



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR
E DO DESENVOLVIMENTO

Kamila Voges

Especiação em *Anopheles (Kerteszia) cruzii* e *Anopheles (Kerteszia) bellator*, os principais vetores de malária na Mata Atlântica brasileira

Florianópolis

2023

Kamila Voges

Especiação em *Anopheles (Kerteszia) cruzii* e *Anopheles (Kerteszia) bellator*, os principais vetores de malária na Mata Atlântica brasileira

Tese submetida ao Programa de Pós-graduação em Biologia Celular e do Desenvolvimento, do Centro de Ciências Biológicas da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutora em Biologia Celular e do Desenvolvimento.

Orientadora: Luísa Damazio Rona Pitaluga.

Coorientador: Antonio Bernardo de Carvalho.

Florianópolis

2023

Voges, Kamila

Especiação em *Anopheles* (*Kerteszia*) *cruzei* e *Anopheles* (*Kerteszia*) *bellator*, os principais vetores de malária na Mata Atlântica brasileira / Kamila Voges ; orientador, Luísa Damazio Rona Pitaluga, coorientador, Antonio Bernardo de Carvalho, 2023. 212 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento, Florianópolis, 2023.

Inclui referências.

1. Biologia Celular e do Desenvolvimento. 2. Malária, *Anopheles*, especiação. I. Pitaluga, Luísa Damazio Rona . II. de Carvalho, Antonio Bernardo . III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento. IV. Título.

Kamila Voges

Especiação em *Anopheles (Kerteszia) cruzii* e *Anopheles (Kerteszia) bellator*, os principais vetores de malária na Mata Atlântica brasileira

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 30 de agosto de 2023, pela banca examinadora composta pelos seguintes membros:

Dr. Leonardo Barbosa Koerich

Instituição: Universidade Federal de Minas Gerais

Dr. Guilherme de Toledo e Silva

Instituição: Universidade Federal de Santa Catarina

Dr. Paulo Eduardo Martins Ribolla

Instituição: Universidade Estadual Paulista “Júlio de Mesquita Filho” Campus de Botucatu

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutora em Biologia Celular e do Desenvolvimento.

Coordenação do Programa de Pós-Graduação

Orientadora: Dra Luisa Damazio Rona Pitaluga

Instituição: Universidade Federal de Santa Catarina

*Dedico este
trabalho aos meus pais,*

Vera e Fidêncio.

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse uma gota”.

Madre Teresa de Calcutá

Agradecimentos

Sem dúvida, neste momento, sinto predominantemente um sentimento de gratidão. Sempre gostei muito do meu trabalho em si. Mas além disso, ao longo do meu doutorado, tive a oportunidade de conhecer pessoas e lugares que tornaram essa experiência ainda mais gratificante. Além de contribuírem para a conclusão desta tese, essas pessoas e lugares me engradeceram muito como pessoa.

Eu gostaria de começar agradecendo a minha família e em especial, aos meus pais que me deram a vida, me apoiaram e sempre me ensinaram a lutar pelo meu sonho.

Aos meus orientadores Luísa e Bernardo.

Luísa que me acolheu no laboratório já lá em 2012, obrigada por tudo que tem me ensinado desde então! Me ensinou desde minha primeira PCR e a fazer ciência. É para mim um exemplo de cientista e de força.

Muitas vezes eu ouvi Luísa dizer: Eu sou fã do Bernardo. E hoje, com certeza eu faço parte desse fã clube! Quando entrei no doutorado minha familiaridade com bioinformática era quase nula. No entanto, o Bernardo sempre demonstrou muita paciência e dedicação, me ensinando e ajudando-me constantemente com os comandos e scripts. Com Bernardo, também aprendi e aprendo muito para além da academia, tem um exemplo em especial: Quando me lembro do “dia do Ipê”. Antes, eu passava todos os dias pelo mesmo caminho da UFRJ sem reparar no ipê. No entanto, em uma carona, Bernardo gentilmente me sinalizou sua beleza. A partir desse momento, iniciamos uma conversa sobre a importância de perceber as coisas simples da vida. E esse é apenas um dos exemplos das lições de vida que aprendo com ele.

Gostaria de agradecer aos amigos, sem eles com certeza essa tese também não teria sido escrita. Eles que me apoiaram, aguentaram meus desabafos, meus chororôs e também riram comigo. Tornaram o processo mais leve. Entre eles cito aqui os amigos de longa data: Marcela, Fernanda, Felipe e Jho. E também os amigos que conheci no Rio, pessoas incríveis que já tenho o prazer de chamar de amigos: Em especial a Val (que se tornou mãemiga! Que

me acolhe, briga comigo quando precisa, que ri de doer a barriga, que festa, pula e tem uma energia incrível). A Carol, Tati, Fernanda, Paula, Catão, Lu e Márcia pelos momentos de alegria e reflexões.

E a Dona Lúcia, que é alguém que eu admiro muito, e que gentilmente me acolheu. Em uma ocasião, perguntei a ela qual era o "segredo da juventude", já que, aos quase 90 anos, tem mais pique do que eu. E ela me respondeu: É ser otimista e Deus (e esse também é só mais um exemplo do quanto eu aprendo com ela).

Agradeço a Mônica pelo carinho e pela ótima companhia nas coletas. Por permitir a mim e aos mosquitinhos fazerem parte do seu diário de expedições.

A Elisa, por me mostrar a maior e mais linda estrela do mar que já vi.

Agradeço aos amigos do laboratório do Rio. Em especial ao Guilherme e ao Edu pelos ensinamentos sobre montagens de genomas e filogenias, e por toda ajuda e apoio. Agradeço também ao Thyago, Fabiana, Felipe, Anna, Julia e Isa por tamanho companheirismo.

Ao pessoal do Rona Pitaluga Lab pela parceria e por toda ajuda nas coletas. Obrigada André Pitaluga, Akira, Nati Valério, Nati Lourenço de Oliveira, Iara, Sabrina, João, Lucas, Larissa.

Obrigada prof Carlos, pela ajuda com a identificação das genitálias, pelos ensinamentos e por todas as risadas.

Também acho justo agradecer as centenas de mosquitinhos que foram coletados para esta tese, sem eles ela também não teria sido realizada.

Agradeço a banca pela disponibilidade de contribuir com meu trabalho. Fico muito feliz com cada membro escolhido para este momento tão importante.

Agradeço a UFSC, UFRJ, PPGBCD, LAMEB, LAPOGE, CAPES, CNPq, Welcome Trust e Royal Society.

Meu mais sincero muito obrigada a cada um que contribuiu direta ou indiretamente para este trabalho ou para minha jornada acadêmica e pessoal.

Sumário

Introdução Geral.....	11
Malária	11
Malária no Brasil	13
Gênero <i>Anopheles</i>	15
Breve histórico da descoberta de <i>An. cruzii</i>	17
Subgênero <i>Kerteszia</i>	20
Complexo de Espécies	21
<i>Anopheles (Kerteszia) bellator</i>	24
<i>Anopheles (Kerteszia) cruzii</i>	26
Por que estudar complexos de espécies crípticas?	33
Capítulo I - Estudo genômico da especiação críptica em <i>Anopheles (Kerteszia)</i> <i>cruzii</i> , principal vetor da malária na Mata Atlântica brasileira.....	36
Resumo	40
Abstract	42
Objetivos.....	44
Objetivo Geral	44
Objetivos específicos	44
Materiais e métodos	45
Locais de coleta	45
Isolamento de DNA.....	46
Características morfológicas da genitália masculina.....	47
Identificação Bocaina Grupo 1 e Bocaina Grupo 2	47
Sequenciamento, montagem do genoma e remoção de contaminantes.....	50
Anotação gênica e busca por ortólogos	53
Inferências filogenômicas	53
Análises de F_{ST}	55
Datação.....	55
Resultados.....	56
Montagem e qualidade dos genomas.....	56
Diferenciação de Bocaina Grupo 1 e Bocaina Grupo 2	60
Diferenciação Genética (F_{ST}).....	61
Inferências filogenômicas	66

.....	67
Tempos de Divergência	68
Morfologia da genitália masculina	69
Discussão	70
Montagem e qualidade de genomas	70
<i>Anopheles (Kerteszia) cruzii</i> s.l.	72
<i>Anopheles (Kerteszia) laneanus</i> :	75
<i>Kerteszia</i> San M1: Uma espécie complexa ou parte de um complexo de espécies?	76
<i>Anopheles (Kerteszia) bellator</i> e <i>Anopheles (Kerteszia) homunculus</i> :	79
Tempos de divergência e hipóteses para especiação	81
Conclusões.....	85
Perspectivas.....	85
Material Suplementar.....	86
Capítulo II - Novas evidências moleculares da estrutura populacional de <i>Anopheles (Kerteszia) bellator</i> na Mata Atlântica brasileira	123
Resumo.....	125
Abstract.....	126
Objetivos.....	127
Objetivo Geral	127
Objetivo Específicos	127
Materiais e métodos	128
Locais de coleta	128
Dados moleculares	129
Análise das sequências	130
Resultados.....	131
Discussão	135
Dados suplementares	138
Referências	147
Apêndice 1.....	169
Apêndice 2.....	205

Introdução Geral

Malária

A malária é uma doença parasitária, endêmica em 84 países. Nos dias atuais, esta patologia ainda é um grande problema de saúde pública, acometendo milhões de pessoas todos os anos. Segundo o último relatório da Organização Mundial de Saúde, no ano de 2021 ocorreram 247 milhões de casos de malária e 619.000 mortes globalmente em decorrência da doença (WHO, 2022). Além disso, estimou-se que quase metade da população global esteve sob risco de contrair malária no mesmo ano (HAY et al., 2004).

Esta é uma doença que afeta a humanidade desde tempos remotos, e há registros de febres intermitentes e calafrios semelhantes aos seus sintomas desde povos antigos. Um exemplo disso é a descrição desses sintomas em papiros egípcios datados de 1570 a.C. (COX, 2010).

A malária também foi citada em obras como A Ilíada, escrita por Homero por volta de 750 a.C., onde sintomas como febre relacionados à doença são descritos. Além disso, importantes filósofos como Aristóteles (384-322 a.C.) e Platão (428-347 a.C.) também fizeram referências à malária em seus trabalhos (ARROW; PANOSIAN; GELBAND, 2004).

Diante de seus impactos para a saúde pública, esta tem sido uma patologia importante ao longo da história. Começou a ser melhor compreendida com os estudos de Charles Louis Alphonse Laveran, que em 1880 descreveu o protozoário do gênero *Plasmodium* como seu causador (STANLEY et al, 1991).

A introdução dos protozoários deste gênero nas Américas, ainda é uma questão controversa. Rodrigues et al., (2018) utilizando mitogenomas elucidaram melhor essa questão na América do Sul. Estes autores demonstraram que *Plasmodium vivax* foi introduzido na América do Sul ao longo das primeiras migrações humanas, provenientes do sul da Ásia e da Oceania, entre 25 mil e 10 mil anos atrás. Em contrapartida, *Plasmodium falciparum* encontrados nos dias atuais na América do Sul parecem ser originários diretamente do continente africano.

Atualmente, sabe-se que diferentes espécies de *Plasmodium* são

predominantemente relacionadas a malária humana. São elas: *P. falciparum*, *P. vivax*, *Plasmodium ovale*, *Plasmodium malariae*, e *Plasmodium knowlesi* (MILNER, 2018). Destas, as que ocasionam o maior número de casos são *P. vivax* e *P. falciparum* (ASHLEY; PYAE PHYO; WOODROW, 2018).

Para que ocorra efetivamente a transmissão da doença é necessário que o parasita da malária complete com êxito todo o seu ciclo de vida. Este ciclo tem início quando o mosquito se alimenta de sangue contendo gametócitos de *Plasmodium*. Os gametócitos se desenvolvem no intestino médio do mosquito e sofrem fecundação para formar os zigotos. Estes, dão origem a oocinetos, que são formas móveis que atravessam o epitélio do intestino médio. Então o oocineto transforma-se em oocisto e sofre ciclos de esporogonia, resultando em esporozoítos. Em seguida, essas formas alcançam a hemocele e as glândulas salivares do vetor. Desta forma, o mosquito carregando as formas infectantes do parasita é capaz de transmiti-las para os humanos e outros animais quando realiza a alimentação sanguínea (LEVASHINA, 2004; CLAYTON; DONG; DIMOPOULOS, 2014). E é então por meio da picada, que os esporozítos invadem a corrente sanguínea e então atingem os hepatócitos, onde sofrem multiplicação assexuada resultando nos merozoítos. Estes merozoítos invadem os eritrócitos, onde sofrem nova etapa de multiplicação. Esse processo ocorre de forma contínua e está relacionado as manifestações clínicas da malária, como febres e calafrios. Alguns merozoítos passam por um processo de desenvolvimento e se transformam em gametócitos machos ou fêmeas. Os gametócitos permanecem no sangue até serem ingeridos por uma fêmea de mosquito anofelino, quando esta se alimenta de sangue infectado (COX, 2010). Desta forma, *Plasmodium* é transmitido por meio da picada de diferentes espécies de mosquitos do gênero *Anopheles*, família Culicidae e subfamília Anophelinae (ASHLEY; PYAE PHYO; WOODROW, 2018).

O fato de que este parasita é transmitido por mosquitos aos humanos, foi elucidado apenas em 1898, por diferentes malariologistas: Camillo Golgi, Amico Bignami, Giovanni Battista Grassi, Giuseppe Bastianelli, Angelo Celli e Ettore Marchiafava (COX, 2010).

Além dos humanos a malária pode acometer também outros animais, como aves, répteis, roedores e primatas não humanos (SMITH; STYCZYNSKI, 2018). E ainda, um aspecto de suma importância, é que esta também pode

ocorrer como uma zoonose. Este ciclo envolve a tríplice primata não humano-humano- mosquito, e nele humanos podem ser infectados com malárias símias, ocasionada por protozoários como: *P. cynomolg*, *P. knowlesi*, *P. brasilianum* e *P. simium* (HANG ET AL, 2021).

Atualmente existem diversas estratégias que têm como objetivo evitar a transmissão do *Plasmodium* e diminuir o número de casos da doença. A maioria delas se centrada no controle do vetor, que pode envolver a aplicação de inseticidas ou a introdução de predadores biológicos. Além disso, medidas individuais também podem ser adotadas, como o uso de mosquiteiros, telas em portas e janelas e uso de roupas longas (WHO, 2023). Recentemente houve um grande avanço na prevenção da malária, com o desenvolvimento de uma vacina. Em 2021, a OMS passou a recomendar aplicação da vacina RTS,S/AS01 contra a malária em crianças que habitam regiões em que ocorre alta transmissão por *P. falciparum*. Estudos demonstraram que a vacina reduziu significativamente a incidência de malária e a sua forma grave em crianças, sendo extremamente promissora como uma ferramenta de saúde pública (LAURENS, 2019; WHO, 2023).

Todavia, cabe ainda ressaltar aqui que esta vacina é eficiente apenas contra *P. falciparum*, não prevenindo as infecções causadas por *P. vivax* (a espécie responsável pela imensa maioria dos casos no Brasil) (OLIVEIRA-FERREIRA et al., 2010). Desta forma, mais estudos de ciência básica envolvendo vetores e parasita se fazem necessários para a criação de uma vacina que previna casos ocasionados por *P. vivax*.

Malária no Brasil

No Brasil apenas em 2021 ocorreram aproximadamente 140 mil casos de malária, a grande maioria autóctones, onde 23.408 foram ocasionados por *P. falciparum* e malária mista e 114.449 foram de malária por *P. vivax* e outras espécies parasitárias (MS, 2022).

No país, os primeiros relatos de malária foram reportados em 1587 em que foram descritas febres que ocorriam em ciclos de 48 e 72 horas, acometendo indígenas (DEANE, 1986). Todavia, análises moleculares apontam que o primeiro caso de malária por *P. falciparum* pode ter ocorrido anterior a isso,

possivelmente em 1560 decorrente do comércio de escravos africanos (YALCINDAG et al., 2012; GRIFFING et al., 2015).

Após o término do período colonial, alguns fatores foram determinantes para o aumento do número de casos, incluindo a expansão da indústria da borracha e a construção de ferrovias (DEANE, 1986). Durante a década de 1870, um grande número de migrantes procurou emprego na indústria da borracha na região amazônica, a fim de escapar da seca que assolava o nordeste brasileiro. Nesse contexto, muitos desses migrantes foram acometidos pela doença. (GRIFFING et al., 2015).

No início o século XX houve uma outra grande epidemia amazônica de malária, em decorrência da migração de pessoas para construção da Estrada de Ferro Madeira-Mamoré (CAMARGO, 2003). A segunda era da malária no Brasil deu-se com a introdução do mosquito *Anopheles gambiae*, em território nacional. Este evento ocasionou altas taxas de óbito em locais como Ceará e Rio Grande do Norte (GRIFFING et al., 2015).

Anterior a isso, os principais transmissores de malária eram vetores locais. Na região amazônica, por exemplo, o *Anopheles darlingi* era identificado como um vetor importante. Davis (1931) comprovou que em 1931, o *An. darlingi* era o responsável pela transmissão da malária em Belém. Além disso, nas áreas cobertas pela Mata Atlântica, no final do século XIX, os *Kerteszia* eram os principais causadores de transmissão do *Plasmodium* (GADELHA, 1994). Sendo a malária endêmica em diferentes locais, como São Paulo, Paraná, Santa Catarina e Rio Grande do Sul. E nestas regiões, os principais vetores eram os mosquitos *An. cruzii* e *An. bellator* (RACHOU, 1958).

Durante a década de 1940, o Serviço Nacional de Malária e o Regime de Combate à Malária foram estabelecidos em todo o país, o que resultou no controle da doença em grande parte do território brasileiro. Por meio de diversas medidas, a incidência da malária no país diminuiu consideravelmente, passando de cerca de 5 milhões de casos antes das campanhas para apenas 52 mil casos confirmados em 1970 (BRASIL, 1941, 1985; BARATA, 1988) .

Atualmente, a grande maioria dos casos ocorre na região Amazônica. Todavia, também são registrados casos extra-amazônicos, em estados como: Bahia, São Paulo, Rio de Janeiro, Santa Catarina, entre outros (totalizando 510 casos em regiões não amazônicas, apenas em 2021) (MS, 2022). Estes casos

apesar de corresponderem a uma menor porcentagem, merecem atenção e podem ter consequências importantes, como ocasionar surtos em locais em que a transmissão é baixa ou ainda acarretar em reintrodução da endemia em áreas em que a malária não ocorre atualmente (AHMED et al., 2020; DE LEMOS et al., 2021; GARCIA et al., 2022). Ao mesmo tempo, nessas regiões extra-amazônicas a taxa de letalidade devido a malária é alarmantemente maior quando comparada aos casos amazônicos, em decorrência de erros ou atraso no diagnóstico dos pacientes. Estudos mostram que em 2019 a taxa de mortalidade foi 123 vezes maior na região extra-amazônica quando comparada a região amazônica (DE PINA-COSTA et al., 2014; GARCIA et al., 2022).

Assim, apesar de casos extra-amazônicos corresponderem a uma pequena parcela dos casos no país, estes podem ser mais mortais e serem responsáveis pela reintrodução de casos em áreas livres de malária. Esses fatos ressaltam ainda mais a importância de estudos para melhor compreensão de diagnóstico, transmissão, biologia, genética e evolução dos vetores em regiões extra-amazônicas.

Estes casos que ocorrem fora da região amazônica, acometem principalmente locais cobertos por Mata Atlântica, região com abundância de plantas da família bromeliaceae, criadouros dos mosquitos do subgênero *Kerteszia*, os principais vetores dos casos autóctones que ocorrem ao longo da costa sudeste do Atlântico (BUERY et al., 2021).

Gênero *Anopheles*

O gênero *Anopheles*, pertencente à família *Culicidae* e subfamília *Anophelinae*, foi descrito pelo entomologista alemão Johann Wilhelm Meigen em 1818 (HARBACH, 2013). No Brasil, esses mosquitos são comumente conhecidos por diversos nomes populares, tais como: carapanãs, muriçocas e mosquitos-prego. Possuem tamanho aproximado de um centímetro de comprimento ou envergadura, além de pernas longas e corpo delgado, o que também lhes confere a designação de pernilongos (BRASIL, 2005).

Os Anofelinos são facilmente distinguíveis de mosquitos de outros gêneros a olho nu, devido à sua postura em repouso e ao se alimentarem, quando pousam com o corpo e a probóscide em linha reta, quase formando um

ângulo reto com o substrato. Eles geralmente têm hábitos preferencialmente crepusculares e noturnos, com a hematofagia ocorrendo durante o horário crepuscular, ao entardecer. As larvas de *Anopheles* não possuem sifão e por este motivo, também são prontamente identificáveis por sua posição de repouso, que é paralela à superfície da água (CONSOLI; LOURENÇO DE OLIVEIRA, 1994), em contraste com a maioria dos outros mosquitos, cujas larvas ficam em posição aproximadamente perpendicular em relação a superfície da água.

Os anofelinos são cosmopolitas e tem distribuição em todos os continentes, com exceção da Antártica. Este gênero consta com aproximadamente 500 espécies descritas, das quais aproximadamente 70 são potencialmente capazes de transmitir a malária (KRZYWINSKI; BESANSKY, 2003; SINKA et al., 2012; HARBACH, 2018).

Os mosquitos do gênero *Anopheles* são insetos com metamorfose completa, com quatro estágios de desenvolvimento: ovo, larva, pupa e adulto. Possuem uma fase aquática (que engloba diferentes instares larvares: L0, L1, L2, L3 e L4), e outra terrestre com o mosquito adulto. Os ovos são postos na água e possuem flutuadores laterais que permitem seu deslocamento para as margens. O tempo de eclosão dos ovos depende da temperatura: em torno de 30 oC, os ovos eclodem em dois a três dias. Em contrapartida a cerca de 16 oC, por volta de sete a quatorze dias. As larvas precisam de cinco a dez dias para se desenvolverem. A fase de pupa pode durar de dois a três dias. Todavia, esses períodos variam entre as diferentes espécies. Os adultos de ambos os sexos se alimentam de néctar e frutos, sendo que somente a fêmea que realiza hematofagia, que é essencial para a ovoposição (FORATTINI, O.P., ISHIATA, G.K., RABELLO, E.X. & COTRIM, 1973; CONSOLI; LOURENÇO DE OLIVEIRA, 1994; FORATTINI, 2002).

O gênero *Anopheles* compreende oito subgêneros: *Anopheles*, *Baimaia*, *Cellia*, *Christya*, *Kerteszia*, *Lophopodomys*, *Nyssorhynchus* e *Stethomyia* (HARBACH, 2008). Dentre estes, merecem destaque *Nyssorhynchus* e *Kerteszia* pois, são os principais responsáveis pelos casos de malária no país. *Nyssorhynchus* é o principal transmissor de malária na região amazônica e *Kerteszia*, é o responsável pelos casos da chamada malária-bromélia, na região de Mata Atlântica (CONSOLI; LOURENÇO DE OLIVEIRA, 1994).

A descoberta de casos de malária em regiões de Mata Atlântica e que estes eram causados por um *Kerteszia* foi bastante interessante. Pois, grande parte das espécies de *Anopheles* em suas formas imaturas, costumam habitar água parada e limpa. Como por exemplo em represas, poços, lagoas, depressões de terrenos e em valas (RACHOU, 1958; ARCOS et al., 2018). Por este motivo, foi uma grande surpresa quando fora detectada malária em região de Mata Atlântica, no local em que estava sendo duplicada a estrada de ferro entre São Paulo e Santos. Pois, estes casos ocorriam tanto no alto da Serra, como nas vertentes íngremes, destituídas de pântanos. Desta forma, esta era uma malária diferente pois, seus vetores não tinham em suas formas larvais habitats comuns. Foi então que Lutz médico e pesquisador brasileiro, em 1903, desvendou esse mistério. Essa malária, (hoje chamada de bromélia-malária) era causada por um *Kerteszia* que em sua fase larval habita a água que se acumula no interior de bromélias. Mais detalhes sobre essa interessante descoberta serão abordados no tópico a seguir (LUTZ, 1903; BENCHIMOL; SÁ, 2005).

Baseado em análises filogenéticas de genes codificadores de proteínas mitocondriais, Foster (2017) e colaboradores propuseram elevar *Kerteszia*, *Lophopodomyia*, *Nyssorhynchus* e *Stethomyia* a gênero. Entretanto, autores como Harbach (2018) discutem que isso deve ser visto com cautela, apesar de o trabalho mostrar resultados notáveis. Harbach, relata que essa alteração de status pode ser irrelevante uma vez que, entre outros motivos, estes grupos são totalmente diagnosticáveis. Este autor, ainda acrescenta que se reportar a *Kerteszia* como gênero ou subgênero pode ser considerada uma decisão subjetiva e questão de preferência. Portanto neste trabalho, continuaremos a nos referir à *Kerteszia* como subgênero.

Breve histórico da descoberta de *An. cruzii*

Anopheles cruzii foi descrito em 1901, recebendo inicialmente o nome de *Anopheles lutzii* Theobald 1901. No entanto, este nome já havia sido previamente utilizado para outra espécie. Então posteriormente, o nome da espécie foi substituído por *An. cruzii*, por Dyar & Knab (WRBU, 2023). Nos parágrafos subsequentes serão abordados estes aspectos em maior

detalhamento.

No final do século XIX Adolpho Lutz, coletou os primeiros espécimes de mosquitos que atualmente conhecemos como *An. cruzii*. Nesta época o médico e cientista brasileiro Adolpho Lutz foi chamado ao local onde estava sendo duplicada a estrada de ferro entre São Paulo e Santos. Neste local uma doença febril estava acometendo grande parte dos operários (LUTZ, 1903; BENCHIMOL; SÁ, 2005). Como o próprio Lutz relatou:

“Na construção da linha nova foram empregados milhares de operários, alojados em ranchos localizados em plena mata e comunicando-se uns com os outros por meio de picadas provisórias, abertas na ocasião. Não tardou muito em surgirem numerosos casos de febre intermitente entre os trabalhadores. Foram atingidos, de preferência, os moradores dos ranchos da zona mais baixa e mais quente, mas a moléstia alastrou-se durante a estação cálida até o alto da serra. O número de vítimas foi considerável, atingindo, muitas vezes, em poucos dias, a maioria dos moradores de um rancho. Devido ao caráter benigno da doença e ao uso abundante de quinino, a incapacidade para o trabalho era geralmente passageira. As recaídas eram muito frequentes, mesmo depois de abandonados os focos de infecção, surgindo ocasionalmente estados anêmicos e caquéticos, mas praticamente não houve casos fatais. Os engenheiros, que não dormiam nas zonas infectadas, escaparam todos, tanto quanto eu saiba, ao mal. Tive oportunidade de examinar bom número de operários doentes e pude verificar que se tratava verdadeiramente de malária e quase sempre (se não sempre), da forma benigna, comprovada pelos resultados dos exames microscópicos, mostrando plasmódios grandes e relativamente escassos” (LUTZ, 1903; BENCHIMOL; SÁ, 2005).

Conforme relatado a cima, ao constatar a presença de *Plasmodium* no sangue dos doentes, Lutz prontamente percebeu que a patologia febril que estava afetando os trabalhadores se tratava de malária. Esse fato despertou sua curiosidade e o levou a se hospedar no local para investigar mais a fundo a questão. Então ele relata que logo na primeira noite observou diversos mosquitos conhecidos seus, mas, também um que não havia nunca visto. Este que era pequeno, mas era um sugador voraz (LUTZ, 1903; BENCHIMOL; SÁ, 2005).

E foi assim que Lutz percebeu que este pequeno mosquito com atividade hematófaga crepuscular, voraz e promíscua, conhecido hoje como *An. cruzii* era o vetor da doença no local. Após a descoberta deste anofelino e confirmação de que era este o responsável pela malária em São Paulo, uma questão ainda permanecia em aberto: Aonde seria o criadouro destes mosquitos em suas formas imaturas?

Nestas regiões, brejos, poças e água parada eram incomuns. No entanto, Lutz já sabia previamente que algumas plantas tinham a propriedade de armazenar água, devido a observações anteriores feitas no Havaí. Além disso, ele sabia também que essas plantas poderiam ser habitat de crustáceos. Diante dessas informações, ele rapidamente conectou os pontos e levantou a hipótese de que a água acumulada no interior das bromélias, poderia ser um possível ambiente para o desenvolvimento das larvas do anofelino em questão. Então, ele verificou que *An. cruzii* e outros mosquitos presentes na floresta, passavam a fase larval na água das bromeliáceas (BENCHIMOL; SÁ, 2005).

Também foi Adolpho Lutz que mais tarde sugeriu que este mesmo mosquito seria o responsável pelos casos de malária em Santa Catarina, que vinham ocorrendo desde o ano de 1820. Todavia, a oposição a esta hipótese por Frederick Knab limitaria o controle desse vetor até 1940 (KNAB, 1912; GADELHA, 1994; GRIFFING et al., 2015).

Em 1899, Lutz enviou a Theobald três espécimes pertencentes à espécie que ele incriminara como responsável pela transmissão de malária aos trabalhadores da via férrea. Em reconhecimento a sua contribuição, Theobald nomeou esse anofelino como *Anopheles lutzii* em 1901 (THEOBALD, 1901).

O mosquito incriminado por Lutz foi reclassificado em 1908 por Dyar e Knab, passando a chamar-se, como o conhecemos atualmente: *An. cruzii* uma vez que, outras espécies de anofelinos já haviam sido nomeadas em

homenagem ao pesquisador, como o *Anopheles (Nyssorhynchus) lutzii* Cruz, 1901 (DYAR; KNAB, 1908; WRBU, 2023).

Subgênero *Kerteszia*

Mosquitos do subgênero *Kerteszia* tem distribuição nas Américas, desde o México até o Sul do Brasil não tendo sido relatados no Chile, Uruguai e ilhas do Caribe, com exceção de Trindade. Grande parte das espécies de *Kerteszia* tem distribuição nas regiões costeiras de Mata Atlântica, onde bromélias são observadas em abundância (ZAVORTINK, 1973; COLLUCCI; SALLUM, 2003).

Larvas e pupas da maioria das espécies de *Kerteszia* desenvolvem-se nas axilas das folhas de bromélias (com exceção de *An. bambusicolus* que como o próprio nome sugere, tem como habitat o interior de bambus). O principal caráter morfológico que caracteriza o subgênero é a presença de quatro faixas escuras dispostas longitudinalmente no escudo, também apresentam as pernas listradas em cor preta e branca (CONSOLI; LOURENÇO DE OLIVEIRA, 1994; NEVES, 2016).

Este subgênero inclui atualmente doze espécies válidas, são elas: *Anopheles auyantepuiensis*, *Anopheles bambusicolus*, *Anopheles bellator*, *Anopheles boliviensis*, *Anopheles cruzii*, *Anopheles gonzalezrinconesi*, *Anopheles homunculus*, *Anopheles laneanus*, *Anopheles lepidotus*, *Anopheles neivai*, *Anopheles pholidotus*, e *Anopheles rollai* (COLLUCCI; SALLUM, 2003).

Dentro deste subgênero as espécies *An. cruzii*, *An. bellator* e *An. homunculus*, são as mais importantes em termos de transmissão de malária nas regiões cobertas por Mata Atlântica como, nos estados do Sul e Sudeste do Brasil (MARRELLI et al., 2007).

Apesar de grande parte das espécies de *Kerteszia* serem consideradas não sinantrópicas (i.e., não estão intimamente associadas ao ambiente humano) trabalhos têm demonstrado que estas podem alimentar-se tanto no ambiente silvestre bem como adentrar em residências em busca de sangue humano sendo então consideradas antropofílicas (DEANE, 1986). Essas espécies tem maior atividade e tendem a se alimentar ao entardecer e nas primeiras horas da noite

(ZAVORTINK, 1973; CONSOLI; LOURENÇO DE OLIVEIRA, 1994; GUIMARÃES et al., 2000).

Trabalhos anteriores sugerem que algumas espécies do subgênero *Kerteszia* podem fazer parte de complexos de espécies. Esta questão será melhor exposta nos tópicos abaixo.

Complexo de Espécies

No gênero *Anopheles*, os complexos de espécies são comuns, e muitos deles incluem os principais vetores de malária do mundo. Entre as quase 500 espécies oficialmente descritas, existem aproximadamente 50 membros que fazem parte de complexos de espécies crípticas (SINKA et al., 2010, 2012). Os membros desses complexos são caracterizados por serem indistinguíveis morfológicamente. Todavia, apresentam diferenças genéticas e comportamentais que podem influenciar na transmissão da doença (DELLA TORRE et al., 2002).

Na região africana, onde ocorrem mais de 95% dos casos de malária, os principais transmissores da doença são do complexo *Anopheles gambiae*, que é o complexo de espécies mais bem estudado atualmente. Esse complexo inclui diferentes espécies crípticas descritas até o momento, dentre estas destacam-se *An. gambiae sensu stricto* e *An. arabiensis*, que são duas das espécies mais importantes em termos de transmissão da doença (SINKA et al., 2012; TENNESSEN et al., 2021; WHO, 2023).

Inicialmente, o complexo *An. gambiae sensu lato* (s.l) incluía seis espécies: *An. gambiae sensu stricto* (s.s), *An. arabiensis*, *An. bwambae*, *An. melas*, *An. merus* e *An. quadriannulatus*. No entanto, posteriormente, foram descobertas duas outras espécies, *An. comorensis* e *An. quadriannulatus B*.

Essas espécies não são distinguíveis por meio de caracteres morfológicos, todavia essa diferenciação pode ser observada avaliando-se inversões cromossômicas (COLUZZI et al., 1977; DORN et al., 2011; COETZEE et al., 2013). Essas espécies de mosquitos podem ocupar *habitats* distintos. Três delas são encontradas em ambientes de água doce: *An. gambiae s.s*, *An. arabiensis* e *An. quadriannulatus*. Já *An. melas* e *An. merus* são encontradas em ambientes de água salobra. Ademais, *An. bwambae* é uma espécie que habita

águas minerais salobras, provenientes de fontes geotermiais (COETZEE et al., 2013). Com relação a *An. gambiae* s.s, este já foi identificado como sendo composto por duas formas moleculares distintas, denominadas M e S. Em um estudo realizado por Turner et al (2005), foram identificadas regiões genômicas que diferenciam essas formas moleculares, sendo localizadas no cromossomo X e as outras duas no cromossomo 2. Nessas regiões analisadas, foram encontradas diferenças fixas entre as formas moleculares. Mais recentemente, a forma M foi reclassificada como *Anopheles coluzzii*, enquanto a forma S permaneceu com o nome original de *An. gambiae*, e ainda, *An. quadriannulatus* B passou a se chamar *An. amharicus*. Desta forma, agora englobam o complexo *An. gambiae* s.l as seguintes espécies: *An. arabiensis*, *An. bwambae*, *An. melas*, *An. merus*, *An. quadriannulatus*, *An. gambiae* s.s., *An. coluzzii* e *An. amharicus* e *An. fontenillei* (COETZEE et al., 2013; BARRÓN et al., 2019).

O complexo *An. gambiae* s.s. engloba diferentes formas cromossômicas. Essas formas cromossômicas também tendem a ocupar diferentes nichos. (DELLA TORRE et al., 2002). Acredita-se que as inversões cromossômicas contribuam para a capacidade de explorar diferentes nichos ecológicos. Isso pode ser observado por meio da distribuição das frequências de inversão ao longo das áreas geográficas indicando que algumas das inversões podem ser produto da seleção que permite as diferentes espécies de *An. gambiae*, s.s, sobreviver e explorar uma grande variedade de nichos (TOURÉ et al., 1998; COLUZZI et al., 2002; LEE et al., 2009).

Essas espécies do complexo podem diferir em ocupação de habitats, distribuição geográfica e temporal (DORN et al., 2011), e provavelmente em um dos aspectos mais importantes: a preferência alimentar. Grande parte das espécies crípticas do complexo *An. gambiae* s.l., podem ser vetoras de malária humana sendo elas altamente antropofílicas, com exceção de *An. quadriannulatus* que é zoofílica, se alimentando principalmente de sangue bovino. E este fato é de suma importância pois pode interferir na transmissão da doença, já que essa espécie diferente das outras do complexo não tem implicância na transmissão da doença (COLUZZI et al., 1979).

O complexo *An.gambiae* s.l é amplamente estudado, devido à sua grande importância epidemiológica (como comentado acima), mas existem outros complexos de espécies dentro do gênero *Anopheles* que também merecem

destaque. Um exemplo é o complexo *Maculipennis*, que foi um dos primeiros complexos descritos para o gênero *Anopheles*. *Maculipennis* foi descoberto ainda durante a década seguinte à Primeira Guerra Mundial, sendo apontado como um complexo de espécies irmãs, incluindo diferentes espécies crípticas (VAN THIEL, 1927; LINTON et al., 2003).

A compreensão do complexo de espécies de *An. maculipennis* permitiu desvendar o mistério epidemiológico que ocorreu na Europa nos anos vinte. Naquela época, este mosquito foi identificado como o vetor de malária na Europa, já que o retorno de soldados infectados durante a Primeira Guerra Mundial levou a surtos locais em diversas partes do continente. No entanto, em alguns locais com altas densidades deste anofelino, não havia transmissão de malária, o que ficou conhecido como “anofelismo sem malária” (HACKETT, 1934, 1937; LINTON et al., 2003). Foi demonstrado que algumas espécies deste complexo, não são boas transmissoras de malária. Como é o caso *An. messeae*, que é descrita como sendo altamente zoofílica, alimentando-se por exemplo, de gado e cães (TAKKEN et al., 2002; BERTOLA et al., 2022).

No Brasil, há também uma variedade de complexos de *Anopheles*, como *Anopheles albitarsis*, que é composto por pelo menos quatro espécies distintas: *Anopheles albitarsis* s.s., *Anopheles deaneorum*, *Anopheles marajoara* e *Anopheles albitarsis* sp. B. Além disso, existe o complexo *Anopheles triannulatus*, que engloba *An. triannulatus* s.s, *An. halophylus* e *An. triannulatus* C, e o complexo *Anopheles oswaldoi*, que inclui *Anopheles oswaldoi* s.s, *Anopheles oswaldoi* A, *Anopheles oswaldoi* B e *Anopheles Konderi*, entre outros (CONN et al., 2002; SILVA-DO-NASCIMENTO et al., 2011; RUIZ-LOPEZ et al., 2013).

Todos os exemplos citados acima pertencem ao subgênero *Nyssorhynchus*. Todavia, também existem complexos brasileiros em outros subgêneros como é o caso de *Kerteszia*, o objeto de estudo deste trabalho.

No início dos anos 1940 até os anos 1970, *An. cruzii* s.l., *An. bellator* e *An. homunculus*, que ocorrem em simpatria ao longo na Mata Atlântica, foram os principais vetores da "malária das bromélias" neste bioma, causando significativo impacto na saúde pública (RACHOU, 1958; DEANE, 1988; MARRELLI et al., 2007). Mas apesar da importância epidemiológica dessas espécies, os dados disponíveis a respeito delas são insuficientes. Todavia,

existem alguns trabalhos utilizando inversões cromossômicas, marcadores moleculares e isoenzimas, que sugeriram que *An. bellator* e *An. cruzii* podem fazer parte de complexos de espécies crípticas (BRANQUINHO et al., 1997; RAMIREZ; DESSEN, 2000a; CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2003, 2004; RONA et al., 2009, 2010; RONA; CARVALHO-PINTO; PEIXOTO, 2010; RONA; CARVALHO-PINTO, 2013; DE REZENDE DIAS et al., 2018; VOGES et al., 2019). Este assunto será detalhado nas seções subsequentes.

Anopheles (Kerteszia) bellator

Anopheles bellator é predominantemente encontrado no litoral, nas regiões costeiras do leste da Venezuela ao sul do Brasil e em Trinidad e Tobago (ARAGÃO, 1964; MARRELLI et al., 2007).

Embora seja raro nas matas, este é predominante encontrado em locais abertos, tendo preferência por locais com maior exposição solar e sujeitos a maior taxa de evaporação. Desta forma, podem ser observados em bromélias de grande porte com maior acúmulo de água. Estas podem ser bromélias rupestres, epífitas ou terrestres. As fêmeas picam ao anoitecer, com maior atividade hematofágica no início da noite (CONSOLI; LOURENÇO DE OLIVEIRA, 1994).

Em 1942, este anofelino foi incriminado como o responsável pelos casos de malária em Trinidad, afetando os trabalhadores das plantações de cacau. Downs e Pittendrigh (1946), conduziram um estudo sobre a ecologia desses vetores e seu papel na transmissão da malária em áreas de plantação de cacau na ilha de Trinidad. Durante a pesquisa, eles constataram que neste local essa espécie era pouco presente dentro de residências. Diante dessas constatações, era plausível considerar que a melhor estratégia para controlar o vetor seria a remoção manual das bromélias e a aplicação de produtos químicos.

Em Santa Catarina, na mesma época, *An. bellator* também foi responsável por casos de malária. No entanto, a diferença fundamental entre o problema da malária em Santa Catarina e em Trinidad era o padrão de transmissão. Enquanto em Trinidad a transmissão ocorria em regiões de mata, no sul do Brasil, o problema também tinha dimensão urbana, com os mosquitos saindo da mata e entrando nas residências (RACHOU, 1958). Além disso, a

malária afetava uma área e uma população muito maiores, abrangendo não apenas Santa Catarina, mas também outros estados como Rio Grande do Sul e Paraná.

Para combater a malária no Sul do Brasil, foram adotadas algumas técnicas além da remoção das bromélias, incluindo o desmatamento em regiões próximas a locais com alto número de casos, a aplicação de DDT e o uso de medicamentos específicos contra a doença. Graças a essas medidas, em 1951, a situação da malária em Santa Catarina estava mais controlada. Em Florianópolis, por exemplo, não foram registrados nenhum caso nesse ano, ao contrário de 1944, quando foram registrados 112 casos (SMITH, 1952).

Um dos estudos pioneiros sobre esta espécie, avaliava seus cromossomos e foi executado por Ramírez e Dessen (1996). Neste trabalho, compararam os cromossomos de *An. cruzii* e *An. bellator*. Estas autoras identificaram várias regiões de homologia entre esses cromossomos, além de constatarem que *An. cruzii* exibe um grau de polimorfismo cromossômico maior do que *An. bellator*. Especificamente, a média de inversões por indivíduo em *An. bellator* foi de 0,63, enquanto em *An. cruzii* foi de 4,15.

Ademais, alguns trabalhos sugerem que este Anofelino possa fazer parte de um complexo de espécies crípticas: Carvalho-Pinto e Lourenço-de-Oliveira (2003) realizaram estudos utilizando métodos isoenzimáticos em populações de *An. bellator* provenientes dos estados brasileiros de Santa Catarina, São Paulo e Bahia, bem como de uma população da Ilha de Trinidad (República de Trinidad e Tobago). De acordo com os resultados, o fluxo gênico entre as populações brasileiras com a população de Trinidad é relativamente baixo, o que sugere um possível processo de especiação. Os autores também observaram que as populações de *An. bellator* do Sul e Sudeste do Brasil são geneticamente mais próximas entre si quando comparadas com a população do Nordeste.

Mais recentemente, Voges et al. (2019) realizaram estudos utilizando genes do relógio biológico como marcadores moleculares e observaram forte estruturação populacional em populações brasileiras de *An. bellator*. Os pesquisadores identificaram dois grupos distintos na Mata Atlântica, denominados *An. bellator A* e *B*. O grupo *An. bellator A* é encontrado nas regiões Sul e Sudeste do Brasil, enquanto o grupo *An. bellator B* ocorre no Nordeste. Em

conjunto, essas observações sugerem que *An. bellator* pode ser um complexo de espécies crípticas.

Anopheles (Kerteszia) cruzii

An. cruzii ocorre desde o estado do Rio Grande do Sul até Sergipe, habitando áreas de Mata Atlântica, tanto em regiões costeiras quanto em locais montanhosos (ZAVORTINK, 1973; MARRELLI et al., 2007). Durante seu estágio imaturo (que compreende as fases larval e de pupa) habita a água acumulada no interior de bromélias, tanto em copas de árvores quanto próximas ao solo. Essas bromélias podem conter desde 5 ml de água até mais de dois litros (VELOSO; MOURA; KLEIN, 1956; CONSOLI; LOURENÇO DE OLIVEIRA, 1994).

Os mosquitos adultos desta espécie sobrevivem em torno de 35 a 56 dias na natureza (FERREIRA et al., 1969). Além disso, alguns estudos demonstram que em condições laboratoriais, o ciclo que vai do ovo até o estágio adulto dura aproximadamente 35 dias (WILKERSON; PEYTON, 1991; CHAHAD - EHLERS; LOZOVEI; MARQUES, 2007). Também em ambiente laboratorial fêmeas ovipositam em torno de 30 ovos por postura. Estes medem cerca de 0,5mm de comprimento e em sua estrutura possuem flutuadores, que fazem com que estes flutuem na superfície da água (WILKERSON; PEYTON, 1991; CALADO; NAVARRO-SILVA, 2005; CHAHAD - EHLERS; LOZOVEI; MARQUES, 2007).

Para oviposição é necessária a alimentação sanguínea, e mosquitos fêmeas de *An. cruzii* tendem a alimentar-se preferencialmente durante o período crepuscular, realizando hematofagia no crepúsculo vespertino e matutino. Esses anofelinos são antropofílicos, todavia, podem ser bastante promíscuos quanto a preferências alimentares tendo fontes de alimentação sanguínea diversas, tais como sangue de aves, humanos, e outros mamíferos. Essa espécie é exofílica, sendo encontrada no ambiente silvestre entretanto, pode adentrar residências em busca de sangue humano (FORATTINI; LOPES; RABELLO, 1968; DEANE, 1986; CONSOLI; LOURENÇO DE OLIVEIRA, 1994; GUIMARÃES et al., 2000).

Ademais, *An. cruzii* apresenta uma forte tendência acrodendrofílica, ou seja, tem a propensão de viver e se alimentar com maior frequência no dossel

das árvores. Porém, há registros de mobilidade entre o solo e a copa das árvores realizando hematofagia em níveis mais elevados de florestas e próximo ao solo (DEANE, 1986; UENO; FORATTINI; KAKITANI, 2007).

Em virtude desse comportamento alimentar promíscuo este é o único anofelino já encontrado naturalmente infectado por protozoário humanos e simianos das espécies *P. falciparum*, *P. vivax*, *P. malarie*, *P. simium* e *P. brasilianum*. Desta forma, este anofelino é responsável pela transmissão tanto de malária humana como de malária simiana nas regiões de Mata Atlântica ou próximo a ela (DEANE; DEANE; FERREIRA NETO, 1966; DEANE et al., 1970, 1971; BRANQUINHO et al., 1997; GUIMARÃES et al., 2012; DUARTE et al., 2013; NEVES et al., 2013; LAPORTA et al., 2015; LORENZ; PATANÉ; SUESDEK, 2015; CARLOS et al., 2019).

Diversos estudos sugeriram que *An. cruzii* é um complexo de espécies crípticas (Tabela 1). Zavortink por exemplo, já em (1973) observou diferenças morfológicas entre larvas do Rio de Janeiro e de Brusque (SC) e questionou sobre a coespecificidade destes espécimes.

Em 2000 Ramirez e Dessen notaram diferenças nos padrões de inversão dos cromossomos X e 3L entre *An. cruzii* do Sul e Sudeste do Brasil, e também indicaram a existência de unidades evolutivas independentes para esta espécie (RAMIREZ; DESSEN, 2000a, 2000b).

Além disso, análises de isoenzimas de populações dos estados de Santa Catarina, São Paulo, Rio de Janeiro e Bahia sugeriram que esta última (do Nordeste do Brasil) é geneticamente distante das outras populações do Sul e Sudeste do Brasil (CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2004).

MALAFRONTA et al. (2007) constataram diferenciação das sequências de ITS2 entre *An. cruzii* capturados em diferentes locais, incluindo Jucituba e Cananéia. Todavia, os autores destacam que as informações descritas no artigo, não seriam suficientes por si só para determinar a ocorrência de um complexo de espécies.

Mais recentemente, estudos usando uma abordagem multilocus (i) sugeriram que *An. cruzii* é um complexo de espécies que consiste em um grupo principal com ampla distribuição costeira do Sul ao Sudeste do Brasil, (ii) reforçaram que a população da Bahia é uma espécie diferente dentro desse complexo e (iii) sugeriram a existência de outra espécie irmã localizada nas

montanhas do Sudeste do Brasil (RONA et al., 2009, 2010; RONA; CARVALHO-PINTO; PEIXOTO, 2010; RONA; CARVALHO-PINTO, 2013; DE REZENDE DIAS et al., 2018). O trabalho de Dias (2018) utilizando dois genes sugere que, nessas montanhas, existem dois grupos geneticamente distintos. O primeiro, estes autores denominaram Grupo 1, que é encontrado em regiões no Sul do Brasil e nas regiões costeiras da Serra do Mar (Florianópolis-SC, Guapimirim-RJ, Tinguá-RJ, and Sana-RJ). Já o Grupo 2, ocorre na parte continental da Serra do Mar (em Bocaina-SP) e da Serra da Mantiqueira (Itatiaia – RJ). Curiosamente, em Bocaina, que está localizada no meio caminho entre o litoral da Serra do Mar e a Serra da Mantiqueira, ambos os grupos ocorrem em simpatria com isolamento reprodutivo parcial (DE REZENDE DIAS et al., 2018).

No presente trabalho, utilizando um número expressivamente maior de genes, observamos a existência de quatro grupos distintos para *An. cruzii*. Desta forma, para evitar confusão utilizaremos esta nomenclatura adotada por Dias (2018) apenas para nos referirmos às amostras provenientes de Bocaina (SP).

Ademais, Kirchgatter et al (2020) realizaram análises filogenéticas com genes mitocondriais, no qual obtiveram resultados similares aos de Rezende Dias (2018), em que houve agrupamento de *An. cruzii* em clados distintos: (i) Serra do Mar, (ii) Serra da Mantiqueira e (iii) Serra da Cantareira.

Esses estudos sugerem que *An. cruzii* é um complexo de espécies. Esse é o assunto principal do presente trabalho.

Tabela 1: Alguns estudos que sugerem que *An. cruzii* é um complexo de espécies crípticas.

Característica	Populações amostradas	Populações diferentes	Comentários	Referência
Larva	Rio de Janeiro (RJ), Brusque (SC)	Rio de Janeiro (RJ) e Brusque (SC).	Autor relatou que as larvas do Rio de Janeiro diferiam das de Brusque (SC) e questionou a coespecificidade.	(ZAVORTINK, 1973)
Inversões cromossômicas	São Francisco do Sul (SC), Cananéia (SP), Guaratuba (SP)	Ocorrência de duas formas divergentes na espécie, cada uma carregando um conjunto diferente de inversões, as quais foram nomeadas A e B.	Ramirez e Dessen, estudando o padrão de bandejamento cromossômico politênico de <i>An. cruzii</i> de São Paulo e Santa Catarina, observaram a existência de duas formas cromossômicas distintas e deficiência de heterozigotos.	(RAMIREZ; DESSEN, 2000a)
Inversões cromossômicas	Boracéia (SP) e Juquitiba (SP)	A forma "C" pode ocorrer em simpatria com a forma A em alguns locais.	Encontraram uma terceira forma cromossômica em duas áreas de São Paulo - pelo menos três espécies crípticas.	(RAMIREZ; DESSEN, 2000b)
Izoenzimas	Florianópolis (SC), Cananéia (SP), Nova Iguaçu (RJ), Jaguaripe (BA)	Bahia diferente do Sul e do Sudeste.	Espécimes do estado da Bahia podem ser considerados como pertencentes a uma espécie distinta das outras três populações estudadas.	(CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2004)
<i>ITS2</i>	São Francisco do Sul (SC), Cananéia (SP), Guaratuba (SP), Peruíbe (SP), Juquitiba (SP), Bertioga (SP).		Demonstraram a existência de quatro sequências ITS2 com alto grau de polimorfismo entre os <i>An. cruzii</i> capturados nos estados de SP e SC. Mas, essas informações por si só não são suficientes para determinar a ocorrência de um complexo de espécies.	(MALAFRONTTE et al., 2007)
<i>timeless</i>	Florianópolis (SC), Cananéia (SP), Juquitiba (SP), Itatiaia	Pelo menos duas espécies crípticas, uma ocorrendo no nordeste (da Bahia) e outra no sul e sudeste do Brasil.		(RONA et al., 2009)

	(RJ), Santa Teresa (ES), Itaparica (BA)			
<i>cpr</i>	Florianópolis (SC), Cananéia (SP), Juquitiba (SP), Itatiaia (RJ), Santa Teresa (ES)	A amostra de Itatiaia representa uma nova espécie irmã neste complexo.	O gene <i>cpr</i> revelou valores muito altos de FST e diferenças fixas entre Itatiaia e as outras quatro populações estudadas (Florianópolis, Cananéia, Juquitiba e Santa Teresa).	(RONA; CARVALHO-PINTO; PEIXOTO, 2010)
<i>Timeless, Clock, cycle, Rp49, RpS2, RpS29</i>	Florianópolis (SC) e Itaparica (BA)	Os resultados corroboraram a hipótese de que as populações do Sul e do Nordeste constituem duas espécies crípticas do complexo.		(RONA et al., 2010)
<i>timeless, Clock, cycle, Rp49, RpS2, RpS29</i>	Florianópolis (SC) e Itatiaia (RJ)	Florianópolis e Itatiaia representam diferentes espécies de <i>An. cruzii</i> . E Itatiaia é composta por duas espécies incipientes simpátricas A e B.		(RONA; CARVALHO-PINTO, 2013)
Genoma mitocondrial completo	São Paulo (SP), Cananéia (SP), Itatiaia (RJ), Maquiné (RS).		Espécimes de quatro populações de <i>An. cruzii</i> apresentaram diferenças na composição dos códons.	(OLIVEIRA et al., 2016)
<i>cpr e clock</i>	Bocaina (SP), Tinguá (RJ), Guapimirim (RJ), Sana (RJ), Itatiaia (RJ), Florianópolis (SC)	Duas espécies irmãs habitam altitudes maiores (acima de 900 m) da Serra do Mar e da Serra da Mantiqueira (esta última é outra cadeia montanhosa do Sudeste do Brasil).	Sugeriram a existência de espécies simpátricas na Bocaina SP.	(DE REZENDE DIAS et al., 2018)
COI cytochrome c oxidase subunit I gene; ND4 NADH dehydrogenase subunit 4 gene	Itatiaia (RJ) (Serra da Mantiqueira); Parque Estadual da Cantareira (SP) (Serra da Cantareira); Cananéia (SP), Caraguatatuba(SP),	<i>Anopheles cruzii</i> agrupou em três clados separados de acordo com as localidades de coleta: Serra do Mar, Serra da Mantiqueira e Serra da Cantareira. Confirmando <i>An. cruzii</i> como um complexo de espécies, composto por pelo menos três espécies putativas.		(KIRCHGATTER et al., 2020)

	Itanhaém (SP), e São Paulo (SP) (Todos na Serra do Mar).			
Genitália masculina	Florianópolis (SC), Guapimirim (RJ), Bocaina (SP), Itatiaia (RJ), Santa Teresa (ES)	(i) Sul/Sudeste do Brasil, aqui representado por Florianópolis-SC e Guapimirim-RJ (região costeira da Serra do Mar); (ii) Grupo Bocaina 1; (iii) Grupo Bocaina 2 (região continental da Serra do Mar, no Sudeste do Brasil) e Itatiaia - RJ (Serra da Mantiqueira, Sudeste do Brasil); (iv) Santa Teresa - ES (Sudeste do Brasil).	Observamos diferenças na genitália masculina apenas na amostra de Santa Teresa.	Este estudo
Genoma completo	Florianópolis (SC), Guapimirim (RJ), Bocaina (SP), Itatiaia (RJ), Santa Teresa (ES)	(i) Sul/Sudeste do Brasil, aqui representado por Florianópolis-SC e Guapimirim-RJ (região costeira da Serra do Mar); (ii) Grupo Bocaina 1; (iii) Grupo Bocaina 2 (região continental da Serra do Mar, no Sudeste do Brasil) e Itatiaia - RJ (Serra da Mantiqueira, Sudeste do Brasil); (iv) Santa Teresa - ES (Sudeste do Brasil).	Duas espécies simpátricas em Bocaina.	Este estudo

Genomas em anofelinos

O genoma de *An. gambiae* foi publicado na revista Science em outubro de 2002, representando um grande marco para a pesquisa de vetores de malária. Neste trabalho, os autores fizeram descobertas cruciais, identificando genes imunes envolvidos na resposta ao *Plasmodium*, além de genes associados à resistência a inseticidas, entre outros achados relevantes. Este genoma tem tamanho de 278 milhões de pares de bases e cerca de 14 mil genes (HOLT et al., 2002).

Até recentemente, não haviam muitos genomas disponíveis para outros anofelinos. Isso limitava as comparações genômicas entre estes mosquitos, tornando difícil a investigação de aspectos essenciais, como características associadas à capacidade vetorial ou resistência a inseticidas em todo o gênero. Portanto, tornou-se evidente que o sequenciamento dos genomas de várias espécies de *Anopheles* era essencial para melhor compreensão destes e outros aspectos. Diante disso, em 2015, foi realizado o sequenciamento de genomas de 16 espécies de diferentes anofelinos de diversos locais, com o propósito de investigar a base genômica da capacidade vetorial e buscar abordagens inovadoras para o controle de vetores (NEAFSEY et al., 2015). Estes genomas apresentam uma variação de tamanho de 142 Mb para *An. maculatus* até 376 Mb para a segunda versão da montagem de *An. sinensis*. A montagem mais contígua pertence a *An. albimanus*, composta por 204 scaffolds com um N50 de 18 Kb. O número de genes variou de 10,738 em *An. christyi* a 16,149 genes em *An. melas* (NEAFSEY et al., 2015).

Mais recentemente, em 2017, o consórcio genômico *The Anopheles gambiae 1000 Genomes Consortium* avaliou 765 amostras de *An. gambiae* e *An. coluzzii* de diferentes locais da África e encontraram mais de 50 milhões de polimorfismos de nucleotídeo único dentro do genoma (THE ANOPHELES GAMBIAE 1000 GENOMES CONSORTIUM, 2017). Em 2020, o mesmo consórcio avaliou a estrutura populacional de 1142 anophelinos, utilizando genomas completos. Os achados deste trabalho de 2020 forneceram novos marcadores genéticos que podem auxiliar no manejo e vigilância da resistência a inseticidas. Outro ponto importante é que os autores também puderam avaliar

melhor o fluxo gênico entre os vetores de malária africanos, o que é de suma importância pois, o fluxo gênico pode proporcionar a disseminação da resistência a inseticidas entre espécies (ANOPHELES GAMBIAE 1000 GENOMES CONSORTIUM, 2020). Em ambos os trabalhos citados acima a metodologia utilizada para sequenciamento dos genomas foi Illumina.

Atualmente, já se tem genomas long reads de anofelinos como *An. coluzzii* e *An. arabiensis*, sequenciados com Nanopore. Estes genomas têm 3 três scaffolds em escala cromossômica, correspondentes aos cromossomos X, 2, 3. Essas novas montagens genômicas possibilitaram a identificação de ponto de quebra de inversões cromossômicas em *An. coluzzii* e *An. arabiensis* (ZAMYATIN et al., 2021).

Por que estudar complexos de espécies crípticas?

A investigação e identificação de espécies crípticas desempenham papéis de extrema importância para melhor entendimento da biologia, ecologia e estratégia de vida dos membros destes complexos. São também essenciais para uma identificação mais precisa das espécies que de fato atuam como vetores de doenças, uma vez que já foi comprovado que nem todos os membros dos complexos de espécies crípticas desempenham papel na transmissão de malária, por exemplo. Desta forma em última instância, tais pesquisas também podem contribuir para o desenvolvimento de estratégias eficazes de manejo e controle de doenças relacionadas a esses vetores (DEANE, 1992; BOURKE et al., 2023).

Krzywinski & Besansky (2003) ressaltam, que o reconhecimento das diferentes espécies dentro de cada complexo pode permitir uma explicação para a endemia da malária em alguns países e a ausência dela em outros, mesmo quando o vetor comprovado da doença está normalmente presente. As espécies do complexo *An. gambiae*, por exemplo, apresentam diferenças ecológicas, que tem como consequência diferenças na capacidade de transmissão da doença (DELLA TORRE et al., 2002).

Como mencionado anteriormente, as espécies presentes nos complexos *An. cruzii* e *An. bellator* são consideradas as principais vetoras de malária em

regiões de Mata Atlântica, sendo as responsáveis por vários casos da doença que são reportados todos os anos, em diferentes estados brasileiros das regiões Sul e Sudeste (DE PINA-COSTA et al., 2014; SINAN, 2023). Cabe ressaltar ainda, que *An. cruzii* é também vetor natural de malária simiana, transmitindo *Plasmodium* de macacos para o homem, o que lhe confere ainda maior importância epidemiológica (DEANE et al., 1970, 1971).

Apesar dessa importância epidemiológica tanto de *An. cruzii* bem como de *An. bellator*, e da grande prevalência desses mosquitos em regiões cobertas pela Mata Atlântica (como nos estados de Santa Catarina, Paraná, São Paulo, Rio de Janeiro e Espírito Santo), atualmente não há uma vigilância entomológica direcionada a eles. A transmissão de malária nessas áreas evidencia a importância de acompanhar de perto essas espécies.

Desta forma, obter um melhor entendimento da estrutura genética desses complexos de espécies, como *An. cruzii* e *An. bellator*, pode ser de suma importância para identificar quais membros destes complexos atuam como vetores efetivos de malária.

Sabe-se de longa data que espécies crípticas constituem grande desafio taxonômico (BICKFORD et al., 2007). Este desafio pode afetar negativamente o manejo adequado dessas espécies e o controle dos casos da doença. Isso ocorre porque esses complexos englobam espécies distintas que podem ter comportamentos distintos na transmissão da malária, mas, morfologicamente semelhantes e que por isso são erroneamente agrupadas sob uma única designação (DELLA TORRE et al., 2002; KORSHUNOVA et al., 2019). No entanto, o advento do sequenciamento de DNA rápido e com preços mais acessíveis, atenuou este desafio. Agora é possível detectar e distinguir espécies crípticas de forma mais eficiente. Essa inovação tecnológica abriu caminho para uma abordagem mais precisa na identificação e diferenciação das espécies pertencentes a complexos (BICKFORD et al., 2007).

Neste contexto, o objetivo do presente estudo foi avaliar a estruturação genética de diferentes populações brasileiras dos complexos *An. cruzii* e *An. bellator*. Inicialmente foi realizado um estudo genético utilizando dois genes para análise da diferenciação em *An. bellator*. Em seguida, conduzimos um novo estudo que envolveu a análise de várias espécies de *Kerteszia*, incluindo *An.*

cruzii e *An. bellator*, agora com diversos genomas completos, utilizados para uma abordagem filogenômica.

Capítulo I - Estudo genômico da especiação críptica em *Anopheles (Kerteszia) cruzii*, principal vetor da malária na Mata Atlântica brasileira

Neste estudo, foi avaliada a diferenciação genética entre populações de *Anopheles (Kerteszia) cruzii* provenientes de Mata Atlântica, utilizando uma abordagem filogenômica. Este trabalho abrange o maior número de genes utilizados para avaliar a estruturação genética de *An. cruzii* s.l. e corrobora a hipótese da especiação críptica (Apêndice 1).

Lista de Tabelas

Tabela 1: Alguns estudos que sugerem que <i>An. cruzii</i> é um complexo de espécies crípticas.....	29
Tabela 2: Estatísticas básicas do sequenciamento e montagem genômicas..	51
Tabela 3: Resultados do BUSCO v4.1.4 para as 40 amostras <i>Kerteszia</i>	58
Tabela 4: Resultados do BUSCO v4.1.4 para as amostras sequenciadas com Nanopore.....	59
Tabela 5: Cálculos das frequências genotípicas esperadas sob o equilíbrio de Hardy-Weinberg.	60
Tabela 6: Média dos valores de F_{ST}	62

Lista de Figuras

Figura 1: Locais de coleta de <i>Kerteszia</i> na Mata Atlântica brasileira. (a) Mapa do Brasil e (b) Detalhe mostrando o local onde as amostras de <i>Kerteszia</i> foram coletadas. Os mapas da Figura 1 foram produzidos utilizando os pacotes: maptools, maps e GISTools do R Software (R CORE TEAM, 2021).	45
Figura 2: A- Alinhamento das sequências de nucleotídeos do gene <i>cpr</i> e as posições dos primers. Alinhamento do fragmento do gene <i>cpr</i> de uma sequência de <i>An. cruzii</i> s.l. de Bocaina Grupo 1 e duas sequências de Bocaina Grupo 2 (segundo Dias et al. 2018). Os números de acesso do GeneBank KT724974.1 e KT724976.1 correspondem ao Bocaina Grupo 1, e KT724986.1 e KT724996.1 ao Bocaina Grupo 2. O triângulo invertido representa a posição da diferença de nucleotídeo único entre os dois primers forward que distinguem o Grupo Bocaina 1 (Boc1F em azul) e Grupo Bocaina 2 (Boc2F em verde). O primer reverso (BocR) está em roxo. B- Visualização em gel de agarose (1%) das amostras classificadas como Bocaina Grupo 1 (G1) e Bocaina Grupo 2 (G2). P1: <i>primer forward</i> projetado para Bocaina Grupo 1 (Boc1F: 5'GTGTAATATGGTAAGCGAACG). P2: <i>primer forward</i> projetado para Bocaina Grupo 2 (Boc2F: 5'GTGTAATATGGTAAGCGAAACG). (Esses primers amplificaram um fragmento de ~180 pb do gene <i>cpr</i>) C-: controle negativo (uso de água ao invés de DNA).	49
Figura 3: Análise de F_{ST} da comparação Bocaina Grupo 1 versus Bocaina Grupo 2, por cromossomo. Os valores de F_{ST} estão indicados no (eixo Y). Os genes foram classificados em braços cromossômicos com base em a localização de seus ortólogos no genoma de <i>An. cruzii</i> do <i>Anopheles Reference Genomes Project</i> (acesso do GenBank: GCA_943734635.1).	64
Figura 4: Diferenciação genética das espécies irmãs de <i>An. cruzii</i> s.l. Os valores de F_{ST} (eixo X) foram calculados entre comparações pareadas de <i>An. cruzii</i> s.l., e também para <i>An. bellator</i> (<i>An. cruzii</i> acima e <i>An. bellator</i> abaixo).....	65
Figura 5: Árvore de espécies inferida com a abordagem de Máxima Verossimilhança (ML) a partir de 2.051 genes presentes em pelo menos 20 das 40 amostras de <i>Kerteszia</i> . <i>Anopheles (Kerteszia) bellator</i> e <i>Anopheles</i>	

(*Kerteszia*) *homunculus* são espécies diferentes reconhecidas e foram usadas para enraizar a árvore. Essas análises filogenômicas sugerem fortemente que *An. cruzii* s.l. (conforme identificada morfológicamente) é composta por pelo menos quatro espécies crípticas: Santa Teresa, Bocaina Grupo 1, Bocaina Grupo 2 + Itatiaia e Florianópolis + Guapimirim + Maquiné (GCA_943734635.1). A evidência de especiação críptica é especialmente convincente no caso das amostras de Bocaina, que vieram do mesmo local e ainda formam dois grupos bastante distantes filogeneticamente. Os números nos nós representam valores de bootstrap (ML). Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Cam: Camacan; Gua: Guapimirim; Ilh: Ilha Grande; San: Santa Teresa; Itp: Itaparica. M: machos; F: fêmeas.....67

Figura 6: Tempos de divergência estimados com RelTime, no MEGA 11. Na legenda, estão ilustrados os tempos absolutos. *An. bellator* foi utilizado como grupo externo.69

Figura 7: Nos histogramas os valores mostrados são as porcentagens de observações para FST para espécies e populações..... 74

Lista de Figuras Suplementares

Figura suplementar 1: Árvores ML e MSC: Árvore de espécies inferida com abordagens de Máxima Verossimilhança (ML) e Coalescência Multiespécies (MSC) de 2.051 genes presentes em pelo menos 20 das 40 amostras de *Kerteszia*. *Anopheles (Kerteszia) bellator* e *Anopheles (Kerteszia) homunculus* são espécies diferentes reconhecidas. Essas análises filogenômicas sugerem fortemente que *An. cruzii* s.l. é composta por pelo menos quatro espécies crípticas: Santa Teresa, Bocaina Grupo 1, Bocaina Grupo 2, Itatiaia e Florianópolis + Guapimirim + Maquiné. A evidência de especiação críptica é especialmente convincente no caso das amostras de Bocaina, que vieram do mesmo local e ainda formam dois grupos bastante distantes filogeneticamente. Os números nos nós representam valores de bootstrap (ML) e probabilidades posteriores (MSC). Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; San: Santa Teresa; Itp: Itaparica. M: machos; F: fêmeas.92

Figura suplementar 2: Estruturas morfológicas: Fotografias das estruturas essenciais à identificação ao nível da espécie (cabeça, asa, pata e escudo), das 39 amostras (incluindo as três amostras sequenciadas individualmente com Nanopore): 1- *An. bellator* Ilh F1, 2- *An. bellator* Ilh M1, 3- *An. bellator* Itp F1, 4- *An. bellator* Itp M1), 5- *An. bellator* Cam F11, 6- *An. bellator* Cam F14, 7- *An. bellator* Cam F25, 8- *An. bellator* Cam F33, 9- *An. bellator* Cam F34, 10- *An. bellator* Cam F36, 11- *An. homunculus* Cam F1, 12- *An. homunculus* Cam F31, 13- *An. homunculus* Flo F1, 14- *An. homunculus* San F2, 15- *An. homunculus* San M2, 16- *Kerteszia* Boc F1, 17- *Kerteszia* Boc F2, 18- *Kerteszia* Boc M1, 19- *Kerteszia* Boc M2, 20- *Kerteszia* Boc M3, 21- *Kerteszia* Boc M4, 22- *Kerteszia* Boc M5, 23- *Kerteszia* Boc M6, 24- *Kerteszia* Boc F3 N, 25- *Kerteszia* Flo F3 N_1, 26- *Kerteszia* Flo F3 N_2, 27- *Kerteszia* Flo F2, 28- *Kerteszia* Flo M1, 29- *Kerteszia* Flo M2, 30- *Kerteszia* Flo M3, 31- *Kerteszia* Gua F1, 32- *Kerteszia* Gua

F2, 33- <i>Kerteszia</i> Itt F1, 34- <i>Kerteszia</i> Itt M1, 35- <i>Kerteszia</i> Itt M2, 36- <i>Kerteszia</i> Itt M3, 37- <i>Kerteszia</i> Itt M4, 38- <i>Kerteszia</i> San F1, 39- <i>Kerteszia</i> San M1, nesta ordem.	112
Figura suplementar 3: Caracteres morfológicos da genitália masculina: Imagens da genitália masculina. A- <i>Kerteszia</i> Boc M4 (Bocaina Grupo 1); B- <i>Kerteszia</i> Boc M2 (Bocaina Grupo 2), C- <i>Kerteszia</i> San M1; <i>Kerteszia</i> D- Flo M2.	113

Lista de Tabelas Suplementares

Tabela suplementar 1: Informações detalhadas sobre as amostras utilizadas neste estudo.	87
Tabela suplementar 2: Resultados de Busco V3 comparando os montadores SPAdes e Platanus em sete amostras.	88
Tabela suplementar 3: estatísticas das montagens de <i>Kerteszia</i> produzidas com QUASt.	89
Tabela suplementar 4: Média dos valores de F_{ST} das comparações entre todas as populações.	90

Resumo

Introdução: A malária é uma doença tropical causada pelo protozoário *Plasmodium*, que afeta milhões de pessoas anualmente. Embora no Brasil a maioria dos casos ocorra na Amazônia, ainda há transmissão fora dessa região, como na Mata Atlântica, onde o principal vetor é o *Anopheles (Kerteszia) cruzii*. Dados de citogenética, isoenzimas e marcadores moleculares sugeriram que *An. cruzii* é um complexo de espécies crípticas. Aqui usamos uma abordagem genômica para confirmar a existência de espécies crípticas e investigar as relações evolutivas dentro deste complexo e sua relação com outros *Kerteszia*.

Metodologia: Para realizar análises de diferenciação genética e inferências filogenômicas, sequenciamos 38 genomas completos de diferentes *An. cruzii* senso lato e espécies relacionadas de *Kerteszia*, usando as tecnologias Illumina e Nanopore.

Resultados: Relatamos os primeiros *drafts* de genoma de várias espécies de *Kerteszia*. As análises filogenômicas e F_{ST} indicam que *An. cruzii* s.l. é composta por pelo menos quatro espécies crípticas: (i) Sul/Sudeste do Brasil, aqui representado por Florianópolis-SC e Guapimirim-RJ (região costeira da Serra do Mar); (ii) Grupo Bocaina 1; (iii) Grupo Bocaina 2 (região continental da Serra do Mar, no Sudeste do Brasil) e Itatiaia - RJ (Serra da Mantiqueira, Sudeste do Brasil); (iv) Santa Teresa - ES (Sudeste do Brasil). A evidência de especiação críptica é particularmente convincente nos dois grupos encontrados na Bocaina: apesar da simpatria, esses grupos mostram um F_{ST} muito alto e falta de heterozigotos entre os grupos, indicando isolamento reprodutivo completo ou quase completo. Por outro lado, as populações de Florianópolis e Guapimirim, distantes aproximadamente 1000 km, não apresentaram separação genética clara e podem ser consideradas uma única espécie. Também descobrimos uma nova espécie de Santa Teresa - ES, que é geneticamente e morfologicamente distinguível dos outros quatro grupos.

Conclusões: Este trabalho demonstra inequivocamente a existência de espécies crípticas em *An. cruzii* s.l., e estimou a diferenciação genética entre as populações utilizando o maior número de genes até o momento.

Palavras-chave: Malária, *Anopheles (Kerteszia) cruzii*, análise genômica, complexo de espécies, especiação críptica.

Abstract

Background: Malaria is a tropical disease caused by *Plasmodium* protozoans, that affects millions of people annually. Although in Brazil most cases occur in the Amazon, there is still transmission outside this region, as in the Atlantic Forest where the main vector is *Anopheles (Kerteszia) cruzii*. Data from cytogenetics, isoenzymes, and molecular markers suggested that *An. cruzii* is a complex of cryptic species. Here we used a genomic approach to confirm the existence of cryptic species, and investigate the evolutionary relationships within this complex, and their relationship with other *Kerteszia*.

Methodology: In order to perform genetic differentiation analysis and phylogenomic inferences, we sequenced 38 complete genomes from different *An. cruzii* *sensu lato* groups and related *Kerteszia* species, using Illumina and Nanopore technologies.

Results: We report the first genome drafts for several species of *Kerteszia*. The phylogenomic and F_{ST} analyses indicates that *An. cruzii* *s.l.* is composed of at least four cryptic species: (i) South/Southeast Brazil, represented here by Florianópolis-SC and Guapimirim-RJ (Serra do Mar coastal region); (ii) Bocaina Group 1; (iii) Bocaina Group 2 (Serra do Mar continental region, in Southeast Brazil) and Itatiaia - RJ (Serra da Mantiqueira, Southeast Brazil); (iv) Santa Teresa - ES (Southeast Brazil). The evidence of cryptic speciation is particularly compelling in the two Bocaina groups: despite sympatry, these groups show a very high F_{ST} and lack of inter-group heterozygotes, indicating complete or nearly complete reproductive isolation. On the other hand, the Florianópolis and Guapimirim populations, which are ~1000 km apart, showed no clear genetic separation and could be considered a single species. We also uncovered a new species from Santa Teresa - ES, which is genetically and morphologically distinguishable from the other four groups.

Conclusions:

This work unequivocally demonstrates the existence of cryptic species in *An. cruzii* *s.l.*, and estimated genetic differentiation between populations using the largest number of genes to date.

Key words: Malaria, *Anopheles (Kerteszia) cruzii*, genomic analysis, species complex, cryptic speciation.

Objetivos

Objetivo Geral

O objetivo principal deste trabalho foi elucidar a existência de um complexo de espécies crípticas dentro do táxon *An. cruzii*, utilizando uma abordagem filogenômica.

Objetivos específicos

- Realizar coletas de *An. cruzii* em diferentes locais como: Itaparica- BA (13°03'36.1"S, 38°46'35.1"W), Camacan- BA (15°23'30.9"S, 39°33'54.6"W), Santa Teresa- ES (19°57'54.6"S, 40°32'24.2"W), Guapimirim- RJ (22°30'22.2"S, 43°00'59.2"W), Ilha Grande- RJ (23°07'04.5"S, 44°16'29.9"W), Itatiaia- RJ (22°27'07.3"S, 44°36'44.9"W), Bocaina- SP (22°43'9.95"S, 44°36'47.1"W) e Florianópolis-SC (27°31'52.2"S., 48°30'44.2"W).
- Sequenciar o genoma completo de mosquitos das populações acima citadas, utilizando Illumina e/ou Nanopore;
- Realizar estudos populacionais e filogenômicos com as diferentes populações, a fim de avaliar estruturação genética e especiação críptica para *An. cruzii*.

Materiais e métodos

Locais de coleta

Os mosquitos utilizados neste estudo foram coletados ao longo da Mata Atlântica brasileira nos seguintes locais: Itaparica (Bahia BA), Camacan (BA), Santa Teresa (ES), Guapimirim (RJ), Ilha Grande (RJ), Itatiaia (RJ), Bocaina (SP) e Florianópolis (SC) (Figura 1).

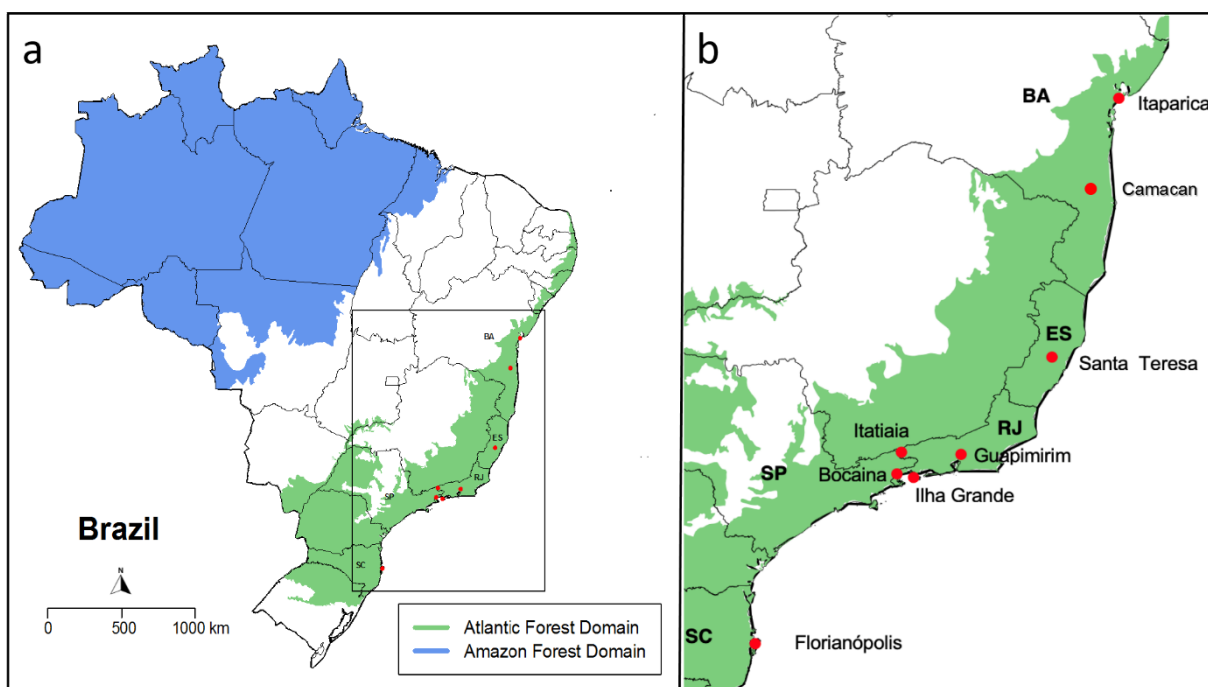


Figura 1: Locais de coleta de *Kerteszia* na Mata Atlântica brasileira. (a) Mapa do Brasil e (b) Detalhe mostrando o local onde as amostras de *Kerteszia* foram coletadas. Os mapas da Figura 1 foram produzidos utilizando os pacotes: maptools, maps e GISTools do R Software (R CORE TEAM, 2021).

Os mosquitos adultos e formas imaturas (larvas e pupas) foram coletados, tratados e preservados conforme descrito em Dias et al.(2018).

Informações mais detalhadas sobre a coleta das amostras, como coordenadas geográficas, sexo dos mosquitos e estágio de vida em que as amostras foram coletadas (adulto ou larva) podem ser visualizados na Tabela Suplementar 1.

A identificação das espécies foi realizada de acordo com Consoli e Lourenço- de-Oliveira (1994) e Forattini (FORATTINI, 2002). As estruturas

morfológicas importantes para a identificação em nível de espécie foram fotografadas com estereomicroscópio Olympus SZX16 para cada indivíduo sequenciado.

Isolamento de DNA

Para o isolamento de DNA dos mosquitos individuais, foi empregado um método não destrutivo, modificado de Santos et al., (2018). Este método permite a extração de DNA preservando a morfologia das partes taxonomicamente relevantes dos animais (por exemplo: exoesqueleto e genitália masculina). As amostras foram previamente preservadas em etanol 100% a -20°C, e o DNA foi extraído com o kit Qiagen Puregene core (número de catálogo Qiagen: 158667) como descrito a seguir. Primeiramente, os animais foram separados na junção abdômen-tórax e colocados em tubos individuais contendo 100µl de solução de lise e 1µl de proteinase K. As amostras foram incubadas em banho seco nesta solução por 3 dias a 45°C e posteriormente incubadas em gelo por 1 minuto. Em seguida, 33µl de solução de precipitação foram adicionados a cada amostra e homogeneizados por inversão. As amostras foram recolocadas no gelo por 5 minutos e centrifugadas por 3 minutos a 21,130 rcf. Transferimos o sobrenadante para um novo tubo, descartamos o tubo antigo contendo o *pellet* e adicionamos 0,5 µl de RNase (4 ug/mL, número de catálogo da Qiagen: 158922). Os tubos foram incubados (i) a 65°C por 15 minutos, (ii) a 37°C por 15 minutos; 2 µl de proteinase K foram adicionados (20 mg/ml, número de catálogo da Qiagen: 158918), e (iii) eles foram incubados a 65°C por 15 minutos. As amostras foram colocadas no gelo por 1 minuto. Em seguida, foram adicionados 33 µl de solução de precipitação, homogeneizados por inversão e colocados no gelo por 5min. Os tubos foram centrifugados novamente a 21.130 rcf por 3 minutos. O sobrenadante foi colocado em um tubo novo com 1 µl de coprecipitante GlycoBlue™ Invitrogen™ (15 mg/ml, número de catálogo: AM9515) homogeneizado e, em seguida, 100 µl de isopropanol foram adicionados e homogeneizados por inversão. Os tubos foram centrifugados a 21.130 rcf por 5 minutos, o sobrenadante foi descartado e os pellets azuis foram deixados secando brevemente à temperatura ambiente. O pellet foi lavado com 100 µl de etanol 70% e centrifugado a 21.130 rcf por 1 minuto. O sobrenadante foi então

descartado e o pellet foi deixado secando em temperatura ambiente por aproximadamente 10 minutos. Em seguida, foram adicionados 50µl da solução de hidratação do DNA e os tubos foram incubados a 65°C por 1 hora e deixados em temperatura ambiente durante a noite (para solubilização do DNA). A concentração de DNA foi estimada com Qubit utilizando os reagents *dsDNA quantitation, high sensitivity* (número de catálogo Invitrogen Q32851). Os tubos foram então armazenados a -20°C até o sequenciamento. Seguimos os mesmos procedimentos gerais para sequenciar amostras tanto com Illumina quanto com Nanopore.

Características morfológicas da genitália masculina

A análise da genitália masculina é um dos métodos mais precisos para identificação morfológica de espécies em *Anopheles* (SALLUM et al., 2020) e foi usada no presente estudo. Modificamos o protocolo de Consoli & Lourenço-de-Oliveira (1994) da seguinte forma.

As genitálias foram conservadas após o processo de extração de DNA. Após a extração, foram cuidadosamente cortadas do abdômen e posteriormente clareadas em KOH 10% por um período de 12 horas. Posteriormente, as amostras foram submetidas a um processo de desidratação utilizando uma série graduada de etanol, começando com uma solução de 70%, seguida de 80% e 90% por 15 minutos. Seguiu-se a imersão em solução de etanol 95% e absoluto, por 10 minutos cada. Para obter visibilidade ótima, uma segunda etapa de clarificação foi realizada com xileno por 60 minutos. As lâminas foram montadas em Entellan.

Identificação Bocaina Grupo 1 e Bocaina Grupo 2

Para identificar as duas espécies simpátricas de Bocaina, foi utilizado um fragmento parcial do gene *cpr*, uma vez que o gene já se mostrou um bom marcador para diferenciar esses dois grupos (DE REZENDE DIAS et al., 2018). Desenhamos dois *primers forward* específicos para cada espécie irmã, denominados Boc1F (5'GTGTAATATGGTAAGCGAACG) e Boc2F (5'GTGTAATATGGTAAGCGAAACG), que em combinação com o *primer*

reverso compartilhado BocR (5' TTTCTCGATGTCTTTCAGCT) permitiu a amplificação alelo-específica de um fragmento do gene *cpr* (Figura 2).

A

```

          ▼
KT724974.1  GTGTAATATGGTAAGCGAA-CG----AGAGAGAGTCCTCGCCTATACGGTGACGCCGGCG
KT724976.1  GTGTAATATGGTAAGCGAA-CG----AGAGAGAGTCCTCGCCTATACGGTGACGCCGGCG
KT724986.1  GTGTAATATGGTAAGCGAAACGAGAGGGAGAGAGTCCGCGCCTATACGGTGACGCCGGCG
KT724996.1  GTGTAATATGGTAAGCGAAACGAGAGAGAGAGAGTCCGCGCCTATACGGTGACGCCGGCG
*****

KT724974.1  GG----CGCGCCAGCATGTTGTAATCCGTTCCGTTCCACTCTCTCTCTCTGTGCACAC
KT724976.1  GG----CGCGCCAGCATGTTGTAATCCGTTCCGTTCCACTCTCTCTCTCTGTGCACAC
KT724986.1  GCGGGCGGGCCAGCATGTTGTAATCCGTTCCGTTCCAC-----TCTCTCTGTGCACAC
KT724996.1  GCGGGCGGAGCCAGCATGTTGTAATCCGTTCCATTCCAC-----TCTCTCTGTGCACAC
**          ** *****

KT724974.1  GCAGGAAGAGCTGTTGCAGCTGAAAGACATCGAGAAATC
KT724976.1  GCAGGAAGAGCTGTTGCAGCTGAAAGACATCGAGAAATC
KT724986.1  GCAGGAAGAGCTGTTGCAGCTGAAAGACATCGAGAAATC
KT724996.1  GCAGGAAGAGCTGTTGCAGCTGAAAGACATCGAGAAATC
*****

```

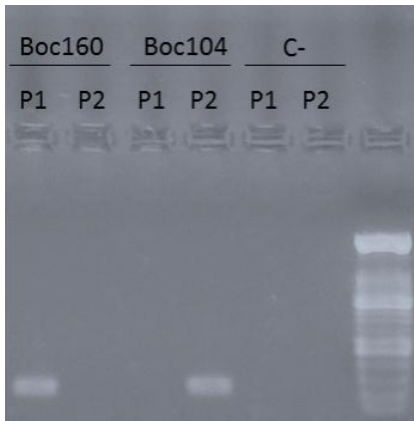


Figura 2: A- Alinhamento das sequências de nucleotídeos do gene *cpr* e as posições dos primers. Alinhamento do fragmento do gene *cpr* de uma sequência de *An. cruzii* s.l. de Bocaina Grupo 1 e duas sequências de Bocaina Grupo 2 (segundo Dias et al. 2018). Os números de acesso do GeneBank KT724974.1 e KT724976.1 correspondem ao Bocaina Grupo 1, e KT724986.1 e KT724996.1 ao Bocaina Grupo 2. O triângulo invertido representa a posição da diferença de nucleotídeo único entre os dois primers forward que distinguem o Grupo Bocaina 1 (Boc1F em azul) e Grupo Bocaina 2 (Boc2F em verde). O primer reverso (BocR) está em roxo. **B-** Visualização em gel de agarose (1%) das amostras classificadas como Bocaina Grupo 1 e Bocaina Grupo 2. P1: primer forward projetado para Bocaina Grupo 1 (Boc1F: 5'GTGTAATATGGTAAGCGAACG). P2: primer forward projetado para Bocaina Grupo 2 (Boc2F: 5'GTGTAATATGGTAAGCGAAACG). (Esses primers amplificaram um fragmento de ~180 pb do gene *cpr*) C-: controle negativo (uso de água ao invés de DNA).

As ampliações por PCR foram realizadas em um Termociclador Applied Biosystems Veriti 96-Well, Modelo 9902, usando GoTaq® Hot Start Polymerase (Promega, número de catálogo: M500B) e as seguintes condições: 95°C por 9 minutos, seguindo 40 ciclos a 90 °C por 30 segundos, 62°C por 30 segundos e 72°C por 20 segundos, seguindo um ciclo de 72°C por 7 minutos. Os produtos de PCR foram aplicados em géis de agarose (1%), corados com brometo de etídio e fotografados em transiluminador UV.

Sequenciamento, montagem do genoma e remoção de contaminantes

Bibliotecas *Illumina 150 bp paired-end TruSeq Nano DNA* ou *Nextera XT* (tamanhos de inserto 350bp) foram produzidas e sequenciadas com HiSeq 2000 na Macrogen, Coreia (Tabela 2).

Tabela 2: Estatísticas básicas do sequenciamento e montagem genômicas.

Dados das amostras				Resultados do GenomeScope				Tamanho da montagem fornecido pelo QUAST	
ID da Amostra	Nome da amostra	Sexo	Localidade	Biblioteca	Unic (%)	Hetero (%)	Cobertura	Tamanho (bp)	Tamanho da montagem
An. bellator Ilh F1	boni2	F	Ilha Grande (RJ)	Illumina TruSeq Nano	87,1	1,42	33,4	176,260,269	128,742,914
An. bellator Ilh M1	Boni_8	M	Ilha Grande (RJ)	Illumina TruSeq Nano	86,9	1,65	49,2	168,050,736	151,900,482
An. bellator Itp F1	Itaparica1-2_F	F	Itaparica (BA)	Illumina TruSeq Nano	89,6	1,36	96,2	170,153,789	160,130,369
An. bellator Itp M1	Itaparica_1_M	M	Itaparica (BA)	Illumina TruSeq Nano	88,7	1,64	57,8	164,714,621	155,423,177
An. bellator Cam F11	Cam F11	F	Camaca (BA)	Illumina TruSeq Nano	87,6	1,64	48,4	174,910,021	146,891,939
An. bellator Cam F14	Cam F14	F	Camaca (BA)	Illumina TruSeq Nano	86,8	1,63	46,8	176,476,552	145,526,116
An. bellator Cam F25	Cam F25	F	Camaca (BA)	Illumina TruSeq Nano	86,8	1,63	43	176,681,770	141,022,429
An. bellator Cam F33	Cam F33	F	Camaca (BA)	NEXTERA XT	82,6	2,2	132,6	174,077,126	233,797,627
An. bellator Cam F34	Cam F34	F	Camaca (BA)	Illumina TruSeq Nano	87,6	1,67	51	174,982,685	147,865,918
An. bellator CamF36	Cam F36	F	Camaca (BA)	Illumina TruSeq Nano	87,4	1,61	46	174,894,682	145,243,795
An. homunculus Flo F1	c40_UCAD	F	Florianópolis (SC)	Illumina TruSeq Nano	87,9	2,05	111,4	172,358,632	163,810,124
An. homunculus San F2	Vv_bio3	F	Santa Teresa (SC)	NEXTERA XT	81,2	3,02	71,6	156,007,070	141,298,410
An. homunculus San M2	Vv_bio1	M	Santa Teresa (ES)	NEXTERA XT	81,3	3,06	72,2	148,028,886	132,845,753
An. homunculus Cam F1	Cam F1	F	Camaca (BA)	Illumina TruSeq Nano	89,7	2,25	59,8	171,669,693	148,53,0252
An. homunculus Cam F31	Cam F31	F	Camaca (BA)	Illumina TruSeq Nano	90,3	2,4	56,4	170,145,392	148,108,851
Kerteszia Boc F1	boc104	F	Bocaina (SP)	Illumina TruSeq Nano	88,7	2	44,8	174,805,481	138,751,828
Kerteszia Boc F2	boc140	F	Bocaina (SP)	Illumina TruSeq Nano	85,9	1,78	44	181,281,242	143,652,064
Kerteszia Boc M1	boc56	M	Bocaina (SP)	Illumina TruSeq Nano	86,5	2,3	99,8	170,424,035	164,236,966
Kerteszia Boc M2	Boc146	M	Bocaina (SP)	Illumina TruSeq Nano	86,8	2,02	115,4	170,962,065	166,174,346
Kerteszia Boc M3	boc160	M	Bocaina (SP)	Illumina TruSeq Nano	86,6	1,89	48,8	170,286,278	150,640,365
Kerteszia Boc M4	boc66	M	Bocaina (SP)	Illumina TruSeq Nano	87,5	2,23	38,8	170,170,567	122,343,481
Kerteszia Boc M5	boc85	M	Bocaina (SP)	Illumina TruSeq Nano	86,3	2,24	36,4	173,220,294	120,988,892
Kerteszia Boc M6	boc86	M	Bocaina (SP)	Illumina TruSeq Nano	85,2	2,21	34,6	175,013,931	121,318,591
Kerteszia Boc F3 N	Boc F3 N	F	Bocaina (SP)	Nanopore	NA	NA	63*	NA	183,033,428
Kerteszia Flo F3 N	Flo F3 N	F	Florianópolis (SC)	Nanopore	NA	NA	25*	NA	170,347,651
Kerteszia Flo F2	UCAD 7.5	F	Florianópolis (SC)	NEXTERA XT	76,3	2,73	72,4	160,085,452	144,104,707
Kerteszia Flo M1	UCAD13-7m	M	Florianópolis (SC)	Illumina TruSeq Nano	88,8	2,29	141,4	165,105,180	165,555,787
Kerteszia Flo M2	UCAD 10.11	M	Florianópolis (SC)	Illumina TruSeq Nano	86,9	2,23	36,2	171,607,016	119,500,679
Kerteszia Flo M3	UCAD 6.2	M	Florianópolis (SC)	Illumina TruSeq Nano	87,9	2,23	35,6	169,405,078	115,214,550
Kerteszia Gua F1	Guapi F1	F	Guapimirim (RJ)	Illumina TruSeq Nano	88,4	2,12	109	171,837,213	164,019,820
Kerteszia Gua F2	Gua F2	F	Guapimirim (RJ)	Illumina TruSeq Nano	88,6	2,15	46	173,909,753	137,993,792
Kerteszia Itt F1	M_lago_8_1_F	F	Itatiaia (RJ)	Illumina TruSeq Nano	89,2	2,1	107,6	170,589,846	163,573,267
Kerteszia Itt M1	CdoPesq1	M	Itatiaia (RJ)	Illumina TruSeq Nano	87,6	2,36	48	168,545,110	145,297,595
Kerteszia Itt M2	M_lago_11_1_M	M	Itatiaia (RJ)	Illumina TruSeq Nano	88,5	2,44	99,8	165,763,796	163,756,844
Kerteszia Itt M3	M_lago 2.1	M	Itatiaia (RJ)	Illumina TruSeq Nano	87	2,48	32,2	173,151,099	107,189,928
Kerteszia Itt M4	M_lago 12.1	M	Itatiaia (RJ)	Illumina TruSeq Nano	87,2	2,33	33,8	171,400,296	114,856,556
Kerteszia San F1	Pousada_1	F	Santa Teresa (ES)	NEXTERA XT	79,8	2,58	53,2	167,510,171	130,157,996
Kerteszia San M1	ResSLBf3_1	M	Santa Teresa (ES)	Illumina TruSeq Nano	85,8	2,08	52,6	174,146,596	156,672,719

Unic (%): porcentagem do genoma que é única (não repetitiva). Hetero. (%): taxa de heteroziguidade. Cobertura: Cobertura estimada. Tamanho (pb): comprimento total do genoma inferido com base em K-mer. Flo: Florianópolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica. M: machos; F: fêmeas.*Para as amostras sequenciadas por Nanopore, a cobertura média foi estimada pelo montador Flye.

As reads foram montadas com SPAdes-3.12.0 (BANKEVICH et al., 2012) e Platanus 1.2.4. (KAJITANI et al., 2014). Em seguida, comparamos a qualidade dessas montagens empregando BUSCO v3 (SIMÃO et al., 2015) e o conjunto de referência de ortólogos Diptera (odb9, baixado de <https://BUSCO.ezlab.org/> e contendo 2.799 genes). Depois de escolher a melhor montagem, usamos Blobtools (LAETSCH; BLAXTER, 2017) para detectar e remover contigs contaminantes na montagem primária (contigs "não artrópode"; tipicamente de origem bacteriana). BUSCO v4.1.4 (MANNI et al., 2021) e o conjunto de referência Diptera odb10 foram usados para avaliar a integridade final de todas as amostras. QUASt (GUREVICH et al., 2013) foi utilizado para estimar outras estatísticas de qualidade de montagem (por exemplo: N50). Jellyfish-2.3.0 (MARÇAIS; KINGSFORD, 2011) e GenomeScope (VURTURE et al., 2017) foram empregados para estimar o tamanho do genoma com base na cobertura de reads brutos.

Três indivíduos foram sequenciados usando a tecnologia Nanopore no equipamento MinION Mk1C, com o sequenciamento realizado usando reagentes SQK-LSK110, e flow cells MinION R9.4.1 (FLO-MIN106). Para avaliar o desempenho de diferentes montadores, usamos Canu 2.1.1 (KOREN et al., 2017) e Flye 2.9.1 (KOLMOGOROV et al., 2019). Ao final do processo de montagem, o Flye 2.9.1 fornece diversas estimativas de qualidade, incluindo N50 e cobertura. Para as duas amostras coletadas em Florianópolis, esses valores foram baixos, claramente devido à cobertura insuficiente. Consequentemente, decidimos combinar as reads dessas duas amostras e remontá-las, como descrito acima. Referimo-nos a esta montagem combinada como Flo F3 N e à outra amostra como Boc F3 N.

A fim de aumentar a precisão das montagens Nanopore, nós as polimos usando três rodadas de Pilon 1.24. Para isso utilizamos as reads Illumina da mesma subespécie (WALKER et al., 2014). Para o polimento da montagem Boc F3 N usamos reads Illumina das seguintes amostras: Boc F2, Boc M3 e Boc M2. Para a montagem Flo F3 N, empregamos reads de Flo M2, Flo M1, Flo M3 e Flo F2. Usamos BUSCO v4.1.4 (MANNI et al., 2021) e QUASt (GUREVICH et al., 2013) para avaliar a qualidade dessas montagens.

Além dos 38 genomas que foram sequenciados neste estudo, também utilizamos dois genomas de referência do *Projeto Genomas de Referência de Anopheles* (acesso no GenBank: GCA_943734635.1 e GCA_943735745.1 para *An. cruzii* coletado em Maquiné (RS), e *An. bellator* de Cananéia (SP), respectivamente).

Anotação gênica e busca por ortólogos

Usamos BUSCO v4.1.4 para anotar ortólogos de cópia única nas 40 montagens (veja acima), procurando por 3285 genes conservados entre Diptera do banco de dados OrthoDB (odb10, baixado de <https://BUSCO.ezlab.org/>). Este procedimento combina duas etapas necessárias à análise filogenômica: a anotação gênica e a identificação de genes ortólogos (WATERHOUSE et al., 2018).

Esses ortólogos de cópia única identificados pelo BUSCO foram utilizados na análise filogenômica, conforme descrito por Waterhouse et al (2018), e modificado conforme Dias et al (2020). Resumidamente, os ortólogos de cópia única foram anotados pelo BUSCO v4.1.4 e processados com os scripts `fix_busco_CDS_frame.txt` (DIAS et al., 2020) e `BUSCO_cleaning_pipeline.txt` (DIAS et al., 2020) que corrigem erros de anotação que podem influenciar as análises filogenômicas. O script `translatorex_vLocal.pl` (ABASCAL; ZARDOYA; TELFORD, 2010) foi utilizado para alinhar sequências de nucleotídeos com base nas sequências de proteínas que eles codificam.

Inferências filogenômicas

O programa BUSCO tenta anotar todos os 3285 genes representados no banco de dados odb10, mas inevitavelmente alguns genes estão faltando em algumas montagens, principalmente devido a erros de montagem. Esses dados ausentes podem ser tratados de maneira mais conservadora, como incluir nas análises filogenômicas apenas genes presentes em pelo menos 35 de 40 amostras, ou de maneira menos conservadora, como incluir apenas genes presentes em pelo menos 20 de 40 amostras. Idealmente, uma amostra mais

equilibrada de genes (a abordagem conservadora) seria preferível, pois ajuda a evitar vieses na estimativa de árvores de espécies. Uma vez que, dados ausentes ("missing data") podem distorcer as inferências filogenéticas (SMITH et al., 2020).

Em contrapartida, a desvantagem da abordagem conservadora é que ela resulta em um número menor de genes, o que pode reduzir o poder estatístico do conjunto de dados. Para minimizar esses problemas, realizamos análises filogenômicas usando dois conjuntos de dados: (i) um conjunto de dados "menos conservador" composto por 2.051 ortólogos presentes em pelo menos 20 de 40 amostras e (ii) um conjunto de dados "conservador" composto por 1.467 ortólogos presentes em pelo menos 35 de 40 amostras. Para ambos os conjuntos de dados, filogenias de máxima verossimilhança foram inferidas para cada gene usando modelos de melhor ajuste escolhidos no IQ-Tree 1.6.1 (NGUYEN et al., 2015).

Em seguida, aplicamos o TreeShrink (Mai e Mirarab, 2018) às árvores de genes produzidas pelo IQ-Tree (MAI; MIRARAB, 2018). Essa etapa removeu sequências discrepantes com ramificações muito longas, que poderiam surgir de erros de anotação (por exemplo, identificação incorreta de um parólogo como o ortólogo desejado) ou causas biológicas reais (por exemplo, um gene com evolução rápida). Essas sequências discrepantes têm o potencial de influenciar as análises e, portanto, foram eliminadas. Posteriormente, as árvores de espécies foram inferidas a partir dos conjuntos de dados "conservadores" e "menos conservadores" usando duas abordagens: máxima verossimilhança (ML) e coalescência multiespécies (MSC). O TreeShrink forneceu alinhamentos "filtrados" (sem sequências potencialmente problemáticas) e árvores de genes "podadas" (sem *outliers*). Os alinhamentos "filtrados" foram concatenados e usados para a inferência de máxima verossimilhança (ML) no IQ-TREE 1.6.1 (NGUYEN et al., 2015). As árvores de genes "podadas" foram utilizadas para a inferência de coalescência multiespécies (MSC) no ASTRAL (MIRARAB et al., 2014), empregando configurações padrão.

Para enraizar as árvores foram utilizadas as sequências de *An. bellator* e *An. homunculus*. Essas sequências representam diferentes espécies do mesmo subgênero de *An. cruzii*. As linhas de comando específicas usadas para realizar as inferências filogenômicas podem ser encontradas no arquivo adicional 1.

Análises de F_{ST}

As análises de F_{ST} das sequências nucleotídicas foram calculadas a partir dos alinhamentos filtrados de genes ortólogos usando o pacote do R hierfstat (GOUDET, 2005), com o script *pairwise_FST_cruzii_1.R* (Arquivo adicional 2). O F_{ST} foi usado para quantificar a diferenciação genética entre pares de populações.

Os genes foram mapeados em cromossomos com base na localização de seus ortólogos no genoma de *An. cruzii* do “Reference Genomes Project” (acesso no GenBank: GCA_943734635.1). O mapeamento foi feito usando o programa BlastN, tendo como "query" os genes de cópia única anotados por BUSCO v4.1.4 de seis amostras do complexo *An. cruzii* s.l concatenadas (GCA_943734635.1, *Kerteszia* Flo M1, *Kerteszia* Flo M2, *Kerteszia* Flo M3, *Kerteszia* Flo F2, *Kerteszia* Flo F3 N) e como base de dados a montagem GCA_943734635.1. O script *FST_cruzii.sync* (Arquivo adicional 3) foi usado para produzir os gráficos de F_{ST} , utilizando o SYSTAT.

Datação

Os tempos de divergência foram inferidos pela aplicação do método RelTime (TAMURA; TAO; KUMAR, 2018). A árvore filogenética com os comprimentos de ramos foi calculada usando o método máxima verossimilhança e *An. bellator* foi utilizado como grupo externo.

Esta análise envolveu 40 sequências genômicas. E utilizamos apenas a 3ª posição dos códons para minimizar a interferência da seleção natural. Havia um total de 748327 posições no conjunto de dados final. Estas análises foram realizadas no MEGA11 (TAMURA; STECHER; KUMAR, 2021).

A taxa de substituição utilizada neste estudo, de 19,9 substituições/kbp/milhão de anos, foi estimada por Dias et al. (dados não publicados) para a 3ª posição do códon de drosofilídeos. Essa taxa foi obtida no programa MCMCTree (REIS; YANG, 2011) através de uma calibração utilizando fósseis.

Resultados

Este estudo produziu os primeiros “*drafts*” de genoma para vários *Kerteszia*, e com eles pudemos obter vários resultados relevantes sobre estes mosquitos. Análises apontam ao menos quatro espécies distintas em *An. cruzii*: (i) Florianópolis-SC, Guapimirim-RJ, Maquiné- RS; (ii) Bocaina Grupo 1; (iii) Bocaina Grupo 2, Itatiaia – RJ; (IV) Santa Teresa – ES. A evidência de especiação é especialmente contundente nos grupos de Bocaina: embora ocorram em simpatria, possuem alta diferenciação genética e deficiência de heterozigotos, indicando isolamento reprodutivo. Por outro lado, populações de Florianópolis e Guapimirim, distantes 1000 km, não mostram separação genética clara, sugerindo pertencerem a uma única espécie. Surpreendentemente, também encontramos uma nova espécie em Santa Teresa - ES, distinta não apenas geneticamente, mas também morfológicamente das demais. Sugerimos também a possibilidade de um complexo para outra espécie deste subgênero: *An. bellator*. Nos parágrafos abaixo comentaremos com maiores detalhes esses resultados.

Montagem e qualidade dos genomas

Aqui, apresentamos os primeiros “*drafts*” de genoma de várias espécies de *Kerteszia*, como diferentes espécies de *An. cruzii* s.l., *An. bellator* s.l e *An. homunculus*. Um total de 38 genomas foram sequenciados, de oito populações diferentes. Ao total foram 23 amostras do complexo de *An. cruzii* s.l., 10 do complexo de *An. bellator* e 5 de *An. homunculus* (maiores informações sobre essas amostras podem ser encontradas na Tabela suplementar 1). De acordo com o GenomeScope, a cobertura dos genomas Illumina variou de 32x a 141x, e de acordo com o Flye, os genomas Nanopore tiveram coberturas de ~63x para *Kerteszia* Boc F3 N e ~25x para *Kerteszia* Flo F3 N (Tabela 2).

Utilizando o GenomeScope para a análise dos reads Illumina brutos, os tamanhos médio dos genoma foram: Para *An. bellator* ~ 173 Mb (intervalo: 164,7 - 176,6 Mb), para *An. homunculus* ~ 163 Mb (intervalo: 148 - 172,3 Mb), e para

An. cruzii s.l. ~ 170 Mb (intervalo: 160 - 181 Mb). Os tamanhos dos genomas obtidos nas montagens foram geralmente menores do que as estimativas baseadas nos reads brutos: para *An. bellator* foi ~ 155 Mb (intervalo: 128,7 - 233,7 Mb), para *An. homunculus* foi ~ 147 Mb (intervalo: 132,8 - 163,8 Mb), e para *An. cruzii* s.l. foi ~ 140 Mb (intervalo: 107,1 - 166,1 Mb), provavelmente devido ao colapso de regiões repetitivas durante a montagem. Para as amostras de *An. cruzii* s.l. sequenciadas com Nanopore, o tamanho estimado do genoma montado por Flye 2.9.1 (Kolmogorov et al., 2019) foi semelhante aos resultados da análise Illumina. Segundo o QUAST foram: para *Kerteszia* Flo F3 N ~170 Mb, e para *Kerteszia* Boc F3 N foi ~183 Mb (Tabela 2). O GenomeScope relatou uma alta heterozigosidade (média: ~ 2%), o que é de certa forma esperado para amostras de campo (Tabela 2).

Comparando os resultados do BUSCO v3 entre 7 amostras montadas com Platanus e SPAdes, um aumento de ortólogos completos acima de 20% foi obtido para a maioria das amostras montadas com Platanus em relação aos SPAdes (Tabela suplementar 2). Por esse motivo, o montador Platanus foi utilizado para montar todos os 36 genomas sequenciados com Illumina deste trabalho. Após as montagens com Platanus, avaliamos a qualidade usando o programa QUAST. Os valores de N50 variaram de 2,5 a 87,7 kb, o maior contig de cada genoma variou de 55 kb a 612 kb, e BUSCO v4.1.4 apresentou valores de ortólogos completos acima de 90% em 23 das 36 amostras (Tabela 3, Tabela suplementar 3). Cinco amostras foram sequenciadas com Illumina Nextera XT devido à baixíssima quantidade de DNA (< 85ng), que era insuficiente para o sequenciamento Illumina TruSeq. Estas 5 amostras (*Kerteszia* Flo F2, *Kerteszia* San F1, *An. homunculus* San F2 e *An. homunculus* San M2, *Kerteszia* Cam F33), apresentaram os menores valores de BUSCO v4.1.4 para seqüências ortólogas completas (36% – 67%).

Tabela 3: Resultados do BUSCO v4.1.4 para as 40 amostras *Kerteszia*.

ID da amostra	Nome da amostra	Completos	Completos cópia única	Completos duplicados	Fragmentados	Ausentes
An. bellator Ilh F1	boni2	3010 (91.6)	2999 (91.3)	11 (0.3)	136 (4.1)	139 (4.3)
An. bellator Ilh	Boni_8	3160 (96.2)	3145 (95.7)	15 (0.5)	59 (1.8)	66 (2.0)
M1 An. bellator Itp	Itaparica1-2_F	3209 (97.7)	3192 (97.2)	17 (0.5)	32 (1.0)	44 (1.3)
F1 An. bellator Itp	Itaparica_1_M	3194 (97.2)	3177 (96.7)	17 (0.5)	38 (1.2)	53 (1.6)
M1	Cam F11	3128 (95.3)	3116 (94.9)	12 (0.4)	63 (1.9)	94 (2.8)
An. bellator Cam F11	Cam F14	3106 (94.5)	3095 (94.2)	11 (0.3)	76 (2.3)	103 (3.2)
An. bellator Cam F14	Cam F25	3066 (93.3)	3056 (93.0)	10 (0.3)	119 (3.6)	100 (3.1)
An. bellator Cam F25	Cam F33	1553 (47.3)	1524 (46.4)	29 (0.9)	836 (25.4)	896 (27.3)
An. bellator Cam F33	Cam F34	3126 (95.1)	3118 (94.9)	8 (0.2)	69 (2.1)	90 (2.8)
An. bellator Cam F34	Cam F36	3118 (94.9)	3108 (94.6)	10 (0.3)	83 (2.5)	84 (2.6)
An. bellator CamF36	An. bellator Can	3238 (97.8)	3212 (97.8)	26 (0.8)	17 (0.5)	30 (0.9)
An. bellator GCA_943735745.1	c40_UCAD	3186 (96.9)	3165 (96.3)	21 (0.6)	44 (1.3)	55 (1.8)
An. homunculus Flo F1	Vv_bio3	1670 (50.8)	1657 (50.4)	13 (0.4)	751 (22.9)	864 (26.3)
An. homunculus San F2	Vv_bio1	1175 (35.8)	1160 (35.3)	15 (0.5)	834 (25.4)	1276 (38.8)
An. homunculus San M2	Cam F1	3040 (92.5)	3023 (92.0)	17 (0.5)	126 (3.8)	119 (3.7)
An. homunculus Cam F1	Cam F31	2904 (88.4)	2889 (87.9)	15(0.5)	196(6.0)	185 (5.6)
An. homunculus Cam F31	boc104	2955 (90.0)	2940 (89.5)	15 (0.5)	153 (4.7)	177 (5.3)
Kerteszia Boc F1	boc140	3056 (93.0)	3046 (92.7)	10 (0.3)	116 (3.5)	113 (3.5)
Kerteszia Boc F2	boc56	3176 (96.7)	3152 (96.0)	24 (0.7)	55 (1.7)	54 (1.6)
Kerteszia Boc M1	Boc146	3195 (97.2)	3177 (96.7)	18 (0.5)	34 (1.0)	56 (1.8)
Kerteszia Boc M2	boc160	3118 (94.9)	3107 (94.6)	11 (0.3)	71 (2.2)	96 (2.9)
Kerteszia Boc M3	boc66	2800 (85.2)	2790 (84.9)	10 (0.3)	259 (7.9)	226 (6.9)
Kerteszia Boc M4	boc85	2724 (82.9)	2717 (82.7)	07 (0.2)	291 (8.9)	270 (8.2)
Kerteszia Boc M5	boc86	2724 (82.9)	2714 (82.6)	10 (0.3)	283 (8.6)	278 (8.5)
Kerteszia Boc M6	Boc F3 N	3241 (98.6)	3158 (96.1)	83 (2.5)	19 (0.6)	25 (0.8)
Kerteszia Boc F3 N	UCAD 7.5	2120 (64.6)	2098 (63.9)	22 (0.7)	590 (18)	575 (17.4)
Kerteszia Flo F2	Flo F3 N	3236 (98.5)	3153 (96.0)	83 (2.5)	17 (0.5)	32 (1.0)
Kerteszia Flo F3 N	UCAD13-7m	3195 (97.3)	3170 (96.5)	25 (0.8)	42 (1.3)	48 (1.4)
Kerteszia Flo M1	UCAD 10.11	2718 (82.7)	2707 (82.4)	11 (0.3)	299 (9.1)	268 (8.2)
Kerteszia Flo M2	UCAD 6.2	2664 (81.1)	2654 (80.8)	10 (0.3)	337 (10.3)	284 (8.6)
Kerteszia Flo M3	Guapi	3166 (96.4)	3141 (95.6)	25 (0.8)	58 (1.8)	61 (1.8)
Kerteszia Gua F1	Gua F2	2809 (85.5)	2797 (85.1)	12 (0.4)	251 (7.6)	225 (6.9)
Kerteszia Gua F2	M_lago_8_1_F	3189 (97.1)	3174 (96.6)	15 (0.5)	39 (1.2)	57 (1.7)
Kerteszia Itt F1	CdoPesq1	2914 (88.7)	2901 (88.3)	13 (0.4)	205 (6.2)	166 (5.1)
Kerteszia Itt M1	M_lago_11_1_M	3165 (96.4)	3146 (95.8)	19 (0.6)	58 (1.8)	62 (1.8)
Kerteszia Itt M2	M_lago 2.1	2455 (74.7)	2445 (74.4)	10 (0.3)	432 (13.2)	398 (12.1)
Kerteszia Itt M3	M_lago 12.1	2622 (79.8)	2614 (79.6)	08 (0.2)	374 (11.4)	289 (8.8)
Kerteszia Itt M4	Pousada_1	2215 (67.5)	2203 (67.1)	12 (0.4)	562 (17.1)	508 (15.4)
Kerteszia San F1	ResSLBf3_1	3145 (95.7)	3127 (95.2)	18 (0.5)	75 (2.3)	65 (2.0)
Kerteszia San M1	Maq F	3242 (98.7)	3180 (96.8)	62 (1.9)	20 (0.6)	23 (0.7)

Kerteszia GCA_943735745.1

Anotação automática de montagens de 3285 genes ortólogos conservados entre Diptera do banco de dados OrthoDB. O número de genes em cada genoma com sequências completas, genes duplicados, genes com sequências fragmentadas ou genes ausentes em cada genoma é mostrado nesta tabela. Os valores entre parênteses são a porcentagem de BUSCOs em cada montagem. Flo: Florianópolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica. M: machos; F: fêmeas.

As duas amostras de Nanopore foram montadas com Canu e Flye. Observou-se que uma das montagens feitas com Canu apresenta uma proporção considerável de genes duplicados (22%). Essa duplicação pode ser causada pela alta heterozigosidade das amostras utilizadas. As montagens com Flye demonstraram maior contiguidade (com um valor de N50 mais elevado) e completude (com um maior número de Buscos completos de cópia única). Por esses motivos, o Flye foi utilizado para todas as análises destes genomas (Tabela 4). Em relação aos valores de N50, para a amostra *Kerteszia* Boc F3 N, o Canu resultou em 692,69 kb, enquanto o Flye produziu 1381,16 kb. No caso da amostra *Kerteszia* Flo F3 N, o Canu resultou em 207,71 kb, enquanto o Flye atingiu 446,87 kb (Tabela 4). Quanto aos valores de Busco completos de cópia única, para *Kerteszia* Boc F3 N, o Canu apresentou 77,1%, enquanto o Flye alcançou 96,1%. Já para *Kerteszia* Flo F3 N, o Canu obteve 95,5%, enquanto o Flye alcançou 96% (Tabela 4).

Tabela 4: Resultados do BUSCO v4.1.4 para as amostras sequenciadas com Nanopore.

ID da amostra	Nome da montagem	Montador	BUSCOS completos cópia única	Completos duplicados	Tamanho total	# contigs	Maior contig	GC (%)	N50
Boc F3N	Boc F3 N_canu_pilon3	Canu	2533 (77.1%)	715 (21.8%)	262112757	1824	7803337	48.51	692687
Boc F3N	Boc F3 N_flye_pilon3	Flye	3158 (96.1%)	83 (2.5%)	183033428	2296	13817620	49.07	1381163
Flo F3 N	Flo F3 N_canu_pilon3	Canu	3138 (95.5%)	97 (3.0%)	184789539	2735	1295927	48.86	207712
Flo F3 N	Flo F3 N _flye_pilon3	Flye	3153 (96.0%)	83 (2.5%)	170347651	2087	1969744	49.23	446870

Anotação automática de montagens de 3285 genes ortólogos conservados entre Diptera do banco de dados OrthoDB (odb10, baixado de <https://BUSCO.ezlab.org/>). O número de genes em cada genoma com sequências completas de cópia única e duplicados é mostrado em cada genoma nesta Tabela. Os valores entre parênteses são a porcentagem de BUSCOs em cada montagem. Estatísticas de montagens produzidas com QUAST. Tamanho total: número total de bases na montagem. # contigs: número total de contigs na montagem. Maior contig: comprimento do contig mais longo da montagem. GC (%): número total de nucleotídeos G e C na montagem dividido pelo comprimento total. N50: integridade da montagem do genoma (comprimento da sequência do contig mais curto em 50% do comprimento total do genoma). Boc: Bocaina; Flo: Florianópolis; F: Fêmea.

Diferenciação de Bocaina Grupo 1 e Bocaina Grupo 2

Dias et al (2018) coletou *An. cruzii* s.l. em Bocaina e encontrou dois haplótipos distintos do gene *cpr*, com déficit de heterozigotos, o que sugere especiação críptica. Investigamos aqui mais profundamente essa questão desenhando primers de PCR alelo específicos para o gene *cpr* (Figura 2), genotipando para o gene *cpr* uma amostra de 145 mosquitos capturados em Bocaina e sequenciando quatro representantes de Bocaina 1 e cinco de Bocaina 2. Nesta seção, relatamos os resultados da genotipagem para *cpr*, os resultados do sequenciamento foram tratados nas próximas seções.

Uma vez que o gene *cpr* está ligado ao cromossomo X em *An. cruzii* s.l. (baseado na montagem GCA_943735745.1), devemos levar em consideração o sexo para calcular as frequências alélicas (Tabela 5). A frequência do alelo *cpr1* entre as fêmeas foi de 3,45%; usando esta frequência calculamos as frequências genotípicas esperadas sob o equilíbrio de Hardy-Weinberg (Tabela 5, coluna 3). Da mesma forma que Dias et al (2018), encontramos um forte déficit de heterozigotos, o que é estatisticamente significativo (($P < 10^{-5}$, Teste Exato de Fisher; (SANTOS; LEMES; OTTO, 2020)). A explicação mais provável para esta deficiência de heterozigoto em simpatria é o isolamento reprodutivo (ou seja, especiação críptica), embora formalmente também possa ser causada por erros grosseiros na genotipagem ou seleção muito forte contra heterozigotos. Podemos descartar essas explicações e confirmar a hipótese da especiação críptica porque sequenciamos os genomas de quatro indivíduos portadores do alelo *cpr1* e cinco portadores do *cpr2* e, conforme detalhado nas próximas seções, encontramos a mesma forte diferenciação em todo o genoma. Dois grupos geneticamente diferenciados coexistindo em simpatria só podem ser espécies diferentes.

Tabela 5: Cálculos das frequências genotípicas esperadas sob o equilíbrio de Hardy-Weinberg.

Genótipo	Freq.	HW esperado
<i>cpr¹ / Y</i>	4	-
<i>cpr² / Y</i>	54	-
<i>cpr¹ / cpr¹</i>	3	0.1
<i>cpr¹ / cpr²</i>	0	5.8

cpr^2 / cpr^2	84	81.1
-----------------	----	------

Diferenciação Genética (F_{ST})

Os dois grupos de Bocaina foram classificados com base na genotipagem de *cpr* relatada na seção anterior. Em todas as comparações pareadas, valores elevados de F_{ST} foram observados (Figura 3, Figura 4, Tabela 6, Tabela suplementar 4).

Observe que valores de F_{ST} muito altos ocorrem em todos os cromossomos, com maiores valores no X (Tabela 6). O caso dos dois grupos simpátricos da Bocaina são de particular interesse. Os valores médios de F_{ST} foram em torno de 0,6 para autossomos e 0,9 para o cromossomo X (Figura 3, Tabela 6). Isso mostra que a diferenciação de *cpr* observada pela primeira vez por Dias et al (2018) realmente corresponde a duas espécies biológicas. Os dados descartam todas as explicações alternativas, como a seleção contra heterozigotos, porque não se espera que tenham efeitos em todo o genoma.

Vale ressaltar que Bocaina Grupo 1 e Grupo 2 são simpátricas e possuem altos valores de diferenciação (média total de F_{ST} de 0,6), o que indica que são espécies diferentes. Em contrapartida, foram obtidos baixos valores de F_{ST} entre Bocaina Grupo 2 e Itatiaia (média de 0,22). Da mesma forma, as populações de Florianópolis e Guapimirim, também apresentaram baixos valores de F_{ST} (média de 0,29) e podem ser consideradas uma única espécie. Ademais, obtivemos altos valores de F_{ST} em todas as comparações com *An. cruzii* de Santa Teresa (ES) (todos valores acima de 0,7). Estes resultados são, compatíveis com as posições desses grupos nas inferências filogenômicas (próxima seção).

Na Figura 4 é possível observar um padrão diferente para o gráfico de F_{ST} de *An. bellator* na comparação entre Ilha Grande e Itaparica, interessante padrão similar não ocorre em nenhuma das outras comparações. Em todas as comparações, constata-se um padrão semelhante a uma distribuição normal (em forma de sino). No entanto, na análise entre Ilha Grande e Itaparica, é possível observar que alguns genes demonstram valores elevados, enquanto outros exibem valores menores, distribuídos de maneira mais desordenada. É possível que este padrão possa ser devido a inversões cromossômicas, que podem

ocorrer nessa espécie (RAMÍREZ; DESSEN, 1996).

Ademais, também foram identificados altos valores de F_{ST} ao comparar as populações de *An. bellator*, sugerindo desta forma, a existência de espécies distintas para este complexo: uma em Cananéia (SP) e Ilha Grande (RJ) e outras duas espécies na Bahia (uma em Camacan e outra em Itaparica). Nestas análises, a média de F_{ST} entre Ilha Grande e Itaparica foi de 0,55, enquanto os valores entre Camacan e Itaparica, foram ainda mais elevados, apresentando uma média de 0,79 (esses resultados corroboram as análises filogenéticas). Não foi possível calcular o F_{ST} para Cananéia uma vez que, temos apenas um genoma desta população.

Para todas as comparações, os maiores valores de F_{ST} foram encontrados no cromossomo X (média de 0,9). Para os autossomos obtivemos média de 0,7 (Tabela suplementar 4). Isso sugere um forte efeito do cromossomo X no processo de diferenciação genética. Um resultado semelhante foi obtido com o complexo *An. gambiae* (NEAFSEY et al., 2015).

Por outro lado, diferente do observado para *An. cruzii* e *An. bellator*, não encontramos evidências de diferenciação genética significativa para *An. homunculus*. Isso é ilustrado, por exemplo, pelo valor de F_{ST} de 0,20 entre as populações de *An. homunculus* de Camacan e Santa Teresa (Tabela 6, Tabela suplementar 4).

Tabela 6: Média dos valores de F_{ST} .

População	Média do F_{ST}	N média	F_{ST} ChrX	N ChrX	F_{ST} Chr2	N Chr2	F_{ST} Chr3	N Chr3
BOC1-BOC2	0.625	1.967	0.895	184	0.599	716	0.595	1.034
BOC1-FLO	0.667	1.995	0.917	187	0.655	728	0.633	1.047
BOC1-GUA	0.687	1.589	0.931	178	0.670	565	0.647	823
BOC1-ITT	0.601	1.986	0.867	185	0.581	724	0.570	1.043
BOC1-SAN	0.800	1.183	0.923	111	0.787	435	0.789	611
BOC2-FLO	0.482	1.984	0.865	187	0.441	727	0.442	1.038
BOC2-GUA	0.456	1.591	0.867	179	0.395	571	0.409	819
BOC2-ITT	0.227	1.981	0.537	187	0.180	725	0.203	1.036
BOC2-SAN	0.723	1.180	0.917	109	0.693	437	0.711	609
FLO-GUA	0.293	1.605	0.691	178	0.248	575	0.239	829
FLO-ITT	0.459	2.001	0.838	188	0.423	734	0.419	1.046
FLO-SAN	0.751	1.193	0.932	111	0.734	441	0.732	616
GUA-ITT	0.427	1.597	0.828	180	0.375	570	0.376	824

GUA-SAN	0.734	967	0.911	110	0.710	343	0.713	496
ITT-SAN	0.701	1.193	0.895	110	0.674	443	0.688	614
b_CAM-b_IG	0.794	1.700	0.927	147	0.766	613	0.791	910
b_CAM-b_ITA	0.797	1.813	0.937	154	0.767	675	0.795	955
b_IG-b_ITA	0.553	1.672	0.810	149	0.459	608	0.574	886
h_CAM-h_STA	0.202	399	0.393	61	0.068	125	0.231	206

Média dos valores de F_{ST} das comparações entre as populações. N: Número de genes.

As amostras foram nomeadas da seguinte forma: Amostras do complexo *An. bellator*: "b_ITA": Itaparica (incluindo as amostras *An. bellator* ltp M1, *An. bellator* ltp F1), "b_IG": Ilha Grande (*An. bellator* Ilh M1, *An. bellator* Ilh F1), "b_CAM": Camaca (*An. bellator* CamF11, *An. bellator* CamF14, *An. bellator* CamF25, *An. bellator* CamF33, *An. bellator* CamF34, *An. bellator* CamF36). Amostras do complexo *An. cruzii*: "FLO": Florianópolis (*Kerteszia* Flo F3 N, *Kerteszia* Flo M1, *Kerteszia* Flo M2, *Kerteszia* Flo F2, *Kerteszia* Flo M3), "SAN": Santa Teresa (*Kerteszia* San M1, *Kerteszia* San F1), "GUA": Guapimirim (*Kerteszia* Gua F1, *Kerteszia* Gua F2), "ITT": Itatiaia (*Kerteszia* Itt M2, *Kerteszia* Itt M3, *Kerteszia* Itt M4, *Kerteszia* Itt M1, *Kerteszia* Itt F1), "BOC1": Bocaina Grupo 1 (*Kerteszia* Boc M2, *Kerteszia* Boc M3, *Kerteszia* Boc F2, *Kerteszia* Boc F3 N), "BOC2": Bocaina Grupo 2 (*Kerteszia* Boc F1, *Kerteszia* Boc M1, *Kerteszia* Boc M4, *Kerteszia* Boc M5, *Kerteszia* Boc M6). Amostras da espécie *An. homunculus*: h_CAM: Camaca (*An. homunculus* CamF31, *An. homunculus* CamF1), "h_FLO": Florianópolis (*An. homunculus* Flo F1) "h_STA": Santa Tereza (*An. homunculus* San M2, *An. homunculus* San F2).

Bocaina Grupo 1 x Bocaina Grupo 2

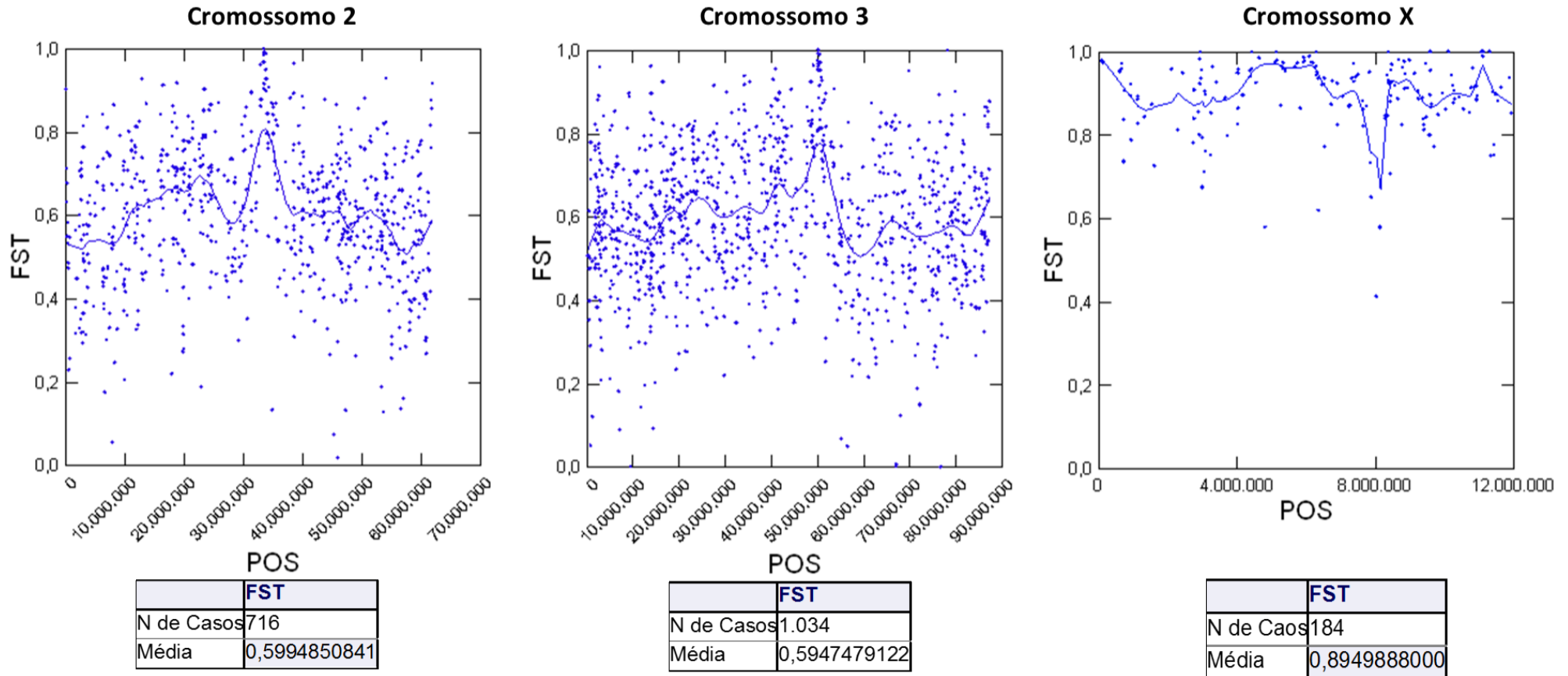


Figura 3: Análise de FST da comparação Bocaina Grupo 1 versus Bocaina Grupo 2, por cromossomo. Os valores de FST estão indicados no (eixo Y). Os genes foram classificados em braços cromossômicos com base em a localização de seus ortólogos no genoma de *An. cruzii* do *Anopheles Reference Genomes Project* (acesso do GenBank: GCA_943734635.1).

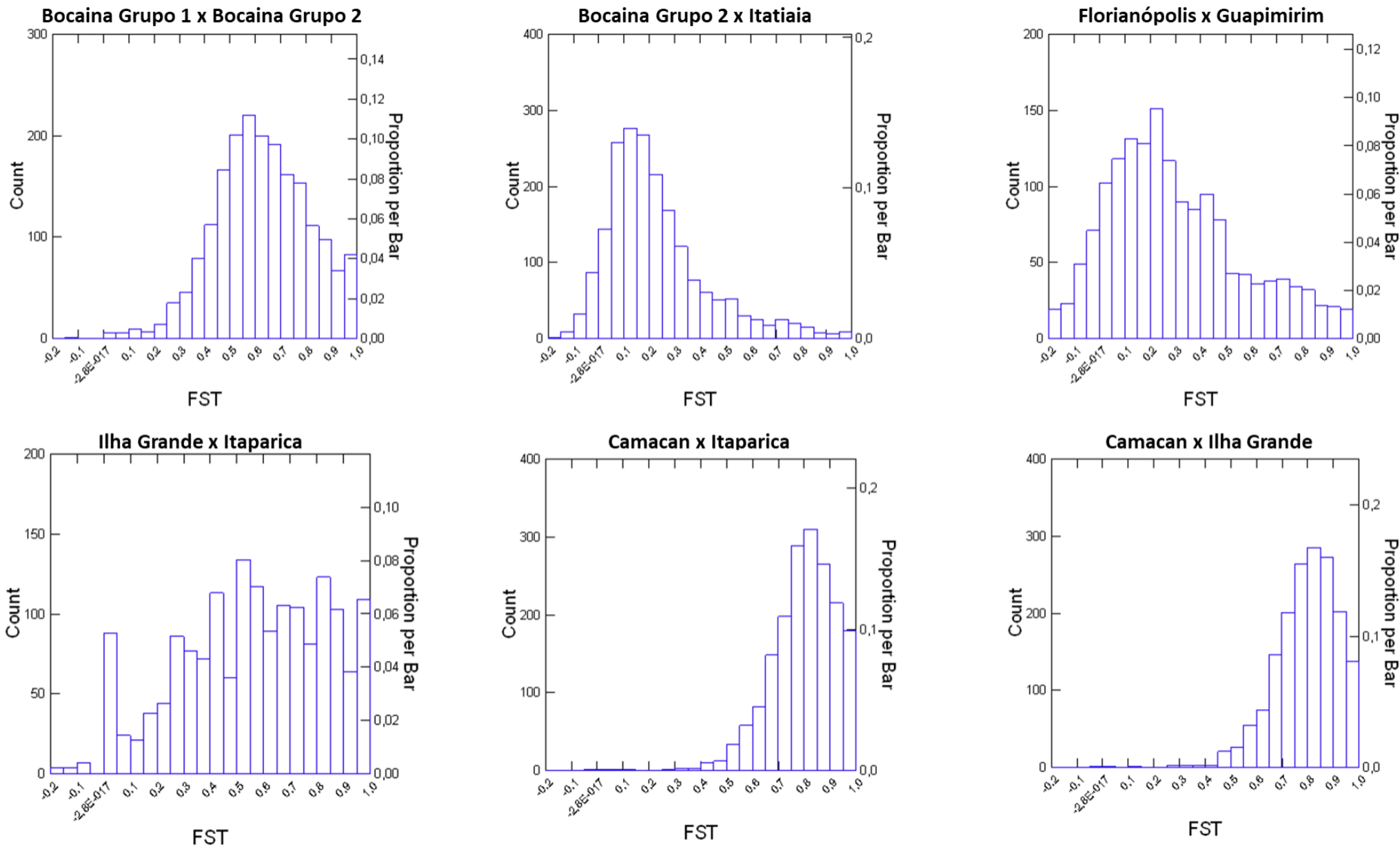


Figura 4: Diferenciação genética das espécies irmãs de *An. cruzii* s.l. Os valores de FST (eixo X) foram calculados entre comparações pareadas de *An. cruzii* s.l. e também para *An. bellator* (*An. cruzii* acima e *An. bellator* abaixo).

Inferências filogenômicas

Árvores de espécies foram inferidas para as 40 amostras de *An. cruzii* s.l. e espécies relacionadas usando os métodos de máxima verossimilhança e coalescência multiespécie com (i) 2.051 ortólogos presentes em pelo menos 20 das 40 amostras (Figura 5, Figura suplementar 1) e (ii) 1.467 ortólogos presentes em pelo menos 35 das 40 amostras (dados não mostrados). Todas as análises filogenômicas são consistentes e sugerem fortemente que *An. cruzii* s.l. é composta por pelo menos quatro espécies crípticas da seguinte forma: (i) Sul/Sudeste do Brasil, aqui representado por Florianópolis-SC, Guapimirim-RJ e Maquiné-RS (região costeira da Serra do Mar); (ii) Grupo Bocaina 1 e (iii) Grupo Bocaina 2, Itatiaia-RJ (Serra da Mantiqueira, Sudeste do Brasil), (iv) Santa Teresa-ES (Sudeste do Brasil). Apesar de ocorrerem no mesmo local, Bocaina Grupo 1 e Bocaina Grupo 2 não foram agrupados e apresentaram alta distância filogenética, o que mostra que pelo menos duas linhagens reprodutivamente isoladas (isto é, duas espécies crípticas) ocorrem em simpatria em Bocaina-SP.

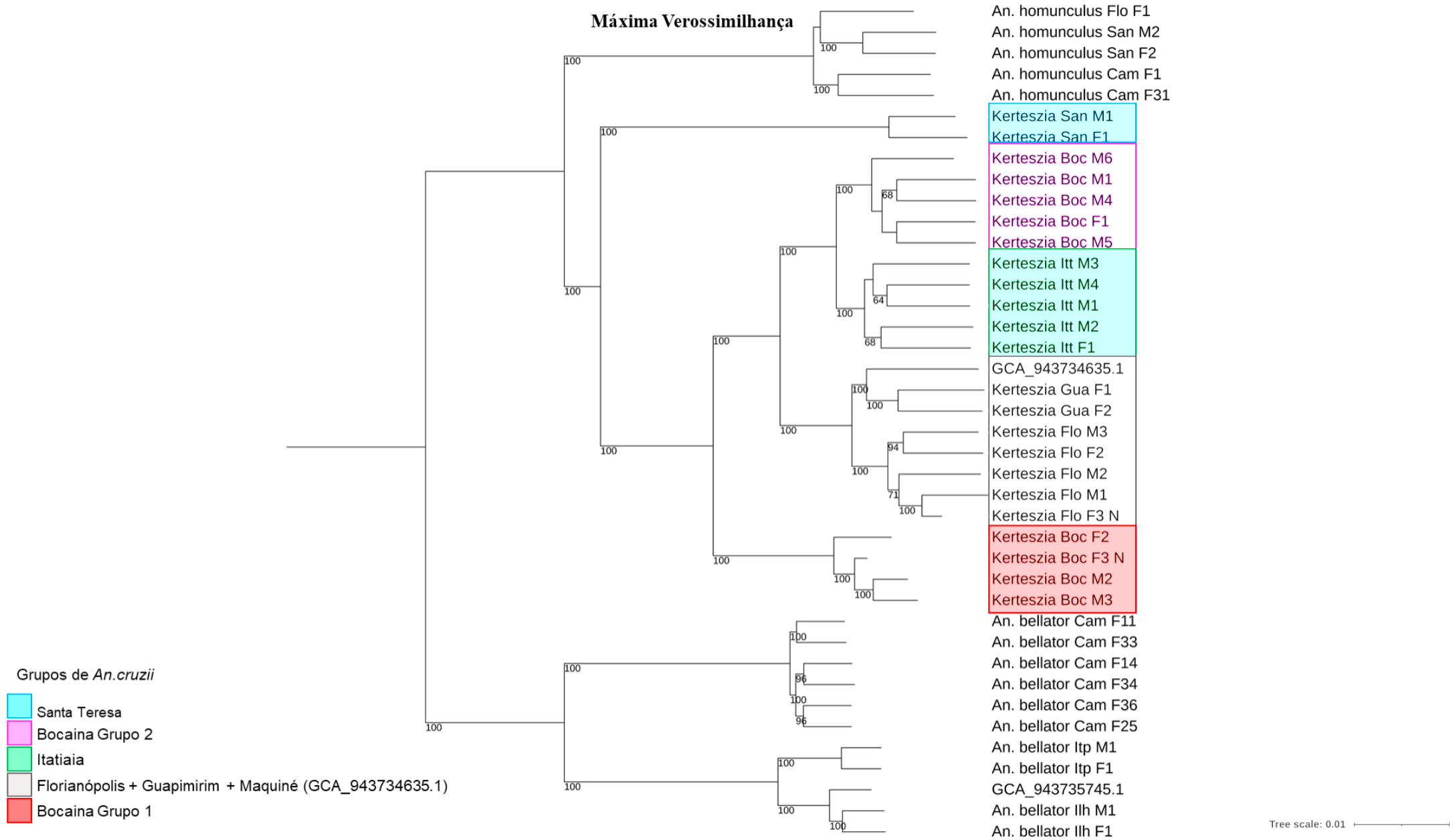


Figura 5: Árvore de espécies inferida com a abordagem de Máxima Verossimilhança (ML) a partir de 2.051 genes presentes em pelo menos 20 das 40 amostras de *Kerteszia*. *Anopheles (Kerteszia) bellator* e *Anopheles (Kerteszia) homunculus* são espécies diferentes reconhecidas e foram usadas para enraizar a árvore. Essas análises filogenômicas sugerem fortemente que *An. cruzii* s.l. (conforme identificada morfológicamente) é composta por pelo menos quatro espécies crípticas: Santa Teresa, Bocaina Grupo 1, Bocaina Grupo 2 + Itatiaia e Florianópolis + Guapimirim + Maquiné (GCA_943734635.1). A evidência de especiação críptica é especialmente convincente no caso das amostras de Bocaina, que vieram do mesmo local e ainda formam dois grupos bastante distantes filogeneticamente. Os números nos nós representam valores de *bootstrap* (ML). Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Cam: Camacan; Gua: Guapimirim; Ilh: Ilha Grande; San: Santa Teresa; Itp: Itaparica. M: machos; F: fêmeas.

Além disso, três grupos estruturados de *An. bellator* também foram encontrados neste trabalho: o primeiro parece estar localizado nas regiões Sudeste em Cananéia-SP (*An. bellator* GCA_943735745.1) e Ilha Grande-RJ (*An. bellator* Ilh F1, *An. bellator* Ilh M1). O segundo e o terceiro grupos são do Nordeste do Brasil: um de Itaparica-BA (*An. bellator* ltp F1, *An. bellator* ltp M1), e outro de Camacan-BA (*An. bellator* Cam F11, *An. bellator* Cam F14, *An. bellator* Cam F25, *An. bellator* Cam F33, *An. bellator* Cam F34, *An. bellator* Cam F36) (Figura 5, Imagem suplementar 1).

Tempos de Divergência

Os tempos de divergência foram calculados para a filogenia com 2051 genes, no software MEGA 11 utilizando o método RelTime, com uma taxa de substituição fixa de 19,9 substituições/kbp/milhão de anos (Dias et al., dados ainda não publicados).

As datações indicam que a separação entre Itatiaia e Bocaina Grupo 2 das populações do Sul e Sudeste do Brasil ocorreu há aproximadamente 1,25 milhões de anos (Figura 6).

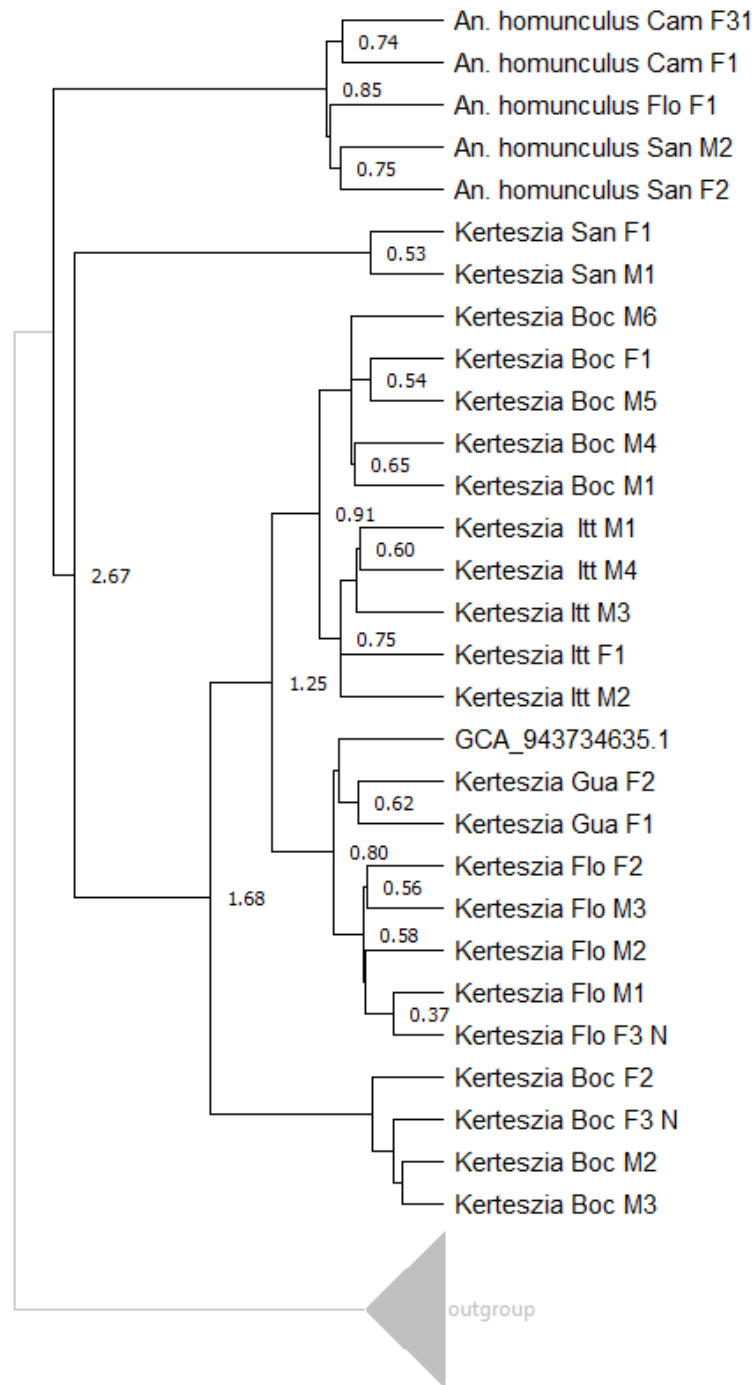


Figura 6: Tempos de divergência estimados com RelTime, no MEGA 11. Na legenda, estão ilustrados os tempos absolutos. *An. bellator* foi utilizado como grupo externo.

Morfologia da genitália masculina

As genitálias masculinas de 16 amostras sequenciadas foram examinadas: 2 delas eram de *An. bellator*, sendo 1 de Ilha Grande (RJ) e 1 de Itaparica (BA); e 14 eram de *An. cruzii* s.l., sendo 6 de Bocaina (SP) (2 de Boc1,

4 de Boc2), 4 de Itatiaia (RJ), 3 de Florianópolis (SC) e 1 de Santa Teresa (ES). Além disso, 14 espécimes foram inicialmente classificados como *An. cruzii*: Seis de Bocaina (duas de Boc1, quatro de Boc2), quatro de Itatiaia (RJ), três de Florianópolis (SC) e uma de Santa Teresa (ES). Note que nessa amostra há representantes de várias linhagens com alta diferenciação genética e que, portanto, presumivelmente pertencem a espécies diferentes. No entanto, foi encontrada diferença morfológica consistente apenas no claspeto ventral da *Kerteszia* San M1, amostra de Santa Teresa (ES). Esta estrutura em *Kerteszia* San M2 é distinta do típico claspeto de *An. cruzii* e se assemelha a um taco de golfe. As genitálias masculinas de todos os machos das outras três espécies do complexo são idênticas. Também não foram observadas diferenças morfológicas nas demais estruturas (Figura suplementar 2 e 3).

Em relação ao complexo *An. bellator*, a genitália masculina dos indivíduos de Itaparica (BA) e Ilha Grande (RJ) são idênticas. Infelizmente, só temos fêmeas de Camacan (BA), portanto não foi possível analisar essa característica diagnóstica nas amostras provenientes deste local.

Discussão

Montagem e qualidade de genomas

Neste trabalho, realizamos o sequenciamento e montagens de genomas de mosquitos *Kerteszia*, compreendendo 38 indivíduos de diferentes espécies. A maior parte destes indivíduos, pertence a espécie *An. cruzii* s.l., o principal vetor da malária na região da Mata Atlântica. A maioria dessas espécies não tinha suas sequências genômicas conhecidas, o que reforça a importância de nossos resultados. Estes dados permitiram aplicar uma abordagem filogenômica para entender melhor a história evolutiva desses anofelinos.

Dos 38 genomas, 36 foram sequenciados com tecnologia Illumina, e nestes o montador Platanus produziu melhores montagens que o SPAdes. O desempenho superior de Platanus provavelmente foi causado pelas altas taxas de heteroziguidade das amostras de campo utilizadas neste estudo (1,3% - 3,0%) que é alta, quando comparada com colônias de insetos de laboratório (heteroziguidade < 1%) (NORRIS et al., 2001; KINGAN et al., 2019; GUAN et

al., 2020). Cabe ressaltar que Platanus foi projetado especificamente para lidar com genomas altamente heterozigotos (> 1%) (KAJITANI et al., 2014), enquanto o SPAdes foi desenvolvido para pequenos genomas (por exemplo, genomas haploides bacterianos) (BANKEVICH et al., 2012).

Nossos resultados sugerem que o tamanho do genoma de *An. cruzii* s.l. é de aproximadamente 170 Mb, o que é semelhante a *Anopheles (Nyshorrynchus) darlingi* (173,9 Mb; MARINOTTI et al., 2013), mas menor que *Anopheles (Cellia) gambiae* (278 Mb; HOLT et al., 2002). Deve-se notar que os tamanhos dos genomas calculados a partir de montagens (*An. cruzii* s.l.: ~140 Mb) tendem a ser menores do que aqueles estimados a partir das reads brutas (*An. cruzii* s.l.: ~170 Mb) devido ao colapso de regiões repetitivas durante a montagem. O sequenciamento com Nanopore produziu tamanhos de genoma de 170 Mb e 183 Mb, corroborando as estimativas baseados em reads brutas. E se assemelhando também ao tamanho do genoma de *An. cruzii* com reads longas (~184 Mb) relatado pelo projeto *Anopheles Reference Genomes* (acesso: GCA_943734635.1). Isso fornece suporte adicional para a confiabilidade de nossos dados de tamanho do genoma apresentados.

Quanto as estimativas de contiguidade, os genomas Illumina apresentaram valores de N50 variando de ~2,5 kb a ~87 kb. Destes, dez genomas apresentaram valor de N50 de mais de 37 kb, sendo próximos do que os obtidos para *An. darlingi* (N50 = ~38 kb) e *An. gambiae* s.s. (N50 = ~42 kb) (LAWNICZAK et al., 2010; MARINOTTI et al., 2013).

Os valores de N50 estimados pelo QUAST para os genomas de sequências longas (*Kerteszia* Flo F3 N e *Kerteszia* Boc F3 N) foram comparativamente menores do que os do genoma GCA_943734635.1. Especificamente, os valores do N50 para nossos genomas de sequências longas variaram de 447 a 1381 kb, enquanto o GCA_943734635.1 apresentou um valor muito maior de 62076 kb. Vale ressaltar que essa disparidade pode ser atribuída ao fato de que o genoma GCA_943734635.1 foi sequenciado utilizando várias tecnologias, incluindo dados Genomics Chromium, dados Arima Hi-C e sequenciamento PacBio HiFi.

Além dos índices de contiguidade, também foi estimada a completude nos genomas sequenciados neste estudo. Das 36 amostras sequenciadas, 23 apresentaram um alto número de ortólogos completos encontrados usando

BUSCO v4.1.4 (90% - ~99%), o que está próximo de conjuntos de leituras longas de *Anopheles coluzzi* e *Anopheles arabiensis* (~99%) (ZAMYATIN et al., 2021), e *An. cruzii*, *An. bellator* (~99% e ~98%: GCA_943734635.1 e GCA_943735745.1, respectivamente), apontando boa qualidade das sequências produzidas neste estudo.

A maior parte das montagens com baixos valores de completude foram sequenciadas com bibliotecas Nextera, que é a única opção disponível quando temos uma pequena quantidade de DNA que é insuficiente para o sequenciamento Illumina TruSeq Nano. De acordo com as instruções da MacroGen, bibliotecas TruSeq Nano requirem ao menos 100 ng de DNA, enquanto que Nextera XT requer 1 ng.

***Anopheles (Kerteszia) cruzii* s.l.**

Nossos resultados filogenômicos corroboram estudos anteriores que sugeriram especiação críptica dentro de *An. cruzii* s.l. (RAMIREZ; DESSEN, 2000a, 2000b; CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2004; RONA et al., 2009, 2010; RONA; CARVALHO-PINTO, 2013; DE REZENDE DIAS et al., 2018; KIRCHGATTER et al., 2020).

Genes completos de cópia única identificados com BUSCO v4.1.4 foram utilizados para inferências filogenômicas em duas abordagens distintas: (i) coalescência multiespécies e (ii) concatenação. Embora a concatenação seja um método amplamente utilizado para análise filogenômica, ela pode gerar árvores de espécies discordantes quando há substancial heterogeneidade entre as árvores de genes. Nesses casos, a abordagem de coalescência multiespécies alcança uma árvore de espécies mais precisa (LIU; WU; YU, 2015; DIAS et al., 2020; JIANG; EDWARDS; LIU, 2020; SHEN; STEENWYK; ROKAS, 2021). As duas abordagens foram usadas aqui para inferir as relações filogenéticas em *An. cruzii* s.l.

Estas análises incluíram 40 genomas, com representantes de *An. cruzii* s.l., *An. bellator* s.l. e *An. homunculus* e foram inferidas com (i) os ortólogos presentes em pelo menos 20 dos 40 genomas ou (ii) aqueles presentes em pelo menos 35 dos 40. Todas estas análises mostraram as mesmas topologias e sugerem que *An. cruzii* s.l. é composta por pelo menos quatro espécies crípticas,

que ocorrem em: I) Sul e Sudeste do Brasil, representadas por Florianópolis (SC), Guapimirim (RJ) e Maquiné (RS); II) Itatiaia e Bocaina 2, III) Bocaina 1 e IV) Santa Teresa (ES) (uma discussão mais detalhada sobre esta última será realizada em tópico posterior). Esses quatro grupos e suas relações foram recuperadas tanto nas árvores de espécies concatenadas quanto nas de coalescência multiespécies.

Dias e colaboradores (2018) estudaram 2 genes (*cpr* e *clock*) em diversas populações. Segundo estes autores *An. cruzii* do Sul/Sudeste do Brasil compreende pelo menos duas espécies crípticas: (i) Grupo 1, ocorrendo no sul do Brasil e nas regiões costeiras da Serra do Mar (Florianópolis-SC, Guapimirim-RJ, Tinguá-RJ, e Sana-RJ); e (ii) Grupo 2, ocorrendo na parte continental da Serra do Mar (Bocaina-SP) e da Serra da Mantiqueira (Itatiaia – RJ). Em Bocaina, que fica no meio do caminho entre Itatiaia e a parte costeira da Serra do Mar, ambos os grupos ocorrem em simpatria com baixa frequência de heterozigotos, sugerindo que há isolamento reprodutivo entre eles (DE REZENDE DIAS et al., 2018). Nossos resultados de inferências filogenômicas também suportam essas descobertas, mostrando que os dois grupos da Bocaina permanecem filogeneticamente distantes quando incluímos um número muito maior de genes.

Nossos resultados também mostram que Bocaina Grupo 2 agrupa com Itatiaia e que os indivíduos do Sul do Brasil (Florianópolis-SC) agruparam-se com aqueles da parte costeira da Serra do Mar (Guapimirim-RJ). Mas diferente de Dias et al. (2018), os indivíduos aqui classificados como Bocaina 1 Grupo estão bastante distantes daqueles provenientes das regiões litorâneas Sul e Sudeste. No entanto, essa discrepância não é inesperada, pois Dias et al. (2018) utilizaram apenas dois fragmentos gênicos, tornando seus resultados mais suscetíveis a vieses decorrentes da heterogeneidade das árvores de genes.

Os valores de F_{ST} obtidos neste estudo corroboram as análises filogenéticas. Ao comparar as espécies de *An. cruzii* s.l., observamos baixos valores de F_{ST} entre Bocaina Grupo 2 e Itatiaia (F_{ST} médio ~0,22), o que é consistente com a estreita relação encontrada nas análises filogenômicas desses grupos. Isso também é válido para os grupos simpátricos de Bocaina Grupo 1 e Bocaina Grupo 2, mostrando que a distância geográfica por si só não explica os níveis de diferenciação observados entre esses grupos de anofelinos.

Observamos uma média total de F_{ST} de 0,6 entre esses dois últimos grupos, valor muito acima do limiar de 0,35 que geralmente indica diferentes espécies (HEY; PINHO, 2012). Mas, note na Figura 7, que há grande sobreposição entre o que define populações ou espécies distintas. Ainda assim, operacionalmente utilizamos aqui o limiar sugerido por Hey e Pinho (2012).

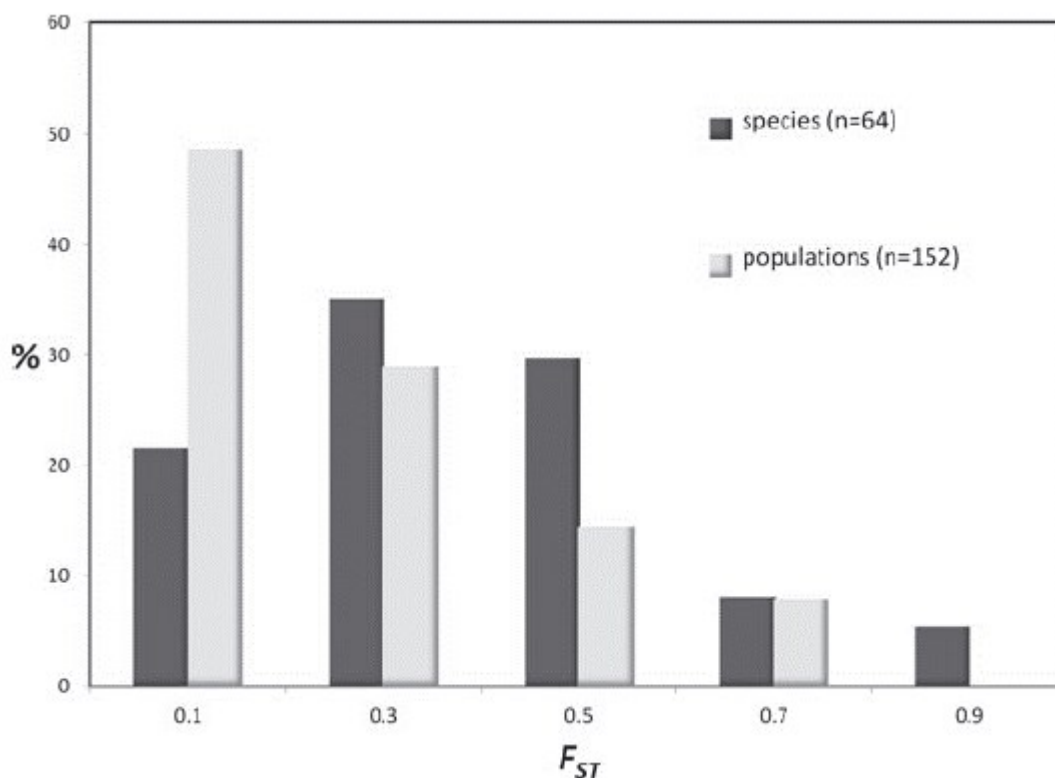


Figura 7: Nos histogramas os valores mostrados são as porcentagens de observações para F_{ST} para espécies e populações.

É importante destacar que a evidência de especiação críptica é particularmente convincente no caso das amostras provenientes de Bocaina: apesar da simpatria, esses grupos se desviam do equilíbrio de Hardy-Weinberg, e apresentam deficiência de heterozigotos, indicando isolamento reprodutivo completo ou quase completo.

Por outro lado, as populações de Florianópolis e Guapimirim, separadas por aproximadamente 1000 km, apresentam baixos valores de F_{ST} (média: 0,29) e podem ser consideradas como uma única espécie.

Os maiores valores de F_{ST} para todas as comparações foram encontrados em loci do cromossomo X (mapeados de acordo com o genoma de *An. cruzii* s.l. GCA_943735745.1 produzido pelo *Anopheles Reference Genomes Project*),

sugerindo que esse cromossomo desempenha um papel importante no processo de diferenciação genética das espécies de *An. cruzii* s.l. Esse mesmo padrão também foi observado para o complexo *An. gambiae* s.l., que possui região no cromossomo X associada ao processo de especiação (ABOAGYE-ANTWI et al., 2015; NEAFSEY et al., 2015; CAMPOS et al., 2021). De fato, estudos genômicos encontraram maior divergência genética em genes do cromossomo X em comparação com aqueles dos autossomos (LEE et al., 2013). Campos et al., (2021) por exemplo, avaliaram a divergência genética entre *An. gambiae* e *An. coluzzii*, e também observaram genes com maiores valores de F_{ST} localizados predominantemente no cromossomo X, principalmente na parte centromérica. Isto pode ser em razão de que em espécies com determinação do sexo X/Y (por exemplo, *Anopheles* spp.), espera-se que os genes do cromossomo X estejam sob seleção mais eficiente (evolução mais rápida do X) do que os autossomos. Isto ocorre devido à expressão de mutações raras e recessivas nos machos (VICOSO; CHARLESWORTH, 2006).

Além disso, as inversões cromossômicas também contribuem para o processo de especiação em dípteros (por exemplo, *Drosophila* spp.), através do desequilíbrio de ligação entre os genes dentro das inversões (NOOR et al., 2001). De fato, alguns complexos de espécies possuem inversões cromossômicas fixas no cromossomo X que distinguem os clados irmãos, como é o caso de *An. cruzii* s.l. (BRANQUINHO et al., 1997; RAMIREZ; DESSEN, 2000a; TURNER; HAHN; NUZHIDIN, 2005).

***Anopheles (Kerteszia) laneanus*:**

Anopheles (Kerteszia) laneanus e *An. cruzii* s.l. são espécies geneticamente muito próximas. Estudos usando marcadores moleculares têm apoiado a hipótese de que *An. laneanus* pertence ao complexo *An. cruzii* (FOSTER et al., 2017) ou que formam um clado irmão (LORENZ; PATANÉ; SUESDEK, 2015; VOGES et al., 2019). Além disso, apesar de *An. cruzii* e *An. laneanus* serem distinguíveis por características morfológicas, como nos tarsos, essas diferenças morfológicas são sutis tornando a identificação destas espécies bastante complexa (FORATTINI, 1962; CONSOLI; LOURENÇO DE OLIVEIRA, 1994). Portanto, a diferenciação morfológica entre eles é mais confiável usando

a genitália masculina (FORATTINI, 1962). Assim, a possibilidade de que uma das espécies dentro do complexo *An. cruzii* s.l., pudesse ser da espécie *An. laneanus* foi considerada e, finalmente, rejeitada após examinar a genitália masculina de espécimes de *An. cruzii* s.l. dos quatro grupos irmãos estudados neste trabalho (não foram observadas diferenças morfológicas). Fizemos uma coleta em Campos do Jordão (SP) em busca de *An. laneanus*; entretanto, não foi possível coletar nenhum indivíduo desta espécie nesta ocasião (CORRÊA; CERQUEIRA, 1944). Todavia, mesmo sem o genoma dessa espécie, conseguimos confirmar que não havia indivíduos dessa espécie em nossas amostras, analisando a genitália masculina.

Já a amostra de Santa Teresa (ES) (*Kerteszia* San M1) apresentou diferenças morfológicas na genitália masculina, mais precisamente no claspeto ventral, e isto será discutido em maiores detalhes no tópico abaixo.

***Kerteszia* San M1: Uma espécie complexa ou parte de um complexo de espécies?**

A amostra de Santa Teresa (ES) (*Kerteszia* San M1) apresentou diferenças morfológicas da genitália masculina no claspeto ventral, que tipicamente tem expansão lateral variando de arredondada a sinuosa na margem lateral, em *An. cruzii* s.l. (SALLUM et al., 2020). E também não se parece com o claspeto ventral de *An. laneanus* que é bastante espiculado medialmente, com muitas espículas curtas estendendo-se lateralmente até a margem lateral (SALLUM et al., 2020). O claspeto ventral de *Kerteszia* San M1 não é arredondado ou sinuoso, mas se assemelha a um taco de golfe, bastante diferente do típico claspeto ventral de *An. cruzii* s.l., bem como das outras espécies de *Kerteszia* (por exemplo: *An. laneanus*) (SALLUM et al., 2020). Cabe destacar, que todas as outras características morfológicas são consistentes com *An. cruzii* s.l. segundo Consoli e Lourenço (1994) e Forattini (2002). Ainda, é importante observar que, com base em nossa análise filogenômica, *Kerteszia* San M1 e *Kerteszia* San F1 (ambas provenientes de Santa Teresa, ES) compreendem um cluster exclusivo dentro do complexo *An. cruzii* s.l., indicando assim uma nova espécie de *Kerteszia*, não identificada até o momento. Todavia, diante destes achados, cabe aqui o seguinte questionamento (previamente

levantado no título deste tópico): *Kerteszia* San M1, seria mesmo uma nova espécie (uma espécie complexa, devido a diferenças morfológicas na genitália masculina) ou parte do complexo *An. cruzii*? Para melhor esclarecer esta questão faz-se importante lembrar que até os dias atuais, o conceito de espécies continua a ser objeto de debates. Efetivamente, existem mais de 30 conceitos alternativos de espécies. Podemos citar o conceito biológico, ecológico, filogenético e muitos outros (ZACHOS, 2016).

Diante disso, o Código Internacional de Nomenclatura Zoológica estabelece uma regra fundamental para a atribuição de nomes a espécies, conhecida como "diagnosticabilidade". De acordo com essa regra, é necessário declarar os caracteres que diferenciam o táxon de outros táxons com os quais é provável que seja confundido (ICZN, 1999).

Desta forma, a morfologia ainda ocupa papel crucial na definição das espécies na taxonomia (DANTAS-TORRES, 2018). Entretanto, tal fato pode representar um desafio para espécies crípticas, resultando frequentemente na impossibilidade de atribuir-lhes uma designação formal (STRUCK et al., 2018).

Além disso, assim como o conceito de espécies, o próprio conceito de complexo de espécies crípticas também pode ser controverso e ter múltiplas definições (KORSHUNOVA et al., 2019). Mayr (1970) por exemplo, define espécies crípticas como populações naturais que apresentam semelhanças morfológicas ou são idênticas entre si, mas são reprodutivamente isoladas. Este termo também tem sido utilizado para se referir a táxons que não podem ser diferenciados com base em características morfológicas, mas existem evidências que sugerem que eles estão seguindo trajetórias evolutivas distintas (STRUCK et al., 2018). Bickford e colaboradores (2007) definem espécies crípticas de modo semelhante, como: diferentes espécies que são erroneamente classificadas sob única definição. Esse conceito, que tem como base princípios taxonômicos, é utilizado para ressaltar que as espécies crípticas são morfologicamente indistinguíveis umas das outras. E ainda, uma quarta definição é que estas espécies crípticas podem apresentar pouca diferenciação morfológica, mas considerável divergência genética (STRUCK et al., 2018).

Diante dessa miríade de definições optamos aqui por adotar este último conceito. Os três clados de *An. cruzii* para os quais nenhuma diferença morfológica foi encontrada (Flo-Gua, Boc2 e Boc2-Itt) são obviamente espécies

crípticas. Quanto ao clado de Santa Teresa (San M1 e San F1), este apresenta diferenciação apenas na genitália masculina, e nos demais caracteres morfológicos não foi observada diferenciação, sendo ainda, as fêmeas indiferenciáveis das outras espécies do complexo *An. cruzii* s.l. Todavia, as amostras provenientes de Santa Teresa apresentam significativa diferenciação genética, como pode ser observado nas filogenias e nos altos valores de F_{ST} . Assim, consideramos San M1 como uma nova espécie críptica, parte do complexo *An. cruzii* s.l.

Como suporte para a classificação adotada acima, podemos citar, exemplos de outros complexos no qual os membros apresentam diferenças morfológicas. O Complexo de *Anopheles Nuneztovari* (Diptera: Culicidae) por exemplo, inclui quatro espécies: *An. nuneztovari* Gabaldon, *An. goeldii* Rozeboom & Gabaldon, *An. dunhami* Causey e *An. nuneztovari species A*, que apresentam diferenças no comprimento do folíolo do edeago (uma parte da genitália masculina) (DOS SANTOS et al., 2019). Um outro exemplo em Diptera é *Lutzomyia cruzii*, que vários autores consideram pertencer ao complexo *Lu. Longipalpis*, mesmo com os machos dessas espécies apresentando pequenas diferenças na genitália (YOUNG; DUNCAN, 1994). Todavia, ainda há debate sobre o *status* taxonômico de *Lu. Longipalpis*. Alguns autores argumentam que *Lu. longipalpis* é uma única espécie, enquanto outros sugerem que representa um complexo de espécies irmãs (HAMILTON et al., 1996; MUKHOPADHYAY et al., 1998; LAMPO et al., 1999; ARRIVILLAGA et al., 2003; BAUZER et al., 2007).

Adicionalmente, também já foram encontradas diferenças na genitália masculina nos complexos de espécies de *Drosophila*. Um exemplo é o subgrupo *willistoni*, que é um complexo de seis espécies crípticas: *D. willistoni*, *D. paulistorum*, (esta constituída por várias semiespécies), *D. equinoxialis*, *D. tropicalis*, *D. insularis* e *D. pavlovskiana* (EHRMAN; POWELL, 1982). Estas espécies podem ser identificadas através da morfologia da genitália masculina (SPASSKY, 1957).

***Anopheles (Kerteszia) bellator* e *Anopheles (Kerteszia) homunculus*:**

Durante o início dos anos 1940 até os anos 1970, *An. cruzii* s.l., *An. bellator* e *An. homunculus*, que ocorrem em simpatria na Mata Atlântica, foram os principais vetores da "malária das bromélias" neste bioma, causando significativo impacto na saúde pública (RACHOU, 1958; DEANE, 1988; MARRELLI et al., 2007). Mas apesar de sua importância epidemiológica juntamente com o *An. cruzii* s.l., existem poucos estudos sobre *An. bellator* e *An. homunculus*. Alguns estudos sugeriram que *An. bellator* pode fazer parte de um complexo de espécies (CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2003; VOGES et al., 2019). Voges et al. (2019) usando marcadores moleculares, sugeriram a existência de dois grupos de *An. bellator* na Mata Atlântica brasileira: *An. bellator* A é amplamente distribuído nas regiões Sul e Sudeste (Ilha do Mel-PR, Cananéia-SP e Ilha Grande-RJ), e *An. bellator* B é encontrado no Nordeste do Brasil (Camacan-BA). Usando isoenzimas, Carvalho-Pinto e Lourenço-de-Oliveira (2003), também verificaram que populações de *An. bellator* do sul e sudeste do Brasil (Florianópolis-SC e Cananéia-SP) são geneticamente mais próximas entre si em comparação com a amostra do Nordeste (Itaparica-BA). É importante notar que, embora esses dois estudos tenham sugerido que as populações da Bahia (Camacan e Itaparica) são geneticamente distantes daquelas encontradas nas regiões mais ao Sul do Brasil, a população da Bahia (Camacan) utilizada por Voges et al. (2019) está a 300 Km ao Sul de Itaparica (BA), utilizada por Carvalho-Pinto e Lourenço-de-Oliveira (2003). Sendo assim, uma pergunta surge: *An. bellator* de Camacan e Itaparica pertencem ao mesmo grupo ou são geneticamente distantes? Nossos resultados indicam fortemente que *An. bellator* é composto por pelo menos três espécies distintas. A primeira espécie é encontrada no Sudeste do Brasil, nas regiões de Cananéia (SP) e Ilha Grande (RJ), e provavelmente corresponde à *An. bellator* A, conforme sugerido por Voges et al. (2019). As outras duas espécies são encontradas na Bahia: uma em Camacan e outra em Itaparica.

Surpreendentemente, apesar da considerável distância geográfica entre Itaparica e Ilha Grande (mais de 1300 Km), a espécie de Itaparica apresenta maior proximidade genética com *An. bellator* A do que com a espécie de Camacan, que está a apenas cerca de 300 km de distância. Essa divergência

entre as populações de Camacan e Itaparica também é corroborada pelo alto valor de F_{ST} , com uma média de 0,79 na comparação entre essas duas localidades na Bahia. Por outro lado, a comparação entre Itaparica e Ilha Grande exibe um valor de F_{ST} mais baixo, em torno de 0,55. Então, como foi encontrado para *An. cruzii* s.l., a distância geográfica por si só não explica os níveis de diferenciação observados entre os grupos de *An. bellator*, pois Itaparica e Camacan distam apenas ~300 Km uma da outra, e formam dois grupos distantes filogeneticamente.

Intrigantemente observamos um padrão diferente para o gráfico de F_{ST} de *An. bellator* na comparação entre Ilha Grande e Itaparica. Observase – se que alguns genes demonstram valores elevados, enquanto outros exibem valores menores, distribuídos de maneira irregular, padrão similar não ocorreu em nenhuma das outras comparações. É possível que este padrão seja decorrente a inversões cromossômicas em alguma destas populações. Uma vez que, Ramírez e Dessen (RAMÍREZ; DESSEN, 1996) já haviam relatado que apesar de muito mais raras do que em *An. cruzii*, inversões também podem ser encontradas em *An. bellator*.

Quanto a *An. homunculus*, os resultados obtidos aqui diferem dos encontrados para *An. cruzii* s.l. e *An. bellator*. As amostras de *An. homunculus* utilizadas neste estudo apresentaram baixa distância genética e podem ser consideradas uma única espécie ao longo de toda a Mata Atlântica, desde Florianópolis (SC, no Sul do Brasil), até Santa Teresa (ES, no Sudeste do Brasil), e Camacan (BA, no Nordeste do Brasil) (o que está em acordo com o valor de F_{ST} de 0,20 entre as populações de *An. homunculus* de Camacan e Santa Teresa). Esses resultados em conjunto corroboram as descobertas de Cardoso et al. (2012) usando marcadores moleculares.

A pequena diferenciação genética encontrada para *An. homunculus* na filogenia pode ser devido à variação intraespecífica, como esperado que ocorra entre indivíduos da mesma espécie (ROCHES et al., 2018a, 2021a), como também foi visto para *An. cruzii* s.l. do Sul e Sudeste do Brasil, aqui representadas por Florianópolis (SC), Guapimirim (RJ) e Maquiné (RS). De fato, espera-se que populações coespecíficas geograficamente distantes apresentem algum nível de diferenciação genética devido à deriva genética, adaptação local, etc. (VIOLLE et al., 2012; ROCHES et al., 2018b). No entanto, observe que *An.*

cruzii s.l. de Florianópolis (SC) e Guapimirim (RJ) ou *An. homunculus* de Florianópolis (SC) e Camacan (BA) que estão geograficamente muito distantes (~1.000 km e 1.500 km, respectivamente) apresentam apenas níveis moderados de diferenciação na filogenia, enquanto *An. cruzii* s.l. de Bocaina Grupo 1 e Bocaina Grupo 2 (simpátricos) formam dois grupos bastante distantes filogeneticamente.

Tempos de divergência e hipóteses para especiação

Os resultados de datação obtidos neste estudo indicam que as populações de Itatiaia e Bocaina Grupo 2 (consideradas como pertencentes à mesma espécie) divergiram das populações do Sul e Sudeste do Brasil há aproximadamente ~1,25 milhões de anos.

Rona et al. (2013) também estimaram os tempos de divergência para as espécies do complexo *An. cruzii*. Estes autores apontam que o evento de cladogênese que teria separado as populações de Itatiaia de populações localizadas no Sul do Brasil ocorreu há ~0,75 milhões de anos (variando de 0,51 a 1,1 milhões de anos).

A hipótese levantada por estes autores, é que estes tempos de divergência entre essas espécies crípticas, correspondem a alterações climáticas ocorridas no Pleistoceno. Nesse período, glaciações resultariam em uma redução da temperatura global, diminuição do volume das chuvas, períodos mais secos e conseqüentemente, à fragmentação da Mata Atlântica. Ainda, de acordo com Carnaval et al (2009), alguns locais, como na Bahia e Sudeste, teriam atuado como refúgios durante esses períodos de resfriamento, preservando a floresta. Assim, a fragmentação da floresta teria propiciado a origem de espécies endêmicas da Mata Atlântica, como *An. cruzii* e outros animais como anfíbios (CARNAVAL et al., 2009) e ainda, de outros mosquitos (LOAIZA et al., 2012).

Nossa datação (1,25 milhões de anos) difere de Rona et al (2013) (0,75 milhões de anos, com intervalo de 0,51 a 1,1 milhões de anos). Essa discrepância pode ser atribuída a diferenças nas abordagens adotadas nos dois

estudos. Os dois estudos usaram taxas de evolução molecular de *Drosophila*, obtidas, porém, de formas diferentes.

Enquanto este trabalho utilizou uma calibração baseada em fósseis de *Drosophila*, Rona et al. (2013) utilizou taxas de substituição sinônimas e não sinônimas de *Drosophila* para genes nucleares (LI, 1997). Estas taxas estimadas por Li (1997) foram calculadas com base no tempo de divergência entre *D. melanogaster* e *D. obscura*, que foi inicialmente estimada em 30 milhões de anos (MORIYAMA; GOJOBORI, 1992). Todavia, existe certo debate sobre o tempo de divergência destas espécies: Russo et al. (1995) por exemplo, aponta que esta divergência tenha ocorrido a 25 milhões de anos. Russo et al (1995) calibraram suas análises usando informações sobre a emergência das ilhas do Havaí. No entanto, alguns autores argumentam que essa calibração pode introduzir vieses ao fornecer tempos mais antigos em comparação com estimativas de tempo baseadas em fósseis. Isso se deve ao fato de que as *Drosophilas* podem não ter colonizado as ilhas imediatamente após sua formação (OBBARD et al., 2012). Por este motivo, a calibração baseada em fósseis forneceria valores mais precisos de tempo de divergência.

Neste contexto, empregamos no presente estudo uma taxa de substituição de 19,9 substituições/kbp/milhão de anos, estimada para a 3ª posição do códon de drosofilídeos por Dias et al. (dados não publicados). Essa taxa foi estimada no programa MCMCTree (REIS; YANG, 2011) a partir de uma calibração com fósseis.

Outra razão para discrepância entre nossos resultados pode ser atribuída ao número de genes utilizados. Enquanto Rona (2013) utilizou seis genes, neste estudo foram utilizados cerca de 2000 genes. É sabido que heterogeneidade de histórias evolutivas pode ocorrer entre os genes de um mesmo genoma. Dois fenômenos que podem contribuir para essa heterogeneidade são a introgressão e "incomplete lineage sorting". Esses processos podem levar a diferentes trajetórias evolutivas entre os genes de um mesmo genoma, e consequentemente em divergências nos resultados obtidos em relação à datação (TAJIMA, 1983; PAMILO; NEI, 1988; LEACHÉ et al., 2014).

Como mencionado acima, nossos resultados sugerem que Itatiaia e Bocaina tenham divergido das populações do Sul e Sudeste do Brasil há ~1,25 Mya. Neste mesmo período, a aproximadamente há 1,25 milhões de anos,

ocorreu o início um processo chamado de transição do Pleistoceno Médio (MPT), marcada pela mudança dos ciclos glaciais do Pleistoceno, que passaram de uma periodicidade de 41 mil anos para cerca de 100 mil anos. Ainda existe considerável debate na área da climapaleontologia sobre a causa dessa alteração na periodicidade (BERENDS et al., 2021).

Ademais, a compreensão exata de como e por que essas mudanças do MTP resultam em distúrbios bióticos generalizados ainda não é clara. No entanto, uma possível explicação pode estar relacionada à escala das perturbações ambientais. Mudanças em larga escala podem exceder as tolerâncias de espécies e ainda, eliminar refúgios (HAYWARD et al., 2012).

Esses fenômenos citados acima poderiam explicar a divergência entre Itatiaia e Bocaina das populações do Sul e do Sudeste. Mas, não podem explicar a simpatria observada na Bocaina.

Anteriormente, de Rezende Dias (2018) utilizando dois genes do relógio biológico, já havia sugerido previamente a existência de duas linhagens distintas na Bocaina. Agora, com um maior número de genes corrobora-se a hipótese de especiação simpátrica. Neste trabalho, das 145 amostras analisadas encontramos apenas 7 amostras da Bocaina grupo 1 e deficiência de heterozigotos. Mas, o que estaria causando esse processo de especiação?

Observamos essa ausência de heterozigotos para o gene *cpr*, o que indica a ocorrência de uma diminuição ou falta de fluxo gênico. Uma hipótese para essa redução no fluxo gênico pode ser atribuída ao acasalamento seletivo.

Estudos sobre especiação simpátrica sugerem que genes envolvidos no acasalamento seletivo entre espécies em processo de divergência podem estar localizados em regiões genômicas que apresentam supressão da recombinação (como inversões cromossômicas e regiões pericentroméricas) (ABOAGYE-ANTWI et al., 2015).

As espécies *An.coluzzii* e *An.gambiae* s.s., que fazem parte do complexo *An. gambiae*, são simpátricas em várias regiões. Em algumas dessas regiões, uma das principais causas de divergência entre as espécies é o acasalamento seletivo, resultando na ocorrência pouco frequente de híbridos (DABIRE et al., 2013; LEE et al., 2013b). Trabalhos apontam que a divergência entre essas espécies está associada a regiões do genoma com baixa taxa de recombinação, conhecidas como "ilhas de especiação". Essas regiões específicas abrigam

genes responsáveis pelo processo de especiação (TURNER; HAHN; NUZHIDIN, 2005; WHITE et al., 2010).

Estes achados podem ajudar a melhor compreender a simpatria observada nas amostras da Bocaina. Assim como ocorre entre as espécies *An. coluzzii* e *An.gambiae* s.s., as amostras da Bocaina também apresentam divergência genética, ocorrendo em simpatria e exibindo uma ausência de heterozigotos. Além disso, ao analisar os gráficos de F_{ST} , observa-se um padrão de maiores valores de F_{ST} em determinadas regiões dos autossomos, que provavelmente correspondem à região centromérica. Neste cenário, é possível que estas “ilhas de especiação” se encontrem nestas regiões também nas amostras da Bocaina e que esteja ocorrendo alguma forma de acasalamento preferencial.

Um outro aspecto que poderia estar envolvido na simpatria das amostras da Bocaina é a preferência por certos locais de alimentação. *An. cruzii* é descrito essencialmente como sendo acrodendrofílico (ou seja, tem tendência a viver e realizar alimentação nas copas das árvores) (CONSOLI; LOURENÇO DE OLIVEIRA, 1994). Todavia, esses mosquitos também já foram encontrados realizando alimentação próximo ao solo, como ocorre nas cercanias de Joinville (SC), local no qual já foram documentados casos de malária tanto em macacos quanto em seres humanos, Deane et al. (1971) coletaram mais de 40% desses mosquitos próximos ao solo e o restante nas copas das árvores. Em Santa Teresa (ES), Buery et al. (2018) também encontraram a grande maioria destes anofelinos com armadilhas CDC nas copas das árvores. No entanto, os pesquisadores também coletaram um número menor de mosquitos *An. cruzii* utilizando armadilhas Shannon e CDC próximas ao solo.

Diante disso, é plausível considerar que o Bocaina Grupo1 e o Bocaina Grupo2 possam possuir certa preferência por se alimentarem próximos ao solo ou nas copas das árvores, o que pode resultar em encontros menos frequentes entre esses anofelinos e, conseqüentemente, contribuir para a diminuição do fluxo gênico entre eles.

Conclusões

Neste trabalho, ~2.000 genes foram utilizados para avaliar a diferenciação genética de *An. cruzii* s.l., aplicando análise de F_{ST} e inferências filogenômicas. Nossos resultados sugerem fortemente que *An. cruzii* s.l. é um complexo composto por pelo menos quatro espécies crípticas: i) Sul e Sudeste do Brasil (representadas aqui por Florianópolis, Guapimirim e Maquiné), ii) Itatiaia e Bocaina Grupo 2 iii) Bocaina Grupo 1, iv) Santa Teresa. A evidência de especiação críptica é inequívoca no caso das amostras de Bocaina, que são provenientes do mesmo local e ainda assim, formam dois grupos distantes filogeneticamente. Além de *An. cruzii* s.l., sugerimos também que *An. bellator* é um complexo de pelo menos três espécies: A primeira espécie é encontrada no Sudeste do Brasil, nas regiões de Cananéia (SP) e Ilha Grande (RJ), e as outras duas espécies são encontradas na Bahia: uma em Camacan e outra em Itaparica. Por outro lado, as populações de *An. homunculus* não apresentaram separação genética clara e podem ser consideradas uma única espécie distribuída do sul ao nordeste do país.

Perspectivas

- Incluir amostras de *Anopheles laneanus*.
- Incluir mais amostras da Bocaina.
- Identificar e anotar os genes ligados ao Y.

Material Suplementar

Tabela suplementar 1: Informações detalhadas sobre as amostras utilizadas neste estudo.

ID da amostra	Nome da amostra	Sexo	Cidade	Estado	Data	Coleta	Latitude	Longitude
An. bellator Ilh F1	boni2	F	Ilha Grande	RJ	02 . aug . 2019	Bromelia	23°07'04.5"S	44°16'29.9"W
An. bellator Ilh M1	Boni_8	M	Ilha Grande	RJ	02 . aug . 2019	Bromelia	23°07'04.5"S	44°16'29.9"W
An. bellator Itp F1	Itaparica1-2_F	F	Itaparica	BA	04 . may . 2019	Bromelia	13°03'36.1"S	38°46'35.1"W
An. bellator Itp M1	Itaparica_1_M	M	Itaparica	BA	04 . may . 2019	Bromelia	13°03'36.1"S	38°46'35.1"W
An. bellator Cam F11	Cam F11	F	Camacan	BA	17 . dez . 2022	CDC + CO2	15°23'30.9"S	39°33'54.6"W
An. bellator Cam F14	Cam F14	F	Camacan	BA	17 . dez . 2023	CDC + CO2	15°23'30.9"S	39°33'54.6"W
An. bellator Cam F25	Cam F25	F	Camacan	BA	17 . dez . 2024	CDC + CO2	15°23'30.9"S	39°33'54.6"W
An. bellator Cam F33	Cam F33	F	Camacan	BA	17 . dez . 2026	CDC + CO2	15°23'30.9"S	39°33'54.6"W
An. bellator Cam F34	Cam F34	F	Camacan	BA	17 . dez . 2027	CDC + CO2	15°23'30.9"S	39°33'54.6"W
An. bellator Cam F36	Cam F36	F	Camacan	BA	17 . dez . 2028	CDC + CO2	15°23'30.9"S	39°33'54.6"W
An. bellator GCA_943735745.1	An. bellator Can	-	Cananéia	SP	20 . oct . 2007	.	25°1'0.001"S	47°57'0.00"W
An. homunculus Flo F1	c40_UCAD	F	Florianópolis	SC	24 . aug . 2019	CDC + CO2	27°31'52.2"S	48°30'44.2"W
An. homunculus San F2	Vv_bio3	F	Santa Teresa	ES	25 . jul . 2019	Bromelia	19°57'54.6"S	40°32'24.2"W
An. homunculus San M2	Vv_bio1	M	Santa Teresa	ES	25 . jul . 2019	Bromelia	19°57'54.6"S	40°32'24.2"W
An. homunculus Cam F1	Cam F1	F	Camacan	BA	17 . dez . 2021	CDC + CO2	15°23'30.9"S	39°33'54.6"W
An. homunculus Cam F31	Cam F31	F	Camacan	BA	17 . dez . 2025	CDC + CO2	15°23'30.9"S	39°33'54.6"W
Kerteszia Boc F1	boc104	F	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc F2	boc140	F	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc M1	boc56	M	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc M2	Boc146	M	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc M3	boc160	M	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc M4	boc66	M	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc M5	boc85	M	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc M6	boc86	M	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Boc F3 N	Boc F3 N	F	Bocaina	SP	27 . feb . 2019	Bromelia	22°43'9.95"S	44°36'47.1"W
Kerteszia Flo F2	UCAD 7.5	F	Florianópolis	SC	01 . fev . 2019	Família	27°31'52.2"S	48°30'44.2"W
Kerteszia Flo F3 N	Flo F3 N	F	Florianópolis	SC	01 . fev . 2019	Família	27°31'52.2"S	48°30'44.2"W
Kerteszia Flo M1	UCAD13-7m	M	Florianópolis	SC	01 . fev . 2019	Família	27°31'52.2"S	48°30'44.2"W
Kerteszia Flo M2	UCAD 10.11	M	Florianópolis	SC	01 . fev . 2019	Família	27°31'52.2"S	48°30'44.2"W
Kerteszia Flo M3	UCAD 6.2	M	Florianópolis	SC	01 . fev . 2020	Família	27°31'52.2"S	48°30'44.2"W
Kerteszia Gua F1	Guapi F1	F	Guapimirim	RJ	01 . oct . 2019	Bromelia	22°30'22.2"S	43°00'59.2"W
Kerteszia Gua F2	Gua F2	F	Guapimirim	RJ	01 . oct . 2019	Bromelia	22°30'22.2"S	43°00'59.2"W
Kerteszia Itt F1	M_lago_8_1_F	F	Itatiaia	RJ	04 . jul . 2019	Bromelia	22°27'07.3"S	44°36'44.9"W
Kerteszia Itt M1	CdoPesq1	M	Itatiaia	RJ	04 . jul . 2019	Bromelia	22°27'07.3"S	44°36'44.9"W
Kerteszia Itt M2	M_lago_11_1_M	M	Itatiaia	RJ	04 . jul . 2019	Bromelia	22°27'07.3"S	44°36'44.9"W
Kerteszia Itt M3	M_lago 2.1	M	Itatiaia	RJ	04 . jul . 2019	Bromelia	22°27'07.3"S	44°36'44.9"W
Kerteszia Itt M4	M_lago 12.1	M	Itatiaia	RJ	04 . jul . 2019	Bromelia	22°27'07.3"S	44°36'44.9"W
Kerteszia San F1	Pousada_1	F	Santa Teresa	ES	25 . jul . 2019	Bromelia	19°57'54.6"S	40°32'24.2"W
Kerteszia San M1	ResSLBf3_1	M	Santa Teresa	ES	25 . jul . 2019	Bromelia	19°57'54.6"S	40°32'24.2"W
Kerteszia GCA_943735745.1	Maq F	F	Maquiné	RS	16 . jan . 2011	.	29°35'22.4"S	50°15'45.5"W

Sexo: sexo de cada amostra – M para machos e F para fêmeas. Fonte: Bromélia - imaturos coletados na água das bromélias e criados em laboratório; CDC + CO2 - adultos coletados com armadilhas CDC com CO2; Família - geração F1 de mosquitos fêmeas coletadas no campo. Flo: Florianópolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica.

Tabela suplementar 2: Resultados de Busco V3 comparando os montadores SPAdes e Platanus em sete amostras.

ID da amostra	Nome da amostra	Montador	Completo	Completo cópia única	Completo duplicado	Fragmentado	Ausente	Tipo de arquivo
An. bellator Itp F1	Itaparica1-2_F	Spades	2343 (83.7)	2330 (83.2)	13 (0.5)	325 (11.6)	131 (4.7)	C
		Platanus	2689 (96.1)	2681 (95.8)	08 (0.3)	59 (2.1)	51 (1.8)	C
An. bellator Ilh M1	Boni_8	Spades	2314 (82.7)	2300 (82.2)	14 (0.5)	354 (12.6)	131 (4.7)	B
		Platanus	2621 (93.7)	2613 (93.4)	08 (0.3)	108 (3.9)	70 (2.4)	B
Kerteszia Boc M1	boc56	Spades	1717 (61.4)	1668 (59.6)	49 (1.8)	740 (26.4)	342 (12.2)	C
		Platanus	2654 (94.9)	2644 (94.5)	10 (0.4)	93 (3.3)	52 (1.8)	C
Kerteszia Boc M2	Boc146	Spades	2081 (74.3)	2061 (73.6)	20 (0.7)	487 (17.4)	231 (8.3)	C
		Platanus	2665 (95.2)	2657 (94.9)	08 (0.3)	80 (2.9)	54 (1.9)	C
Kerteszia Flo M1	UCAD13-7m	Spades	1747 (62.4)	1704 (60.9)	43 (1.5)	731 (26.1)	321 (11.5)	C
		Platanus	2652 (94.8)	2642 (94.4)	10 (0.4)	94 (3.4)	53 (1.8)	C
Kerteszia Itt M1	CdoPesq1	Spades	1654 (59.1)	1631 (58.3)	23 (0.8)	777 (27.8)	368 (13.1)	B
		Platanus	2382 (85.1)	2376 (84.9)	06 (0.2)	266 (9.5)	151 (5.4)	B
Kerteszia Itt M2	M_lago_11_1_M	Spades	1668 (59.6)	1634 (58.4)	34 (1.2)	777 (27.8)	354 (12.6)	B
		Platanus	2637 (94.2)	2632 (94.0)	05 (0.2)	107 (3.8)	55 (2.0)	B

Anotação automática de 2799 ortólogos de cópia única do conjunto de referência de ortólogos Diptera (odb9, baixado de <https://BUSCO.ezlab.org/>). O número de genes com sequências completas, genes duplicados, genes com sequências fragmentadas, ou genes ausentes em cada genoma é mostrado nesta tabela. Os valores entre parênteses são a porcentagem de BUSCOs em cada montagem. Tipo de arquivo: arquivo antes ou depois de remover os contaminantes – B; arquivos onde os contaminantes ainda não foram removidos e C; arquivos nos quais os contaminantes já foram removidos. Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Ilh: Ilha Grande; Itp: Itaparica. M: machos; F: fêmeas.

ID da amostra	Nome da amostra	Tamanho total	# contigs	Maior contig	GC (%)	N50
An. bellator Ilh F1	boni2	128742914	25396	148881	49.59	9518
An. bellator Ilh M1	Boni_8	151900482	9880	380859	49.01	37061
An. bellator Itp F1	Itaparica1-2_F	160130369	7377	587565	48.85	87751
An. bellator Itp M1	Itaparica_1_M	155423177	7694	612450	48.92	62214
An. bellator Cam F11	Cam F11	146891939	14660	241254	48.89	21793
An. bellator Cam F14	Cam F14	145526116	14663	257829	48.93	21231
An. bellator Cam F25	Cam F25	141022429	18377	220071	49.06	14990
An. bellator Cam F33	Cam F33	233797627	143877	143376	47.64	2532
An. bellator Cam F34	Cam F34	147865918	13852	379019	48.84	23972
An. bellator CamF36	Cam F36	145243795	16091	200488	48.95	18637
An. bellator GCA_943735745.1	An. bellator Can	169575294	2985	85293694	48.78	85293694
An. homunculus Flo F1	c40_UCAD	163810124	8325	482870	49.12	76692
An. homunculus San F2	Vv_bio3	141298410	39013	70947	48.82	8068
An. homunculus San M2	Vv_bio1	132845753	54680	55030	49.01	4473
An. homunculus Cam F1	Cam F1	148530252	18385	174182	49.23	17348
An. homunculus Cam F31	Cam F31	148108851	20086	220308	49.16	16612
Kerteszia Boc F1	boc104	138751828	24166	159463	49.51	10727
Kerteszia Boc F2	boc140	143652064	21663	170163	49.55	13373
Kerteszia Boc M1	boc56	164236966	9911	382945	49.20	60439
Kerteszia Boc M2	Boc146	166174346	8318	495670	49.20	75317
Kerteszia Boc M3	boc160	150640365	15830	190016	49.49	20928
Kerteszia Boc M4	boc66	122343481	33291	133900	49.94	6215
Kerteszia Boc M5	boc85	120988892	34908	91723	49.97	5722
Kerteszia Boc M6	boc86	121318591	34588	136539	49.99	5706
Kerteszia Boc F3 N	Boc F3 N	183033428	2296	13817620	49.07	1381163
Kerteszia Flo F2	UCAD 7.5	144104707	54347	81108	49.32	6283
Kerteszia Flo F3 N	Flo F3 N	170347651	2087	1969744	49.23	446870
Kerteszia Flo M1	UCAD13-7m	165555787	8128	581900	49.09	77815
Kerteszia Flo M2	UCAD 10.11	119500679	34884	103900	49.88	5780
Kerteszia Flo M3	UCAD 6.2	115214550	36067	91655	49.91	5349
Kerteszia Gua F1	Guapi F1	164019820	8833	441808	49.09	67669
Kerteszia Gua F2	Gua F2	137993792	26022	353605	49.39	9836
Kerteszia Itt F1	M_lago_8_1_F	163573267	9429	431426	49.21	65949
Kerteszia Itt M1	CdoPesq1	145297595	21776	194459	49.43	12956
Kerteszia Itt M2	M_lago_11_1_M	163756844	9303	374593	49.21	62238
Kerteszia Itt M3	M_lago 2.1	107189928	39211	110223	49.96	4328
Kerteszia Itt M4	M_lago 12.1	114856556	36506	125009	50.01	5050
Kerteszia San F1	Pousada_1	130157996	43169	104739	49.49	6590
Kerteszia San M1	ResSLBf3_1	156672719	16060	315229	49.51	21569
Kerteszia GCA_943735745.1	Maq F	184069349	5088	87567467	48.97	62075725

Tabela suplementar 3: estatísticas das montagens de *Kerteszia* produzidas com QUAST.

Comprimento total: número total de bases na montagem. # contigs: número total de contigs na montagem. Maior contig: comprimento do contig mais longo da montagem. GC (%): número total de nucleotídeos G e C na montagem dividido pelo comprimento total. N50: integridade da montagem do genoma (comprimento da sequência do contig mais curto em 50% do comprimento total do genoma). Flo: Florianópolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica. M: machos; F: fêmeas

Tabela suplementar 4: Média dos valores de F_{ST} das comparações entre todas as populações.

População	Média do F_{ST}	N média	F_{ST} ChrX	N ChrX	F_{ST} Chr2	N Chr2	F_{ST} Chr3	N Chr3
BOC1-BOC2	0.625	1.967	0.895	184	0.599	716	0.595	1.034
BOC1-FLO	0.667	1.995	0.917	187	0.655	728	0.633	1.047
BOC1-GUA	0.687	1.589	0.931	178	0.670	565	0.647	823
BOC1-ITT	0.601	1.986	0.867	185	0.581	724	0.570	1.043
BOC1-SAN	0.800	1.183	0.923	111	0.787	435	0.789	611
BOC1-h_CAM	0.782	1.572	0.930	144	0.773	571	0.764	833
BOC1-h_STA	0.824	505	0.963	77	0.785	162	0.807	257
BOC2-FLO	0.482	1.984	0.865	187	0.441	727	0.442	1.038
BOC2-GUA	0.456	1.591	0.867	179	0.395	571	0.409	819
BOC2-ITT	0.227	1.981	0.537	187	0.180	725	0.203	1.036
BOC2-SAN	0.723	1.180	0.917	109	0.693	437	0.711	609
BOC2-h_CAM	0.705	1.567	0.923	145	0.676	572	0.687	826
BOC2-h_STA	0.755	506	0.952	78	0.691	163	0.735	256
FLO-GUA	0.293	1.605	0.691	178	0.248	575	0.239	829
FLO-ITT	0.459	2.001	0.838	188	0.423	734	0.419	1.046
FLO-SAN	0.751	1.193	0.932	111	0.734	441	0.732	616
FLO-h_CAM	0.734	1.583	0.930	146	0.715	578	0.713	836
FLO-h_STA	0.774	511	0.957	78	0.725	165	0.749	259
GUA-ITT	0.427	1.597	0.828	180	0.375	570	0.376	824
GUA-SAN	0.734	967	0.911	110	0.710	343	0.713	496
GUA-h_CAM	0.702	1.283	0.910	141	0.679	461	0.673	665
GUA-h_STA	0.764	414	0.964	75	0.688	130	0.738	203
ITT-SAN	0.701	1.193	0.895	110	0.674	443	0.688	614
b_CAM-BOC1	0.852	1.943	0.957	173	0.839	704	0.843	1.033
b_CAM-BOC2	0.816	1.926	0.952	173	0.796	699	0.807	1.022
b_CAM-FLO	0.830	1.953	0.957	176	0.815	712	0.819	1.033
b_CAM-GUA	0.838	1.551	0.958	169	0.821	551	0.825	809
b_CAM-ITT	0.807	1.947	0.941	174	0.790	708	0.797	1.032
b_CAM-SAN	0.843	1.160	0.943	103	0.828	427	0.838	604
b_CAM-b_IG	0.794	1.700	0.927	147	0.766	613	0.791	910
b_CAM-b_ITA	0.797	1.813	0.937	154	0.767	675	0.795	955
b_CAM-h_CAM	0.827	1.539	0.943	134	0.816	556	0.818	825
b_CAM-h_STA	0.859	487	0.963	68	0.830	155	0.848	255
b_IG-BOC1	0.871	1.725	0.962	155	0.864	624	0.861	916
b_IG-BOC2	0.803	1.714	0.949	155	0.782	617	0.792	912

b_IG-FLO	0.826	1.731	0.959	157	0.814	627	0.813	917
b_IG-GUA	0.830	1.395	0.966	151	0.812	488	0.815	734
b_IG-ITT	0.789	1.732	0.932	156	0.771	626	0.777	920
b_IG-SAN	0.843	1.044	0.921	93	0.825	389	0.843	538
b_IG-b_ITA	0.553	1.672	0.810	149	0.459	608	0.574	886
b_IG-h_CAM	0.817	1.371	0.923	124	0.804	493	0.807	732
b_IG-h_STA	0.855	436	0.966	63	0.801	143	0.859	222
b_ITA-BOC1	0.873	1.833	0.970	161	0.863	681	0.864	962
b_ITA-BOC2	0.807	1.821	0.955	161	0.784	675	0.798	956
b_ITA-FLO	0.830	1.844	0.966	164	0.816	689	0.818	963
b_ITA-GUA	0.833	1.465	0.979	156	0.810	530	0.820	758
b_ITA-ITT	0.791	1.836	0.938	162	0.772	684	0.781	961
b_ITA-SAN	0.846	1.106	0.931	97	0.826	419	0.846	565
b_ITA-h_CAM	0.817	1.451	0.931	129	0.800	536	0.811	765
b_ITA-h_STA	0.857	463	0.972	64	0.800	152	0.861	240
h_CAM-ITT	0.686	1.578	0.902	146	0.659	579	0.667	829
h_CAM-SAN	0.728	968	0.857	93	0.711	363	0.717	494
h_CAM-h_STA	0.202	399	0.393	61	0.068	125	0.231	206
h_STA-ITT	0.737	510	0.932	78	0.683	164	0.711	259
Média total			0.906		0.694		0.708	

Média dos valores de F_{ST} das comparações entre as populações. N: Número de genes. As amostras foram nomeadas da seguinte forma: Amostras do complexo *An. bellator*: "b_ITA": Itaparica (incluindo as amostras *An. bellator* Itp M1, *An. bellator* Itp F1), "b_IG": Ilha Grande (*An. bellator* Ilh M1, *An. bellator* Ilh F1), "b_CAM": Camaca (*An. bellator* CamF11, *An. bellator* CamF14, *An. bellator* CamF25, *An. bellator* CamF33, *An. bellator* CamF34, *An. bellator* CamF36). Amostras do complexo *An. cruzii*: "FLO": Florianópolis (*Kerteszia* Flo F3 N, *Kerteszia* Flo M1, *Kerteszia* Flo M2, *Kerteszia* Flo F2, *Kerteszia* Flo M3), "SAN": Santa Teresa (*Kerteszia* San M1, *Kerteszia* San F1), "GUA": Guapimirim (*Kerteszia* Gua F1, *Kerteszia* Gua F2), "ITT": Itatiaia (*Kerteszia* Itt M2, *Kerteszia* Itt M3, *Kerteszia* Itt M4, *Kerteszia* Itt M1, *Kerteszia* Itt F1), "BOC1": Bocaina Grupo 1 (*Kerteszia* Boc M2, *Kerteszia* Boc M3, *Kerteszia* Boc F2, *Kerteszia* Boc F3 N), "BOC2": Bocaina Grupo 2 (*Kerteszia* Boc F1, *Kerteszia* Boc M1, *Kerteszia* Boc M4, *Kerteszia* Boc M5, *Kerteszia* Boc M6). Amostras da espécie *An. homunculus*: h_CAM: Camaca (*An. homunculus* CamF31, *An. homunculus* CamF1), "h_FLO": Florianópolis (*An. homunculus* Flo F1) "h_STA": Santa Tereza (*An. homunculus* San M2, *An. homunculus* San F2).

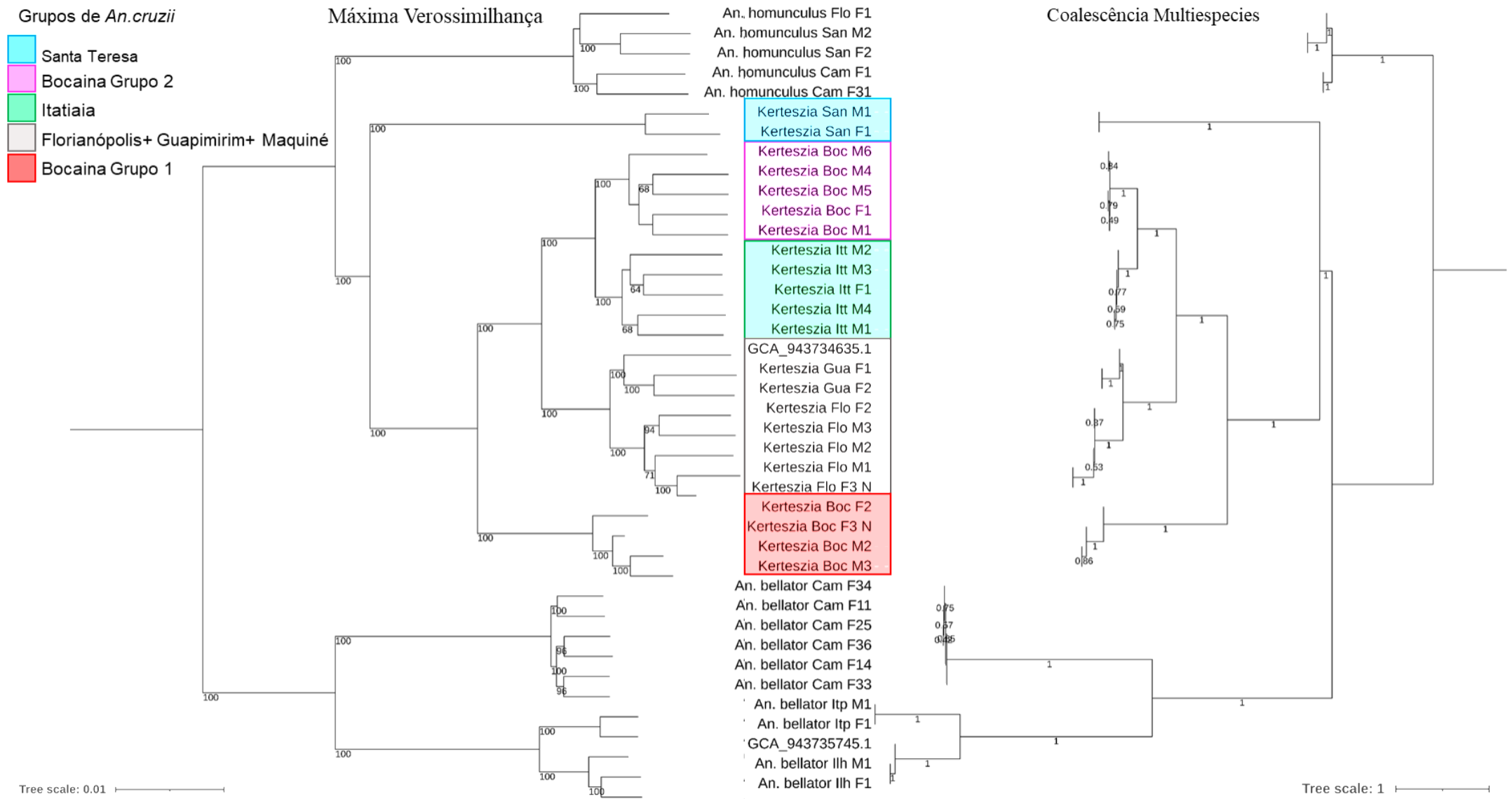
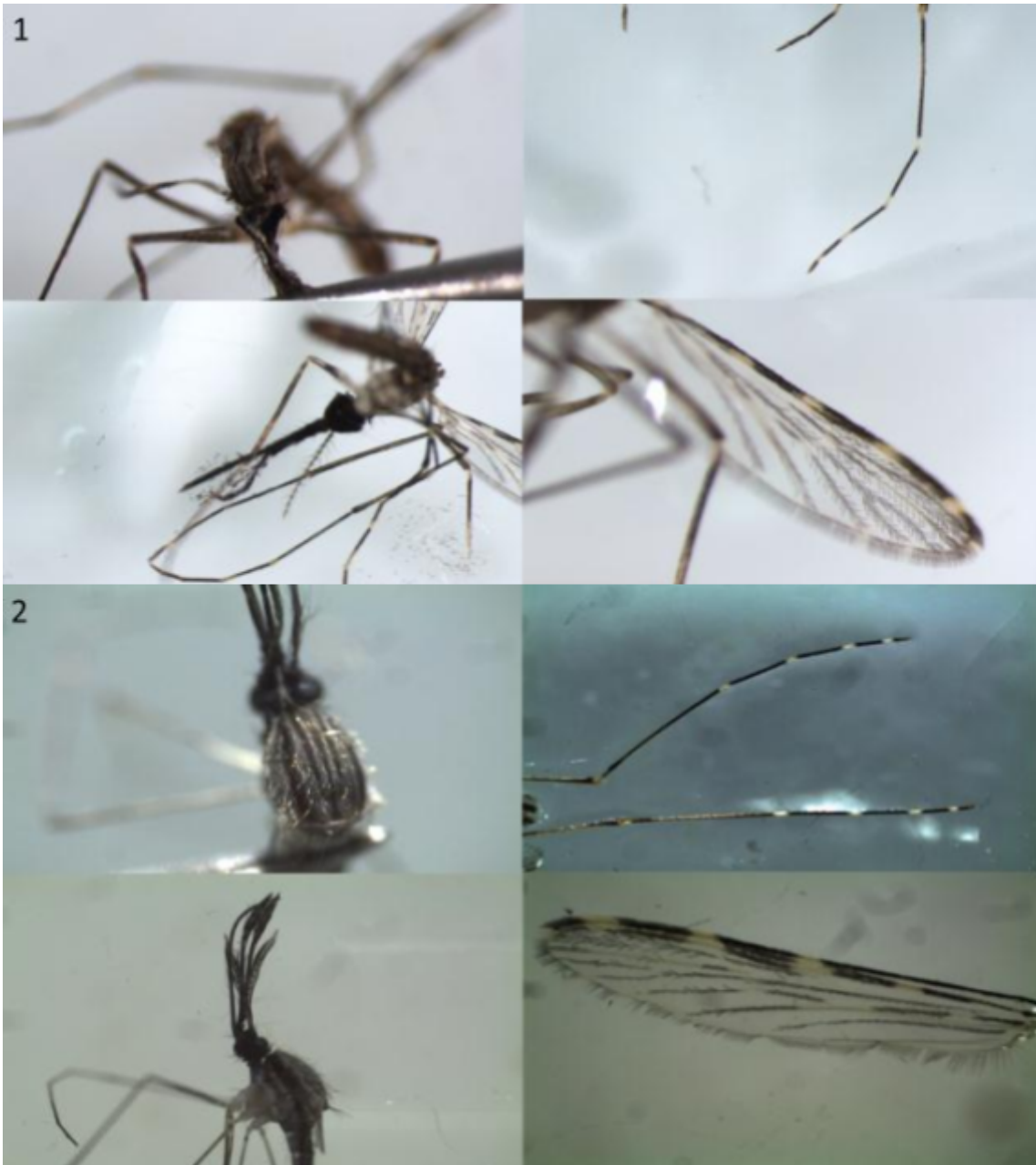


Figura suplementar 1: Árvores ML e MSC: Árvore de espécies inferida com abordagens de Máxima Verossimilhança (ML) e Coalescência Multiespécies (MSC) de 2.051 genes presentes em pelo menos 20 das 40 amostras de *Kerteszia. Anopheles (Kerteszia) bellator* e *Anopheles (Kerteszia) homunculus* são espécies diferentes reconhecidas. Essas análises filogenômicas sugerem fortemente que *An. cruzii* s.l. é composta por pelo menos quatro espécies crípticas: Santa Teresa, Bocaina Grupo 1, Bocaina Grupo 2, Itatiaia e Florianópolis + Guapimirim + Maquiné. A evidência de especiação críptica é especialmente convincente no caso das amostras de Bocaina, que vieram do mesmo local e ainda formam dois grupos bastante distantes filogeneticamente. Os números nos nós representam valores de *bootstrap* (ML) e probabilidades posteriores (MSC). Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; San: Santa Teresa; Itp: Itaparica. M: machos; F: fêmeas.









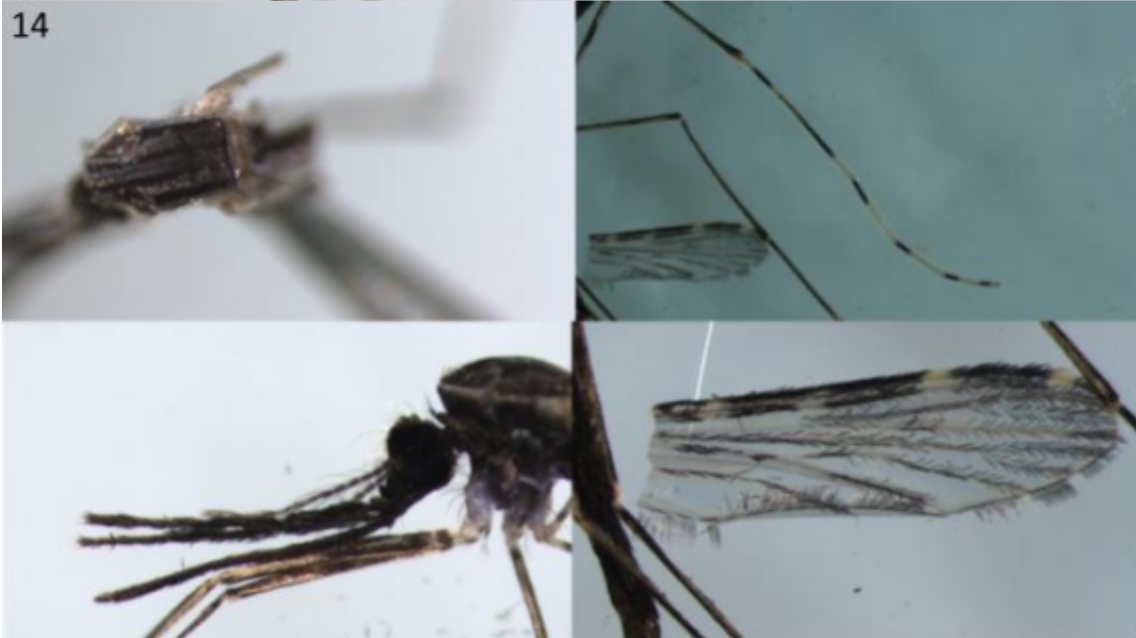




13



14

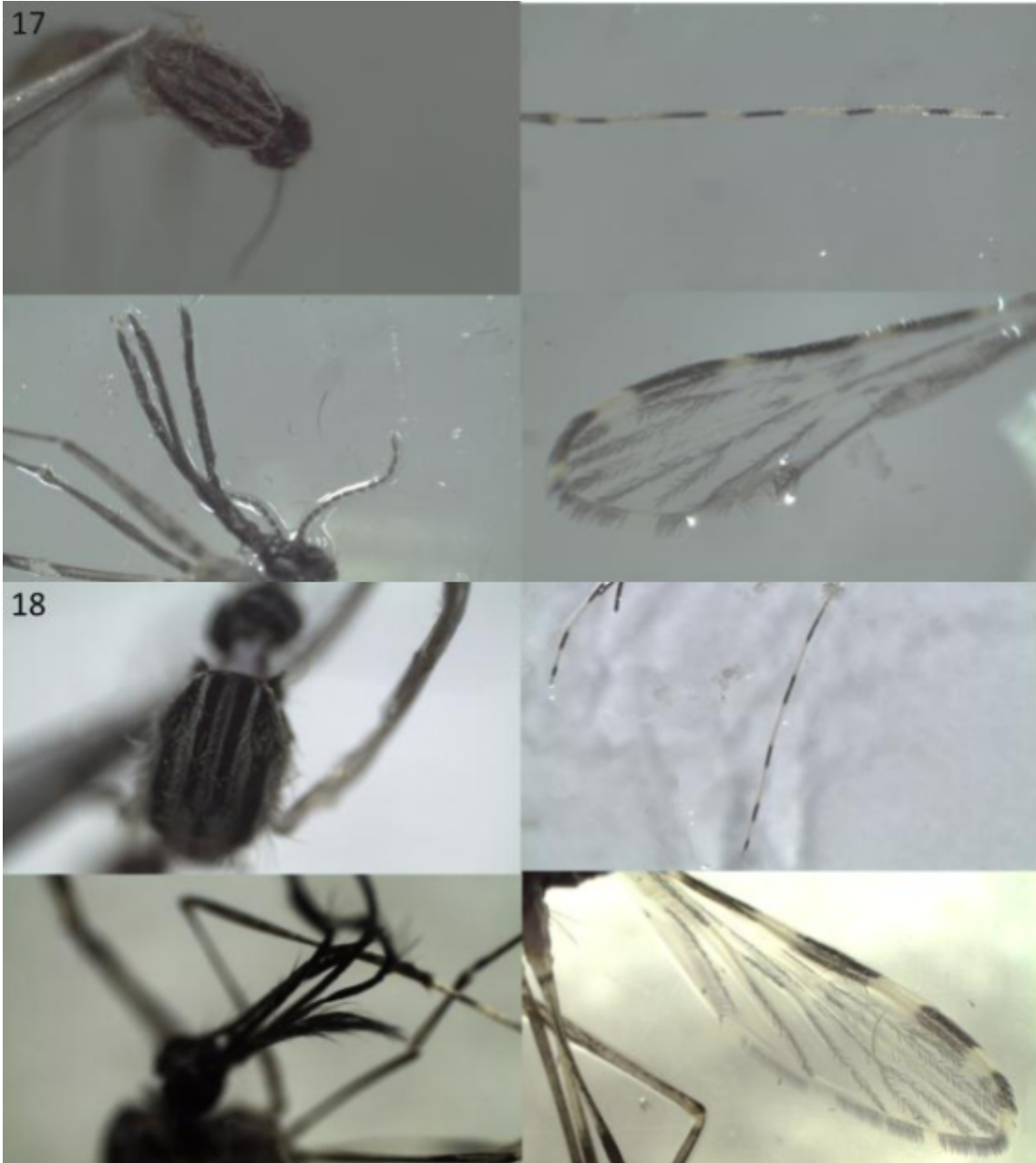


15



16



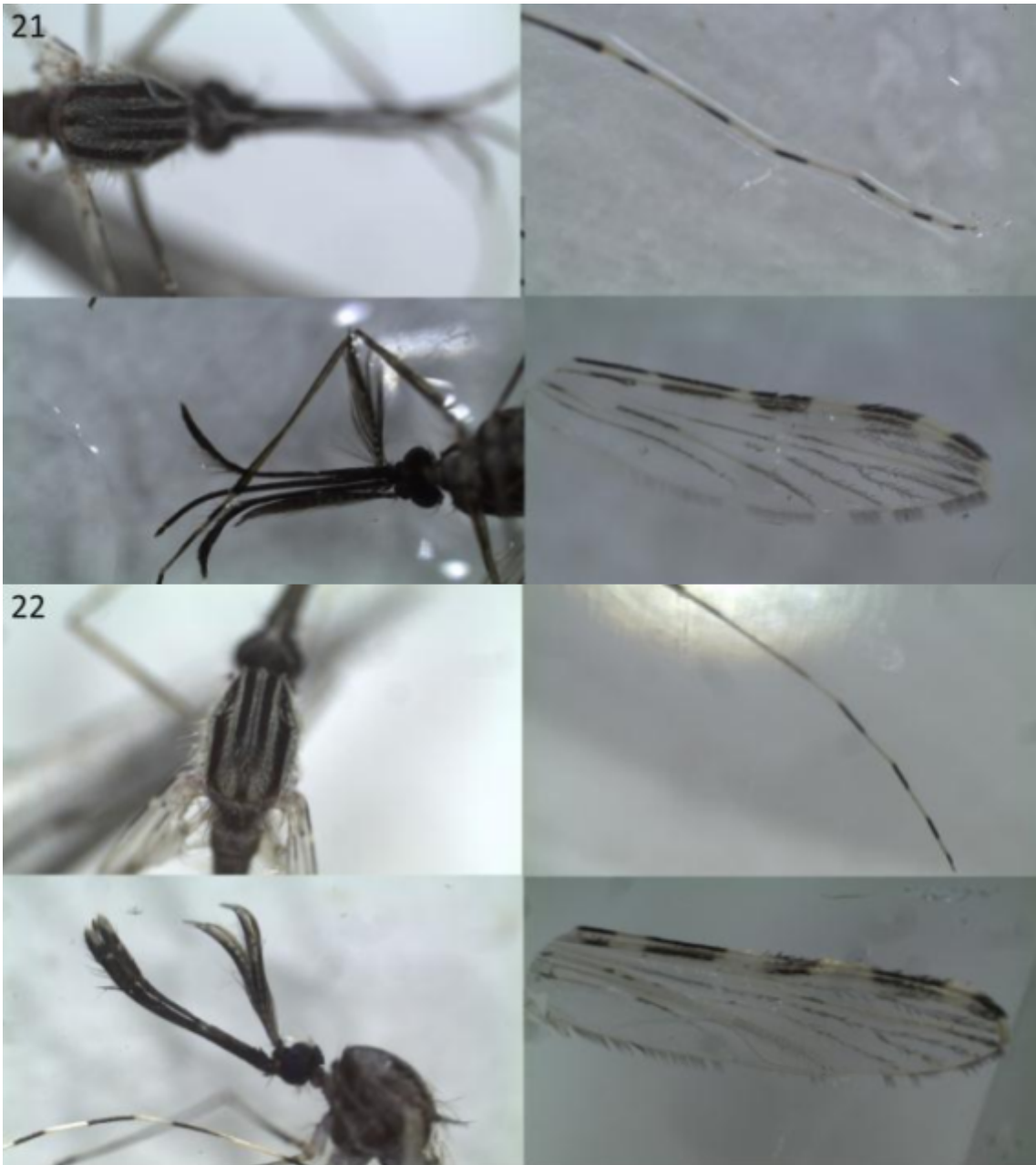


19



20





23



24



25



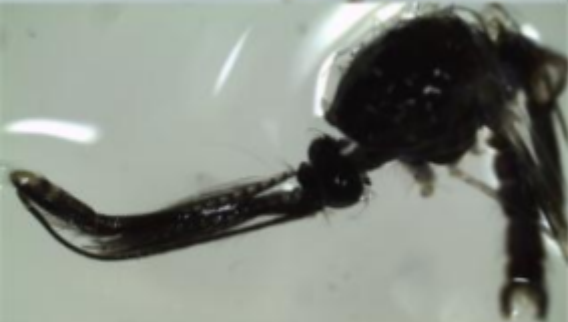
26



27



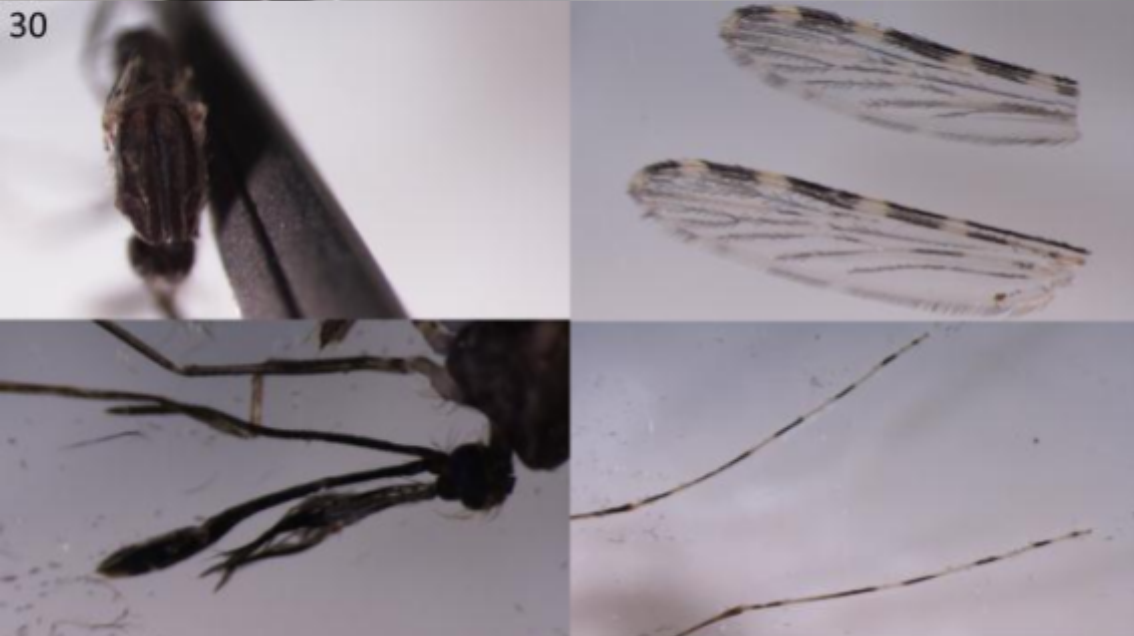
28



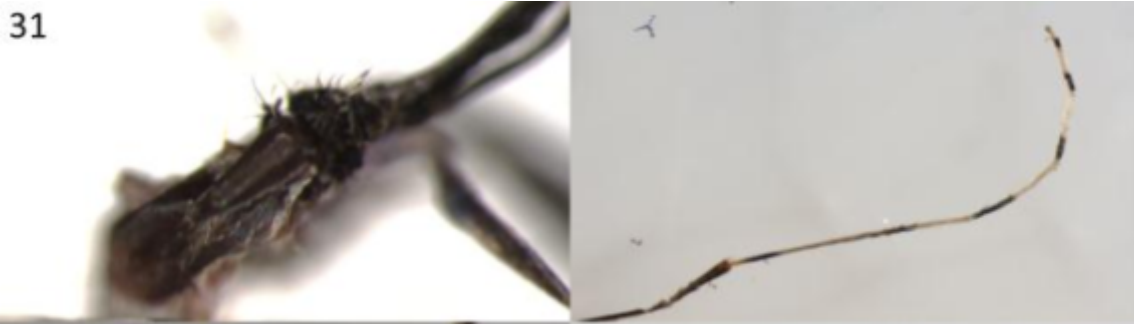
29



30



31



32



33



34



35



36



37



38





Figura suplementar 2: Estruturas morfológicas: Fotografias das estruturas essenciais à identificação ao nível da espécie (cabeça, asa, pata e escudo), das 39 amostras (incluindo as três amostras sequenciadas individualmente com Nanopore): 1- *An. bellator* Ilh F1, 2- *An. bellator* Ilh M1, 3- *An. bellator* Itp F1, 4- *An. bellator* Itp M1), 5- *An. bellator* Cam F11, 6- *An. bellator* Cam F14, 7- *An. bellator* Cam F25, 8- *An. bellator* Cam F33, 9- *An. bellator* Cam F34, 10- *An. bellator* Cam F36, 11- *An. homunculus* Cam F1, 12- *An. homunculus* Cam F31, 13- *An. homunculus* Flo F1, 14- *An. homunculus* San F2, 15- *An. homunculus* San M2, 16- *Kerteszia* Boc F1, 17- *Kerteszia* Boc F2, 18- *Kerteszia* Boc M1, 19- *Kerteszia* Boc M2, 20- *Kerteszia* Boc M3, 21- *Kerteszia* Boc M4, 22- *Kerteszia* Boc M5, 23- *Kerteszia* Boc M6, 24- *Kerteszia* Boc F3 N, 25- *Kerteszia* Flo F3 N_1, 26- *Kerteszia* Flo F3 N_2, 27- *Kerteszia* Flo F2, 28- *Kerteszia* Flo M1, 29- *Kerteszia* Flo M2, 30- *Kerteszia* Flo M3, 31- *Kerteszia* Gua F1, 32- *Kerteszia* Gua F2, 33- *Kerteszia* Itt F1, 34- *Kerteszia* Itt M1, 35- *Kerteszia* Itt M2, 36- *Kerteszia* Itt M3, 37- *Kerteszia* Itt M4, 38- *Kerteszia* San F1, 39- *Kerteszia* San M1, nesta ordem.

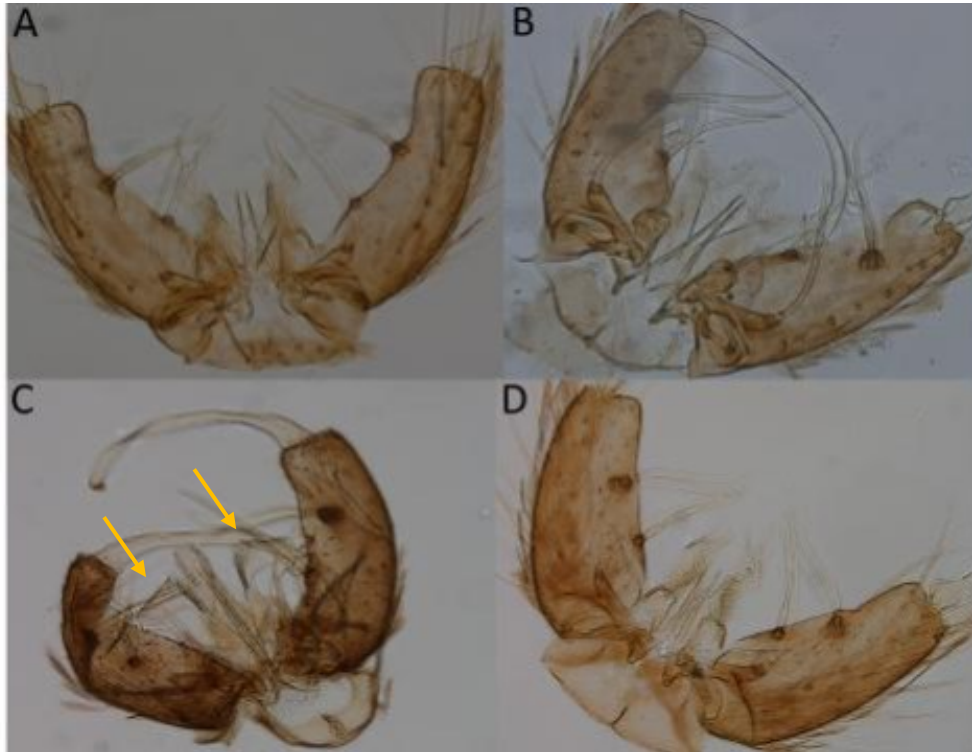


Figura suplementar 3: Caracteres morfológicos da genitália masculina: Imagens da genitália masculina. A- *Kerteszia* Boc M4 (Bocaina Grupo 1); B- *Kerteszia* Boc M2 (Bocaina Grupo 2), C- *Kerteszia* San M1; *Kerteszia* D-Flo M2. As flechas amarelas indicam o claspeto em formato similar a um taco de golfe.

Arquivo adicional 1: Linhas de comando específicas para inferências filogenéticas.

```
##### now aligning with translatorX
#####

cd working_directory/40_at_least_20

cd working_directory/40_at_least_20

nohup awk
'(1==1){file="working_directory/multifasta/bo146_bo56_boni8_c40ucad_cdop_gua_itaF_itaM
_mla8_mla11_ressl_ucad137_bo160_bo104_bo140_bo66_bo85_bo86_boni2_mla21_mla121_
ucad1011_ucad75n_ucad62_Pou1n_Vv1n_Vv3n_ca1_ca11_ca14_ca25_ca31_ca33_ca34_ca36
_Gua2_belS_cruS_ucaNan_bocNan_" $1 ".fna"; system("perl ../translatorx_vLocal.pl -g 1 -t F
-i " file " -o " $1 ".trx ")}'
working_directory/40_at_least_20/40atleast20speciesCV10.gene_name > nohup_trx.out &
#217228
```

#will generate the files trx.nt_cleanali.fasta

```
#####Single gene ML tree
inference##### para inferir uma arvore para cada
gene
```

I created a list file to be able to run the autoiqtree

```
ls *at7147.trx.nt_cleanali.fasta > list #2051 list
```

```
cd working_directory/40_at_least_20
```

```
mv *nt_cleanali.fasta cleanali.fasta/
```

```
mkdir cleanali.fasta
```

```
cp -p working_directory/40_at_least_20/*cleanali.fasta
working_directory/40_at_least_20/cleanali.fasta
```

```
#252432
```

```
nohup perl AutoIQTree.pl & #
```

```
#####TreeS
hrink#####
##
```

```
mkdir working_directory/TreeShrink/
```

```
mkdir working_directory/TreeShrink/Input_dir_40_at_least_20
```

```
nohup nice -10 perl makedirectories.pl &
```

```
nohup nice -10 perl copydirectories.pl & #272374
```

```
nohup nice -10 perl correctname.pl & #274562
```

```
nohup nice -10 python working_directory/run_treeshrink.py -i
working_directory/TreeShrink/Input_dir_40_at_least_20 -t tree_file -a cleanali_file -o
/draft5/kamila/Fil36genomas_at_least_3_per_spp/TreeShrink/Input_dir_40_at_least_20 &
```

```
#317795
```

```
cd working_directory/TreeShrink/TreeShrink
```

```
mv nohup.out nohup_TreeShrink.out
```

```
##### I ran in treeshrink all gene mode. Part of treeshrink nohup:
```

```
# Species mla8_8216at7147 only exists in 1 gene trees
```

```
# Species bo56_8216at7147 only exists in 1 gene trees
# Species bo85_8216at7147 only exists in 1 gene trees
# Species mla11_8216at7147 only exists in 1 gene trees
# Species bo104_8216at7147 only exists in 1 gene trees
# Species mla121_8216at7147 only exists in 1 gene trees
# Species bo66_8216at7147 only exists in 1 gene trees
# There are species with low occupancy in the dataset. TreeShrink will run in 'All-genes' mode
# Writing output ...

# Output files written to working_directory/TreeShrink/Input_dir_40_at_least_20 with prefix
output.
```

```
nohup perl change_align_names.pl & #320904
```

```
##### ASTRAL
(MSC)#####
#Multispecies coalescent method for species tree inference
```

```
cat *at7147*/*output.nwk > list_treefile
```

```
nohup java -jar /home/tools/Astral/astral.5.7.3.jar -i list_treefile.nwk & #24460
```

```
#####IQ-TREE
(ML)#####
```

```
ls *at7147*/*cleanali_file_shrunk0.05 > list_40_alin_shrunk0.05
```

```
nohup perl Concat_MissingData.pl & #144010
```

I had to create a nexus file with partition information. I used Concat.info and edited it in notepad++ in order to create a nexus file (Concat27genAtleast.nex)

EXAMPLE OF NEXUS PARTITION FILE com os resultados do Concat.info:

```
# #nexus
```

```
# begin sets;
```

```
  # charset 3753at7147 = 1700764-1702953;
```

```
    # charset 55301at7147 = 2344006-2344650;
```

```
    # charset 42904at7147 = 1932826-1933704;
```

```
    # charset 34044at7147 = 1529890-1530903;
```

```
    # charset 44737at7147 = 2010205-2011176;
```

```
    # charset 48609at7147 = 2136511-2137212;
```

```
    # charset 51524at7147 = 2235973-2237451;
```

```
    # charset 29906at7147 = 1330666-1332609;
```

```
    # charset 48155at7147 = 2117911-2118621;
```

```
# end;
```

```
nohup /home6/tools/iqtree-1.6.12-Linux/bin/iqtree -s Concat.fas -spp Concat_40gen.nex -m MFP -b 1000 -alrt 1000 -nt 40 &
```

Arquivo adicional 2: Pairwise_Fst_cruzii_1 .R, script utilizado para as análises de F_{ST} das sequências nucleotídicas. Estas foram calculadas a partir dos alinhamentos filtrados de genes ortólogos usando o pacote do R hierfstat (GOUDET, 2005).

```
# R script for calculating Fst in An cruzii. 16jun2023 v3 added try v4: print pops in the same order.
```

```
library(apex)
```

```
library(adeigenet)
```

```
library(hierfstat)
```

```
# setwd("/home6/bernardo/projects/Fst_toy")
```

```
# fileConn<-file("output.txt")
```

```
# writeLines(c("locus","pop_a$","pop_b$","num_a","num_b","Fst\n"), fileConn, sep="t")
```

```
# write(c("locus","pop_a$","pop_b$","num_a","num_b","Fst\n"), file="output.txt",append = FALSE, sep = "\t")
```

```
cat("locus$","pop_a$","pop_b$","num_a","num_b","Fst\n", file="output.txt",append=FALSE, sep="\t")
```

```
fileNames <- Sys.glob("*.cleanali_file_shrunk0.05")
```

```
for (current_fasta in fileNames) {
```

```
    # library("apex")
```

```
    geneData.multidna <- read.multiFASTA(current_fasta)
```

```
    # getLocusNames(geneData.multidna)
```

```
    (setLocusNames(geneData.multidna) <- gsub(".cleanali_file_shrunk0.05", "", getLocusNames(geneData.multidna)))
```

```
    locus <- getLocusNames(geneData.multidna)
```

```
    population_names <- geneData.multidna@labels
```

```
    population_names <- gsub("boni.*|Boni.*", "b_IG", population_names)
```

```
    population_names <- gsub("ita.*", "b_ITA", population_names)
```

```
    population_names <- gsub("^ca31$|^ca1$", "h_CAM", population_names)
```

```
    population_names <- gsub("^ca11$|^ca33$|^ca36$|^ca14$|^ca34$|^ca25$", "b_CAM", population_names)
```

```
    population_names <- gsub("c40ucad", "h_FLO", population_names)
```

```
    population_names <- gsub("ucaNan", "FLO", population_names)
```

```

population_names <- gsub(".*ucad.*", "FLO", population_names)
population_names <- gsub("ressl|Pou.*", "SAN", population_names)
population_names <- gsub("Gua.*|gua", "GUA", population_names)
population_names <- gsub("cruS", "MAQ", population_names)
population_names <- gsub("mla.*|cdop", "ITT", population_names)
population_names <- gsub("bo146|bo160|bo140|bocNan", "BOC1",
population_names)
population_names <- gsub("bo104|bo56|bo66|bo85|bo86", "BOC2",
population_names)
population_names <- gsub("belS", "b_CAN", population_names)
population_names <- gsub("Vv.*", "h_STA", population_names)
# population_names
# geneData.gid <- multidna2genind(geneData.multidna, mlst = TRUE)
geneData.gid <- multidna2genind(geneData.multidna, mlst = FALSE)
# geneData.gid

# library("adegenet")
my_strata <- data.frame(populations = population_names)
strata(geneData.gid) <- my_strata
setPop(geneData.gid) <- ~populations
# geneData.gid
# print(geneData.gid@tab)

# library("hierfstat")
#pairwise.WCfst(geneData.gid,diploid=FALSE)
dat<-genind2hierfstat(geneData.gid)
dat<-dat[order(dat[,1]),]
pops_count <- table(dat[,1])
pops<-unique(dat[,1])
pops_original<-unique(dat[,1])
npop<-length(pops)
dat[,1]<-as.numeric(dat[,1])

```

```

pops<-as.numeric(pops)
for(a in 2:npop){
  pop_a <- as.character(pops_original[a])
  num_a <- pops_count[pop_a]
  for(b in 1:(a-1)){
    pop_b <- as.character(pops_original[b])
    num_b <- pops_count[pop_b]
    subdat <- dat[dat[,1] == pops[a] | dat[,1]==pops[b],]
    result <- try({fst_ab <- wc(subdat,diploid=FALSE)$FST}, silent = FALSE,
outFile = "error.txt" ) # captures errors in wc
    if (class(result) == "try-error") {
      fst_ab <- "."
      cat(locus,pop_a,pop_b,num_a,num_b,fst_ab,"\n",
file="error.txt",append=TRUE, sep="\t")
    }
    # fst_ab <- wc(subdat,diploid=FALSE)$FST
    if (is.nan(fst_ab)) {fst_ab <- "." } # SYSTAT missing data symbol is "."
    if (pop_a <= pop_b){
      cat(locus,pop_a,pop_b,num_a,num_b,fst_ab,"\n",
file="output.txt",append=TRUE, sep="\t")
      cat(locus,pop_a,pop_b,num_a,num_b,fst_ab,"\n", sep="\t")
    }
    else{
      cat(locus,pop_b,pop_a,num_b,num_a,fst_ab,"\n",
file="output.txt",append=TRUE, sep="\t")
      cat(locus,pop_b,pop_a,num_b,num_a,fst_ab,"\n", sep="\t")
    }
    # writeLines(c(locus,pop_a,pop_b,num_a,num_b,fst_ab,"\n"),
fileConn,sep="\t")
    # write(c(locus,pop_a,pop_b,num_a,num_b,fst_ab), file
="output.txt",append = TRUE, sep = "\t")
  }
}

```



```

        rm(geneData.multidna, geneData.gid, dat, subdat, fst_ab, pops_original, pops, pops_count
    )
}

# close(fileConn)

```

Arquivo adicional 3: Script utilizado para plotar as análises de F_{ST} das sequências nucleotídicas, no SYSTAT.

```

USE F:\Fst_com_Bernardo\21jun23\Fst_40pop_3268genes.syz
format 3
select min_num >=2
CLASSIC ON
by pops$
cstatistics Fst / n mean
REM chr X
select (( min_num >=2) AND (chr$ == "chrX" ) )
CLASSIC ON
by pops$
cstatistics Fst / n mean

REM chr2
select (( min_num >=2) AND (chr$ == "chr2" ) )
CLASSIC ON
by pops$
cstatistics Fst / n mean

REM chr3
select (( min_num >=2) AND (chr$ == "chr3" ) )
CLASSIC ON
by pops$
cstatistics Fst / n mean

REM adding chromosome data (same output.txt Fst data) 21jun2023
IMPORT working_directory\Fst_40pop_3268genes.txt / TYPE=ASCII
LET POPPS$ = POP_A$ + "-" + POP_B$
LET min_num= min(num_a,num_b)
DEFVAR num_a,num_b, min_num,pos/ TYPE=number DISPLAY={12,0}
sort pops$
ESAVE working_directory\Fst_40pop_3268genes.syz
by
select min_num >=2 AND pops$="BOC1-BOC2"
select min_num >=2 AND pops$="FLO-GUA"
select min_num >=2 AND pops$="BOC1-ITT"
density Fst / histo bars=24 xmin=-0.2 xmax=1 xtick=12
select min_num >=2
by pops$
REM cstatistics Fst / n mean median
density Fst / histo bars=24 xmin=-0.2 xmax=1 xtick=12
REM analysis per chromosome
select min_num >=2 AND chr$ = "chrX"
select min_num >=2 AND chr$ = "chr2"
REM plotting th positions along the chromosoems
by
select min_num >=2 AND pops$="BOC1-BOC2" AND chr$="chr2"

```

```
plot Fst*pos / size=0.5 fill=1 smooth=lowess tension=0.1
select min_num >=2 AND pops$="BOC1-BOC2" AND chr$="chr3"
plot Fst*pos / size=0.5 fill=1 smooth=lowess tension=0.1
select min_num >=2 AND pops$="BOC1-BOC2" AND chr$="chrX"
plot Fst*pos / size=0.5 fill=1 smooth=lowess tension=0.1
plot Fst*pos / size=0.5 fill=1 smooth=median tension=0.1
```

Capítulo II - Novas evidências moleculares da estrutura populacional de *Anopheles (Kerteszia) bellator* na Mata Atlântica brasileira

Este trabalho é o resultado das nossas análises sobre a diferenciação genética de *Anopheles (Kerteszia) bellator*, gerando um artigo científico que foi publicado em 2019 (Apêndice 2). Neste estudo, a variação molecular e o grau de diferenciação genética entre diferentes populações de *An. bellator* foram analisados utilizando dois fragmentos dos genes envolvidos do relógio biológico: *timeless* e *Clock*.

Referência:

Voges K, da Rosa MPC, Westphal-Ferreira B, Navarro-Silva MA, Pontes CLM, Pitaluga AN, de Carvalho-Pinto CJ, Rona LD. Novel molecular evidence of population structure in *Anopheles (Kerteszia) bellator* from Brazilian Atlantic Forest. Mem Inst Oswaldo Cruz. 2019;114:e180598 (Apêndice 2).

Lista de imagens

- Imagem 1:** Locais de coleta de *Anopheles bellator* na Mata Atlântica brasileira. Mapa do Brasil e ampliação do quadrado preto, exibindo os locais onde *An. bellator* foram coletadas. Os eixos x e y desenhados no mapa (B) exibem a longitude e a latitude, respectivamente. 129
- Imagem 2:** Árvores de máxima verossimilhança das sequências *Timeless* e *Clock* do *Anopheles bellator* (modelos TIM3e+G+I e HKY+G, respectivamente). Essas filogenias mostram que dois grupos estruturados foram identificados nas populações de *An. bellator* do Brasil: *An. bellator* B ocorre na Mata Atlântica do Nordeste (Camacan) e *An. bellator* A é encontrado nas demais regiões mais meridionais do Brasil. Os números nos nós representam os valores percentuais de bootstrap com base em 1.000 replicações. Letras minúsculas nos nomes dos haplótipos especificam os dois alelos (A ou B) obtidos de cada indivíduo. 134

Lista de Tabelas

- Tab. 1:** Diferenciação genética entre todas as populações de *Anopheles bellator*. 132

Lista de arquivos adicionais

- Arquivo adicional 1:** Alinhamento das sequências de DNA do fragmento do gene *Clock*. 143
- Arquivo adicional 2:** Alinhamento das sequências de DNA do fragmento do gene *Timeless*. 144

Lista de Tabelas suplementares

- Tab. suplementar 1:** Este arquivo contém as informações de coleta de campo para cada amostra utilizada neste estudo. 139
- Tab. suplementar 2:** Polimorfismos de todas as populações. 142

Lista de imagens suplementares

- Imagem suplementar 1:** Árvore de máxima verossimilhança de *Anopheles bellator*, *An. cruzii* s.l. e *An. laneanus*. Sequências de *timeless* (modelo K80+G). A filogenia do gene *Timeless* mostra claramente a monofilia entre *An. bellator* A e B quando comparadas com outras espécies de *Kerteszia* estreitamente relacionadas. Os números nos nós representam os valores percentuais de bootstrap com base em 1.000 replicações. Letras minúsculas nos nomes dos haplótipos especificam os dois alelos (A ou B) obtidos de cada indivíduo. 146

Resumo

A malária é uma doença parasitária causada por protozoários *Plasmodium*, transmitida a humanos através da picada do mosquito *Anopheles*. *Anopheles (Kerteszia) bellator* é um dos principais vetores desta doença nas regiões brasileiras cobertas pela Mata Atlântica. Neste trabalho, foram utilizados fragmentos dos genes do relógio biológico *timeless* e *Clock*, para avaliar a diferenciação genética de populações brasileiras de *An. bellator*. Cinco populações de diferentes estados (Bahia, Rio de Janeiro, São Paulo e Paraná) foram analisadas. Os resultados revelaram diferenciação genética significativa entre as populações analisadas. Isso pode ser evidenciado principalmente pelos altos valores de F_{ST} e diferenças genéticas fixas observadas entre a população da Bahia quando comparada as populações do Rio de Janeiro, São Paulo e Paraná. Além disso, a análise filogenética agrupou claramente as sequências da Bahia em um cluster distinto, apoiado por altos valores de bootstrap. Esses resultados sugerem que as populações de *An. bellator* estudadas constituem um complexo de espécies.

Palavras-chave: *Anopheles (Kerteszia) bellator*, bromélia-malária, Mata Atlântica brasileira.

Abstract

Malaria is a parasitic disease caused by *Plasmodium* protozoa, transmitted to humans through the bite of the *Anopheles* mosquito. *Anopheles (Kerteszia) bellator* is one of the main vectors of this disease in Brazilian regions covered by the Atlantic Forest. In this work, fragments of the *timeless* and *Clock* biological clock genes were used to evaluate the genetic differentiation of Brazilian populations of *An. bellator*. Five populations from different states (Bahia, Rio de Janeiro, São Paulo and Paraná) were analyzed. The results revealed significant genetic differentiation between the analyzed populations. This can be evidenced mainly by the high values of F_{ST} and fixed genetic differences observed between the population of Bahia when compared to the populations of Rio de Janeiro, São Paulo and Paraná. Furthermore, the phylogenetic analysis clearly grouped the sequences from Bahia into a distinct cluster, supported by high bootstrap values. These results suggest populations of *An. bellator* studied constitute a complex of species.

Key words: *Anopheles (Kerteszia) bellator*, bromeliad-malaria, Brazilian Atlantic Forest.

Objetivos

Objetivo Geral

Avaliar a diversidade genética de populações de *Anopheles bellator*, a fim de determinar a existência de fluxo gênico entre essas populações.

Objetivo Específicos

- Realizar coletas de mosquitos *An. bellator* em diferentes regiões, como: Ilha Grande (Rio de Janeiro), Cananéia (São Paulo), Ilha do Mel (Paraná) e Camacan (BA);
- Investigar a presença de fluxo gênico e avaliar o grau de divergência genética entre essas populações;
- Verificar se *An. bellator* faz parte de um complexo de espécies crípticas.

Materiais e métodos

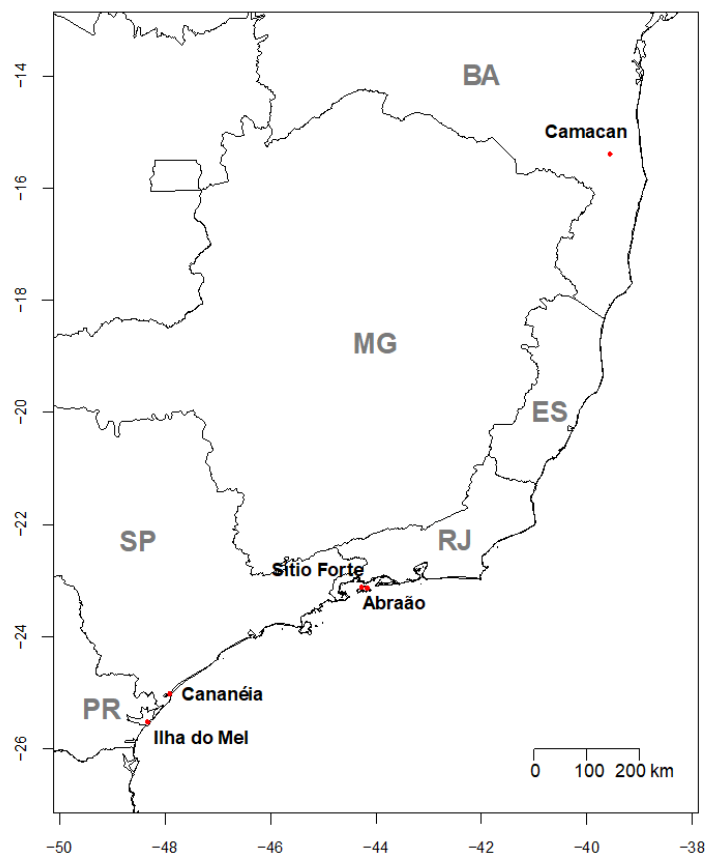
Locais de coleta

Os anofelinos utilizados neste trabalho foram coletados em diferentes locais: Ilha do Mel (PR) (25°30'58.2"S/48°20'2.3"W), Cananéia (SP) (25°01'S/47°55'W), Abraão (23°8'5.14"S/44°10'12.07"W), Sítio Forte (23°7'4.75"S/44°16'29.81"W) (RJ) e Camacan (BA) (15°23'30.9"S/39°33'54.6"W) (Imagem 1). As larvas de *An. bellator* foram coletadas na água acumulada no interior de plantas da família Bromeliaceae. Estas larvas foram levadas para o laboratório e criadas até o estágio adulto. Também foram coletados mosquitos adultos no horário crepuscular, no início da noite final de tarde (compreendendo o período das 17:00 horas às 20:00 horas). A identificação dos mosquitos adultos foi realizada conforme Consoli & Oliveira (1994). Após a identificação, essas amostras foram acondicionadas em etanol 95% e armazenadas em -20°C até o momento da extração do DNA. As informações sobre estágio de vida das amostras e local de coleta podem ser visualizadas no dados suplementares, na Tabela 1 (Tab. suplementar 1). O mapa da imagem 1 foi feito com a utilização dos pacotes maptools, maps e GISTools (BRUNSDON; CHEN, 2014; BECKER et al., 2017; BIVAND; LEWIN-KOH, 2017) do programa R.

Imagem 1: Locais de coleta de *Anopheles bellator* na Mata Atlântica brasileira. Mapa do Brasil e ampliação do quadrado preto, exibindo os locais onde *An. bellator* foram coletadas. Os eixos x e y desenhados no mapa (B) exibem a longitude e a latitude, respectivamente.



Population	Latitude	Longitude
Ilha do Mel (PR)	-25.516167	-48.333972
Cananéia (SP)	-25.016667	-47.916667
Sítio Forte (RJ)	-23.117986	-44.274947
Abraão (RJ)	-23.134761	-44.170019
Camacan (BA)	-15.391917	-39.565167



Dados moleculares

Sequências gênicas parciais dos genes *timeless* e *Clock* de *An. bellator* foram obtidas por reação em cadeia da polimerase (PCR), clonagem e sequenciamento conforme descrito em Dias et al (2018). As reações de PCR foram realizadas em um termociclador Eppendorf Mastercycler® usando DNA polimerase proofreading Pfu (Biotools) e primers descritos anteriormente (RONA et al., 2009, 2010; RONA; CARVALHO-PINTO; PEIXOTO, 2010). Os produtos de PCR foram purificados com QIAquick PCR Purification Kit (QIAGEN) e clonados usando o CloneJET™ PCR Cloning Kit (Fermentas Life Sciences). O sequenciamento foi executado em sequenciador de DNA ABI3130 da Myleus

Biotechnology. Para cada mosquito, pelo menos oito clones foram sequenciados para mitigar erros de PCR e permitir a identificação dos dois alelos. Sequências consenso representando os dois alelos foram geradas. Os indivíduos foram classificados como homozigotos quando apenas um haplótipo foi observado entre as oito sequências. As sequências foram submetidas ao GenBank sob os números de acesso: MG755641 - MG755734. Um alinhamento de todas as sequências de *An. bellator* para cada gene são apresentadas em arquivos suplementares (Arquivo adicional 1 e 2).

A identidade dos fragmentos clonados foi determinada por análise BlastX contra o banco de dados NCBI [<http://www.ncbi.nlm.nih.gov/BLAST/>].

Sequências do fragmento do gene *timeless* de *Anopheles (Kerteszia) laneanus* e *An. cruzii s.l.* foram utilizadas para verificar a relação de *An. bellator* com outras espécies de *Kerteszia* relacionadas. As sequências de *An. laneanus* foram obtidos por PCR, clonagem e sequenciamento conforme descrito acima e submetidos ao GenBank sob os números de acesso: KT803975 - KT803978. As sequências de *An. cruzii s.l.* foram as publicadas anteriormente por Rona et al. (2009) (números de acesso do GenBank FJ408732 , FJ408735 , FJ408746 , FJ408786 , FJ408788 , FJ408813 , FJ408816 , FJ408832 , FJ408736 , FJ408738 para população SC e FJ408838 , FJ408840 , FJ408848 , FJ408854 , FJ408855, FJ408856 , FJ408862 , FJ408864 para população BA).

Análise das sequências

As sequências de DNA foram alinhadas com MAFFT (KATO; STANDLEY, 2013) e as árvores filogenéticas foram construídas para cada gene sob o método de Máxima Verossimilhança usando os pacotes ape e phangorn (PARADIS; CLAUDE; STRIMMER, 2004; SCHLIEP, 2011) do Software R versão 3.3.2 (R CORE TEAM, 2014). Os modelos de substituição de melhor ajuste HKY+G (*Clock*) e TIM3e+G+I (*timeless*) foram selecionados seguindo o critério AIC usando a função modelTest do pacote phangorn (SCHLIEP, 2011).

O software DnaSP v5 (ROZAS et al., 2003) foi usado para calcular estatísticas intrapopulacionais, incluindo diversidade de nucleotídeos e número de sítios segregantes. Este software também foi usado para garantir que todos os loci atendessem ao modelo de neutralidade calculando o teste D de Tajima

(TAJIMA, 1989). O software PROSEQ v 2.91 (FILATOV, 2002) foi implementado para realizar análises interpopulacionais, incluindo o número de polimorfismos fixos e compartilhados e as estimativas pareadas de diferenciação populacional.

Resultados

No total 94 sequências de dois genes do relógio biológico foram analisadas neste estudo (Dados suplementares, Tab. suplementar 2). O fragmento parcial do gene *Clock* consistia em 164 pb, enquanto o fragmento do gene *timeless* compreendia 368 pb (Dados suplementares, arquivo adicional 1 e 2).

Análises de diversidade nucleotídica (valores θ e π), indicaram que *Clock* tem maior diversidade genética quando comparado ao *timeless* (Dados suplementares, tab. suplementar 2). Para os testes D de Tajima não obtivemos resultados significativos para ambos os *loci*, indicando que eles seguem o pressuposto da neutralidade (dados suplementares, tab. suplementar 2).

A Tabela 1 apresenta valores de F_{ST} , polimorfismos compartilhados (S_s), diferenças fixas (S_f), divergência bruta (D_{xy}) e divergência líquida (D_a) entre as populações analisadas. D_{xy} é a medida de divergência entre as sequências das duas populações, baseado no número de diferenças par-a-par, que não exclui a variabilidade intrapopulacional, ou seja, é a divergência bruta. D_a é um índice que mede apenas a diferença entre as populações, ou seja, a diferença líquida.

Quanto ao F_{ST} , nas análises realizadas com a população da BA, obtivemos valores altos de diferenciação (F_{ST} variando entre 0,62 e 0,67). Além disso, haviam poucos polimorfismos compartilhados e um número substancial de diferenças fixas (Tab. 1), algumas inclusive ocasionaram troca de aminoácidos (Dados suplementares, arquivo suplementar 1 e 2). Ademais, o número médio de substituições de nucleotídeos por sítio (D_{xy}) e o número líquido de substituições de nucleotídeos por sítio entre as populações (D_a) apresentaram maiores valores nas comparações envolvendo a população BA. Por outro lado, entre as populações do Sul e do Sudeste, polimorfismos compartilhados foram identificados e não foram encontradas diferenças fixas (Tab.1).

Tab. 1: Diferenciação genética entre todas as populações de *Anopheles bellator*.

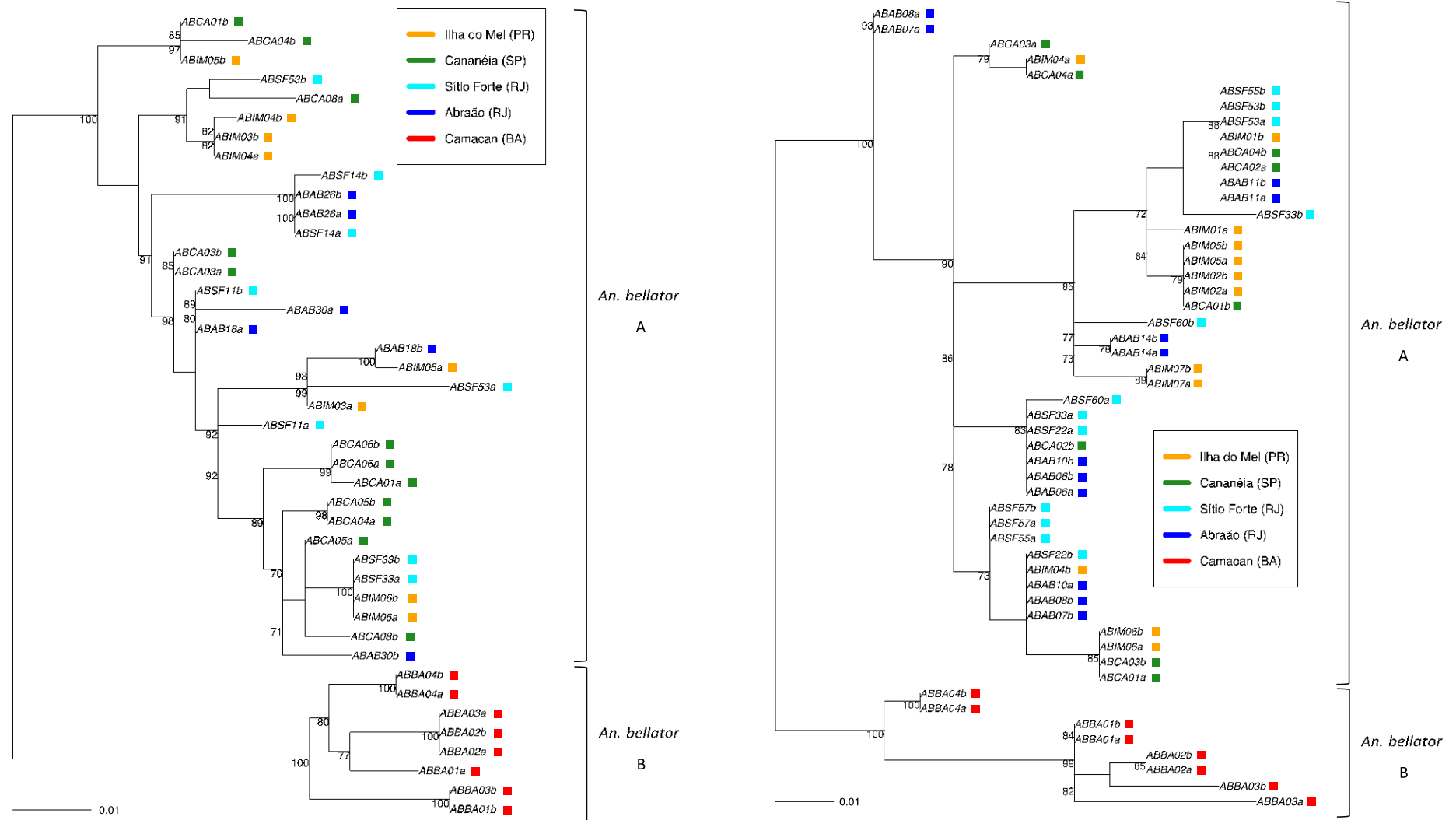
Populações	Gene	Km	F_{ST}	D_{xy}	D_a	S_s	S_f
Abraão x Sítio Forte	<i>timeless</i>	10.89	-0.1039	0.0177	-0.0018	11	00
	<i>Clock</i>		-0.0157	0.0322	-0.0005	08	00
Ilha do Mel x Abraão	<i>timeless</i>	496.5	0.0411	0.0215	0.0009	09	00
	<i>Clock</i>		0.1134	0.0404	0.0046	07	00
Ilha do Mel x Sítio Forte	<i>timeless</i>	488.6	0.0695	0.0224	0.0016	11	00
	<i>Clock</i>		0.0904	0.0377	0.0034	08	00
Ilha do Mel x Cananéia	<i>timeless</i>	70.1	0.1147	0.0233	0.0027	10	00
	<i>Clock</i>		-0.0508	0.0376	-0.0019	10	00
Cananéia x Sítio Forte	<i>timeless</i>	425.7	0.1158*	0.0221	0.0026	08	00
	<i>Clock</i>		0.0001	0.0364	0.0000	09	00
Cananéia x Abraão	<i>timeless</i>	434.1	0.1311*	0.0222	0.0029	06	00
	<i>Clock</i>		0.0086	0.0383	0.0003	08	00
Camacan (BA) x Ilha do Mel	<i>timeless</i>	1449	0.6471*	0.0621	0.0402	00	07
	<i>Clock</i>		0.6482*	0.1168	0.0757	02	05
Camacan (BA) x Cananéia	<i>timeless</i>	1379	0.6694*	0.0622	0.0417	00	09
	<i>Clock</i>		0.6287*	0.1163	0.0731	02	05
Camacan (BA) x Abraão	<i>timeless</i>	987.1	0.6722*	0.0628	0.0422	01	10
	<i>Clock</i>		0.6236*	0.1050	0.0655	02	03
Camacan (BA) x Sítio Forte	<i>timeless</i>	990.1	0.6758*	0.0641	0.0434	00	09
	<i>Clock</i>		0.6562*	0.1105	0.0725	02	05

F_{ST} : estimativas pareadas de diferenciação populacional. A significância dos valores de F_{ST} foi avaliada por 1.000 permutações aleatórias (* $p < 0,05$). D_{xy} : número médio de substituições de nucleotídeos por sítio entre as populações; D_a : número de substituições de nucleotídeos líquido por sítio entre as populações. S_s : número de polimorfismos compartilhados entre as duas populações. S_f : número de diferenças fixas entre as duas populações. Km: as distâncias geográficas aproximadas entre localidades em km.

As árvores filogenéticas não apresentaram separação distinta entre as sequências da Ilha do Mel, Cananéia, Abraão e Sítio Forte. No entanto, ocorreu claramente agrupamento das sequências da população da BA em um cluster separado, com altos valores de bootstrap (imagem 2).

Também foi avaliada a relação filogenética de *An. bellator* com *An. cruzii* s.l. e *An. laneanus* (dados complementares, imagem suplementar 1). Essa topologia revelou uma maior proximidade filogenética entre *An. cruzii* e *An. laneanus* em comparação com *An. bellator*.

Imagem 2: Árvores de máxima verossimilhança das sequências de *timeless* e *Clock* do *Anopheles bellator* (modelos TIM3e+G+I e HKY+G, respectivamente). Essas filogenias mostram que dois grupos estruturados foram identificados nas populações de *An. bellator* do Brasil: *An. bellator* B ocorre na Mata Atlântica do Nordeste (Camacan) e *An. bellator* A é encontrado nas demais regiões mais meridionais do Brasil. Os números nos nós representam os valores percentuais de bootstrap com base em 1.000 replicações. Letras minúsculas nos nomes dos haplótipos especificam os dois alelos (A ou B) obtidos de cada indivíduo.



Discussão

Nossos achados demonstraram um nível significativo de divergência genética entre a população da BA e populações das regiões sul e sudeste do Brasil, provenientes do PR, SP e RJ. Isso sugere fortemente a existência de pelo menos dois grupos diferentes de *An. bellator* nas regiões cobertas pela Mata Atlântica brasileira (aqui denominados “*An. bellator A*” e “*An. bellator B*”).

Esse fato fica bastante evidente ao observarmos os valores de análises da diferenciação genética, como o índice de F_{ST} , onde obtivemos um valor médio de 0,65 na comparação entre a Bahia e as populações do sul e sudeste brasileiro. Esses valores de diferenciação ultrapassam o padrão geralmente observado dentro de uma única espécie. E está em concordância com a constatação de Hey & Pinho (2012) de que valores de F_{ST} maiores que 0,35 indicam a presença de espécies distintas. Ademais, também foram observadas diferenças genéticas fixas entre a população da BA e as populações do sul. Além disso, os maiores valores de D_a e D_{xy} também foram encontrados nas comparações realizadas com a população da Bahia, o que confirma as medidas de F_{ST} .

Em concordância com esses resultados, as filogenias dos genes *timeless* e *Clock* não apresentaram separação entre as sequências de PR, SP e RJ (referido aqui como *An. bellator A*) todavia, agrupam as sequências de BA em um cluster separado (chamado de *An. bellator B*), com valores de suporte de ramos altos.

Para avaliar a relação filogenética de *An. bellator* com outras espécies de *Kerteszia* (*An. cruzii s.l.* e *An. laneanus*), foi estimada uma árvore de Máxima Verossimilhança usando apenas as regiões codificantes do gene *timeless*. A árvore filogenética resultante mostrou que *An. cruzii* e *An. laneanus* são filogeneticamente mais próximas quando comparadas a *An. bellator*, resultado semelhante ao de Foster et al (2017) e Lorenz et al (2015) utilizando sequências mitocondriais e genômicas.

Além disso, nossos dados também corroboram os resultados de Carvalho-Pinto e Lourenço de Oliveira (2003), no sentido de que existem grupos geneticamente diferentes para *An. bellator*. Estes autores utilizaram de métodos

isoenzimáticos para avaliar estruturação genética entre populações dos estados de Santa Catarina, São Paulo, Bahia e da ilha de Trinidad, e concluíram que havia maiores distâncias genéticas nas comparações entre a população de BA quando comparada as de SC e SP (CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2003).

Do mesmo modo, observa-se um padrão semelhante para outro anofelino do subgênero *Kerteszia*: *An. cruzii*. Estudos anteriores também sugerem a existência de linhagens distintas para *An. cruzii*. De acordo com um trabalho realizado por Carvalho-Pinto e Lourenço-de-Oliveira (2004), com populações de *An. cruzii* coletadas em diferentes regiões do Brasil, constatou-se que as populações de São Paulo, Santa Catarina e Rio de Janeiro são geneticamente mais semelhantes entre si em comparação com a população da Bahia. Ademais, estudos conduzidos por Rona et al. (2009, 2010; 2010; 2013), através de uma análise multilocus com genes que codificam proteínas ribossomais e genes do relógio biológico, confirmaram as descobertas de Carvalho-Pinto e Lourenço-de-Oliveira (2004). No qual sugerem que estes anofelinos provenientes da Bahia formam linhagens geneticamente distintas das populações do sul/sudeste do Brasil.

Uma hipótese para explicar esse mesmo padrão entre essas duas diferentes espécies de *Kerteszia* foi sugerida por Rona (2009; 2010). Estes autores relatam que os tempos de divergência entre essas espécies correspondem ao mesmo período em que ocorreram mudanças climáticas no Pleistoceno. Nesse contexto, eventos glaciais causaram uma redução na temperatura, levando a uma diminuição nas chuvas, o que resultou em períodos longos de seca e fragmentação da Mata Atlântica. Contudo, de acordo com Carnaval et al. (2009), durante esses períodos de resfriamento, alguns locais como a Bahia, mantiveram as florestas intactas. Conseqüentemente, essa fragmentação da floresta teria impulsionado esse processo de especiação de diferentes espécies presentes na Mata Atlântica. Como: *An. cruzii*, *An. bellator*, e ainda outros mosquitos transmissores de malária (LOAIZA et al., 2012).

Uma outra hipótese que poderia explicar essa estruturação genética observada entre BA e as populações do Sul/Sudeste em nosso trabalho, poderia ser simplesmente distância geográfica entre elas (pelo menos 980 Km de distância) (Tab. 1). Mas neste caso, seria razoável pressupor também uma

maior diferenciação genética entre Ilha do Mel (PR) e Cananéia (SP) em relação as populações do RJ (Abraão e Sítio Forte), considerando a distância de cerca de ~500 km entre elas. No entanto, foi obtido F_{ST} médio de apenas 0,07. Surpreendentemente entre as populações do sul e sudeste, foram observados polimorfismos compartilhados e ausência de diferenças fixas. Além disso, todos os valores de F_{ST} foram, em geral, não significativos e menores que $F_{ST} = 0,1$, sugerindo que pertencem à uma única espécie.

Por fim, este estudo revela a existência de dois grupos geneticamente distintos na Mata Atlântica brasileira: *An. bellator* A, que está amplamente distribuído nas regiões Sul e Sudeste, e *An. bellator* B, encontrado no nordeste do Brasil. É curioso observar que, apesar dos casos de malária serem registrados frequentemente na BA, *An. bellator* nunca foi identificado como vetor da doença neste estado, ao contrário do que ocorre nas regiões do Sul do Brasil. Portanto, seria plausível sugerir que a diferenciação genética observada entre BA e as populações do Sul poderia estar relacionada a suposta competência vetorial diferente entre populações de *An. bellator* no Brasil.

Dados suplementares

Tab. Suplementar 1: Este arquivo contém as informações de coleta de campo para cada amostra utilizada neste estudo.

ID do indivíduo	País	Estado	Localidade	Latitude	Longitude	Ano	Mês	Sexo	Fonte
ABAB06	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	M	larva
ABAB07	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	M	larva
ABAB08	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	M	larva
ABAB10	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	M	larva
ABAB11	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	M	larva
ABAB14	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	M	larva
ABAB18	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	F	larva
ABAB26	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	F	larva
ABAB30	Brasil	Rio de Janeiro (RJ)	Abraão, Ilha Grande, Angra dos Reis	- 23134761	- 44170019	2012	Agosto	F	larva
ABBA01	Brasil	Bahia (BA)	Camacan	- 15391917	- 39565167	2015	Março	F	larva
ABBA02	Brasil	Bahia (BA)	Camacan	- 15391917	- 39565167	2015	Março	F	larva

ABBA03	Brasil	Bahia (BA)	Camacan	-	-	2015	Março	F	larva
				15391917	39565167				
ABBA04	Brasil	Bahia (BA)	Camacan	-	-	2015	Março	F	larva
				15391917	39565167				
ABCA01	Brasil	São Paulo (SP)	Cananéia	-	-	2003	Agosto	F	larva
				25016667	47916667				
ABCA02	Brasil	São Paulo (SP)	Cananéia	-	-	2003	Agosto	F	larva
				25016667	47916667				
ABCA03	Brasil	São Paulo (SP)	Cananéia	-	-	2003	Agosto	F	larva
				25016667	47916667				
ABCA04	Brasil	São Paulo (SP)	Cananéia	-	-	2003	Agosto	F	larva
				25016667	47916667				
ABCA05	Brasil	São Paulo (SP)	Cananéia	-	-	2003	Agosto	F	larva
				25016667	47916667				
ABCA06	Brasil	São Paulo (SP)	Cananéia	-	-	2003	Agosto	F	larva
				25016667	47916667				
ABCA08	Brasil	São Paulo (SP)	Cananéia	-	-	2003	Agosto	F	larva
				25016667	47916667				
ABIM01	Brasil	Paraná (PR)	Paranaguá, Ilha do Mel	-	-	2014	Janeiro	F	larva
				25516167	48333972				
ABIM02	Brasil	Paraná (PR)	Paranaguá, Ilha do Mel	-	-	2014	Janeiro	F	larva
				25516167	48333972				
ABIM03	Brasil	Paraná (PR)	Paranaguá, Ilha do Mel	-	-	2014	Janeiro	F	larva
				25516167	48333972				
ABIM04	Brasil	Paraná (PR)	Paranaguá, Ilha do Mel	-	-	2014	Janeiro	F	larva
				25516167	48333972				
ABIM05	Brasil	Paraná (PR)	Paranaguá, Ilha do Mel	-	-	2014	Janeiro	F	larva
				25516167	48333972				

ABIM06	Brasil	Paraná (PR)	Paranaguá, Ilha do Mel	- 25516167	- 48333972	2014	Janeiro	M	larva
ABIM07	Brasil	Paraná (PR)	Paranaguá, Ilha do Mel	- 25516167	- 48333972	2014	Janeiro	?	Larva
ABSF11	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	F	Larva
ABSF14	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	F	Larva
ABSF22	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	F	Larva
ABSF33	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	F	Larva
ABSF53	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	F	Larva
ABSF55	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	F	Larva
ABSF57	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	F	Larva
ABSF60	Brasil	Rio de Janeiro (RJ)	Sítio Forte, Ilha Grande, Angra dos Reis	- 23117986	- 44274947	2012	Agosto	?	Larva

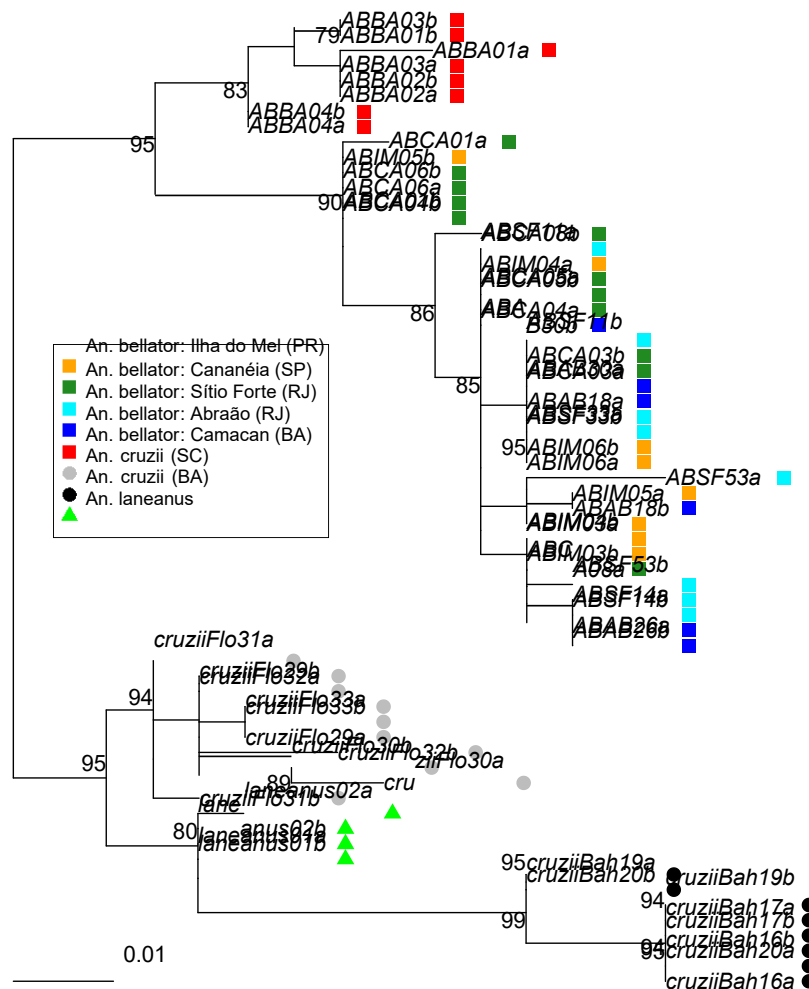
ID do indivíduo, a identidade de cada amostra; país, estado, localidade, latitude e longitude, detalhes dos locais de coleta para cada ponto de coleta; ano e mês, datas de coleta; fonte, se os espécimes foram coletados como adultos ou larvas. Os imaturos coletados foram criados em laboratório até se tornarem adultos. Em Cananeia (SP) e Camacan (BA), foram realizadas capturas apenas de fêmeas adultas; nas outras três localidades, foram realizadas apenas capturas de imaturos. M: machos; F: fêmeas; ?: sexo indeterminado.

Tab. Suplementar 2: Polimorfismos de todas as populações.

População	Gene	<i>N</i>	<i>S</i>	θ	π	D_T
Abraão (RJ)	<i>timeless</i>	06	22	0.0266	0.0258	-0.1963
	<i>Clock</i>	12	11	0.0224	0.0288	1.1698
Camacan (BA)	<i>timeless</i>	08	17	0.0178	0.0199	0.6018
	<i>Clock</i>	08	15	0.0375	0.0354	-0.2837
Cananéia (SP)	<i>timeless</i>	12	18	0.0174	0.0194	0.5138
	<i>Clock</i>	08	12	0.0285	0.0346	1.0643
Ilha do Mel (PR)	<i>timeless</i>	08	20	0.0230	0.0243	0.2898
	<i>Clock</i>	12	11	0.0224	0.0311	1.6025
Sítio Forte (RJ)	<i>timeless</i>	08	17	0.0215	0.0197	-0.4378
	<i>Clock</i>	12	11	0.0278	0.0330	0.7890

N: número de sequências de DNA de cada população; S: número de polimórficos (segregando) locus; θ : diversidade de nucleotídeos com base no número total de mutações; π : diversidade de nucleotídeos com base no número médio de diferenças pareadas; D_T : Tajima's D, com base no total número de mutações. Em nenhum caso os valores D de Tajima foram significativos ($p > 0,10$ em todos os casos).

Imagem suplementar 1: Árvore de máxima verossimilhança de *An. bellator*, *An. cruzii* s.l. e *An. laneanus*. Sequências de *timeless* (modelo K80+G). A filogenia do gene *timeless* mostra claramente a monofilia entre *An. bellator* A e B quando comparadas com outras espécies de *Kerteszia* estreitamente relacionadas. Os números nos nós representam os valores percentuais de bootstrap com base em 1.000 replicações. Letras minúsculas nos nomes dos haplótipos especificam os dois alelos (A ou B) obtidos de cada indivíduo.



Referências

ABASCAL, F.; ZARDOYA, R.; TELFORD, M. J. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, v. 38, n. SUPPL. 2, p. 7–13, 2010.

ABOAGYE-ANTWI, F. et al. Experimental swap of *Anopheles gambiae* 's assortative mating preferences demonstrates key role of X-chromosome divergence island in incipient sympatric speciation. *PLOS Genetics*, v. 11, n. 4, p. e1005141, 16 abr. 2015. Disponível em: <<https://dx.plos.org/10.1371/journal.pgen.1005141>>.

AHMED, S. et al. Travel is a key risk factor for malaria transmission in pre-elimination settings in Sub-Saharan Africa: A review of the literature and meta-analysis. *American Journal of Tropical Medicine and Hygiene*, v. 103, n. 4, p. 1380–1387, 2020.

ANOPHELES GAMBIAE 1000 GENOMES CONSORTIUM. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. *Genome Research*, v. 30, n. 10, p. 1533–1546, 2020.

ARAGÃO, M. B. Distribuição geográfica e abundância de espécies de *Anopheles (Kerteszia)* (Diptera, Culicidae). *Rev Bras Malariol Doenças Trop*, v. 16, n. 73–109, 1964.

ARCOS, A. N. et al. Characterization of artificial larval habitats of *Anopheles darlingi* (Diptera: Culicidae) in the Brazilian Central Amazon. *Revista Brasileira de Entomologia*, v. 62, n. 4, p. 267–274, out. 2018. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S2221169115302483>>.

ARRIVILLAGA, J. et al. The taxonomic status of genetically divergent populations of *Lutzomyia longipalpis* (Diptera: Psychodidae) based on the distribution of mitochondrial and isozyme variation. *Journal of Medical Entomology*, v. 40, n. 5, p. 615–627, 2003.

ARROW, K. J.; PANOSIAN, C.; GELBAND, H. *Saving lives, buying time:*

economics of malaria drugs in an age of resistance - A brief history of malaria.
[s.l: s.n.]

ASHLEY, E. A.; PYAE PHYO, A.; WOODROW, C. J. Malaria. *The Lancet*, v. 391, n. 10130, p. 1608–1621, 2018.

BANKEVICH, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, v. 19, n. 5, p. 455–477, 2012.

BARATA, R. *Malária e seu controle*. 1. ed. São Paulo: Editora Hucitec, 1988.

BARRÓN, M. G. et al. A new species in the major malaria vector complex sheds light on reticulated species evolution. *Scientific Reports*, v. 9, n. 1, p. 14753, 14 out. 2019. Disponível em: <<https://www.nature.com/articles/s41598-019-49065-5>>.

BAUZER, L. G. et al. *Lutzomyia longipalpis* in Brazil: a complex or a single species? A mini-review. *Memórias do Instituto Oswaldo Cruz*, v. 102, n. 1, p. 1–12, fev. 2007. Disponível em:
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762007000100001&lng=en&tlng=en>.

BECKER, R. et al. *Maps: draw geographical maps. R package version 3.2.0.*, 2017. . Disponível em: <<https://cran.r-project.org/package=maps>>.

BENCHIMOL, J. L.; SÁ, R. M. *Adolpho Lutz – Obra Completa – v. 2 Febre Amarela, Malária e Protozoologia Yellow Fever, Malaria and Protozoology*. [s.l: s.n.]

BERENDS, C. J. et al. On the cause of the mid-pleistocene transition. *Reviews of Geophysics*, v. 59, n. 2, 16 jun. 2021. Disponível em:
<<https://onlinelibrary.wiley.com/doi/10.1029/2020RG000727>>.

BERTOLA, M. et al. Updated occurrence and bionomics of potential malaria vectors in Europe: a systematic review (2000–2021). *Parasites and Vectors*, v. 15, n. 1, p. 1–34, 2022. Disponível em: <<https://doi.org/10.1186/s13071-022-05204-y>>.

BICKFORD, D. et al. Cryptic species as a window on diversity and

conservation. *Trends in Ecology & Evolution*, v. 22, n. 3, p. 148–155, mar. 2007.

Disponível em:

<<https://linkinghub.elsevier.com/retrieve/pii/S0169534706003703>>.

BIVAND, R.; LEWIN-KOH, N. *Maptools: tools for reading and handling spatial objects. R package version 0.9-2. 2017*, 2017. . Disponível em: <<https://cran.r-project.org/package=maptools>>.

BOURKE, B. P. et al. High levels of diversity in *Anopheles* Subgenus *Kerteszia* revealed by species delimitation analyses. *Genes*, v. 14, n. 2, p. 1–14, 2023.

BRANQUINHO, M. S. et al. Infection of *Anopheles (Kerteszia) cruzii* by *Plasmodium vivax* and *Plasmodium vivax* variant VK247 in the municipalities of São Vicente and Juquitiba, São Paulo. *Revista panamericana de salud publica = Pan American journal of public health*, v. 2, n. 3, p. 189–93, 1997. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9445765>>.

BRASIL. *O controle de endemias no Brasil: de 1979 a 1984*. [s.l.: s.n.].

BRASIL. *Manual para profissionais da saúde na atenção básica*. Brasília, DF Ministério da Saúde. Secretaria de vigilância em saúde., , 2005. .

BRASIL, P. et al. Outbreak of human malaria caused by *Plasmodium simium* in the Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation. *The Lancet Global Health*, v. 5, n. 10, p. e1038–e1046, out. 2017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2214109X17303339>>.

BRUNSDON, C.; CHEN, H. *GISTools: some further GIS capabilities for R*. R package version 0.7-4, , 2014. .

BUERY, J. C. et al. Ecological characterisation and infection of Anophelines (Diptera: Culicidae) of the Atlantic Forest in the southeast of Brazil over a 10 year period: has the behaviour of the autochthonous malaria vector changed? *Memórias do Instituto Oswaldo Cruz*, v. 113, n. 2, p. 111–118, 2 fev. 2018.

Disponível em:

<<https://malariajournal.biomedcentral.com/articles/10.1186/s12936-019-2744-8>>.

- BUERY, J. C. et al. Atlantic forest malaria: A review of more than 20 years of epidemiological investigation. *Microorganisms*, v. 9, n. 1, p. 1–14, 2021.
- CALADO, D. C.; NAVARRO-SILVA, M. a. Identificação de *Anopheles (Kerteszia) cruzii* Dyar & Knab e *Anopheles (Kerteszia) homunculus* Komp (Diptera, Culicidae, Anophelinae) através de marcadores moleculares (RAPD e RFLP). *Revista Brasileira de Zoologia*, v. 22, n. 4, p. 1127–1133, 2005.
- CAMARGO, E. P. Malária, maleita, paludismo. *Ciênc. cult. (São Paulo)*, v. 55, n. 1, p. 26–30, 2003.
- CAMPOS, M. et al. Unravelling population structure heterogeneity within the genome of the malaria vector *Anopheles gambiae*. *BMC Genomics*, v. 22, n. 1, p. 422, 8 dez. 2021. Disponível em: <<https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-021-07722-y>>.
- CARDOSO, J. D. C. et al. New records of *Anopheles homunculus* in central and serra do mar biodiversity corridors of the atlantic forest, Brazil. *Journal of the American Mosquito Control Association*, v. 28, n. 1, p. 1–5, mar. 2012. Disponível em: <<http://www.bioone.org/doi/abs/10.2987/11-6205.1>>.
- CARLOS, B. C. et al. A comprehensive analysis of malaria transmission in Brazil. *Pathogens and Global Health*, v. 113, n. 1, p. 1–13, 2 jan. 2019. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/20477724.2019.1581463>>.
- CARNAVAL, A. C. et al. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, v. 323, n. 5915, p. 785–789, 2009.
- CARVALHO-PINTO, C. J.; LOURENÇO-DE-OLIVEIRA, R. Isoenzymatic Analysis of Four *Anopheles (Kerteszia) bellator* Dyar & Knab (Diptera: Culicidae) Populations. *Memórias do Instituto Oswaldo Cruz*, v. 98, n. 8, p. 1045–1048, 2003.
- CARVALHO-PINTO, C. J. de; LOURENÇO-DE-OLIVEIRA, R. Isoenzymatic analysis of four *Anopheles (Kerteszia) cruzii* (Diptera: Culicidae) populations of Brazil. *Memórias do Instituto Oswaldo Cruz*, v. 99, n. 5, p. 471–475, 2004.

CHAHAD-EHLERS, S.; LOZOVEI, A. L.; MARQUES, M. D. Reproductive and Post-Embryonic Daily Rhythm Patterns of the Malaria Vector *Anopheles* (*Kerteszia*) *cruzei*: Aspects of the Life Cycle. *Chronobiology International*, v. 24, n. 2, p. 289–304, 7 jan. 2007. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/07420520701282174>>.

CLAYTON, A. M.; DONG, Y.; DIMOPOULOS, G. The *Anopheles* innate immune system in the defense against malaria infection. *Journal of Innate Immunity*, v. 6, n. 2, p. 169–181, 2014.

COETZEE, M. et al. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa*, v. 3619, n. 3, p. 246–274, 2013.

COLLUCCI, E.; SALLUM, M. Phylogenetic analysis of the subgenus *Kerteszia* of *Anopheles* (Diptera: Culicidae: Anophelinae) based on morphological characters. *Insect Syst.*, v. 34, p. 361–372, 2003.

COLUZZI, M. et al. Behavioural divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. *Nature*, v. 266, p. 832–833, 1977.

COLUZZI, M. et al. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Royal society of tropical medicine and hygiene*, v. 73, 1979.

COLUZZI, M. et al. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science*, v. 298, n. 5597, p. 1415–1418, 2002.

CONN, J. E. et al. Emergence of a new neotropical malaria vector facilitated by human migration and changes in land use. *American Journal of Tropical Medicine and Hygiene*, v. 66, n. 1, p. 18–22, 2002.

CONSOLI, R. A. G. B.; LOURENÇO DE OLIVEIRA, R. *Principais mosquitos de importância médica no Brasil*. [s.l.: s.n.]. Único

CORRÊA, R. R.; CERQUEIRA, F. M. C. Descrição de *Anopheles* (*Kerteszia*) *laneanus*, nova espécie de anofelino de Campos do Jordão (Diptera, Culicidae). *Arquivos de higiene e saúde pública*, v. 9, p. 111–117, 1944.

COX, F. E. History of the discovery of the malaria parasites and their vectors. *Parasites and Vectors*, v. 3, n. 1, p. 1–9, 2010.

DABIRE, K. R. et al. Assortative mating in mixed swarms of the mosquito *Anopheles gambiae* s.s . M and S molecular forms, in Burkina Faso, West Africa. *Medical and Veterinary Entomology*, v. 27, n. 3, p. 298–312, set. 2013. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2915.2012.01049.x>>.

DANTAS-TORRES, F. Species concepts: What about ticks? *Trends in Parasitology*, v. 34, n. 12, p. 1017–1026, dez. 2018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1471492218302137>>.

DAVIS, N. C. A Note on the Malaria-carrying Anophelines in Belém, Para, and in Natal, Rio Grande do Norte, Brazil. *Rivista di Malariologia*, v. 10, p. 43–51, 1931.

DE LEMOS, A. B. et al. Reemergence of human malaria in Atlantic Forest of Rio Grande do Sul, Brazil. *Memórias do Instituto Oswaldo Cruz*, v. 116, 2021. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762021000100403&tlng=en>.

DE PINA-COSTA, A. et al. Malaria in Brazil: What happens outside the Amazonian endemic region. *Memorias do Instituto Oswaldo Cruz*, v. 109, n. 5, p. 618–633, 2014.

DE REZENDE DIAS, G. et al. Cryptic diversity in an Atlantic Forest malaria vector from the mountains of South-East Brazil. *Parasites and Vectors*, v. 11, n. 1, p. 1–11, 2018.

DEANE, L. *Malaria vectors in Brazil* Mem. Inst. Oswaldo Cruz, 1986. .

DEANE, L. Malaria Studies and Control in Brazil. *Am J Trop Med Hyg*, v. 38, n. 2, p. 223–230, 1988.

DEANE, L. Os grandes marcos na história do controle da malária. *Rev Bras Med Trop*, v. 25, p. 12–22, 1992.

DEANE, L. M. et al. *Anopheles (Kerteszia) cruzii*, a natural vector of the monkey malaria parasites, *Plasmodium simium* and *Plasmodium brasilianum* .

Transactions of the Royal Society of Tropical Medicine and Hygiene, v. 64, n. 4, p. 647, 1970.

DEANE, L. M. et al. On the transmission of simian malaria in Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo*, v. 13, n. 5, p. 311–319, 1971.

DEANE, L. M.; DEANE, M. P.; FERREIRA NETO, J. Studies on transmission of simian malaria and on a natural infection of man with *Plasmodium simium* in Brazil. *Bulletin of the World Health Organization*, v. 35, n. 5, p. 805–808, 1966.

DELLA TORRE, A. et al. Speciation within *Anopheles gambiae* - The glass is half full. *Science*, v. 298, n. 5591, p. 115–117, 2002.

DIAS, G. R. et al. A phylogenomic study of Steganinae fruit flies (Diptera: Drosophilidae): strong gene tree heterogeneity and evidence for monophyly. *BMC Evolutionary Biology*, v. 20, n. 1, p. 1–12, 2020. Disponível em: <<https://doi.org/10.1186/s12862-020-01703-7>>.

DORN, P. L. et al. *Genetics of major insect vectors*. [s.l.: s.n.]

DOS SANTOS, M. M. M. et al. Morphological identification of species of the *Nuneztovari* Complex of *Anopheles* (Diptera: Culicidae) from an area affected by a Brazilian hydroelectric plant. *Zootaxa*, v. 4565, n. 2, p. 235–244, 2019.

DOWNS, W. G.; PITTENDRIGH, C. S. Bromeliad malaria in Trinidad, British West Indies. *The American journal of tropical medicine and hygiene*, v. 26, p. 47–66, jan. 1946. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21018584>>.

DUARTE, A. M. R. C. et al. Natural infection in anopheline species and its implications for autochthonous malaria in the Atlantic forest in Brazil. *Parasites and Vectors*, v. 6, n. 1, p. 1–6, 2013.

DYAR, H. G.; KNAB, F. Forthcoming monograph. *Proceedings of the United States National Museum*, v. 35, n. 1, p. 53–70, 1908.

EHRMAN, L.; POWELL, J. R. *Drosophila willistoni* species group. *The Genetics and Biology of Drosophila*, v. 3b, p. London, 1982.

FERREIRA, E. et al. Longevity of *Anopheles* (*Kerteszia*) *cruzzi* and *Anopheles*

(Kerteszia) bellator under natural conditions. *Rev Bras Malariol Doencas Trop*, v. 4, p. 823–8, 1969.

FILATOV, D. A. PROSEQ: A software for preparation and evolutionary analysis of DNA sequence data sets. *Molecular Ecology Notes*, v. 2, n. 4, p. 621–624, 2002.

FORATTINI, O.P., ISHIATA, G.K., RABELLO, E.X. & COTRIM, M. .
Observações sobre os mosquitos Culex da cidade de São Paulo, Brasil. *Rev. Saúde Pública*, v. 7, p. 315–30, 1973.

FORATTINI, O. P. *Entomologia Médica*. Volume 1 ed. São Paulo: Faculdade de Higiene e Saúde Pública, 1962.

FORATTINI, O. P. *Culicidologia médica*. 2. ed. São Paulo: Ed. Universidade de São Paulo, 2002.

FORATTINI, O. P.; LOPES, O. S.; RABELLO, E. X. Investigações sobre o comportamento de formas adultas de mosquitos silvestres no Estado de São Paulo, Brasil. *Revista de saúde pública*, v. 2, n. 2, p. 111–173, 1968.

FOSTER, P. G. et al. Phylogeny of anophelinae using mitochondrial protein coding genes. *Royal Society Open Science*, v. 4, n. 11, 2017.

GADELHA, P. From “forest malaria” to “bromeliad malaria”: a case-study of scientific controversy and malaria control. *Parassitologia*, v. 36, p. 175–95, 1994.

GARCIA, K. K. S. et al. Malaria time series in the extra-Amazon region of Brazil: epidemiological scenario and a two-year prediction model. *Malaria Journal*, v. 21, n. 1, p. 1–11, 2022. Disponível em: <<https://doi.org/10.1186/s12936-022-04162-1>>.

GOUDET, J. Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, v. 5, n. 1, p. 184–186, mar. 2005.
Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1471-8286.2004.00828.x>>.

GRIFFING, S. M. et al. A historical perspective on malaria control in Brazil. *Memorias do Instituto Oswaldo Cruz*, v. 110, n. 6, p. 701–718, 2015.

GUAN, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, v. 36, n. 9, p. 2896–2898, 1 maio 2020.

Disponível em:

<<https://academic.oup.com/bioinformatics/article/36/9/2896/5714742>>.

GUIMARÃES, A. É. et al. Ecology of Mosquitoes (Diptera: Culicidae) in Areas of Serra do Mar State Park, State of São Paulo, Brazil. II -Habitat Distribution.

Memorias do Instituto Oswaldo Cruz, v. 95, n. 1–2, p. 17–28, 2000.

GUIMARÃES, L. O. et al. The genetic diversity of *Plasmodium malariae* and *Plasmodium brasilianum* from human, simian and mosquito hosts in Brazil.

Acta Tropica, v. 124, n. 1, p. 27–32, 2012. Disponível em:

<<http://dx.doi.org/10.1016/j.actatropica.2012.05.016>>.

GUREVICH, A. et al. QUAST: Quality assessment tool for genome assemblies.

Bioinformatics, v. 29, n. 8, p. 1072–1075, 2013.

HACKETT, L. W. The present status of our knowledge of the sub-species of

Anopheles maculipennis. *Transactions of the Royal Society of Tropical*

Medicine and Hygiene, p. 109– 128, 1934.

HACKETT, L. W. *Malaria in Europe. An Ecological Study*. London: Oxford

University Press, 1937.

HAMILTON, J. G. C. et al. Comparison of the sex-pheromone components of

Lutzomyia longipalpis (Diptera: Psychodidae) from areas of visceral and

atypical cutaneous leishmaniasis in Honduras and Costa Rica. *Annals of*

Tropical Medicine & Parasitology, v. 90, n. 5, p. 533–541, 15 jan. 1996.

Disponível em:

<<https://www.tandfonline.com/doi/full/10.1080/00034983.1996.11813079>>.

HANG ET AL. Zoonotic malaria : Non- laverania *Plasmodium* Biology and

Invasion Mechanisms. *Pathogens*, p. 1–22, 2021.

HARBACH, R. E. *Mosquito taxonomic inventory*. Disponível em:

<[https://mosquito-taxonomic-](https://mosquito-taxonomic-inventory.myspecies.info/simpletaxonomy/term/6047/)

[inventory.myspecies.info/simpletaxonomy/term/6047/](https://mosquito-taxonomic-inventory.myspecies.info/simpletaxonomy/term/6047/)>. Acesso em: 28 maio.

2023.

HARBACH, R. E. The Phylogeny and Classification of *Anopheles*. *Anopheles mosquitoes - New insights into malaria vectors*, p. 3–55, 2013.

HARBACH, R. E. An *Anopheles* by any other name? *Journal of Medical Entomology*, v. 55, n. 5, p. 1069–1070, 2018.

HAY, S. I. et al. The global distribution and population at risk of malaria: Past, present, and future. *Lancet Infectious Diseases*, v. 4, n. 6, p. 327–336, 2004.

HAYWARD, B. . et al. *The Last Global Extinction (Mid-Pleistocene) of Deep-Sea Benthic Foraminifera (Chrysalogoniidae, Ellipsoidinidae, Glandulodosariidae, Plectofrondiculariidae, Pleurostomellidae, Stilostomellidae), their Late Cretaceous–Cenozoic History and Taxonomy*. [s.l.] Cushman Foundation for Foraminiferal Research Special Publication, 2012.

HEY, J.; PINHO, C. Population genetics and objectivity in species diagnosis. *Evolution*, v. 66, n. 5, p. 1413–1429, 2012.

HOLT, R. A. et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*, v. 298, n. 5591, p. 129–149, 4 out. 2002. Disponível em: <<https://www.science.org/doi/10.1126/science.1076181>>.

ICZN. *International Code of Zoological Nomenclature*. London: The International Trust for Zoological Nomenclature, 1999.

JIANG, X.; EDWARDS, S. V; LIU, L. The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. *Systematic Biology*, v. 69, n. 4, p. 795–812, 1 jul. 2020. Disponível em: <<https://academic.oup.com/sysbio/article/69/4/795/5721375>>.

KAJITANI, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, v. 24, n. 8, p. 1384–1395, 2014.

KATOH, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, v. 30, n. 4, p. 772–780, 2013.

KINGAN, S. et al. A high-quality de novo genome assembly from a single

mosquito using PacBio sequencing. *Genes*, v. 10, n. 1, p. 62, 18 jan. 2019. Disponível em: <<http://www.mdpi.com/2073-4425/10/1/62>>.

KIRCHGATTER, K. et al. Phylogeny of Anopheles (Kerteszia) (Diptera: Culicidae) Using Mitochondrial Genes. *Insects*, v. 11, n. 5, p. 324, maio 2020.

KNAB, F. *Reported discussion*, 1912. .

KOLMOGOROV, M. et al. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, v. 37, n. 5, p. 540–546, 1 maio 2019. Disponível em: <<https://www.nature.com/articles/s41587-019-0072-8>>.

KOREN, S. et al. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, v. 27, n. 5, p. 722–736, 2017.

KORSHUNOVA, T. et al. Multilevel fine-scale diversity challenges the ‘cryptic species’ concept. *Scientific Reports*, v. 9, n. 1, p. 1–23, 2019.

KRZYWINSKI, J.; BESANSKY, N. J. Molecular Systematics of *Anopheles* : From Subgenera to Subpopulations. *Annual Review of Entomology*, v. 48, p. 111–139, 2003.

LAETSCH, D. R.; BLAXTER, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research*, v. 6, p. 1287, 2017.

LAMPO, M. et al. Occurrence of sibling species of *Lutzomyia longipalpis* (Diptera: Psychodidae) in Venezuela: first evidence from reproductively isolated sympatric populations. *The American Journal of Tropical Medicine and Hygiene*, v. 61, n. 6, p. 1004–1009, 1 dez. 1999. Disponível em: <<http://www.ajtmh.org/content/journals/10.4269/ajtmh.1999.61.1004>>.

LAPORTA, G. Z. et al. *Plasmodium falciparum* in the southeastern Atlantic forest: A challenge to the bromeliad-malaria paradigm? *Malaria Journal*, v. 14, n. 1, p. 1–12, 2015.

LAURENS, M. B. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Human Vaccines and Immunotherapeutics*, v. 16, n. 3, p. 480–489, 2019. Disponível em: <<https://doi.org/10.1080/21645515.2019.1669415>>.

- LAWNICZAK, M. K. N. et al. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, v. 330, n. 6003, p. 512–514, 22 out. 2010. Disponível em: <<https://www.science.org/doi/10.1126/science.1195755>>.
- LEACHÉ, A. D. et al. The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, v. 63, n. 1, p. 17–30, 1 jan. 2014. Disponível em: <<https://academic.oup.com/sysbio/article/63/1/17/1687129>>.
- LEE, Y. et al. Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector *Anopheles gambiae* sensu stricto. *Malaria Journal*, v. 8, n. 1, p. 1–15, 2009.
- LEE, Y. et al. Chromosome inversions, genomic differentiation and speciation in the African malaria mosquito *Anopheles gambiae*. *PLoS ONE*, v. 8, n. 3, p. e57887, 20 mar. 2013a. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0057887>>.
- LEE, Y. et al. Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, *Anopheles gambiae*. *Proceedings of the National Academy of Sciences*, v. 110, n. 49, p. 19854–19859, 3 dez. 2013b. Disponível em: <<https://pnas.org/doi/full/10.1073/pnas.1316851110>>.
- LEVASHINA, E. A. Immune responses in *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology*, v. 34, n. 7, p. 673–678, 2004.
- LI, W. *Molecular evolution*. [s.l.] Sunderland: Sinauer Associates, 1997.
- LINTON, Y. M. et al. Morphological and molecular characterization of *Anopheles (Anopheles) maculipennis* Meigen, type species of the genus and nominotypical member of the *Maculipennis* Complex. *Systematic Entomology*, v. 28, n. 1, p. 39–56, 2003.
- LIU, L.; WU, S.; YU, L. Coalescent methods for estimating species trees from phylogenomic data. *Journal of Systematics and Evolution*, v. 53, n. 5, p. 380–390, 2015.
- LOAIZA, J. R. et al. Review of genetic diversity in malaria vectors (Culicidae: Anophelinae). *Infection, Genetics and Evolution*, v. 12, n. 1, p. 1–12, 2012.

Disponível em: <<http://dx.doi.org/10.1016/j.meegid.2011.08.004>>.

LORENZ, C.; PATANÉ, J. S. L.; SUESDEK, L. Morphogenetic characterisation, date of divergence, and evolutionary relationships of malaria vectors *Anopheles cruzii* and *Anopheles homunculus*. *Infection, Genetics and Evolution*, v. 35, p. 144–152, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.meegid.2015.08.011>>.

LUTZ, A. Mosquitos da floresta e malária silvestre *. *Archiv*, p. 759–768, 1903.

MAI, U.; MIRARAB, S. TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, v. 19, n. Suppl 5, 2018.

MALAFRONTI, R. D. S. et al. Intraspecific variation of second internal transcribed spacer of nuclear ribosomal DNA among populations of *Anopheles* (*Kerteszia*) *cruzii* (Diptera: Culicidae). *Journal of Medical Entomology*, v. 44, n. 3, p. 538–542, 1 maio 2007. Disponível em: <<https://academic.oup.com/jme/article-lookup/doi/10.1093/jmedent/44.3.538>>.

MANGABEIRA, O. Sobre a sistemática e biologia dos flebótomos do Ceará. *Revista Brasileira de Malariologia e Doenças Tropicais*, v. 21, p. 3–26, 1969.

MANNI, M. et al. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, v. 38, n. 10, p. 4647–4654, 2021.

MARÇAIS, G.; KINGSFORD, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, v. 27, n. 6, p. 764–770, 2011.

MARINOTTI, O. et al. The Genome of *Anopheles darlingi*, the main neotropical malaria vector. *Nucleic Acids Research*, v. 41, n. 15, p. 7387–7400, ago. 2013. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt484>>.

MARRELLI, M. T. et al. *Kerteszia* subgenus of *Anopheles* associated with the Brazilian Atlantic rainforest: Current knowledge and future challenges. *Malaria*

Journal, v. 6, p. 1–8, 2007.

MAYR, E. *Populations, species, and evolution: An abridgment of animal species and evolution*. [s.l.] Belknap Press of Harvard University Press, 1970.

MIGUEL, R. B. et al. Malaria in the state of Rio de Janeiro, Brazil, an Atlantic Forest area: an assessment using the health surveillance service. *Memórias do Instituto Oswaldo Cruz*, v. 109, n. 5, p. 634–640, 13 ago. 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762014000500634&lng=en&tlng=en>.

MILNER, D. A. Malaria pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, v. 8, n. 1, p. 1–11, 2018.

MIRARAB, S. et al. ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics*, v. 30, n. 17, p. 541–548, 2014.

MORIYAMA, E. N.; GOJOBORI, T. Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. *Genetics*, v. 130, n. 4, p. 855–864, 1 abr. 1992. Disponível em: <<https://academic.oup.com/genetics/article/130/4/855/6007208>>.

MS. *Malária em 2021: Buscando o caminho para a eliminação da malária no Brasil*. Disponível em: <<https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2022/boletim-epidemiologico-vol-53-no17.pdf>>. Acesso em: 28 maio. 2023.

MUKHOPADHYAY, J. et al. Genetic polymorphism of morphological and biochemical characters in a Natal, Brazil, population of *Lutzomyia longipalpis* (Diptera: Psychodidae). *Journal of the American Mosquito Control Association*, v. 14, n. 3, p. 277–82, set. 1998. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9813825>>.

NEAFSEY, D. E. et al. Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. *Science*, v. 347, n. 6217, jan. 2015.

NEVES, A. et al. Malaria outside the Amazon region: Natural *Plasmodium* infection in anophelines collected near an indigenous village in the Vale do Rio Branco, Itanhaém, SP, Brazil. *Acta Tropica*, v. 125, n. 1, p. 102–106, 2013.

Disponível em: <<http://dx.doi.org/10.1016/j.actatropica.2012.08.014>>.

NEVES, D. P. *Parasitologia humana*. São Paulo: Atheneu, 2016.

NGUYEN, L. T. et al. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, v. 32, n. 1, p. 268–274, 2015.

NOOR, M. A. F. et al. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, v. 98, n. 21, p. 12084–12088, 2001.

NORRIS, D. E. et al. Microsatellite DNA Polymorphism and Heterozygosity Among Field and Laboratory Populations of *Anopheles gambiae* s.s. (Diptera: Culicidae). *Journal of Medical Entomology*, v. 38, n. 2, p. 336–340, 1 mar. 2001. Disponível em: <<https://academic.oup.com/jme/article-lookup/doi/10.1603/0022-2585-38.2.336>>.

OBBARD, D. J. et al. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Molecular Biology and Evolution*, v. 29, n. 11, p. 3459–3473, nov. 2012. Disponível em: <<https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss150>>.

OLIVEIRA-FERREIRA, J. et al. Malaria in Brazil: an overview. *Malaria Journal*, v. 9, n. 1, p. 115, 30 dez. 2010. Disponível em: <<https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-9-115>>.

OLIVEIRA, T. M. P. et al. Mitochondrial Genomes of *Anopheles* (Kerteszia) (Diptera: Culicidae) From the Atlantic Forest, Brazil. *Journal of Medical Entomology*, v. 53, n. 4, p. 790–797, jul. 2016. Disponível em: <<https://academic.oup.com/jme/article-lookup/doi/10.1093/jme/tjw001>>.

PAMILO, P.; NEI, M. Relationship between gene trees and species trees. *Molecular Biology and Evolution*, v. 5, n. 5, p. 568–583, 1988.

PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, v. 20, n. 2, p. 289–290, 2004.

PFEIFER, B. et al. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology and Evolution*, v. 31, n. 7, p. 1929–

1936, jul. 2014. Disponível em: <<https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu136>>.

R CORE TEAM. *R: A language and environment for statistical computing* Vienna, Austria R Foundation for Statistical Computing, , 2014. .

R CORE TEAM. *A language and environment for statistical computing*. Vienna, Austria R Foundation for Statistical Computing, , 2021. .

RACHOU, R. Anofelinos do Brasil: Comportamento das espécies vetoras de malária. *Rev Bras Malariol Doenças Trop*, v. 10, p. 145–181, 1958.

RAMIREZ, C. C. L.; DESSEN, E. M. B. Chromosomal evidence for sibling species of the malaria vector *Anopheles cruzii*. *Genome*, v. 43, n. 1, p. 143–151, 2000a.

RAMÍREZ, C. C. L.; DESSEN, E. M. B. *The polytene chromosomes of the mosquito Anopheles bellator compared with those of Anopheles cruzii* Brazilian *Journal of Genetics*, 1996. .

RAMIREZ, C.; DESSEN, E. Chromosome differentiated populations of *Anopheles cruzii*: evidence for a third sibling species. *Genetica*, v. 1, n. 108, p. 73–80, 2000b.

REIS, M. d.; YANG, Z. Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. *Molecular Biology and Evolution*, v. 28, n. 7, p. 2161–2172, 1 jul. 2011. Disponível em: <<https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msr045>>.

ROCHES, S. Des et al. The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, v. 2, n. January, 2018a. Disponível em: <<http://dx.doi.org/10.1038/s41559-017-0402-5>>.

ROCHES, S. Des et al. The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, v. 2, n. January, 2018b.

ROCHES, S. Des et al. Conserving intraspecific variation for nature's contributions to people. *Nature Ecology & Evolution*, v. 5, n. May, 2021a. Disponível em: <<http://dx.doi.org/10.1038/s41559-021-01403-5>>.

- ROCHES, S. Des et al. Conserving intraspecific variation for nature's contributions to people. *Nature Ecology & Evolution*, v. 5, n. May, 2021b.
- RODRIGUES, P. T. et al. Human migration and the spread of malaria parasites to the New World. *Scientific Reports*, v. 8, n. 1, p. 1–13, 2018.
- RONA, L. D. et al. Assessing the molecular divergence between *Anopheles (Kerteszia) cruzii* populations from Brazil using the timeless gene: Further evidence of a species complex. *Malaria Journal*, v. 8, n. 1, p. 1–10, 2009.
- RONA, L. D.; CARVALHO-PINTO, C. J. Evidence for the occurrence of two sympatric sibling species within the *Anopheles (Kerteszia) cruzii* complex in southeast Brazil and the detection of asymmetric introgression between them using a multilocus analysis. *BMC Evolutionary Biology*, v. 13, n. 1, p. 1, 2013. Disponível em: <BMC Evolutionary Biology>.
- RONA, L. D.; CARVALHO-PINTO, C. J.; PEIXOTO, A. A. Molecular evidence for the occurrence of a new sibling species within the *Anopheles (Kerteszia) cruzii* complex in south-east Brazil. *Malaria Journal*, v. 9, n. 1, p. 1–9, 2010.
- RONA, L. D. P. et al. Estimation of divergence time between two sibling species of the *Anopheles (Kerteszia) cruzii* complex using a multilocus approach. *BMC Evolutionary Biology*, v. 10, n. 1, 2010.
- ROZAS, J. et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, v. 19, n. 18, p. 2496–2497, 2003.
- RUIZ-LOPEZ, F. et al. Systematics of the oswaldoi complex *Anopheles, Nyssorhynchus* in South America. *Parasites and Vectors*, v. 6, n. 1, p. 1–13, 2013.
- RUSSO, C. A. M.; TAKEZAKI, N.; NEI, M. Molecular phylogeny and divergence times of drosophilid species. *Molecular Biology and Evolution*, v. 12, n. 3, p. 391–404, 1995.
- SALLUM, M. A. M. et al. Identification key to the *Anopheles* mosquitoes of South America (Diptera: Culicidae). III. Male genitalia. *Parasites and Vectors*, v. 13, n. 1, p. 1–23, 2020. Disponível em: <<https://doi.org/10.1186/s13071-020-04300-1>>.

SANTOS, D. et al. A non-destructive enzymatic method to extract DNA from arthropod specimens: Implications for morphological and molecular studies. *PLoS ONE*, v. 13, n. 2, p. 1–11, 2018.

SANTOS, F. A. B.; LEMES, R. B.; OTTO, P. A. HW_TEST, a program for comprehensive HARDY-WEINBERG equilibrium testing. *Genetics and Molecular Biology*, v. 43, n. 2, 2020a. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572020000400604&tlng=en>.

SANTOS, F. A. B.; LEMES, R. B.; OTTO, P. A. HW_TEST, a program for comprehensive HARDY-WEINBERG equilibrium testing. *Genetics and Molecular Biology*, v. 43, n. 2, 2020b.

SCHLIEP, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics*, v. 27, n. 4, p. 592–593, 2011.

SHEN, X.-X.; STEENWYK, J. L.; ROKAS, A. Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. *Systematic Biology*, v. 70, n. 5, p. 997–1014, 11 ago. 2021. Disponível em: <<https://academic.oup.com/sysbio/article/70/5/997/6146422>>.

SHERLOCK, I. A.; SHERLOCK, V. A. On the experimental infection of *Phlebotomus longipalpis* by *Leishmania donovani*. *Revista brasileira de biologia*, v. 22, p. 409–18, dez. 1961. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14039201>>.

SILVA-DO-NASCIMENTO, T. F. et al. Molecular divergence in the timeless and cpr genes among three sympatric cryptic species of the *Anopheles triannulatus* complex. *Memorias do Instituto Oswaldo Cruz*, v. 106, n. SUPPL. 1, p. 218–222, 2011.

SIMÃO, F. A. et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, v. 31, n. 19, p. 3210–3212, 2015.

SINAN. *Doenças e Agravos de Notificação - 2007 em diante - Malária*. Disponível em: <<https://datasus.saude.gov.br/acesso-a-informacao/doencas-e-agravos-de-notificacao-de-2007-em-diante-sinan/>>. Acesso em: 2 fev. 2023.

- SINKA, M. E. et al. The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: Occurrence data, distribution maps and bionomic précis. *Parasites and Vectors*, v. 3, n. 1, p. 117, 2010. Disponível em: <<http://www.parasitesandvectors.com/content/3/1/117>>.
- SINKA, M. E. et al. A global map of dominant malaria vectors. *Parasites and Vectors*, v. 5, n. 1, p. 69, 2012. Disponível em: <<http://www.parasitesandvectors.com/content/5/1/69>>.
- SMITH, B. T. et al. Uneven missing data skew phylogenomic relationships within the lorises and lorikeets. *Genome Biology and Evolution*, v. 12, n. 7, p. 1131–1147, 2020.
- SMITH, L. B. *Bromeliad malaria* Annu Rep Smithson Inst, , 1952. .
- SMITH, M. L.; STYCZYNSKI, M. P. Systems Biology-Based Investigation of Host–Plasmodium Interactions. *Trends in Parasitology*, v. 34, n. 7, p. 617–632, 2018. Disponível em: <<https://doi.org/10.1016/j.pt.2018.04.003>>.
- SMITH, S. A.; BROWN, J. W. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, v. 105, n. 3, p. 302–314, 14 mar. 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/ajb2.1019>>.
- SPASSKY, B. Morphological differences between sibling species of *Drosophila*. *University of Texas Publications*, v. 5721, p. 48–61, 1957.
- STANLEY C. OAKS, JR., VIOLAINE S. MITCHELL, G. W.; PEARSON, and C. C. J. C. *Malaria — obstacles and opportunities*. [s.l.: s.n.]v. 8
- STRUCK, T. H. et al. Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, v. 33, n. 3, p. 153–163, mar. 2018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169534717302902>>.
- TAJIMA, F. Evolutionary relationship of dna sequences in finite populations. *Genetics*, v. 105, n. 2, p. 437–460, 1 out. 1983. Disponível em: <<https://academic.oup.com/genetics/article/105/2/437/5996285>>.
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, v. 123, n. 3, p. 585–595, 1989.

TAKKEN, W. et al. Distribution and Dynamics of Larval Populations of *Anopheles messeae* and *A. atroparvus* in the Delta of the Rivers Rhine and Meuse, The Netherlands. *AMBIO: A Journal of the Human Environment*, v. 31, n. 3, p. 212–218, maio 2002. Disponível em:

<<http://www.bioone.org/doi/abs/10.1579/0044-7447-31.3.212>>.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, v. 38, n. 7, p. 3022–3027, 25 jun. 2021. Disponível em:

<<https://academic.oup.com/mbe/article/38/7/3022/6248099>>.

TAMURA, K.; TAO, Q.; KUMAR, S. Theoretical Foundation of the RelTime Method for Estimating Divergence Times from Variable Evolutionary Rates.

Molecular Biology and Evolution, v. 35, n. 7, p. 1770–1782, 1 jul. 2018.

Disponível em: <<https://academic.oup.com/mbe/article/35/7/1770/4947869>>.

TENNESSEN, J. A. et al. A population genomic unveiling of a new cryptic mosquito taxon within the malaria-transmitting *Anopheles gambiae* complex.

Molecular Ecology, v. 30, n. 3, p. 775–790, 10 fev. 2021. Disponível em:

<<https://onlinelibrary.wiley.com/doi/10.1111/mec.15756>>.

THE ANOPHELES GAMBIAE 1000 GENOMES CONSORTIUM. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*, v. 552, p. 96–100, 2017.

THEOBALD, F. V. *A Monograph of the Culicidae or Mosquitoes*. 1901. Mainly Compiled from the Collections Received at the British Museum from Various Parts of the World, 1901.

TOURÉ, Y. T. et al. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia*, v. 40, p. 477–511, 1998.

TURNER, T. L.; HAHN, M. W.; NUZHDIN, S. V. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, v. 3, n. 9, p. 1572–1578, 2005.

UENO, H. M.; FORATTINI, O. P.; KAKITANI, I. Vertical and seasonal distribution of *Anopheles (Kerteszia)* in Ilha Comprida, Southeastern Brazil. *Revista de Saude Publica*, v. 41, n. 2, p. 269–275, 2007.

- VAN THIEL, P. H. *Sur l'origine des variations de taille de l'Anopheles maculipennis dans les Pays-Bas*. [s.l.: s.n.].
- VELOSO, H. P.; MOURA, J. V. de; KLEIN, R. M. *Delimitação ecológica dos anofelíneos do subgênero Kerteszia na região costeira do sul do Brasil* Memórias do Instituto Oswaldo Cruz, 1956. .
- VICOSO, B.; CHARLESWORTH, B. Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews Genetics*, v. 7, n. 8, p. 645–653, ago. 2006. Disponível em: <<https://www.nature.com/articles/nrg1914>>.
- VIOLLE, C. et al. The return of the variance : intraspecific variability in community ecology. v. 27, n. 4, 2012.
- VOGES, K. et al. Novel molecular evidence of population structure in *Anopheles (Kerteszia) bellator* from brazilian atlantic forest. *Memorias do Instituto Oswaldo Cruz*, v. 114, n. 4, p. 1–5, 2019.
- VURTURE, G. W. et al. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, v. 33, n. 14, p. 2202–2204, 2017.
- WALKER, B. J. et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, v. 9, n. 11, p. e112963, 19 nov. 2014. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0112963>>.
- WATERHOUSE, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, v. 35, n. 3, p. 543–548, 2018.
- WHITE, B. J. et al. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*, v. 19, n. 5, p. 925–939, mar. 2010. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2010.04531.x>>.
- WHO. *Malaria World Report 2022*. [s.l.: s.n.].
- WHO. *Malaria, Key facts*. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/malária>>. Acesso em: 28 maio. 2023.

WILKERSON, R.; PEYTON, E. The Brazilian malaria vector *Anopheles (Kerteszia) cruzii*: life stages and biology (Diptera: Culicidae). *Mosquito Systematics*, v. 23, n. 2, p. 110–122, 1991. Disponível em: <<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA512864>>.

WRBU. *Anopheles cruzii* species page. Disponível em: <<https://wrbu.si.edu/vectorspecies/mosquitoes/cruzii>>.

YALCINDAG, E. et al. Multiple independent introductions of *Plasmodium falciparum* in South America. *Proceedings of the National Academy of Sciences of the United States of America*, v. 109, n. 2, p. 511–516, 2012.

YOUNG, D. G.; DUNCAN, M. A. *Guide to the identification and geographic distribution of Lutzomyia sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae)*. Gainesville, FL.: Memories American Entomological Institute No. 54, Associated Publishers., 1994.

ZACHOS, F. E. *Species Concepts in Biology*. Cham: Springer International Publishing, 2016.

ZAMYATIN, A. et al. Chromosome-level genome assemblies of the malaria vectors *Anopheles coluzzii* and *Anopheles arabiensis*. *GigaScience*, v. 10, n. 3, 15 mar. 2021. Disponível em: <<https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giab017/6170950>>.

ZAVORTINK, T. A review of the subgenus *Kerteszia* of *Anopheles*. Contributions of the American Entomological Institute. *Contributions of the American Entomological Institute*, v. 9, p. 1–54, 1973.

Apêndice 1

A genomic study of cryptic speciation in *Anopheles (Kerteszia) cruzii*, the main malaria vector in the Brazilian Atlantic Forest

Kamila Voges¹, Guilherme de Rezende Dias², Eduardo Guimarães Dupim², André Nóbrega Pitaluga^{3,4}, Tiago Vanderlinde², Carlos José de Carvalho Pinto^{4,5}, Helder Ricas Rezende⁶, Sarah Jayne Forrester⁷, James Chong⁷, A. Bernardo Carvalho^{†,2,4}, Luísa D. P. Rona^{†,1,4,§}

¹Universidade Federal de Santa Catarina, Departamento de Biologia Celular, Embriologia e Genética, Florianópolis, Brasil

²Universidade Federal do Rio de Janeiro, Departamento de Genética, Rio de Janeiro, Brasil

³Laboratório de Biologia Molecular de Parasitas e Vetores, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brasil

⁴Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular, Rio de Janeiro, Brasil

⁵Universidade Federal de Santa Catarina, Departamento de Microbiologia, Imunologia e Parasitologia, Florianópolis, Brasil

⁶Secretaria de Estado da Saúde do Espírito Santo, Vitória, ES, Brasil

⁷Department of Biology, University of York, Wentworth Way, Heslington, York YO10 5DD, UK.

[§]Corresponding author: luisa.rona@ufsc.br

[†]Equal senior authors

Abstract

Background: Malaria is a tropical disease caused by *Plasmodium* protozoan, that affects millions of people annually. Although in Brazil most cases occur in the Amazon, there is still transmission outside this region, as in the Atlantic Forest where the main vector is *Anopheles (Kerteszia) cruzii*. Data from cytogenetics, isoenzymes, and molecular markers suggested that *An. cruzii* is a complex of cryptic species. Here we used a genomic approach to confirm the existence of cryptic species, and investigate the evolutionary relationships within this complex, and their relationship with other *Kerteszia*.

Methodology: In order to perform genetic differentiation analysis and phylogenomic inferences, we sequenced 38 complete genomes from different *An. cruzii* sensu lato groups and related *Kerteszia* species, using Illumina and Nanopore technologies.

Results: We report the first genome drafts for several species of *Kerteszia*. The phylogenomic and F_{ST} analyses indicates that *An. cruzii* s.l. is composed of at least four cryptic species: (i) South/Southeast Brazil, represented here by Florianópolis-SC and Guapimirim-RJ (Serra do Mar coastal region); (ii) Bocaina Group 1 and (iii) Bocaina Group 2 (Serra do Mar continental region, in Southeast Brazil); Itatiaia - RJ (Serra da Mantiqueira, Southeast Brazil); (iv) Santa Teresa - ES (Southeast Brazil). The evidence of cryptic speciation is particularly compelling in the two Bocaina groups: despite sympatry, these groups show a very high F_{ST} and lack of inter-group heterozygotes, indicating complete or nearly complete reproductive isolation. On the other hand, the Florianópolis and Guapimirim populations, which are ~1000 km apart, showed no clear genetic separation and could be considered a single species. We also uncovered a new species from Santa Teresa - ES, which is genetically and morphologically distinguishable from the other three groups mentioned above.

Conclusions: This work encompasses the largest number of genes used to evaluate the genetic differentiation of *An. cruzii* s.l. and corroborates the cryptic speciation hypothesis.

Key words: Malaria, *Anopheles (Kerteszia) cruzii*, genomic analysis, species complex, cryptic speciation.

Background

Malaria is a serious public health problem, which affects millions of people globally every year. According to the latest report of the World Health Organization, 247 million cases of the disease occurred in 2021 (WHO, 2022). Although in Brazil most malaria cases occur in the Amazon region, autochthonous events are also reported in areas of Atlantic Forest in southern and southeastern Brazil, mainly in the States of Rio de Janeiro, São Paulo, and Santa Catarina. The main malaria vector in these places is *Anopheles (Kerteszia) cruzii*, a bromeliad breeding mosquito, which transmits both human and simian malaria (LUTZ, 1903; DEANE, 1986; MARRELLI et al., 2007; DE PINA-COSTA et al., 2014; BRASIL et al., 2017). This anopheline occurs from the Rio Grande do Sul (South Brazil) to Sergipe (Northeast Brazil) States, where bromeliads are abundant (CONSOLI; LOURENÇO DE OLIVEIRA, 1994), and is also known as a “bromeliad malaria” vector (RACHOU, 1958; GADELHA, 1994).

Many genetic studies have suggested that *An. cruzii* is a complex of cryptic species (Table 1). First, differences in inversion patterns of the *X* and *3L* chromosomes among *An. cruzii* populations from South and Southeast Brazil indicated the existence of independent evolutionary units for this species (RAMIREZ; DESSEN, 2000a, 2000b). Second, isoenzyme analyses using populations from Santa Catarina, São Paulo, Rio de Janeiro, and Bahia States suggested that the latter (from Northeast Brazil) is genetically distant from the other populations from South and Southeast Brazil (CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2004). Third, studies using a different genes (i) suggested that the populations of *An. cruzii* located on the lower coastal s.l.opes of the Serra do Mar mountains (in altitudes below 600 m) from South to Southeast Brazil belong to the same species. These populations span a range of ~1000 km. (ii) reinforced the hypothesis that the Bahia population is a different species within the *An. cruzii* complex, and (iii) uncovered at least three other sibling species inhabiting higher altitudes (above 900 m) of Serra do Mar and Serra da Mantiqueira (the later is another mountain range in Southeast Brazil). Strikingly, the Bocaina population in the continental side of Serra do Mar mountains was found to contain two reproductively isolated groups in sympatry (RONA et al., 2009, 2010; RONA; CARVALHO-PINTO; PEIXOTO, 2010; RONA; CARVALHO-PINTO, 2013; DE REZENDE DIAS et al., 2018). It is interesting to note that Dias et al (DE REZENDE DIAS et al., 2018) studied two nuclear loci, *cpr* (presumably X-linked) and *clock* (presumably autosomal) and found that the genetic differentiation was usually much stronger for *cpr*, hinting on the possible existence of "speciation islands", as was found in *An. gambiae* s.l. (TURNER; HAHN; NUZHIDIN, 2005).

Table 1: Some studies that suggest that *An. cruzii* is a cryptic species complex.

Trait	Populations sampled	Differentiated populations	Comments	Ref
Larvae	Rio de Janeiro (RJ), Brusque (SC)	Rio de Janeiro (RJ) and Brusque (SC).	The author reported that the larvae from Rio de Janeiro differed from those from Brusque (SC) and questioned the cospecificity.	(ZAVORTINK, 1973)
Chromosomal inversions	São Francisco do Sul (SC), Cananéia (SP), Guaratuba (SP)	Occurrence of two divergent forms in the species, each carrying a different set of inversions, which were named A and B.	Ramirez and Dessen, studying the polytene chromosomal banding pattern of <i>An. cruzii</i> from São Paulo and Santa Catarina, observed the existence of two distinct chromosomal forms and the absence of heterozygotes.	(RAMIREZ; DESSEN, 2000a)
Chromosomal inversions	Boracéia (SP) and Jquitiba (SP)	And the “C” form can occur in sympatry with the A form in some places.	They found a third chromosome form in two areas of São Paulo - at least three cryptic species.	(RAMIREZ; DESSEN, 2000b)
Isoenzymes	Florianópolis (SC), Cananéia (SP), Nova	Bahia is different from the South and Southeast.	Specimens from the state of Bahia can be considered as	(CARVALHO-PINTO;

	Iguaçu (RJ), Jaguaripe (BA)		belonging to a species distinct from three other nearby populations.	LOURENÇO-DE-OLIVEIRA, 2004)
ITS2	São Francisco do Sul (SC), Cananéia (SP), Guaratuba (SP), Peruíbe (SP), Juquitiba (SP), Bertioga (SP).		They demonstrated the existence of four ITS2 sequences with a high degree of polymorphism among <i>An. cruzii</i> captured in the states of SP and SC. But this information alone is not enough to determine the occurrence of a species complex.	(MALAFRONTTE et al., 2007)
<i>Timeless</i>	Florianópolis (SC), Cananéia (SP), Juquitiba (SP), Itatiaia (RJ), Santa Teresa (ES), Itaparica (BA)	At least two cryptic species, one occurring in the northeast (of Bahia) and another in the south and southeast of Brazil.		(RONA et al., 2009)
<i>cpr</i>	Florianópolis (SC), Cananéia (SP), Juquitiba (SP), Itatiaia (RJ), Santa Teresa (ES)	The sample from Itatiaia represents a new sister species in this complex.	The <i>cpr</i> gene revealed very high F_{ST} values and fixed differences between Itatiaia and the other four studied populations (Florianópolis, Cananéia, Juquitiba and Santa Teresa).	(RONA; CARVALHO-PINTO; PEIXOTO, 2010)

<i>Timeless, Clock, cycle, Rp49, RpS2, RpS29</i>	Florianópolis (SC) and Itaparica (BA)	The results corroborate the hypothesis that the populations of the South and Northeast constitute two cryptic species of the complex.		(RONA et al., 2010)
<i>timeless (tim), Clock (Clk) and cycle (cyc); and three code for ribosomal proteins: Rp49 (Ribosomal protein 49, known also as RpL32– Ribosomal protein L32), RpS2 (Ribosomal protein S2) and RpS29 (Ribosomal protein S29)</i>	Florianópolis (SC) and Itatiaia (RJ)	Florianópolis and Itatiaia represent different species of <i>An. cruzii</i> . And Itatiaia is composed of two sympatric incipient species A and B.		(RONA; CARVALHO-PINTO, 2013)
Mitochondrial genome	São Paulo (SP), Cananéia (SP), Itatiaia (RJ), Maquiné (RS).		Specimens from four populations of <i>An. cruzii</i> showed differences in codon composition.	(OLIVEIRA et al., 2016)
<i>cpr</i> and <i>clock</i>	Bocaina (SP), Tinguá (RJ), Guapimirim (RJ), Sana (Sana), Itatiaia (RJ), Florianópolis	Two sister species inhabit higher altitudes (above 900 m) in Serra do Mar and Serra da Mantiqueira (the latter is another mountain range in southeastern Brazil).		(DE REZENDE DIAS et al., 2018)

COI cytochrome c oxidase subunit I gene; ND4 NADH dehydrogenase subunit 4 gene	Itatiaia (RJ) (Serra da Mantiqueira); Parque Estadual da Cantareira (SP) (Serra da Cantareira); Cananéia (SP), Caraguatatuba(SP), Itanhaém (SP), and São Paulo (SP) (all in Serra do Mar).	<i>Anopheles cruzii</i> grouped into three separate clades according to collection locations: Serra do Mar, Serra da Mantiqueira and Serra da Cantareira. Confirming <i>An. cruzii</i> as a species complex, composed of at least three putative species.		(KIRCHGATTER et al., 2020)
Male genitalia	Florianópolis (SC), Guapimirim (RJ), Bocaina (SP), Itatiaia (RJ), Santa Teresa (ES)	(i) South/Southeast Brazil, represented here by Florianópolis-SC and Guapimirim-RJ (coastal region of Serra do Mar); (ii) Bocaina Group 1 and (iii) Bocaina Group 2 (mainland region of Serra do Mar, in Southeast Brazil) and Itatiaia - RJ (Serra da Mantiqueira, Southeast Brazil); (iv) Santa Teresa - ES (Southeastern Brazil).	We observed differences in male genitalia only in the sample from Santa Teresa.	This study
Complete genome	Florianópolis (SC), Guapimirim (RJ), Bocaina (SP), Itatiaia (RJ), Santa Teresa (ES)	(i) South/Southeast Brazil, represented here by Florianópolis-SC and Guapimirim-RJ (coastal region of Serra do Mar); (ii) Bocaina Group 1 and (iii) Bocaina Group 2 (mainland region of Serra do Mar, in Southeast Brazil) and Itatiaia - RJ (Serra da Mantiqueira, Southeast Brazil); (iv) Santa Teresa - ES (Southeastern Brazil).	Two sympatric species in Bocaina.	This study

In the well-known species complex *Anopheles gambiae* s.l. (from Africa), the sibling species differ in malaria transmission competence (COLUZZI et al., 1979; DELLA TORRE et al., 2002). Hence, the co-occurrence of more than one lineage of *An. cruzii* in the Atlantic Forest may have important implications for malaria transmission and control in Brazil. It is interesting to note that all historic and current autochthonous cases of human malaria in the S/SE of Brazil (e.g, Florianópolis, Guapimirim; match the distribution of only one of the *An. cruzii* cryptic species detected by molecular methods (DE PINA-COSTA et al., 2014; MIGUEL et al., 2014; LORENZ; PATANÉ; SUESDEK, 2015). Hence, the knowledge of the population genomic structure of this species complex is relevant to the identification of which members are true malaria vectors.

In this work, the relationship among different lineages of *An. cruzii* from the Brazilian Atlantic Forest was dissected using a genomic approach.

Methodology

Mosquito collection - The mosquitoes used in this study were captured along the Brazilian Atlantic Forest in the following locations: Itaparica (Bahia State-BA), Camacan (Bahia State-BA), Santa Teresa (Espírito Santo State-ES), Guapimirim (Rio de Janeiro State-RJ), Ilha Grande (RJ), Itatiaia (RJ), Bocaina (São Paulo State-SP), and Florianópolis (Santa Catarina State-SC) (Figure 1).

The adult and immature mosquitoes were collected, treated, and preserved as described in Dias et al. (2018), Detailed information about the sample collection, such as coordinates, sex of mosquitoes, and the life stage at which samples were collected (adult or larvae) are provided in Additional file 1. Species identification was performed according to Consoli and Lourenço-de-Oliveira (1994) and Forattini, (2002). The morphological structures important for species-level identification were photographed with a stereomicroscope Olympus SZX16 for each sequenced individual.

DNA isolation - We used a non-destructive enzymatic method to extract DNA from individual mosquitoes, modified from Santos et al. (2018). This method allows the extraction of DNA while preserving the morphology of the taxonomically relevant parts of the animals (e.g.: exoskeleton, and male genitalia). The samples were previously preserved in ethanol 100% at -20°C, and DNA was extracted with the Qiagen Puregene core kit (Qiagen catalogue number: 158667) as follows. First, the animals were split at the abdomen-thorax junction and placed in individual tubes containing 100µl of lysis solution and 1µl of proteinase K (20 mg/mL, Qiagen catalogue number: 158918). The samples were incubated in this solution for 3 days at 45°C and subsequently incubated on ice for 1 minute. Then, 33µl of precipitation solution was added to each sample and homogenized by inversion. The samples were put back on the ice for 5 minutes and centrifuged for 3 minutes at 21.130 ref . We transferred the supernatant to a new tube, discarded the old tube containing the pellet, and added 0.5 µl of RNase (4 ug/mL, Qiagen catalogue number: 158922). The tubes were incubated (i) at 65°C for 15 minutes, (ii) at 37°C for 15 minutes; 2 µl of proteinase K was added, and (iii) they were incubated at 65°C for 15 minutes. The samples were placed on ice for 1 minute. Then, 33 µl of precipitation solution was added, homogenized by inversion, and placed on ice for 5 min. The tubes were centrifuged again at 21.130 ref for 3 minutes . The supernatant was placed in a fresh tube with 1 µl of Invitrogen™ GlycoBlue™ Coprecipitant (15 mg/mL, catalogue number: AM9515), mixed, and then 100 µl of isopropanol was added and

homogenized by inversion. The tubes were centrifuged at 21.130 rcf for 5 minutes, the supernatant was discarded, and the blue pellets were left to dry briefly at room temperature. The pellet was washed with 100 µl of ethanol 70% and centrifuged at 21.130 rcf for 1 minute. The supernatant was then discarded, and the pellet was left to dry at room temperature for approximately 10 minutes. Then, 50µl of DNA hydration solution was added, and the tubes were incubated at 65°C for 1 hour and left at room temperature overnight (to solubilize the DNA). The DNA concentration was estimated with Qubit using the dsDNA Quantitation, high sensitivity reagents (Invitrogen catalogue number Q32851), and the tubes were then stored at -20°C until sequencing. We followed the same general procedures to sequence samples with Nanopore.

Morphological characters of the male genitalia - Analysis of male genitalia is the best method for morphological species identification in *Anopheles* (e.g., (SALLUM et al., 2020)), and was used in the present study. We modified the protocol from Consoli & Lourenço-de-Oliveira (1994) as follows. The exoskeletons that were conserved following the process of DNA extraction were carefully clipped from the abdomen, and subsequently cleared in KOH 10% for a period of 12 hours. Subsequently, the samples were submitted to a dehydration process using a graduated series of ethanol, starting with a 70% solution, followed by 80%, and 90% for 15 minutes. This was then followed by immersion in a solution of 95% and absolute ethanol, for ten minutes each. In order to achieve optimum visibility, a second clarification step was carried out using xylene for 60 minutes, slides were mounted in Entellan.

Bocaina Group 1 and Bocaina Group 2 identification - In order to identify the two sympatric species from Bocaina, a partial fragment of the *cpr* gene was used, since the gene has previously been shown to be a good marker for differentiating these two groups (DE REZENDE DIAS et al., 2018). We designed two forward primers specific for each sibling from Bocaina, named Boc1F (5'GTGTAATATGGTAAGCGAACG) and Boc2F (5'GTGTAATATGGTAAGCGAAACG), which in combination with the shared reverse primer BocR (5'TTTCTCGATGTCTTTCAGCT) allowed allele-specific amplification of a *cpr* fragment (Figure 2). PCR amplifications were carried out in a Applied Biosystems Veriti 96-Well Thermal Cycler, Model 9902. We used GoTaq® Hot Start Polymerase (Promega cat num. M500B) and the following conditions: 95°C for 9 minutes, following 40 cycles at 90°C for 30 seconds, 62°C for 30 seconds, and 72°C for 20 seconds, following a cycle of 72°C for 7 minutes. The PCR products were applied to agarose gels (1%), stained with Ethidium Bromide and photographed in a UV transilluminator.

Sequencing, genome assembly, and removal of contaminants - Illumina 150 bp paired-end TruSeq Nano DNA or Nextera XT (insert size 350bp) libraries were produced and sequenced with a HiSeq 2000 at Macrogen, Korea (Table 2). Reads were assembled with both SPAdes-3.12.0 (BANKEVICH et al., 2012) and Platanus 1.2.4. (KAJITANI et al., 2014). We then compared the quality of these assemblies using BUSCO v3 (SIMÃO et al., 2015) and the Diptera reference set of orthologs (odb9, downloaded from <https://BUSCO.ezlab.org/> and containing 2,799 genes). After choosing the best assembly, we used Blobtools (LAETSCH; BLAXTER, 2017) to detect and remove contaminant contigs in the primary assembly (non-Arthropod; typically of bacterial origin). BUSCO v4.1.4 (MANNI et al., 2021) and the Diptera reference set odb10 were used to evaluate the final completeness of all samples. QUAST (GUREVICH et al., 2013) was used to estimate other assembly quality statistics (e.g.: N50). Jellyfish-2.3.0 (MARÇAIS; KINGSFORD, 2011) and GenomeScope (VURTURE et al., 2017) used to estimate genome size based on the raw read coverage.

Three individuals were sequenced using Nanopore technology on a MinION

Mk1C instrument, with the sequencing performed using SQK-LSK110 reagents and MinION R9.4.1 flow cells (FLO-MIN106). To evaluate the performance of different assemblers, we used both Canu 2.1.1 (KOREN et al., 2017) and Flye 2.9.1 (KOLMOGOROV et al., 2019). At the end of the assembly process, Flye 2.9.1 provides several quality estimates, including N50 and coverage. For the two samples collected in Florianópolis, these values were low, clearly due to insufficient coverage. Consequently, we decided to combine the reads from these two samples, and reassembled them as described above. We refer to this combined assembly as Flo F3 N, and to the other sample as Boc F3 N.

In order to enhance the accuracy of the Nanopore assemblies, we polished them using three rounds of Pilon 1.24 and the Illumina reads from the same subspecies (WALKER et al., 2014). The following Illumina reads were utilized to polish the Boc F3 N assembly: Boc F2, Boc M3, and Boc M2. For the Flo F3 N assembly, we employed Flo M2, Flo M1, Flo M3, and Flo F2 reads. We used BUSCO v4.1.4 (MANNI et al., 2021) and QUAST (GUREVICH et al., 2013) to assess the quality of these assemblies.

In addition to the 38 genomes that were sequenced in this study, we also utilized two reference genomes from the *Anopheles* Reference Genomes Project (GenBank assembly accession: GCA_943734635.1 and GCA_943735745.1 for *An. cruzii* collected in Maquiné (RS), and *Anopheles bellator* from Cananéia (SP), respectively).

Search for orthologs and gene annotation - We used BUSCO v4.1.4 to annotate single-copy orthologs in the 40 assemblies (see above), searching for 3285 genes conserved among Diptera from the OrthoDB database (odb10, downloaded from <https://BUSCO.ezlab.org/>). These single-copy orthologs were used in the phylogenomic analysis, as described by Waterhouse et al (2018), and modified according to Dias (2020). Briefly, the single-copy orthologs annotated by BUSCO v4.1.4 were processed with the scripts `fix_busco_CDS_frame.awk` (DIAS et al., 2020) and `BUSCO_cleaning_pipeline.awk` (DIAS et al., 2020) which corrects annotation errors that might bias the phylogenomic analyses. The script `translatorex_vLocal.pl` (ABASCAL; ZARDOYA; TELFORD, 2010) was used to align nucleotide sequences based on the protein sequences they encode.

Phylogenomic inferences - While BUSCO v4.1.4 attempts to annotate all 3285 genes represented in the odb10 database, inevitably some genes are missing in some assemblies, mostly due to assembly errors. These "missing data" can be handled more conservatively (e.g., by including in the analysis only genes that are present in at least 35 of the 40 samples) or less conservatively (e.g., by including only genes that are present in at least 20 of the 40 samples). All else being equal, a more balanced sample of genes (i.e., the conservative approach) would be better since it avoids several bias in the estimation of species trees (e.g., Smith, et al. (2018)). The trade-off is that the conservative approach results in a smaller number of genes, which may reduce the statistical power of the dataset. We used both approaches by performing phylogenomic analyses with two datasets: (i) a "less conservative" one, with 2051 orthologs present in at least 20 of the 40 samples and (ii) a "conservative" one, composed of 1467 orthologs that are present in at least 35 of the 40 samples. For both datasets, maximum-likelihood phylogenies were inferred for each gene using best-fit models chosen in IQ-Tree 1.6.1(NGUYEN et al., 2015). We then applied TreeShrink (MAI; MIRARAB, 2018) to the gene trees produced by IQ-Tree; this removes sequences deemed as outliers for having very long branches. These outlier sequences can result from annotation errors (e.g., a paralog was misidentified as the desired ortholog), or from real biological causes (e.g., a gene with very fast evolution); in both cases, they may bias the analyses, and were therefore removed. After removing outliers, we inferred species trees from the "conservative" and "less conservative"

datasets using two approaches: maximum-likelihood (ML) and multispecies coalescent (MSC). TreeShrink generates (i) "filtered" alignments (without possibly problematic sequences) and (ii) "pruned" gene trees (without the outliers). The "filtered" alignments were concatenated and used for the Maximum-Likelihood (ML) inference in IQ-TREE 1.6.1 (NGUYEN et al., 2015). The "pruned" gene trees were used for the Multispecies Coalescent (MSC) inference in ASTRAL (MIRARAB et al., 2014), with default settings.

In order to root the resulting trees, we included sequences from *Anopheles (Kerteszia) bellator* and *Anopheles (Kerteszia) homunculus*, which are recognized as different species from the same subgenus of *An. cruzii*. The command lines used to perform the phylogenomic inferences are detailed in the file *phylogenetic_inferences.txt* (Additional folder).

F_{ST} analysis - The *F_{ST}* of nucleotide sequences were calculated from the filtered alignments of orthologous genes using the R package *PopGenome* (PFEIFER et al., 2014), with the script *table_summary.r* (Additional folder). This population-genetic measure was used to quantify the differentiation between pairs of the following *An. cruzii* sibling groups: (i) Florianópolis + Guapirim Group + Maquiné (samples Flo F2, Flo M1, Flo M2, Flo M3, Gua F1, Gua F2, Flo F3 N, GCA_943734635.1); (ii) Itatiaia Group (samples Itt F1, Itt M1, Itt M2, Itt M3, Itt M4), (iii) Bocaina Group 1 (samples Boc F2, Boc F3 N Boc M2, Boc M3), and (iv) Bocaina Group 2 (samples Boc F1, Boc M1, Boc M4, Boc M5, Boc M6).

The genes were classified into chromosomal arms based on the location of their orthologs in the *An. cruzii* genome from the *Anopheles* Reference Genomes Project (GenBank assembly accession: GCA_943734635.1). This classification was achieved by a BlastN search using as queries the single copy genes annotated by BUSCO v4.1.4 of the Flo M1 sample, and as the database the chromosome-level assembly GCA_943734635.1. The *graphic_fst_R.txt* script (Additional folder) was used to produce the *F_{ST}* plots.

Results

The first genome drafts for some *Kerteszia* provided here reveal interesting findings for some species of this subgenus. Analyses indicate at least four distinct species in *An. cruzii*: (i) Florianópolis-SC, Guapimirim-RJ, Maquiné-RS; (ii) Bocaina Group 1; (iii) Bocaina Group 2 and Itatiaia – RJ; (IV) Santa Teresa – ES. The evidence for speciation is especially overwhelming in the Bocaina groups: although they live side by side, they have high genetic differentiation and heterozygote deficiency, indicating reproductive isolation. On the other hand, populations from Florianópolis and Guapimirim, 1000 km away, do not show clear genetic separation, suggesting a single species. Surprisingly, we also found a new species in Santa Teresa - ES, distinct not only genetically but also morphologically from the others. Not enough, we also suggest the possibility of a complex for another species of this subgenus: *An. bellator*. In the paragraphs below we will comment in more detail on these findings.

Assembly and quality of genomes - Here, we present the first drafts for several species of *Kerteszia*, such as different species of the *An. cruzii* s.l. and *An. homunculus*. A total of 38 genomes were sequenced, from eight different populations. 23 samples from *An. cruzii* s.l. complex, ten from *An. bellator* complex, and five from *An. homunculus* (Additional file 1).

According to GenomeScope, the Illumina genomes coverage varied from 32.2x to 141.4x, and according to Flye 2.9.1, the Nanopore genomes displayed coverages of ~63x for *Kerteszia* Boc F3 N, and ~25x for *Kerteszia* Flo F3 N (Table 2).

Using the GenomeScope analysis of the raw reads, the average genome size for

An. bellator Illumina sequencing reads was ~173 Mb (range: 164.7 - 176.6 Mb), for *An. homunculus* was ~163 Mb (range: 148 - 172.3 Mb), and for *An. cruzii* s.l. was ~170 Mb (range: 160 - 181 Mb). However, the genome sizes calculated from the genome assembly data were generally smaller than the estimates based on the raw reads: for *An. bellator* was ~155 Mb (range: 128.7 - 233.7 Mb), for *An. homunculus* was ~147 Mb (range: 132.8 - 163.8 Mb), and for *An. cruzii* s.l. was ~140 Mb (range: 107.1 - 166.1 Mb), probably due to collapsing of repetitive regions during assembly. Interestingly, for the long reads samples of *An. cruzii* s.l., the estimated genome size assembled by Flye 2.9.1 and estimated by QUAST (KOLMOGOROV et al., 2019) was similar to the Illumina GenomeScope analysis results: for *Kerteszia* Flo F3 N was ~170 Mb, and for *Kerteszia* Boc F3 N was ~183 Mb (Table 2). Also, GenomeScope reported a high heterozygosity (average: ~2%), which is somehow expected for field samples (Table 2).

Comparing the BUSCO results among 07 samples assembled with both Platanus and SPAdes, an increase of complete orthologs above 20% was achieved for most samples assembled with Platanus in relation to SPAdes (Additional file 2). For this reason, the Platanus assembler was used to assemble all 36 genomes sequenced with Illumina of this work. Once the genomes were assembled with Platanus, the quality of all assemblies was evaluated. The N50 values ranged from 2.5 to 87.7 kb, the largest contig of each genome varied from 55 kb to 612 kb, and BUSCO v4.1.4 showed values of complete orthologs above 74% in 31 of 36 samples (Table 3, Additional file 3). The five samples sequenced using Nextera XT (*Kerteszia* Flo F2, *Kerteszia* San F1, *An. homunculus* San F2, and *An. homunculus* San M2, *Kerteszia* Cam F33) due to the very low amount of DNA (< 85 ng), showed the lowest BUSCO v4.1.4 values for complete orthologs sequences (36% – 67%).

The two Nanopore samples were assembled with Canu and Flye. The largest contig for the two Nanopore samples was 1969.74 kb for *Kerteszia* Flo F3 N and 13817.62 kb for *Kerteszia* Boc F3 N. Furthermore, the N50 values were 446.87 kb for *Kerteszia* Flo F3 N and 1381.16 kb for *Kerteszia* Boc F3 N. The BUSCO v4.1.4 results for complete orthologs were 98.5% and 98.6% for *Kerteszia* Flo F3 N and *Kerteszia* Boc F3 N, respectively (see Table 3 and Additional file 3 for more details).

One of the Canu assemblies has a large amount of duplicated genes (22%), possibly due to the high heterozygosity of our samples. Flye assemblies were more contiguous (have higher N50) and complete (have higher Busco "complete single copy"; Table 4) and were used in all analysis.

Differentiation of Bocaina Group 1 and Bocaina Group 2 - Dias et al (2018) found in *An. cruzii* s.l. collected at Bocaina two distinct haplotypes of the *cpr* gene, with a strong heterozygote deficit, which suggested cryptic speciation. We further investigated this issue here by designing allele-specific PCR primers for the *cpr* gene, screening a larger sample of 145 wild-caught mosquitoes from Bocaina, and sequencing four representatives of Bocaina 1 and five Bocaina 2. In this section we report the *cpr* genotyping results; the sequencing results were dealt with in the next sections.

Since the *cpr* gene is X-linked in *An. cruzii* s.l. (the *Kerteszia* GCA_943735745.1 sample), we should take sex into account to calculate allele frequencies (Table 5). The frequency of the *cpr*¹ allele among females was 3.45%; using this frequency we calculated the expected genotypic frequencies under the Hardy-Weinberg equilibrium (Table 5, column 3). Similarly to Dias et al (2018), we found a strong deficit of heterozygotes, which is statistically significant ($P < 10^{-5}$, Fisher's Exact Test; ref Santos et al (SANTOS; LEMES; OTTO, 2020)). The most likely explanation for this heterozygote deficiency in sympatry is reproductive isolation (i.e., cryptic speciation), although formally it can also be caused by gross errors in genotyping or very strong selection against heterozygotes.

We could rule out these explanations and confirm the cryptic speciation hypothesis because we sequenced the genomes of four individuals carrying the *cpr*¹ allele and five carrying *cpr*² and, as detailed in the next sections, we found the same strong differentiation across the genome. Two genetically differentiated groups co-existing in sympatry can only be different species.

Table 5: Calculations of expected genotype frequencies under Hardy-Weinberg equilibrium.

Genotype	Freq.	Expected HW
<i>cpr</i> ¹ / Y	4	-
<i>cpr</i> ² / Y	54	-
<i>cpr</i> ¹ / <i>cpr</i> ¹	3	0.1
<i>cpr</i> ¹ / <i>cpr</i> ²	0	5.8
<i>cpr</i> ² / <i>cpr</i> ²	84	81.1

Genetic Differentiation (F_{ST}) - We estimated the F_{ST} values among all *An. cruzii* s.l. groups that have at least three individuals: (i) South/Southeast Brazil (Florianópolis + Guapimirim + Maquiné), (ii) Itatiaia, (iii) Bocaina Group 1, and (iv) Bocaina Group 2. The two Bocaina groups were classified based on the *cpr* genotyping reported in the previous section. In all pairwise comparisons, high F_{ST} values were observed (Figure 3, Figure 4, Additional file 4). The case of the two sympatric Bocaina groups is of particular interest. Note that very high F_{ST} values occur across all chromosomes, which shows that the *cpr* differentiation first observed by Dias et al (2018) indeed correspond to two biological species. The data rule out all alternative explanations such as selection against heterozygotes because they are not expected to have genome-wide effects.

The highest genetic differentiation was detected in the comparisons including Bocaina Group 1, which showed mean F_{ST} values around 0.6 in the autosomes ($F_{ST} > 0.35$, which according to Hey and Pinho (2012) it is the threshold value for species diagnosis). The lowest F_{ST} values obtained were between Bocaina Group 2 and Itatiaia, which is consistent with the close position of these groups in the phylogenomic inference estimated in this study (next section). However, even between Bocaina Group 2 and Itatiaia high F_{ST} values were observed: from 1988 genes, 436 had $F_{ST} > 0.35$. For all comparisons, the highest F_{ST} values were found on the X chromosome, suggesting a strong effect of this chromosome on the genetic differentiation process (Figure 3, Additional file 4). A similar result was obtained with the *An. gambiae* complex (NEAFSEY et al., 2015).

Phylogenomic Inferences – Species trees were inferred for the 40 samples of *An. cruzii* s.l. and related species using both maximum likelihood and multispecies coalescence methods with (i) 2,051 orthologues present in at least 20 of the 40 samples (Figure 5) and (ii) 1,467 orthologues present in at least 35 out of 40 samples (data not shown). The same topology and similar levels of support were obtained in both analysis. All phylogenomic analyses are consistent and strongly suggest that *An. cruzii* s.l. is composed of at least four cryptic species as follows: (i) South/Southeast Brazil, represented here by Florianópolis-SC, Guapimirim-RJ and Maquiné-RS (Serra do Mar coastal region); (ii) Bocaina Group 1 and (iii) Bocaina Group 2 (Serra do Mar continental region, in Southeast Brazil) and Itatiaia-RJ (Serra da Mantiqueira, Southeast Brazil), (iv)

Santa Teresa (Southeast Brazil). Despite occurring in the same location, Bocaina Group 1 and Bocaina Group 2 did not cluster together and showed a high phylogenetic distance, which are strong evidences of speciation.

Additionally, three *An. bellator* groups were also uncovered in this work: the first one seems to be widespread in southern/southeastern regions from Cananéia-SP (*An. bellator* GCA_943735745.1) and Ilha Grande-RJ (samples Ilh F1, Ilh M1). The second and the third groups are from the North-East Brazil: one from Itaparica-BA (Itp F1, Itp M1), and the other from Camacan-BA (Cam F11, Cam F14, Cam F25, Cam F33, Cam F34, Cam F36) (Figure 5, Additional file 5).

Morphology of the male genitalia: Male genitalia of all 16 sequenced samples were examined. Two of them were from *An. bellator*: one from Ilha Grande (RJ) and one from Itaparica (BA). Additionally, 14 specimens were initially classified as *An. cruzii*: six from Bocaina (SP) (two from Boc1, four from Boc2), four from Itatiaia (RJ), three from Florianópolis (SC), and one from Santa Teresa (ES). However, morphological difference was found only in the ventral claspette of the *Kerteszia* San M1, sample from Santa Teresa (ES). This structure in *Kerteszia* San M1 is distinct from the typical claspette of *An. cruzii* and resembles a golf club. As far as we could see, the male genitalia of all males from the other four groups are identical. No morphological differences were observed in the other structures (Additional file 6 and Additional file 7). Regarding the *An. bellator* samples, the male genitalia of individuals from Itaparica (BA) and Ilha Grande (RJ) seem to be identical. Unfortunately, we only have females from Camacan (BA), so it is not possible to analyze this diagnostic characteristic in these samples.

Discussion

Assembly and quality of genomes

Here we present the genome assemblies of *Kerteszia* mosquitoes, comprising 38 individuals from different species including *An. cruzii* s.l., the primary malaria vector in the Atlantic Forest region. Most of these genome sequences have not been published yet, which underscores the significance of our results, allowing us to apply a phylogenomic approach to better understand the evolutionary history of these anophelines.

In this study, 38 complete genomes were sequenced: 36 with Illumina (coverage: 32.2x - 141.4x) and two with Nanopore technology (coverage: ~ 25x and 63x). We employed Flye and Canu to assemble the long read sequences, and tested Platanus and SPades for the short reads. The Platanus and Flye assemblers generated genomes with better BUSCO results than SPAdes and Canu. Platanus recovering a higher number of genes with complete sequences (> 20%). And Flye recovering a higher number of genes with complete single copy sequences. For Boc F3 N, it was approximately 20% higher, while for Flo F3 N, the difference was more subtle, only around 0.5%. Given the improved quality indices (e.g., BUSCO) obtained with Flye assemblies, we have decided to utilize this assembler for all samples.

In the case of Illumina reads, this may have been caused by the heterozygosity rates of the field samples used in this study (1.3% - 3.0%), which is high when compared with laboratory insect colonies (heterozygosity < 1%) (NORRIS et al., 2001; KINGAN et al., 2019; GUAN et al., 2020), since the Platanus was designed to deal with highly heterozygous genomes (> 1%) (KAJITANI et al., 2014), while SPAdes was developed for small ones (e.g., bacterial haploid genomes) (BANKEVICH et al., 2012).

Our analysis revealed that the genome size of *An. cruzii* s.l. based on short reads is approximately 170 Mb, which is similar to *Anopheles (Nyshorhynchus) darlingi* (173.9 Mb (MARINOTTI et al., 2013); but smaller than *Anopheles (Cellia) gambiae* (278 Mb)

(GARCIA et al., 2022). It should be noted that genome sizes calculated from assemblies (*An. cruzii* s.l.: ~140 Mb) tend to be smaller than those based on raw reads (*An. cruzii* s.l.: ~170 Mb) due to the collapsing of repetitive regions during assembly. Nanopore sequencing produced genome sizes of 170 Mb and 183 Mb, in agreement with our Illumina assemblies based on raw reads and a previous long-read estimation of *An. cruzii* genome size (~184 Mb) reported by the *Anopheles* Reference Genomes project (accession: GCA_943734635.1). This provides additional support for the reliability of our presented genome size data.

Regarding the contiguity estimates, the Illumina genomes exhibited N50 values ranging from ~2.5 kb to ~87 kb. Among these, ten genomes showed an N50 value exceeding 37 kb, which is close to those obtained for *An. darlingi* (N50 = ~38 kb) and *An. gambiae* s.s. (N50 = ~42 kb) (LAWNICKZAK et al., 2010; MARINOTTI et al., 2013). The N50 values estimated by QUAST for the long-read genomes (*Kerteszia* Flo F3 N, and *Kerteszia* Boc F3 N) were found to be comparatively lower than those of the GCA_943734635.1 genome. Specifically, the N50 values for the former ranged from 447 to 1381 kb, while the latter boasted a much higher value of 62076 kb. It is noteworthy to mention that this disparity can be attributed to the fact that the GCA_943734635.1 genome underwent assembly using multiple technologies, including Genomics Chromium data, Arima Hi-C data, and PacBio HiFi sequencing. In addition to contiguity indices, completeness was also verified in the genomes sequenced in this study. From the 38 samples analysed, 33 showed high numbers of complete orthologs found using BUSCO (~75% - 99%), which is close to long reads assemblies of *Anopheles coluzzi*, *Anopheles arabiensis* (~99%) (ZAMYATIN et al., 2021), and *An. cruzii*, *An. bellator* (~99% and ~98%: GCA_943734635.1 and GCA_943735745.1, respectively), confirming the good quality sequences produced in this study.

***Anopheles (Kerteszia) cruzii* s.l.**

Our phylogenomic results corroborate previous studies that suggested cryptic speciation within *An. cruzii* s.l. (CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2004; RONA et al., 2009, 2010; RONA; CARVALHO-PINTO; PEIXOTO, 2010; RONA; CARVALHO-PINTO, 2013; DE REZENDE DIAS et al., 2018; KIRCHGATTER et al., 2020).

Complete single-copy genes identified with BUSCO were used for phylogenomic inferences with two approaches: (i) Multispecies Coalescent and (ii) concatenation. Although concatenation is a method widely used for phylogenomic analysis, it may produce incorrect trees when there is substantial heterogeneity between gene trees. In these cases, the multispecies coalescent approach achieves a more accurate species tree (LIU; WU; YU, 2015; DIAS et al., 2020; JIANG; EDWARDS; LIU, 2020; SHEN; STEENWYK; ROKAS, 2021). For this reason, these two approaches were used here to infer the *An. cruzii* s.l. species tree.

The analyses of the 40 genomes, which included *An. cruzii* s.l., *An. bellator*, and *An. homunculus* and used (i) the orthologs present in at least 20 of the 40 genomes or (ii) those present in at least 35 of the 40, showed similar topologies and suggest that *An. cruzii* s.l. is composed of at least four cryptic species, which occur in: I) South and Southeast Brazil, represented by Florianópolis (SC), Guapimirim (RJ), and Maquiné (RS); II) Itatiaia and Bocaina Group 2. In Bocaina, two sympatric groups are found, the other one, is III) Bocaina Group 1 and IV) Santa Teresa (Southeast Brazil). These four clusters and their relationships were recovered in both the concatenation and Multispecies Coalescent species trees. According to Dias et al. (2018), *An. cruzii* populations from the South/Southeast Brazil comprises at least two cryptic species: (i) Group 1, occurring in southern Brazil and in the coastal regions of the Serra do Mar mountain (Florianópolis-

SC, Guapimirim–RJ, Tinguá–RJ, and Sana–RJ); and (ii) Group 2, occurring in the continental part of the Serra do Mar mountain (Bocaina-SP) and the Serra da Mantiqueira mountain (Itatiaia – RJ). In Bocaina, which is halfway between Itatiaia and the coastal part of Serra do Mar, both groups occur in sympatry with a low frequency of heterozygotes, suggesting that there is reproductive isolation between them (DE REZENDE DIAS et al., 2018). Our results support these findings. Our results also show that Bocaina Group 2 is a sister taxon of the Itatiaia Group and that South Brazil individuals (Florianópolis–SC) clustered with those from the coastal part of Serra do Mar (Guapimirim–RJ). But different from Dias et al. (2018), the individuals classified here as Bocaina Group 1 are quite distant from those coming from South and Southeast coastal regions. However, this discrepancy is not unexpected since Dias et al. (2018) used only two gene fragments, making their results more susceptible to bias resulting from the heterogeneity of gene trees.

Among all pairwise comparisons for *An. cruzii* s.l. species, the lowest F_{ST} values obtained were between Bocaina Group 2 and Itatiaia (mean F_{ST} ~0.2), which is consistent with the close relationship recovered in the phylogenomic analyses for these groups. However, even among them, ~400 loci (from the ~2000 analysed) provided F_{ST} above 0.35, threshold that usually distinguishes different species (HEY; PINHO, 2012). Furthermore, the F_{ST} levels we observed among the other pairwise comparisons among *An. cruzii* s.l. siblings are well above the 0.35 threshold (HEY; PINHO, 2012), even between the sympatric Bocaina Group 1 and Bocaina Group 2, which means that geographic distance *per se* does not explain the levels of differentiation we observed among these anopheline groups. Hence, all available data support the conclusion that there are at least 4 cryptic species in *An. cruzii* s.l.

The highest F_{ST} values in all pairwise comparisons were found in loci from the *X* chromosome (based on *An. cruzii* s.l. GCA_943735745.1 genome from the *Anopheles* Reference Genomes Project), suggesting that this chromosome may play an important role in the process of genetic differentiation of the *An. cruzii* s.l. species. This same pattern was also observed for the *An. gambiae* s.l. complex, which has a region on the *X* chromosome associated with the speciation process (ABOAGYE-ANTWI et al., 2015; CAMPOS et al., 2021). Indeed, genomic studies found higher genetic divergence in genes from the *X* chromosome compared with those from the autosomes (LEE et al., 2013) since in species with X/Y sex determination (e.g. *Anopheles* spp.) genes from the *X* chromosome are expected to be under more efficient selection (faster-*X* evolution) than those from the autosomes, due to the full expression of rare recessive mutations in the males (VICOSO; CHARLESWORTH, 2006). Moreover, chromosomal inversions also contribute to the speciation process in dipterans (e.g. *Drosophila* spp.), through linkage disequilibrium among genes inside the inversions (NOOR et al., 2001). In fact, some complex of species have fixed chromosomal inversions on *X* chromosome that distinguish the siblings, and this is the case of *An. cruzii* s.l. (RAMIREZ; DESSEN, 2000b; TURNER; HAHN; NUZHIDIN, 2005).

***Anopheles (Kerteszia) laneanus*:**

Anopheles (Kerteszia) laneanus and *An. cruzii* s.l. are genetically quite close to each other. Studies using molecular markers have supported the hypothesis that *An. laneanus* belongs to the *An. cruzii* complex (FOSTER et al., 2017), or that they form a sister clade (LORENZ; PATANÉ; SUESDEK, 2015; VOGES et al., 2019). Also, besides *An. cruzii* and *An. laneanus* be distinguishable by morphological features on the tarsus, these morphological differences are subtle making species identification quite complex (FORATTINI, 1962; CONSOLI; LOURENÇO DE OLIVEIRA, 1994). Therefore, the morphological differentiation between them is more reliable using male genitalia

(FORATTINI, 1962). So, the possibility that one of the *An. cruzii* s.l. species could be *An. laneanus* was considered, and ultimately rejected after examining the male genitalia of *An. cruzii* s.l. specimens from the four siblings disclosed in this work, since no morphological differences were found indicating they all belong to the *An. cruzii* s.l. *An. cruzii* typically has lateral expansion ranging from rounded to sinuous at the lateral margin, in *An. cruzii* s.l. (SALLUM et al., 2020). In contrast, *An. laneanus* that is quite spiculated medially, with many short spicules extending laterally to the lateral margin (SALLUM et al., 2020). The morphological characteristics pertaining to the genitalia of all mosquitoes previously identified as *An. cruzii* were consistent with those morphological character of this specie (with exception of Santa Teresa, as will be commented below).

Collections were made in Campos do Jordão (SP) in search of *An. laneanus*. However, it was not possible to collect any individuals on this occasion. Nevertheless, even without the genome of this species, we were able to confirm that there were no individuals of this species in our samples by analyzing the male genitalia. But the sample from Santa Teresa (ES) (*Kerteszia* San M1) showed male genitalia morphological differences in the ventral claspette, which typically has lateral expansion ranging from rounded to sinuous at the lateral margin, but not curved posteriorly in *An. cruzii* s.l. (SALLUM et al., 2020). However, the *Kerteszia* San M1 ventral claspette is not rounded or sinuous, but instead resembles a golf club, which is very distant from the typical *An. cruzii* s.l. ventral claspette, as well as those of other *Kerteszia* species (e.g.: *An. laneanus*) (SALLUM et al., 2020). It should be noted that all other morphological characteristics are consistent with *An. cruzii* s.l. according to Forattini (1962), Consoli & Lounrenço-de-Oliveira (1994). It is noteworthy to observe that, based on our phylogenomic analysis *Kerteszia* San M1 and *Kerteszia* San F1 comprise an exclusive cluster within the *An. cruzii* s.l. group, thus indicating the likelihood of a novel *Kerteszia* species that remains unidentified until now.

Morphological differentiation can also be observed in other species complexes, such as *Lutzomyia longipalpis*, where males can exhibit one or two pale spots on the abdominal tergites, classified as 1S (with one spot on the third tergite) or 2S (with two spots, one on the third tergite and another on the fourth tergite), respectively. Some studies have noted that individuals from Ceará (Northeast Brazil) only have one spot, whilst those collected in Pará (North Brazil) have a pair of spots (SHERLOCK; SHERLOCK, 1961; MANGABEIRA, 1969). There is still debate regarding the taxonomic status of this species. Some authors argue that *Lu. longipalpis* is a single species, while others suggest that it represents a complex of sibling species (ARRIVILLAGA et al., 2003; BAUZER et al., 2007).

***Anopheles (Kerteszia) bellator* and *Anopheles (Kerteszia) homunculus*:**

During the early 1940s until the 1970s, *An. cruzii* s.l., *An. bellator*, and *An. homunculus*, which occur in synmpatry all along the Atlantic Forest, were the main "bromeliad malaria" vectors in this biome causing significant public health burden (RACHOU, 1958; DEANE, 1988; MARRELLI et al., 2007). But despite their epidemiological importance together with *An. cruzii* s.l., there are very few studies concerning *An. bellator* and *An. homunculus*. Some studies have suggested that *An. bellator* may be part of a species complex (CARVALHO-PINTO; LOURENÇO-DE-OLIVEIRA, 2003; VOGES et al., 2019). Voges et al.(2019) using molecular markers, disclosed two different *An. bellator* groups in the Brazilian Atlantic Forest: *An. bellator* A is widespread in Southern and Southeastern regions (Ilha do Mel-PR, Cananéia-SP and Ilha Grande-RJ), and *An. bellator* B is found in the north-east Brazil (Camacan-BA). Using isoenzymes, Carvalho-Pinto and Lourenço-de-Oliveira (2003), also suggested that

An. bellator populations from south and southeast Brazil (Florianópolis-SC, and Cananéia-SP) were genetically close to each other, and quite distant from the Northeast sample (Itaparica-BA). It is important to note that although these two studies have suggested that the Bahia populations (Camacan and Itaparica) are genetically distant from those found in the southern Brazilian regions, the BA population (Camacan) used by Voges et al. (VOGES et al., 2019) is 300 Km away (to the south direction) from Itaparica (BA) used by Carvalho-Pinto and Lourenço-de-Oliveira (2003). So, before our study there was no cue about if *An. bellator* from Camacan and Itaparica belong to the same group, and their relationship with southern populations. Accordingly, our results strongly suggest that *An. bellator* is a complex of at least three species – the first one is from southeast Brazil (Cananéia – SP + Ilha Grande – RJ) and probably belongs to *An. bellator* A proposed by Voges et al. (2019), and the others are from the Bahia State: one from Camacan, and the other from Itaparica, which despite the great geographical distance (> 1300 Km), is genetically closer to *An. bellator* A. So, as was found for *An. cruzii* s.l., geographical distance *per se* does not explain the levels of differentiation observed among *An. bellator* groups, since Itaparica and Camacan are only ~300 Km apart, and form two groups quite distant phylogenetically. However, in contrast with the high differentiation found for *An. cruzii* s.l. and *An. bellator* populations, *An. homunculus* samples used in this study showed low genetic distance and could be considered a single species all along the Atlantic Forest, from Florianópolis-SC (South of Brazil) to Camacan-BA (Northeast Brazil), which is consistent with Cardoso et al. (CARDOSO et al., 2012) findings using molecular markers. The small genetic differentiation found for *An. homunculus* in the phylogeny could be due to intraspecific variation, which typically occurs among individuals of the same species (ROCHES et al., 2018b, 2021b), as was also seen for *An. cruzii* s.l. from South and Southeast Brazil, represented here by Florianópolis (SC), Guapimirim (RJ), and Maquiné (RS). Indeed, co-specific populations that are geographically distant are expected to display some level of genetic differentiation due to genetic drift, local adaptation, etc (VIOLLE et al., 2012; ROCHES et al., 2018a). However, note that *An. cruzii* s.l. from Florianópolis (SC) and Guapimirim (RJ) or *An. homunculus* from Florianópolis (SC) and Camacan (BA) that are geographically very distant (~1,000 km and 1500 km apart, respectively) show only moderate levels of differentiation in the phylogeny, whereas *An. cruzii* s.l. from Bocaina Group 1 and Bocaina Group 2 (sympatric) form two groups quite distant phylogenetically.

Conclusions

In this work, ~2,000 genes were used to assess the genetic differentiation of *An. cruzii* s.l., applying F_{ST} analysis and phylogenomic inferences. Our results strongly suggest that *An. cruzii* s.l. is a complex composed of at least four cryptic species: i) South and Southeast Brazil (represented here by Florianópolis, Guapimirim and Maquiné), ii) Itatiaia and Bocaina Group 2, iii) Bocaina Group 1, iv) Santa Teresa (Southeast Brazil). The evidence of cryptic speciation is especially compelling in the case of the Bocaina samples, which came from the same location and yet form two groups that are quite distant phylogenetically. In addition to *An. cruzii* s.l., we also suggest that *An. bellator* is a complex of at least three species, but *An. homunculus* populations showed no clear genetic separation and may be considered a single species.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by CAPES, CNPq, Royal Society and Wellcome Trust.

Authors' contributions

KV, LDPR, CJCP, ANP and HRR participated in mosquitoes capturing. KV and CJCP carried the morphological identification. KV was responsible for experiments. KV, LDPR, BC, GRD, ED, and TV were responsible for assembled genomes, data generation and analysis. KV drafted the manuscript and LDPR, ABC and GRD helped in the paper drafting by critically reading the original manuscript. LDPR and ABC, participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

The authors are indebted to Natália Valério de Souza, Iara Carolini Pinheiro, André Akira Gonzaga Yoshikawa, Sabrina Fernandes Cardoso, João Victor Costa Guesser and Gabriela Pinto Guimarães for their logistic support during the field work, and to LAMEB - Federal University of Santa Catarina for the use of its microscopy facilities and for its technical assistance.

References

- ABASCAL, F.; ZARDOYA, R.; TELFORD, M. J. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, v. 38, n. SUPPL. 2, p. 7–13, 2010.
- ABOAGYE-ANTWI, F. et al. Experimental swap of *Anopheles gambiae*'s assortative mating preferences demonstrates key role of X-chromosome divergence island in incipient sympatric speciation. *PLOS Genetics*, v. 11, n. 4, p. e1005141, 16 abr. 2015. Disponível em: <<https://dx.plos.org/10.1371/journal.pgen.1005141>>.
- AHMED, S. et al. Travel is a key risk factor for malaria transmission in pre-elimination settings in Sub-Saharan Africa: A review of the literature and meta-analysis. *American Journal of Tropical Medicine and Hygiene*, v. 103, n. 4, p. 1380–1387, 2020.
- ANOPHELES GAMBIAE 1000 GENOMES CONSORTIUM. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. *Genome Research*, v. 30, n. 10, p. 1533–1546, 2020.
- ARAGÃO, M. B. Distribuição geográfica e abundância de espécies de *Anopheles* (*Kerteszia*) (Diptera, Culicidae). *Rev Bras Malariol Doenças Trop*, v. 16, n. 73–109, 1964.
- ARCOS, A. N. et al. Characterization of artificial larval habitats of *Anopheles darlingi* (Diptera: Culicidae) in the Brazilian Central Amazon. *Revista Brasileira de Entomologia*, v. 62, n. 4, p. 267–274, out. 2018. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S2221169115302483>>.
- ARRIVILLAGA, J. et al. The taxonomic status of genetically divergent populations of *Lutzomyia longipalpis* (Diptera: Psychodidae) based on the distribution of mitochondrial and isozyme variation. *Journal of Medical Entomology*, v. 40, n. 5, p.

615–627, 2003.

ARROW, K. J.; PANOSIAN, C.; GELBAND, H. *Saving lives, buying time: economics of malaria drugs in an age of resistance - A brief history of malaria*. [s.l: s.n.]

ASHLEY, E. A.; PYAE PHYO, A.; WOODROW, C. J. Malaria. *The Lancet*, v. 391, n. 10130, p. 1608–1621, 2018.

BANKEVICH, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, v. 19, n. 5, p. 455–477, 2012.

BARATA, R. *Malária e seu controle*. 1. ed. São Paulo: Editora Hucitec, 1988.

BARRÓN, M. G. et al. A new species in the major malaria vector complex sheds light on reticulated species evolution. *Scientific Reports*, v. 9, n. 1, p. 14753, 14 out. 2019. Disponível em: <<https://www.nature.com/articles/s41598-019-49065-5>>.

BAUZER, L. G. et al. *Lutzomyia longipalpis* in Brazil: a complex or a single species? A mini-review. *Memórias do Instituto Oswaldo Cruz*, v. 102, n. 1, p. 1–12, fev. 2007. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762007000100001&lng=en&tlng=en>.

BECKER, R. et al. *Maps: draw geographical maps. R package version 3.2.0.*, 2017. . Disponível em: <<https://cran.r-project.org/package=maps>>.

BENCHIMOL, J. L.; SÁ, R. M. *Adolpho Lutz – Obra Completa – v. 2 Febre Amarela, Malária e Protozoologia Yellow Fever, Malaria and Protozoology*. [s.l: s.n.]

BERENDS, C. J. et al. On the cause of the mid- pleistocene transition. *Reviews of Geophysics*, v. 59, n. 2, 16 jun. 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1029/2020RG000727>>.

BERTOLA, M. et al. Updated occurrence and bionomics of potential malaria vectors in Europe: a systematic review (2000–2021). *Parasites and Vectors*, v. 15, n. 1, p. 1–34, 2022. Disponível em: <<https://doi.org/10.1186/s13071-022-05204-y>>.

BICKFORD, D. et al. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, v. 22, n. 3, p. 148–155, mar. 2007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169534706003703>>.

BIVAND, R.; LEWIN-KOH, N. *Maptools: tools for reading and handling spatial objects. R package version 0.9-2.* 2017, 2017. . Disponível em: <<https://cran.r-project.org/package=maptools>>.

BOURKE, B. P. et al. High levels of diversity in *Anopheles* Subgenus *Kerteszia* revealed by species delimitation analyses. *Genes*, v. 14, n. 2, p. 1–14, 2023.

BRANQUINHO, M. S. et al. Infection of *Anopheles (Kerteszia) cruzii* by *Plasmodium vivax* and *Plasmodium vivax* variant VK247 in the municipalities of São Vicente and Jucitaba, São Paulo. *Revista panamericana de salud publica = Pan American journal of public health*, v. 2, n. 3, p. 189–93, 1997. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9445765>>.

BRASIL. *O controle de endemias no Brasil: de 1979 a 1984*. [s.l: s.n.].

BRASIL. *Manual para profissionais da saúde na atenção básica*. Brasília, DF Ministério da Saúde. Secretaria de vigilância em saúde., , 2005. .

BRASIL, P. et al. Outbreak of human malaria caused by *Plasmodium simium* in the

Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation. *The Lancet Global Health*, v. 5, n. 10, p. e1038–e1046, out. 2017. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2214109X17303339>>.

BRUNSDON, C.; CHEN, H. *GISTools: some further GIS capabilities for R.R package version 0.7-4*, , 2014. .

BUERY, J. C. et al. Ecological characterisation and infection of Anophelines (Diptera: Culicidae) of the Atlantic Forest in the southeast of Brazil over a 10 year period: has the behaviour of the autochthonous malaria vector changed? *Memórias do Instituto Oswaldo Cruz*, v. 113, n. 2, p. 111–118, 2 fev. 2018. Disponível em: <<https://malariajournal.biomedcentral.com/articles/10.1186/s12936-019-2744-8>>.

BUERY, J. C. et al. Atlantic forest malaria: A review of more than 20 years of epidemiological investigation. *Microorganisms*, v. 9, n. 1, p. 1–14, 2021.

CALADO, D. C.; NAVARRO-SILVA, M. a. Identificação de *Anopheles (Kerteszia) cruzii* Dyar & Knab e *Anopheles (Kerteszia) homunculus* Komp (Diptera, Culicidae, Anophelinae) através de marcadores moleculares (RAPD e RFLP). *Revista Brasileira de Zoologia*, v. 22, n. 4, p. 1127–1133, 2005.

CAMARGO, E. P. Malária, maleita, paludismo. *Ciênc. cult. (São Paulo)*, v. 55, n. 1, p. 26–30, 2003.

CAMPOS, M. et al. Unravelling population structure heterogeneity within the genome of the malaria vector *Anopheles gambiae*. *BMC Genomics*, v. 22, n. 1, p. 422, 8 dez. 2021. Disponível em: <<https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-07722-y>>.

CARDOSO, J. D. C. et al. New records of *Anopheles homunculus* in central and serra do mar biodiversity corridors of the atlantic forest, Brazil. *Journal of the American Mosquito Control Association*, v. 28, n. 1, p. 1–5, mar. 2012. Disponível em: <<http://www.bioone.org/doi/abs/10.2987/11-6205.1>>.

CARLOS, B. C. et al. A comprehensive analysis of malaria transmission in Brazil. *Pathogens and Global Health*, v. 113, n. 1, p. 1–13, 2 jan. 2019. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/20477724.2019.1581463>>.

CARNAVAL, A. C. et al. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, v. 323, n. 5915, p. 785–789, 2009.

CARVALHO-PINTO, C. J.; LOURENÇO-DE-OLIVEIRA, R. Isoenzymatic Analysis of Four *Anopheles (Kerteszia) bellator* Dyar & Knab (Diptera: Culicidae) Populations. *Memórias do Instituto Oswaldo Cruz*, v. 98, n. 8, p. 1045–1048, 2003.

CARVALHO-PINTO, C. J. de; LOURENÇO-DE-OLIVEIRA, R. Isoenzymatic analysis of four *Anopheles (Kerteszia) cruzii* (Diptera: Culicidae) populations of Brazil. *Memórias do Instituto Oswaldo Cruz*, v. 99, n. 5, p. 471–475, 2004.

CHAHAD- EHLERS, S.; LOZOVEI, A. L.; MARQUES, M. D. Reproductive and Post- Embryonic Daily Rhythm Patterns of the Malaria Vector *Anopheles (Kerteszia) cruzii*: Aspects of the Life Cycle. *Chronobiology International*, v. 24, n. 2, p. 289–304, 7 jan. 2007. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/07420520701282174>>.

CLAYTON, A. M.; DONG, Y.; DIMOPOULOS, G. The *Anopheles* innate immune system in the defense against malaria infection. *Journal of Innate Immunity*, v. 6, n. 2, p. 169–181, 2014.

- COETZEE, M. et al. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa*, v. 3619, n. 3, p. 246–274, 2013.
- COLLUCCI, E.; SALLUM, M. Phylogenetic analysis of the subgenus *Kerteszia* of *Anopheles* (Diptera: Culicidae: Anophelinae) based on morphological characters. *Insect Syst.*, v. 34, p. 361–372, 2003.
- COLUZZI, M. et al. Behavioural divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. *Nature*, v. 266, p. 832–833, 1977.
- COLUZZI, M. et al. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Royal society of tropical medicine and hygiene*, v. 73, 1979.
- COLUZZI, M. et al. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science*, v. 298, n. 5597, p. 1415–1418, 2002.
- CONN, J. E. et al. Emergence of a new neotropical malaria vector facilitated by human migration and changes in land use. *American Journal of Tropical Medicine and Hygiene*, v. 66, n. 1, p. 18–22, 2002.
- CONSOLI, R. A. G. B.; LOURENÇO DE OLIVEIRA, R. *Principais mosquitos de importância médica no Brasil*. [s.l.: s.n.]v. Único
- CORRÊA, R. R.; CERQUEIRA, F. M. C. Descrição de *Anopheles* (*Kerteszia*) *laneanus*, nova espécie de anofelino de Campos do Jordão (Diptera, Culicidae). *Arquivos de higiene e saúde pública*, v. 9, p. 111–117, 1944.
- COX, F. E. History of the discovery of the malaria parasites and their vectors. *Parasites and Vectors*, v. 3, n. 1, p. 1–9, 2010.
- DABIRE, K. R. et al. Assortative mating in mixed swarms of the mosquito *Anopheles gambiae* s.s. M and S molecular forms, in Burkina Faso, West Africa. *Medical and Veterinary Entomology*, v. 27, n. 3, p. 298–312, set. 2013. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2915.2012.01049.x>>.
- DANTAS-TORRES, F. Species concepts: What about ticks? *Trends in Parasitology*, v. 34, n. 12, p. 1017–1026, dez. 2018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1471492218302137>>.
- DAVIS, N. C. A Note on the Malaria-carrying Anophelines in Belém, Para, and in Natal, Rio Grande do Norte, Brazil. *Rivista di Malariologia*, v. 10, p. 43–51, 1931.
- DE LEMOS, A. B. et al. Reemergence of human malaria in Atlantic Forest of Rio Grande do Sul, Brazil. *Memórias do Instituto Oswaldo Cruz*, v. 116, 2021. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762021000100403&tlng=en>.
- DE PINA-COSTA, A. et al. Malaria in Brazil: What happens outside the Amazonian endemic region. *Memorias do Instituto Oswaldo Cruz*, v. 109, n. 5, p. 618–633, 2014.
- DE REZENDE DIAS, G. et al. Cryptic diversity in an Atlantic Forest malaria vector from the mountains of South-East Brazil. *Parasites and Vectors*, v. 11, n. 1, p. 1–11, 2018.
- DEANE, L. *Malaria vectors in Brazil* Mem. Inst. Oswaldo Cruz, 1986. .
- DEANE, L. Malaria Studies and Control in Brazil. *Am J Trop Med Hyg*, v. 38, n. 2, p.

223–230, 1988.

DEANE, L. Os grandes marcos na história do controle da malária. *Rev Bras Med Trop*, v. 25, p. 12–22, 1992.

DEANE, L. M. et al. *Anopheles (Kerteszia) cruzii*, a natural vector of the monkey malaria parasites, *Plasmodium simium* and *Plasmodium brasilianum*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, v. 64, n. 4, p. 647, 1970.

DEANE, L. M. et al. On the transmission of simian malaria in Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo*, v. 13, n. 5, p. 311–319, 1971.

DEANE, L. M.; DEANE, M. P.; FERREIRA NETO, J. Studies on transmission of simian malaria and on a natural infection of man with *Plasmodium simium* in Brazil. *Bulletin of the World Health Organization*, v. 35, n. 5, p. 805–808, 1966.

DELLA TORRE, A. et al. Speciation within *Anopheles gambiae* - The glass is half full. *Science*, v. 298, n. 5591, p. 115–117, 2002.

DIAS, G. R. et al. A phylogenomic study of Steganinae fruit flies (Diptera: Drosophilidae): strong gene tree heterogeneity and evidence for monophyly. *BMC Evolutionary Biology*, v. 20, n. 1, p. 1–12, 2020. Disponível em: <<https://doi.org/10.1186/s12862-020-01703-7>>.

DORN, P. L. et al. *Genetics of major insect vectors*. [s.l: s.n.]

DOS SANTOS, M. M. M. et al. Morphological identification of species of the *Nuneztovari* Complex of *Anopheles* (Diptera: Culicidae) from an area affected by a Brazilian hydroelectric plant. *Zootaxa*, v. 4565, n. 2, p. 235–244, 2019.

DOWNS, W. G.; PITTENDRIGH, C. S. Bromeliad malaria in Trinidad, British West Indies. *The American journal of tropical medicine and hygiene*, v. 26, p. 47–66, jan. 1946. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21018584>>.

DUARTE, A. M. R. C. et al. Natural infection in anopheline species and its implications for autochthonous malaria in the Atlantic forest in Brazil. *Parasites and Vectors*, v. 6, n. 1, p. 1–6, 2013.

DYAR, H. G.; KNAB, F. Forthcoming monograph. *Proceedings of the United States National Museum*, v. 35, n. 1, p. 53–70, 1908.

EHRMAN, L.; POWELL, J. R. *Drosophila willistoni* species group. *The Genetics and Biology of Drosophila*, v. 3b, p. London, 1982.

FERREIRA, E. et al. Longevity of *Anopheles (Kerteszia) cruzii* and *Anopheles (Kerteszia) bellator* under natural conditions. *Rev Bras Malariol Doencas Trop*, v. 4, p. 823–8, 1969.

FILATOV, D. A. PROSEQ: A software for preparation and evolutionary analysis of DNA sequence data sets. *Molecular Ecology Notes*, v. 2, n. 4, p. 621–624, 2002.

FORATTINI, O.P., ISHIATA, G.K., RABELLO, E.X. & COTRIM, M. . Observações sobre os mosquitos *Culex* da cidade de São Paulo, Brasil. *Rev. Saúde Pública*, v. 7, p. 315–30, 1973.

FORATTINI, O. P. *Entomologia Médica*. Volume 1 ed. São Paulo: Faculdade de Higiene e Saúde Pública, 1962.

FORATTINI, O. P. *Culicidologia médica*. 2. ed. São Paulo: Ed. Universidade de São Paulo, 2002.

- FORATTINI, O. P.; LOPES, O. S.; RABELLO, E. X. Investigações sobre o comportamento de formas adultas de mosquitos silvestres no Estado de São Paulo, Brasil. *Revista de saúde pública*, v. 2, n. 2, p. 111–173, 1968.
- FOSTER, P. G. et al. Phylogeny of anophelinae using mitochondrial protein coding genes. *Royal Society Open Science*, v. 4, n. 11, 2017.
- GADELHA, P. From “forest malaria” to “bromeliad malaria”: a case-study of scientific controversy and malaria control. *Parassitologia*, v. 36, p. 175–95, 1994.
- GARCIA, K. K. S. et al. Malaria time series in the extra-Amazon region of Brazil: epidemiological scenario and a two-year prediction model. *Malaria Journal*, v. 21, n. 1, p. 1–11, 2022. Disponível em: <<https://doi.org/10.1186/s12936-022-04162-1>>.
- GOUDET, J. Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, v. 5, n. 1, p. 184–186, mar. 2005. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1471-8286.2004.00828.x>>.
- GRIFFING, S. M. et al. A historical perspective on malaria control in Brazil. *Memorias do Instituto Oswaldo Cruz*, v. 110, n. 6, p. 701–718, 2015.
- GUAN, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, v. 36, n. 9, p. 2896–2898, 1 maio 2020. Disponível em: <<https://academic.oup.com/bioinformatics/article/36/9/2896/5714742>>.
- GUIMARÃES, A. É. et al. Ecology of Mosquitoes (Diptera: Culicidae) in Areas of Serra do Mar State Park, State of São Paulo, Brazil. II -Habitat Distribution. *Memorias do Instituto Oswaldo Cruz*, v. 95, n. 1–2, p. 17–28, 2000.
- GUIMARÃES, L. O. et al. The genetic diversity of *Plasmodium malariae* and *Plasmodium brasilianum* from human, simian and mosquito hosts in Brazil. *Acta Tropica*, v. 124, n. 1, p. 27–32, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.actatropica.2012.05.016>>.
- GUREVICH, A. et al. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, v. 29, n. 8, p. 1072–1075, 2013.
- HACKETT, L. W. The present status of our knowledge of the sub-species of *Anopheles maculipennis*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, p. 109–128, 1934.
- HACKETT, L. W. *Malaria in Europe. An Ecological Study*. London: Oxford University Press, 1937.
- HAMILTON, J. G. C. et al. Comparison of the sex-pheromone components of *Lutzomyia longipalpis* (Diptera: Psychodidae) from areas of visceral and atypical cutaneous leishmaniasis in Honduras and Costa Rica. *Annals of Tropical Medicine & Parasitology*, v. 90, n. 5, p. 533–541, 15 jan. 1996. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00034983.1996.11813079>>.
- HANG ET AL. Zoonotic malaria : Non- laverania *Plasmodium* Biology and Invasion Mechanisms. *Pathogens*, p. 1–22, 2021.
- HARBACH, R. E. *Mosquito taxonomic inventory*. Disponível em: <<https://mosquito-taxonomic-inventory.myspecies.info/simpletaxonomy/term/6047/>>. Acesso em: 28 maio. 2023.
- HARBACH, R. E. The Phylogeny and Classification of *Anopheles*. *Anopheles mosquitoes - New insights into malaria vectors*, p. 3–55, 2013.

- HARBACH, R. E. An Anopheles by any other name? *Journal of Medical Entomology*, v. 55, n. 5, p. 1069–1070, 2018.
- HAY, S. I. et al. The global distribution and population at risk of malaria: Past, present, and future. *Lancet Infectious Diseases*, v. 4, n. 6, p. 327–336, 2004.
- HAYWARD, B. . et al. *The Last Global Extinction (Mid-Pleistocene) of Deep-Sea Benthic Foraminifera (Chrysalogoniidae, Ellipsoidinidae, Glandulonodosariidae, Plectofrondiculariidae, Pleurostomellidae, Stilostomellidae), their Late Cretaceous–Cenozoic History and Taxonomy*. [s.l.] Cushman Foundation for Foraminiferal Research Special Publication, 2012.
- HEY, J.; PINHO, C. Population genetics and objectivity in species diagnosis. *Evolution*, v. 66, n. 5, p. 1413–1429, 2012.
- HOLT, R. A. et al. The genome sequence of the malaria mosquito *Anopheles gambiae* . *Science*, v. 298, n. 5591, p. 129–149, 4 out. 2002. Disponível em: <<https://www.science.org/doi/10.1126/science.1076181>>.
- ICZN. *International Code of Zoological Nomenclature*. London: The International Trust for Zoological Nomenclature, 1999.
- JIANG, X.; EDWARDS, S. V; LIU, L. The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. *Systematic Biology*, v. 69, n. 4, p. 795–812, 1 jul. 2020. Disponível em: <<https://academic.oup.com/sysbio/article/69/4/795/5721375>>.
- KAJITANI, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, v. 24, n. 8, p. 1384–1395, 2014.
- KATO, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, v. 30, n. 4, p. 772–780, 2013.
- KINGAN, S. et al. A high-quality de novo genome assembly from a single mosquito using PacBio sequencing. *Genes*, v. 10, n. 1, p. 62, 18 jan. 2019. Disponível em: <<http://www.mdpi.com/2073-4425/10/1/62>>.
- KIRCHGATTER, K. et al. Phylogeny of Anopheles (Kerteszia) (Diptera: Culicidae) Using Mitochondrial Genes. *Insects*, v. 11, n. 5, p. 324, maio 2020.
- KNAB, F. *Reported discussion*, 1912. .
- KOLMOGOROV, M. et al. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, v. 37, n. 5, p. 540–546, 1 maio 2019. Disponível em: <<https://www.nature.com/articles/s41587-019-0072-8>>.
- KOREN, S. et al. Canu: Scalable and accurate long-read assembly via adaptive κ -mer weighting and repeat separation. *Genome Research*, v. 27, n. 5, p. 722–736, 2017.
- KORSHUNOVA, T. et al. Multilevel fine-scale diversity challenges the ‘cryptic species’ concept. *Scientific Reports*, v. 9, n. 1, p. 1–23, 2019.
- KRZYWINSKI, J.; BESANSKY, N. J. Molecular Systematics of *Anopheles* : From Subgenera to Subpopulations. *Annual Review of Entomology*, v. 48, p. 111–139, 2003.
- LAETSCH, D. R.; BLAXTER, M. L. BlobTools: Interrogation of genome assemblies. *F1000Research*, v. 6, p. 1287, 2017.
- LAMPO, M. et al. Occurrence of sibling species of *Lutzomyia longipalpis* (Diptera:

- Psychodidae) in Venezuela: first evidence from reproductively isolated sympatric populations. *The American Journal of Tropical Medicine and Hygiene*, v. 61, n. 6, p. 1004–1009, 1 dez. 1999. Disponível em: <<http://www.ajtmh.org/content/journals/10.4269/ajtmh.1999.61.1004>>.
- LAPORTA, G. Z. et al. *Plasmodium falciparum* in the southeastern Atlantic forest: A challenge to the bromeliad-malaria paradigm? *Malaria Journal*, v. 14, n. 1, p. 1–12, 2015.
- LAURENS, M. B. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Human Vaccines and Immunotherapeutics*, v. 16, n. 3, p. 480–489, 2019. Disponível em: <<https://doi.org/10.1080/21645515.2019.1669415>>.
- LAWNICZAK, M. K. N. et al. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, v. 330, n. 6003, p. 512–514, 22 out. 2010. Disponível em: <<https://www.science.org/doi/10.1126/science.1195755>>.
- LEACHÉ, A. D. et al. The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, v. 63, n. 1, p. 17–30, 1 jan. 2014. Disponível em: <<https://academic.oup.com/sysbio/article/63/1/17/1687129>>.
- LEE, Y. et al. Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector *Anopheles gambiae* sensu stricto. *Malaria Journal*, v. 8, n. 1, p. 1–15, 2009.
- LEE, Y. et al. Chromosome inversions, genomic differentiation and speciation in the African malaria mosquito *Anopheles gambiae*. *PLoS ONE*, v. 8, n. 3, p. e57887, 20 mar. 2013a. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0057887>>.
- LEE, Y. et al. Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, *Anopheles gambiae*. *Proceedings of the National Academy of Sciences*, v. 110, n. 49, p. 19854–19859, 3 dez. 2013b. Disponível em: <<https://pnas.org/doi/full/10.1073/pnas.1316851110>>.
- LEVASHINA, E. A. Immune responses in *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology*, v. 34, n. 7, p. 673–678, 2004.
- LI, W. *Molecular evolution*. [s.l.] Sunderland: Sinauer Associates, 1997.
- LINTON, Y. M. et al. Morphological and molecular characterization of *Anopheles (Anopheles) maculipennis* Meigen, type species of the genus and nominotypical member of the *Maculipennis* Complex. *Systematic Entomology*, v. 28, n. 1, p. 39–56, 2003.
- LIU, L.; WU, S.; YU, L. Coalescent methods for estimating species trees from phylogenomic data. *Journal of Systematics and Evolution*, v. 53, n. 5, p. 380–390, 2015.
- LOAIZA, J. R. et al. Review of genetic diversity in malaria vectors (Culicidae: Anophelinae). *Infection, Genetics and Evolution*, v. 12, n. 1, p. 1–12, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.meegid.2011.08.004>>.
- LORENZ, C.; PATANÉ, J. S. L.; SUESDEK, L. Morphogenetic characterisation, date of divergence, and evolutionary relationships of malaria vectors *Anopheles cruzii* and *Anopheles homunculus*. *Infection, Genetics and Evolution*, v. 35, p. 144–152, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.meegid.2015.08.011>>.
- LUTZ, A. Mosquitos da floresta e malária silvestre *. *Archiv*, p. 759–768, 1903.

- MAI, U.; MIRARAB, S. TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, v. 19, n. Suppl 5, 2018.
- MALAFRONTI, R. D. S. et al. Intraspecific variation of second internal transcribed spacer of nuclear ribosomal DNA among populations of *Anopheles (Kerteszia) cruzii* (Diptera: Culicidae). *Journal of Medical Entomology*, v. 44, n. 3, p. 538–542, 1 maio 2007. Disponível em: <<https://academic.oup.com/jme/article-lookup/doi/10.1093/jmedent/44.3.538>>.
- MANGABEIRA, O. Sobre a sistemática e biologia dos flebótomos do Ceará. *Revista Brasileira de Malariologia e Doenças Tropicais*, v. 21, p. 3–26, 1969.
- MANNI, M. et al. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, v. 38, n. 10, p. 4647–4654, 2021.
- MARÇAIS, G.; KINGSFORD, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, v. 27, n. 6, p. 764–770, 2011.
- MARINOTTI, O. et al. The Genome of *Anopheles darlingi*, the main neotropical malaria vector. *Nucleic Acids Research*, v. 41, n. 15, p. 7387–7400, ago. 2013. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt484>>.
- MARRELLI, M. T. et al. *Kerteszia* subgenus of *Anopheles* associated with the Brazilian Atlantic rainforest: Current knowledge and future challenges. *Malaria Journal*, v. 6, p. 1–8, 2007.
- MAYR, E. *Populations, species, and evolution: An abridgment of animal species and evolution*. [s.l.] Belknap Press of Harvard University Press, 1970.
- MIGUEL, R. B. et al. Malaria in the state of Rio de Janeiro, Brazil, an Atlantic Forest area: an assessment using the health surveillance service. *Memórias do Instituto Oswaldo Cruz*, v. 109, n. 5, p. 634–640, 13 ago. 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762014000500634&lng=en&tlng=en>.
- MILNER, D. A. Malaria pathogenesis. *Cold Spring Harbor Perspectives in Medicine*, v. 8, n. 1, p. 1–11, 2018.
- MIRARAB, S. et al. ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics*, v. 30, n. 17, p. 541–548, 2014.
- MORIYAMA, E. N.; GOJOBORI, T. Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. *Genetics*, v. 130, n. 4, p. 855–864, 1 abr. 1992. Disponível em: <<https://academic.oup.com/genetics/article/130/4/855/6007208>>.
- MS. *Malária em 2021: Buscando o caminho para a eliminação da malária no Brasil*. Disponível em: <<https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2022/boletim-epidemiologico-vol-53-no17.pdf>>. Acesso em: 28 maio. 2023.
- MUKHOPADHYAY, J. et al. Genetic polymorphism of morphological and biochemical characters in a Natal, Brazil, population of *Lutzomyia longipalpis* (Diptera: Psychodidae). *Journal of the American Mosquito Control Association*, v. 14, n. 3, p. 277–82, set. 1998. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9813825>>.
- NEAFSEY, D. E. et al. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. *Science*, v. 347, n. 6217, jan. 2015.

- NEVES, A. et al. Malaria outside the Amazon region: Natural *Plasmodium* infection in anophelines collected near an indigenous village in the Vale do Rio Branco, Itanhaém, SP, Brazil. *Acta Tropica*, v. 125, n. 1, p. 102–106, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.actatropica.2012.08.014>>.
- NEVES, D. P. *Parasitologia humana*. São Paulo: Atheneu, 2016.
- NGUYEN, L. T. et al. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, v. 32, n. 1, p. 268–274, 2015.
- NOOR, M. A. F. et al. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, v. 98, n. 21, p. 12084–12088, 2001.
- NORRIS, D. E. et al. Microsatellite DNA Polymorphism and Heterozygosity Among Field and Laboratory Populations of *Anopheles gambiae* s.s. (Diptera: Culicidae). *Journal of Medical Entomology*, v. 38, n. 2, p. 336–340, 1 mar. 2001. Disponível em: <<https://academic.oup.com/jme/article-lookup/doi/10.1603/0022-2585-38.2.336>>.
- ORBARD, D. J. et al. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Molecular Biology and Evolution*, v. 29, n. 11, p. 3459–3473, nov. 2012. Disponível em: <<https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss150>>.
- OLIVEIRA-FERREIRA, J. et al. Malaria in Brazil: an overview. *Malaria Journal*, v. 9, n. 1, p. 115, 30 dez. 2010. Disponível em: <<https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-9-115>>.
- OLIVEIRA, T. M. P. et al. Mitochondrial Genomes of Anopheles (Kerteszia) (Diptera: Culicidae) From the Atlantic Forest, Brazil. *Journal of Medical Entomology*, v. 53, n. 4, p. 790–797, jul. 2016. Disponível em: <<https://academic.oup.com/jme/article-lookup/doi/10.1093/jme/tjw001>>.
- PAMILO, P.; NEI, M. Relationship between gene trees and species trees. *Molecular Biology and Evolution*, v. 5, n. 5, p. 568–583, 1988.
- PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, v. 20, n. 2, p. 289–290, 2004.
- PFEIFER, B. et al. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology and Evolution*, v. 31, n. 7, p. 1929–1936, jul. 2014. Disponível em: <<https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu136>>.
- R CORE TEAM. *R: A language and environment for statistical computing* Vienna, Austria R Foundation for Statistical Computing, , 2014. .
- R CORE TEAM. *A language and environment for statistical computing*. Vienna, Austria R Foundation for Statistical Computing, , 2021. .
- RACHOU, R. Anofelinos do Brasil: Comportamento das espécies vetoras de malária. *Rev Bras Malariol Doenças Trop*, v. 10, p. 145–181, 1958.
- RAMIREZ, C. C. L.; DESSEN, E. M. B. Chromosomal evidence for sibling species of the malaria vector *Anopheles cruzii*. *Genome*, v. 43, n. 1, p. 143–151, 2000a.
- RAMÍREZ, C. C. L.; DESSEN, E. M. B. *The polytene chromosomes of the mosquito Anopheles bellator compared with those of Anopheles cruzii* Brazilian Journal of

Genetics, 1996. .

RAMIREZ, C.; DESSEN, E. Chromosome differentiated populations of *Anopheles cruzii*: evidence for a third sibling species. *Genetica*, v. 1, n. 108, p. 73–80, 2000b.

REIS, M. d.; YANG, Z. Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. *Molecular Biology and Evolution*, v. 28, n. 7, p. 2161–2172, 1 jul. 2011. Disponível em: <<https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msr045>>.

ROCHES, S. Des et al. The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, v. 2, n. January, 2018a. Disponível em: <<http://dx.doi.org/10.1038/s41559-017-0402-5>>.

ROCHES, S. Des et al. The ecological importance of intraspecific variation. *Nature Ecology & Evolution*, v. 2, n. January, 2018b.

ROCHES, S. Des et al. Conserving intraspecific variation for nature's contributions to people. *Nature Ecology & Evolution*, v. 5, n. May, 2021a. Disponível em: <<http://dx.doi.org/10.1038/s41559-021-01403-5>>.

ROCHES, S. Des et al. Conserving intraspecific variation for nature's contributions to people. *Nature Ecology & Evolution*, v. 5, n. May, 2021b.

RODRIGUES, P. T. et al. Human migration and the spread of malaria parasites to the New World. *Scientific Reports*, v. 8, n. 1, p. 1–13, 2018.

RONA, L. D. et al. Assessing the molecular divergence between *Anopheles (Kerteszia) cruzii* populations from Brazil using the timeless gene: Further evidence of a species complex. *Malaria Journal*, v. 8, n. 1, p. 1–10, 2009.

RONA, L. D.; CARVALHO-PINTO, C. J. Evidence for the occurrence of two sympatric sibling species within the *Anopheles (Kerteszia) cruzii* complex in southeast Brazil and the detection of asymmetric introgression between them using a multilocus analysis. *BMC Evolutionary Biology*, v. 13, n. 1, p. 1, 2013. Disponível em: <BMC Evolutionary Biology>.

RONA, L. D.; CARVALHO-PINTO, C. J.; PEIXOTO, A. A. Molecular evidence for the occurrence of a new sibling species within the *Anopheles (Kerteszia) cruzii* complex in south-east Brazil. *Malaria Journal*, v. 9, n. 1, p. 1–9, 2010.

RONA, L. D. P. et al. Estimation of divergence time between two sibling species of the *Anopheles (Kerteszia) cruzii* complex using a multilocus approach. *BMC Evolutionary Biology*, v. 10, n. 1, 2010.

ROZAS, J. et al. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, v. 19, n. 18, p. 2496–2497, 2003.

RUIZ-LOPEZ, F. et al. Systematics of the oswaldoi complex *Anopheles, Nyssorhynchus* in South America. *Parasites and Vectors*, v. 6, n. 1, p. 1–13, 2013.

RUSSO, C. A. M.; TAKEZAKI, N.; NEI, M. Molecular phylogeny and divergence times of drosophilid species. *Molecular Biology and Evolution*, v. 12, n. 3, p. 391–404, 1995.

SALLUM, M. A. M. et al. Identification key to the *Anopheles* mosquitoes of South America (Diptera: Culicidae). III. Male genitalia. *Parasites and Vectors*, v. 13, n. 1, p. 1–23, 2020. Disponível em: <<https://doi.org/10.1186/s13071-020-04300-1>>.

- SANTOS, D. et al. A non-destructive enzymatic method to extract DNA from arthropod specimens: Implications for morphological and molecular studies. *PLoS ONE*, v. 13, n. 2, p. 1–11, 2018.
- SANTOS, F. A. B.; LEMES, R. B.; OTTO, P. A. HW_TEST, a program for comprehensive HARDY-WEINBERG equilibrium testing. *Genetics and Molecular Biology*, v. 43, n. 2, 2020a. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572020000400604&tlng=en>.
- SANTOS, F. A. B.; LEMES, R. B.; OTTO, P. A. HW_TEST, a program for comprehensive HARDY-WEINBERG equilibrium testing. *Genetics and Molecular Biology*, v. 43, n. 2, 2020b.
- SCHLIEP, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics*, v. 27, n. 4, p. 592–593, 2011.
- SHEN, X.-X.; STEENWYK, J. L.; ROKAS, A. Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data. *Systematic Biology*, v. 70, n. 5, p. 997–1014, 11 ago. 2021. Disponível em: <<https://academic.oup.com/sysbio/article/70/5/997/6146422>>.
- SHERLOCK, I. A.; SHERLOCK, V. A. On the experimental infection of *Phlebotomus longipalpis* by *Leishmania donovani*. *Revista brasileira de biologia*, v. 22, p. 409–18, dez. 1961. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14039201>>.
- SILVA-DO-NASCIMENTO, T. F. et al. Molecular divergence in the timeless and cpr genes among three sympatric cryptic species of the *Anopheles triannulatus* complex. *Memorias do Instituto Oswaldo Cruz*, v. 106, n. SUPPL. 1, p. 218–222, 2011.
- SIMÃO, F. A. et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, v. 31, n. 19, p. 3210–3212, 2015.
- SINAN. *Doenças e Agravos de Notificação - 2007 em diante - Malária*. Disponível em: <<https://datasus.saude.gov.br/aceso-a-informacao/doencas-e-agravos-de-notificacao-de-2007-em-diante-sinan/>>. Acesso em: 2 fev. 2023.
- SINKA, M. E. et al. The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: Occurrence data, distribution maps and bionomic précis. *Parasites and Vectors*, v. 3, n. 1, p. 117, 2010. Disponível em: <<http://www.parasitesandvectors.com/content/3/1/117>>.
- SINKA, M. E. et al. A global map of dominant malaria vectors. *Parasites and Vectors*, v. 5, n. 1, p. 69, 2012. Disponível em: <<http://www.parasitesandvectors.com/content/5/1/69>>.
- SMITH, B. T. et al. Uneven missing data skew phylogenomic relationships within the lorries and lorikeets. *Genome Biology and Evolution*, v. 12, n. 7, p. 1131–1147, 2020.
- SMITH, L. B. *Bromeliad malaria* Annu Rep Smithson Inst, , 1952. .
- SMITH, M. L.; STYCZYNSKI, M. P. Systems Biology-Based Investigation of Host–Plasmodium Interactions. *Trends in Parasitology*, v. 34, n. 7, p. 617–632, 2018. Disponível em: <<https://doi.org/10.1016/j.pt.2018.04.003>>.
- SMITH, S. A.; BROWN, J. W. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, v. 105, n. 3, p. 302–314, 14 mar. 2018. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1002/ajb2.1019>>.

- SPASSKY, B. Morphological differences between sibling species of *Drosophila*. *University of Texas Publications*, v. 5721, p. 48–61, 1957.
- STANLEY C. OAKS, JR., VIOLAINE S. MITCHELL, G. W.; PEARSON, and C. C. J. C. *Malaria — obstacles and opportunities*. [s.l: s.n.]v. 8
- STRUCK, T. H. et al. Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, v. 33, n. 3, p. 153–163, mar. 2018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0169534717302902>>.
- TAJIMA, F. Evolutionary relationship of dna sequences in finite populations. *Genetics*, v. 105, n. 2, p. 437–460, 1 out. 1983. Disponível em: <<https://academic.oup.com/genetics/article/105/2/437/5996285>>.
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, v. 123, n. 3, p. 585–595, 1989.
- TAKKEN, W. et al. Distribution and Dynamics of Larval Populations of *Anopheles messeae* and *A. atroparvus* in the Delta of the Rivers Rhine and Meuse, The Netherlands. *AMBIO: A Journal of the Human Environment*, v. 31, n. 3, p. 212–218, maio 2002. Disponível em: <<http://www.bioone.org/doi/abs/10.1579/0044-7447-31.3.212>>.
- TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, v. 38, n. 7, p. 3022–3027, 25 jun. 2021. Disponível em: <<https://academic.oup.com/mbe/article/38/7/3022/6248099>>.
- TAMURA, K.; TAO, Q.; KUMAR, S. Theoretical Foundation of the RelTime Method for Estimating Divergence Times from Variable Evolutionary Rates. *Molecular Biology and Evolution*, v. 35, n. 7, p. 1770–1782, 1 jul. 2018. Disponível em: <<https://academic.oup.com/mbe/article/35/7/1770/4947869>>.
- TENNESSEN, J. A. et al. A population genomic unveiling of a new cryptic mosquito taxon within the malaria- transmitting *Anopheles gambiae* complex. *Molecular Ecology*, v. 30, n. 3, p. 775–790, 10 fev. 2021. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/mec.15756>>.
- THE ANOPHELES GAMBIAE 1000 GENOMES CONSORTIUM. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*, v. 552, p. 96–100, 2017.
- THEOBALD, F. V. *A Monograph of the Culicidae or Mosquitoes*. 1901. Mainly Compiled from the Collections Received at the British Museum from Various Parts of the World, 1901.
- TOURÉ, Y. T. et al. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. . *Parassitologia*, v. 40, p. 477–511, 1998.
- TURNER, T. L.; HAHN, M. W.; NUZHDIN, S. V. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, v. 3, n. 9, p. 1572–1578, 2005.
- UENO, H. M.; FORATTINI, O. P.; KAKITANI, I. Vertical and seasonal distribution of *Anopheles (Kerteszia)* in Ilha Comprida, Southeastern Brazil. *Revista de Saude Publica*, v. 41, n. 2, p. 269–275, 2007.
- VAN THIEL, P. H. *Sur l'origine des variations de taille de l'Anopheles maculipennis dans les Pays-Bas*. [s.l: s.n.].

- VELOSO, H. P.; MOURA, J. V. de; KLEIN, R. M. *Delimitação ecológica dos anofelíneos do subgênero Kerteszia na região costeira do sul do Brasil* Memórias do Instituto Oswaldo Cruz, 1956. .
- VICOSO, B.; CHARLESWORTH, B. Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews Genetics*, v. 7, n. 8, p. 645–653, ago. 2006. Disponível em: <<https://www.nature.com/articles/nrg1914>>.
- VIOLLE, C. et al. The return of the variance : intraspecific variability in community ecology. v. 27, n. 4, 2012.
- VOGES, K. et al. Novel molecular evidence of population structure in *Anopheles (Kerteszia) bellator* from brazilian atlantic forest. *Memorias do Instituto Oswaldo Cruz*, v. 114, n. 4, p. 1–5, 2019.
- VURTURE, G. W. et al. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, v. 33, n. 14, p. 2202–2204, 2017.
- WALKER, B. J. et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, v. 9, n. 11, p. e112963, 19 nov. 2014. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0112963>>.
- WATERHOUSE, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, v. 35, n. 3, p. 543–548, 2018.
- WHITE, B. J. et al. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology*, v. 19, n. 5, p. 925–939, mar. 2010. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2010.04531.x>>.
- WHO. *Malaria World Report 2022*. [s.l: s.n.].
- WHO. *Malaria, Key facts*. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/malária>>. Acesso em: 28 maio. 2023.
- WILKERSON, R.; PEYTON, E. The Brazilian malaria vector *Anopheles (Kerteszia) cruzii*: life stages and biology (Diptera: Culicidae). *Mosquito Systematics*, v. 23, n. 2, p. 110–122, 1991. Disponível em: <<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA512864>>.
- WRBU. *Anopheles cruzii species page*. Disponível em: <<https://wrbu.si.edu/vectorspecies/mosquitoes/cruzii>>.
- YALCINDAG, E. et al. Multiple independent introductions of *Plasmodium falciparum* in South America. *Proceedings of the National Academy of Sciences of the United States of America*, v. 109, n. 2, p. 511–516, 2012.
- YOUNG, D. G.; DUNCAN, M. A. *Guide to the identification and geographic distribution of Lutzomyia sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae)*. Gainesville, FL.: Memories American Entomological Institute No. 54, Associated Publishers., 1994.
- ZACHOS, F. E. *Species Concepts in Biology*. Cham: Springer International Publishing, 2016.
- ZAMYATIN, A. et al. Chromosome-level genome assemblies of the malaria vectors *Anopheles coluzzii* and *Anopheles arabiensis*. *GigaScience*, v. 10, n. 3, 15 mar. 2021.

Disponível em:

<<https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giab017/6170950>>.

ZAVORTINK, T. A review of the subgenus *Kerteszia* of Anopheles. Contributions of the American Entomological Institute. *Contributions of the American Entomological Institute*, v. 9, p. 1–54, 1973.

Figure legends:

Figure 1 - *Kerteszia* collection sites in the Brazilian Atlantic Rainforest. Map of Brazil (a) and detail (b) showing the location where *Kerteszia* samples were collected. The maps in Figure 1 were made using the packages *maptools*, *maps*, and *GISTools* from R Software (R Core Team, 2021).

Figure 2- A- Alignment of *cpr* gene nucleotide sequences and primer positions. Alignment of the *cpr* gene fragment of an *An. cruzii* s.l. from Bocaina Group 1 and two sequences from Bocaina Group 2 (according to Dias et al. 2018). The GeneBank accession numbers KT724974.1 and KT724976.1 correspond to Bocaina Group 1, and KT724986.1 and KT724996.1 to Bocaina Group 2. The inverted triangle represents the position of the single nucleotide difference between the two forward primers that distinguish the Bocaina Group 1 (Boc1F in blue) and the Bocaina Group 2 (Boc2F in green). The reverse primer (BocR) is in purple. **B- Visualization in agarose gel (1%) of samples classified as Bocaina Group 1 (G1) and Bocaina Group 2 (G2).** G1: forward primer designed for Bocaina Group 1 (Boc1F: 5'GTGTAATATGGTAAGCGAACG). G2: forward primer designed for Bocaina Group 2 (Boc2F: 5'GTGTAATATGGTAAGCGAAACG) (These primers amplified a ~180 bp fragment of the *cpr* gene) C-: negative control (use of water instead of DNA).

Figure 3 - Genetic differentiation of *An. cruzii* s.l. sibling species using orthologous genes. The F_{ST} -values (Y-axis) were calculated between all pairwise comparisons of *An. cruzii* s.l. sibling species from: (i) South/Southeast Brazil (Flo + Gua: Florianópolis + Guapimirim), (ii) Itatiaia, (iii) Bocaina Group 1, and (iv) Bocaina Group 2. The genes were classified into chromosomal arms based on the location of their orthologs in the *An. cruzii* genome from the *Anopheles* Reference Genomes Project (GenBank assembly accession: GCA_943734635.1) (X-axis).

Figure 4 - Boxplot highlighting averages of F_{ST} values. The F_{ST} -values (Y-axis) were calculated between all pairwise comparisons of *An. cruzii* s.l. sibling species from: (i) South/Southeast Brazil (Flo + Gua: Florianópolis + Guapimirim), (ii) Itatiaia, (iii) Bocaina Group 1, and (iv) Bocaina Group 2.

Figure 5 - Species tree inferred with Maximum-Likelihood (ML) approach from 2051 genes present in at least 20 of the 40 *Kerteszia* samples. *Anopheles (Kerteszia) bellator* and *Anopheles (Kerteszia) homunculus* are recognized different species and were used to root the tree. These phylogenomic analyses strongly suggest that *An. cruzii* s.l. is composed of at least four cryptic species: Bocaina Group 1, Bocaina Group 2 + Itatiaia, Florianópolis + Guapimirim + Maquiné, Santa Teresa. The evidence of cryptic speciation is especially compelling in the case of the Bocaina samples, which came from the same location and yet form two groups quite distant phylogenetically. Numbers on the nodes represent bootstrap values (ML). Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; San: Santa Teresa; Itp: Itaparica. M: males; F: females.

Table 1: Some studies that suggest that *An. cruzii* is a cryptic species complex.

Table 2 - Genomescope results with k-mer size 21. *Library Construction:* Illumina paired-end DNA-seq libraries TruSeq Nano DNA (insert size 350bp) or Nextera DNA XT) produced and sequenced with Hiseq 2000. *Length (bp):* inferred total genome length based on K-mer size. *Total length assembly:* total number of bases in the assembly (Estimated by QUAST). *Hetero. (%):* rate of heterozygosity. *dup (%):* average rate of read duplications. *uniq (%):* percent of the genome that is unique (not repetitive). *Kcov:*

mean k-mer coverage for heterozygous bases. *Prof.* ($kcov \times 2$): depth - the *kcov* multiplied by two. *Error (%)*: error rate of the reads. Flo: Florianópolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica. M: males; F: females.*For the long reads samples the mean coverage was estimated by Flye.

Table 3 - BUSCO v4.1.4 results for the 40 *Kerteszia* samples. Assemblies automatic annotation of 3285 orthologous genes conserved among Diptera from the OrthoDB database (odb10, downloaded from <https://BUSCO.ezlab.org/>). The number of genes in each genome with complete sequences (C and S), genes duplicated (D), genes with fragmented sequences (F), or genes absent in each genome (M) is shown in this table. The values in parentheses are the percentage of BUSCOs in each assembly. Flo: Florianópolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica. M: males; F: females.

Table 4: Results from BUSCO v4.1.4 for samples sequenced with Nanopore. Automatic annotation of 3285 conserved orthologous gene assemblies among Diptera from the OrthoDB database (odb10, downloaded from <https://BUSCO.ezlab.org/>). The number of genes in each genome with complete single-copy and duplicate sequences is shown for each genome in this table. Values in parentheses are the percentage of BUSCOs in each assembly. Statistics of assemblies produced with QUAST. Total Size: Total number of bases in the assembly. # contigs: total number of contigs in the assembly. Longest contig: Length of the longest contig in the assembly. GC (%): total number of G and C nucleotides in the assembly divided by the total length. N50: genome assembly integrity (sequence length of the shortest contig within 50% of the total genome length). Boc: Mouth; Flo: Florianópolis; F: Female.

Table 5: Calculations of expected genotype frequencies under Hardy-Weinberg equilibrium.

Additional file 1 - Detailed information about the samples used in this study. *Sex*: sex of each sample – M for males, and F for females. *Source*: Bromeliad - immatures collected in the water of bromeliads and reared in the laboratory; CDC + CO₂ - adults collected using CDC traps with CO₂; Family - F1 generation of female mosquitoes collected in the field. Flo: Florianópolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica. M: males; F: females.

Additional file 2: Busco V3 results comparing SPAdes and Platanus assemblers in seven samples. Automatic annotation of 2799 single-copy orthologs from the Diptera reference set of orthologs (odb9, downloaded from <https://BUSCO.ezlab.org/>). The number of genes with complete sequences (C and S), genes duplicated (D), genes with fragmented sequences (F), or genes absent in each genome (M) is shown in this table. The values in parentheses are the percentage of BUSCOs in each assembly. File type: file before or after removing contaminants – B; files where contaminants have not yet been removed, and C; files in which contaminants have already been removed. Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Ilh: Ilha Grande; Itp: Itaparica. M: males; F: females.

Additional file 3: *Kerteszia* assemblies statistics produced with QUAST. *Total length*: total number of bases in the assembly. *# contigs*: total number of contigs in the assembly. *Largest contig*: length of the longest contig in the assembly. *GC (%)*: total number of G and C nucleotides in the assembly divided by the total length. *N50*: genome assembly completeness (sequence length of the shortest contig at 50% of the total genome length).

Flo: Florianopolis; San: Santa Teresa; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; Itp: Itaparica. M: males; F: females.

Additional file 4 - F_{ST} -values from the orthologous genes used in this analysis. *Gene ID*: Gene identity for *Anopheles gambiae*. *Orthodb ID*: Gene identity in Orthodb10. The F_{ST} -values were calculated between all pairwise comparisons of *An. cruzii* s.l. siblings from: (i) South/Southeast Brazil (*cruzii*), (ii) Itatiaia (*ita*), (iii) Bocaina Group 1 (*Boc1*), and (iv) Bocaina Group 2 (*Boc2*). The genes were classified into chromosomal arms based on the location of their orthologs in the *An. cruzii* genome from the *Anopheles* Reference Genomes Project (GenBank assembly accession: GCA_943734635.1).

Additional file 5 - ML and MSC trees: Species tree inferred with Maximum-Likelihood (ML) and Multispecies Coalescent (MSC) approaches from 2051 genes present in at least 20 of the 40 *Kerteszia* samples. *Anopheles (Kerteszia) bellator* and *Anopheles (Kerteszia) homunculus* are recognized different species. These phylogenomic analyses strongly suggest that *An. cruzii* s.l. is composed of at least four cryptic species: Bocaina Group 1, Bocaina Group 2 + Itatiaia, and Florianopolis + Guapimirim + Maquiné, Santa Teresa. The evidence of cryptic speciation is especially compelling in the case of the Bocaina samples, which came from the same location and yet form two groups quite distant phylogenetically. Numbers on the nodes represent bootstrap values (ML) and posterior probabilities (MSC). Flo: Florianópolis; Boc: Bocaina; Itt: Itatiaia; Gua: Guapimirim; Ilh: Ilha Grande; San: Santa Teresa; Itp: Itaparica. M: males; F: females.

Additional file 6 - Morphological structures: Photos of the essential structures for identification at the species level (head, wing, paw and scutum), from the 40 samples. 1- *An. bellator* Ilh F1, 2- *An. bellator* Ilh M1, 3- *An. bellator* Itp F1, 4- *An. bellator* Itp M1), 5- *An. bellator* Cam F11, 6- *An. bellator* Cam F14, 7- *An. bellator* Cam F25, 8- *An. bellator* Cam F33, 9- *An. bellator* Cam F34, 10- *An. bellator* Cam F36, 11- *An. homunculus* Cam F1, 12- *An. homunculus* Cam F31, 13- *An. homunculus* Flo F1, 14- *An. homunculus* San F2, 15- *An. homunculus* San M2, 16- *Kerteszia* Boc F1, 17- *Kerteszia* Boc F2, 18- *Kerteszia* Boc M1, 19- *Kerteszia* Boc M2, 20- *Kerteszia* Boc M3, 21- *Kerteszia* Boc M4, 22- *Kerteszia* Boc M5, 23- *Kerteszia* Boc M6, 24- *Kerteszia* Boc F3 N, 25- *Kerteszia* Flo F3 N_1, 26- *Kerteszia* Flo F3 N_2, 27- *Kerteszia* Flo F2, 28- *Kerteszia* Flo M1, 29- *Kerteszia* Flo M2, 30- *Kerteszia* Flo M3, 31- *Kerteszia* Gua F1, 32- *Kerteszia* Gua F2, 33- *Kerteszia* Itt F1, 34- *Kerteszia* Itt M1, 35- *Kerteszia* Itt M2, 36- *Kerteszia* Itt M3, 37- *Kerteszia* Itt M4, 38- *Kerteszia* San F1, 39- *Kerteszia* San M1, in this order.

Additional file 7 - Morphological characters of the male genitalia: Images of male genitalia. A- *Kerteszia* Boc M4 (Bocaina Group 1); B- *Kerteszia* Boc M2 (Bocaina Group 2), C- *Kerteszia* San M1; *Kerteszia* D- Flo M2.

Apêndice 2

SHORT COMMUNICATION

Mem Inst Oswaldo Cruz, Rio de Janeiro, Vol. 114: e180598, 2019

Novel molecular evidence of population structure in *Anopheles (Kerteszia) bellator* from Brazilian Atlantic Forest

**Kamila Voges¹, Marcela Possato Correa da Rosa¹, Betina Westphal-Ferreira²,
Mario Antonio Navarro-Silva², Carime Lessa Mansur Pontes³, André Nóbrega Pitaluga^{4,5,6},
Carlos José de Carvalho-Pinto^{3,6}, Luísa DP Rona^{4,5,6/+}**

¹Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Departamento de Biologia Celular, Embriologia e Genética, Florianópolis, SC, Brasil

²Universidade Federal do Paraná, Setor de Ciências Biológicas, Departamento de Zoologia, Curitiba, PR, Brasil

³Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Departamento de Microbiologia, Imunologia e Parasitologia, Florianópolis, SC, Brasil

⁴Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de Biologia Molecular de Parasitas e Vetores, Rio de Janeiro, RJ, Brasil

⁵Imperial College London, Department of Life Sciences, London, United Kingdom

⁶Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular, Rio de Janeiro, RJ, Brasil

Anopheles bellator is a primary malaria vector in the Atlantic Forest. Partial sequences of *timeless* and *Clock* genes were used to assess the genetic differentiation of five Brazilian populations, which showed strong population structure (e.g. high F_{ST} values and fixed differences) in all pairwise comparisons between Bahia sample and the others from Paraná, São Paulo and Rio de Janeiro states. Also, the resulting phylogenetic trees clearly grouped the sequences from Bahia in a different cluster with high bootstrap values. Among southern and southeastern populations low levels of genetic differentiation were found suggesting a general stability of the genetic structure.

Key words: *Anopheles (Kerteszia) bellator* - bromeliad-malaria - Brazilian Atlantic Forest

Anopheles (Kerteszia) bellator, a mosquito species whose immature stages develop in water accumulated in bromeliads, occurs in Trinidad and Tobago and coastal areas spanning eastern Venezuela to southern Brazil.^(1,2) Interestingly, it is only considered a main human malaria vector in the two extremes of its distribution area,^(3,4) and despite its epidemiological importance, there are few molecular studies concerning this species. Carvalho-Pinto and Lourenço-de-Oliveira⁽⁵⁾ evaluated *An. bellator* isoenzymes in populations from three locations in Brazil [Santa Catarina (SC), São Paulo (SP) and Bahia (BA)], and one from Trinidad Island (Trinidad and Tobago). They concluded that the Trinidad Island population was genetically distant from the Brazilian populations, mostly from BA. Also, regarding the three Brazilian samples, SC and SP were genetically close to each other, and quite distant from BA. Recently, a handful of studies have tried to understand the evolutionary relationship among different species from *Kerteszia* subgenus (including *An. bellator*, *Anopheles cruzii* and *Anopheles homunculus*),^(6,7) rather than investigating the genetic diversity between *An. bellator* populations. Therefore, to date, there have

major group with broad range dispersion from south and south-east Brazil and a distinct sibling species in BA.^(8,9,10,11) This knowledge raises an interesting question: are *An. bellator* populations highly structured like those from *An. cruzii*; or, like *An. homunculus*, is it a species without significant genetic differentiation in its wide range distribution? To answer this question and to better understand the genetic variability and divergence among *An. bellator* populations, partial fragments of *timeless* and *Clock* genes were evaluated in five Brazilian locations: Ilha do Mel (PR), Cananéia (SP), Abraão and Sítio Forte (RJ) and Camacan (BA). *Clock* and *timeless* loci are involved in the circadian rhythms of insects,⁽¹²⁾ and have previously been shown to be good molecular markers to define different species in the *An. cruzii* complex.^(10,11,13,14)

MATERIALS AND METHODS

Sampling sites - The mosquitoes used in this study were captured along the Brazilian Atlantic Forest at the following localities: Ilha do Mel (PR), Cananéia (SP), Abraão and Sítio Forte (RJ), and Camacan (BA) (Fig. 1). The maps found in Fig. 1 were made using the packages *maptools*, *maps* and *GISTools*^(15,16,17) from R Software version 3.3.2.⁽¹⁸⁾

online | memorias.ioc.fiocruz.br

2|5 Kamila Voges et al.

doi: 10.1590/0074-02760180598

Financial support: FIOCRUZ (ANP), CNPq (grant no. 442300/2014-0) and Newton International Fellowship / Royal Society (grant no. NF161472) (LDPR).

+ Corresponding author: luisa.rona@imperial.ac.uk; luisa.rona@ufsc.br

<https://orcid.org/0000-0002-3400-3950>

Received 18 December 2018 Accepted

3 April 2019



been no further investigations regarding the relationship between *An. bellator* populations from Brazil.

From the south of Brazil to the north-east, *An. bellator*, *An. homunculus* and *An. cruzii* are sympatric in the Atlantic Forest.^(1,8) However, while *An. homunculus* populations from these regions showed no clear genetic separation — they were in fact considered a single species — *An. cruzii* was shown to be a complex of at least two siblings: one

An. bellator adults and immatures were obtained, treated and preserved as described in de Rezende Dias et al.⁽¹⁴⁾ The immatures were collected in bromeliads located on rocks, in Restinga areas. Species identification was carried out according to Consoli and Lourenço-de-Oliveira.⁽²⁾ The detailed field collection information, such as sex and the life stage in which the samples were collected (adult or larvae), of each specimen is shown in Supplementary data (Table I).

Molecular data - The *timeless* and *Clock* partial gene sequences from the five *An. bellator* populations were obtained by polymerase chain reaction (PCR), cloning and sequencing as described in de Rezende Dias et al.⁽¹⁴⁾ The primers used in the PCR reactions were those described in Rona et al.^(10,11) The sequencing was carried out in an ABI3130 DNA

sequencer at Myleus Biotechnology. For each mosquito, at least eight clones were sequenced to mitigate PCR errors, and allow the identification of the two alleles. Consensus sequences representing the two alleles were generated. The individuals were classified as homozygotes when only one haplotype was observed among the eight sequences. Sequences were submitted to GenBank under the Accession Numbers: MG755641 - MG755734. An alignment of all *An. bellator* sequences for each gene is presented in Supplementary data (Figs. 1,2).

Sequences from the *timeless* gene fragment from *Anopheles (Kerteszia) laneanus* and *Anopheles*

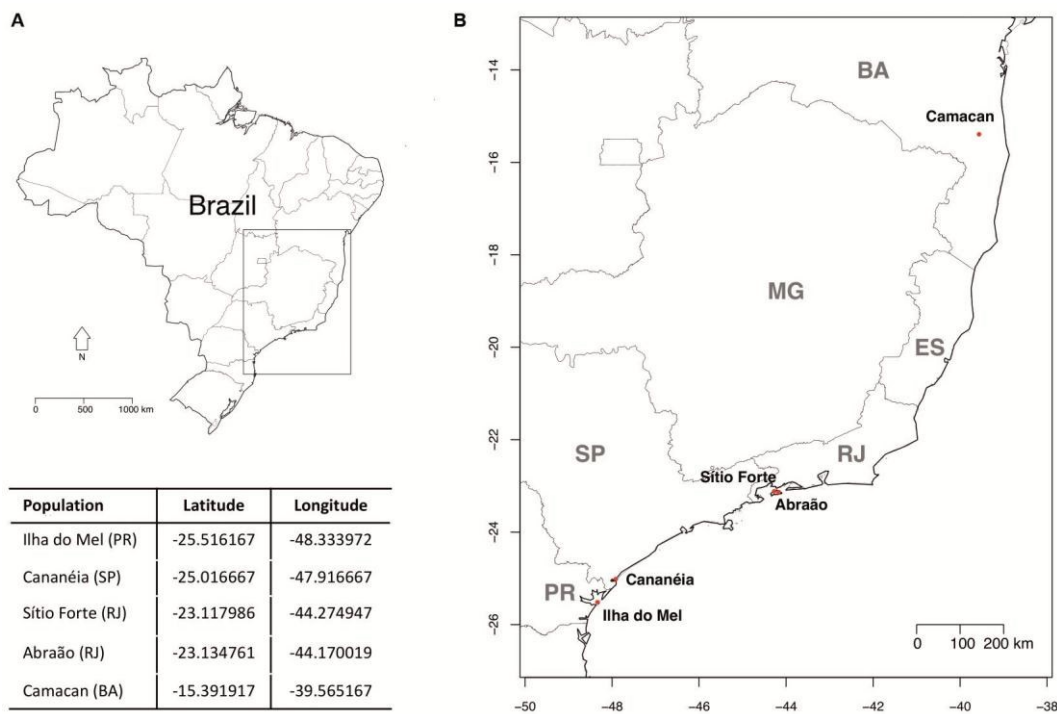
(*Kerteszia*) *cruzii* s.s. were used in order to verify the relationship of *An. bellator* samples analysed in this study with other closely related *Kerteszia* species. The sequences from *An. laneanus* were obtained by PCR, cloning and sequencing as described above and submitted to GenBank under the Accession Numbers: KT803975 - KT803978. The sequences from *An. cruzii* s.s. were those previously published by Rona et al. (10) (GenBank accession numbers FJ408732, FJ408735, FJ408746, FJ408786, FJ408788,

for SC population and FJ408838, FJ408840, FJ408848, FJ408854, FJ408855, FJ408856, FJ408862, FJ408864 for BA population).

DNA sequence analysis - Anopheles bellator DNA sequences were aligned with MAFFT⁽¹⁹⁾ and phylogenetic trees were constructed for each gene under the Maximum Likelihood method using the *ape* and *phangorn* packages^(20,21) in R version 3.3.2.⁽¹⁸⁾ The best-fit substitution models HKY+G (*Clock*) and TIM3e+G+I (*timeless*) were selected following the AIC criterion using the modelTest function from *phangorn*.⁽²¹⁾ An additional *timeless* phylogenetic tree using the coding regions from *An. bellator*, *An. laneanus* and *An. cruzii* s.s. were also constructed (K80+G model) as described above.

DnaSP v5 software⁽²²⁾ was used to calculate intrapopulation statistics, including nucleotide diversity and number of segregating sites, and also to ensure that all loci met the neutrality model by calculating the Tajima's

D.⁽²³⁾ The P_{RO}SEQ v 2.91⁽²⁴⁾ software was used to perform inter-population analyses, including the



1: *Anopheles bellator* collection sites in the Brazilian Atlantic Rainforest. (A): map of Brazil. (B): magnification of the black box from (A) laying the sites where *An. bellator* samples were collected. The x and y axes drawn in the map (B) display the longitude and latitude, ectively.

FJ408813, FJ408816, FJ408832, FJ408736, FJ408738 identification of fixed and shared polymorphisms

and the generation of pair-wise estimates of population differentiation (e.g. F_{ST} , D_{xy} and D_a).

RESULTS AND DISCUSSION

Population structure analysis - A total of 94 sequences (two alleles from each *An. bellator* individual) were analysed for the two loci [Supplementary data (Table II)]. *Clock* and *timeless* fragment sequences were 164 bp and 368 bp long, respectively [Supplementary data (Figs 1,2)]. Polymorphism data from a variety of measures — e.g. nucleotide diversity (θ and π values) — showed that in general the *timeless* gene has lower genetic diversity compared with *Clock* [Supplementary data (Table II)], although the first has a bigger fragment size. None of the Tajima's D results were significant, which means that there is no evidence of natural selection acting in any of the populations [Supplementary data (Table II)].

In all pairwise comparisons with the BA population, very high and significant F_{ST} values (ranging from ~ 0.62 to 0.67) were found, there were few shared polymorphisms and a large number of fixed differences (Table), some of which caused amino acid changes [Supplementary data (Figs 1,2)]. The average number of nucleotide substitutions per site (D_{xy}) and the number of net nucleotide substitutions per site between populations (D_a) was also higher in the BA pairwise comparisons, confirming the F_{ST} measurements. The southern and southeastern populations exhibited limited differentiation, characterized by shared polymorphisms and no fixed differences.

Haplotype genealogies analysis - The genealogical relationship of haplotypes was inferred by Maximum Likelihood for the two genes (Fig. 2). The resulting

trees showed no clear separation between the sequences from PR, SP and RJ (henceforth called *An. bellator* A), however they clearly grouped the sequences from BA (henceforth called *Anopheles bellator* B) in a different cluster with high bootstrap values.

In order to verify the genealogical relationship between *An. bellator* A and B with other closely related *Kerteszia* species (*An. cruzii* s.s. and *An. laneanus*), a Maximum Likelihood tree was constructed using the coding regions of the *timeless* gene fragment, since the introns showed a substantial number of indels in the alignment [Supplementary data (Fig. 3)]. The resulting phylogenetic tree showed that *An. bellator* A and B are a monophyletic group, clustering them in a separated clade with high bootstrap value. The final topology was very similar to those from Foster et al.⁽²⁵⁾ and Lorenz et al.⁽⁶⁾ using mitochondrial and genomic sequences: *An. cruzii* s.s. and *An. laneanus* are genetically closer when compared with *An. bellator*. An attempt was also made to compare *An. bellator* *Clock* sequences with those from different *Kerteszia* species available in the GenBank. Although, the same problem found in the *timeless* noncoding regions (indels) was also disclosed in the *Clock* alignment, but in a more challenging way, since more than 70% of the fragment is composed by an intron.

Anopheles bellator A and B - Two highly *An. bellator* structured groups (*An. bellator* A and B) were disclosed in this study. The population divergence between them is well supported by (i) fixed differences (Table), (ii) high F_{ST} values (average of 0.65) (Table), and (iii) it is the earliest divergent node in both phylogenetic trees with strong

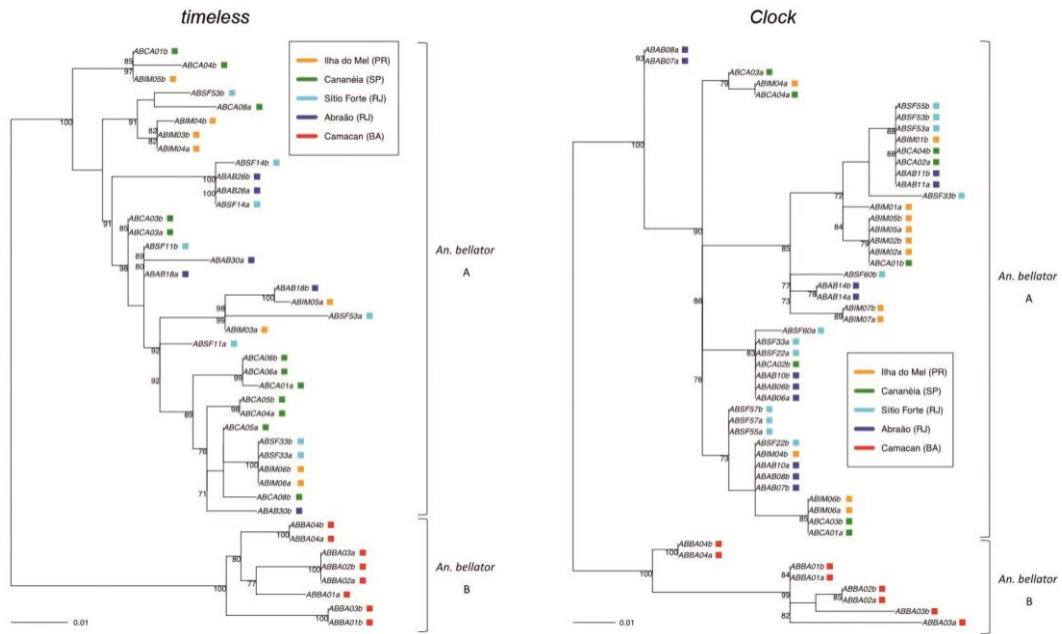


Fig. 2: maximum likelihood trees of *Anopheles bellator* *timeless* and *Clock* sequences (TIM3e+G+I and HKY+G models, respectively). These phylogenies show that two structured groups were identified in the *An. bellator* populations from Brazil: *An. bellator* B occurs in the NorthEast Atlantic Rainforest (Camacan) and *An. bellator* A is found in the other more Southern Brazilian regions. Numbers on the nodes represent the percentage bootstrap values based on 1000 replications. Lowercase letters in the haplotype names specify the two alleles (A or B) obtained from each individual.

4 | 5 Kamila Voges et al.

TABLE
Genetic differentiation between all *Anopheles bellator* populations

Populations	Gene	Km	F_{ST}	D_{xy}	D_a	S_s	S_f
Abraão (RJ) x Sítio Forte (RJ)	<i>timeless</i>	10.89	-0.1039	0.0177	-0.0018	11	00
	<i>Clock</i>		-0.0157	0.0322	-0.0005	08	00
Ilha do Mel (PR) x Abraão (RJ)	<i>timeless</i>	496.5	0.0411	0.0215	0.0009	09	00
	<i>Clock</i>		0.1134	0.0404	0.0046	07	00
Ilha do Mel (PR) x Sítio Forte (RJ)	<i>timeless</i>	488.6	0.0695	0.0224	0.0016	11	00
	<i>Clock</i>		0.0904	0.0377	0.0034	08	00
Ilha do Mel (PR) x Cananéia (SP)	<i>timeless</i>	70.1	0.1147	0.0233	0.0027	10	00
	<i>Clock</i>		-0.0508	0.0376	-0.0019	10	00
Cananéia (SP) x Sítio Forte (RJ)	<i>timeless</i>	425.7	0.1158*	0.0221	0.0026	08	00
	<i>Clock</i>		0.0001	0.0364	0.0000	09	00
Cananéia (SP) x Abraão (RJ)	<i>timeless</i>	434.1	0.1311*	0.0222	0.0029	06	00
	<i>Clock</i>		0.0086	0.0383	0.0003	08	00
Camacan (BA) x Ilha do Mel (PR)	<i>timeless</i>	1449	0.6471*	0.0621	0.0402	00	07
	<i>Clock</i>		0.6482*	0.1168	0.0757	02	05
Camacan (BA) x Cananéia (SP)	<i>timeless</i>	1379	0.6694*	0.0622	0.0417	00	09

			Clock	0.6287*	0.1163	0.0731	02	05
Camacan (BA) x Abraão (RJ)		timeless	987.1	0.6722*	0.0628	0.0422	01	10
			Clock	0.6236*	0.1050	0.0655	02	03
Camacan (BA) x Sítio Forte (RJ)		timeless	990.1	0.6758*	0.0641	0.0434	00	09
			Clock	0.6562*	0.1105	0.0725	02	05

F_{ST} : pair-wise estimates of population differentiation. The significance of F_{ST} values was evaluated by 1,000 random permutations (* $p < 0.05$). D_{xy} : average number of nucleotide substitutions per site between populations; D_a : number of net nucleotide substitutions per site between populations.⁽²⁹⁾ S_s : number of shared polymorphisms between the two populations. S_f : number of fixed differences between the two populations. Km: the approximated geographic distances between localities in km.

statistical support (Fig. 2). The results shown here are in agreement with a former study that used allozymes as genetic markers and showed that among all pairwise comparisons between Brazilian *An. bellator* samples, the highest genetic distances were found between Itaparica Island (BA) and the other more southern populations from Florianópolis (SC) and Cananéia (SP).⁽⁵⁾ However, the BA population (Camacan) used in this study is 300 Km away (to the south direction) from Itaparica Island (BA), which raises new questions like: Does *An. bellator* from Camacan belong to the same group as *An. bellator* from Itaparica Island? Also, are there additional structured *An. bellator* groups in the Atlantic Forest? A more extensive sampling across the geographic distribution of *An. bellator*, mainly focusing in Espírito Santo (ES) and BA, and the use of a larger number of molecular markers are desirable in order to better understand the evolutionary forces that establish and maintain the population structure of *An. bellator*. Accordingly, a number of other studies have shown the same separation pattern between northern and southern Atlantic Forest segments in many fauna (including *An. cruzii* s.s.) and flora groups, which converge around northern ES and southern BA, suggesting a common vicariant event.^(10,11,26)

The high genetic differentiation observed between BA and the more southern populations might also be explained by the geographic distance between them (at least 980 Km away) (Table). However, we would expect higher F_{ST} values in the pairwise comparisons between PR and SP, with the two populations from RJ (average F_{ST} of only 0.07), which do not have a negligible distance apart (~ 500 km). Even so, among the southern and southeastern populations, we found a widespread lack of significant differentiation.

This study uncovers two structured *An. bellator* groups in the Brazilian Atlantic Forest: *An. bellator* A is widespread in Southern and Southeastern regions, and *An. bellator* B is found in the north-east Brazil. Interestingly, in spite of the fact that malaria cases are accounted for yearly in BA, *An. bellator* has never been viewed as a malaria vector in this state^(27,28) as it is for the Southern Brazilian regions.⁽²⁾ So, future studies would be valuable in order to differentiate the two *An. bellator* groups regarding malaria susceptibility.

ACKNOWLEDGEMENTS

To Dr Edmundo Carlos Grisard and Dr Patricia Hermes Stoco for their logistic support during this study; to Bibiana Sgorla, Clisten Fátima Staffen and Mari Dalva Staffen for their technical assistance; to David A Ellis for kindly reviewing the English and to Dr Rosely dos Santos Malafronte and Dr Maria Anice M Sallum for providing *Anopheles bellator* from Cananéia and *Anopheles laneanus*, respectively.

AUTHORS' CONTRIBUTION

LDPR, KV, MPCR and CLMP participated in data generation and analysis; LDPR, BWF, MANS, CJCP and ANP carried mosquito collections and morphological identification; LDPR drafted the manuscript; ANP and CJCP helped in the paper drafting by critically reading the original manuscript; LDPR was the principal investigator, participated in its design and coordination. All authors read and approved the final manuscript.

REFERENCES

1. Aragão MB. Geographic distribution and abundance of *Anopheles* species (*Kerteszia*) (Diptera, Culicidae). Rev Bras Malariol Doenças Trop. 1964; 16: 73-109.
2. Consoli RAGB, Lourenço-de-Oliveira R. Principais mosquitos de importância sanitária no Brasil. Rio de Janeiro: Fiocruz; 1994. 225 pp.

3. Smith LB. Bromeliad malaria. Annual Rep Smithson Inst. 1952; 385-98.
4. Rachou RG, Ferreira MO, Lobo AG, Pires WM. Epidemiology of malaria in Southern Brazil. Rev Bras Malariol Doenças Trop. 1955; 6: 177-88.
5. de Carvalho-Pinto CJ, Lourenço-de-Oliveira R. Isoenzymatic analysis of four *Anopheles (Kerteszia) bellator* Dyar & Knab (Diptera: Culicidae) populations. Mem Inst Oswaldo Cruz. 2003; 98(8): 1045-8.
6. Lorenz C, Patané JS, Suesdek L. Morphogenetic characterisation, date of divergence, and evolutionary relationships of malaria vectors *Anopheles cruzii* and *Anopheles homunculus*. Infect Genet Evol. 2015; 35: 144-52.
7. Oliveira TM, Foster PG, Bergo ES, Nagaki SS, Sanabani SS, Marinotti O, et al. Mitochondrial genomes of *Anopheles (Kerteszia)* (Diptera: Culicidae) from the Atlantic Forest, Brazil. J Med Entomol. 2016; 53(4): 790-7.
8. Cardoso JC, Bergo ES, Oliveira TMP, Sant'ana DC, Motoki MT, Sallum MAM. New records of *Anopheles homunculus* in Central and Serra do Mar biodiversity corridors of the Atlantic Forest, Brazil. J Am Mosq Control Assoc. 2012; 28(1): 1-5.
9. de Carvalho-Pinto CJ, Lourenço-de-Oliveira R. Isoenzymatic analysis of four *Anopheles (Kerteszia) cruzii* (Diptera: Culicidae) populations of Brazil. Mem Inst Oswaldo Cruz. 2004; 99(5):471-5.
10. Rona LD, Carvalho-Pinto CJ, Gentile C, Grisard EC, Peixoto AA. Assessing the molecular divergence between *Anopheles (Kerteszia) cruzii* populations from Brazil using the *timeless* gene: further evidence of a species complex. Malar J. 2009; 8: 60.
11. Rona LD, Carvalho-Pinto CJ, Mazzoni CJ, Peixoto AA. Estimation of divergence time between two sibling species of the *Anopheles (Kerteszia) cruzii* complex using a multilocus approach. BMC Evol Biol. 2010; 10: 91.
Mem Inst Oswaldo Cruz, Rio de Janeiro, Vol. 114, 2019
12. Sakai T, Ishida N. Circadian rhythms of female mating activity governed by clock genes in *Drosophila*. Proc Natl Acad Sci. 2001; 98(16): 9221-5.
13. Rona LDP, Carvalho-Pinto CJ, Peixoto AA. Evidence for the occurrence of two sympatric sibling species within the *Anopheles (Kerteszia) cruzii* complex in southeast Brazil and the detection of asymmetric introgression between them using a multilocus analysis. BMC Evol Biol. 2013; 13: 207.
14. de Rezende Dias G, Fujii TTS, Fogel BF, Lourenço-de-Oliveira R, Silva-do-Nascimento TF, Pitaluga AN, et al. Cryptic diversity in an Atlantic Forest malaria vector from the mountains of SouthEast Brazil. Parasit Vectors. 2018; 11(1): 36.
15. Brunson C, Chen H. GISTools: some further GIS capabilities for R. R package version 0.7-4. 2014. Available from: <https://CRAN.R-project.org/package=GISTools>.
16. Bivand R, Lewin-Koh N. maptools: tools for reading and handling spatial objects. R package version 0.9-2. 2017. Available from: <https://CRAN.R-project.org/package=maptools>.
17. Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A. maps: Draw Geographical Maps. R package version 3.2.0. 2017. Available from: <https://CRAN.R-project.org/package=maps>.
18. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2014.
19. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30(4): 772-80.
20. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004; 20(2): 289-90.
21. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011; 27(4): 592-3.
22. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003; 19(18): 2496-7.
23. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123(3): 585-95.
24. Filatov D. ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. Mol Ecol Notes. 2002; 2: 621-4.
25. Foster PG, de Oliveira TMP, Bergo ES, Conn JE, Sant'Ana DC, Nagaki SS, et al. Phylogeny of *Anophelinae* using mitochondrial protein coding genes. R Soc Open Sci. 2017; 4(11): 170758.
26. Patterson BD, Costa LP. Bones, clones, and biomes: the history and geography of recent neotropical mammals. Chicago: University of Chicago Press; 2012. 432 pp.
27. Rachou RG. Anofelinos do Brasil: Comportamento das espécies vetoras de malária. Rev Bras Malar Doenças Trop. 1958; 10: 145-81.
28. Deane LM. Malaria vectors in Brazil. Mem Inst Oswaldo Cruz. 1986; 81(Suppl. II): 5-14.
29. Nei M, Kumar S. Molecular evolution and phylogenetics. New York: Oxford University Press; 2000. 352 pp.