. Sem grupo externo e inferida por BE



2. Com grupo externo e inferida por BE



### 3. Com grupo externo e inferida por BE

