



Contents lists available at ScienceDirect

## Molecular Phylogenetics and Evolution

journal homepage: [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)

## Phylogeography of Cuban *Rivulus*: Evidence for allopatric speciation and secondary dispersal across a marine barrier



José Luis Ponce de León<sup>a</sup>, Gunnary León<sup>a</sup>, Rodet Rodríguez<sup>a</sup>, Cushla J. Metcalfe<sup>b</sup>, Damir Hernández<sup>c</sup>, Didier Casane<sup>d,e</sup>, Erik García-Machado<sup>c,\*</sup>

<sup>a</sup>Facultad de Biología, Universidad de La Habana, Calle 25, No. 455 entre J e I, Vedado, Ciudad Habana 10400, Cuba

<sup>b</sup>Instituto de Biociências, Universidade de São Paulo, Rua do Matão 277, Cidade Universitária, São Paulo, 05508-090 SP, Brazil

<sup>c</sup>Centro de Investigaciones Marinas, Universidad de La Habana, Calle 16, No. 114 entre 1ra y 3ra, Miramar, Playa, La Habana 11300, Cuba

<sup>d</sup>Laboratoire Evolution Génomes et Spéciation, UPR9034 CNRS, 1 avenue de la terrasse, 91198 Gif-sur-Yvette, France

<sup>e</sup>Université Paris Diderot, Sorbonne Paris Cité, 5 rue Thomas-Mann, 75205 Paris, France

## ARTICLE INFO

## Article history:

Received 1 December 2013

Revised 9 July 2014

Accepted 10 July 2014

Available online 19 July 2014

## Keywords:

Colour variation

Cryptic species

Dispersal

Freshwater fish

Island

Taxonomy

## ABSTRACT

The genus *Rivulus* is currently comprised of two species, *R. cylindraceus* and *R. insulaepinorum*, which are endemic to Cuba. However, the taxonomic status of the latter species remains dubious because of the poor quality of the original description. In addition, a recent barcoding survey suggests that the two species may be conspecific. The aim of this study was to test the hypothesis that the two species represent a single evolutionary clade. To delimit the species and their evolutionary history, we used a combination of molecular phylogenetic analyses, with both mitochondrial and nuclear sequences, tests of phylogeographic hypotheses, combined with morphological measurements and information on known dispersal barriers and species distribution. None of the data sets support *R. insulaepinorum* and *R. cylindraceus* as separate taxa. However, a new species, restricted to the northwestern part of the main island, was identified by phylogenetic analyses, body colour pattern and geographical distribution. The evolutionary distance between the two lineages (cytb,  $d = 15\%$ ; CAM-4,  $d = 2.5\%$ ) indicates a long period of divergence. Phylogeographic analyses shed light on the dispersal history of *R. cylindraceus*, which probably originated on the Isla de la Juventud. They also suggest that each lineage had contrasting histories; *Rivulus* sp. is restricted to a relatively small geographic area whereas *R. cylindraceus* has dispersed considerably and more than once from its centre of origin, probably facilitated by sea level fluctuations. These results strengthen previous findings, i.e. that the diversity of Cuban freshwater fishes is far from well-known and deserves more in-depth studies, and that vicariance and dispersal events have resulted in a complex biogeographical landscape which has had a significant impact on the freshwater fishes of the Caribbean islands.

© 2014 Elsevier Inc. All rights reserved.

## 1. Introduction

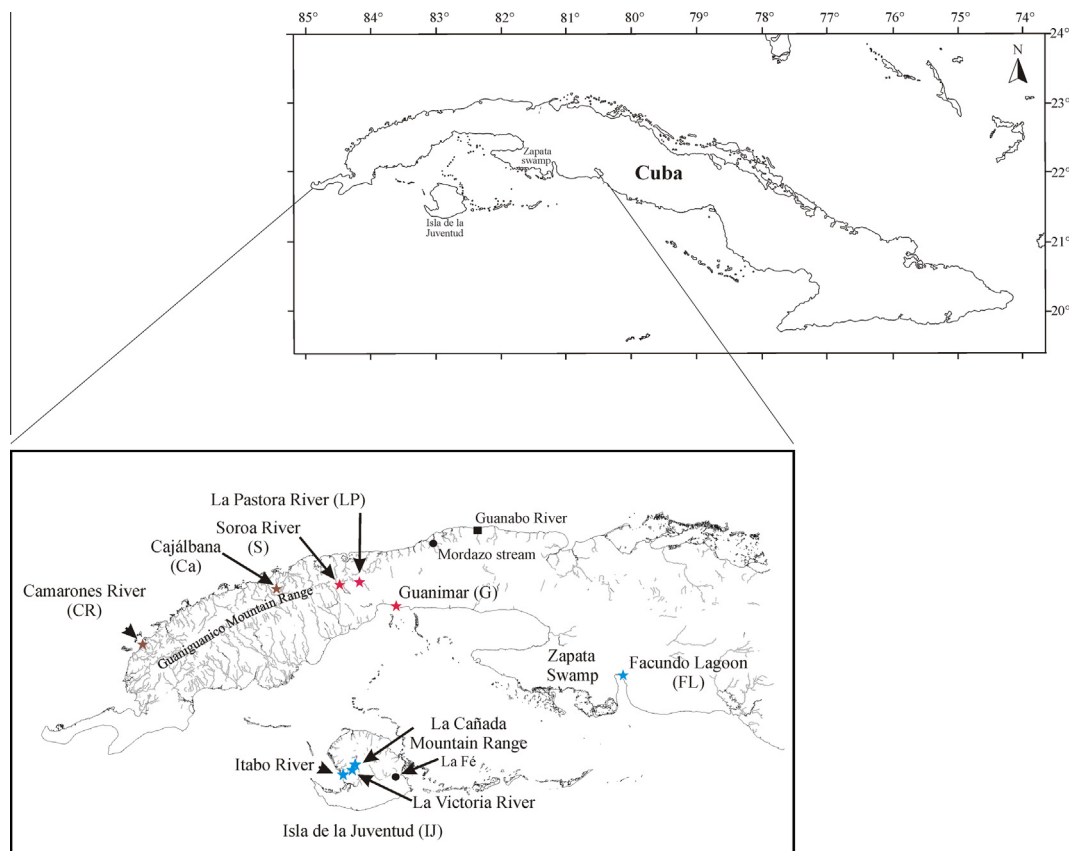
The genus *Rivulus* Poey (1860) belongs to the family Rivulidae, represented by over 350 species occurring between Florida and northeastern Argentina (Costa, 2003, 2006, 2011; Valdesalici et al., 2011). *Rivulus* is endemic to Cuba and is one of the least diverse and among the more geographically restricted of the 35 genera of freshwater fishes in the archipelago (de la Cruz and

Dubitsky, 1976). At present, two species are recognized (Costa, 2011): *R. cylindraceus* Poey, 1860 (Green rivulus) endemic to western Cuba, and *R. insulaepinorum* de la Cruz and Dubitsky, 1976 (Isle of Pines rivulus), endemic to Isla de la Juventud. Both species are restricted to freshwater mountain streams, lowland rivers, and lagoons (Fig. 1).

The delimitation of *R. cylindraceus* and *R. insulaepinorum* was based on relatively few morphological characters: i.e. the number of lateral scales and the relative position of the dorsal fin (Poey, 1860; de la Cruz and Dubitsky, 1976). The taxonomic status of the latter is particularly controversial because of the poor quality of the original description (Huber, 1992) and remains unresolved. Moreover, a recent DNA barcoding study suggests that *R. insulaepinorum* may be conspecific with *R. cylindraceus* (Lara et al., 2010). The estimated COI sequence divergence ( $1.8 \pm 0.4\%$ ) is

\* Corresponding author. Fax: +53 7 2025223.

E-mail addresses: [jotaelepe76@gmail.com](mailto:jotaelepe76@gmail.com) (J.L. Ponce de León), [gunnaryleon@gmail.com](mailto:gunnaryleon@gmail.com) (G. León), [rodetrodriguezsilva@gmail.com](mailto:rodetrodriguezsilva@gmail.com) (R. Rodríguez), [cushlametcalfe@gmail.com](mailto:cushlametcalfe@gmail.com) (C.J. Metcalfe), [damir@cim.uh.cu](mailto:damir@cim.uh.cu) (D. Hernández), [Didier.Casane@legs.cnrs-gif.fr](mailto:Didier.Casane@legs.cnrs-gif.fr) (D. Casane), [egarcia@cim.uh.cu](mailto:egarcia@cim.uh.cu) (E. García-Machado).



**Fig. 1.** Sampling sites of the genus *Rivulus* in Cuba. The sampling sites are indicated by stars. Circles indicate the type localities of *R. cylindraceus* (Mordazo stream) and *R. insulaeipinorum* (La Fé), and a square indicates the sampling site for *Kryptolebias marmoratus*. The colour of the stars corresponds to that of the clades in Fig. 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

low and a COI phylogeny did not indicate reciprocal monophyly (Lara et al., 2010). A population aggregation analysis (PAA) (Davis and Nixon, 1992) also failed to find diagnostic nucleotide sites for *R. insulaeipinorum*.

The delimitation of species has important consequences since biodiversity assessment and conservation programs are largely based on species identification (Agapow et al., 2004). The definition of a species is a contentious field of theoretical discussion in the evolutionary literature, mainly due to the lack of consensus about which properties of a species and the speciation process are universally valid criteria (de Queiroz, 2007). In recent years, several approaches have been proposed to improve species delimitation (Puerto et al., 2001; Templeton, 2001; Wiens and Penkrot, 2002; Sites and Marshall, 2003; Knowles and Carstens, 2007; Rissler and Apodaca, 2007; La Salle et al., 2009; Puillandre et al., 2012). The vast majority of authors recognize that species should be assessed within an explicit statistical framework in which a species represents a hypothesis that needs to be corroborated by supplementary data. For instance, the null hypothesis that organisms being sampled come from a single evolutionary lineage can be tested using phylogenetic and phylogeographic analysis (e.g. Templeton, 2001). Mitochondrial DNA (mtDNA) has often been the marker of choice in spite of controversy about some of its characteristics (Ballard and Whitlock, 2004; Galtier et al., 2009). Species delimitation based on mtDNA analysis can then be further assessed using independent and relevant characters such as nuclear genes, morphology, behavior, mating preference and dispersal ability (Dayrat, 2005; Will et al., 2005; Rissler and Apodaca, 2007; Yeates et al., 2011; Puillandre et al., 2012).

The distribution of tropical freshwater fishes is a function of both the current physical characteristics of the habitat, such as temperature, water level, and stream order (Moyle and Cech, 1996), as well as the effectiveness of barriers and historical climatic changes. Barrier efficiency appears to vary as function of several intrinsic (e.g. tolerance to salinity) and extrinsic (e.g. lack of corridors, presence of predators, distance between favourable habitats, episodic flooding due to strong rain and storms) factors. Under certain conditions even highly habitat – specialised species have been reported to be able to disperse and colonise new territories (Bossuyt et al., 2004; Bruyn and Mather, 2007; García-Machado et al., 2011; Walter et al., 2011). Historical variations in climate and landscape can have profound effects on species distribution and diversification by creating temporary corridors that promote species dispersion (Iturralde-Vinent and MacPhee, 1999; Iturralde-Vinent, 2006) or by imposing barriers or fragmenting territories (Jones and Johnson, 2009).

The current distribution of *R. cylindraceus* is highly fragmented, due to human impact as well as natural factors. In the former case, the destruction of the type locality, the Mordazo stream in Havana City is an obvious example. To our knowledge, *R. cylindraceus* and *R. insulaeipinorum* have low tolerance to salinity. Climate changes during the last few million years has had a profound impact on sea levels and the shape of the Cuban archipelago (Iturralde-Vinent, 2006), resulting in periods where the Isla de la Juventud has been connected to parts of the main island. The distribution of *R. cylindraceus* and *R. insulaeipinorum* on Isla de la Juventud as well as on the main island probably reflects these past changes in the landscape. Similarly, low tolerance to salinity may have

limited dispersal across the Caribbean islands in some freshwater fishes, such as *Girardinus* (Rivas, 1958; Doadrio et al., 2009) and *Poecilia* (Torres-Dowdall et al., 2013).

Using a combination of approaches, and both morphological and molecular characters, we test the hypothesis that *R. cylindraceus* and *R. insulae pinorum* represent a single evolutionary lineage. To estimate the intraspecific diversity of both molecular and morphological characters and to obtain a better estimate of inter-specific genetic divergence we collected a comprehensive data set from a large range of localities, including known and previously unknown localities for both species. We inferred independent mitochondrial and nuclear gene sequence phylogenies. The intra and inter lineage evolutionary distances were estimated and compared with those estimated from species belonging to the same family and for which homologous mtDNA sequences were available. Finally, we tested alternative dispersal pattern hypotheses using Approximate Bayesian Computation.

## 2. Materials and methods

### 2.1. Samples and localities

Two hundred and eighty seven fishes (151 females and 136 males) were collected between 2007 and 2012 using hand nets at six localities on the main island of Cuba and at three localities on the Isla de la Juventud (formerly Isla de Pinos). Five of the localities (Cajálbana, the Camarones River, the La Pastora River, Sierra La Cañada and the Soroa River) are mountain rivers and four (Guanimar, Facundo Lagoon, the Itabo River and the La Victoria River) are within lowland wetlands. The sampling was designed to cover a representative sample of the natural distribution of *Rivulus* (Fig. 1) which is particularly important because the type localities of both species no longer exist, and the preserved types were not available for the present analysis. The type locality of *R. cylindraceus* (Poey, 1860) was a freshwater channel in Havana while that of *R. insulae pinorum* was a rice field in La Fé, Isla de la Juventud (de la Cruz and Dubitsky, 1976) (Fig. 1). However, one individual (BC4, MFP 18.000143, Supplementary material 1) was sampled from the La Victoria River, the original locality of a female paratype described by de la Cruz and Dubitsky (1976). All fish were maintained in a 200 L aquarium with live and dry food provided twice a day. Individuals were collected under field capture permission numbers CH-40-DB (026) 08, CH-40-AN (80) 2009 and CH-8116247-5 from the Cuban Center for Environmental Inspection and Control (CICA).

### 2.2. Collection of DNA sequence data

Genomic DNA (gDNA) was isolated from 121 individuals, including individuals which had been previously been DNA bar-coded (Supplementary material 1). Animals were anesthetized using Tricaine mesylate (MS 222) (0.75%). Fin clips were taken, preserved in ethanol 95% and stored at  $-20^{\circ}\text{C}$ .

Total gDNA was extracted from fin clips using proteinase K (100  $\mu\text{g}/\text{mL}$ ) in 200  $\mu\text{L}$  lysis buffer (100 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 mM NaCl, 0.1% SDS, and 50 mM DTT) at  $50^{\circ}\text{C}$  with slow constant shaking, followed by a phenol:chloroform extraction and purification using a Phase Lock Gel™ (Eppendorf).

The polymerase chain reaction (PCR) was used to amplify the 5' domain of the mitochondrial cytochrome b gene (*cytb*) (754 bp) using the primers Glufish (5'-CCAATGACTTGAAGAACCACCGTTG) (Meyer et al., 1990) and CB3 5'-GCCAAATAGGAARTATCATTC (Palumbi, 1996) and the nuclear calmodulin intron 4 (*CAM-4*) using the primers CalMex4f (5'-CTGACCATGATGGCCAGAAA) and CalMex5r (5'-GTTAGCTTCTC CCCAGGTT) (Chow, 1998)

(Supplementary material 1). Five to 50 ng of gDNA was used as template in a 25  $\mu\text{L}$  PCR reaction with 1 unit of GoTaq DNA polymerase (Promega), 0.2  $\mu\text{M}$  of each primer, 200  $\mu\text{M}$  of dNTPs, and 1.5 mM  $\text{MgCl}_2$ .

PCR products were purified using the Illustra ExoStar 1-Step (General Electric Company) kit following the manufacturer's instructions. 0.8  $\mu\text{L}$  of the purified product was sequenced on an ABI 3100 automated sequencer (Applied Biosystems). Sequences have been deposited in the EMBL database, accession numbers are provided in Supplementary material 1.

### 2.3. Sequence analysis and phylogeny reconstruction

The raw sequences were inspected by eye against the chromatogram using Bioedit Sequence Alignment Editor v5.0.9 (Hall, 1999). The alignments were performed with Clustal W (Thompson et al., 1994) as included in MEGA version 5.10 (Tamura et al., 2011). For the *cytb* gene, "*Rivulus*" *roloffi* (incertae sedis) (349 bp) (Murphy et al., 1999) (Accession No. RRU41780) and an *Kryptolebias marmoratus* individual captured in the Guanababo River (BC107, Lara et al., 2010) (Accession No. LK022680) (Fig. 1, Supplementary material 1) were used as an outgroup. For the *CAM-4* gene, *K. marmoratus* was used as outgroup (Accession No. LK022681).

We used the Bayesian information criterion (BIC; Schwarz, 1978) implemented in the program jModelTest version 0.1.1 (Guindon and Gascuel, 2003; Posada, 2008) to select the nucleotide substitution model that best fits the data. MEGA 5.10 (Tamura et al., 2011) and MrBayes 3.2.1 (Ronquist et al., 2012) were used to infer a maximum likelihood (ML) and Bayesian consensus tree, respectively. The parameters of the selected model, the nucleotide substitution matrix and the gamma-distributed rate variation across sites, were used for tree inference. Stationary base frequencies and substitution rates were optimized during tree inference for the ML phylogeny, and incorporated as priors in the Bayesian inference. The ML tree was obtained by heuristic search with Nearest-Neighbour-Interchange (NNI) algorithm for tree optimization and a weak branch swap filter for an exhaustive search.

The Bayesian analyses consisted initially of two independent runs using four Metropolis coupled Monte Carlo Markov chains for  $1.3 \times 10^6$  generations, with sampling every 200 generations. TRACER v 1.5 (<http://beast.bio.ed.ac.uk/>) was used to inspect convergence of the distribution of the MCMC and evaluate the efficiency of the posterior sampling distribution with the effective sample size (ESS). The MCMC chains were then run for a further  $5 \times 10^6$  generations. The efficiency of the posterior sampling distribution was again evaluated. The first 25% of the sampled trees were discarded as burn-in. The sampled trees from the second run were used to construct a consensus tree.

The robustness of the nodes of the ML trees was assessed using the bootstrap method with 500 replicates. Posterior probabilities of the nodes in the Bayesian consensus tree was obtained from the 95% credible set of trees.

### 2.4. Genetic diversity and dispersal hypothesis testing

For each population *cyt b* haplotype (*h*) and nucleotide diversity ( $\pi$ ), as well as 95% confidence intervals for both, were estimated using DnaSP v. 5.10 with 10,000 coalescence simulations (Librado and Rozas, 2009). Haplotype relationships were examined by network analysis with the Median Joining (MJ) network algorithm (Bandelt et al., 1999) using Network 4.6.1.1 (Fluxus Technology Ltd.). These were post processed using maximum parsimony calculations to reduce the number of superfluous network links.

To explore models that best describe the origin of the geographical distribution of present-day populations, we evaluated the

posterior probabilities of historical scenarios using the Approximate Bayesian Computation approach (ABC) (Beaumont et al., 2002) as implemented in DIY ABC v. 1.0.4.46 (Cornuet et al., 2008).

We took into account the network topology and inferred diversity parameters to propose two alternative scenarios for the centre of origin (Fig. 2). In the first scenario (A) the source population was set as the Guanamar – La Pastora River drainage basin, and in the second scenario (B), on Isla de la Juventud. For both scenarios, the Soroa River and Facundo Lagoon populations were set as more recently derived. We assumed that a bottleneck had occurred for each founding population since it is expected that they would represent a fraction of the source population. We also assumed that the effective number of founders ( $N_{fi}$ ) remained small for a few generations before reaching a larger effective population size ( $N_i$ ). It was also assumed that the effective size of the ancestral population remained constant over time. We chose uniform priors for  $N_i$ ,  $N_{fi}$  and bottleneck duration ( $Db$ ) (Supplementary material 2). The mutation model for the mtDNA sequence was chosen as described above for phylogenetic analysis, default values were used for the remaining parameters. The summary statistics used were: the number of haplotypes, the number of segregating sites, the mean pairwise differences and variance for each population, the number of pairwise differences within and between populations, and the Hudson et al. (1992) estimator of population differentiation ( $F_{ST}$ ). We produced a reference table with four million simulated data sets. This procedure was repeated three times to test for consistency of the results. The posterior probabilities of the two scenarios were estimated using the logistic regression method (Fagundes et al., 2007) from 1% (40,000) of the simulated data sets.

### 2.5. Meristic and morphometric variables

We measured two meristic and three morphometric variables that are generally considered reliable characters (Hoedeman, 1959) and commonly used in killifish species descriptions (Costa, 2004; Valdesalici et al., 2011; Valdesalici and Schindler, 2011). The meristic variables used were: the number of pre-dorsal scales (NPS) and the number of lateral scales (NLS). The frontal squamation pattern of *R. cylindraceus* and *R. insulaepinorum* is d-type (Hoedeman, 1958). To measure the NPS we followed Hoedeman (1958) and counted the scales along the mid row between the cephalic scale b (including b) and the first dorsal fin ray. The morphometric variables used were: Pre-dorsal proportion (PP), that is the ratio between pre-dorsal length (PDL: from snout to the base of the first dorsal ray) and Total length (TL); Anal proportion (AP), that is the ratio between pre-anal length (PAL: from snout to the first anal ray) and TL; and Inter dorsal-anal Proportion (Id-aP), that

is the ratio between inter dorsal-anal distance (IDAL: PDL-PAL) and TL. These five variables were analysed for a variable number of male (M) and female (F) individuals from each site: the La Pastora River (32 M and 35 F), the Soroa River (26 M and 26 F), Guanamar (8 M and 18 F), Facundo Lagoon (22 M and 23 F), Cajálbana (15 M and 16 F), Isla de la Juventud (33 M and 33 F). Scale counts were done with a stereoscopic microscope (40 $\times$ ). All morphological variables were measured using a Vernier Caliper to the nearest 0.05 mm. Body colour patterns on males and females were also documented (i.e.: main body colouration, fin colouration, presence/absence of stripes and spots).

### 2.6. Statistical analysis of morphological data

Meristic and morphometric data do not fulfill the assumptions of normality and variance homogeneity, even after data transformation (1/x, Sqr, Log and Sqr (arcsine)), we therefore used non parametric tests. We compared the central tendency of every meristic and morphometric variable between sexes of each population using the Mann Whitney *U* test. We then used the Kruskal–Wallis test and a pairwise comparison test (Dunn's test) to conduct central tendency comparisons of each variable separately for males and females. All these statistical analyses were done using Statistica 6.1 (Statsoft, 2004, Tulsa, IL, USA). To determine if individuals from the sampled localities could be differentiated using a combination of morphological characters, a Principal Components Analysis (Correlation type) was performed using the program Past 1.99 (Hammer et al., 2001).

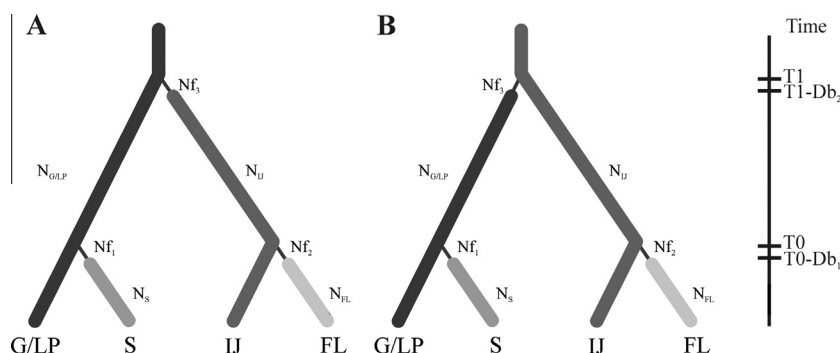
## 3. Results

### 3.1. Cytochrome *b* phylogeny

The final *cyt b* sequence alignment is from 121 individuals and is 754 bp long, with 115 variable sites. Twenty nucleotide mutations result in inferred amino acid substitutions. Two substitutions are shared by all individuals from the Cajálbana and Camarones Rivers, one is unique to samples from the Soroa River, and one is shared by all individuals from the Isla de la Juventud and Facundo Lagoon.

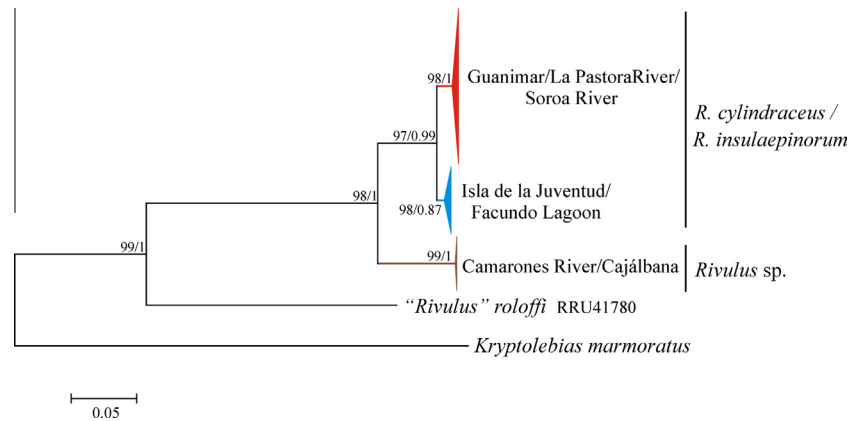
The HKY +  $\Gamma$  evolutionary model was selected by jModeltest and used in the ML and Bayesian analyses. The gamma shape parameter ( $\alpha$ ) inferred in the ML tree reconstruction ( $\alpha = 0.67$ ) was used as prior for Bayesian analysis.

Both methods resulted in well resolved and fully congruent trees (Fig. 3). The robustness of the nodes (bootstrap value and posterior probability) was high (greater than 98 for all the main nodes). Two major clades were identified. The first is composed



**Fig. 2.** Alternative hypotheses tested using the approximate Bayesian computation approach based on *cytb* sequences of the species *R. cylindraceus*. Model (A) establishes that Guanamar–La Pastora drainage basin represents the source population, while model (B) proposes that the point origin was on the Isla de la Juventud. FL, G, IJ, S and LP indicate: Facundo Lagoon, Guanamar, Isla de la Juventud, Soroa River and La Pastora River, respectively.  $N_i$  indicates the effective population size of a given population;  $Db$  is bottleneck duration, and  $N_{fi}$  is the effective number of founders after a dispersal event.





**Fig. 3.** Maximum likelihood phylogenetic tree ( $\text{Log} L = -2407.12$ ) based on partial *cytb* sequences. Two clades were identified, a smaller clade of *R. cylindraceus* individuals endemic to the northwestern region of the main island (Camarones River and Cajalbana), renamed *Rivulus* sp., and a second, larger, clade of *R. cylindraceus* and *R. insulaeipinorum* individuals from the rest of the localities, named *R. cylindraceus*. The HKY model with gamma correction ( $\alpha = 0.67$ ) was used. ML bootstrap values and Bayesian analysis posterior probabilities are represented at each node, in that order. For sake of clarity, the three main clades are collapsed and are represented by triangles whose width is proportional to the number of individuals.

of individuals identified as *Rivulus cylindraceus* / *R. insulaeipinorum* from most of the sampled localities, hereafter the *R. cylindraceus* clade. The second is composed of *Rivulus cylindraceus* samples from Cajalbana and the Camarones River, hereafter the *Rivulus* sp. clade. In the first clade two well supported subclades were recovered, the first subclade is composed of samples from the La Pastora River, Guanimar and the Soroa River, and second subclade of samples from Isla de la Juventud and the Facundo Lagoon (Fig. 1). This second subclade has been previously identified by COI sequence analyses (Lara et al., 2010) and indicates a close relationship between individuals identified geographically as *R. insulaeipinorum* (Isla de la Juventud) and *R. cylindraceus* (Facundo Lagoon, Zapata swamp).

Mean evolutionary distances within samples from different localities (using Tamura-Nei correction) were low and ranged from 0.03% to 0.55% (Table 1). The mean divergence estimates within the two *R. cylindraceus* subclades was also low:  $d = 0.3\% \pm 0.1$  for Guanimar–the La Pastora River–the Soroa River, and  $d = 0.5\% \pm 0.1$  for the Isla de la Juventud – the Facundo Lagoon. Between the two subclades a  $d = 2.6\% \pm 0.6$  was obtained. In contrast, the mean sequence divergence between *R. cylindraceus* and *Rivulus* sp. was high,  $d = 15.0\% \pm 2.5$ .

### 3.2. CAM-4 phylogeny

CAM-4 sequences were obtained for a subset of 24 individuals representing all localities. The sequence alignment was 550 bp long with 20 variable sites and indels that ranged in length from 2 to 174 bp (Table 2). The 174 bp sequence indel was absent in individuals from Guanimar, the Soroa River and the Facundo Lagoon. Several individuals were heterozygous for nucleotide variants and four for allele size (2 bp indel). For the nucleotide variant,

phylogenetic analysis was carried out using a consensus allele in which the polymorphic sites were replaced by the IUPAC code (A and G = R; T and C = Y). For the size allele variations the longest allele was used. Maximum likelihood and Bayesian phylogenetic reconstructions were conducted for the CAM-4 sequences using the complete deletion option in MEGA to exclude gap regions.

The TN model was selected by jModeltest. The unrooted ML tree ( $\text{Log} L = -894.33$ ) has a similar topology to the one obtained with the *cytb* sequences (Fig. 4). However the Bayesian inference produced a basal polytomy (not shown). In the ML tree the first clade, the *R. cylindraceus* clade, was recovered with high bootstrap value (95%), while the second clade, the *Rivulus* sp. Clade, was recovered with a relatively low bootstrap support (74%) but high posterior probabilities (0.97) (Fig. 4). Five nucleotide changes and two indels, involving two and eight nucleotides respectively, were identified as diagnostic in distinguishing the two clades. However, it should be noted that the two individuals from the Camarones River (*Rivulus* sp. clade) shared five nucleotide substitutions with individuals from the *R. cylindraceus* clade.

In contrast with the *cytb* phylogeny, the two subclades within the *R. cylindraceus* clade were not recovered. The Camarones River and Cajalbana individuals showed a low evolutionary distance ( $d = 0.7\% \pm 0.2$ , Tamura-Nei correction), similar to the *R. cylindraceus* intraclade divergence ( $d = 0.2\% \pm 0.17$ ). However, the sequence divergence between the *R. cylindraceus* and *Rivulus* sp. clades was relatively high ( $d = 2.5\% \pm 0.8$ ).

### 3.3. Morphological analysis

There was no consistent sex difference pattern for the meristic and morphometric variables analysed (Supplementary material 3).

**Table 1**  
Estimation of the mean pairwise genetic distance (in %) among the different sampled localities. Below and above the diagonal are Tamura-Nei distance and standard error estimates. On the diagonal (on bold) are the intra-locality estimates ( $p$ ). Highlighted on light grey are pairwise inter-locality distances, and in darker grey are pairwise inter-species distances.

Locality	Soroa River	La Pastora River	Guanimar	Facundo Lagoon	Isla de la Juventud	Cajalbana	Camarones River
Soroa River	<b>0.06</b>	0.2	0.2	0.6	0.6	2.6	2.7
La Pastora River	0.31	<b>0.03</b>	0.0	0.6	0.6	2.7	2.9
Guanimar	0.37	0.10	<b>0.16</b>	0.6	0.6	2.7	2.9
Facundo Lagoon	2.54	2.54	2.59	<b>0.08</b>	0.2	2.7	2.6
Isla de la Juventud	2.63	2.64	2.69	0.59	<b>0.55</b>	2.7	2.6
Cajalbana	14.41	14.90	14.92	14.87	15.26	<b>0.22</b>	0.3
Camarones River	15.41	15.91	15.92	15.09	15.47	1.01	<b>0.27</b>

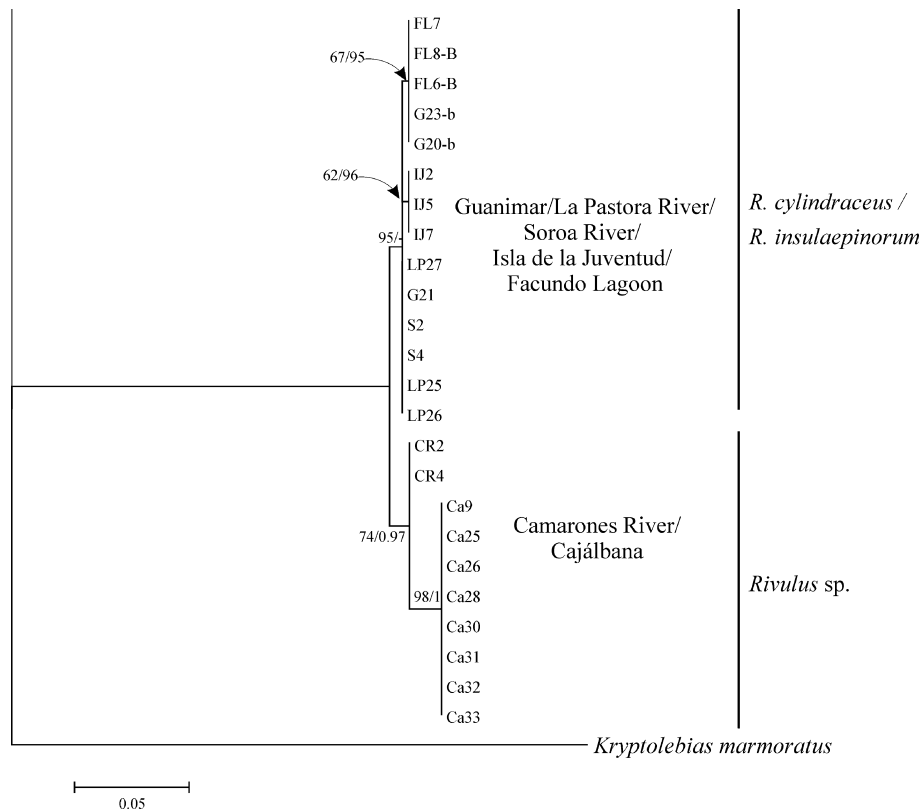
**Table 2**  
Polymorphism statistics summary.

Clade/locus	Sampling location	<i>n</i>	Sequence length <sup>a</sup>	<i>S</i> <sup>b</sup>	Number of haplotypes <sup>b</sup>	$\pi \pm SD$	<i>h</i> $\pm$ SD	<i>k</i> $\pm$ SD
<i>R. cylindraceus</i> cytb	Guanimar	24	754	11	8	0.0016 $\pm$ 0.00054	0.612 $\pm$ 0.113	1.217 $\pm$ 0.228
	Isla de la Juventud	16	754	17	10	0.0053 $\pm$ 0.00102	0.917 $\pm$ 0.049	4.000 $\pm$ 0.743
	La Pastora River	25	754	2	4	0.0003 $\pm$ 0.00015	0.230 $\pm$ 0.110	0.237 $\pm$ 0.077
	Soroa River	18	754	4	4	0.0006 $\pm$ 0.00030	0.314 $\pm$ 0.138	0.444 $\pm$ 0.134
	Facundo Lagoon	14	754	2	3	0.0008 $\pm$ 0.00022	0.560 $\pm$ 0.125	0.626 $\pm$ 0.192
	All	99	754	51	30	0.0118 $\pm$ 0.00091	0.830 $\pm$ 0.032	8.916 $\pm$ 0.590
CAM-4	All	14	373–550	2	3	0.0023 $\pm$ 0.0004	0.692 $\pm$ 0.065	0.857 $\pm$ 0.237
<i>Rivulus</i> sp. cytb	Cajalbana	22	754	4	5	0.0013 $\pm$ 0.00024	0.623 $\pm$ 0.080	0.779 $\pm$ 0.176
	Camarones River	2	754	2	2	0.0027 $\pm$ 0.0013	1.000 $\pm$ 0.500	2.000 $\pm$ 1.453
	All	24	754	9	7	0.0024 $\pm$ 0.0007	0.685 $\pm$ 0.077	1.46 $\pm$ 0.263
CAM-4	All	10	534–540	9	2	0.0060 $\pm$ 0.0027	0.356 $\pm$ 0.159	3.20 $\pm$ 0.796

*n* is the number of sequenced individuals; *S* is the number of segregating sites;  $\pi$  is the nucleotide diversity (Nei, 1987); *h* is the haplotypic diversity; *k* is the mean number of differences between haplotypes.

<sup>a</sup> Sequence length is the maximum length of the aligned sequences. CAM-4 intron sequences were variable in length.

<sup>b</sup> *S* and the number of haplotypes were determined excluding the indels for CAM-4.



**Fig. 4.** Maximum likelihood phylogenetic tree (LogL = -894.33) based on CAM-4 sequences. Two clades were identified, a smaller clade of *R. cylindraceus* individuals endemic to the northwestern region of the main island (Camarones River and Cajalbana), renamed *Rivulus* sp., and a second, larger, clade of *R. cylindraceus* and *R. insulaepinorum* individuals from the rest of the localities, named *R. cylindraceus*. The Tamura-Nei model was used. ML bootstrap values and Bayesian analysis posterior probabilities are represented at each node, in that order. Only bootstrap values greater than 60% are shown.

The number of lateral scales, number of predorsal scales, dorsal proportion, anal proportion and inter dorsal and anal proportions overlap extensively between both sexes within localities (Supplementary material 3.1 and 3.2). For most variables, significant differences in median values between sexes were only found for the La Pastora River samples. The dorsal proportion was significantly different for the Facundo Lagoon and Guanimar samples, while samples from most localities showed significant differences between individuals of both sexes for the anal proportion, except for those from the Soroa River and Cajalbana (Supplementary material 3.3). However, there was a high degree of overlap in the

variability of the meristic and morphometric variables analysed for both male and female among localities (Supplementary material 3.1 and 3.2), and consequently, no meristic or morphometric variable was identified as being diagnostic for the *R. cylindraceus* or *Rivulus* sp. clades.

The results of the Principal Component Analysis (PCA) are shown in Fig. 5. The first two principal components account for 64.22% of the total variance. The PCA scatter plot indicates that most of the individuals distributed are around the PC intercept, with a marked overlap between the six sampled localities.

Body colouration varies significantly among localities (Fig. 6A), particularly in males, resulting in four well distinguished morphs. Females showed variation in body colouration, but it much less pronounced than in males. One morph, characterized by a greenish pale brown body, with scarce orange dots on the flanks and black fins edges (Fig. 6A) was found in the La Pastora and Soroa Rivers (Fig. 6A). Some individuals also have black marks in the mouth. The second morph was found in the Facundo Lagoon and Guanimar (Fig. 6A). The males were dark brownish olive green with several orange dots on the flanks. The third morph was only found on the Isla de la Juventud (Fig. 6A). Fish were bright yellowish green with several red dots on the flanks. The fourth morph was found exclusively in the Cajálbana and the Camarones River (Fig. 6A). In this case, the individuals presented a black to deep grey stripe along the median line, and especially in males, pale orange colouration on ventral part of the body from mouth to caudal peduncle, pelvic and anal fins, and irregular deep orange dots densely spread on the belly and anal fins.

### 3.4. Haplotype relationships

Twenty eight *cytb* haplotypes were found within the *R. cylindraceus* clade and seven in *Rivulus* sp. clade. A network of haplotype relationships shows that, in the *R. cylindraceus* clade, with the exception of samples from the Guanimar and La Pastora River, all the localities have unshared (diagnostic) haplotypes, and that the mean number of differences between haplotypes is relatively high ( $k = 8.9 \pm 0.59$ ) (Fig. 6A, Table 2). It is interesting to note that haplotypes from the Facundo Lagoon are nested within the larger haplotype diversity of the Isla de la Juventud (Fig. 6A). Haplotype frequency analysis between Guanimar and the La Pastora River rejected the expectation of frequency homogeneity between these two localities ( $\chi^2 = 13.31$ ; d.f. = 10;  $p = 0.02$ , using the CHIRXC program with 10,000 randomizations (Zaykin and Pudovkin, 1993)), suggesting some degree of genetic differentiation between these localities. The long branch connecting the *R. cylindraceus* and *Rivulus* sp. clade haplotypes has a minimum of 61 nucleotide changes.

### 3.5. Genetic diversity

Estimates of the within sample locality genetic diversity for both genes is shown in Table 2. Mean values were in general higher

for *cytb* than for the *CAM-4* though estimates of nucleotide diversity and the mean number of haplotypic diversity show an inverse trend in the case of the *Rivulus* sp. clade. Samples from the Isla de la Juventud and Guanimar were remarkable for their high *cytb* nucleotide and haplotypic diversities. In particular, the Isla de la Juventud has the highest values ( $\pi = 0.005 \pm 0.001$ ,  $h = 0.917 \pm 0.049$ ) and the highest mean differences between haplotypes ( $k = 4 \pm 0.74$ ). The Cajálbana and Camarones River samples show moderate nucleotide and haplotypic diversity; however there are only two Camarones River samples.

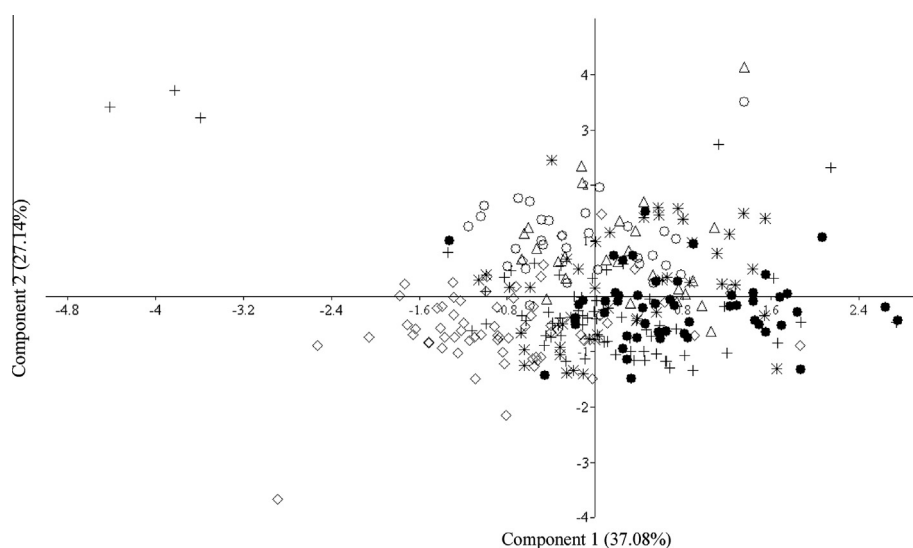
### 3.6. Tests of alternative hypotheses on the origin and dispersal of *R. cylindraceus*

The Isla de la Juventud and Guanimar have the highest levels of genetic and haplotype differences. We hypothesized that one of these may be the centre of origin from which *R. cylindraceus* dispersed. In order to simplify evaluation of various scenarios we pooled together samples from Guanimar and the La Pastora River because, despite significant differences in haplotypes frequencies (see above), they share the most common haplotype at each locality, indicating gene flow or recent common ancestry (see discussion).

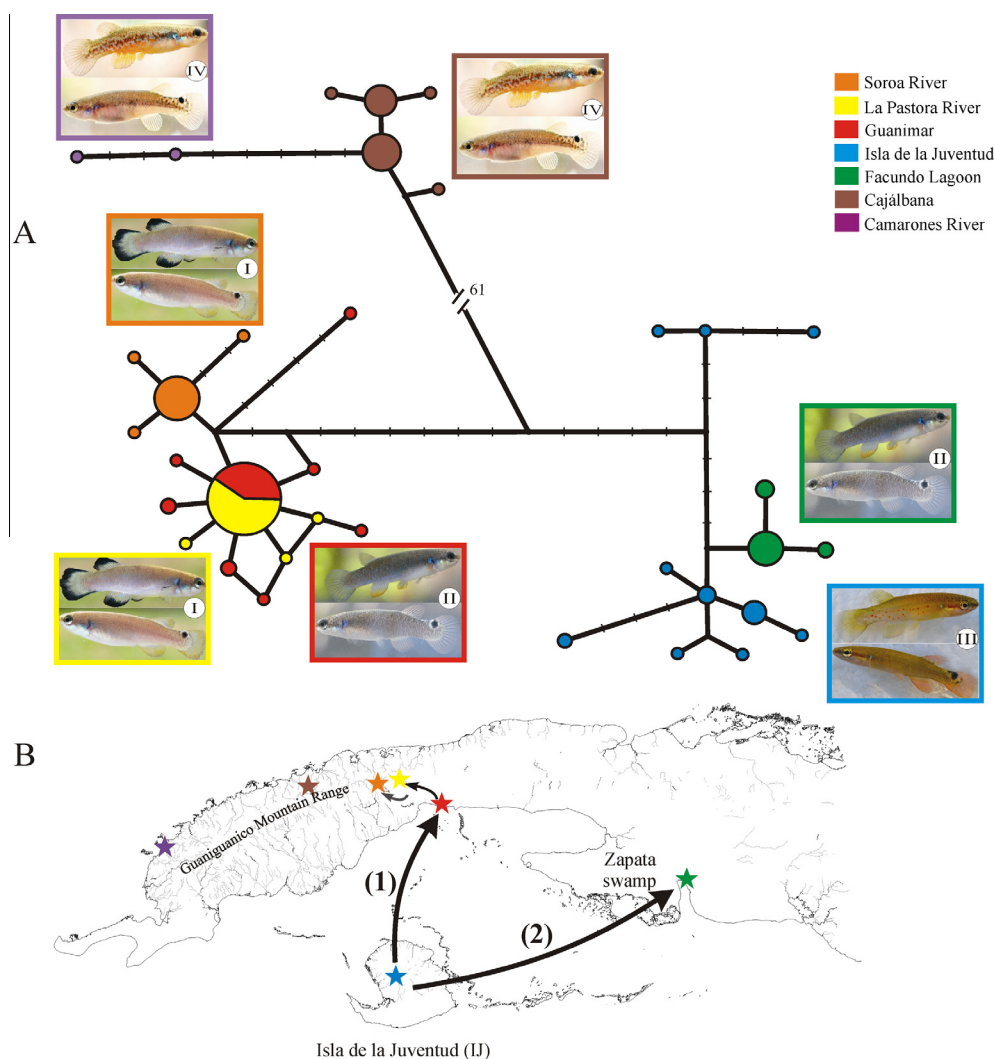
An evaluation of the scenarios (Fig. 2) using the ABC approach suggests with a strong probability (higher than 89% in all three independent runs) that the centre of origin was the Isla de la Juventud (Fig. 6B). The results also suggest that there were at least two separate dispersal events from the Isla de la Juventud to the main island: an earlier one to Guanimar and the La Pastora River and a second to the Facundo Lagoon. A further, more recent, dispersal event can be inferred from ABC analysis from Guanimar and the La Pastora River to the Soroa River.

## 4. Discussion

We used phylogenetic analysis (mtDNA and nuclear sequences) as well as phenotypic traits to investigate the systematics, distribution and diversity of the poorly-understood *Rivulus* genus. The genus *Rivulus* is currently comprised of two species endemic to Cuba, *R. cylindraceus* (Poey, 1860) and *R. insulaepinorum* (de la Cruz and Dubitsky, 1976). We tested the hypothesis that *R. cylindraceus* and *R. insulaepinorum* represent a single evolutionary lineage. Our



**Fig. 5.** Principal component analysis (PCA) based on two meristic (number of lateral scales and number of pre-dorsal scales) and three morphometric variables (dorsal proportion, anal proportion and inter dorsal-anal proportion) taken from 287 *Rivulus* individuals (males and females) collected at six localities in Cuba. Symbols for localities: Cajálbana (open circle); Facundo Lagoon (asterisk); Guanimar (triangle); Isla de la Juventud (diamond); La Pastora River (+); Soroa River (black circle).



**Fig. 6.** (A) Haplotype network of *cytb* partial sequences from *Rivulus* in Cuba. The size of the circles (haplotypes) is proportional to the frequency of the haplotypes. Thin vertical lines crossing the connecting lines indicate a single mutation step. The broken line with the 61 next to it indicates the number of changes inferred to connect *Rivulus* sp. to *R. cylindraceus* haplotypes. (B) The most likely hypothesis about the origin and pattern of dispersal events inferred for *R. cylindraceus* obtained by *cytb* sequence variation ABC analysis. (1) First dispersal; (2) Second dispersal. The different colour morphs are depicted using the colour code (boxes) for each haplotype group and geographic location. Phenotypic variants of *Rivulus*. *R. cylindraceus* from La Pastora and Soroa rivers (I); Guanimar and Facundo Lagoon (II); Isla de la Juventud (III). *Rivulus* sp. from Cajálbana and Camarones River (IV). Males and females are on the top and the bottom, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

analyses identified two independent evolutionary lineages within *Rivulus*, a smaller clade of *Rivulus* sp. individuals endemic to the northwestern region of the main island and a second, larger, clade of all *R. cylindraceus* and *R. insulaepinorum* individuals from the remaining area of distribution of *Rivulus* in Cuba. These results indicate that the smaller clade is a newly identified species while *R. insulaepinorum* is not a valid species. We propose using *R. cylindraceus* for the larger clade, and *Rivulus* sp. for the smaller clade. To distinguish between the two uses of the term *R. cylindraceus* we will use '*R. cylindraceus*' to describe prior use of the term, and *R. cylindraceus* for the larger clade identified here. We also show that the complex paleogeographic processes that characterise the formation of the Cuban archipelago during the last few thousand to million years has probably shaped the dispersion and colonisation of *R. cylindraceus* from the Isla de la Juventud to the main island.

#### 4.1. *Rivulus insulaepinorum* is not a valid species

The original description of the species *R. insulaepinorum* was based solely on morphological analysis of four female individuals from the Isla de la Juventud (de la Cruz y Dubitsky, 1976). Huber

(1992) and Murphy and Collier (1996) considered that this species was probably conspecific with *R. cylindraceus* and suggested that the species needed verification. At present, one of the major difficulties in approaching this problem is the loss of the type locality, a rice field on Isla de la Juventud.

Our mtDNA *cytb* sequence analyses indicate that *Rivulus* samples from the Isla de la Juventud are genetically similar to some samples from the main island (i.e. those from the Facundo Lagoon in the Zapata swamp). The *cytb* divergence (2.6%) was lower than values for congeneric species pairwise comparisons for several genera of rivulids. To establish a range of sequence divergence between species and genera, we constructed a ML phylogenetic tree from 67 reported *cytb* nucleotide sequences of rivulids, including two *Rivulus*, from the Guanimar and from the Cajálbana. As all genera were not monophyletic in the ML tree, for each genus we identified the clade which contained most of the species and discarded the few, if any, species which were outside the clade. We estimated *cytb* sequence divergence between species for 13 genera [*Anablepsoides* ( $n = 9$ ), *Atlantirivulus* ( $n = 3$ ), *Austrofundulus* ( $n = 2$ ), *Austrolebias* ( $n = 2$ ), *Cynodonichthys* ( $n = 7$ ), *Campellolebias* ( $n = 2$ ), *Gnatholebias* ( $n = 2$ ), *Hypsolebias* ( $n = 2$ ), *Laimosemion* ( $n = 10$ ),



*Leptolebias* ( $n = 2$ ), *Melanorivulus* ( $n = 2$ ), *Moema* ( $n = 2$ ), *Trigonectes* ( $n = 2$ ) (data not shown). The Tamura-Nei corrected distance ( $d_{TN}$ ) estimates ranged from a minimum of  $d_{TN} = 4\%$  to a maximum of  $d_{TN} = 36.5\%$  (uncorrected distances:  $p = 3.9$  and  $p = 27.2$ , respectively). For *Rivulus* we estimate an overall mean  $d_{TN} = 5.6\%$  (uncorrected distance:  $p = 4.7$ ), which is higher than the lowest limit estimated above. Lara et al. (2010), using COI sequences, found that the genetic divergence between '*R. cylindraceus*' and *R. insulaepinorum* ( $d = 1.8\%$  with K2P correction) was below the 2–3% threshold that is used to delimit many freshwater fish sister species (Hubert et al., 2008; April et al., 2011) and similar to the divergence observed at the intraspecific level in *R. cylindraceus* ( $1.6 \pm 0.4\%$ ).

The nuclear *CAM-4* sequences and morphological data also failed to distinguish *R. insulaepinorum* from '*R. cylindraceus*', but in contrast to the *cytb* tree, the *CAM-4* tree did not identify subclades within *R. cylindraceus* associated with locality (Fig. 4). A similar result has been previously reported based on allozyme loci (Rodríguez, 2009). Neither meristic nor morphometric variables discriminated between '*R. cylindraceus*' and *R. insulaepinorum* (Fig. 5). All these results support the hypothesis that *R. insulaepinorum* should be considered a synonym of the more widely distributed '*R. cylindraceus*' (Huber, 1992; Murphy and Collier, 1996; Lara et al., 2010). However, the level of divergence of *cytb* sequences ( $d_{TN} = 2.6\%$ ) between the two subclades of *R. cylindraceus* approaches that of subspecies level. The Facundo Lagoon and Isla de la Juventud individuals could also therefore be considered '*R. insulaepinorum*' with a larger geographic distribution than previously thought.

#### 4.2. A new species in the north of Pinar del Río

Based on molecular data the Cajalbana and Camarones River lineage, *Rivulus* sp., is reciprocally monophyletic ("exclusive" *sensu* Wiens and Penkrot, 2002) with respect to *R. cylindraceus* (Figs. 3 and 4). These lineages show 15% *cytb* and 2.5% *CAM-4* sequence divergence, which is more than three times the minimum divergence observed between *cytb* sequences of rivulid sister species (see above). Additionally, although *Rivulus* sp. cannot be distinguished from '*R. cylindraceus*' based on most morphological variables, they can be distinguished on adult male body colour pattern (Fig. 6). Finally, the non-overlapping geographic distributions of *R. cylindraceus* and *Rivulus* sp., the separate phylogeographic history of *R. cylindraceus* (see below), morphological features found in *Rivulus* sp. that are not found in *R. cylindraceus*, such as the dark stripe in both sexes, also found in juveniles, and the pale orange colouration on the ventral part of the body, all indicate that *Rivulus* sp. should be considered a new species and that *R. cylindraceus* and *Rivulus* sp. have a relatively long history of allopatric divergence.

#### 4.3. Origin and dispersal of *R. cylindraceus*

Phylogeographic analyses can shed light on gene flow barriers and past events that may have influenced the present geographic distribution of species and populations (Templeton, 1998; Puorto et al., 2001; Gifford et al., 2004). In both *R. cylindraceus* and *Rivulus* sp. individuals from geographically close drainage basins have closely related haplotypes (Fig. 1 and Fig. 6) suggesting recent ancestry, followed by geographic isolation, complete allele sorting and limited or no gene flow after colonisation (Avice et al., 1987). The only gene flow detected was between Guanímar and the La Pastora River individuals. Guanímar, a lowland wetland, and the La Pastora River, a mountain river, are on the same (South) slope and during strong rains or hurricanes they could become confluent. However, it should be noted that other data suggest that gene flow is limited, only one high frequency haplotype is shared by both localities, dif-

ferences in haplotype frequencies are statistically significant and there are also differences in colouration.

Two waves of dispersal and colonisation were inferred from the pattern of geographic distribution of haplotypes and from the test of phylogeographic hypotheses. The Isla de la Juventud as the source population for *R. cylindraceus* in the Cuban archipelago was strongly supported. The differentiation observed between groups of haplotypes of both islands appears to result mostly from the barrier represented by the sea that separates the Isla de la Juventud and the main Cuban island. To our knowledge *R. cylindraceus* has low salt water tolerance, which would effectively reduce its dispersal ability across marine barriers. The Cuban archipelago was formed during the latest Eocene–early Oligocene (35–33 Ma). During most of this time the western part of Cuba was two major lands masses which today are the Isla de la Juventud and the Guaniguanico mountain range. These territories have varied widely in size over time but remained unconnected until 3–2 Ma when the low lands between them were periodically exposed. During the last several thousand years there have been several glacial and interglacial cycles with repeated low-land exposure and inundation. The maximum land area exposed during this time was 25–20 ka (Iturralde-Vinent, 2003, 2006).

During periods of low-land exposure the connection between Isla de la Juventud and the Guaniguanico mountain range (Furrazola-Bermúdez et al., 1964) could have allowed the dispersal of *R. cylindraceus* to the southern part of the western Cuba (*i.e.* Guanimar) and more recently from Isla de la Juventud to the south of the Matanzas province (*i.e.* Facundo Lagoon). This second event appears to correlate with the more recent emergence of the Zapata swamp basin (Iturralde-Vinent, 2003). Additionally, these dispersal events appear to have occurred only once since there is no evidence of intermixing (reciprocal monophyly) of haplotype lineages at any of these localities. Some authors have proposed similar scenarios based on observed inter-basin exchanges and patterns of phylogenetic relationships between populations of some freshwater vertebrate species in the Caribbean. Alexander et al. (2006) (*Poecilia reticulata*), and Walter et al. (2011) (*Anablepsoides hartii*) have suggested that the close relationship between haplotypes from neighbouring territories is most likely related to cycles of gene flow and isolation events as a consequence of glacial/interglacial cycles and associated sea level fluctuations during the Pleistocene. Alonso et al. (2012) suggests that the observed recent and widespread gene flow within *P. peltocephala*, including expansion into formerly inundated areas, and the presence of this species on the Isla de la Juventud, may be the result of the separation of the island from mainland Cuba during the Pleistocene or Holocene.

The investigation of the geographic pattern of speciation involves information on phylogeny, divergence times and species distribution (Barraclough and Vogler, 2000). Theoretically, species that have diverged allopatrically should show a higher level of geographic overlap as time from divergence increases, while species that have diverged sympatrically should show lower geographic overlap as time increases from divergence (Lynch, 1989). Although in this case no information is available about the ancestral distribution of the two lineages, haplotype divergence and the inferred progression of the dispersal of *R. cylindraceus* from Isla de la Juventud to the Cuba main island then to the west of the island suggests long term allopatric divergence between the two main lineages (*R. cylindraceus* and *Rivulus* sp.). A similar scenario has been observed for the blind cavefish genus *Lucifuga*, in which several evolutionary lineages are confined to narrow allopatric geographic areas (*e.g.* *L. simile*, *L. dentata holguinensis*) while others (*i.e.* *Lucifuga dentata*) show progressively long range dispersal events that have resulted in geographical overlap with other highly divergent species (García-Machado et al., 2011). A complex pattern of population differentiation was also shown for *A. hartii* across Trinidad

(Jowers et al., 2007; Walter et al., 2011), suggesting that post-colonisation diversification on islands may be more common than thought. The fragmented nature of freshwater ecosystems may prevent gene flow between populations, thereby increasing population differentiation overtime (April et al., 2011).

The *Rivulus* sp. lineage is found in the north-western region of Cuba, which is a relatively narrow area of the north drainage slope delimited in the south by the Guaniguanico mountain range (see Fig. 1). However, its distribution does not seem to be determined by the direction of water flow within the drainage basins across the island, since *R. cylindraceus* was first recorded from the Mor-dazo stream, in the north of La Havana, and more recently was found by the authors at two localities: San Claudio river and Las Terrazas, both from the northern slope in the eastern part of the Guaniguanico mountain range. Historical factors other than the direction of water flow within the drainage basins may have determined its isolation and divergence.

We lack data (i.e. mutation rates, fossil or biogeographical events) to calibrate sequence divergence, but the 1–2% mutation rate per one million years widely used for fish (Brown et al., 1979; Bermingham et al., 1997; Near and Benard, 2004; García et al., 2012) would place the split between *Rivulus* sp. and *R. cylindraceus* to 4 and 8 Ma, during the late Miocene. Assuming that at that time the distribution of *Rivulus* sp. was restricted to the north of the Guaniguanico mountain range, and *R. cylindraceus* to the Isla de la Juventud, it seems likely that a common ancestor colonised both territories during this period. However, the high genetic divergence estimated between these lineages raises the possibility that there may have been more than one independent colonisation event. This has been described for *A. hartii* for which multiple intra and inter-island colonisation best explains haplotype relationships between different Caribbean islands and South America (Jowers et al., 2007; Walter et al., 2011).

## 5. Concluding remarks

Based on the analysis of several data sets and independent of the criteria used to delimit species, we have demonstrated the invalidity of *R. insulaepinorum* and, have identified *Rivulus* sp. as a new species. High genetic divergence, differences in body colouration, non-overlapping geographic distributions, and patterns of population dispersion, all support the hypothesis that *R. cylindraceus* and *Rivulus* sp. are independent evolutionary units with no evidence of gene flow between them, and thus should be recognised as separate species.

Our results are also consistent with a growing body of evidence suggesting that complex biogeographical scenarios resulting from a combination of both vicariance and dispersal events have had an important impact on the freshwater fishes of the Caribbean islands (Murphy and Collier, 1997; Chakrabarty, 2006; Jowers et al., 2007; Doadrio et al., 2009; Costa, 2010; Hulsey et al., 2011; García-Machado et al., 2011; Walter et al., 2011), and have promoted diversification at different taxonomic and geographical scales. This study is concordant with a previous analysis of the freshwater fishes of Cuba (Lara et al., 2010), indicating that the biodiversity of freshwater fishes in this archipelago is poorly understood.

## Acknowledgements

We want to thank Adrian Gerhart for his help in preparing the map and geospatial positioning of the sampling sites and Yuriem Lezcano for her valuable assistance in finding for rare articles. We would also like to thank two anonymous reviewers for their useful comments and suggestions to improve the manuscript.

We thank the University of Paris-Sud, the CNRS, the Embassy of France in Cuba and the University of Havana for financial and logistic support during this investigation.

The University of Paris-Sud and the Laboratoire Evolution Génomes et Spéciation, UPR9034 CNRS provided laboratory facilities and support for laboratory analysis.

The Centro de Investigaciones Marinas and the Faculty of Biology of the University of Havana provided transportation for sampling and support for laboratory analysis.

The Embassy of France in Cuba provided financial support for exchanges of professionals and laboratory consumables.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2014.07.007>.

## References

- Agapow, P.-M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman, J.L., Mace, G.M., Marshall, J.C., Purvis, A., 2004. The impact of species concept on biodiversity studies. *Q. Rev. Biol.* 79, 161–179.
- Alexander, H.J., Taylor, J.S., Wu, S.S.-T., Breden, F., 2006. Parallel evolution and vicariance in the guppy (*Poecilia reticulata*) over multiple spatial and temporal scales. *Evolution* 60 (11), 2352–2369.
- Alonso, R., Crawford, A.J., Bermingham, E., 2012. Molecular phylogeny of an endemic radiation of Cuban toads (Bufonidae: Peltophyryne) based on mitochondrial and nuclear genes. *J. Biogeogr.* 39, 434–451.
- April, J., Mayden, R.L., Hanner, R.H., Bernatchez, L., 2011. Genetic calibration of species diversity among North America's freshwater fishes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10602–10607.
- Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* 18, 489–522.
- Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* 13, 729–744.
- Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
- Barracough, T.G., Vogler, A.P., 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* 155, 419–434.
- Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
- Bermingham, E., MacCafferty, S., Martin, A.P., 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: Kocher, T.D., Stepien, C.A. (Eds.), *Molecular Systematics of Fishes*. Academic Press, San Diego, CA, pp. 113–128.
- Bossuyt, F., Meegaskumbura, M., Beenaerts, N., Gower, D.J., Pethiyagoda, R., Roelants, K., Mannaert, A., Wilkinson, M., Bahir, M.M., Manamendra-Arachchi, K., Ng, P.K.L., Schneider, C.J., Oommen, O.V., Milinkovitch, M., 2004. Local endemism within the Western Ghats-Sri Lanka biodiversity hotspot. *Science* 306, 479–481.
- Brown Jr., W.M., George Jr., M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 76, 1967–1971.
- Bruyn, M., Mather, P.B., 2007. Molecular signatures of Pleistocene sea-level changes that affected connectivity among freshwater shrimp in Indo-Australian waters. *Mol. Ecol.* 16, 4295–4307.
- Chakrabarty, P., 2006. Systematics and historical biogeography of greater Antillean Cichlidae. *Mol. Phylogenet. Evol.* 39, 619–627.
- Chow, S., 1998. Universal PCR primer for calmodulin gene intron in fish. *Fish. Sci.* 64, 999–1000.
- Cornuet, J.M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.M., Balding, D.J., Guillemaud, T., Estoup, A., 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24, 2713–2719.
- Costa, W.J.E.M., 2003. A new species of the genus *Rivulus* Poey, 1860 from the Parnaíba river basin, northeastern Brazil (Teleostei: Cyprinodontiformes: Rivulidae). *Boletim do Museu Nacional, Rio de Janeiro* 511, 1–7.
- Costa, W.J.E.M., 2004. *Kryptolebias*, a substitute name for *Cryptolebias* Costa, 2004 and *Kryptolebitinae*, a substitute name for *Cryptolebitinae* Costa, 2004 (Cyprinodontiformes: Rivulidae). *Neotrop. Ichthyol.* 3 (1), 69–82.
- Costa, W.J.E.M., 2006. Relationships and taxonomy of the killifish genus *Rivulus* (Cyprinodontiformes: Aplocheiloidi: Rivulidae) from the Brazilian Amazonas river basin, with notes on historical ecology. *Aquat. J. Ichthyol. Aquat. Biol.* 11, 133–175.
- Costa, W.J.E.M., 2010. Historical biogeography of cynolebiasine annual killifishes inferred from dispersal–vicariance analysis. *J. Biogeogr.* 37 (10), 1995–2004.

- Costa, W.J.E.M., 2011. Phylogenetic position and taxonomic status of *Anablepsoides*, *Atlantirivulus*, *Cynodonichthys*, *Laimosemion* and *Melanorivulus* (Cyprinodontiformes: Rivulidae). *Ichthyol. Explor. Freshw.* 22 (3), 233–249.
- Davis, J.I., Nixon, K.C., 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Syst. Biol.* 41 (4), 421–435.
- Dayrat, B., 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85, 407–415.
- de la Cruz, J., Dubitsky, A.M., 1976. Dos nuevas especies de peces dulceacuicolas del género *Rivulus* Poey (Cyprinodontidae) de Cuba e Isla de Pinos. *Poeyana* 155, 1–5.
- de Queiroz, K., 2007. Species concepts and species delimitation. *Syst. Biol.* 56, 879–886.
- Doadrio, I., Perea, S., Alcaraz, L., Hernández, N., 2009. Molecular phylogeny and biogeography of the Cuban genus *Girardinus* Poey, 1854 and relationships within the tribe Girardinini (Actinopterygii, Poeciliidae). *Mol. Phylogenet. Evol.* 50, 16–30.
- Fagundes, N.J.R., Ray, N., Beaumont, M., Neuenchwander, S., Salzano, F.M., Bonatto, S.L., Excoffier, L., 2007. Statistical evaluation of alternative models of human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17614–17619.
- Furrazola-Bermúdez, G., Judoley, C.M., Mijailóvskaya, M.S., Novojastky, I.P., Núñez-Jiménez, A., Solsona, J.B., 1964. *Geología de Cuba*. Editorial Nacional de Cuba.
- Galtier, N., Nabholz, B., Glémin, S., Hurst, G.D.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18, 4541–4550.
- García, G., Gutiérrez, V., Vergara, J., Calviño, P., Duarte, A., Loureiro, M., 2012. Patterns of population differentiation in annual killifishes from the Paraná-Uruguay-La Plata Basin: the role of vicariance and dispersal. *J. Biogeogr.* 39, 1707–1719.
- García-Machado, E., Hernández, D., García-Debrás, A., Chevalier-Monteagudo, P., Metcalfe, C., Bernatchez, L., Casane, D., 2011. Molecular phylogeny and phylogeography of the Cuban cave-fishes of the genus *Lucifuga*: evidence for cryptic allopatric diversity. *Mol. Phylogenet. Evol.* 61 (2), 470–483.
- Gifford, M.E., Powell, R., Larson, A., Gutberlet Jr., R.L., 2004. Population structure and history of a phenotypically variable teiid lizard (*Ameiva chrysolasma*) from Hispaniola: the influence of a geologically complex island. *Mol. Phylogenet. Evol.* 32, 735–748.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 41, 95–98.
- Hammer, O., Harper, D.A.T., Ryan, P.D., 2001 4 (1), 9.
- Hoedeman, J.J., 1958. The frontal scalation pattern in some groups of toothcarps (Pisces, Cyprinodontiformes). *Bull. Aquat. Biol.* 1, 23–28.
- Hoedeman, J.J., 1959. Rivulid fishes of Surinam and other Guyanas. With a preliminary review of the genus *Rivulus*. *Studies on the fauna of Surinam and other Guyanas* 7, 44–98.
- Huber, J.H., 1992. Review of *Rivulus*: ecobiogeography relationship the most widespread Neotropical cyprinodont genus. *Société Française d'Ichtyologie*, Paris.
- Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Laviolette, N., Burrigge, M., Watkinson, D., Curry, A., Bentzen, P., Zhang, J., April, J., Bernatchez, L., 2008. Identifying Canadian freshwater fishes through DNA barcodes. *Plos ONE* 3, e2490.
- Hudson, R.R., Boos, D.D., Kaplan, N., 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9, 138–151.
- Hulsey, C.D., Keck, B.P., Hollingsworth, P.R., 2011. Species tree estimation and the historical biogeography of heroine cichlids. *Mol. Phylogenet. Evol.* 58, 124–131.
- Iturralde-Vinent, M., 2003. *Ensayo sobre la paleogeografía del Cuaternario de Cuba*. Memorias, Resúmenes y Trabajos, V Congreso Cubano de Geología y Minería, CD ROM, ISBN 959-7117-II-8.
- Iturralde-Vinent, M., 2006. Meso-Cenozoic Caribbean paleogeography: implications for the historical biogeography of the region. *Int. Geol. Rev.* 48, 791–827.
- Iturralde-Vinent, M., MacPhee, R.D.E., 1999. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Am. Museum Nat. History Bull.* 238, 1–95.
- Jones, C.P., Johnson, J.B., 2009. Phylogeography of the livebearer *Xenophallus umbratilis* (Teleostei: Poeciliidae): glacial cycles and sea level change predict diversification of a freshwater tropical fish. *Mol. Ecol.* 18, 1640–1653.
- Jowers, M.J., Cohens, B.L., Downie, J.R., 2007. The cyprinodont fish *Rivulus* (Aplocheiloidei: Rivulidae) in Trinidad and Tobago: molecular evidence for marine dispersal, genetic isolation and local differentiation. *J. Zool. Syst. Evol. Res.* 46 (1), 48–55.
- Knowles, L.L., Carstens, B.C., 2007. Estimating a geographically explicit model of population divergence. *Evolution* 61, 477–493.
- La Salle, J., Wheeler, Q., Jackway, P., Winterton, S., Hobern, D., Lovell, D., 2009. Accelerating taxonomic discovery through automated character extraction. *Zootaxa* 2217, 43–55.
- Lara, A., Ponce de León, J.L., Rodríguez, R., Casane, D., Côté, G., Bernatchez, L., García-Machado, E., 2010. DNA barcoding of Cuban freshwater fishes: evidence for cryptic species and taxonomic conflicts. *Mol. Ecol. Res.* 10 (3), 421–430.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Lynch, J.D., 1989. The gauge of speciation: on the frequency of modes of speciation. In: Otte, D., Endler, J.A. (Eds.), *Speciation and its Consequences*. Sinauer, Sunderland, Massachusetts, USA, pp. 527–553.
- Meyer, A., Kocher, T.D., Basasibwaki, P., Wilson, A.C., 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 347, 550–553.
- Moyle, P.B., Cech Jr., J.J., 1996. *Fishes. An introduction to Ichthyology*, third ed. Prentice-Hall, Upper Saddle River, New Jersey, 590pp.
- Murphy, W.J., Collier, G.E., 1996. Phylogenetic relationships within the Aplocheiloid fish genus *Rivulus* (Cyprinodontiformes, Rivulidae): implications for Caribbean and Central America Biogeography. *Mol. Biol. Evol.* 13 (5), 642–649.
- Murphy, W.J., Collier, G.E., 1997. A molecular phylogeny for Aplocheiloid fishes (Atherinomorpha: Cyprinodontiformes): the role of vicariance and the origins of annualism. *Mol. Biol. Evol.* 14 (8), 790–799.
- Murphy, W.J., Thomason, J.E., Collier, G.E., 1999. Phylogeny of the Neotropical Killifish Family Rivulidae (Cyprinodontiformes, Aplocheiloidei) inferred from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 13, 289–301.
- Near, T.J., Benard, M.F., 2004. Rapid allopatric speciation in logperch darters (Percidae: Percina). *Evolution* 58, 2798–2808.
- Nei, M., 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Palumbi, S.R., 1996. *Nucleic acids II: the polymerase chain reaction*. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), *Molecular Systematics*. Sinauer Associates Inc, Massachusetts, USA, pp. 205–247.
- Poey, F., 1860. *Memorias sobre la historia natural de la isla de Cuba, acompañadas de sumarios latinos y extractos en francés*. Habana, Imprenta de Barcina, Cuba.
- Posada, D., 2008. JModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256.
- Puillandre, N., Modica, M.V., Zhang, Y., Sirovich, L., Boisselier, M.-C., Cruaud, C., Holford, M., Samadi, S., 2012. Large-scale species delimitation method for hyperdiverse groups. *Mol. Ecol.* 21, 2671–2691.
- Puerto, G., Da Graca Salomao, M., Theakston, R.D.G., Thorpe, R.S., Warrell, D.A., Wuster, W., 2001. Combining mitochondrial DNA sequences and morphological data to infer species boundaries: phylogeography of lanceheaded pitvipers in the Brazilian Atlantic forest, and the status of *Bothrops pradoi* (Squamata: Serpentes: Viperidae). *J. Evol. Biol.* 14, 527–538.
- Rissler, L.J., Apodaca, J.J., 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Syst. Biol.* 56, 924–942.
- Rivas, L.R., 1958. The origin, evolution, dispersal and geographical distribution of the Cuban poeciliid fishes of the tribe Girardinini. *Am. Philos. Soc.* 102 (3), 281–320.
- Rodríguez, R., 2009. *Análisis de la validez taxonómica de la especie Rivulus insulaeiporum* (Teleostei: Cyprinodontiformes). Master in Science Thesis. Faculty of Biology, University of Havana.
- Ronquist, F., Teleslenko, M., Van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61 (3), 539–542.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Statist.* 6, 461–464.
- Sites Jr., J.W., Marshall, J.C., 2003. Delimiting species: a renaissance issue in systematic biology. *Trends Ecol. Evol.* 18, 462–470.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony method. *Mol. Biol. Evol.* 28, 2731–2739.
- Templeton, A.R., 1998. Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* 7, 381–397.
- Templeton, A.R., 2001. Using phylogeographic analyses of gene trees to test species status and processes. *Mol. Ecol.* 10, 779–791.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Res.* 22, 4673–4680.
- Torres-Dowdall, J., Dargent, F., Handelsman, C.A., Ramnarine, I., Ghalambor, K., 2013. Ecological correlates of the distribution limits of two poeciliid species along a salinity gradient. *Biol. J. Linn. Soc.* 108, 790–805.
- Valdesalici, S., Schindler, I., 2011. Description of a new killifish of the genus *Rivulus* (Teleostei: Cyprinodontiformes: Rivulidae) from south eastern Peru. *Vertebr. Zool.* 61, 313–320.
- Valdesalici, S., García, J.R., Tavares, D., 2011. *Rivulus albae*, a new species of killifish (Teleostei: Cyprinodontiformes: Rivulidae) from northeastern Brazil. *Vertebr. Zool.* 61 (1), 105–108.
- Walter, R.P., Blum, M.J., Snider, S.B., Paterson, I.G., Bentzen, P., Lamphere, B.A., Gilliam, J.F., 2011. Isolation and differentiation of *Rivulus hartii* across Trinidad and neighbouring islands. *Mol. Ecol.* 20, 601–618.
- Wiens, J.J., Penkrot, T.A., 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Syst. Biol.* 51, 69–91.
- Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54, 844–851.
- Yeates, D.K., Seago, A.Y., Nelson, L., Cameron, S.L., Joseph, L., Trueman, J.W.H., 2011. Integrative taxonomy, or iterative taxonomy? *Syst. Entomol.* 36, 209–217.
- Zaykin, D.V., Pudovkin, A.I., 1993. Two programs to estimate Chi-square values using pseudo-probability test. *J. Hered.* 84, 152.