Demographic Expansion and Contraction in a Neotropical Fish during the Late Pleistocene-Holocene

Carolina L. N. Costa¹, S. Ivan Perez², José Louvise³, Carlos H. Tonhatti⁴, Rute B. G. Clemente-Carvalho⁵, Ana C. Petry⁶, S. F. dos Reis⁷

¹Programa de Pós-Graduação em Ecologia, Universidade Estadual de Campinas, São Paulo, Brazil
²División Antropología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, CONICET, La Plata, Argentina
³Faculdade do Vale do Jaguaribe, Ceará, Brazil
⁴Programa de Pós-Graduação em Genética, Universidade Estadual de Campinas, São Paulo, Brazil
⁵Department of Biology, Queen’s University, Ontario, Canada
⁶Núcleo em Ecologia e Desenvolvimento Sócio-Ambiental de Macaé, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
⁷Departamento de Biologia Animal, Universidade Estadual de Campinas, São Paulo, Brazil

Email: lemes.carol@gmail.com

Abstract

Demographic changes during the late Pleistocene-Holocene left signatures in the DNA of contemporary populations. These signatures reveal demographic phenomena like the increase or decrease in effective population size. In this paper we searched for signatures of demographic change in the DNA of the Neotropical freshwater fish Poecilia vivipara. Also, we investigated whether demographic changes are correlated with palaeoclimatic events of the late Pleistocene-Holocene, in particular, if changes in effective population size are correlated with expansion and contraction of available habitats, induced by global ice-volume changes and sea-level fluctuations. We used Bayesian Skyline Plot (BSP) analysis with sequences from the mitochondrial gene cytochrome b to estimate the ancestral demography of the Neotropical freshwater fish P. vivipara. To test the assumptions of neutrality and absence of population structure we used Tajima’s D and Spatial Analysis of Molecular Variance (SAMOVA), respectively. Effective population size of P. vivipara remained stable until 75,000 years ago, increased by 10-fold reaching a maximum at approximately 25,000 years ago, then suddenly declined at the Pleistocene-Holocene boundary. Variation in effective population size in P. vivipara correlates with expansion and contraction of habitats induced by sea-level fluctuations, caused by the advance and retreat of ice sheets during the Last
Glacial Maximum (LGM).

**Keywords**
Ancestral Population Dynamics, Coalescence, Bayesian Skyline Plot, Mitochondrial DNA, Last Glacial Maximum

1. Introduction

Late Pleistocene-Holocene global climatic oscillations markedly influenced the demographic history of Earth’s biota [1] [2] [3]. From the standpoint of demography, departures from stability, experienced as expansion or contraction in population size, are relevant events [4] [5]. Signatures of changes in effective population size are written in the DNA of contemporary populations because of ancestor-descendant transmission of genetic material and the processes of mutation and substitution [6] [7] [8]. Consequently, from evolutionary and ecological perspectives, it is important to assess whether temporal changes in effective population size are correlated with episodic palaeoclimatic events [9]. Of the several known cases, the steppe bison is probably one of the richest for the complexity of its demographic response to the Pleistocene-Holocene climate changes [10] [11] [12] [13]. Bison effective population size increased from 150 thousand years before present (ka BP), reaching a maximum between 30 - 45 ka BP [11]. After the maximum, effective population size declined precipitously, reaching a minimum at approximately 10 ka BP [11]. Following the minimum, effective population size increased and levelled off, and then decreased again [11]. Minimum and maximum effective population sizes in the bison correlate, among other factors, with the contraction and expansion of available habitat, driven by the advance and retreat of ice sheets during the Last Glacial Maximum (LGM) and Holocene, respectively [11]. As demonstrated by coalescent simulation, effective population size correlates with the amount of habitat available [14]. Therefore, a causal association between demographic expansion and contraction and palaeoclimatic cyclic events is tenable.

The expansion and contraction of habitats in continental areas in the Northern hemisphere is primarily a consequence of the cyclical retreat and advance of ice sheets [15]. In non-glaciated regions, however, the dynamics of habitat availability is locally induced by glacio-eustatic sea-level changes [16]. Sea-level low stands, caused by the advance of ice sheets, expose the continental shelf increasing habitat availability. Conversely, in non-glaciated regions the retreat of ice sheets induces an inverse effect on habitat availability. Such glacio-eustatic sea-level changes did have a major impact on coastal areas of South America [17] [18] [19]. Here, we investigate the potential consequences of the late Pleistocene-Holocene climatic oscillations for the historical demography of a poeciliid fish, *Poecilia vivipara*, inhabiting a non-glaciated region in South America. Poeciliids are well-known model organisms [20] [21], and *P. vivipara* is a fresh-
water species that also colonizes brackish and salt-water environments near the mouth of rivers [22]. *Poecilia vivipara* thus has the potential to track environmental changes such as the expansion or contraction of habitats associated with glacio-eustatic sea-level changes. Given the environmental changes of the late Pleistocene-Holocene and the ecological and life history attributes of *P. vivipara*, we expect contemporary populations of *P. vivipara* will harbour signatures of their historical demography. We used sequences of the mitochondrial gene cytochrome *b* (*cyt-b*), sampled from southeastern Brazil, and the coalescent-based Bayesian Skyline Plot (BSP) [11] to explore an association between historical variation in effective population size in *P. vivipara* and climatic oscillations during the late Pleistocene-Holocene. We believe our results do contribute relevant information to the ongoing research on the interplay between late Pleistocene-Holocene climatic oscillations and ecological and evolutionary processes in South America.

2. Methods

2.1. Study System

We sampled a total of 206 individuals of *P. vivipara* using a fine mesh net from 20 lagoons, which are hydrogeologically associated with the Paraíba do Sul River in northern Rio de Janeiro state in southeastern Brazil (Table 1). These lagoons vary in salinity from freshwater to brackish and saltwater [22].

2.2. DNA Processing

DNA was extracted from the caudal peduncle using the salting-out method [23]. The entire mitochondrial *cyt-b* (1140 base pairs) gene was amplified using primers and reaction parameters defined in Schories et al. [24]. Purified Polymerase Chain Reaction (PCR) products were outsourced to Macrogen Inc. (South Korea) for sequencing (using the BigDyeTM terminator kit and run on an ABI 3730XL). PCR products were sequenced in both directions with the same primers used for PCR amplification and subjected to Basic Local Alignment Search Tool (BLAST) in GenBank to verify that the desired sequences had been amplified. The sequences were deposited in GenBank (accession numbers: MG266475-MG266680).

2.3. Tests of Neutrality and Population Structure

We used Tajima’s *D* statistic [25] to test for the departure from the mutation-drift equilibrium and to evaluate whether the *cyt-b* gene conformed to neutral expectations using program DNA Sequence Polymorphism (DnaSP) v4.50.02 [26]. We estimated the genealogical relationships among haplotypes using the statistical parsimony method, developed by Templeton, Crandall and Sing (TCS) [27], with a 95% probability that no multiple substitutions have occurred, as implement in the software TCS v1.21 [28]. The assumption of a panmictic population was tested with Spatial Analysis of Molecular Variance...
Table 1. Geographical locations of the 20 sampled populations of *Poecilia vivipara*. The number of individuals (*n*) analysed for cyt-\(b\) mitochondrial DNA (mtDNA) marker and the number of haplotypes (*h*) are given for each location.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sample</th>
<th>Latitude S</th>
<th>Latitude W</th>
<th>n</th>
<th>h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Açú</td>
<td>21.99416’</td>
<td>41.00001’</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Atafona</td>
<td>21.62836’</td>
<td>41.02417’</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Cabiunas</td>
<td>22.30140’</td>
<td>41.69300’</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Cacimba</td>
<td>21.63940’</td>
<td>41.09540’</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Catingosa</td>
<td>22.18850’</td>
<td>41.41540’</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Campelo</td>
<td>21.66463’</td>
<td>41.20097’</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>Cima</td>
<td>21.78693’</td>
<td>41.51622’</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>Comércio</td>
<td>21.57530’</td>
<td>41.06740’</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>Feia</td>
<td>21.94553’</td>
<td>41.34551’</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>Barra do Furado</td>
<td>22.08086’</td>
<td>41.11850’</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>Garça</td>
<td>22.12140’</td>
<td>41.49220’</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>Grussaí</td>
<td>21.71979’</td>
<td>41.03783’</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>13</td>
<td>Heliporto</td>
<td>22.03078’</td>
<td>41.06392’</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>14</td>
<td>Iquipari</td>
<td>21.80019’</td>
<td>41.02122’</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>Lagamar</td>
<td>22.06027’</td>
<td>41.08317’</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>16</td>
<td>Praia I</td>
<td>21.57534’</td>
<td>41.06043’</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>Praia II-Gargau</td>
<td>21.57440’</td>
<td>41.05867’</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>Santa Clara</td>
<td>21.52043’</td>
<td>41.07912’</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>Tai Pequeno</td>
<td>21.80620’</td>
<td>41.12870’</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>Três Vendas</td>
<td>21.68240’</td>
<td>41.33580’</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

(SAMOVA) [29], which requires no *a priori* definition of groups of populations. SAMOVA iteratively searches for the partitions of \(K\) groups of populations that maximize the \(F_{CT}\) index, which is the proportion of the total genetic variance due to difference between groups of populations for any number of putative groups. SAMOVA was performed with 18 partitions for the 20 population samples for \(K\) varying from two to 19 groups of population samples. The \(F_{CT}\) index was computed for each partition. The cases in which all population samples belong to the same group (\(K = 1\)) or each population is itself a group (\(K = 20\)) were not included in the analysis. For each \(K\) value, 10,000 simulated annealing steps were executed starting from 200 sets of initial conditions.

### 2.4. Demographic Analysis

We estimated historical variation in the effective population size of *P. vivipara* using sequences of the cyt-\(b\) gene and the BSP method, which is based on the variable population size coalescent model [11]. Starting with a sequence alignment from a sample, the BSP method co-estimates the phylogeny and the effective population size through time [11]. This method is based on assumptions that samples come from a panmictic population and that sequences are orthologous, non-recombining, and evolve neutrally [11]. We used the software Bayesian...
Evolutionary Analysis Utility (BEAUti) [30] to generate the XML input files for the Bayesian Evolutionary Analysis Sampling Trees software (BEAST v.1.8.2) [30]. We used unlinked strict clock rates and the General Time Reversible (GTR) model of nucleotide of substitution, which best fitted the cyt-b sequences according to jModelTestv.2.1.10 [31]. We ran 400 million Markov chain Monte Carlo (MCMC) iterations, sampling every 40,000 runs. The first 10% runs were discarded as burn-in. We verified convergence to a stationary distribution with Tracer v.1.5 [32]. Effective sample size for parameter estimates was always >200. Demographic history through time was reconstructed also using Tracer v.1.5 [32]. Estimation of absolute ages of demographic events was based on a substitution rate of 0.9 × 10^{-8} per site per year, which we calculated using estimates of absolute divergence times derived by Meredith et al. [33].

3. Results

The inference of ancestral effective population size using DNA sequences and the BSP technique require testing the assumptions of neutrality and absence of population structure [11]. The assumption of neutrality was tested with Tajima’s D statistics [25], which was not significantly different from zero (D = −0.1523; P > 0.10). This result indicates that the assumption of mutation-drift is not violated and that the cyt-b gene is evolving neutrally in the sampled populations of P. vivipara. We used the statistical parsimony network [27] and the spatial analysis of molecular variance [29] to test for population structure. Thirty-eight unique haplotypes were identified among the 206 cyt-b sequences analysed. All haplotypes were connected into a single network, which has an overall star-like pattern (Figure 1). The network is characterized by three high-frequency haplotypes, several intermediate, and rare haplotypes. The haplotypes are divided in two main groups, with each group connected by a maximum of three mutational steps, whereas a minimum of nine steps separates each group. In SAMOVA, the F_{CT} statistics estimates the total genetic variance due to differences between groups of samples for partitions of K groups of populations. F_{CT} values decrease smoothly as K increases from 2 to 8 and then levels off for values of K from 9 through 19 (Figure 2). This result indicates that no single value of K is maximized, indicating a lack of population structure in the populations sampled. An example of group composition for K = 5 resulted from the simulated anealing steps is shown in Table 2.

Modelling the effective population size over time with the BSP using the cyt-b sequences revealed a complex demographic history for P. vivipara (Figure 3(a)). In this plot, the y-axis is expressed as female effective population size (N_e) × generation time and the x-axis is calendar time in years. Effective population size remained stable approximately until 75 ka BP and is followed by a steep, approximately 10-fold increase reaching a maximum value at the LGM (approximately 25 ka BP). There follows a sharp and sudden decline in effective population size starting around the Pleistocene-Holocene boundary, a trend that pers-
ists to the present. A sample of the MCMC iterations is shown in Figure 4. The trace is stable indicating good parameter mixing and convergence of the likelihood estimation (Figure 4(a)). Also the plot of the marginal density for the likelihood estimation shows a high degree of symmetry (Figure 4(b)).

Table 2. Group composition of sampled populations of *Poecilia vivipara* in one of the 10,000 simulated annealing steps for *K* = 5 in SAMOVA analysis.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of populations</th>
<th>Lagoons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>Açú, Cacimba, Comércio, Feia, Furado, Grussai, Heliporto, Iquipari, Lagamar, Praia I, Praia II, Santa Clara, Tai Pequeno Três Vendas, Catingosa, Garça,</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>Cima</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Campelo</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>Cabiunas</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Atafona</td>
</tr>
</tbody>
</table>

**Figure 1.** Statistical parsimony network, as implement in the software TCS v1.21, showing genealogical relationships among the 38 mtDNA haplotypes found (*h*1 to *h*38). Colors represent samples of different lagoons. Size of circles is proportional to haplotype frequencies. Black bars and lines denote mutational steps separating haplotypes.

**Figure 2.** Spatial analysis of molecular variance (SAMOVA) of the 20 populations samples of *Poecilia vivipara*. The proportion of the total genetic variance explained by the among-group variation (*F_{CT}* index) for any number of putative groups (*K* = 2 to *K* = 17) was lower (mean value of 33%) than the within-sample variation (mean value of 67.8%), implying no evidence of structure in the sample.
Figure 3. (a) Bayesian skyline plot derived from a sample ($n = 206$) of the cytochrome $b$ sequences from $P. vivipara$. The $x$-axis is calendar time in years. The $y$-axis is equal to female effective population size ($N_e$) × generation time. (b) Eustatic sea-level curve since 130 ka BP (based on bathymetric data from Pillans et al. [46] and Nicol [47]).

Figure 4. Samples of MCMC iterations visualized in Tracer. (a) Trace of likelihood, showing a convergence. The clearer blue line represent the first 10% runs which were discarded as burn-in. (b) The marginal density for the likelihood estimation, showing a high degree of symmetry.
4. Discussion

The cyt-b sequences of the sampled populations of *P. vivipara* do have variation, which allows the estimation of variation in ancestral effective population size over time. However, it is well established that population structure can influence the estimation of effective population size [34] [35]. In our study, no evidence of geographic structure can be gleaned from the haplotype network, that is, no discernable association between haplotypes and locations is evident. Furthermore, the evidence from the spatial analysis of molecular variance points to the absence of structure in the populations of *P. vivipara*. Equally relevant, the sequences of the cyt-b gene appear to be evolving neutrally. Therefore, fundamental assumptions of the BSP method [11] appear not to be violated by our samples.

The BSP revealed a marked increase in the effective population size of *P. vivipara*, which reached a maximum at approximately 25 ka BP ([Figure 3(a)]). This demographic event coincides approximately with the LGM, which is dated between 26.5 ka BP and 19 - 20 ka BP [36] [37]. During the LGM, the sea level of the coast of Brazil was ~120 meters below its current level [37] ([Figure 3(b)]), exposing, as a consequence, much of the continental shelf ([Figure 5(a)]), arguably increasing the availability of both freshwater and brackish habitats for *P. vivipara*. A second striking feature is the sharp demographic decrease after the maximum effective population size ([Figure 3(a)]). This demographic trend is conceivably associated with a sea-level rising after the LGM ([Figure 3(b)]), which probably induced a contraction of habitats previously available during the LGM ([Figure 5(b)]).

How might the ancestral demography of *P. vivipara* revealed by the Bayesian skyline plot be interpreted? Idiosyncrasies of the ecology and life history of contemporary populations of *P. vivipara* may shed light onto the dynamics of effective population size and the climatic-related fluctuations associated with the LGM. *Poecilia vivipara* successfully colonizes rivers and inland habitats varying in salinity from 0 to 36 parts per thousand [22]; the latter figure actually being equal to the salinity of sea water. Populations of *P. vivipara* at either extreme of the salinity gradient differ significantly in traits such as body size and shape, fecundity, and reproductive allotment, which are likely determined genetically [38] [39]. *Poecilia vivipara* also varies markedly in personality traits including boldness, activity, and sociability/shoaling behaviour, which are modulated by environmental factors such as water transparency/visibility, salinity, and dissolved oxygen [40]. Variability in personality traits vanishes, however, under homogenous laboratory conditions, indicating that plasticity underlies the behavioural differences observed in nature [40]. The available data thus suggest that genetics and plasticity play fundamental roles in the distribution of ecological and behavioural traits of *P. vivipara*, which are potentially relevant for the colonization of heterogeneous environments subjected to climatic fluctuations. In particular, plasticity, a major factor in the adaptive process [41], may be a key
mechanism underlying the increase (decrease) in effective population size associated with habitat expansion (contraction) driven by glacio-eustatic sea level oscillations during the LGM.

The ancestral demography of *P. vivipara* appears strongly correlated with environmental changes driven by the LGM. This is indeed the major task of historical demography, that is, to test for correlation between demographic phenomena, as revealed by variation in DNA sequences and palaeoclimatic episodic events [9]. However, a correlation between variation in effective population size and climatic oscillations does not always emerge. For example, in the Northern hemisphere effective population size increased monotonically in the lizard *Podarcis filfolensis* in the Maltese archipelago [42], whereas effective population size increased and then levelled off in the tree frog *Hylasarda* in the Corsica-Sardinian island system [43]. For these systems, the observed demographic expansion is actually the reverse trend expected under the classic expansion-contraction model, which predicts demographic contraction during glaciations [42]. In the Southern hemisphere, in estuarine habitats of the coastal plains of southeastern Brazil four species of fishes experienced demographic stability, whereas three other species underwent demographic expansion followed by an
asymptote in effective population size [44]. In none of these cases, however, the demographic changes did coincide with the LGM. Global climatic oscillations coupled with local geomorphological features are probably the major exogenous factors contributing to the origin of the observed diversity of demographic regimes. However, as pointed out by Ruzzante et al. [45], the life history and ecology of individual species is doubtless a fundamental endogenous determinant of the spatiotemporal demographic dynamics of contemporary populations. In this connection, Ruzzante et al. [45] showed that two sympatric species of Patagonian freshwater fishes differ in their historical demography. *Galaxias platei* experienced a bottleneck during the LGM, whereas *Percichthys truch* experienced sustained growth in effective population size for the last 100 ka BP. However, the two species differ in physiological traits, habitat preferences, feeding ecology, ability to colonize new habitats, and predation pressure [45], which conceivably explain the observed demographic dynamics of current populations of *G. platei* and *P. truch* [45]. The diversity of factors potentially affecting the demography of populations in a variety of species possibly explain the complexity of outcomes as the cited above, with different factors playing a more important role in each system.

5. Conclusion

The emerging complexity and diversity of demographic regimes in terrestrial and aquatic species during the late Pleistocene-Holocene both in the Northern and in Southern hemispheres, in particular regarding the LGM, is remarkable. The pattern of variation in effective population size we uncovered for *P. vivipara*, particularly with respect the role of the LGM, adds to the already known complexity. In this connection, comparative studies of species differing in life history and ecology, but sharing common environments, will be instrumental in uncovering the processes underlying the historical demography of contemporary populations. Such an approach might lead to the discovery of the factors that determine whether the demographic dynamics coincides or not with the episodic climatic events of the Late Pleistocene-Holocene.

Acknowledgements

We are grateful to Diego Gobbo for helping with Figure 2. João M. S. de Souza kindly helped with fieldwork. We are indebted to Mathias M. Pires for critically reviewing the manuscript. Carolina L. N. Costa was supported by a scholarship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, FinanceCode 001). José Louvise was supported by a fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Carlos H. Tonhatti was supported by a scholarship from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Ministério do Meio Ambiente (Sistema de Autorização e Informação em Biodiversidade) issued collecting permits. Research supported by CNPq and FAPESP.
Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References


