

CARLOS HENRIQUE TONHATTI

Análise da variação molecular de *Poecilia
vivipara* (Cyprinodontiformes: Poeciliidae)

CAMPINAS
2012

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA



CARLOS HENRIQUE TONHATTI

Análise da variação molecular de *Poecilia vivipara*

(Cyprinodontiformes: Poeciliidae)

Este exemplar corresponde à redação final da tese defendida pelo(a) candidato (a) <u>CARLOS HENRIQUE TONHATTI</u> <u>Sérgio Furtado</u> e aprovada pela Comissão Julgadora.
--

Dissertação apresentada ao Instituto de Biologia para obtenção do Título de Mestre em Genética e Biologia Molecular, na área de Genética Animal e Evolução.

A handwritten signature in black ink, appearing to read "Sérgio Furtado dos Reis".

Orientador: Sérgio Furtado dos Reis

CAMPINAS

2012

FICHA CATALOGRÁFICA ELABORADA POR
ROBERTA CRISTINA DAL' EVEDOVE TARTAROTTI – CRB8/7430
BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

T614a	<p>Tonhatti, Carlos Henrique, 1983- Análise da variação molecular de <i>Poecilia vivipara</i> (Cypronodontiformes: Poecillidea) / Carlos Henrique Tonhatti. – Campinas, SP: [s.n.], 2012.</p> <p>Orientador: Sérgio Furtado dos Reis. Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.</p> <p>1. <i>Poecilia vivipara</i>. 2. Genética de populações. 3. Região de controle. 4. DNA mitocondrial. 5. Haplótipos. I. Reis, Sérgio Furtado dos, 1952-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.</p>
-------	---

Informações para Biblioteca Digital

Título em Inglês: Analysis of molecular variation of *Poecilia vivipara*
(Cypronodontiformes: Poecillidea)

Palavras-chave em Inglês:

Poecilia vivipara

Population genetics

Control region

Mitochondrial DNA

Haplotypes

Área de concentração: Genética Animal e Evolução

Titulação: Mestre em Genética e Biologia Molecular

Banca examinadora:

Sérgio Furtado dos Reis [Orientador]

Anete Pereira de Souza

Márcio Silva Araújo

Data da defesa: 20-07-2012

Programa de Pós Graduação: Genética e Biologia Molecular

Campinas, 20 de julho de 2012.

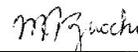
BANCA EXAMINADORA

Prof. Dr. Sérgio Furtados dos Reis (Orientador)



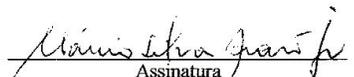
Assinatura

Profª. Dra. Maria Imaculada Zucchi



Assinatura

Prof. Dr. Márcio Silva Araújo



Assinatura

Profª. Dra. Selma Giorgio

Assinatura

Prof. Dr. Wesley Augusto Conde de Godoy

Assinatura

Agradecimentos

Aos Joões, Marias, Josés e Chicas que mesmo sem me conhecerem custearam meus estudos.

A_{GR}ADEÇO primeiramente à minha família que sempre apoiou as minhas escolhas dando condições para que eu continuasse estudando mesmo sem entender muito bem o que eu fazia. Pai, mãe e irmã, tios e tias e a todos outros muito obrigado por permitirem pelo ambiente familiar e, normalmente, amistoso no qual me desenvolvi. À minha segunda família (grupo e tropa escoteira Jr) que também me apoiou sempre, e que sentiu minha falta nestes últimos anos.

Aos amigos do Norte, do Sul, do Leste e do Oeste, daqui e de lá que infelizmente não terei espaço aqui para nomear, mas que me ajudaram (e continuam me ajudando) com vossas amizades, com vossos exemplos. Espero poder retribuir a todos vocês a confiança e amizade que vocês tem em mim, muito mais que em algumas linhas que se seguem. De qualquer modo preciso colocar alguns nomes aqui.

Aos responsáveis pela minha escolha profissional: João Batista Alves de Oliveira, técnico da biologia animal e juiz de futebol; e Francisco Carlos Motta, professor de ciências. Que durante meu ensino fundamental me mostraram quão bela e interessante é a biologia. Interesse esse que me leva até hoje a continuar nessa profissão. Nesse ponto não posso deixar de agradecer ao Bisa (Cláudio José Junior) que além da amizade reforçou o quão belo pode ser até mesmo um lofoforado.

Aos amigos com quem dividi o teto e se tornaram como irmãos, com quem aprendi a conviver com as diferenças e me acolheram tão bem. Tanto da graduação com no mestrado.

Aos amigos de trabalho de ontem e de hoje. profa Eneida que abriu as portas desse mundo para mim quando eu ainda nem era “bixo”. Ao Bispo de quem herdei meu primeiro projeto de IC e o gosto por *UNIX e \LaTeX com os quais escrevo este trabalho. À Marcela Benedetti que tanto me aturou e quem me ensinou o básico de genética molecular. E aos outros profissionais que tive o prazer de conviver em tantos lugares que passei: Biomembranas, Boldrini, Cipoi, DIFe, etc.

Ao José Louvise Gomes-Jr que tanto me ajudou sobre os desconhecidos “barrigudinhos”, e a Rute Beatriz Clemente Carvalho que tanto me guiou no laboratório. Sem vocês seria impossível concluir este trabalho.

Ao professor Sérgio que mesmo sem me conhecer confiou em mim dando condições para levar a diante um plano de estudos que foi se montando com o passar do tempo com simplicidade e paciência.

À Unicamp e ao IB que foram minha casa durante estes últimos 10 anos e que permitiram conhecer todos estas maravilhosas pessoas.

*— I was like a boy playing on
the sea-shore, and diverting
myself now and then finding a
smoother pebble or a prettier
shell than ordinary, whilst the
great ocean of truth lay all
undiscovered before me.
— Standing on the shoulders of
giants.*

Isaac Newton, 1642–1727

Resumo

Os poecilídeos são um excelente sistema modelo para estudos de evolução de história de vida seleção natural e sexual, evolução e coevolução experimental e evolução fenotípica em gradientes ecológicos. Os poecilídeos são também excelentes modelos para o estudo de processos ecológicos e evolutivos associados como a invasão e colonização de novos ambientes.

As populações de *Poecilia vivípara* que ocorrem no sistema lagunar de Campos de Goytacazes no norte do estado do Rio de Janeiro são um exemplo notável de invasão e colonização de novos ambientes. Nesse sistema, a origem das lagoas deve-se a processos geomorfológicos associados com a formação do delta do rio Paraíba do Sul durante o Holoceno.

Para este sistema formulamos a hipótese que a história geológica da região influenciou a variação genética atual em *Poecilia vivípara*. Deste modo, uma população de uma outra bacia hidrográfica seria diferente das populações da bacia do rio Paraíba do Sul. Uma população do rio Paraíba do Sul de uma região com formação mais antiga seria diferente das populações de regiões com história mais recente. Dentre as mais recentes as que vivem na área de influência marinha seriam diferentes das que vivem na área fluvial.

A hipótese foi testada usando sequências da região de controle da replicação mitocondrial de 8 populações com 30 indivíduos cada. Os resultados mostraram uma grande diversidade genética dentro e entre as populações, estruturação genética entre as populações antigas \times recentes do rio Paraíba do Sul e que não houve mudanças no tamanho efetivo das populações recentemente.

A partir dos resultados a hipótese formulada não foi refutada. Assim, existe relação entre a história geológica da região e a variação genética atual do *P. vivípara*. Há evidências que o regime de inundação característico da região também age sobre a variação genética destas populações aumentando o fluxo genético entre as mesmas.

Abstract

The fishes of Poeciliidae family are an excellent model system for studies of life history evolution of natural and sexual selection, experimental evolution and coevolution in phenotypic evolution and ecological gradients. These are also excellent models for the study of ecological and evolutionary processes associated as the invasion and colonization of new environments.

Populations of *Poecilia vivipara* occurring in the lagoon system of Goyta-cazes fields in northern Rio de Janeiro state are a notable example of invasion and colonization of new environments. In this system, the origin of the lakes due to geomorphological processes associated with the formation of the delta of the River Paraíba do Sul during the Holocene.

For this system we hypothesized that the geological history of the region influenced the genetic variation present in *Poecilia vivipara*. Thus, a population of another watershed would be different populations of river basin Paraíba do Sul A population of Paraíba do Sul River in a region with older formation would be different regions with populations of more recent history. Among the most recent ones that live in the area of marine influence would be different from those who live in the river.

The hypothesis was tested using sequences of the control region of mitochondrial replication of 8 populations with 30 individuals each. The results showed a high genetic diversity within and among populations, genetic structure among populations older *times* recent Paraíba do Sul river and that there were no changes in effective population size recently.

From the results the hypothesis was not refuted. Thus, there is a relationship between the geological history of the region and genetic variation of the current *P. vivipara*. There is evidence that the flooding regime characteristic of the region also acts on the genetic variation of these populations increasing gene flow between them.

Sumário

1	Introdução	1
1.1	Poecilídeos	1
1.1.1	<i>Poecilia vivipara</i>	1
1.2	Zonas litorâneas	3
1.2.1	Formação do delta do rio Paraíba do Sul	4
1.3	As populações de <i>Poecilia vivipara</i> do delta do rio Paraíba do Sul	8
2	Materiais e métodos	10
2.1	Amostragem das populações	10
2.2	Obtenção de dados	15
2.2.1	Extração do DNA	15
2.2.2	Padronização da reação de amplificação	15
2.2.3	Purificação dos fragmentos amplificados	16
2.2.4	Sequenciamento, edição e montagem das sequências	16
2.3	Análise de dados	17
2.3.1	Variação genética	17
2.3.2	Estruturação genética	18
2.3.3	Demografia histórica	21
3	Resultados	23
3.1	Padronização da reação de amplificação	23
3.2	Variação genética	24
3.3	Estruturação genética	28
3.3.1	Redes de haplótipos	28
3.3.2	Diferenciação genética	32
3.4	Demografia histórica	35

<i>SUMÁRIO</i>	ix
4 Discussão	38
5 Conclusão	41
A Função índices moleculares	49
A.1 Página de ajuda	49
A.2 Função mol.indice	51
B <i>Scripts</i>	54

Capítulo 1

Introdução

1.1 Poecilídeos

Os peixes vivíparos da família Poeciliidae são um grupo endêmico do Novo Mundo e compreendem cerca de 220 espécies incluídas em 28 gêneros [34] [35]. Os poecilídeos são pequenos não atingindo mais que 200 mm de comprimento sendo que a maioria tem menos da metade deste tamanho. Dentro dessa família estão os menores e mais polimórficos vertebrados viventes [55]. Estes apresentam uma grande tolerância a ambientes de água salgada e salobra e muitos deles são essencialmente de água doce [55]. Os poecilídeos exibem modificações sexuais secundárias. Nos machos a nadadeira anal sofre modificações para poder transferir esperma à genitália da fêmea (figura 1.1a). O tamanho da prole varia de poucos a mais de 300 indivíduos de cada vez [29]. A dieta também varia de invertebrados aquáticos e terrestres, detritos, plantas e principalmente larvas de insetos [55].

Por causa dessas características os poecilídeos são um excelente sistema modelo para estudos de evolução de história de vida [5], seleção natural e sexual [28], evolução e coevolução experimental [42][7] e evolução fenotípica em gradientes ecológicos [27] [11] [30]. Os poecilídeos são também excelentes modelos para o estudo de processos ecológicos e evolutivos associados com a invasão e colonização de novos ambientes [66].

1.1.1 *Poecilia vivipara*

Poecilia vivipara Bloch e Schneider 1801 é uma espécie de pequeno porte (2–5 cm de comprimento) (figuras 1.1a e 1.1b) amplamente distribuída por



(a) Macho



(b) Fêmea

Figura 1.1: Fotos de *P. vivipara* (a) macho adulto, (b) fêmea adulta — cortesia de José Louvise Gomes Junior.

toda a costa da América do Sul habitando o curso baixo de grandes rios [45]. Os fatores que influenciam a distribuição desta espécie não são conhecidos e, provavelmente, *P. vivipara* não seja apenas uma espécie mas um complexo de espécies [9]. Como qualquer outro da mesma família o *P. vivipara* é um bom modelo experimental sendo de fácil manutenção em laboratório. Por isso é utilizado como modelo experimental para estudo de parasitas [60] [58] [59] [59], ensaios sobre de toxicologia ambiental [37] sendo sugerido como uma ferramenta de monitoramento ambiental [46].

1.2 Zonas litorâneas

As zonas litorâneas em qualquer parte do mundo são ambientes heterogêneos pois possuem níveis diferentes de complexidade. Estes níveis de complexidade ocorrem devido à combinação de processos morfológicos, tectônicos, abrasivos e deposicionais que formaram cada setor da costa [1]. Assim, para entender a ordem e a importância desses processos na formação de um litoral qualquer é necessário considerar o espaço total costeiro que envolve sempre a faixa que se estende da linha da costa até a retro-terra costeira. Para Aziz Nacib Ab'Saber [1].

...Devido essa ampliação do espaço-objeto de estudo, estamos mais preparados para retrair a sequência dos fatos acontecidos na zona costeira ao longo do Quaternário. O que permite dizer que os litorais na sua aparente simplicidade paisagística e na sua dinâmica habitual exigem considerações similares ou mais complexas do que os espaços interiores, já que envolvem sérias questões relacionadas com as variações do nível do mar, paleoclimas e história vegetacional. Ou seja, o litoral, tal como outras áreas dotadas de paisagens ecológicas, pode ser considerado sempre como uma herança de processos remodelados pela dinâmica costeira hoje prevalecente.

Portanto, as zonas litorâneas são resultantes de processos que geraram uma grande quantidade de ambientes complexos e diferentes como praias arenosas, detritos calcáreos, manguezais, costões, grutas de abrasão, restingas isoladas ou múltiplas, lagunas e lagoas fragmentados por deltas intralagunares, deltas e barras de rios, campos de dunas, estuários, recifes, etc [1].

Os 7367 km de extensão do litoral brasileiro podem ser classificados geomorfologicamente em quatro grandes setores. A saber: Brasil Equatorial Atlântico (Amazônia Atlântica); Brasil Atlântico Semi-Árido (Costa Cearense/Potiguar); Brasil Tropical Atlântico Oriental (Costa dos recifes, barreiras e tabuleiros); Brasil Tropical Atlântico de Sudeste (Costa dos esporões da Serra do Mar); e, Brasil Subtropical Atlântico (Costa gaúcha/sul catarinense [1]. A mesma região é dividida em uma escala menor com 49 setores. Destes setores menores o setor do delta do rio Paraíba do Sul se destaca como a maior planície deltáica arqueada do litoral brasileiro de origem subatual,

com mudança da embocadura do rio para o setor norte da planície deltáica e com restingas e feixes de restingas envolvendo o setor sul deltaico [15].

O delta do rio Paraíba do Sul está dentro da região do litoral norte fluminense que corresponde a cerca de 180 km de litoral. Trata-se de uma região de baixada contendo várias lagoas e amplas áreas de várzea representando grandes espaços territoriais de áreas inundáveis. A zona estuarina do rio Paraíba do Sul onde se encontram manguezais bem como dos rios Itabapoana e Macaé. Os quatro municípios da região são São João da Barra, Campos, Quicamã e Macaé [15].

Os principais ecossistemas presentes são as baías e estuários seguidos das lagunas, dos mangues, praias e costões. Estes ecossistemas encontram-se altamente ameaçados devido às atividades econômicas como a agroindústria, indústria sucro-alcooleira, extração de petróleo, mineração e ocupação desordenada do solo. Estas atividades ameaçam os ecossistemas por causa da geração de resíduos industriais e de pesticidas usados na agricultura que são altamente poluidores para os recursos hídricos da região. Além disso, a exploração da bacia petrolífera de Campos é uma constante fonte de risco ambiental para a região [15]. Outra ameaça aos ecossistemas são os projetos de saneamento implantados a partir de 1933 que construíram canais de drenagem, irrigação e diques [8].

1.2.1 Formação do delta do rio Paraíba do Sul

O termo delta é bastante antigo sendo proposto por Heródoto 400 anos a.C. que usou a semelhança entre a letra grega maiúscula e planície aluvial situada entre os dois distributários principais do rio Nilo. Atualmente um delta é definido como uma protuberância na linha de costa nos locais onde os rios adentram os oceanos, mares e lagos. São constituídos por sedimentos transportados pelos rios que o alimentam [64].

O delta do rio Paraíba do Sul é resultado de dois processos concomitantes e intrinsecamente associados: a planície formada de aluviões transportados pelo rio Paraíba do Sul da zona cristalina, e a planície marinha resultante de movimentos oceânicos de avanço e retrocesso [36] [8].

Resumidamente, há mais de um milhão de anos o rio Paraíba do Sul desembocava em uma grande baía rasa de mar aberto na qual depositava os sedimentos resultantes do processo de erosão do seu leito. Na penúltima elevação do nível do mar há 123 mil anos atrás houve a erosão dos depósitos continentais anteriores e durante o rebaixamento do nível do mar que se se-

guiu foram construídos terraços marinhos arenosos e pequenas elevações do solo. O último avanço do mar (5100 anos antes do presente) (figura 1.2) erodiu os terraços marinhos formando sistemas lagunares. Estes perderam contato direto com o mar devido às ilhas-barreiras que se formaram e, com a desembocadura do rio nessas lagunas foram desenvolvidos deltas intralagunares. Com o subsequente abaixamento do nível relativo do mar ocorreu a transformação gradual das lagunas em lagoas de água doce e, em alguns casos, em pântanos. As variações no nível relativo do mar após 5100 anos antes do presente foram de pequena amplitude e curta duração [36] [8]. Todo este processo é responsável pela heterogeneidade presente nessa porção de litoral ao norte do estado do Rio de Janeiro e sul de Espírito Santo (figura 1.3a).

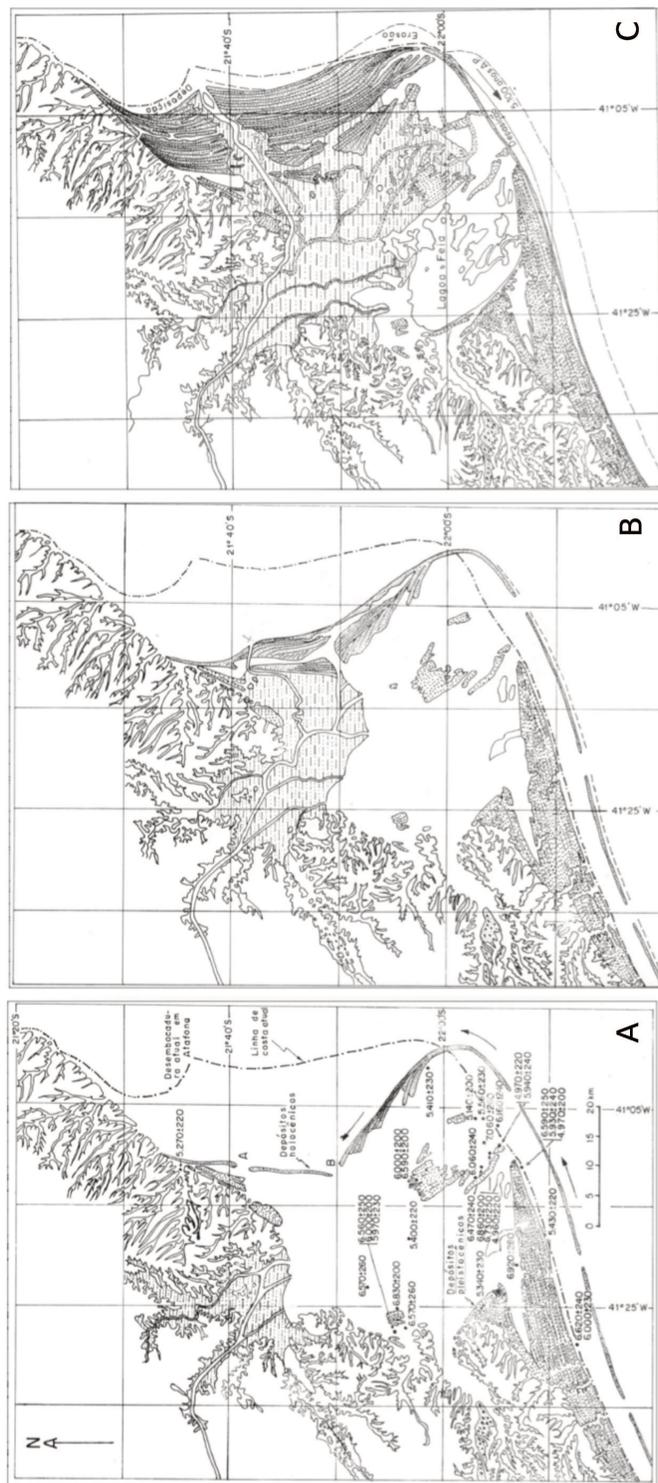
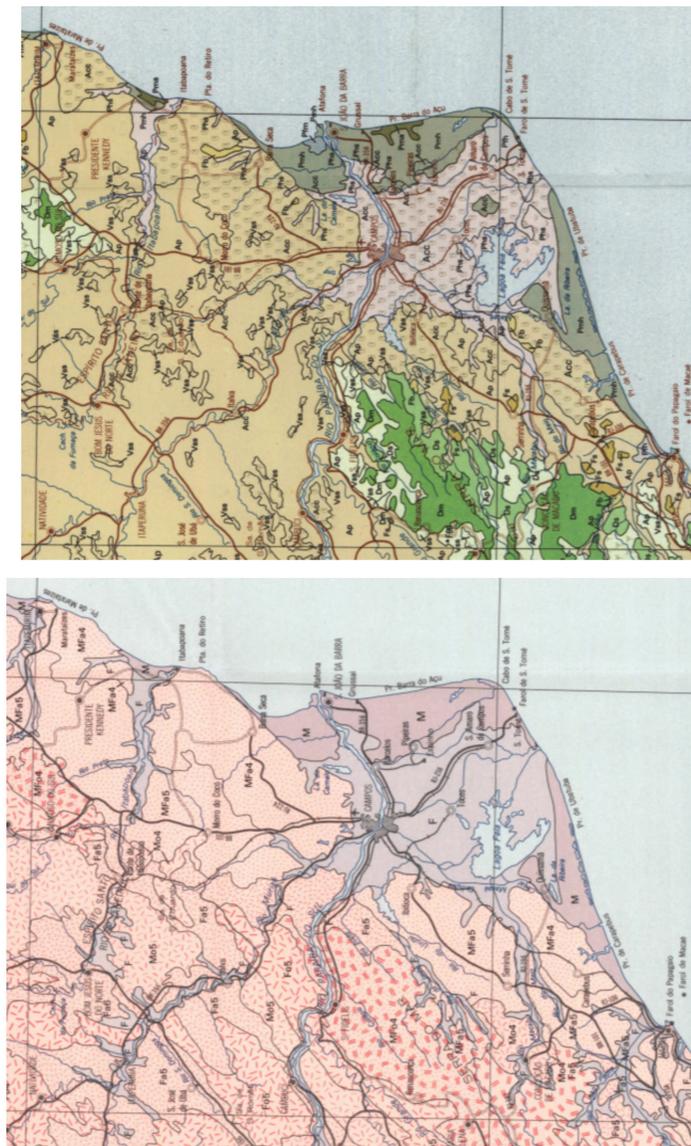


Figura 1.2: Formação do delta do Rio Paraíba do Sul. A— planície costeira aproximadamente 5100 anos A.P.; B— ilhas-barreiras delimitam uma laguna, C— planície costeira entre 5100 e 4000 anos A.P., C— planície costeira de 2500 anos até hoje. Fonte [36].



(a) relevo

(b) vegetação

Figura 1.3: Recortes dos mapas da região do estado Rio de Janeiro e sul de Espírito Santo: (a) relevo, Categoria de acumulação: M—marinha, F—fluvial. Fonte original; (b) vegetação, Cobertura vegetal: Pma— influência marinha (arbórea), Pmh— influência marinha (herbácea), Pfh— influência fluviomarina (herbácea), Phs— influência fluvial (herbácea sem palmeiras), Acc— agricultura culturas cíclicas. Fonte original [50].

1.3 As populações de *Poecilia vivipara* do delta do rio Paraíba do Sul

As populações de *Poecilia vivipara* que ocorrem no delta do Rio Paraíba do Sul e no sistema lagunar de Campos de Goytacazes no norte do estado do Rio de Janeiro [9] são um exemplo notável de invasão e colonização de novos ambientes. Estas populações mostram uma grande variação fenotípica [38] e também variações no tamanho populacional com crescimento em períodos de menor salinidade e maiores temperaturas [9].

Nesse sistema, a origem das lagoas deve-se a processos geomorfológicos associados com a formação do delta do rio Paraíba do Sul durante o Holoceno [36]. Essas lagoas diferem em área e distância do mar e têm gradientes de salinidade que influenciam marcadamente a composição da vegetação marginal e da comunidade de peixes. Sistemas como esses de origem Holocênica são particularmente relevantes pelo fato das mudanças ecológicas e evolutivas estarem intimamente associadas e poderem ocorrer nas mesmas escalas de tempo [27]. A colonização do complexo lagunar de Campos de Goytacazes por *P. vivipara* a partir de populações do Rio Paraíba do Sul resultou em um experimento natural em uma curta escala de tempo (< 6.000 anos; [36]), gerando uma rápida diversificação fenotípica associada a gradientes de salinidade e predação [38]. A evolução fenotípica observada em *P. vivipara*, particularmente em resposta ao gradiente de predação, é semelhante a de outros poecilídeos sob condições ecológicas equivalentes e parece ser evidência da ação de seleção natural [30].

Estudos de evolução fenotípica, particularmente em função de gradientes ambientais ecológicos, são importantes para o conhecimento da dinâmica ecológica do processo de colonização de novos ambientes. Todavia, informação molecular é fundamentalmente necessária para a compreensão da dinâmica do processo de colonização [33] [31].

Deste modo, é possível construir a hipótese de que os processos geomorfológicos que geraram os ambientes (neste caso as lagoas dentro do delta) influenciaram a diversidade abiótica e biótica hoje presente. Assim, a diversidade geológica hoje presente (e.g. as planícies de acumulação fluvial e marinha, gradiente de salinidade, etc) é resultado dos processos que a geraram. Do mesmo modo a diversidade biológica presente pode ser, de forma direta ou indireta, resultado dos mesmos processos geomorfológicos.

Para testar esta hipótese é necessário, primeiramente, descrever a variação

molecular dentro e entre as populações que ocupem ambientes que tenham passado por processos diferentes e por isso sejam diferentes hoje. Com a variação descrita é possível o segundo passo que é analisar a estruturação desta variação para num terceiro passo inferir quais foram os principais processos geradores da variação observada e compará-los com os processos que geraram a diversidade ambiental.

Capítulo 2

Materiais e métodos

Para testar a hipótese declarada anteriormente que a variação molecular em *P. vivipara* esta relacionada aos processos geomorfológicos foi necessário amostrar populações de lagoas geradas por processos diferentes. Destas populações amostradas extrair o material genético e descrever a sua variação. Se existisse variação procurar por padrões. Se suficientes estes padrões serviriam para inferir quais os processos geraram a variação molecular observada e relacioná-los com os processo geomorfológicos já descritos para a região.

2.1 Amostragem das populações

A amostragem das populações foi feita numa área que compreende as lagoas que se formaram com delta do rio Paraíba do Sul durante o Quaternário e outros dois pontos com histórias de formação diferentes do delta. No total oito populações foram amostradas em oito localidades diferentes; duas externas ao delta, e seis internas (figura 2.1).

Barra do Itabapoana não faz parte do delta do rio Paraíba do Sul. Situa-se na desembocadura do rio Itabapoana.

Lagoa de Cima tem origem no rio Paraíba do Sul porém de formação mais antiga que a planície quaternária. Situa-se a oeste da planície e recebe atualmente águas dos rios Imbé e Urubu e passa atualmente pelo processo de eutrofização.

Lagoa do Campelo tem origem a partir da planície quaternária. Situa-se na região de aluvial ao norte do rio Paraíba do Sul estando atualmente

ligada apenas a este. Esteve conectada à lagoas e rios mais ao norte. É uma das maiores e mais importantes entre os rios Paraíba do Sul e Itabapoana.

Lagoa Comércio tem origem a partir da planície quaternária. Situa-se na região de restinga ao norte do rio Paraíba do Sul em meio ao centro econômico de Gargaú. Está atualmente poluída e eutrofizada.

Lagoa Taí tem origem a partir da planície quaternária. Situa-se na região de restinga ao sul do rio Paraíba do Sul. Devido às obras de saneamento e prevenção de inundações hoje esta lagoa ser tornou um brejo.

Lagoa Grussaí tem origem a partir da planície quaternária. Situa-se na região de restinga ao sul do rio Paraíba do Sul. Recebia águas do rio Paraíba do Sul até a abertura do canal Quitingute que cortou a conexão e diminuiu o volume de água. Está atualmente assoreada, eutrofizada e poluída.

Lagoa Feia tem origem a partir da planície quaternária. Situa-se na região aluvial ao sul do rio Paraíba do Sul. É conectada com várias outras lagoas através de canais. É a maior lagoa da planície sendo uma das mais antigas, teve seu formato muito alterado no último século por obras de saneamento e prevenção de inundações.

Lagoa Açú tem origem a partir da planície quaternária. Situa-se na região de restinga ao sul do rio Paraíba do Sul. Também conhecida como rio Açú recebia água da lagoa Feia e do rio Paraíba do Sul, estas conexões foram cortadas pelas obras de saneamento e prevenção de inundações. Com o fim das conexões tornou-se uma lagoa com alto grau de salinidade.

Tabela 2.1: Características ambientais e históricas das localidades amostradas:

Localidade	Sigla	Idade (anos) ¹	Acumulação ²	Salinidade (ppm) ³	Coordenadas (Graus, decimais)	
					Latitude	Longitude
Barra do Itabapoana	BI	–	marinha	12	-21,33602	-40,97192
Lagoa de Cima	CI	10000	terraço fluvial	0	-21,78693	-41,51622
Lagoa do Campelo	CA	4000	fluvial	0	-21,6317	-41,191490
Lagoa Comércio	CO	1000	marinha	0	-21,57653	-41,06701
Lagoa Taí	TA	2000	marinha	0	-21,78748	-41,14363
Lagoa Grussaí	GR	1500	marinha	2,3	-21,71979	-41,03783
Lagoa Feia	FE	7000	fluviolacustre	0	-21,94813	-41,33949
Lagoa Açú	AC	3000	marinha	13	-21,99416	-41,00001

Fontes: 1—[36], 2—[50],3—[39].

Trinta indivíduos de *P. vivipara* de cada lagoa foram coletados. A coleta foi feita pelo Dr. José Louvise Gomes Jr., como parte de um projeto sobre a ecologia evolutiva de *P. vivipara* com uma rede de arrasto de malha fina (tela de mosquiteiro) seguida por anestesia com gelo e da secção do pedúnculo caudal que foi fixado em etanol absoluto em temperatura ambiente. No laboratório o material foi mantido a -20°C até a extração do DNA.



Figura 2.1: Mapa da áreas de estudo com a localização dos pontos de amostragem.

2.2 Obtenção de dados

2.2.1 Extração do DNA

O DNA genômico foi extraído utilizando o protocolo de extração com sal [3]: Uma amostra de tecido de aproximadamente 2 mm² de músculo do pedúnculo caudal de *P. vivipara* foi dissecada e incubada em 110 µl de tampão de lise (20 µl de SDS 10 %, 10 µl de proteinase K 20 mg/ml e 80 µl de TNE 1×) em *overnight* a 55 °C. Depois foi adicionada 5 µl RNase 10 mg/ml e incubado a temperatura ambiente por 10 minutos seguida da adição de 35 µl NaCl 5 M e centrifugação por 15 minutos a 14000 RPM.

O sobrenadante foi transferido para um novo tubo no qual foi misturado com 330 µl de etanol absoluto gelado e incubado a -20 °C em *overnight*. Depois centrifugado por 15 minutos a 14000 RPM descartando-se o sobrenadante. Em seguida foi adicionado 150 µl de etanol 70 % e novamente centrifugada por três minutos a 13000 RPM. O sobrenadante foi descartado e o tubo deixado secar em temperatura ambiente. Depois de seco, um volume apropriado (35–50 µl) de TE fraco foi adicionado para ressuspender o DNA e incubado a 55 °C por duas horas.

2.2.2 Padronização da reação de amplificação

Para a região de controle da replicação do DNA mitocondrial de *P. vivipara* estavam disponíveis na literatura três combinações diferentes de *primers* L15926 × MRT-2, L15926 × H16498, L15926 × H00651 [2]. A primeira e a segunda combinação foram desenhadas para o gênero *Poecilia* e a terceira é universal para vertebrados. Todas estas combinações foram testadas.

Para as duas primeiras combinações de *primers* as reações foram preparadas com: 2,5 µl de tampão 10×, 3,25 µl de dNTP 2,0 mM, 3 µl de MgCl₂ 25 mM, 0,63 µl de cada *primer* 10 mM, 0,25 µl de Taq polimerase 5 U/µl, 3 µl de DNA 10 ng/µl e água q.s.p. 25 µl. Para a terceira combinação de *primers* a reação foi preparada com: 2,5 µl de tampão 10×, 3,25 µl de dNTP 2,0 mM, 3,75 µl de BSA, 3 µl de MgCl₂ 25 mM, 0,63 µl de cada *primer* 10 mM, 0,25 µl de Taq polimerase 5 U/µl, 3 µl de DNA 10 ng/µl e água q.s.p. 25 µl.

Os programas das reações foram os mesmos sugeridos inicialmente pelos autores (94 °C 5 min, 31× (94 °C 1 min, 50 °C 30 s, 72 °C 1 min 30 s), 72 °C 7 min, 4 °C *forever*) “ciclo curto”₁++₂. Para as duas primeiras combinações de *primers* foi testado adicionalmente um programa com maior número de

ciclos e duas temperaturas de hibridação [24]: (96 °C 2 min, 53 °C 45 s, 72 °C 2 min, 39×(94 °C 30 s, 50 °C 45 s, 72 °C 2 min) 72 °C 10 min, 5 °C *forever*) “ciclo longo”.

2.2.3 Purificação dos fragmentos amplificados

Os produtos de PCR foram purificados utilizando o kit *Wizard Sv Gel and PCR Clean-up System* da *Promega Corporation*. Ao produto de PCR foi adicionado um volume igual de *Membrane Binding Solution*. Em seguida a solução foi transferida para a coluna e deixada em repouso por um minuto antes de ser centrifugada a 14000 RPM por um minuto sendo o eluído descartado. A coluna foi lavada com 700 µl de *Membrane Wash Solution* centrifugando a 14000 RPM por um minuto repetindo o processo com 500 µl por cinco minutos. Para secar a coluna uma centrifugação adicional foi realizada por um minuto na mesma velocidade. Para a eluição do DNA 35 µl de água estéril foi aplicado ao centro da coluna a qual foi deixada em repouso por um minuto antes de ser centrifugada por um minuto a 14000 RPM.

2.2.4 Sequenciamento, edição e montagem das sequências

O sequenciamento do DNA foi terceirizado sendo realizado pela Macrogen Inc. — Coreia do Sul. Os *primers* utilizados para sequenciamento foram os mesmos que os usados na amplificação.

A edição e montagem das sequências foi feita usando o programa *ChromasPro* versão 1.5 — Technelysium Pty Ltd, Tewantin QLD, Australia. A sequência de ida foi alinhada ao reverso complementar da sequência de volta através do algoritmo de alinhamento global. A edição manual das sequências foi feita com os pareamentos incorretos foram resolvidos analisando o cromatograma e a atribuição foi feita considerando o pico com melhor nota *Q* para aquela posição [17]. Por terem baixa qualidade, pares de bases do início e do final de cada sequência foram descartados (aproximadamente 80pb de cada sequência).

2.3 Análise de dados

Para descrever a variação genética dentro e entre as populações amostradas de *P. vivipara* foram utilizadas estatísticas de **Diversidade genética**. Para avaliar como essa variação é organizada foram utilizadas estatísticas de **Estruturação genética**. Para inferência de processos históricos demográficos foram utilizadas estatísticas de **Demografia histórica**. Abaixo estas estatísticas são explicadas de forma sucinta e como foram calculadas.

2.3.1 Variação genética

A variação genética pode ser estimada em vários níveis e, em cada nível por vários métodos [41]. A variação no nível do DNA pode ser medida comparando as sequências alinhadas de de diferentes indivíduos.

A partir do alinhamento é possível reconhecer as sequências idênticas entre os indivíduos. Por serem indivíduos de uma mesma população amostrada supõe-se que a identidade entre as sequências seja consequência do processo de herança. Assim, cada sequência diferente pode ser definida como um haplótipo diferente. Haplótipo é definido como um trecho de DNA de uma única molécula o qual é herdado como uma única unidade [57].

A variação dos haplótipos em uma população é normalmente estimada pela **diversidade haplotípica** (h) sendo calculada pela equação [40]:

$$h = 1 - \sum_{i=1}^m x_i^2 \quad (2.1)$$

onde x_i é frequência do haplótipo i e m é o número de haplótipos encontrados. Assim, h representa a probabilidade de dois alelos escolhidos ao acaso dentro de uma população serem diferentes sendo equivalente a heterozigosidade para locos haplóides.

A variação nos sítios entre os haplótipos pode ser estimada pelos **sítios segregantes** (S) que é definida como o número de sítios de uma amostra de sequências que estão ocupados por 2 ou mais nucleotídeos diferentes [26]. Porém, por depender do tamanho amostral S não é uma boa estatística para estimar a variação nos sítios [63]. Uma melhor estimativa dessa variação pode ser feita usando a **diversidade nucleotídica** (π) que é definida como número médio de diferenças nucleotídicas ou substituições por sítio de um grupo de sequências de DNA amostrados de uma população [41] que pode

ser calculado usando a equação:

$$\pi = \sum f_i f_j p_{ij} \quad (2.2)$$

onde f_i e f_j representam a frequência dos haplótipos i e j na população e p_{ij} representa a divergência entre eles. Normalmente este parâmetro é normalizado pelo comprimento da sequência, ou seja, a diversidade nucleotídica por sítio¹.

A identificação e contagem dos haplótipos foi feita usando o programa *Arlequin* versão 3.5 [18]. A diversidade haplotípica, os sítios segregantes e a diversidade nucleotídica foram calculados com o uso dos pacotes *Ape* [44] e *Pegas* [43] no ambiente de programação *R* [51].

2.3.2 Estruturação genética

A maioria das populações estão organizadas em pequenos grupos o que interfere na determinação dos parceiros sexuais. Muitos organismos naturalmente formam subpopulações como famílias, colônias, bandos, cardumes e outros tipos de agrupamentos o que gera uma estrutura dentro da população [26]. Portanto, é o modo pelo qual os parceiros sexuais se encontram que gera a estruturação em uma população.

A subdivisão em subpopulações separadas significa que a variação genética dentro de uma população existe em dois níveis primários: dentro das subpopulações e entre subpopulações [4]. Por isso algumas populações tem uma estrutura populacional hierárquica o que significa que as subpopulações podem ser agrupadas de forma progressiva de modo que níveis mais baixos são incorporados a níveis mais altos [26].

Para dados haplotípicos um modo de descrever a estrutura genética é utilizar **redes de haplótipos**. Outro modo mais geral é estimar a **diferenciação genética** dentro e entre as subpopulações e, se possível, verificar a existência de hierarquia entre o níveis.

Redes de haplótipos

A filogenia intraespecífica é uma poderosa ferramenta para testar uma variedade de hipóteses em diversas áreas: filogeografia, evolução molecular,

¹Na literatura os símbolos utilizados para a diversidade nucleotídica com e sem correção são os mesmos. Para este trabalho usaremos π para simbolizar apenas a diversidade nucleotídica por sítio.

sistemática biológica, taxonomia, entre outras [49] [57]. Esta filogenia pode ser reconstruída por vários métodos diferentes.

Os métodos de reconstrução filogenética tradicionais foram desenvolvidos para estimar relações entre espécies diferentes (interespecífica). Por isso, estes métodos têm algumas de suas premissas violadas quando aplicados a dados intraespecíficos: baixa variação entre os indivíduos como consequência da pouca idade da população, os estados ancestral e derivado estão presentes na população e, a possibilidade da existência de multifurcações [49] [57].

Nos últimos anos foram desenvolvidos vários métodos de reconstrução filogenética adaptados para dados intraespecíficos. Destes, os três mais comuns são: “parcimônia estatística” [65] [13], “rede de junção por medianas” como implementado no programa *Network* [6] e o “rede geradora mínima” como implementado no programa *Arlequin* [18]. Todos estes métodos de reconstrução resultam em genealogias que são chamadas de redes pois podem conter reticulações. Relações reticuladas em uma população podem ser geradas por processos evolutivos que agem no nível da população como a recombinação entre genes, a hibridação entre linhagens e homoplasia [49]. Todos estes processos não são possíveis para dados haplotípicos.

Recentemente o desempenho dos métodos de reconstrução filogenética com dados haplótipos foi avaliado [57]. Apesar de ter algumas premissas violadas os métodos tradicionais de reconstrução têm um melhor desempenho que os métodos adaptados a filogenia intraespecífica. O melhor desempenho ocorre principalmente quando há migração entre as populações [57]. Entre os métodos tradicionais o de parcimônia tem o melhor desempenho seguido por verossimilhança máxima e junção de vizinhos. Entre os métodos adaptados o de parcimônia estatística tem o melhor desempenho seguido por rede de junção por medianas e a rede geradora mínima [57].

A genealogia entre os haplótipos foi reconstruída pelos métodos tradicionais: parcimônia implementado no *DNAPARS* [20], verossimilhança máxima implementado no programa *DNAML* [20] e pelo método de junção de vizinhos implementado no programa *Neighbor* [20] e foram visualizadas no programa *Haplotype Viewer*². E também reconstruída por parcimônia estatística implementado no programa *TCS* [13]. As representações finais das genealogias foram desenhadas no programa *Inkscape*.

²Distribuído pelo autor em <http://www.cibiv.at/~greg/haploviewer>

Diferenciação genética

Quando há estrutura em uma população há inevitavelmente alguma diferenciação genética entre as subpopulações [21]. Assim é possível avaliar a estrutura de toda a população medindo a diferenciação entre as subpopulações. Neste contexto, diferenciação genética é quanto as frequências dos alelos (ou haplótipos) se tornam diferentes entre as subpopulações [26].

A diferenciação genética é tradicionalmente estudada medindo os desvios das frequências alélicas (ou haplotípicas) observadas daquelas esperadas num sistema panmítico. Muitos procedimentos de estimação de estrutura relacionados com a estatística F de Wright foram propostos para o tratamento de sistemas polimórficos [19]. Wright propôs o F_{ST} como medida de diferenciação entre subpopulações [69]. Uma abordagem usual é calcular o F_{ST} pareado, ou seja, de duas em duas subpopulações e assim verificar a diferenciação entre cada par de subpopulação. Uma limitação dessa medida é que ela não considera a identidade entre os alelos (ou haplótipos) e com isso há um enviesamento pois não é considerado o grau de relação genealógica entre os mesmos [4].

Uma medida que considera a identidade entre os alelos (ou haplótipos) é a ϕ_{ST} que é estimado pela análise de variância molecular (AMOVA). Por este método é possível testar possíveis relações hierárquicas entre as subpopulações usando qualquer tipo de informação *a priori* como hipóteses de hierarquia [19].

A estruturação de uma população pode ocorrer devido ao isolamento geográfico entre as subpopulações. Este isolamento pode ocorrer devido a barreiras naturais que dificultem a trânsito de indivíduos entre as subpopulações como rios, vales, montanhas, etc. Um outro impedimento ao trânsito de indivíduos é a distância geográfica entre as subpopulações.

A estimação das distâncias genéticas entre as subpopulações foi feita usando o programa *Arlequin* versão 3.5 [18]. A AMOVA foi realizada usando o pacote *Ade4* [12] e as distâncias geográficas usando o pacote *gstudio* [16] no ambiente de programação *R* [51]. As correlações entre as distâncias foram estimadas no mesmo ambiente de programação.

2.3.3 Demografia histórica

A variação genética atual pode ser resultado tanto de fatores contemporâneos como históricos. Muitos estudos examinam o nível de diferenciação entre populações para determinar quanto o fluxo genético está sendo limitado. No entanto, a diferença genética entre populações atuais pode ser resultado de um fluxo genético historicamente restrito ao invés da fragmentação recente do habitat [14] e, da mesma forma a falta de diferença pode ser gerada pela histórica conectividade ou recente ancestralidade ao invés do atual fluxo genético [10][68]. Ademais, alguns estudos enfatizam a importância da perspectiva histórica na análise da variação genética e na estrutura populacional [53].

Dentre os processos históricos que podem influenciar a variação genética atual estão os eventos demográficos que ajam sobre o tamanho populacional. Mudanças no tamanho da população podem deixar um sinal particular que pode ser eventualmente detectado na sequência de DNA. Um evento demográfico recente como um crescimento populacional, causa que a maioria dos eventos de coalescência ocorram antes do crescimento e, conseqüentemente, amostras dessas populações têm genealogias de genes esticadas perto dos nós externos e comprimidas perto da raiz (genealogias estrelas) [52].

Uma primeira abordagem para detectar estes sinais é correlacionar a diversidade nucleotídica com a diversidade haplotípica: alta diversidade nucleotídica e baixa haplotípica é provável uma divergência entre as subpopulações geograficamente distintas, baixas diversidades nucleotídica e haplotípica sugerem um gargalo recente ou efeito fundador, altas diversidades nucleotídica e haplotípica sugerem populações grandes e estáveis por longo período e, uma baixa diversidade nucleotídica e alta haplotípica sugere um gargalo seguido de crescimento [26] [25]. Uma segunda abordagem é utilizar testes estatísticos para detectar o crescimento populacional. Estes testes podem ser classificados com base no tipo de informação usada [52]:

Classe I usam a informação da frequência de mutações (sítios segregantes).

Essas estatísticas são mais apropriadas para distinguir entre a hipótese de crescimento populacional e a hipótese de tamanho constante pois a primeira gera um excesso de mutação nos ramos externos da genealogia e então um excesso de haplótipos únicos [63] [62] [61].

Classe II usa informação a partir da distribuição dos haplótipos. É uma das estatísticas com melhor desempenho na detecção de crescimento

populacional[52].

Classe III usam informação a partir das diferenças entre sequências pareadas. São mais conservadoras que as outras [52].

Os testes da classe I e II foram originalmente desenvolvidos para testar desvios da neutralidade (exceção do R_2 [52]). A grande quantidade de testes nestas classes se justifica pois é improvável que apenas um teste estatístico seja suficiente para detectar todos os tipos de forças evolutivas que podem influenciar o padrão de polimorfismo em uma população [22]. Portanto, é necessário comparar os resultados de vários testes para poder distinguir entre os processos evolutivos que geraram o padrão observado [56].

Para este trabalho foram utilizados os testes D de Tajima [63], D^* , F^* [23], R_2 [52] da classe I, o F_s de Fu [22] da classe II e análise da distribuição das dissimilaridades [61] [54] da classe III implementados no programa DnaSP v5 [32].

Capítulo 3

Resultados

3.1 Padronização da reação de amplificação

As combinações de *primers* 1 e 2 resultaram em fragmentos de aproximadamente 900 e 400 pb, respectivamente, sem amplificação inespecífica. Para estas combinações o programa com maior número de ciclos e duas temperaturas de hibridação resultou numa concentração maior de produto (figura 3.1a). A combinação 3 resultou na amplificação de fragmentos de diversos tamanhos (figura 3.1c) dificultando o isolamento da região de controle.

Assim, para obtenção do maior fragmento mais concentrado as condições ótimas para a reação de amplificação foram estabelecidas: combinação de *primers* 1 com o programa com maior número de ciclos e duas temperaturas de hibridação gerando fragmentos de 900pb. Para atender as exigências de concentração para a reação de sequenciamento foi necessário amplificar cada indivíduo em duplicata, estas foram misturadas por indivíduo antes de serem purificadas.

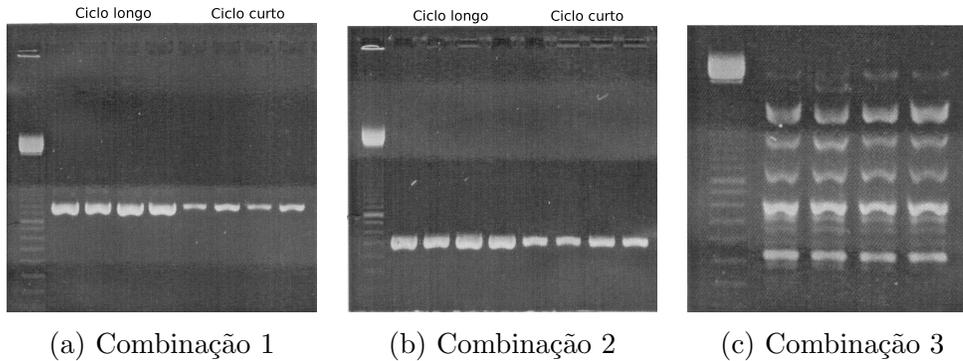


Figura 3.1: Padronização das reações de amplificação realizada com quatro indivíduos diferentes da mesma população em cada condição: *primers* (a) combinação 1 (ciclo longo e curto), (b) combinação 2 (ciclo longo e curto), (c) combinação 3.

3.2 Variação genética

Sequências de 911pb de comprimento da região de controle de replicação mitocondrial de 233 indivíduos de *P. vivipara* foram obtidas (tabela 3.2). O alinhamento destas sequências revelou 38 sítios polimórficos com 39 substituições definindo 32 haplótipos. Nenhuma inserção ou deleção foi encontrada. A composição de nucleotídeos dessas sequências foi rica em A-T (A : 33,12%, T : 32,92%, G : 12,54%, C : 21,42%) e as mutações consistiram predominantemente em transições (transições/transversões = 18,5).

O haplótipo mais abundante foi compartilhado por 62 indivíduos de 6 localidades (exceto BI e CI) e o segundo mais abundante foi compartilhado por 61 indivíduos de 6 localidades (exceto BI e CO) (figura 3.2). Os haplótipos menos abundantes foram aqueles presentes em apenas um indivíduo. No total 14 indivíduos de todas as populações amostradas apresentaram haplótipos não compartilhados (tabela 3.1). As sequências desses haplótipos estão disponíveis no *NCBI sequence database* sob os números de acesso JX415210-JX415241.

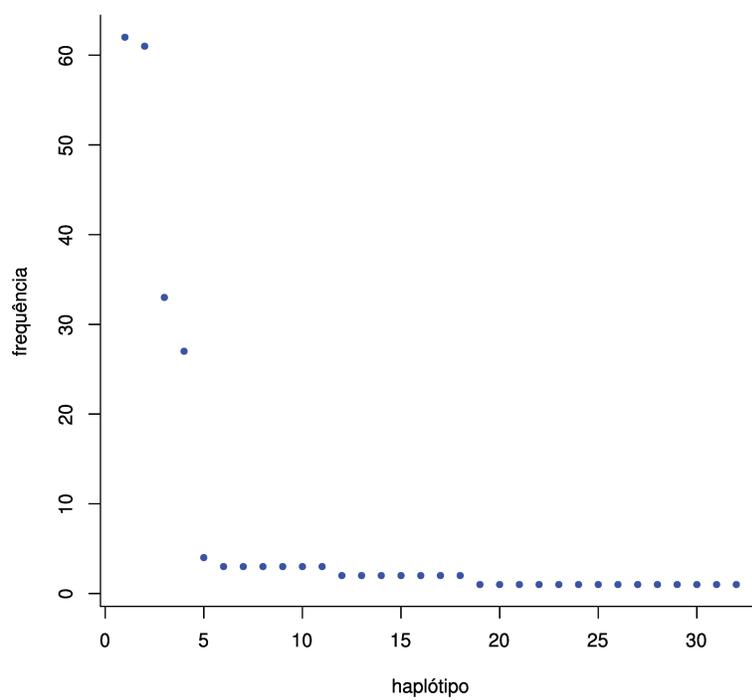


Figura 3.2: Frequência global dos haplótipos das oito populações amostradas de *P.vivipara*.

Altos níveis de diversidade haplotípica e nucleotídica foram detectados dentro de cada população amostrada (exceto BI), o que demonstra um alto nível de diversidade genética nas populações amostradas de *P. vivipara* (tabela 3.2). A diversidade haplotípica variou de $0,0741 \pm 0,0652$ em BI a $0,8253 \pm 0,0472$ em FE e a diversidade nucleotídica de $0,055\% \pm 0,053$ em BI a $0,851\% \pm 0,455$ em AC. Para todas as populações combinadas a diversidade haplotípica e nucleotídica foram $0,8287 \pm 0,0137$ e $0,828\% \pm 0,429$; respectivamente.

Tabela 3.2: Índices de diversidade molecular das oito populações amostradas de *P. vivipara*.

ID	Localidade	n	nh	h	π (%)
BI	Bara do Itabapoama	28	2	0,0741(0,0652)	0,055(0,053)
CI	Cima	30	7	0,5517(0,1044)	0,117(0,088)
CA	Campelo	30	8	0,7471(0,0603)	0,346(0,205)
CO	Comércio	29	5	0,5936(0,0646)	0,765(0,412)
TA	Tai	29	7	0,6700(0,0849)	0,778(0,418)
GR	Grussai	29	8	0,7906(0,0408)	0,784(0,421)
FE	Feia	30	10	0,8253(0,0472)	0,832(0,444)
AC	Açu	28	10	0,8148(0,0529)	0,851(0,455)
	<i>Total</i>	233	32	0,8287(0,0137)	0,828(0,429)

Identificação, localidade, tamanho amostral(n), número de haplótipos(nh), diversidade haplotípica(h) e nucleotídica(π), desvio-padrão entre parenteses.

3.3 Estruturação genética

3.3.1 Redes de haplótipos

A relação entre os diferentes haplótipos foi obtida com a construção de redes haplotípicas. As redes foram construídas utilizando diferentes métodos: parcimônia estatística, parcimônia, verossimilhança máxima e junção de vizinhos (figura 3.3). Todos os métodos empregados resultaram em redes com características em comum: i— Não disjunção, todos os haplótipos estão ligados a uma só rede; ii— Agrupamento, presença de 2 grupos de haplótipos; iii— Longo ramo interno, um grande número de passos em um mesmo ramo interno.

O método de parcimônia estatística resultou numa rede com muitas reticulações. Estas podem ser interpretadas de modos diferentes: como a presença de topologias ótimas conflitantes o que evidencia uma falha do método em decidir entre as alternativas de conexões entre os haplótipos ou como eventos de recombinação representando um cenário evolutivo real. No entanto, não existe um modo estatístico de lidar com as reticulações, ou seja, de afirmar se as reticulações presentes numa rede são um problema analítico ou um evento real [57].

O método de parcimônia resultou em 64 árvores com a mesma pontuação. Como o número de árvores é muito grande e não existe um método de escolha entre estas optou-se resumir todas as topologias usando o método de árvore *sensu stricto*. Por ser mais conservadora a árvore *sensu stricto* normalmente exhibe mais multifurcações que são interpretadas como falta de resolução. Assim a rede resultante tem mais multifurcações que as outras resultantes de outros métodos de construção de redes, distorcendo as relações entre os haplótipos.

O método de verossimilhança máxima resultou em uma única rede sem reticulações. Porém, do mesmo modo que o método de parcimônia, este método envolve busca no espaço de árvores sendo necessária uma avaliação criteriosa desta busca pois existe a possibilidade da topologia resultante ser a que possui a verossimilhança máxima apenas localmente.

O método de junção de vizinhos gerou em uma única rede sem reticulações. Por não ter reticulações, não envolver busca e possuir uma topologia semelhante as demais consideramos a rede gerada por este método a mais próxima da verdadeira. A partir desta rede foram desenhadas redes separadas por população (figura 3.4). Assim é possível analisar a estrutura

dos haplótipos dentro de cada população.

A rede de BI apresenta um longo ramo interno mas com apenas dois haplótipos ligados. A rede de CI apresenta haplótipos muito próximos sem o longo ramo interno e com formato semelhante a uma estrela. As demais redes são de populações que ocorrem na planície quaternária e apresentam um topologia muito semelhante entre elas e com a topologia da rede com todas as populações. A semelhança ocorre pois todas têm haplótipos nos dois grupos ligados por um longo ramo interno. Os haplótipos que estão do lado direito das redes são exclusivos de populações da planície quaternária. As redes das populações mais ao sul (FE e AC) apresentam uma distribuição mais simétrica de haplótipos ente os dois grupos e há uma tendência a perder esta simetria nas redes mais ao norte.

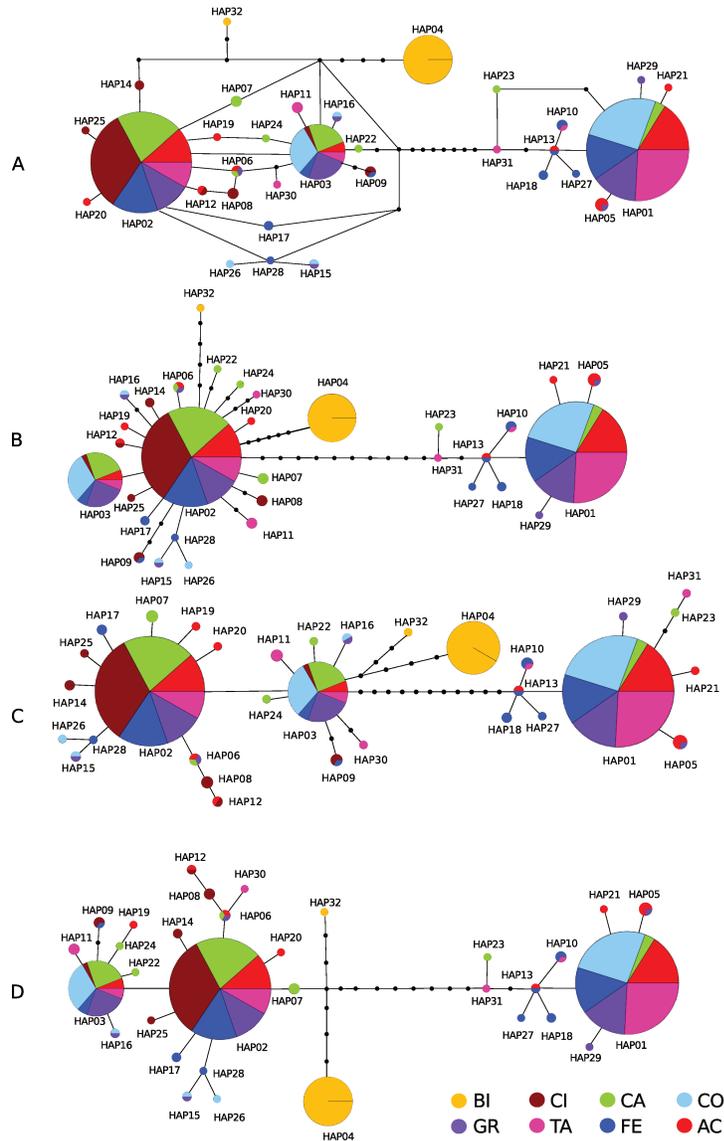


Figura 3.3: Redes de haplótipos de *P. vivipara* construídas por vários métodos: A- parcimônia estatística, B- parcimônia *sestrictu* $\uparrow+\uparrow\downarrow$, C- verossimilhança máxima, D- junção de vizinhos. O tamanho do círculo representa a frequência do haplótipo.

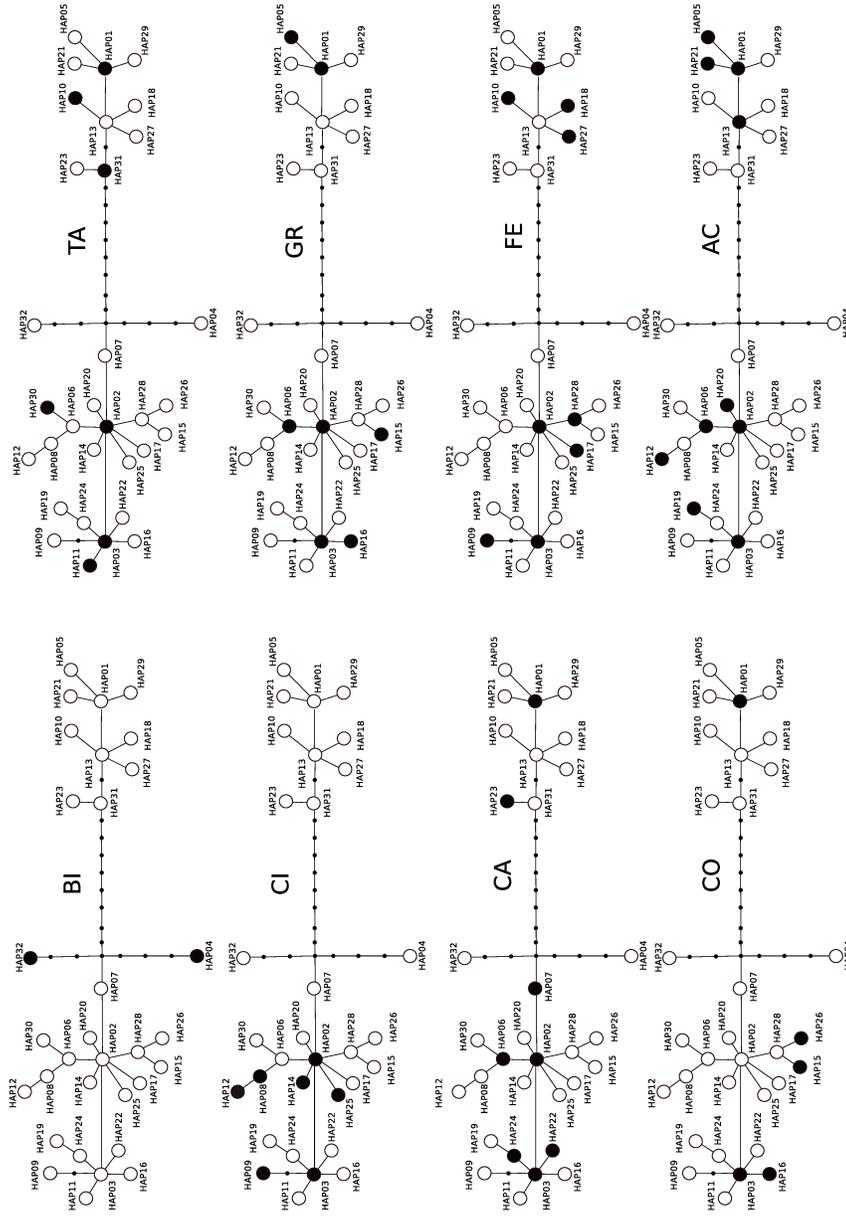


Figura 3.4: Redes de haplótipos de *P. vivipara* separadas por população amostrada geradas por junção de vizinhos. O tamanho do círculo não representa a frequência do haplótipo. Em preto os haplótipos presentes na população.

3.3.2 Diferenciação genética

As divergências genéticas entre as oito populações amostradas foram estimadas (tabela 3.3). Estas divergências foram menores para o mesmo conjunto de populações amostradas: CO, GR, TA, FE, AC. Isto é consistente com a árvore construída pelo método de junção de vizinhos que mostrou três grupos diferentes de populações (figura 3.5). Estes três grupos recuperam parcialmente a estrutura espacial prevista com base nos processos geomorfológicos (figura 1.2).

Tabela 3.3: Divergências entre as populações amostradas de *P. vivipara*: F_{ST} pareado.

	BI	CI	CA	CO	TA	GR	FE	AC
BI	–	–	–	–	–	–	–	–
CI	0,682*	–	–	–	–	–	–	–
CA	0,583*	0,075	–	–	–	–	–	–
CO	0,664*	0,420*	0,230*	–	–	–	–	–
TA	0,625*	0,308*	0,185*	0,059	–	–	–	–
GR	0,565*	0,191*	0,038	0,053	0,049	–	–	–
FE	0,543*	0,134*	0,055	0,125	0,036	0,009	–	–
AC	0,556*	0,177*	0,078	0,095	0,016	0,001	-0,010	–

* valor significativo ($\alpha = 0,05$) com a correção sequencial de Bonferroni ($p < 0,002$).

Para melhor avaliar a estrutura das populações foi realizada uma AMOVA hierárquica. Os dois níveis hierárquicos utilizados foram populações dentro de grupos. Os grupos foram definidos segundo as divergências genéticas como mostrado na figura 3.5: o primeiro grupo foi o conjunto das cinco populações amostradas com a menor divergência entre elas (CO, GR, TA, FE, AC); o segundo pelo conjunto das outras duas populações com a menor divergência entre elas (CI, CA) e o último por apenas BI. A avaliação mostrou que a maior parte da variação molecular está dentro das populações (65,49%) e do restante 4,17% esta entre populações dentro de um mesmo grupo e 30,34% entre os grupos (tabela 3.4).

Para testar se a distância geográfica influencia a diversidade genética foi estimada a relação entre as distâncias genéticas e geográficas das populações. A relação foi marginalmente significativa (teste de mantel, $p = 0,0577$) para

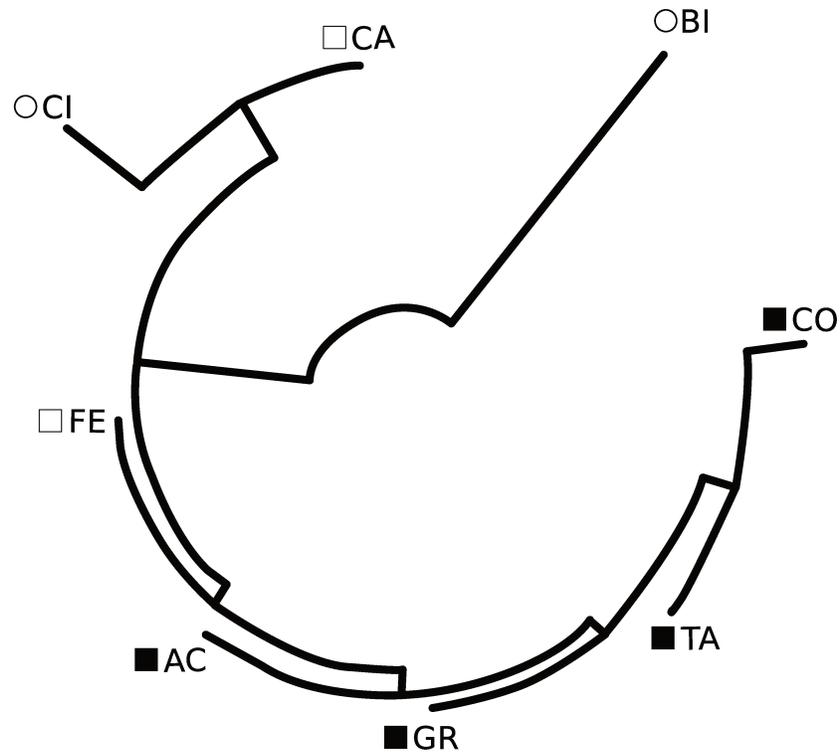
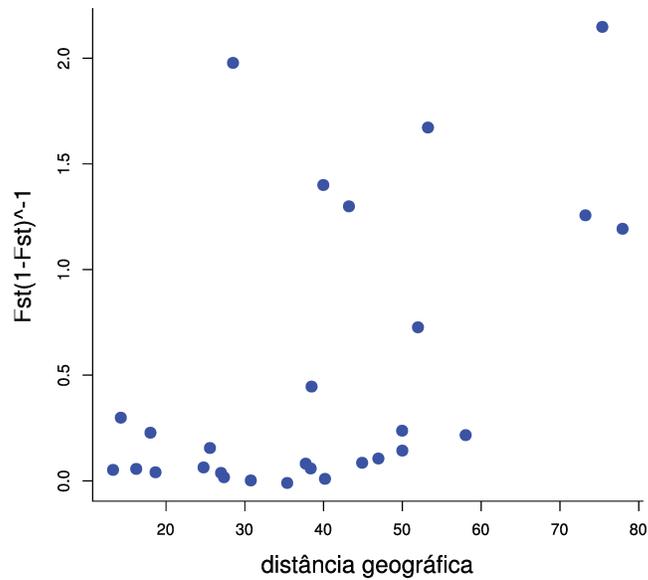


Figura 3.5: Árvore construída pelo método de junção de vizinhos a partir do F_{ST} pareado. Círculo, não faz parte da planície quaternária; Quadrado vazio, esta na planície quaternária sob influência fluvial; Quadrado cheio, esta na planície quaternária sob influência marítima.

o conjunto de todas populações. Assim, a distância geográfica não tem influência sobre a distância genética.

Tabela 3.4: Análise de variância molecular hierárquica da estrutura genética das oito populações amostradas de *P. vivipara*.

Fonte de variação	g.l.	Soma dos quadrados	Componentes de variância	% de variância
Entre grupos	2	40.103	0,294	30,34*
Entre populações dentro dos grupos	5	9.100	0,040	4,17*
Dentro das populações	225	143,04	0,635	65,49*
Total	232	192,24	0,828	100
		$\phi_{ST} = 0,345$	$\phi_{CT} = 0,0599$	$\phi_{SC} = 0,303$

* $p < 0,005$ Figura 3.6: Relação entre as distâncias geográficas e genéticas das oito populações amostradas de *P. vivipara*.

3.4 Demografia histórica

Numa primeira abordagem a correlação entre as diversidades nucleotídica e haplotípica das oito populações amostradas foi realizada (figura 3.7). A correlação para BI indica que esta população passou por uma redução no tamanho populacional recente (gargalo) ou efeito fundador. As correlações para CI e CA indicam que estas populações passaram por um gargalo recente seguido de um rápido crescimento com acúmulo de novas mutações. Para as demais populações as correlações indicam que estas permaneceram grandes e estáveis por um longo período. Notar que mais uma vez as populações da planície quaternária têm o mesmo comportamento com exceção apenas de CA.

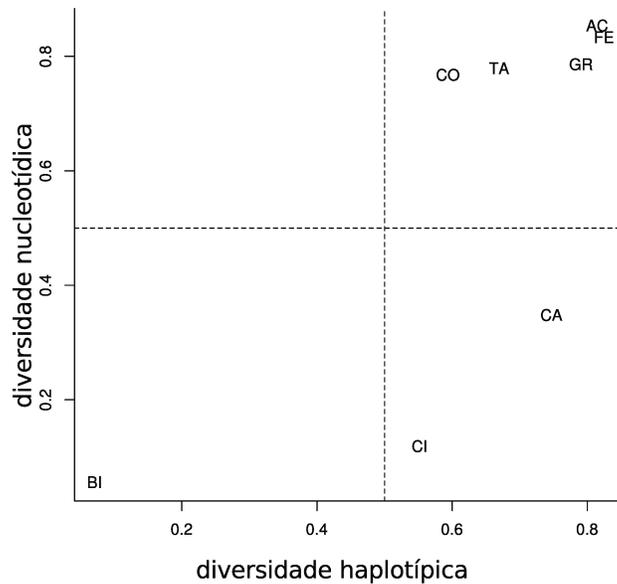


Figura 3.7: Correlação entre a diversidade nucleotídica e haplotípica das oito populações amostradas de *P. vivipara*. As linhas tracejadas são referências que dividem os eixos em alto e baixo para peixes marinhos [25].

Numa segunda abordagem foram realizados os testes de demografia para cada população isolada e para as populações combinadas segundo a hierar-

quia sugerida pela análise de estrutura. A maioria dos testes de neutralidade (classe I e classe II) não foram significativamente diferente de zero o que sugere o ajuste dos dados ao modelo de evolução neutra (tabela 3.5). Ou seja, o padrão da variação genética observada na região controle da replicação mitocondrial das populações de *P. vivipara* amostradas é resultado da mutação e deriva que agem sobre as populações.

A população BI apresentou desvios significativos da neutralidade, este perfil de desvio é um indicativo de seleção de “background”. A população CI apresentou desvios marginalmente significativos a neutralidade, este perfil é indicativo de crescimento populacional que não é corroborado pelos testes de classe III. Nenhuma outra população isolada apresentou desvios significativos a neutralidade. A combinação das populações CI e CA se comportou de modo semelhante às populações separadas com exceção para o teste R_2 que foi significativo para este conjunto o que sugere crescimento populacional para esta combinação de populações contrariando os resultados dos outros testes. A combinação das populações CO, TA, GR, FE, AC se comportou do mesmo modo que as populações separadas.

Tabela 3.5: Valores dos testes de demografia para as oito populações amostradas de *P. vivipara* separadas e agrupadas. Comparação com a expectativa no caso de crescimento.

Pop.	Neutrality test					MMD		
	D-tajima	D*	F*	R ₂	Fs	r	θ	τ
BI	-2,17	-3,52*	-3,63*	0,17	1,60	0,87	1,66	0,00
CI	-1,17	0,58	0,06	0,08+	-2,52+	0,06	0,64	0,42
CA	-0,75	0,44	0,08	0,09	0,24	0,10	4,41	0,00
CO	2,11	0,52	1,20	0,20	7,73	0,40	5,90	1,08
TA	1,90	0,59	1,17	0,19	4,73	0,25	6,01	1,08
GR	1,66	0,00	0,62	0,18	3,60	0,10	5,97	1,17
FE	2,01	0,64	1,26	0,20	2,15	0,13	5,73	1,86
AC	1,80	0,10	0,75	0,19	1,97	0,09	6,27	1,49
(CI, CA)	0,00	0,59	-0,21	0,051*	0,14	0,04	3,48	0,00
(CO, TA, GR, FE,AC)	0,00	-1,66	0,68	0,12	0,05	0,08	5,19	1,40
cresc.	< 0	ns	ns	baixo	*	-	-	-

* $p < 0,05$; + $p = 0,052$; ns = não significativa.

Capítulo 4

Discussão

Nosso principal objetivo foi testar a hipótese que os processos que geraram os diferentes ambientes influenciaram a diversidade abiótica e biótica hoje presente. Para tanto, nós descrevemos e analisamos a variação molecular da região de controle da replicação mitocondrial em populações de *Poecilia vivipara* do norte do estado do Rio de Janeiro. E assim, mostramos que a variação molecular presente nas populações atuais está relacionada com os processos históricos e espaciais que ocorreram naquele sistema. Esta relação é indicada pelos resultados que, de um modo geral, mostra que as populações de lagoas que tiveram a mesma história de formação possuem uma diversidade semelhante.

A diversidade molecular da região de controle da replicação mitocondrial de *P. vivipara* foi alta (exceto em BI). Parte dessa variação é possivelmente anterior ao início da formação das lagoas da região do delta do rio Paraíba do Sul e, assim, pertence a uma população ancestral. É improvável que toda a variação hoje observada tenha sido gerada num espaço de tempo tão curto. Ou seja, a variação molecular observada decorre, ao menos em parte, da divisão aleatória da variação molecular pré-existente — deriva. Um resultado que corrobora esta explicação é a rede de haplótipos, a ancestralidade de um haplótipo pode ser inferida por sua frequência nas populações amostradas. O haplótipo 1 e o haplótipo 2 são os mais frequentes com 62 e 61, respectivamente. Assim podem ser considerados os mais antigos entre os haplótipos amostrados. Enquanto o haplótipo 2 e seus descendentes diretos estão presentes em todas as populações (exceto BI) o haplótipo 1 e seus descendentes diretos estão restritos às lagoas da planície quaternária.

A deriva ocorreu devido ao processo de formação do delta do rio Paraíba

do Sul que aos poucos limitou o fluxo genético com a construção de barreiras. A limitação do fluxo imposta pelas barreiras foi diminuída pela migração entre as diferentes lagoas. A migração ocorreu principalmente pelas várias comunicações permanentes entre as lagoas existentes até a metade do século XX mas também pelas inundações constantes na região.

O processo de formação do delta gerou um ambiente heterogêneo num pequeno espaço. Esta heterogeneidade é perceptível nos solos, no relevo e na salinidade das lagoas formando gradientes ambientais que condicionaram a ocupação do espaço pelos seres vivos. Este condicionamento direcionou a distribuição dos tipos de vegetação (figura 1.3b) e também o início da ocupação humana [8]. O mesmo condicionou a diversidade morfológica de *P. vivipara* [38].

O tamanho de uma população influencia a variação que a mesma pode apresentar. Populações grandes suportam uma grande variação e populações pequenas não. Alterações no tamanho populacional também modificam a diversidade atual. As populações da planície quaternária mostraram uma mesma tendência: todas permaneceram grandes e estáveis por um longo período de tempo, uma exceção parece ser CA que pode ter passado por um gargalo seguido por crescimento. CI passou por um crescimento populacional possivelmente após um gargalo. A população BI provavelmente foi colonizada duas vezes por poucos indivíduos em épocas diferentes sendo o haplótipo mais frequente vindo na primeira colonização dominando o habitat antes da chegada do segundo haplótipo.

A ocupação humana da região e a degradação ambiental gerada pelas atividades econômicas provavelmente não influenciaram a variação molecular encontrada em *P. vivipara*. Esta ocupação iniciou-se com o povoamento indígena por povos Puris e Goytacás mas foi apenas a partir do século XVIII que esta ocupação ocorreu de forma mais impactante sendo apenas no último século o impacto da industrialização e do uso de pesticidas [8]. Estas alterações ambientais seriam extremamente recentes para alterar o padrão de variação descrito para um loco como a região de controle. Ademais, o *P. vivipara* é muito resistente a condições ambientais adversas, resistindo a mudanças de temperatura, salinidade e também a toxicidade.

Entretanto, alterações na estrutura populacional por ação humana num curto intervalo de tempo e numa pequena escala já foi descrita. *Poecilia mexicana* que ocorre em cavernas do sul do México tem populações com sensibilidade diferentes ao tóxico usado pelos índios para a pesca. Estas populações estão a poucos quilômetros de distância [67]. No mesmo sistema de

cavernas as populações são geneticamente estruturadas por um outro gradiente de toxidade natural. Por ter origem vulcânica algumas cavernas têm altas concentrações de H_2S [47]. Esta estrutura genética foi modificada devido ao aumento de chuvas na região num período de 24hs em 2007 [48]. A estruturação em *P. mexicana* foi avaliada em todos os casos por locos mitocondriais (cit b) e nucleares (microssatélites) o que permite avaliar níveis sutis de estruturação ao contrário da região controle usada nesse trabalho para *P. vivipara*. Assim a diversidade morfológica encontrada para *P. vivipara* nas lagoas do delta do rio Paraíba do Sul poderá ser mapeada com a utilização de locos que evoluam mais rápido com os locos microssatélites.

Capítulo 5

Conclusão

A estruturação da variação genética nas populações de *P. vivipara* que ocorrem nos lagoas da planície quaternária do delta do rio Paraíba do Sul é resultado dos mesmos processos que geraram os diferentes ambientes encontrados hoje . Para investigar a influência de outros processos mais recentes é necessário utilizar marcadores moleculares que evoluam mais rápido que a região de controle da replicação mitocondrial como locos nucleares: micros-satélites e polimorfismo de um único nucleotídeo.

Referências Bibliográficas

- [1] Ab'Saber, A. N.: *Fundamentos da geomorfologia costeira do Brasil Atlântico inter e subtropical*. Revista Brasileira de Geomorfologia, 1:27–43, 2000.
- [2] Alexander, H. J., J. S. Taylor, S. T. Wu e F. Breden: *Parallel evolution and vicariance in the guppy (*Poecilia reticulata*) over multiple spatial and temporal scales*. Evolution, 60:2352–2369, 2006.
- [3] Aljanbi, S. M. e I. Martinez: *Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques*. Nucleic Acids Research, 25:4692–4693, 1997.
- [4] Allendorf, F. W. e G. Luikart: *Conservation and the genetics of populations*. Blackwell, 2007.
- [5] Auer, S. K., J. D. Arendt, R. Chandramouli e D. N. Reznick: *Juvenile compensatory growth has negative consequences for reproduction in Trinidadian guppies (*Poecilia reticulata*)*. Ecology Letters, 13:998–1007, 2010.
- [6] Bandelt, H. J., P. Foster e A. Roehl: *Median -joining networks for inferring intraspecific phylogenies*. Molecular Ecology, 16:37–48, 1999.
- [7] Bassar, R. D., M. C. Marshall, A. López-Sepulcre, E. Zandonata, S. K. Auer, J. Travis, C. M. Pringle, A. S. Flecker, S. A. Thomas, D. F. Fraser e D. N. Reznick: *Local adaptation in Trinidadian guppies alters ecosystem processes*. Proceedings of the National Academy of Sciences, 107:3616–3621, 2010.
- [8] Bidegain, P.: *Lagoas do Norte Fluminense - perfil ambiental*. Fundação Superintendência de Rios e Lagoas (SERLA), Rio de Janeiro, 2002.

- [9] Bizerril, C. R. S. F. e P. B. da S Primo: *Peixes de águas Interiores do Estado do Rio de Janeiro*. Fundação de Estudos do Mar, Rio de Janeiro, 2001.
- [10] Bulgin, N. L., H. L. Gibbs, P. Vickery e A. JBaker: *Ancestral polymorphisms in genetic markers obscure detection of evolutionarily distinct populations in the endangered Florida grass-hopper sparrow (*Ammodramus savannarum floriadanus*)*. *Molecular Ecology*, 12:831–844, 2003.
- [11] Burns, J. G., P. D. Nardo e F. H. Rodd: *The role of predation in variation in body shape in guppies *Poecilia reticulata*: a comparison of field and common garden phenotypes*. *Journal of Fish Biology*, 75:1144–1157, 2009.
- [12] Chessel, D., A. B. Dufour e J. Thioulouse: *The ade4 package-I- One-table methods*. *R news*, 4:5–10, 2004.
- [13] Clement, M., D. Posada e K. A. Crandall: *TCS: a computer program to estimate gene genealogies*. *Molecular Ecology*, 9:1657–1659, 2000.
- [14] Cunningham, M. e C. Moritz: *Genetic effects of forest fragmentation on a rainforest restricted lizard (*Scincidae: Gnypetoscincus queenslandidae*)*. *Biological Conservation*, 83:19–30, 1998.
- [15] Diegues, A. C. e P. C. C. Rosman: *Caracterização dos ativos ambientais em áreas selecionadas da Zona Costeira brasileira*. Ministério do Meio Ambiente, 1998.
- [16] Dyer, R. J.: *gstudio: GeneticStudio packages for R.*, 2012. <http://CRAN.R-project.org/package=gstudio>, R package version 0.8.
- [17] Ewing, B. e P. Green: *Base-calling of automated sequencer traces using phred. II. Error probabilities*. *Genome Research*, 8(3):186–194, 1998.
- [18] Excoffier, L. e H. E. L. Lischer: *Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows*. *Molecular Ecology Resources*, 10:564–567, 2010.
- [19] Excoffier, L., P. E. Simouse e J. M. Quattro: *Analysis of molecular variance inferred from metric distances among DNA haplotypes: application*

- to human mitochondrial DNA restriction data.* Genetics, 131:479–491, 1992.
- [20] Felsenstein, J.: *PHYLIP- Phylogeny Inference Package (Version 3.2)*. Cladistics, 5:164–166, 1989.
- [21] Freeland, J. R.: *Molecular Ecology*. Jonh Willey & Sons, Reino Unido, 2005.
- [22] Fu, Y. X.: *Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.* Genetics, 147:915–925, 1997.
- [23] Fu, Y. X. e W. H. Li: *Statistical tests of neutrality of mutations.* Genetics, 133:693–709, 1993.
- [24] Goebel, A. M., J. M. Donnelly e M. Atz: *PCR Primers and Amplification Methods for 12S Ribosomal DNA, the Control Region, Cytochrome Oxidase I, and Cytochrome b in Bufonids and Other Frogs, and an Overview of PCR Primers which have Amplified DNA in Amphibians Successfully.* Molecular Phylogenetics and Evolution, 11:163–199, 1999.
- [25] Grant, W. A. S. e B. W. Bowen: *Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation.* Journal of Heredity, 89(5):415–426, 1998.
- [26] Hartl, D. L. e A. G. Clark: *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts, 4^a ed., 2007.
- [27] Hendry, A. P., M. L. Kelly, M. T. Kinnison e D. N. Reznick: *Parallel evolution of the sexes? Effects of predation and habitat features on the size and shape of wild guppies.* Journal of Evolution Biology, 19:741–754, 2006.
- [28] Houde, A. E.: *Sexual Selection and Mate Choice in Guppies*. Princeton University Press, Princeton, New Jersey, 1997.
- [29] Krumholz, L. A.: *Reproduction in the western mosquito fish, Gambusia affinis affinis (Baird & Girard), and its use in mosquito control.* Ecological Monographs, 1:1–43, 1948.

- [30] Langerhans, R. B.: *Predicting evolution with generalized models of divergent selection: a case study with poeciliid fish*. Integrative and Comparative Biology, 50:1167–1184, 2010.
- [31] Lee-Yaw, J., A. Davidson, B. McRae e D. Green: *Do landscape processes predict phylogeographic patterns in the wood frog?* Molecular Ecology, 18:1863–1874, 2009.
- [32] Librado, P. e J. Rozas: *DnaSP v5: A software for comprehensive analysis of DNA polymorphism data*. Bioinformatics, 25:1451–1452, 2009.
- [33] Lindholm, A. K., F. Breden, H. J. Alexander, W. K. Chan, S. G. Thakurta e R. Brooks: *Invasion success and genetic diversity of introduced populations of guppies *Poecilia reticulata* in Australia*. Molecular Ecology, 14:3671–3682, 2005.
- [34] Lucinda, P. H. F.: *Check List of the Freshwater Fishes of South and Central America*, cap. Family Poeciliidae, pp. 555–581. EDIPUCRS, Porto Alegre, 2003.
- [35] Lucinda, P. H. F. e P. E. Reis: *Systematics of the subfamily Poeciliinae Bonaparte (Cyprinodontiformes: Poeciliidae), with an emphasis on the tribe Cnesterodontini Hubbs*. Neotropical Ichthyology, 3:1–60, 2005.
- [36] Martin, L., K. Suguio, J. M. L. Dominguez e J. M. Flexor: *Geologia do quaternário costeiro do litoral Norte do Rio de Janeiro e do Espírito Santo*. CPRM, Belo Horizonte, 1997.
- [37] Mattos, J. J., M. N. Siebert, K. H. Luchmann, N. Granucci, T. Dorrington, P. H. Stoco, E. C. Grisard e A. C. D. Bainy: *Differential gene expression in *Poecilia vivipara* exposed to diesel oil water accommodated fraction*. Marine Environmental Research, 69 Suppl:S31–S33, 2010.
- [38] Monteiro, L. R. e J. L. Gomes: *Morphological divergence patterns among populations of *Poecilia vivipara* (Teleostei: Poeciliidae): test of an ecological paradigm*. Biological Journal of Linnean Society, 93:799–812, 2008.
- [39] Monteiro, L. R. e J. L. Gomes-Jr.: *Morphological divergence rate tests for natural selection: Uncertainty of parameter estimation and robustness of results*. Genetics and Molecular Biology, 28(2):345–355, 2005.

- [40] Nei, M.: *Analysis of gene diversity in subdivided populations*. Proceedings of the National Academy of Sciences, 70:3321–3323, 1973.
- [41] Nei, M.: *Molecular Evolutionary Genetics*. Columbia University Press, 1987.
- [42] Palkovacs, E. P., M. C. Marshall, B. A. Lamphere, B. R. Lynch, D. J. Weese, D. F. Fraser, D. N. Reznick, C. M. Pringle e M. T. Kinnison: *Experimental evaluation of evolution and coevolution as agents of ecosystem change in Trinidadian streams*. Philosophical Transactions of Royal Society B, 364:1617–1628, 2009.
- [43] Paradis, E.: *pegas: an R package for population genetics with an integrated-modular approach*. Bioinformatics, 26:419–420, 2010.
- [44] Paradis, E., J. Claude e K. Strimmer: *APE: analyses of phylogenetics and evolution in R language*. Bioinformatics, 20:289–290, 2004.
- [45] Parenti, L. R. e M. Rauchenberger: *The Ecology and Evolution of Poeciliid Fishes (Poeciliidae) (Meffe, A. & Snelson, F. F., eds), pp. 3–12. New Jersey: Prentice Hall., cap. Systematic overview of the Poeciliines, pp. 3–12. Prentice Hall, New Jersey, 1989.*
- [46] Paulo, D. V., F. M. Fontes e F. Flores-Lopes: *Histopathological alterations observed in the liver of Poecilia vivipara (Cyprinodontiformes: Poeciliidae) as a tool for the environmental quality assessment of the Cachoeira River, BA*. Brazilian Journal of Biology, 72(1):131–140, fev. 2012.
- [47] Plath, M., J. S. Hauswaldt, K. Moll, M. Tobler, F. J. G. D. León, I. Schlupp e R. Tiedemann: *Local adaptation and pronounced genetic differentiation in an extremophile fish, Poecilia mexicana, inhabiting a Mexican cave with toxic hydrogen sulphide*. Molecular Ecology, 16(5):967–976, mar. 2007.
- [48] Plath, M., B. Hermann, C. Schröder, R. Riesch, M. Tobler, F. J. G. de León, I. Schlupp e R. Tiedemann: *Locally adapted fish populations maintain small-scale genetic differentiation despite perturbation by a catastrophic flood event*. BMC Evolutionary Biology, 10:256, 2010.

- [49] Posada, D. e K. A. Crandall: *Intraspecific gene genealogies: trees grafting into networks*. Trends in Ecology and Evolution, 16:37–45, 2001.
- [50] Projeto RADAMBRASIL: *Folhas SF. 23/24 Rio de Janeiro/Vitória; geologia, geomorfologia, pedologia, vegetação e uso potencial da terra*. Ministério de Minas e Energia, 1983.
- [51] R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. <http://www.R-project.org/>.
- [52] Ramos-Onsins, S. E. e J. Rozas: *Statistical properties of new neutrality tests against population growth*. Molecular Biology and Evolution, 12:2092–2100, 2002.
- [53] Reding, D. M., , L. A. Freed, R. L. Cann e R. C. Fleischer: *Spatial and temporal patterns of genetic diversity in an endangered Hawaiian honeycreeper, the Hawaii Akepa (*Loxops coccineus coccineus*)*. Conservation Genetics, 11:225–240, 2010.
- [54] Rogers, A. R. e H. Harpending: *Population growth makes waves in the distribution of pairwise genetic differences*. Molecular Biology and Evolution, 9:552–569, 1992.
- [55] Rosen, D. E. e R. M. Balley: *The poeciliid fishes (Cyprinodontiformes), their structure, zoogeography, and systematics*. Bulletin of the American Museum of Natural History, 126:1–176, 1963.
- [56] Russel, A. L., R. A. Medellíni e G. F. McCracken: *Genetic variation and migration in the Mexican free-tailed bat (*Tadarida brasiliensis mexicana*)*. Molecular Ecology, 14:2207–2222, 2005.
- [57] Salzburger, W., G. B. Ewing e A. von Haeseler: *The performance of phylogenetic algorithms in estimating haplotype genealogies with migration*. Molecular Ecology, 20:1952–1963, 2011.
- [58] Santos, C. P., S. B. E. Simões, H. S. Barbosa e T. Scholz: *Redescription of *Ascocotyle (Ascocotyle) felippei* Travassos, 1928 (Digenea: Heterophiyidae) with new synonymies*. Journal of Parasitology, 93(6):1468–1475, Dec 2007.

- [59] Santos, E. G. N., R. A. Cunha e C. P. Santos: *Behavioral responses of Poecilia vivipara (Osteichthyes: Cyprinodontiformes) to experimental infections of Acanthocolaritrema umbilicatum (Digenea: Cryptogonimidae)*. *Experimental Parasitology*, 127(2):522–526, Feb 2011.
- [60] Simões, S. B. E., H. S. Barbosa e C. P. Santos: *The life history of Pygidiopsis macrostomum Travassos, 1928 (Digenea: Heterophyidae)*. *Memórias do Instituto Oswaldo Cruz*, 104(1):106–111, Feb 2009.
- [61] Slatkin, M. e R. R. Hudson: *Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations*. *Genetics*, 129:555–562, 1991.
- [62] Tajima, F.: *The effect of change in population size on DNA polymorphism*. *Genetics*, 123:597–601, 1989.
- [63] Tajima, F.: *Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism*. *Genetics*, 123:585–595, 1989.
- [64] Teixeira, W., T. R. Fairchild, M. C. M. de Toledo e F. Taioli (eds.): *Decifrando a terra*. Companhia Editorial Nacional, São Paulo, 2ª ed., 2009.
- [65] Templeton, A. R., K. A. Crandall e C. F. Sing: *A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation*. *Genetics*, 132:619–633, 1992.
- [66] Tobler, M.: *Divergence in trophic ecology characterizes colonization of extreme habitats*. *Biology Journal Linnean Society*, 95:517–528, 2008.
- [67] Tobler, M., Z. W. Culumber, M. Plath, K. O. Winemiller e G. G. Rosenthal: *An indigenous religious ritual selects for resistance to a toxicant in a livebearing fish*. *Biology Letters*, 7(2):229–232, Apr 2011.
- [68] Williams, B. L., J. D. Brawn e K. N. Paige: *Landscape scale genetic effects of habitat fragmentation on a high gene flow species: Speyeria idalia (Nymphalidae)*. *Molecular Ecology*, 12:11–20, 2003.
- [69] Wright, S.: *Evolution and Genetics of Populations: II. The Theory of Gene Frequencies*. University of Chicago Press, 1969.

Apêndice A

Função índices moleculares

Função desenvolvida por mim durante o curso “Uso da Linguagem R para Análises de dados Ecológicos- 2012” oferecido pela PPG-Ecologia do Instituto de Biologia da USP.

A.1 Página de ajuda

`mol.index` package:<none> R Documentation

Indices moleculares

Description:

Retorna os índices moleculares a partir de um conjunto de sequências de DNA.

Usage:

```
mol.index(nomearq,formato="fasta")
```

```
plot.mol.index(x)
```

Arguments:

`nomearq` é o nome da arquivo que contem as sequências. As sequências devem

estar alinhadas.

formato é o formato do arquivo das sequencias. Por padrão é fasta, mais detalhes veja read.dna.

x é um lista gerada por mol.index.

Details:

mol.index retorna uma lista dos seguintes parâmetros: número de indivíduos(individuos), comprimento das sequencias(sitios), posição dos sitios segregantes(stios.seg.pos), número de sitios segregantes (sitios.seg), lista de haplotipos (haplotipos), número de haplotipos(nh), frequência de haplotipos (freq.hap), diversidade haplotípica (diver.haplo), diversidade nucleotídica(diver.nuc), theta s (theta.s), theta pi (theta.pi).

Value:

Lista com os parâmetros estimados o um conjunto de sequencias.

Author:

Carlos Henrique Tonhatti (carlostomate@gmail.com)

References:

Nei, M. (1987) *Molecular evolutionary genetics*. New York: Columbia University Press.

See Also:

pegas, read.dna, ape, seqinr

Example:

```
## Criando os dados em fasta
cat("> No305",
    "NTTCGAAAAACACACCCCACTACTAAAANTTATCAGTCACT",
    "> No304",
    "ATTCGAAAAACACACCCCACTACTAAAATTATCAACCACT",
    "> No306",
```

```
"ATTCGAAAAACACACCCCACTACTAAAAATTATCAATCACT",  
> No306",  
"ATTCGAAAAACACACCCCACTACTAAAAATTATCAATCACT",  
file = "exdna.txt", sep = "\n")
```

```
#Executando a função
```

```
a<-mol.index("exdna.txt")  
plot.mol.index(a)
```

A.2 Função mol.indice

```
# Trabalho final ##  
## carlos henrique tonhatti  
#####  
require(ape)  
require(pegas)  
  
# Carregar os dados  
mol.index <- function(nomearq,formato="fasta"){  
  
require(ape)  
require(pegas)  
  
pop<-read.dna(nomearq, format=formato)  
  
resul <- list()  
  
#tamanho da amostra (numero de individuos  
n.ind<-length(labels(pop))  
resul$individuos <-n.ind
```

```
#comprimento das seq
sitios <- length(pop[1,])
resul$sitios<- sitios

# sitios polimorficos
sitios.seg.p <- seg.sites(pop) # localização dos sitios polimorficos
resul$sitios.seg.pos <-sitios.seg.p
sitios.seg<-length(seg.sites(pop)) # o numero de sitios
resul$sitios.seg <- sitios.seg

#numero de haplotipos
hap<-haplotype(pop, labels=NULL)
resul$haplotipos <- hap
hap.l <- attr(hap,"index")
nh <- length(hap.l)
resul$nh <- nh

#diversidade genetica
#heterozigosidade esperada
freq <- rep(NA,length(hap.l))
for(i in 1:length(hap.l)){
  freq[i] <- length(hap.l[[i]])
}
resul$freq.hap <- freq

HE <- heterozygosity(freq, variance=T)
resul$diver.haplo <- HE

#diversidade nucleotidica
#pi
nd <- nuc.div(pop, variance=T)
resul$diver.nuc <-nd

# thetaS
thes <- theta.s(s=sitios.seg,n=n.ind, variance= T)
thep <- theta.pi<- nuc.div(pop)*sitios
```

```
resul$theta.s <- thes
resul$theta.pi <- c(thep,nd[2])

      return(resul)}

plot.mol.index <- function(lista){
temp <-rep(1,lista$sitios.seg)
par(mfrow=c(2,1))

plot(lista$sitios.seg.pos,temp,ann=FALSE, axes=F,cex=1,
      col="blue",pch=20,
      xlim=c(0,lista$sitios))
  mtext("Posição dos sitios segregantes", cex=1.5)
  axis(1)

plot(lista$haplotipos)
mtext("Frequência dos haplótipos",cex=1.5)

}
```

Apêndice B

Scripts

Parte dos *scripts* utilizados nas análises.

```
##Estrutura
#Carregar dados
tabela <- read.csv("table.csv")

rownames(tabela) <- tabela[,1]

tabela[,1] <- NULL
tabela[is.na(tabela)] <- 0

AC <- rep(names(tabela),as.numeric(tabela[1,]))
BI <- rep(names(tabela),as.numeric(tabela[2,]))
CA <- rep(names(tabela),as.numeric(tabela[3,]))
CI <- rep(names(tabela),as.numeric(tabela[4,]))
CO <- rep(names(tabela),as.numeric(tabela[5,]))
FE <- rep(names(tabela),as.numeric(tabela[6,]))
GR <- rep(names(tabela),as.numeric(tabela[7,]))
```

```
TA <- rep(names(tabela),as.numeric(tabela[8,]))

pop <- c(rep("BI",28),rep("CI",30),rep("CA",30),
  rep("CO",29),rep("TA",29),rep("GR",29),
  rep("FE",30),rep("AC",28))
mtDNA <- c(BI,CI,CA,CO,TA,GR,FE,AC)

# dados de GPS
lat <- c(rep(21.33602,28),rep(21.78693,30),
  rep(21.63170,30),rep(21.57653,29),
  rep(21.78748,29),rep(21.71979,29),
  rep(21.94813,30),rep(21.99416,28))
lat <- lat*-1

long<- c(rep(40.97192,28),rep(41.51622,30),
  rep(41.19149,30),rep(41.06701,29),
  rep(41.14363,29),rep(41.03783,29),
  rep(41.33949,30),rep(41.00001,28))
long <- long*-1

require(gstudio)
mtDNA.k <- c()
for(i in 1:length(mtDNA))
  mtDNA.k<- c(mtDNA.k,Locus(mtDNA[i]))

# Criando objeto população
pop.1 <- Population(pop=pop,loc=mtDNA.k,lat=lat,
  long=long)

#Grafico das frequencias
plot(freq.order*233, ylab= "frequencia",
  xlab="haplotipo", pch= 20,col="blue",
  bty="l")

#AMOVA
require(ade4)
tabela.t <- as.data.frame(t(tabela))
```

```
tabela.org <- data.frame(tabela.t[2],tabela.t[4],
                        tabela.t[3],tabela.t[5],
                        tabela.t[8],tabela.t[7],
                        tabela.t[6],tabela.t[1])

# 3 grupos (bi x (ci x ca) x TODOS)
grupo <- as.data.frame(c("a","b","b",rep("d",5)))
colnames(grupo) <- "grupo"
a2 <- amova(tabela.org,distances=NULL,structure=grupo)
randtest.amova(a2,nrepet=1000)

## distancia fisica
d.fis <- stratum.distance(pop.1,stratum="pop",lat="lat",
                        lon="long")

#Colocar a distancia fisica na mesma
#ordem que distancia genetica
ordem <- c(2,4,3,5,8,7,6,1)
d.fis <- d.fis[ordem,ordem]
# Transformar em objeto de dist
d.fis.d <- matrix(NA,ncol=8,nrow=8)
rownames(d.fis.d) <- c("BI","CI","CA","CO","TA","GR",
                    "FE","AC")
d.fis.d[lower.tri(d.fis.d)] <-(d.fis[lower.tri(d.fis)])
d.fis.d <- as.dist(d.fis.d)

# carregar fstpareado

fst <- read.csv("fst.csv",header=F)

rownames(fst) <- c("BI","CI","CA","CO","TA","GR",
                "FE","AC")

colnames(fst) <- c("BI","CI","CA","CO","TA","GR",
                "FE","AC")
fst.d <-as.dist(fst)
```

```
fst.sl <- fst*(1-fst)^-1
rownames(fst.sl) <- c("BI","CI","CA","CO","TA",
                     "GR","FE","AC")
fst.sl <- as.dist(fst.sl)
reg.fis.fst.sl <- lm(fst.sl~d.fis.d)

plot(d.fis.d,fst.sl,
     bty= "l",
     xlab="distância geográfica",
     ylab="Fst(1-Fst)^-1",
     cex=1.5,
     cex.lab=1.5,
     col="blue",
     pch=19
     )
abline(reg.fis.fst.sl,lty=2)
text(70,0.5,"R= 0,2628 \n p= 0,00310",cex=1.5)

#Mantel
require(ade4)
mantel.pop <- mantel.randtest(d.fis.d,fst.sl,nrepet=10000)
```