GENOTYPING AND MITOCHONDRIAL PHYLOGENOMICS OF THE ASEXUAL FISH POECILIOPSIS MONACHA-LATIDENS

A Thesis

by

YUNYANG WANG

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Mariana Mateos
Committee Members,	Manfred Schartl
	Claudio Casola
Head of Department,	Cliff Lamb

December 2021

Major Subject: Wildlife and Fisheries Sciences

ABSTRACT

The freshwater livebearing fish genus *Poeciliopsis* has served as a study system for the evolution of sex, due to the presence of asexually reproducing (gynogenetic and hybridogenetic) biotypes that originated through hybridization. Previous studies inferred the evolutionary origin of most *Poeciliopsis* asexual biotypes, but not of the diploid hybridogen (=hemiclone) *P. monacha-latidens*. We used phylogenetic analyses of whole mitochondrial genomes to infer that *P. monacha-latidens* had multiple independent origins. Within the Rio Mocorito (its southernmost limit), the monophyly of *P. monacha-latidens* and *P. monacha-lucida* and their sharing of identical mitochondrial haplotypes, suggests that they represent the same hemiclonal lineage, utilizing two sperm donor species. We provide the first report of heteroplasmy in *Poeciliopsis*, and of existence of *P. viriosa* x *P. latidens* hybrids. In addition, as an alternative to allozyme-based genotyping for *Poeciliopsis*, we optimized a DNA-based protocol to diagnose the presence or absence of a nuclear *P. latidens* allele.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professor Mariana Mateos [advisor] of the Department of Ecology and Conservation Biology, Professor Claudio Casola of the Department of Ecology and Conservation Biology, and Professor Manfred Schartl from of the Department of Chemistry and Biochemistry, Texas State University, San Marcos. The *Ck-M* gene sequences used to design primer sequences were provided by Professor Mariana Mateos. Most of the samples used were collected by Professor Mariana Mateos and collaborators (see below). Professor Mariana Mateos also contributed to the preparation of the figures. All other work conducted for the thesis was completed by the student independently.

Alejandro Varela, Luis A. Hurtado, Robert C. Vrijenhoek helped in the fieldwork. Renato D. LaTorre helped design primers and troubleshoot PCRs. Samantha Levell and David Reznick provided the putative *P. viriosa* x *P. latidens* male specimen and photograph. Albert van der Heiden contributed the putative *P. viriosa* x *P. latidens* female photograph.

Funding Sources

Part of this study was supported by funding from NSF DEB-9902224 funded to R.C. Vrijenhoek and MM, Monterey Bay Aquarium Research Institute.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I INTRODUCTION

The predominance of sexual reproduction among eukaryotes represents an important paradox of evolutionary biology. By sexual reproduction we mean a form of propagation that typically requires the union of two gametes produced via the process of meiosis with recombination (Holub & Shackelford, 2021). Sexual reproduction is subject to several theoretical "costs", which are generally not incurred by asexual reproduction. One of these is the "cost of males" (Smith, 1978), whereby a population that only invests in the production of females is expected to grow more rapidly than a population that invests half of its resources in the production of males. Another cost to sexual reproduction is the "cost of meiosis" (Williams, 2020), whereby only half of each parent's genes are passed to the progeny. Additional costs incurred by some sexually reproducing organisms include resources and time devoted to reproductive behaviors and sexually selected traits, and exposure to natural enemies during mating (Lewis, 1987). Despite these presumed advantages of asexual reproduction, only 0.1% to 1% of eukaryotic species are asexual (De Meeûs et al., 2007)

Several phenomena may explain why sexual reproduction is more common than asexual reproduction. Muller's ratchet, which is the irreversible accumulation of harmful mutations predicted to occur in the absence of recombination (Muller, 1964), is expected to decrease the mean fitness of a finite asexual population, which can ultimately drive it to extinction (Kondrashov, 1994). Experimental evidence of Muller's ratchet, based on decreased fitness over time of asexual lab lineages, has been documented in diverse

lineages, including ciliated protozoa (Bell, 1989), an RNA bacteriophage (Chao, 1990), the bacterium *Salmonella typhimurium* (Andersson & Hughes, 1996), and yeast (Zeyl et al., 2001). Another approach to examine Muller's ratchet is to compare the proportion of inferred deleterious mutations based on molecular evolution analyses in sexual populations and their asexual counterparts. Evidence for Muller's ratchet at the DNA level has been detected for a few asexual taxa (Bast et al., 2018; Hollister et al., 2015; Lovell et al., 2017; Sharbrough et al., 2018), but not others (Jaron et al., 2021).

Another explanation for the persistence of sexual reproduction is the Red Queen hypothesis, whereby asexually reproducing populations lack the genetic diversity generated by sexual reproduction, that would enable them to avoid extinction in the face of rapidly evolving natural enemies, such as parasites (Ladle, 1992). Numerous empirical studies of host-parasite systems have found evidence consistent with several predictions of the Red Queen Hypothesis, such as local or time-lagged adaptation, rapid parasite evolution, and association between host reproductive mode and parasite infection rate (Gibson et al., 2016; King et al., 2011; Lively et al., 1990; Morran et al., 2014; Park & Bolker, 2019).Other empirical studies have not found evidence consistent with the Red Queen Hypothesis (Gösser et al., 2019; Tobler et al., 2005).

An alternative hypothesis to explain the rarity and short evolutionary longevity of asexual lineages without invoking fitness disadvantages, posits that if new clonal lineages arise at relatively high rate and compete with each other, it could result in an expected young age for extant clonal lineages (Janko et al., 2008). Finally, for systems where asexual lineage formation is tied to a hybridization event, such as vertebrates, the

following hypotheses to explain their rarity have been proposed. The "balance hypothesis" (Moritz et al., 1992) states that asexuality arises only when the genomic distance between the hybridizing species is high enough to disrupt normal gametogenesis, but sufficiently low to enable hybrids to be fertile and viable (R C Vrijenhoek, 1989). The "rare formation" hypothesis states that in addition to the genomic preconditions, the two sexual species must coexist in the same place (Costa & Schlupp, 2020; Matthias Stöck et al., 2010).

Asexuality in vertebrates is restricted to fish, reptiles and amphibians, and has been documented in approximately 100 lineages (Avise, 2015; Neaves & Baumann, 2011). Three main asexual reproduction modes occur in vertebrates: parthenogenesis; gynogenesis; and hybridogenesis. Parthenogenesis is the development of an individual directly from an unfertilized egg without the need of males (Smith, 1978). Gynogenesis is a mechanism in which the sperm from a related sexual species is required to activate embryogenesis, but its genetic material is not incorporated into the egg and thus, barring any new mutations, the genotype does not change over generations (Vrijenhoek, 1994). Hybridogenesis is a hemiclonal reproduction mechanism in which paternal genes are incorporated and expressed in the hybrids but are not heritable; only the maternal genome is transmitted between generations (Schultz, 1969) (Figure 1). Although it is a rare reproduction mode, it has been found in some non-fish vertebrates such as *Ambystoma* salamanders (Bogart et al., 1989), *Bufotes* toads (Stöck et al., 2011), *Pelophylax* water frogs (Dedukh et al., 2019), as well as several fish genera, such as *Squalius* (Alves et al., 2002), *Misgurnus* (Morishima et al., 2008), *Hypesleotris* (Schmidt

et al., 2011), *Hexagrammos* (Suzuki et al., 2020), *Fundulus* (Dalziel et al., 2020), and *Poeciliopsis* (Schultz, 1969).

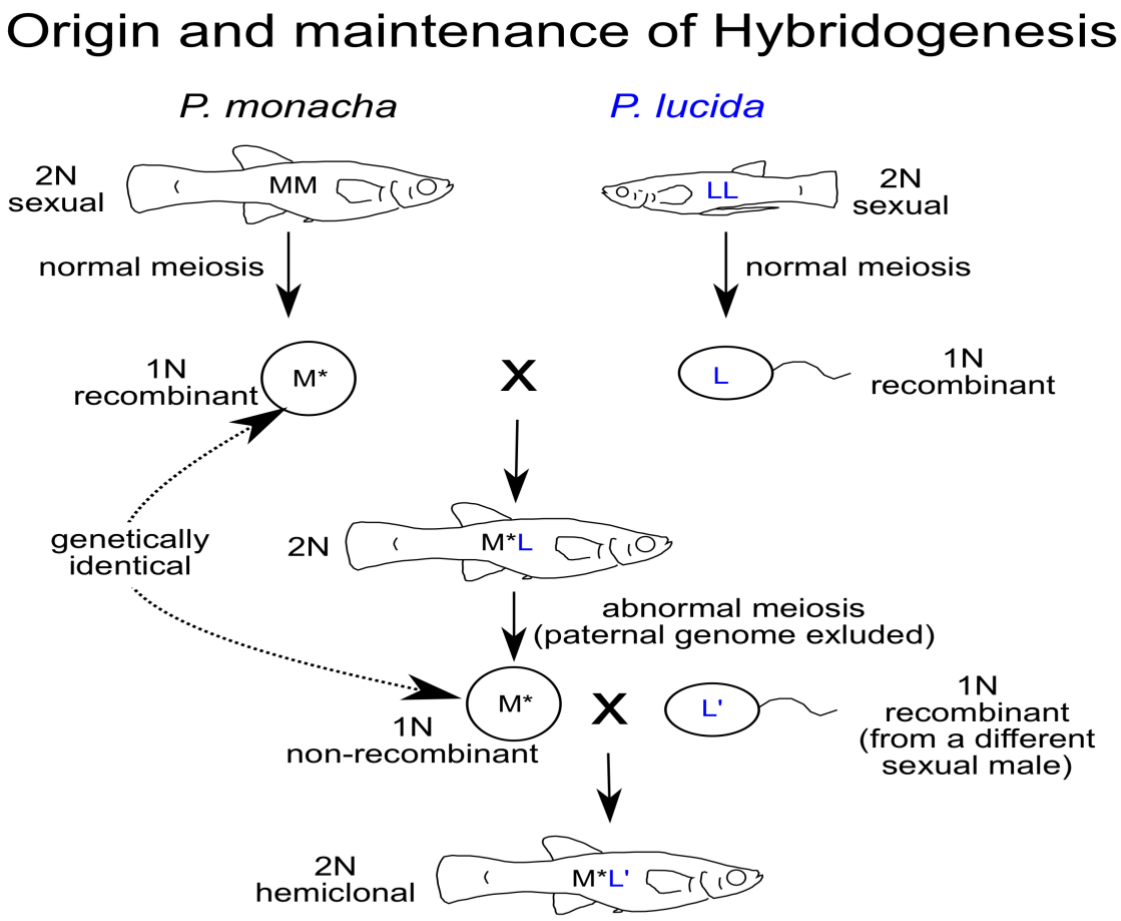

Figure 1 Illustration of the origin and maintenance of hybrid genesis. Freshwater fish biotype *Poeciliopsis monacha-lucida* (ML) is used. A diploid sexual *P. monacha* (MM) female produces a recombinant haploid egg (M) through normal meiosis. A diploid sexual *P. lucida* (LL) male produces a recombinant haploid sperm (L) through normal meiosis. The *P. monacha* (M) egg is fertilized by the *P. lucida* (L) sperm, producing a diploid female F1 hybrid *P. monacha-lucida*. Through abnormal meiosis that excludes paternal (L) chromosomes, this ML female produces a haploid egg containing only the maternal (M) genome, which is therefore genetically identical to the egg produced by her mother (asterisks used to emphasize that they are genetically identical). This (M) haploid egg is then fertilized by a sperm from *P. lucida* (different superscripts represent different sperm/males from *P. lucida*), producing a diploid female *P. monacha-lucida*. *P. lucida* males transfer their genetic material to their daughters but not to their granddaughters. Black font inside fish or gamete symbols depicts *P. monacha* chromosome sets; blue font depicts *P. lucida* chromosome sets.

In the freshwater live-bearing fish genus *Poeciliopsis*, three confirmed (*P. monacha-occidentalis, P. monacha-lucida* and *P. monacha-latidens*) and one putative (*P. monacha-jackschultzi*) hybridogenetic biotypes (all diploid) are known. In addition, three types of gynogenetic biotypes (all triploid) are known (*P. monacha-monacha-lucida, P. monacha-lucida-lucida,* and *P. monacha-lucida-viriosa*), which appear to have arisen by addition of a genome from *P. monacha*, *P. lucida,* and *P. viriosa*, respectively, to an unreduced (2n) *P. monacha-lucida* (ML) egg (Mateos & Vrijenhoek, 2005). Previous studies based on allozymes and mitochondrial DNA have revealed that the maternal progenitor of all asexual hybrids of *Poeciliopsis* examined to date is the sexual species *P. monacha* (Mateos & Vrijenhoek, 2005; Quattro et al., 1991). Whereas *P. monacha* has a very restricted distribution (i.e., headwaters of three adjacent rivers: Rio Mayo; Rio Fuerte and Rio Sinaloa; Figure 2), the asexual hybrids collectively have a much broader range, where they depend on sperm from one of the following sexual species for reproduction: *P. occidentalis*[1], *P. jackschultzi*, *P. lucida,* and *P. latidens.*

In this study, we focused on the evolutionary origin of *P. monacha-latidens*, also referred to as "spotted hybrids" because of their characteristic pigmentation pattern. Whereas *P. latidens* have black vertical bars and spots alongside the body, *P. monacha-latidens* (M-lat) only have spots alongside the body (Schultz, 1969). The evolutionary origins of *P. monacha-lucida* (ML) and *P. monacha-occidentalis* (MO) have been investigated (Mateos & Vrijenhoek, 2002, 2005; Quattro et al., 1991; Quattro et al.,

[1] In this thesis, for simplicity our use of "*P. occidentalis*" is in the *sensu lato* sense. i.e., including both *P. occidentalis sensu stricto* and *P. sonoriensis*, which are recognized as separate species (Miller et al., 2005). Similarly, *P. monacha-occidentalis* also includes *P. monacha-sonoriensis*.

1992), but *P. monacha-latidens* has received little attention. We hereby report on the development of a DNA-based genotyping assay to identify *P. monacha-latidens* (Chapter 2), the use mitochondrial phylogenomics to infer the evolutionary origin of *P. monacha-latidens*, and provide new insights into the evolution of *Poeciliopsis* asexual biotypes (Chapter 3).

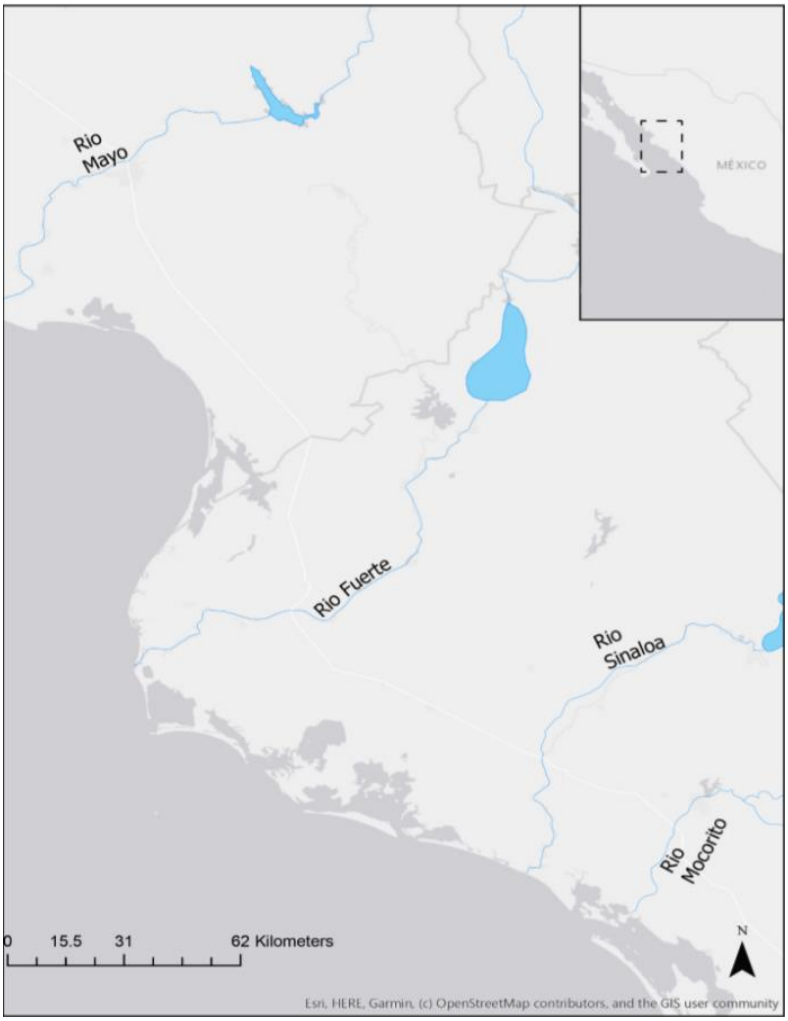

Figure 2 Location of relevant rivers for this study in Northwest Mexico. Dashed rectangle in the inset indicates the location of the relevant rivers.

CHAPTER II DNA-BASED GENOTYPING ASSAY FOR IDENTIFICATION OF POECILIOPSIS MONACHA-LATIDENS

2.1 Introduction

The hybrid nature of asexual vertebrate taxa poses challenges to their identification, as they typically have phenotypes that are intermediate between, but overlapping with, their parental progenitor species. Early studies used morphological traits, lab crosses, and evidence of all-female offspring to discover asexual hybrids (e.g. the identification of the Amazon molly, *Poecilia formosa*) (Hubbs & Hubbs, 1932). The advent of molecular markers including allozymes, RFLPs, microsatellites, and DNA sequencing greatly enhanced the ability to genotype hybrids and identify their parental progenitors (Crespo-López et al., 2006; Gómez & Carvalho, 2000). The initial discovery and determination of reproductive mode and ploidy levels of *Poeciliopsis* gynogenetic and hybridogenetic biotypes by R.J. Schultz relied on a combination of crosses and examination of morphology and chromosome numbers (Schultz, 1961, 1967, 1969; Schultz, 1966). This endeavor benefited in particular from the use of crosses with *P. latidens*, the species of *Poeciliopsis* that is most distinct at numerous traits (i.e., genitalia, dentition, lips, ova size, and pigmentation) from the other species that serve as progenitors of most asexual biotypes (i.e., *P. monacha*, *P. viriosa*, *P. lucida* and *P. occidentalis*) (Schultz, 1966). Morphology alone, however, is insufficient for diagnosis of all *Poeciliopsis* hybrids (Lima et al., 1996; Schultz, 1961; Schultz, 1966; Vrijenhoek & Schultz, 1974), and it was not until the application of allozymes (Vrijenhoek, 1972,

1975; Vrijenhoek, 1979) and mtDNA markers (Quattro et al., 1991; Quattro et al., 1992) that the maternal and paternal progenitors of hybrids could be identified.

Although allozyme markers were instrumental for distinguishing the different sexual and asexual biotypes of *Poeciliopsis*, they have several drawbacks. First, because proteins are generally not as stable as DNA (Hoy, 2003), allozyme genotyping requires fresh or frozen tissue, thereby precluding the use of ethanol-preserved specimens. In many regions, including northern Mexico where asexual *Poeciliopsis* occur, it has become increasingly difficult to obtain dry ice (M. Mateos, pers. comm.). Furthermore, air travel regulations implemented after 2001, prohibit travelling with dry ice in the checked baggage. Finally, dry ice shipping between the US and Mexico is also banned (M. Mateos, pers. comm.). Secondly, different allozyme alleles may not reflect differences at the nucleotide level, as demonstrated by Olsen et al. (2014) in fin whales. Therefore, development of DNA-based genotyping assays is highly desirable. Single nucleotide polymorphisms (SNPs) represent a particularly tractable DNA-based marker.

Discovery of SNPs in a diversity of model and non-model species has been aided greatly by High-Throughput DNA Sequencing (HTS) technologies (e.g. Illumina short reads) (Goodwin et al., 2016; Metzker, 2010) SNP genotyping can be achieved through a variety of approaches, including those based on HTS of entire genomes (whole genome shotgun sequencing) or reduced representation sequencing, where a subset of the genome is selected/enriched prior to HTS (e.g. RADseq, probe capture, etc.) (Andrews et al., 2016). SNP genotyping methods that do not rely on HTS, include microarray-based (Cutler et al., 2001) and PCR-based approaches. Whereas micro-array based

assays are generally employed for genome-scale genotyping, genotyping of a small number of SNPs (e.g. for specimen identification or genotyping a small number of SNPs), might be best achieved with PCR-based approaches. Currently available PCR-based SNP genotypic approaches include: Allele-specific PCR (Liu et al., 2012), Sanger-sequencing-based genotyping, KASP (KBiosciences Competitive Allele-Specific PCR) (Semagn et al., 2014); TaqMan assay (Woodward, 2014); semi-thermal asymmetric reverse PCR (Long et al., 2017); Amplifluor SNP genotyping system (Jatayev et al., 2017); mini-sequencing (Syvänen et al., 1990); and high resolution melt (Taylor et al., 2010). These methods require different platforms and have their own disadvantages. For example, mini-sequencing can achieve high accuracy but is laborious and expensive (Long et al., 2017). Allele-specific PCR, however, can amplify PCR products specific to the SNPs within the 3' end of allele-specific primers and does not require a unique platform. Therefore, it can be an efficient way for genotyping with low cost and minimal steps (Gaudet et al., 2009). However, it has its own drawbacks. One of the challenges of using allele-specific PCR is achieving reliable discrimination, which sometimes can be improved by introducing mismatches within the three bases closest to the 3' end, and optimizing reagent concentrations and temperature profile (Liu et al., 2012). The present study aimed at optimizing the allele-specific PCR approach for *Poeciliopsis* hybrids, and reported on the application of allele-specific PCR to diagnose different sexual and asexual biotypes of *Poeciliopsis*.

2.2 Materials and methods

2.2.1 Sampling

We selected 24 samples to test primer combinations, including sexual species and asexual biotypes, of which 20 were field-collected and identified by allozymes, Sanger-sequencing of a mitochondrial gene, and/or morphology, and four were lab strains (Table 1, Figure 3).

2.2.2 Mitochondrial gene sequences to identify maternal progenitor species

Genomic DNA was extracted from seven specimens (Table 1) that had not been previously genotyped with the DNEasy Tissue extraction kit (Qiagen, Inc., Valencia, CA). The *Cytochrome b* (*cytb*) gene was amplified using the following protocol: initial denaturation at 94°C for 2 min,

Table 1 Specimens examined in Chapter II. Samples examined with the Allele Specific PCR (ASP) with primer combination 3, aimed at detecting the *Ck-A* allele of *P. latidens*: positive (+) and negative (–) amplifications based on presence/absence of band are indicated. A checkmark in the "*Cytb*" column indicates that sample was amplified and Sanger-sequenced for the whole/partial mitochondrial *Cytochrome b* gene in this or a previous study. A checkmark in the "*SH3PX3*" column indicates that the sample was amplified and Sanger-sequenced for the *SH3PX3* gene in the present study. WGS = sample was subjected to whole genome shotgun sequencing (see Chapter 3). "pigment" = pigmentation pattern was used as one means of identification. "allozyme" = at least one allozyme locus was used as a means of identification. "mt genome" = whole mitochondrial genome was assembled and analyzed (see Chapter 3). Species/Biotype was achieved by at least two of the above identification methods.

Num -ber	Species	Voucher	ASP	*Cytb*	*SH3 PX3*	WGS	Identification method
1	*P. latidens*	MVH99-17#5	+	√	√	√	P, MT
2	*P. monacha*	V87-3/4	–				A, Crosses
3	*P. monacha-latidens*	MVH99-15#10	+	√	√	√	P, MT
4	*P. monacha-latidens*	MV00-6A#10	+	√	√	√	P, MT
5	*P. monacha-latidens*	MVH99-17#72	+			√	P, A, MT
6	*P. monacha-latidens*	MVH99-16#59	+			√	P, A, MT
7	*P. monacha-latidens*	MVH99-17#61	+	√		√	P, A, MT
8	*P. monacha-latidens*	MV00-10#45	+			√	P, MT
9	*P. monacha-lucida*	MVH99-12#3	–	√[a]		√	P, A, MT
10	*P. monacha-lucida*	MVH99-18#3	–	√[a]		√	P, A, MT
11	*P. monacha-lucida*	Cz-M61-35	–	√[a]		√	A, MT, Crosses
12	*P. monacha-lucida*	Cx-I-MVH99-16#5	–	√[a]		√	A, MT
13	*P. monacha-lucida*	Cx-II-MVH99-16#48	–	√[a]		√	A, MT

Num -ber	Species	Voucher	ASP	*Cytb*	*SH3 PX3*	WGS	Identification method
14	*P. monacha-lucida*	Cx-SV73-7s	–	√[a]		√	A, MT, Crosses
15	*P. monacha-lucida*	MVH99-18#1	–	√[a]		√	P, A, MT
16	*P. latidens*	VD81-46#1	+	√[b]			P
17	*P. monacha*	MVH99-9.6 D+	–	√[b]			A, Crosses
18	*P. lucida*	MVH99-11#2	–	√[b]		√	A, MT
19	*P. viriosa*	VD81-47#1	–				P
20	*P. presidionis*	MV00-8#31	–	√[b]			P
21	*P. prolifica*	VL78-4#74	–	√[b]			P, A
22	*P. occidentalis*	AV76-7#M1	–			√[c]	P, MT
23	*P. scarlii*	MH00-3#1	–	√[b]			P
24	*P. infans*	VD81-42#2	–	√[b]			P
n/a	*P. monacha-latidens*	MVH99-17#64	n/a	√			A
n/a	*P. monacha-latidens*	MV00-6A#11	n/a	√		√	P, MT
n/a	*P. latidens*	MVH99-17#2	n/a	√	√		P, MT
n/a	suspected *P. viriosa* X *P. latidens* male		n/a	√	√		P

[a] Mateos and Vrijenhoek (2002); [b] Mateos et al. (2002); [c] Mateos et al. (2019b); [U] unpublished data.
Table 1 continued

35 cycles of 0.5 min at 94°C , 1 min at annealing temperature 57°C, and 1 min at 72°C, and a final extension at 72°C for 5 min (Mateos & Vrijenhoek, 2002) and Sanger-sequenced at Eurofins (Luxembourg, Luxembourg). Forward and reverse strands were assembled and edited in Geneious Prime 2021.0.1 (https://www.geneious.com). The consensus sequences were then blasted against standard database by blastn (Altschul et al., 1990) to identify their maternal progenitor species.

Figure 3 Photograph of three lab-born males of *P. viriosa* and *P. latidens*. The *cytb* and *SH3PX3* genes were amplified and sequenced for one of them. The yellowish coloration and the black dorsal fin margin are characteristic of *P. viriosa*; the black wedge above the gonopodium is characteristic of *P. viriosa* and *P. monacha* (Miller, 1960). The round spots are characteristic of *P. latidens* and are present in *P. monacha-latidens*. Although *P. latidens* tends to have an alternation of vertical bars and round spots, the vertical bars do not express in *P. monacha-latidens* (Schultz, 1967, 1977). Photo courtesy: Samantha Levell and David Reznick.

2.2.3 Allele-specific PCR of one nuclear gene to identify presence or absence of a *P. latidens* allele

We targeted a *creatine kinase* gene (*Ck-M*) that is predominantly expressed in muscle. This locus was previously characterized in allozyme studies of *Poeciliopsis*, where it was referred to as *Ck-A* as well as "Muscle Protein 3" because it was also recovered with a simple Ponceau red stain (Vrijenhoek et al., 1992). We used 16 partial cDNA sequences of this locus based on total RNA isolates from muscle (GenBank Accession No. MW678631–MW678646). Intron-exon boundaries were identified by alignment with the *P. occidentalis* genome (PRJNA532900) (Mateos, Kang, et al., 2019). We identified positions in this gene that appear diagnostic among several species of interest, including *P. latidens* and *P. monacha* (Figure 4). These positions were used to design primers whose 3' end sits at a species-diagnostic single nucleotide polymorphism (SNP) (Cha et al., 1992). Positive amplification, as visualized via agarose gel electrophoresis stained with GelRed or Ethidium Bromide, should reflect exact match between the primer and the template. Work in other organisms has shown that introducing one mismatch at the third base from the 3' end can achieve a better discrimination (Drenkard et al., 2000). Therefore, to increase our ability to discriminate, we designed primers with such mismatches (see Table 2). The combinations of primers that we initially targeted (amplicon length range: 86–113bp) are listed in Table 3, but given unsuccessful preliminary attempts with three of these combinations (see Results), most of our efforts were targeted at combination 3, expected to amplify the *Ck-M* allele of *P. latidens* (and its sister *P. fasciata*), but not of any other species of *Poeciliopsis*.

Combination 3 was tested on a variety of samples (Table 1) with both Apex Taq RED Master Mix, 2X (Apex Bioresearch Products, Houston, CA), and HiDi Polymerase (#9001, MyPOLS Biotec, Kontanz, Germany), using the PCR thermocycler model T100 (Bio-Rad Laboratories, Hercules, CA). All PCRs included negative controls (water instead of DNA) and positive controls. Positive controls were developed from the 1/50 dilution of PCR amplicon of *P. latidens* with combination 3.

Table 2 The sequences of the primers for the *Ck-M* gene. Where two primer sequences are shown, the lower one has a mismatch indicated by boldface with respect to the top primer, which represents a perfect match to the template. Universal (blue font) = should amplify all species of *Poeciliopsis*.

Primer	Sequence (5' → 3')
CK_M_exon1_F1 (universal)	GAGTTTCCAGACCTGTCCAAGC
CK_M_Exon_1_monacha_specific_R1	GTGTAACCACTTGGTGTCTGCC (GTGTAACCACTTGGTGTCTTCC)
CK_M_Exon_1_Leptoraphis_specific_R1	GTGTAACCRCTTGGTGTCTGCT (GTGTAACCRCTTGGTGTCTTCT)
CK_M_Exon_1_fasciata_latidens_only_R1	ACATCATCCAAAGTGTAACCGCTC (ACATCATCCAAAGTGTAACCGATC)
CK_M_exon7_all_speceis_F1 (universal)	CACTGCRTCTGTGGGTGGCG
CK_M_exon7_new_sp_only_R1	CAACCACCAGCTGGACCTGCG

Table 3 Details about the four combinations of primers (listed in Table 1) initially tested, the species they are expected to amplify, and the expected amplicon length. Universal (for *Poeciliopsis*) primers are shown in blue font.

Combination Number	Forward Primer	Reverse Primer	Species	Amplicon Length
1	CK_M_exon1_F1	CK_M_Exon_1_monacha_specific_R1	*P. monacha* *P. viriosa*	101
2	CK_M_exon1_F1	CK_M_Exon_1_Leptoraphis_specific_R1	*P. prolifica* *P. jackschultzi* *P. occidentalis* *P. fasciata* *P. baenschi* *P. gracilis* *P. infans* *P. scarlii* *P. presidionis* *P. turneri* *P. elongata*	101
3	CK_M_exon1_F1	CK_M_Exon_1_fasciata_latidens_only_R1	*P. latidens* *P. fasciata*	113
4	CK_M_exon7_all_speceis_F1	CK_M_exon7_new_sp_only_R1	*P. jackschultzi*	86

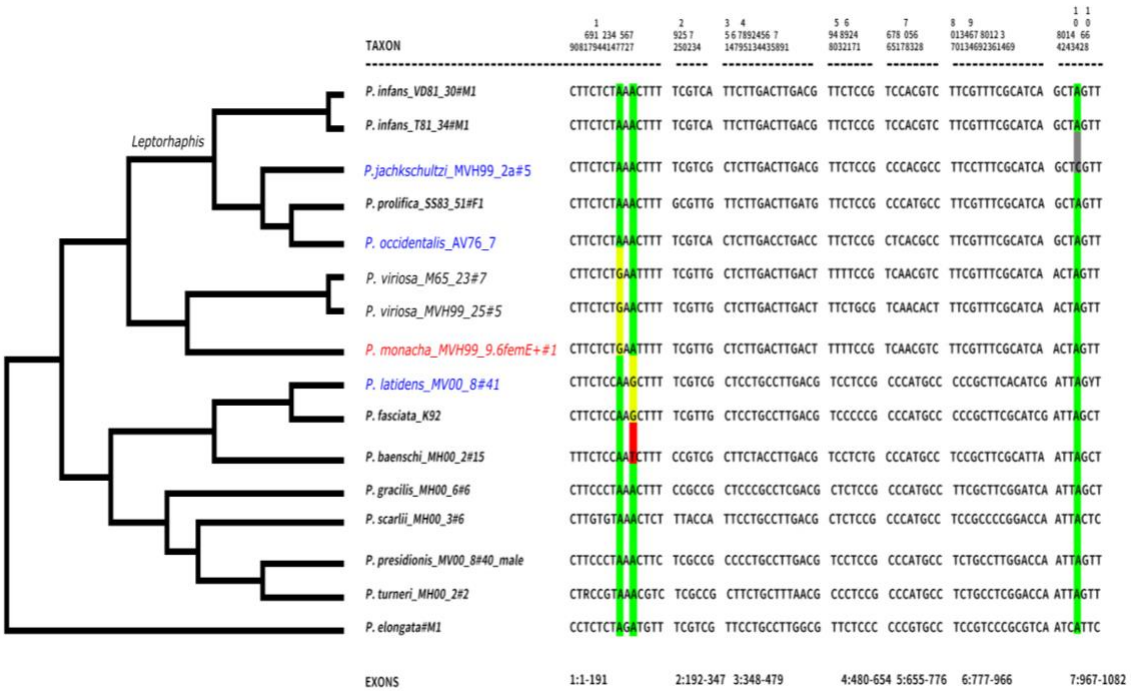

Figure 4 SNPs at the *Ck-M* locus in representatives of *Poeciliopsis* (i.e., all taxa except *P. elongata*, a representative of the sister subgenus *Aulophallus*), based on sequences derived from muscle cDNA (GenBank Accession No. MW678631–MW678646). Highlighted SNPs are those targeted by the diagnostic primers of Table 2. Exons 1–7 are separated by spaces and their coordinates are indicated in the last row. The annotated genome of *P. occidentalis* (Mateos, Kang, et al., 2019) was used to identify the exon boundaries. A phylogram is drawn on the left (Mateos, Domínguez-Domínguez, et al., 2019). Red font taxon label is *P. monacha*, the maternal progenitor of all asexual hybrid biotypes of *Poeciliopsis*. Blue font taxon labels indicate sexual species that serve as natural sexual hosts for hybridogenetic biotypes; the only sexual host species missing is *P. lucida* (a member of *Leptorhaphis*).

2.2.4 PCR amplification and Sanger sequencing of one nuclear gene to distinguish *P. monacha*, *P. latidens*, and their hybrids

To verify the results from the mitochondrial and allele-specific assay that diagnose *P. monacha-latidens*, and to test an additional nuclear genotyping approach, we also examined published sequences of the *SH3PX3* nuclear gene to identify SNPs that distinguish *P. monacha* GenBank Accession No. KJ697543 (and its sister lineage *P. viriosa* KJ697552) from *P. latidens* KJ697541 (and its close relatives *P. fasciata* KJ697538 and *P. baenschi* KJ697534). We then amplified this gene (forward 5'-GACATGCTGGAGTTTCAGGA-3', reverse 5'-ACTTGTTRGCMACTGGGTCAAA-3') using the following protocol: initial denaturation at 94°C, for 2 min, 35 cycles of 0.5 min at 94°C, 1 min at annealing temperature 57°C, and 1 min at 72°C, and a final extension of 5 min at 72°C (Li et al., 2007). PCR amplicons from five samples were subjected to Sanger-sequencing. We expected double chromatogram peaks in the *P. monacha-latidens* samples at the positions that differ between *P. monacha* (KJ697543) and *P. latidens* (KJ697541).

In addition, to verify the expected genotypes at the *SH3PX3* gene, we used Bowtie2 v.2.4.2 (Langmead et al., 2009) to map the Illumina sequences obtained in Chapter 3 to the published *SH3PX3* sequence of *P. monacha* (KJ697543). The resulting mapping files were visually inspected in Geneious Prime.

2.3 Results

2.3.1 Identification of maternal progenitor species with mitochondrial markers

The blast results of the seven newly assembled *cytb* gene Sanger sequences revealed that the closest hits of five specimens (MVH99-15#10, MVH99-17#61, MVH99-17#64, MV00-6A#10 and MV00-6A#11) were sequences assigned to *P. monacha* or its derived asexual hybrids (range of differences 0–47), indicating that their maternal ancestor was *P. monacha* (a mitogenomic phylogeny containing these samples is shown in Chapter 3). The closest hit of specimens MVH99-17#5 and MVH99-17#2 were assigned to *P. latidens* (range of differences 3–25), indicating that the maternal progenitor of these specimens was *P. latidens*. The closest blastn hits of *cytb* sequence (1121 bp) of the suspected *P. viriosa* x *P. latidens* hybrid were two specimens assigned to *P. viriosa* (range of differences 9–15), confirming that the maternal progenitor was *P. viriosa*. For the specimens whose mitochondrial genome was not fully assembled in Chapter 3, we have deposited their *cytb* sequences under GenBank Acc. Nos. MZ576568-MZ576574).

2.3.2 Allele-specific (high discrimination) PCR of the *Ck-M* nuclear gene

The four combinations were tested on a set of samples that were previously identified via morphology, allozymes, lab breeding experiments, and/or mitochondrial gene sequences (Table 4). The temperature profile was as follows: initial denaturation at 96°C for 2 min, 35 cycles of 0.5 min at 96°C, 0.5 min at annealing temperature 61°C, and 0.5 min at 72°C, and a final extension at 72°C for 5 min. As expected based on the diagnostic primer“CK_M_Exon_1_fasciata_latidens_only_R1”, combination 3 consistently amplified a single band close to the 100 bp size marker (expected amplicon

length = 113 bp), in specimens of *P. latidens*, *P. fasciata*, or *P. monacha-latidens*, and consistently failed to amplify all other *Poeciliopsis* specimens tested (Figure 5, Table 4). We also tested, and successfully amplified, combination 3 on a pure *P. latidens* sample at a variety of annealing temperatures (i.e., 57, 58 , 59, and 60°C). In addition, we tested, and successfully amplified, combination 3 using both Apex Taq RED Master Mix and HiDi polymerase (results not shown).

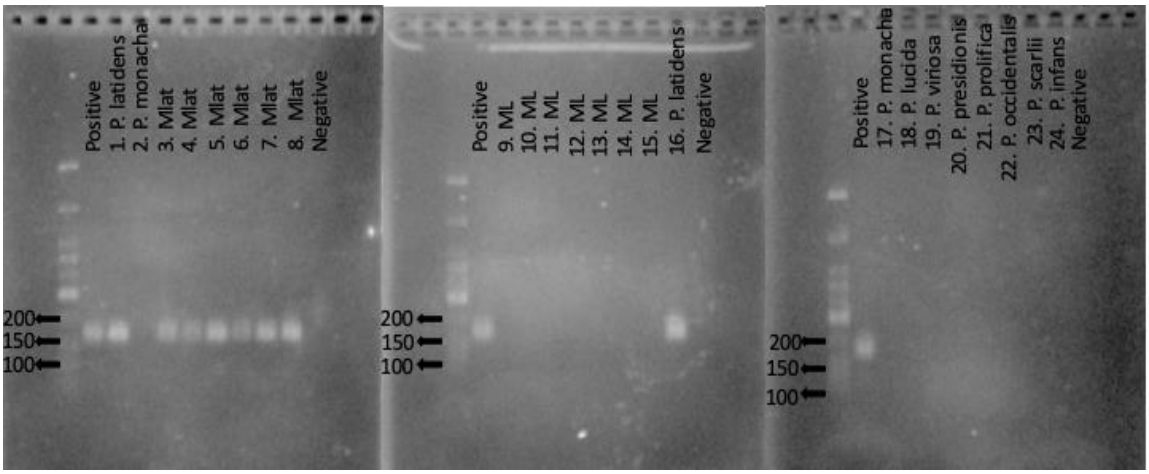

Figure 5 PCR analysis for combination 3. Mlat represents *P. monacha-latidens* and ML represents *P. monacha-lucida*. 100, 150, 200bp ladder bands are indicated by arrows (Low Molecular Weight DNA Ladder, Cat. No. 3233S, New England BioLabs, Inc., Ipswich, MA). Numbers 1–24 correspond to the sample numbers in Table 4. Apex Taq RED Master Mix was used for this PCR.

Our work with the other three primer combinations (combinations 1, 2 and 4) was unsuccessful despite attempts with different annealing temperatures (not shown). These combinations exhibited at least one of the following undesirable outcomes: positive amplifications yielded more than one band size implying amplification of more than one locus; failure to diagnose species-specific alleles (e.g. positive amplification of non-target species/biotypes, or negative amplification of target species/biotypes); and inconsistent results across replicated PCR runs.

2.3.3 Sanger chromatograms of the *SH3PX3* gene to diagnose *P. monacha, P. latidens*, and their hybrids

Based on previously published sequences of the *SH3PX3* gene, we identified eight SNPs that distinguish *P. monacha* + *P. viriosa* vs. *P. latidens* + *P. fasciata* + *P. baenschi* (positions: 34, 115, 358, 508, 511, 529, 572, and 592; Figure 6). In addition, one position (370) was distinct in *P. monacha*, and three positions (376, 658, and 676) were distinct in *P. latidens* vs. *P. monacha* + *P. viriosa*. As proof of concept and to validate the identifications of *P. monacha*, *P. latidens*, and *P. monacha-latidens* based on pigmentation, mitochondrial sequences and the *Ck-M* allele-specific PCR, we amplified and Sanger-sequenced *SH3PX3* sequences from a subset of specimens (Table 4). With the exception of position 376 (addressed below), chromatogram peaks were as expected: a single peak matching the previously sequenced *P. latidens* specimen for the two putative pure *P. latidens*; double peaks reflecting the *P. monacha* and *P. latidens* alleles in the two putative *P. monacha-latidens* specimens; and double peaks reflecting the *P. viriosa* and *P. latidens* alleles in the putative *P. viriosa* X *P. latidens* male. For position

376, the two putative *P. monacha-latidens* specimens had evidence of double peaks (A/G), but the putatively pure *P. latidens* had G (i.e., the base previously reported in *P. monacha*, *P. viriosa*, and *P. fasciata*) instead of the expected A (previously reported in *P. latidens* and *P. baenschi*). The putative *P. viriosa* X *P. latidens* hybrid appeared homozygous for G. These results suggest that position 376 segregates for A/G in *P. latidens*. Indeed, mapping the raw Illumina reads of the only *P. latidens* sample examined revealed a G in that position. Furthermore, as expected *P. monacha* and *P viriosa,* had only one allele matching previously sequenced specimens at the 12 positions listed above, and *P. monacha-latidens* had two alleles that matching previously Sanger results at those positions.

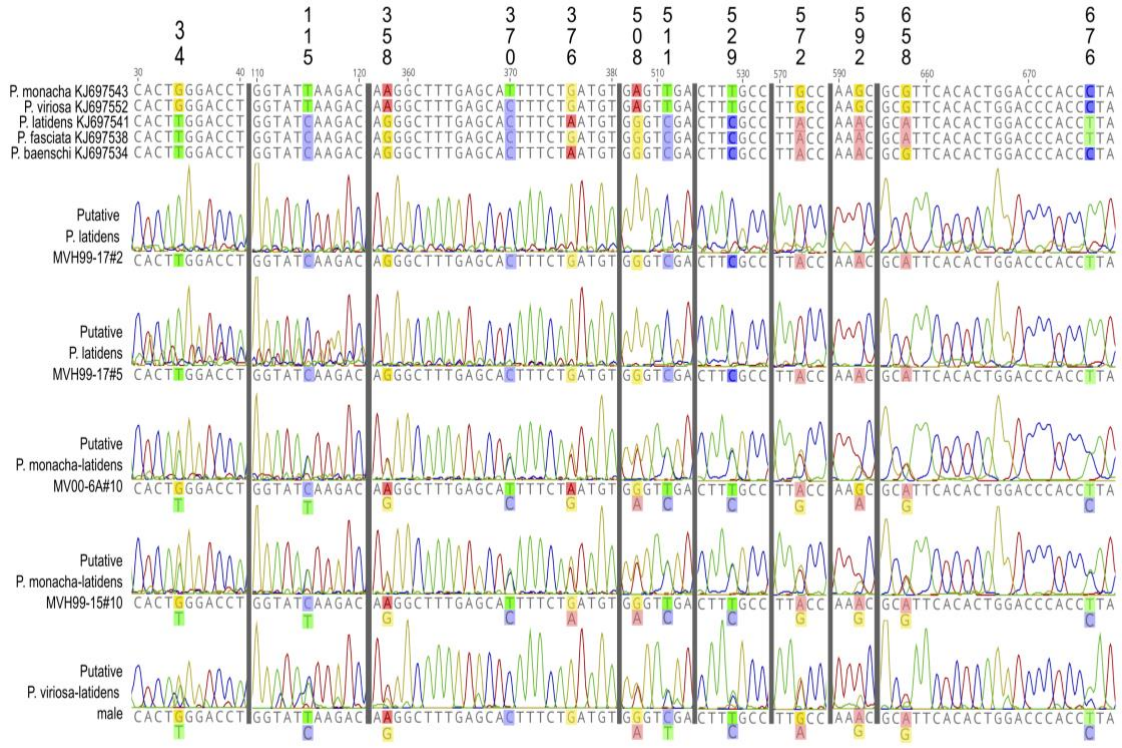

Figure 6 SNPs and chromatograms. Twelve *SH3PX3* SNPs of relevance to distinguishing *P. monacha*, *P. latidens*, and *P. viriosa*. Rows lacking chromatograms are sequences from prior studies (GenBank Accessions are indicated). The rows depicting chromatograms are the newly generated sequences; letters under them are our inferred allele composition (vertical order of bases in inferred heterozygotes is arbitrary). For reference, a few bases before and after each SNPs are also shown to show sequence quality in that region. Vertical gray bars indicate boundaries of the depicted sequence blocks. Numbers above sequences indicate nucleotide position of the SNPs.

2.4 Discussion

Distinction of asexual vertebrate biotypes from their sexual relatives, as well as identification of different clonal lineages, has commonly relied on a combination of mitochondrial gene sequencing and genotyping of nuclear markers, such as microsatellites [e.g. *Fundulus* (Hernández Chávez & Turgeon, 2007) and *Cobitis* (Janko et al., 2012)], allozymes [e.g. *Poeciliopsis* ((Mateos & Vrijenhoek, 2005; Quattro et al., 1991)], or a combination of both [e.g. the frog hybrid complex *Pelophylax esculentus* (Christiansen, 2005; Doležálková-Kaštánková et al., 2021; Tejedo et al., 2000)]. Because of the drawbacks of allozymes, the lack of DNA-based nuclear genotyping for *Poeciliopsis* represents a limitation for examining ethanol-preserved specimens or specimens that have not been allozyme-genotyped. We explored the use of allele-specific PCR to genotype several *Poeciliopsis* asexual biotypes and their sexual relatives. Of our attempted allele-specific PCR primer combinations targeting an exon of the *Ck-M* gene, only one was deemed successful. At the annealing temperature tested (61°C), primer combination 3 consistently amplified a fragment of the target size (i.e., close to the 100bp marker) in every specimen that carried at least one set of chromosomes derived from *P. latidens*, but not in any of the other specimens (Figure 5, Table 4). Successful diagnosis of the *P. latidens* allele occurred with both polymerases tested. In addition, our successful amplification of the pure *P. latidens* sample tested at several lower annealing temperatures (57, 58, 59, and 60°C) suggests that this assay is reliable even under slight variations of the PCR conditions. Our successful diagnosis of the *P. latidens* allele with the less costly polymerase that is not High Discrimination is

encouraging. We consider that although the allele-specific PCR approach is feasible in cases such as combination 3, our failure to successfully optimize the other primer combinations implies that the success of this approach is very dependent on the particular SNPs targeted and may require substantial optimization effort. For the *Poeciliopsis* system, future application of allele-specific PCR for diagnosis of other species and hybrid biotypes would benefit from targeting other loci. The allele-specific PCR approach might be a desirable one where access to Sanger sequencing is limited. Where Sanger sequencing is readily available, using "universal" primers to amplify and sequence a locus containing species-diagnostic SNPs, such as *SH3PX3*, may be more practical.

A particular challenge of the *Poeciliopsis* system that allozyme genotyping helped overcome is the co-occurrence of diploid (hybridogenetic) and triploid (gynogenetic) biotypes. Distinction of these requires knowledge about genome dosage. Several of the allozyme loci developed for *Poeciliopsis* allowed distinction of ML, MML and MLL based on the levels of expression of the *P. monacha* and *P. lucida* alleles (Vrijenhoek, 1975). Other sources of evidence of triploidy were the detection of three different alleles in a single specimen (e.g. in MLV at loci diagnostic for *P. monacha*, *P. viriosa*, and *P. lucida*) (Mateos & Vrijenhoek, 2005). The allele-specific PCR and Sanger sequencing approaches described above are not designed to diagnose allele dosage, but may be useful for tri-allelic SNPs in triploids. Methods that could be optimized for determination of allele dosage in bi-allelic SNPs include High Resolution Melting analysis (e.g. as applied to polyploid plants) (Han et al., 2012), and competitive allele-

specific PCR amplification (KASPar ™) as in (Cuenca et al., 2013), both of which require equipment to measure fluorescence (e.g. quantitative PCR). Albeit more effort-intensive, distinction of ML, MML and MLL may be feasible by comparing dosage of diagnostic genes from whole genome shotgun sequencing data.

2.4.1 Discovery of *P. viriosa* x *P. latidens* hybrids

Collectively, the intermediate pigmentation (e.g., black wedge and yellow pigmentation from *P. viriosa*, and spots from *P. latidens*), the *P. viriosa* mitochondrial gene, and the heterozygosity of the *SH3PX3* gene at the diagnostic *P. viriosa* vs. *P. latidens* SNPs suggest that this male is a hybrid. Although this male was lab-born, it is unclear whether its wild-caught mother was inseminated by *P. latidens* in the field or in the lab. A recently collected female specimen from Sinaloa also has an intermediate pigmentation between *P. viriosa* and *P. latidens* (Figure 7), but adequate DNA for genotyping was not available. These observations raise the possibility of introgression between *P. viriosa* and *P. latidens*. Indeed, R.J. Schultz suspected such a scenario based on his observation that a small proportion (~10%) of the hundreds of *P. latidens* males from the Rio Mocorito he examined had dentition more similar to that of *P. viriosa* (R.J. Shultz, pers. comm. to M. Mateos 2001), and consequently also *P. monacha*. Genome-wide studies of preserved or fresh collections should enable assessment of hybridization and introgression, and whether frequency of hybridization between *P. viriosa* and *P. latidens* is more common at present than historically (e.g. the pollution-mediated hybridization in *Xiphophorus*) (Fisher et al., 2006). Lab crossing experiments should allow determination of degree of reproductive compatibility between these two species.

Figure 7 Wild-collected female from Meseta de Cacaxtla Sinaloa (23°35.386 N, - 106°41.778 W; small pond below Toyhua bridge of Mazatlán-Culiacán Hwy 15D) exhibiting intermediate pigmentation patterns between *P. viriosa* and *P. latidens*; collection date 1 November 2017. Courtesy of Albert van der Heiden (CIAD, Mazatlán, Mexico).

CHAPTER III EVOLUTIONARY ORIGIN OF POECILIOPSIS MONACHA-LATIDENS BASED ON MITOGENOMICS

3.1 Introduction

The fish genus *Poeciliopsis* has served as a useful system to address questions about the evolution of sex (Lively et al., 1990), because of the occurrence of several asexual forms that arose via hybridization (Vrijenhoek, 1994). Previous studies have investigated the origin of the hybridogenetic biotypes *P. monacha-lucida* and *P. monacha-occidentalis*, both of which are evolutionarily young compared to the age of the sexual lineages that gave rise to them (Mateos & Vrijenhoek, 2002; J. M. Quattro et al., 1991; Vrijenhoek et al., 1977, 1978). Vrijenhoek et al. (1977) found the highest allozyme haplotypic diversity in *P. monacha-occidentalis* in the Rio Mayo (its southernmost limit) consistent with multiple hybrid origins, where both progenitor species overlap. Based on mtDNA, Quattro et al. (1992) found evidence for the monophyletic origin of the *P. monacha-occidentalis* in the four northern rivers, Rio Yaqui, Rio Matape, Rio Sonora, and Rio Concepcion, consistent with the stepwise colonization of a single unisexual lineage from Rio Mayo. Quattro et al. (1991) found two types of *P. monacha-lucida* hemiclones widely distributed and some stream-endemic hemiclones in the Rio Fuerte, suggesting multiple asexual hybrid origin events happened in the Rio Fuerte, where both *P. monacha* and *P. lucida* co-occur. Mateos and Vrijenhoek (2002) inferred that *P. monacha-lucida* from Rio Mocorito, where *P. monacha* is absent, is derived from the Rio Fuerte or Rio Sinaloa.

The origin of *P. monacha-latidens*, found in the Rios Fuerte, Sinaloa, and Mocorito, has not been previously examined. Its range overlaps with that of *P. monacha-lucida*, raising the possibility that the two biotypes represent the same hemiclonal (*monacha*) genome, differing only by the sperm that fertilizes them and whose genetic material is incorporated and expressed for a single generation. Of the sexual hosts of asexual *Poeciliopsis*, *P. latidens* is the most distinct genetically, morphologically, and ecologically. Accordingly, comparison of *P. monacha-latidens* and *P. monacha-lucida* "populations" suggests that they are governed by distinct dynamics (Lanza, 1983). For example, *P. monacha-lucida* is suggested to undergo greater sperm limitation than *P. monacha-latidens*. In contrast, *P. monacha-lucida* does not appear to compete with *P. lucida* in terms of habitat and resource use, but *P. monacha-latidens* loses in competition for food with *P. latidens* under food shortage. Another interesting feature of *P. monacha-latidens* is that it is the only biotype where males occur, albeit at low frequencies, both in the field (M. Mateos, pers. comm.) and in the lab (6.7% of offspring born to one female) (Schultz, 1966). Indeed, it was this feature that contributed to the identification of the maternal progenitor prior to the availability of allozyme or DNA genotyping (Schultz, 1967). The main goal of this chapter was to use mitochondrial phylogenomic approach to investigate the evolutionary origin of the hybridogenetic biotype *P. monacha-latidens*, and determine whether male and female *P. monacha-latidens* differ in mitochondrial haplotypes.

3.2 Materials and Methods

3.2.1 Sampling

We examined seven *P. monacha-latidens* specimens (Table 4): three females from the Rio Fuerte (northernmost limit); and four (two females and two males) from the Rio Mocorito (southernmost limit). For comparison, we included representatives of *P. monacha*-derived asexuals with distinct *cytb+ND2* haplotypes from the Rios Fuerte, Sinaloa, and Mocorito that were previously examined (Mateos & Vrijenhoek, 2002, 2005), some of which were lab bred. We also examined the following samples of sexual species: *P. viriosa* (as outgroup), two samples of *P. monacha* (one new, and one published mt genome; Acc. No. KX229692), and *P. latidens*.

Table 4 Specimens examined in Chapter III.

Species or Biotype	Voucher	Haplotype ID[a]	Locality	Sex	Source
P. monacha-monacha-lucida	Cx-B S68-5	2’	JA, Fuerte	Female	lab-born
P. monacha-lucida	MVH99-12#3	Fuerte 1*	SP, Fuerte	Female	wild-caught
P. monacha-lucida	MVH99-18#3	Sinaloa 1*	PA, Sinaloa	Female	wild-caught
P. monacha-lucida	ML IV Cz_M61-35	C	CA, Fuerte	Female	lab-born
P. monacha-lucida	ML Cx I MVH99-16#5	B	LH, Mocorito	Female	lab-born
P. monacha-lucida	MVH99-16#48	A	LH, Mocorito	Female	wild-caught
P. monacha-lucida	Cx SV73-7s	A	CU, Fuerte	Female	lab-born
P. monacha-lucida	MVH99-18#1	Sinaloa 2*	PA, Sinaloa	Female	wild-caught
P. monacha	MVH99-6#7	Not identified	JA, Fuerte	Female	wild-caught
P. monacha-latidens	MVH99-15#10	Not identified	SP, Fuerte	Female	wild-caught
P. monacha-latidens	MV00-6A#10	Not identified	LCU, Fuerte	Female	wild-caught
P. monacha-latidens	MVH99-17#72	Not identified	LH, Mocorito	Female	wild-caught

Species or Biotype	Voucher	Haplotype ID[a]	Locality	Sex	Source
P. monacha-latidens	MVH99-16#59	Not identified	LH, Mocorito	Female	wild-caught
P. monacha-latidens	MVH99-17#61	Not identified	LH, Mocorito	Male	wild-caught
P. monacha-latidens	MV00-10#45	Not identified	LH, Mocorito	Male	wild-caught
P. viriosa	M65-23#6	Not identified	Mocorito	Male	lab-born
P. monacha-lucida	Cw S68-4	1	Fuerte	Female	lab-born
P. monacha-lucida	Cx T70-30#M1	1	Fuerte	Female	lab-born
P. monacha-latidens	MV00-6A#11	Not identified	LCU, Fuerte	Female	wild-caught
P. monacha-lucida-lucida	M61-31#M1	2''	Fuerte	Female	lab-born
P. monacha-lucida-lucida	SV73-4	2'''	Fuerte	Female	lab-born
P. monacha-lucida	MVH99-12#1	2''	SP, Fuerte	Female	wild-caught
P. monacha-lucida	M65-24	Not identified	Fuerte	Female	lab-born
P. monacha-lucida-viriosa	MV00-9#1	D	SB, Mocorito	Female	wild-caught
P. monacha-lucida-viriosa	MV00-10#3	D	LH, Mocorito	Female	wild-caught
P. monacha-lucida-viriosa	MV00-8#17	D	SB, Mocorito	Female	wild-caught

Table 4 Continued

Species or Biotype	Voucher	Haplotype ID[a]	Locality	Sex	Source
P. monacha-lucida-viriosa	MVH99-16#22	D	LH, Mocorito	Female	wild-caught
P. monacha-lucida-viriosa	MVH99-16#9	E	LH, Mocorito	Female	wild-caught
P. monacha-lucida-viriosa	MV00-9#5	F	SB, Mocorito	Female	wild-caught
P. monacha-lucida	Cx II MV00-10#18	A	LH, Mocorito	Female	wild-caught
P. monacha-lucida	Cx I MV00-10#33	A	LH, Mocorito	Female	wild-caught

[a] all haplotype IDs except those with * were assigned by (Mateos & Vrijenhoek, 2002, 2005)
Table 4 Continued

3.2.2 Library Preparation, Sequencing and Mitochondrial Genome Assembly and Annotation

Genomic DNA was extracted from all specimens with the DNeasy Tissue extraction kit (Qiagen, Valencia, CA). The DNA from ~31 specimens was submitted to the Texas A&M Institute for Genome Sciences and Society's Experimental Genomics Core for library preparation (Swift 2s Turbo Whole Genome Library Prep, Swift Biosciences, Ann Arbor, MI) and sequencing (Illumina Nova-Seq, Paired-End, 150bp reads), assuming a genome size of 1,271-1,311 Mb (~1.3 picogram) (Cimino, 1974).

Raw read files were uploaded to a galaxy server (Afgan et al., 2018). We used Trimmomatic v. 0.38 (Bolger et al., 2014) for trimming reads with average quality lower than 20, cutting adapter (Nextera), and discarding reads if they were shorter than 50 bases. The resulting reads "clean reads" were then mapped to the mitochondrial genome of *P. monacha* (GenBank Acc. No. KX229692) using Bowtie2 v.2.4.2 (Langmead et al., 2009). The consensus sequence was generated at 90% threshold using Geneious. Upon alignment of the mitochondrial genomes (see below), singletons and gaps were verified by visually inspecting the mapped reads (SAM file) in Geneious.

For the *P. viriosa* sample, we first computed a consensus sequence of the "clean reads" mapped to the *P. monacha* reference (90% threshold and replacing gaps with N). This consensus did not recover the entire mt genome of *P. viriosa*, as parts of genes appeared to be missing. We then mapped all of the *P. viriosa* "clean reads" to the above consensus and computed a new consensus (at 90% threshold). Using the *P. monacha*

reference annotation, we annotated this second (and final) consensus sequence, which appears to be complete.

3.3.3 Phylogenetic Analyses

The mitochondrial genomes were aligned with MAFFT v.7 (Katoh et al., 2018; Kuraku et al., 2013) and verified visually. We performed analyses with and without an outgroup. As outgroup, we used *P. viriosa*. Given the mitochondrial variation reported among previously examined *P. monacha* and asexual biotypes (Mateos & Vrijenhoek, 2002, 2005), we did not expect a high degree of substitution saturation. Therefore, we used parsimony (heuristic searches; collapsing minimum branch lengths of zero) in PAUP* (Swofford & Sullivan, 2012) to infer the shortest trees(s). A consensus tree was generated by conducting a strict consensus. Node support based on Maximum Likelihood was evaluated with standard (100 replicates), ultrafast bootstrap analyses (1000 replicates) and SH-aLRT branch test (1000 replicates) on the IQ-Tree web server (Trifinopoulos et al., 2016). Pairwise uncorrected p-distances were calculated in MEGA-X (Kumar et al., 2018).

3.3.4 Estimation of radical and conservative amino acid changes

PAMLX (Xu & Yang, 2013) was used to map the synonymous and nonsynonymous changes to the tree by setting RateAncestor rate to 1. Nonsynonymous changes were assigned to a radical vs. conservative category as follows. Seven amino acid classification schemes were used [three from (Zhang, 2000), three from (Hanada et al., 2007) and one from (Grantham, 1974)], to compute the Conservative-Radical Index (CRI) described by (Sharbrough et al., 2018). Under each scheme, amino acid changes

were assigned a value of 1 for radical or 0 for conservative. CRI was then calculated by averaging across the seven schemes. The threshold was set to 0.5, which means changes with CRI > 0.5 were considered as radical changes. We used a CRI python script developed by (Sharbrough et al., 2018) to calculate the CRI value for all the amino acid changes. The output file was then processed by a custom python script (Supplemental Dataset "Python Script") to calculate the number of radical and conservative amino acid changes.

3.3 Results

3.3.1 Sequencing read statistics and quality control

We obtained 40–100 million raw sequence reads per each of the 31 samples (Table S1). Raw reads have been submitted to NCBI under BioProject PRJNA687040. On average 94% of raw reads were retained after the quality control steps (referred to as "clean reads").

0.1–8% of the "clean reads" per sample mapped to the reference mitochondrial genome, achieving an average coverage of 6861X (Supporting Table S1). Alignment of the consensus sequences revealed several unique indels and substitutions. For three individuals, visual inspection of the mapped reads revealed variation at one position (different for each individual), suggesting that they contained two alternative copies (heteroplasmy) of the mitochondrial genome (Table 5).

The final mitochondrial genomes have been deposited in GenBank (Acc. No. MZ681814-681846).

Table 5 Evidence of heteroplasmy in three individuals. Nucleotide position, gene, number and percentage of reads covering each haplotype are provided. The nucleotide under haplotype 1 corresponds to the variant found in all other samples.

	Position (gene)	haplotype 1	haplotype 2
P. monacha-lucida MVH99-16#48	3872 (tRNA)	G: 332 (15.8%)	T: 1772 (84.2%)
P. monacha-latidens MV00-10#45	7224 (COX2)	A: 1849 (43.7%)	No base: 2230[a] (52.7%)
P. monacha-latidens MV00-6A#10	5394 (tRNA)	T: 427 (45.1%)	No base: 515 (54.4%)

[a] causes a frameshift and early stop codon. The % don't add to 100 because 149 reads had a C and 1 read had a G.

3.3.2 Mitochondrial Phylogeny

The alignment including the mitochondrial genome of *P. viriosa* (outgroup) was 16,854 bps long. Gapped positions (annotated in the alignment file) were removed prior to Maximum Likelihood analyses, resulting in 16,765 characters. For the parsimony analyses, 8 gapped positions were retained and gaps were coded as a fifth character state. 1559 characters were variable, of which 1378 were parsimony informative.

The parsimony analyses excluding gapped positions recovered 7 most parsimonious trees (1649 steps, CI 0.96); the strict consensus is shown in (Supporting Figure S1). Inclusion of gapped positions (as a fifth state) resulted in a single most parsimonious tree (1657 steps, CI 0.9602).

The ML (IQ-Tree) analyses recovered a topology almost identical to the MP tree (Figure 7). Rooting with *P. viriosa* recovered four clades of *P. monacha* or *P. monacha*-derived mitogenomes joined at a basal polytomy. One clade (of ML) was exclusive to the Rio Sinaloa, but it received relatively low support. Two clades were exclusive to the Rio Fuerte. Clade "Fuerte exclusive 1" (also with relatively low node support) contained sexual *P. monacha*, which was sister to a clade composed of three *P. monacha-latidens* haplotypes (including the two haplotypes found in individual MV00-6A#10). Clade "Fuerte-exclusive 2" contained the other sexual *P. monacha* individual, several ML, MML and MLL, and one M-lat, which was a divergent sister lineage to the remaining members of this clade.

Two clades exclusive to the Rio Mocorito were recovered. The Mocorito hemiclones (red in Figure 7) clade included ML and M-lat. Two ML and three M-lat

(including one male) individuals shared the same haplotype representing the most recent common ancestor of this clade (Mocorito hemiclone MRCA). Five additional haplotypes were found in this clade differing at 1–3 steps from their MRCA; two of these represented the second variant in the two individuals with evidence of heteroplasmy. The closest relatives to the Mocorito hemiclones clade were *P. monacha-lucida* from the Rio Fuerte. The sister lineage to this "Mocorito hemiclones + Fuerte" clade was the Mocorito gynogens clade (blue in Figure 7), which was comprised exclusively of Mocorito MLV. The Mocorito gynogens clade contained four haplotypes separated from each other by 2–4 steps. Although these results are consistent with the previous study based on two mitochondrial genes (Mateos and Vrijenhoek 2005), examination of the entire mitochondrial genome enabled detection of support for the monophyly of the Mocorito hemiclones with respect to their closest known (Rio Fuerte) relatives.

Whole mitochondrial divergence (i.e., uncorrected pairwise p distance) between the ingroup (sexual *P. monacha* and all asexuals) and the outgroup *P. viriosa* was ~8.2%, whereas the maximum divergence within the ingroup was 0.51% (Supporting Figure S2). Several comparisons between clades identified in Figure 8, had pairwise distance values within the range 0.36–0.51% (e.g. all clades against "Fuerte exclusive 1", "Fuerte exclusive 2" and "Sinaloa exclusive"). In contrast, the divergence between the two Mocorito clades (i.e., hemiclones vs. gynogens) was relatively low (0.12–0.15%). A maximum divergence of 0.03% and 0.02% was observed within the Mocorito hemiclones and gynogens, respectively. The maximum divergences within the Fuerte

exclusive 1, Fuerte exclusive 2, and Sinaloa clades were substantially higher (0.47%, 0.22%, and 0.46%, respectively).

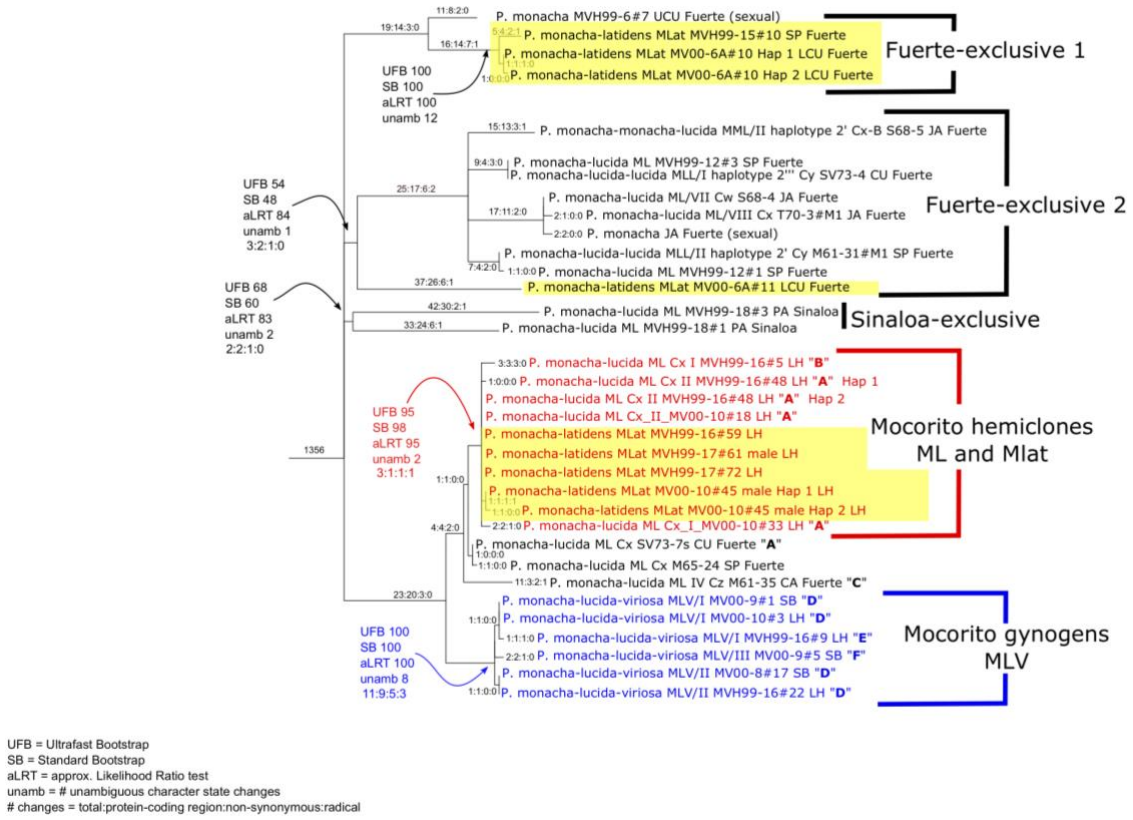

Figure 8 Tree based on mitochondrial genome. Bootstrap consensus (50% majority rule) of the UltraFast Bootstrap ML analysis. Branch lengths represent inferred parsimony steps. The outgroup *P. viriosa* is not shown (separated from the ingroup by 1356 steps). Red taxon labels represent all of the Rio Mocorito hemiclones: *P. monacha-lucida* (ML) and *P. monacha-latidens* (MLat), whereas blue taxon labels represent all of the Mocorito gynogens (*P. monacha-lucida-viriosa*; MLV). Bold-face letters in quotes are the *Cytb+ND2* haplotype labels given to Mocorito clones (Mateos and Vrijenhoek, 2002; 2005). Yellow highlight identifies the biotype *P. monacha-latidens*. Branch values indicate changes, protein coding region changes, nonsynonymous changes and radical changes, from left to right separated by colons. Support values were generated by conducting ultrafast bootstrap, standard bootstrap and approximate likelihood ratio test using IQTREE. Nodes that have no support values shown, received 100% support from the three methods.

3.3.3 Radical and conservative amino acid changes

Our analyses of substitutions at protein-coding mitochondrial genes of the ingroup revealed a lower number of nonsynonymous (range = 0–7) than synonymous (range = 0–28) changes per branch. The number of radical changes per branch was as follows: zero in 23 branches; 1 in 7 branches; 2 in one branch; and 3 in one branch (Figure 8 and Supporting Table S2). For branches where at least one coding substitution occurred, 8 branches had more nonsynonymous than synonymous substitutions (i.e. Kn>Ks). The following branches had a Kn:Ks = 1:0: Mocorito hemiclone monophyly; MLat MV00-10#45 male Hap 1; and MLV MVH99-16#9. The highest Kn:Ks ratio was found in the branch leading to Mocorito hemiclone ML MVH99-16#5_Hap_B (3:0), followed by the branch leading to Fuerte ML MVH99-12#3 and MLL SV73-4 (3:1), Fuerte ML M61-35_Hap_C (2:1), and the branch supporting the monophyly of the Mocorito gynogens MLV (1.25:1). With one exception (i.e., the branch leading to Fuerte ML MVH99-12#3 + MLL SV73-4), the aforementioned branches also had more radical than conservative substitutions (i.e. Kr>Kc). The two terminal branches leading the sexual *P. monacha* specimens had Kn/Ks < 1, and lacked radical substitutions.

The radical substitutions identified occurred in seven genes (*ND1*, *ND2*, *ND4*, *ND5*, *ND6*, *COX1* and *COX2*), which collectively represent 2 electron transport system (ETS) complexes, I and IV (Sazanov, 2015). (Supporting Table S3).

3.4 Discussion

In this study, we used mitochondrial phylogenomics to investigate the origin of *P. monacha-latidens* from the northernmost (Fuerte) and southernmost (Mocorito) rivers where this biotype has been reported. The occurrence of *P. monacha-latidens* in three separate clades (two in the Fuerte and one in the Mocorito) whose closest known relatives are other asexual biotypes or sexual *P. monacha*, provides evidence of at least three independent hybridization events between a female *P. monacha* (or *P. monacha-lucida*; see below) and a male *P. latidens* that gave rise to hybridogenetic biotype *P. monacha-latiden*s. Such independent origins of *P. monacha-latidens* were expected given the overlap of *P. monacha* and *P. latidens* in the Rio Fuerte, where recurrent hybridization events are feasible, akin to the pattern reported for *P. monacha-lucida* in the Rio Fuerte (Quattro et al., 1991) and *P. monacha-occidentalis* in the Rio Mayo (Quattro et al., 1992). In contrast, in the Rio Mocorito, where *P. monacha* is absent, we found a single clade comprised exclusively by all the *P. monacha-lucida* and *P. monacha-latidens* hemiclones sampled to date at that river. Furthermore, the sharing of identical mitochondrial haplotypes between the *P. monacha-lucida* and the *P. monacha-latidens* biotypes of the Rio Mocorito, suggests that they represent the same hemiclonal genome, which uses either *P. lucida* or *P. latidens* as sperm sources (Figure 9). If true, *P. monacha-lucida* and *P. monacha-latidens* individuals that share the same mitochondrial genome are expected to also share an identical hemiclonal nuclear genome. The very low divergence detected within the Mocorito hemiclone clade (maximum p = 0.03%), suggests that the hybridization event that gave rise to it is relatively recent. Indeed, the

haplotype of several specimens matches the haplotype attributed to the most recent common ancestor (MRCA) of this clade.

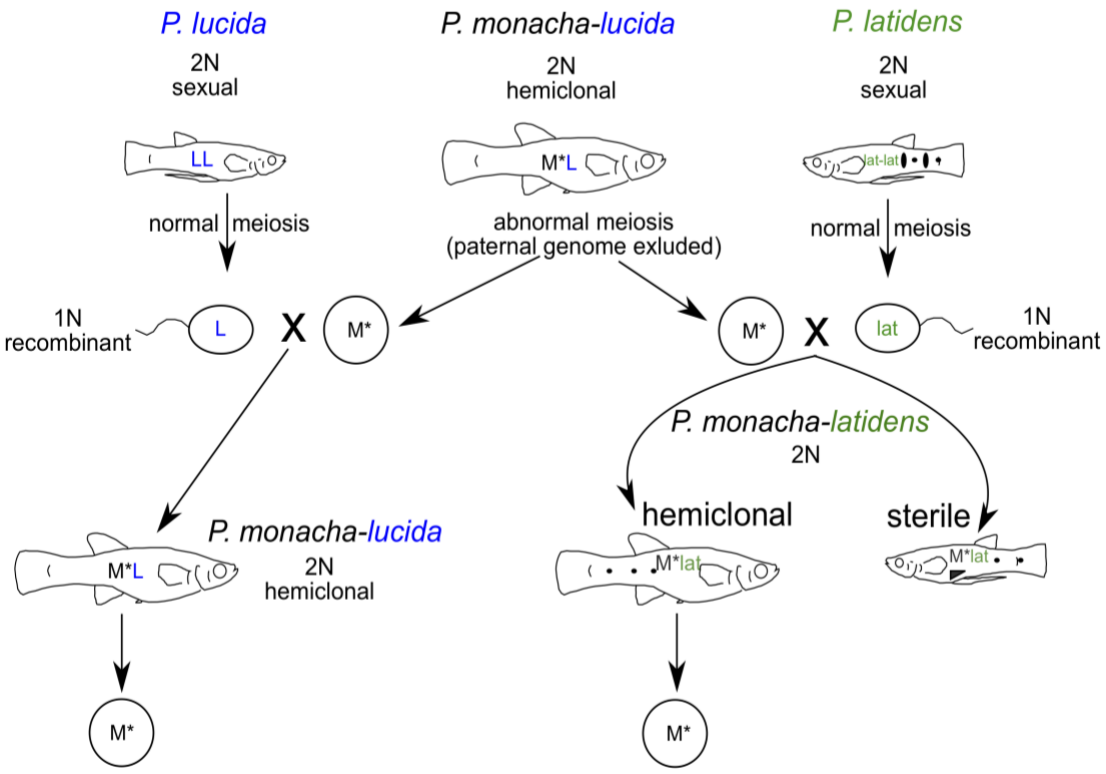

Figure 9 Illustration of a possible formation scenario of *P. monacha-latidens*. A diploid hemiclonal (=hybridogenetic) *P. monacha-lucida* (ML) female produces a haploid (M) eggs through an abnormal meiosis that excludes the paternal (L) chromosomes. On the left, the M egg is fertilized by sperm from the sexual host *P. lucida*, producing a new *P. monacha-lucida* (ML) diploid hemiclonal female. On the right, the M egg is fertilized by sperm from the sexual host *P. latidens*. This cross produces mostly diploid *P. monacha-latidens* (Mlat) hemiclonal females, but occasionally males are produced, which appear to be sterile. Asterisks are used to emphasize that the haploid set of *P. monacha*-derived chromosomes (M) is identical (barring de novo mutations) among descendants of the original hemiclonal female, as long as hemiclonal reproduction persists. *P. monacha-latidens* (Mlat) is also known as "spotted hybrid" due to the characteristic pigmentation pattern inherited from *P. latidens*. The black wedge above the gonopodium (modified anal fin) of the P. monacha-latidens male is unique to (non-nuptial) males of the *P. monacha* and *P. viriosa* (Miller et al., 2005). Blue font depicts *P. lucida* chromosome sets; black font depicts *P. monacha* chromosome sets; and green depicts *P. latidens* chromosome sets.

A previous study of the Mocorito *P. monacha-lucida* based on crossing experiments and morphological traits (pigmentation, dorsolateral scale rows, and dentition), concluded that the hemiclonal haploid genome is comprised of a mix of *P. monacha* and *P. viriosa* genes, rather than being purely derived from *P. monacha* (Vrijenhoek & Schultz, 1974). Under this scenario (Figure 10), referred to as the reticulate origin hypothesis by Mateos and Vrijenhoek (2002), a *P. monacha-lucida* would have mated with a *P. viriosa* male, producing a *P. monacha* x *P. viriosa* F1 hybrid; evidence of such hybrids has been found in nature and they have been recreated in the lab (Mateos & Vrijenhoek, 2002; Vrijenhoek & Schultz, 1974). Such hybrids undergo regular meiosis and produce haploid eggs (or sperm) with a mix of *P. monacha* and *P. viriosa* genes (Vrijenhoek & Schultz, 1974). Subsequent fertilization of such a "reticulate" *monacha/viriosa* (*m/v*) haploid egg by sperm from *P. lucida* could have resulted in a new hemiclonal biotype, whose maternal nuclear genome is mix of *P. monacha* and *P. viriosa* genes. Our finding that *P. monacha-lucida* and *P. monacha-latidens* in the Rio Mocorito share the same mitogenomic haplotypes, implies that the hemiclonal genome of *P. monacha-latidens* might also carry genes acquired from *P. viriosa*. The degree of *P. viriosa* genes present in Mocorito hemiclones awaits to be determined on the basis of nuclear DNA-based markers. Given the evidence for complete incompatibility between *P. viriosa* and *P. lucida* (Vrijenhoek & Schultz, 1974), and the evidence of viable *P. viriosa* x *P. latidens* hybrids (see Chapter 2), implying some degree of compatibility, it is possible that *m/v* haploid eggs containing a greater

complement of *P. viriosa* genes are incompatible with *P. lucida* sperm, but still compatible with *P. latidens* sperm.

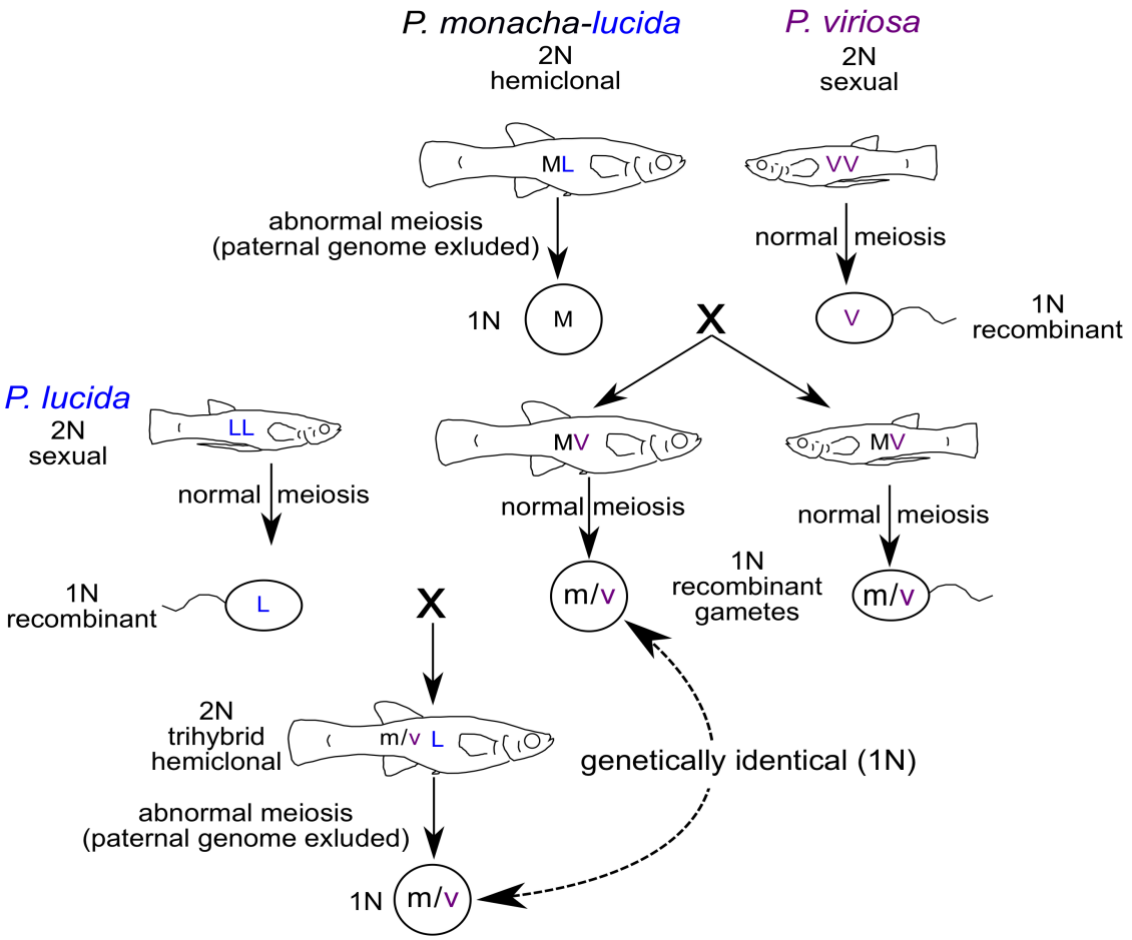

Figure 10 Illustration of recticulate origin hypothesis. Reticulate origin hypothesis for the Mocorito hemiclone (modified from Mateos and Vrijenhoek 2002). A hemiclonal (=hybridogenetic) diploid *P. monacha-lucida* (ML) female produces a haploid (M) egg with only the maternal set of chromosomes (the paternal L set was discarded during an abnormal meiosis; see Figure 1). This egg is fertilized by sperm from *P. viriosa* (a sexual species that is reproductively compatible with *P. monacha*) resulting in *P. monacha-viriosa* F1 hybrids (females and males), which produce genetically different haploid gametes through normal meiosis. "m/v" represents recombinant monacha-viriosa haploid genome. When the m/v egg is fertilized by sperm from *P. lucida*, a diploid female is produced (m/v-L), which reproduces by hybridogenesis, such that the eggs it produces are genetically identical to the eggs produced by her mother (barring any new mutations). Blue "L" depicts a haploid chromosome set from *P. lucida*; black "m" and "M" depict a partial and full, respectively, haploid chromosome set from *P. monacha*; magenta "v" and "V" depict a partial and full, respectively, haploid chromosome set from *P. viriosa*.

An asexual biotype to which *P. viriosa* has undoubtedly contributed substantial genetic material is the Mocorito gynogen *P. monacha-lucida-viriosa* (MLV). Our results also suggest that the Mocorito gynogens, which likely arose from the addition of a *P. viriosa* genome to an unreduced diploid *P. monacha-lucida* (ML) egg (Mateos & Vrijenhoek, 2005), did not derive from the present-day lineage of Mocorito hemiclones. It thus appears that the present-day hemiclonal and gynogenetic biotypes in the Rio Mocorito derive from two independent invasions. Similarly to the Mocorito hemiclone clade, the very low divergence detected within the Mocorito gynogens clade (maximum $p = 0.02\%$) implies a relatively recent origin. The Mocorito gynogens might have arisen by genome addition to a now extinct (or unsampled) *P. monacha-lucida* lineage within the Rio Mocorito. Although it is possible that such *P. monacha-lucida* lineage is still extant in the Rio Mocorito, we consider that taking into account the present data and those reported in Mateos and Vrijenhoek (2002) and Mateos and Vrijenhoek (2005), it is highly unlikely that we have missed it. Collectively the two studies have examined 20 hemiclonal individuals and 13 gynogenetic individuals for at least two mitochondrial genes representing two Rio Mocorito localities (LH and SB) and sampling periods spanning ~20 years (1978 and 1999–2000). An alternative hypothesis is that the *viriosa* genome addition event that gave rise to the Mocorito gynogen occurred in the neighboring Rio Sinaloa, which is the northern limit of *P. viriosa*. Unfortunately, the sites of easiest access in this river have been greatly degraded. The headwater site Arroyo Seco where *P. monacha* was reported prior to 1978 has dried up (Quattro et al., 1991; Vrijenhoek, 1979), whereas the downstream sites where *P. viriosa* was reported in

the 1960s [University of Michigan Museum of Zoology (UMMZ) record no. 179672] have been converted to irrigation canals and/or been invaded by exotic species (M. Mateos, pers. comm.).

Although relatively young, the Mocorito gynogen clade does appear to have accrued a small number of mutations since its origin, with none of the mitochondrial haplotypes identical to the MRCA. Mateos and Vrijenhoek (2005) reported three different MLV clones (I, II, and III) on the basis of allozyme genotypes. Clone III, which had a unique *Cytb+ND2* haplotype (F) in Mateos and Vrijenhoek (2005), has accrued one synonymous and one non-synonymous substitution in its mt genome since the MRCA. In Mateos and Vrijenhoek (2005), all four MLV Clone II individuals examined shared the same *Cytb+ND2* haplotype (D) with seven out of the eight MLV Clone I specimens examined. Our study, which examined two representatives of Clone II, revealed that this lineage has accrued one substitution (synonymous) since the MLV MRCA, and is distinct from Clone I individuals harboring the *Cytb+ND2* haplotype D. Examination of the nuclear genomes of MLV is needed to determine whether additional mutations or recombination (e.g. via the mechanism proposed by Asher and Nace (1971)) has occurred within the genomes of Mocorito gynogens).

Mitochondrial genomes, which are mostly transmitted asexually, can be affected by the transition to asexual reproduction because of the tight linkage to the nuclear genome that they acquire, which reduces effective population size (Normark & Moran, 2000), and thus the efficiency of purifying selection to remove harmful mutations (Birky & Walsh, 1988; W. G. Hill & Robertson, 1966; Kimura & Ohta, 1971). Consistent with

this prediction, asexual lineages of the snail *Potamopyrgus antipodarum* experience higher rate of nonsynonymous substitutions and eliminate radical mutations more slowly than their sexual counterparts (Neiman et al., 2010; Sharbrough et al., 2018). Our mapping of synonymous (Ks), and radical (Kr) vs. conservative (Kc) nonsynonymous (Kn) changes revealed three branches with a Kr>Kc and Kn>Ks and five additional branches with a Kn>Ks, all of which are either internal or terminal branches leading to exclusively asexual taxa. The two terminal branches leading to sexual *P. monacha*, as well as several of the "exclusively asexual" branches, did not exhibit Kr>Kc or Kn>Ks. Such observations may be suggestive that purifying selection is relaxed in some asexual lineages, but there are several caveats. First, some of the "exclusively asexual" branches may contain yet unsampled sexual *P. monacha*, particularly in the Rio Fuerte where de novo asexual hybrid formation is common. Secondly, to adequately assess whether purifying selection is indeed less efficient in the asexual lineages, the expectation that radical (and presumably more harmful) mutations persist longer in asexual than sexual lineages should be verified (Neiman et al., 2010), which would require analyses of polymorphism within "populations" and thus additional sampling of at least sexual *P. monacha*. Examination of putatively older asexuals (e.g. *P. monacha-occidentalis* from the Rio Concepcion) (Quattro et al., 1992) may be necessary, as present-day the Rio Fuerte, Rio Sinaloa, and Rio Mocorito asexuals may be too young to have accrued detectable levels of deleterious mutations. The occurrence of amino acid substitutions, including some deemed radical, in the mitochondrial genome of *Poeciliopsis* asexuals examined here, also raises the question as to whether co-evolution occurs between the

nuclear genes whose products interact with those of these mitochondrial genes (G. E. Hill, 2020).

P. monacha-latidens is the only *Poeciliopsis* asexual biotype where, albeit rarely, males occur both in the field (M. Mateos, pers. comm.) and in the lab (6.7% of offspring born to one female) (Schultz, 1966). *P. monacha-latidens* males are likely sterile because matings to *P. monacha-lucida*, *P. lucida*, *P. latidens*, and *P. monacha-lucida-lucida* (MLL) failed to produce offspring (Schultz 1967). We found little or no variation between the mitochondrial genomes of male and female *P. monacha-latidens* specimen samples (males were only available for the Rio Mocorito). What enables the production of males by *P. monacha-latidens* but not in other *Poeciliopsis* asexual biotypes is unknown. Schultz (1961) examined ovaries and brood data for the biotypes *P. monacha-lucida* and *P. monacha-occidentalis* (referred to as "C_X" and "F_X", respectively by Schultz 1961), which never produce males, and found no evidence of male lethality or sex inversion at any stage. Accordingly, Kallman (1984) proposed that *P. monacha* is homozygous for a particular allele at an autosomal locus that "blocks" the male-determining gene from *P. lucida* and *P. occidentalis*; in other words the *monacha* allele has a dominant epistatic effect on the male determining gene/pathway of *P. lucida* and *P. occidentalis*. Upon matings with *P. latidens* males, the Rio Fuerte *P. monacha-lucida* biotype known as "C_X" produced both females and males (Schultz, 1967). In contrast, "C_Z" (i.e., another "strain" of *P. monacha-lucida*) differed from Rio Fuerte "C_X" in that it produced only males upon mating with *P. latidens* (Schultz, 1967). Similarly, matings of *P. latidens* males with different strains of Mocorito *P. monacha-lucida*, produced

either both sexes, only females, or no offspring (Vrijenhoek and Schultz 1974). The different sex ratios produced by *P. monacha-lucida* matings with *P. latidens* suggests that the *monacha* hemiclonal genomes of these strains differed from each other and/or that variation in the "male-determining gene" was present among or within the *P. latidens* males used. The sex determination mechanism of members of *Poeciliopsis* is unknown, but given the large diversity of mechanisms in poeciliids (Kallman, 1984; Volff & Schartl, 2001), numerous mechanisms that would result in the observed sex ratios among different crosses are possible. For example, crosses between the mosquitofish *Gambusia holbrooki* (with an XX-XY sex chromosome system) and *Gambusia affinis* (with an ZZ-ZW system) result in WY males (i.e., the Y chromosome is dominant over the W chromosome), XZ females (i.e., X is dominant over Z) (Kottler et al., 2020). In contrast, in the platyfish *Xiphophous* maculatus (with an X-Y-Z system), WY fish develop as females (i.e., the W chromosome is dominant over the Y chromosome) (Kallman & Schreibman, 1973; Volff & Schartl, 2002). "Sexomics" (Matthias Stöck et al., 2021) of female and male *P. monacha-latidens*, as well as *P. monacha*, *P. latidens* and the other species involved in asexual *Poeciliopsis*. hybrids should enable identification of the sex determination mechanisms. Comparison of the nuclear hemiclonal genomes of female and male *P. monacha-latidens* is needed to determine whether any *P. monacha* genes are involved in the sex determination of *P. monacha-latidens*. Identification of potential male determining genes in *P. latidens* (e.g. through comparison of the male and female *P. latidens* genomes), will also help understand the sex determination mechanism of *P. monacha-latidens*. The influence of

environmental factors such as temperature on sex ratios, however, should be also considered, as Sullivan and Schultz (1986) reported that one strain of *P. lucida* produced fewer males with decreasing temperatures.

That the Mocorito *P. monacha-lucida* and *P. monacha-latidens* might represent the same hemiclones differing only by the sperm donor is not too surprising, given that Schultz (1977) reported that a pregnant wild-caught *P. monacha-lucida* from the Rio Mocorito produced a mixed brood of six *P. monacha-lucida* and one *P. monacha-latidens* offspring in the lab. A similar phenomenon has been observed in the Rio Concepcion (the northernmost locality for the hybridogenetic biotype *P. monacha-occidentalis*). In addition to *P. monacha-occidentalis* and its sexual host *P. occidentalis*, a small range of the Rio Concepcion also harbors the recently described bisexual species *P. jackschultzi*, and many females carrying mitochondrial haplotypes identical to those of local *P. monacha-occidentalis*, and allozyme genotypes consistent with a *P. monacha x P. jackschultzi* F1 cross (Conway et al., 2019; Schenk, 1991). Several such *P. monacha-jackschultzi* hybrids that were kept in the lab produced all-female offspring, suggesting that this biotype also reproduces via hybridogenesis (Conway et al., 2019), but future work will be needed to rule out other reproductive modes, including normal meiosis or gynogenesis. The ability to exploit two very different sexual host species may have implications on the long-term persistence of the Mocorito (and Rio Concepcion) hemiclones in terms of both, securing inseminations and ability to compete against the sexual hosts. Males of the sexual host species should gain no direct fitness advantages from inseminating hemiclones, as their genes will not be passed on to their

granddaughters. Thus, such males may be under strong selection to avoid matings with hemiclones. Indeed, studies have revealed that *P. monacha-lucida* (and *P. monacha-occidentalis*) populations may be primarily limited by sperm, with *P. lucida* (and *P. occidentalis*, respectively) males exhibiting a strong preference for conspecifics (McKay 1971, Lanza 1983, Lima et al. 1996). In contrast, *P. monacha-latidens* populations (only Rio Fuerte populations were examined) do not appear to be primarily sperm limited, but *P. latidens* males still exhibit a preference for conspecifics (J. Lanza, 1983). The primary limitation for *P. monacha-latidens* appears to be reduced competitive ability with *P. latidens* (J. Lanza, 1983). *P. lucida* and *P. latidens* differ in several ecological and life history traits. *P. lucida* is omnivorous, pool-inhabiting, and has variable but typically female-biased sex ratios. In contrast, *P. latidens* is carnivorous, riffle-inhabiting and typically has male-biased sex ratios. *P. monacha* is carnivorous and restricted to headwater pools. *P. monacha-latidens* is less well adapted to fast water than *P. latidens* and loses in competition for food with *P. latidens* when food is limited (J. Lanza, 1983). Microhabitat-use experiments in sympatry vs. allopatry in artificial streams revealed no competition between *P. monacha-lucida* and *P. lucida*, or with between *P. monacha-lucida* and *P. monacha-latidens* (Janet Lanza, 1983). Collectively, these studies suggest that *P. monacha-lucida* benefit from ecological convergence with *P. lucida* (e.g. appear *lucida*-like to secure inseminations by *P. lucida* males), and *P. monacha-latidens* benefit from ecological divergence from *P. latidens*, to minimize competition (Janet Lanza, 1983). How these seemingly opposing forces may act upon the same hemiclonal genome remains unclear. This is further complicated by the fact that the hemiclonal genome in

the Rio Mocorito can evolve not only by de novo mutations, but also through introgression from *P. viriosa*, via the temporary return to sexuality explained above. Finally, if *P. latidens* males are less "choosy" than *P. lucida* as prior studies suggest, it will be interesting to determine whether *P. latidens* males are more likely to inseminate *P. monacha-lucida,* than the *P. lucida* males are to inseminate *P. monacha-latidens*.

Evidence consistent with heteroplasmy was detected in three hemiclone individuals; one from the Fuerte (M-lat MV00-6A#10), and two from the Mocorito (ML MVH99-16#48 and M-lat MV00-10#45). One had a single base deletion at the COX2 gene predicted to cause a frameshift and early stop codon. The other two had a single base substitution or deletion at a tRNA gene (Table 5). We consider that sequencing error is unlikely to explain these variants because their coverage indicates they occurred at substantially high copy numbers. To completely rule out technical error, however, it would be necessary to demonstrate that these variants are not mutations introduced by the polymerase during the amplification step of the sequencing library preparation (e.g. sequence an independent library prepared from the same sample and/or use PCR and Sanger sequencing to demonstrate that the variants are present). In many taxa where heteroplasmy has been reported, it is attributable to paternal leakage (Ladoukakis & Zouros, 2017). In the case of the hemiclones we examined, however, we can rule out such leakage from *P. lucida* or *P. latidens* fathers because their mitochondrial genomes exhibit substantially higher divergence from *P. monacha*. In contrast, in the hybrid (also hemiclonal) *Pelophylax* frog system of the Balkan Peninsula, heteroplasmy due to

paternal leakage is widespread, and hybridization itself is thought to contribute to the high incidence of heteroplasmy (Radojičić et al., 2015).

An alternative explanation is that the additional mt variants we detected in three individuals actually represent nuclear elements of mtDNA (NUMTs) (Kowal et al., 2020), but the high coverage of these variants suggests that they are present at very high copy numbers, such as mitochondrial genomes. A NUMT of the observed coverage would imply a high number of copies in the nuclear genome. Thus, the most parsimonious explanation is heteroplasmy.

Detection of heteroplasmy in asexual *Poeciliopsis* raises several questions related to their possible causes and consequences. Evaluation of mother-offspring groups would help determine whether these mutations are somatic or heritable. Concerning potential fitness consequences, it has been well documented that heteroplasmy is associated with human diseases (Holt et al., 1990; Stewart & Chinnery, 2015). Research on heteroplasmic disease-causing variants in humans has revealed that the relative copy numbers of the disease-causing variant influences the penetrance and expressivity of the disease (Chinnery et al., 1997). Based on our results, the putatively deleterious COX2 variant represented more than half of the copies. The heteroplasmic state itself, even when the variants do not carry harmful mutations, has been shown to be deleterious (Hoitzing et al., 2019; Sharpley et al., 2012). Future work could compare the fitness or aerobic capacity of individuals with and without heteroplasmy, or with different relative copies of the putatively deleterious variant. If heteroplasmy affects fitness negatively, it could contribute to the demise of the asexual lineage. Finally, it is possible that the

hybrid genomic environment (e.g. through incompatibilities of the divergent maternal and paternal genomes), and/or the predicted higher mutation load of asexual genomes, could themselves lead to a higher incidence of heteroplasmy. Based on a GWAS of humans, Nandakumar et al. (2021) reported a heritable nuclear component to mitochondrial heteroplasmy.

CHAPTER IV CONCLUSION

Chapter 2 reports on the development of the first DNA-based nuclear genotyping protocol for identification of an asexual hybrid biotype of the genus *Poeciliopsis*. This allele-specific protocol diagnoses the presence of the allele of *P. latidens* at the nuclear locus *Ck-M*. Thus, in combination with genotyping at a mitochondrial gene such as *cytb*, the *P. monacha-latidens* biotype can be unequivocally distinguished from its sexual progenitors (*P. monacha* and *P. latidens*), as well as from other co-occurring sexual species (e.g. *P. viriosa* and *P. lucida*) and asexual biotypes (e.g. *P. monacha-lucida*). This study also confirmed that Sanger sequencing of PCR products targeting another nuclear gene (*SH3PX3*) with universal PCR primers serves as an alternative DNA-based genotyping protocol for *P. monacha-latidens*. The allele-specific PCR tends to require more optimization, whereas the Sanger-sequencing based protocol requires access to more specialized equipment (i.e., sequencer), such that each has its advantages and disadvantages depending on the setting. Future work should use already available nuclear gene sequences or generate new sequences to identify species-diagnostic loci to serve as the basis for the development of similar protocols to accurately identify the other *Poeciliopsis* hybrid biotypes. The availability of DNA-based, rather than protein-based, genotyping approaches opens the possibility of examining old ethanol preserved collections and facilitates the collection and preservation of future specimens, as availability of dry ice, liquid nitrogen, or ultra-low freezers will not be necessary.

Chapter 2 also includes the first report on the existence of specimens with external morphology and/or mitochondrial and nuclear genotype consistent with an F1 hybrid between a *P. viriosa* female and a *P. latidens* male, raising the possibility of introgression between these highly divergent species.

Chapter 3 generated whole genome sequencing raw data that were used to assemble the whole mitochondrial genomes of representatives of most of the asexual biotypes of *Poeciliopsis* that occur in three adjacent rivers (Rio Fuerte, Rio Sinaloa, and Rio Mocorito). Phylogenetic analyses of the mitogenomes revealed the following insights. First, we detected evidence of three independent origins of the asexual biotype *P. monacha-latidens* (two within the Rio Fuerte). The Rio Mocorito *P. monacha-latidens* shared identical or very closely related mitochondrial haplotypes with the sympatric biotype *P. monacha-lucida*, suggesting a common origin within the Rio Mocorito or the invasion by a single hemiclonal lineage (the "Mocorito hemiclone") from the Rio Fuerte or Sinaloa. The very low divergence within the Mocorito hemiclone clade suggests a relatively recent origin. These results suggest that *P. monacha-latidens* and *P. monacha-lucida* from the Rio Mocorito may constitute a single hemiclonal lineage that uses two divergent sexual host species. Examination of the nuclear genome is necessary to test this hypothesis. Secondly, the Mocorito-endemic gynogenetic triploid biotype *P. monacha-lucida-viriosa*, which likely arose by the addition of a *viriosa* genome to an unreduced *P. monacha-lucida* diploid egg, did not arise from the present-day Mocorito hemiclones.

Analyses of substitution patterns in the mitochondrial genomes, revealed several asexual lineages that have accrued a small number of amino acid substitutions deemed radical at seven genes that are collectively involved in two of the electron transport system complexes, but examination of variation within populations is needed to determine whether asexual lineages accumulate deleterious mutations at a higher rate that recombining lineages. Such radical substitutions, which could render these mitochondrial genes incompatible with the alleles at nuclear loci that are part of these complexes, could lead to compromised fitness and threaten the long-term persistence of these asexual lineages. Whether the constant heterozygous state of these nuclear loci in hemiclones contributes to alleviate or exacerbate potential mitonuclear incompatibilities is intriguing.

Chapter 3 also revealed the first record of mitochondrial heteroplasmy in *Poeciliopsis*. Separate instances of heteroplasmy were detected in three hemiclonal specimens. Each heteroplasmic individual contained two variants (minimum frequency of ~16% based on read coverage). Variants within each individual, differed by one substitution or one insertion/deletion, and thus are not attributable to paternal leakage (because the sperm donor of hemiclones is from species highly divergent from the maternal progenitor *P. monacha*). Future work will be needed to determine the fitness consequences of the observed heteroplasmy, which depend on whether they are somatic or germline, and whether or not they are deleterious. Deleterious heteroplasmy can occur due to a non-functional product (e.g. one of the heteroplasmic variants is predicted to cause an early stop codon), or to the state of heteroplasmy itself. Finally, our detection of

heteroplasmy in only asexual lineages is interesting, as it is possible that the genomic environment itself (e.g. two highly divergent genomes) might contribute to development of heteroplasmy.

REFERENCES

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1038/nbt0388-282

Alves, M. J., Collares-Pereira, M. J., Dowling, T. E., & Coelho, M. M. (2002). The genetics of maintenance of an all-male lineage in the Squalius alburnoides complex. *Journal of Fish Biology*, *60*(3), 649–662. https://doi.org/10.1006/jfbi.2002.1883

Andersson, D. I., & Hughes, D. (1996). Muller's ratchet decreases fitness of a DNA-based microbe. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(2), 906–907. https://doi.org/10.1073/pnas.93.2.906

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Asher, J. H., & Nace, G. W. (1971). The genetic structure and evolutionary fate of parthenogenetic amphibian populations as determined by markovian analysis.

Integrative and Comparative Biology, *11*(2), 381–398. https://doi.org/10.1093/icb/11.2.381

Avise, J. C. (2015). Evolutionary perspectives on clonal reproduction in vertebrate animals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8867–8873. https://doi.org/10.1073/pnas.1501820112

Bast, J., Parker, D. J., Dumas, Z., Jalvingh, K. M., Van, P. T., Jaron, K. S., Figuet, E., Brandt, A., Galtier, N., & Schwander, T. (2018). Consequences of asexuality in natural populations: Insights from stick insects. *Molecular Biology and Evolution*, *35*(7), 1668–1677. https://doi.org/10.1093/molbev/msy058

Bell, G. (1989). *Sex and death in Protozoa: the history of obsession*. Cambridge University Press. https://doi.org/10.1017/CBO9780511525704

Birky, C. W., & Walsh, J. B. (1988). Effects of linkage on rates of molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(17), 6414–6418. https://doi.org/10.1073/pnas.85.17.6414

Bogart, J. P., Elinson, R. P., & Licht, L. E. (1989). Temperature and sperm incorporation in polyploid salamanders. *Science*, *246*(4933), 1032–1034. https://doi.org/10.1126/science.2587986

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Cha, R. S., Zarbl, H., Keohavong, P., & Thilly, W. G. (1992). Mismatch amplification mutation assay (MAMA): Application to the c-H-ras gene. *Genome Research*, *2*(1),

14–20. https://doi.org/10.1101/gr.2.1.14

Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature*, *348*(6300), 454–455. https://doi.org/10.1038/348454a0

Chinnery, P. F., Howell, N., Lightowlers, R. N., & Turnbull, D. M. (1997). Molecular pathology of MELAS and MERRF. The relationship between mutation load and clinical phenotypes. *Brain: A Journal of Neurology*, *120*(10), 1713–1721. https://doi.org/10.1093/brain/120.10.1713

Christiansen, D. G. (2005). A microsatellite-based method for genotyping diploid and triploid water frogs of the Rana esculenta hybrid complex. *Molecular Ecology Notes*, *5*(1), 190–193. https://doi.org/10.1111/j.1471-8286.2004.00869.x

Cimino, M. C. (1974). The nuclear DNA content of diploid and triploid Poeciliopsis and other poeciliid fishes with reference to the evolution of unisexual forms. *Chromosoma*, *47*(3), 297–307. https://doi.org/10.1007/BF00328863

Conway, K. W., Mateos, M., & Vrijenhoek, R. C. (2019). A new species of the live-bearing fish genus Poeciliopsis from northern Mexico (Cyprinodontiformes, Poeciliidae). *ZooKeys*, *2019*(883), 91–118. https://doi.org/10.3897/zookeys.883.37586

Costa, G. C., & Schlupp, I. (2020). Placing the hybrid origin of the asexual Amazon molly (Poecilia formosa) based on historical climate data. *Biological Journal of the Linnean Society*, *129*(4), 835–843. https://doi.org/10.1093/biolinnean/blaa010

Crespo-López, M. E., Duarte, T., Dowling, T., & Coelho, M. M. (2006). Modes of reproduction of the hybridogenetic fish Squalius alburnoides in the Tejo and

Guadiana rivers: an approach with microsatellites. *Zoology*, *109*(4), 277–286. https://doi.org/10.1016/j.zool.2006.03.008

Cuenca, J., Aleza, P., Navarro, L., & Ollitrault, P. (2013). Assignment of SNP allelic configuration in polyploids using competitive allele-specific PCR: Application to citrus triploid progeny. *Annals of Botany*, *111*(4), 731–742. https://doi.org/10.1093/aob/mct032

Cutler, D. J., Zwick, M. E., Carrasquillo, M. M., Yohn, C. T., Tobin, K. P., Kashuk, C., Mathews, D. J., Shah, N. A., Eichler, E. E., Warrington, J. A., & Chakravarti, A. (2001). High-throughput variation detection and genotyping using microarrays. *Genome Research*, *11*(11), 1913–1925. https://doi.org/10.1101/gr.197201

Dalziel, A. C., Tirbhowan, S., Drapeau, H. F., Power, C., Jonah, L. S., Gbotsyo, Y. A., & Dion-Côté, A. M. (2020). Using asexual vertebrates to study genome evolution and animal physiology: Banded (Fundulus diaphanus) x Common Killifish (F. heteroclitus) hybrid lineages as a model system. *Evolutionary Applications*, *13*(6), 1214–1239. https://doi.org/10.1111/eva.12975

De Meeûs, T., Prugnolle, F., & Agnew, P. (2007). Asexual reproduction: Genetics and evolutionary aspects. *Cellular and Molecular Life Sciences*, *64*(11), 1355–1372. https://doi.org/10.1007/s00018-007-6515-2

Dedukh, D., Litvinchuk, J., Svinin, A., Litvinchuk, S., Rosanov, J., & Krasikova, A. (2019). Variation in hybridogenetic hybrid emergence between populations of water frogs from the Pelophylax esculentus complex. *PLoS ONE*, *14*(11), 1–19. https://doi.org/10.1371/journal.pone.0224759

Doležálková-Kaštánková, M., Mazepa, G., Jeffries, D. L., Perrin, N., Plötner, M., Plötner, J., Guex, G. D., Mikulíček, P., Poustka, A. J., Grau, J., & Choleva, L. (2021). Capture and return of sexual genomes by hybridogenetic frogs provide clonal genome enrichment in a sexual species. *Scientific Reports*, *11*(1), 1–10. https://doi.org/10.1038/s41598-021-81240-5

Drenkard, E., Richter, B. G., Rozen, S., Stutius, L. M., Angell, N. A., Mindrinos, M., Cho, R. J., Oefner, P. J., Davis, R. W., & Ausubel, F. M. (2000). A simple procedure for the analysis of single nucleotide polymorphism facilitates map-based cloning in Arabidopsis. *Plant Physiology*, *124*(4), 1483–1492. https://doi.org/10.1104/pp.124.4.1483

Fisher, H. S., Wong, B. B. M., & Rosenthal, G. G. (2006). Alteration of the chemical environment disrupts communication in a freshwater fish. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1591), 1187–1193. https://doi.org/10.1098/rspb.2005.3406

Gaudet, M., Fara, A. G., Beritognolo, I., & Sabatti, M. (2009). Allele-specific PCR in SNP genotyping. In *Methods in Molecular Biology* (pp. 415–424). Humana Press. https://doi.org/10.1007/978-1-60327-411-1_26

Gibson, A. K., Xu, J. Y., & Lively, C. M. (2016). Within-population covariation between sexual reproduction and susceptibility to local parasites. *Evolution; International Journal of Organic Evolution*, *70*(9), 2049–2060. https://doi.org/10.1111/evo.13001

Gómez, A., & Carvalho, G. R. (2000). Sex , parthenogenesis and genetic structure of

rotifers : microsatellite analysis of contemporary and resting egg bank populations. *Molecular Ecology*, *9*(2), 203–214. https://doi.org/10.1046/j.1365-294x.2000.00849.x

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Gösser, F., Schartl, M., García-De León, F. J., Tollrian, R., & Lampert, K. P. (2019). Red Queen revisited: Immune gene diversity and parasite load in the asexual Poecilia formosa versus its sexual host species P. mexicana. *PLoS ONE*, *14*(7), 7–9. https://doi.org/10.1371/journal.pone.0219000

Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science*, *185*(4154), 862–864. https://doi.org/10.1126/science.185.4154.862

Han, Y., Khu, D. M., & Monteros, M. J. (2012). High-resolution melting analysis for SNP genotyping and mapping in tetraploid alfalfa (Medicago sativa L.). *Molecular Breeding*, *29*(2), 489–501. https://doi.org/10.1007/s11032-011-9566-x

Hanada, K., Shiu, S. H., & Li, W. H. (2007). The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Molecular Biology and Evolution*, *24*(10), 2235–2241. https://doi.org/10.1093/molbev/msm152

Hernández Chávez, C., & Turgeon, J. (2007). Asexual and sexual hybrids between Fundulus diaphanus and F. heteroclitus in the Canadian Atlantic region. *Molecular Ecology*, *16*(7), 1467–1480. https://doi.org/10.1111/j.1365-294X.2007.03239.x

Hill, G. E. (2020). Mitonuclear Compensatory Coevolution. *Trends in Genetics*, *36*(6), 403–414. https://doi.org/10.1016/j.tig.2020.03.002

Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, *8*(3), 269–294. https://doi.org/10.1017/S001667230800949X

Hoitzing, H., Gammage, P. A., Van Haute, L., Minczuk, M., Johnston, I. G., & Jones, N. S. (2019). Energetic costs of cellular and therapeutic control of stochastic mitochondrial dna populations. *PLoS Computational Biology*, *15*(6), 1–27. https://doi.org/10.1371/journal.pcbi.1007023

Hollister, J. D., Greiner, S., Wang, W., Wang, J., Zhang, Y., Wong, G. K. S., Wright, S. I., & Johnsony, M. T. J. (2015). Recurrent loss of sex is associated with accumulation of deleterious mutations in oenothera. *Molecular Biology and Evolution*, *32*(4), 896–905. https://doi.org/10.1093/molbev/msu345

Holt, I. J., Harding, A. E., Petty, R. K. H., & Morgan-Hughes, J. A. (1990). A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *American Journal of Human Genetics*, *46*(3), 428–433.

Holub, A. M., & Shackelford, T. K. (2021). *Sex in the Animal Kingdom*.

Hoy, M. (2003). *Insect molecular genetics: an introduction to principles and applications*. Academic Press.

Hubbs, C. L., & Hubbs, L. C. (1932). Apparent parthenogenesis in nature, in a form of fish of hybrid origin. *Science*, *76*(1983), 628–630. https://doi.org/10.1126/science.76.1983.628

Janko, K., Drozd, P., Flegr, J., & Pannell, J. R. (2008). Clonal turnover versus clonal

decay: A null model for observed patterns of asexual longevity, diversity and distribution. *Evolution*, *62*(5), 1264–1270. https://doi.org/10.1111/j.1558-5646.2008.00359.x

Janko, K., Kotusz, J., de Gelas, K., Šlechtová, V., Opoldusová, Z., Drozd, P., Choleva, L., Popiołek, M., & Baláž, M. (2012). Dynamic Formation of Asexual Diploid and Polyploid Lineages: Multilocus Analysis of Cobitis Reveals the Mechanisms Maintaining the Diversity of Clones. *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0045384

Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M., & Schwander, T. (2021). Genomic Features of Parthenogenetic Animals. *The Journal of Heredity*, *112*(1), 19–33. https://doi.org/10.1093/jhered/esaa031

Jatayev, S., Kurishbayev, A., Zotova, L., Khasanova, G., Serikbay, D., Zhubatkanov, A., Botayeva, M., Zhumalin, A., Turbekova, A., Soole, K., Langridge, P., & Shavrukov, Y. (2017). Advantages of Amplifluor-like SNP markers over KASP in plant genotyping. *BMC Plant Biology*, *17*(Suppl 2). https://doi.org/10.1186/s12870-017-1197-x

Kallman, K. D. (1984). A new look at sex determination in poeciliid fishes. In *In Evolutionary genetics of fishes* (pp. 95–171). Springer. https://doi.org/doi.org/10.1007/978-1-4684-4652-4_3

Kallman, K. D., & Schreibman, M. P. (1973). A sex-linked gene controlling gonadotrop differentiation and its significance in determining the age of sexual maturation and size of the platyfish, Xiphophorus maculatus. *General and Comparative*

Endocrinology, *21*(2), 287–304. https://doi.org/10.1016/0016-6480(73)90061-0

Katoh, K., Rozewicki, J., & Yamada, K. D. (2018). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Kimura, M., & Ohta, T. (1971). On the rate of molecular evolution. *Journal of Molecular Evolution*, *1*(1), 1–17. https://doi.org/10.1007/BF01659390

King, K. C., Jokela, J., & Lively, C. M. (2011). Parasites, sex, and clonal diversity in natural snail populations. *Evolution*, *65*(5), 1474–1481. https://doi.org/10.1111/j.1558-5646.2010.01215.x

Kondrashov, A. S. (1994). Muller's ratchet under epistatic selection. *Genetics*, *136*(4), 1469–1473. https://doi.org/10.1093/genetics/136.4.1469

Kottler, V. A., Feron, R., Nanda, I., Klopp, C., Du, K., Kneitz, S., Helmprobst, F., Lamatsch, D. K., Lopez-Roques, C., Lluch, J., Journot, L., Parrinello, H., Guiguen, Y., & Schartl, M. (2020). Independent Origin of XY and ZW Sex Determination. *Genetics*, *214*(1), 193–209. https://doi.org/10.1534/genetics.119.302698

Kowal, K., Tkaczyk, A., Pierzchała, M., Bownik, A., & Ślaska, B. (2020). Identification of mitochondrial dna (Numts) in the nuclear genome of daphnia magna. *International Journal of Molecular Sciences*, *21*(22), 1–17. https://doi.org/10.3390/ijms21228725

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, *41*(Web Server issue), 22–28. https://doi.org/10.1093/nar/gkt389

Ladle, R. J. (1992). Parasites and sex: Catching the red queen. *Trends in Ecology and Evolution*, *7*(12), 405–408. https://doi.org/10.1016/0169-5347(92)90021-3

Ladoukakis, E. D., & Zouros, E. (2017). Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions. *Journal of Biological Research (Greece)*, *24*(1), 1–7. https://doi.org/10.1186/s40709-017-0060-4

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). https://doi.org/10.1186/gb-2009-10-3-r25

Lanza, J. (1983). Population control of an all-female fish, Poeciliopsis monacha-latidens. *Biotropica*, 100–107. https://doi.org/10.2307/2387951

Lanza, Janet. (1983). Microhabitat use by bisexual and unisexual fishes (Poeciliopsis: Poeciliidae) in an artificial stream. *Oecologia*, *57*(1–2), 142–147. https://doi.org/10.1007/BF00379571

Lewis, W. M. (1987). The cost of sex. In S. C. Stearns (Ed.), *The Evolution of Sex and its Consequences* (pp. 33–57). Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-6273-8_2

Li, C., Ortí, G., Zhang, G., & Lu, G. (2007). A practical approach to phylogenomics: The phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC*

Evolutionary Biology, *7*, 1–11. https://doi.org/10.1186/1471-2148-7-44

Lima, N. R. W., Kobak, C. J., & Vrijenhoek, R. C. (1996). Evolution of sexual mimicry in sperm-dependent all-female forms of Poeciliopsis (Pisces: Poeciliidae). *Journal of Evolutionary Biology*, *9*(2), 185–203. https://doi.org/10.1046/j.1420-9101.1996.9020185.x

Liu, J., Huang, S., Sun, M., Liu, S., Liu, Y., Wang, W., Zhang, X., Wang, H., & Hua, W. (2012). An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods*, *8*(1), 1–9. https://doi.org/10.1186/1746-4811-8-34

Lively, C. M., Craddock, C., & Vrijenhoek, R. C. (1990). Red Queen hypothesis supported by parasitism in sexual and clonal fish. *Nature*, *344*(6269), 864–866. https://doi.org/10.1038/344864a0

Long, Y. M., Chao, W. S., Ma, G. J., Xu, S. S., & Qi, L. L. (2017). An innovative SNP genotyping method adapting to multiple platforms and throughputs. *Theoretical and Applied Genetics*, *130*(3), 597–607. https://doi.org/10.1007/s00122-016-2838-4

Lovell, J. T., Williamson, R. J., Wright, S. I., McKay, J. K., & Sharbel, T. F. (2017). Mutation Accumulation in an Asexual Relative of Arabidopsis. *PLoS Genetics*, *13*(1), 1–14. https://doi.org/10.1371/journal.pgen.1006550

Mateos, M., Domínguez-Domínguez, O., & Varela-Romero, A. (2019). A multilocus phylogeny of the fish genus Poeciliopsis: Solving taxonomic uncertainties and preliminary evidence of reticulation. *Ecology and Evolution*, *9*(4), 1845–1857. https://doi.org/10.1002/ece3.4874

Mateos, M., Kang, D., Klopp, C., Parrinello, H., García-Olazábal, M., Schumer, M., Jue, N. K., Guiguen, Y., & Schartl, M. (2019). Draft Genome Assembly and Annotation of the Gila Topminnow Poeciliopsis occidentalis. *Frontiers in Ecology and Evolution*, *7*, 404. https://doi.org/10.3389/fevo.2019.00404

Mateos, M., Sanjur, O. I., & Vrijenhoek, R. C. (2002). Historical biogeography of the livebearing fish genus Poeciliopsis (Poeciliidae: Cyprinodontiformes). *Evolution*, *56*(5), 972–984. https://doi.org/10.1111/j.0014-3820.2002.tb01409.x

Mateos, M., & Vrijenhoek, R. C. (2002). Ancient versus reticulate origin of a hemiclonal lineage. *Evolution*, *56*(5), 985–992. https://doi.org/10.1111/j.0014-3820.2002.tb01410.x

Mateos, M., & Vrijenhoek, R. C. (2005). Independent origins of allotriploidy in the fish genus Poeciliopsis. *Journal of Heredity*, *96*(1), 32–39. https://doi.org/10.1093/jhered/esi010

McKay, F. E. (1971). Behavioral aspects of population dynamics in unisexual-bisexual Poeciliopsis (Pisces: Poeciliidae). *Ecology*, *52*(5), 778–790. https://doi.org/10.2307/1936025

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, *11*(1), 31–46. https://doi.org/10.1038/nrg2626

Miller, R. R. (1960). *Four new species of viviparous fishes, genus Poeciliopsis, from northwestern Mexico*. https://doi.org/10.1126/science.46.1200.643

Miller, R. R., Minckley, W. L., & Norris, S. M. (2005). *Freshwater fishes of Mexico* (p. 490). Chicago: The University of Chicago Press.

Morishima, K., Nakamura-Shiokawa, Y., Bando, E., Li, Y. J., Boroń, A., Khan, M. M. R., & Arai, K. (2008). Cryptic clonal lineages and genetic diversity in the loach Misgurnus anguillicaudatus (Teleostei: Cobitidae) inferred from nuclear and mitochondrial DNA analyses. *Genetica*, *132*(2), 159–171. https://doi.org/10.1007/s10709-007-9158-1

Moritz, C., Wright, J. W., & Brown, W. M. (1992). Mitochondrial DNA analyses and the origin and relative age of parthenogenetic Cnemidophorus: phylogenetic constraints on hybrid origins. *Evolution*, *46*(1), 184–192. https://doi.org/10.1111/j.1558-5646.1992.tb01993.x

Morran, L. T., Parrish, R. C., Gelarden, I. A., Allen, M. B., & Lively, C. M. (2014). Experimental coevolution: Rapid local adaptation by parasites depends on host mating system. *American Naturalist*, *184*(SUPPL. 1). https://doi.org/10.1086/676930

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *1*(1), 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Nandakumar, P., Tian, C., O'Connell, J., Hinds, D., Paterson, A. D., & Sondheimer, N. (2021). Nuclear genome-wide associations with mitochondrial heteroplasmy. *Science Advances*, *7*(12), 1–10. https://doi.org/10.1126/sciadv.abe7520

Neaves, W. B., & Baumann, P. (2011). Unisexual reproduction among vertebrates. *Trends in Genetics*, *27*(3), 81–88. https://doi.org/10.1016/j.tig.2010.12.002

Neiman, M., Hehman, G., Miller, J. T., Logsdon, J. M., & Taylor, D. R. (2010).

Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Molecular Biology and Evolution*, *27*(4), 954–963. https://doi.org/10.1093/molbev/msp300

Normark, B. B., & Moran, N. A. (2000). Testing for the accumulation of deleterious mutations in asexual eukaryote genomes using molecular sequences. *Journal of Natural History*, *34*(9), 1719–1729. https://doi.org/10.1080/00222930050122147

Olsen, M. T., Pampoulie, C., Daníelsdóttir, A. K., Lidh, E., Bérubé, M., Víkingsson, G. A., & Palsbøll, P. J. (2014). Fin whale MDH-1 and MPI allozyme variation is not reflected in the corresponding DNA sequences. *Ecology and Evolution*, *4*(10), 1787–1803. https://doi.org/10.1002/ece3.1046

Park, S. W., & Bolker, B. (2019). Bridging the gap between theory and data: the Red Queen Hypothesis for sex. *BioRxiv*, 637413. https://doi.org/10.1101/637413

Quattro, J. M., Avise, J. C., & Vrijenhoek, R. C. (1991). Molecular evidence for multiple origins of hybridogenetic fish clones (Poeciliidae: Poeciliopsis). *Genetics*, *127*(2), 391–398. https://doi.org/doi.org/10.1093/genetics/127.2.391

Quattro, Joseph M., Avise, J. C., & Vrijenhoek, R. C. (1992). An ancient clonal lineage in the (Atheriniformes: Poeciliidae. *Molecular Ecology and Evolution: The Organismal Side: Selected Writings from the Avise Laboratory*, *89*(January), 114–118. https://doi.org/10.1142/9789814317764_0002

Radojičić, J. M., Krizmanić, I., Kasapidis, P., & Zouros, E. (2015). Extensive mitochondrial heteroplasmy in hybrid water frog (Pelophylax spp.) populations from Southeast Europe. *Ecology and Evolution*, *5*(20), 4529–4541.

https://doi.org/10.1002/ece3.1692

Sazanov, L. A. (2015). A giant molecular proton pump: Structure and mechanism of respiratory complex I. *Nature Reviews Molecular Cell Biology*, *16*(6), 375–388. https://doi.org/10.1038/nrm3997

Schenk, M. E. (1991). Genetics and Geographical Distribution of a New Clonally Derived Sexual Species of Poeciliopsis (Atheriniformes; Poecilidae). *(Doctoral Dissertation, Rutgers University)*.

Schmidt, D. J., Bond, N. R., Adams, M., & Hughes, J. M. (2011). Cytonuclear evidence for hybridogenetic reproduction in natural populations of the Australian carp gudgeon (Hypseleotris: Eleotridae). *Molecular Ecology*, *20*(16), 3367–3380. https://doi.org/10.1111/j.1365-294X.2011.05206.x

Schultz, R. J. (1961). Reproductive Mechanism of Unisexual and Bisexual Strains of the Viviparous Fish Poeciliopsis. *Evolution*, *15*(3), 302–325. https://doi.org/doi.org/10.2307/2406230

Schultz, R. J. (1967). Gynogenesis and triploidy in the viviparous fish Poeciliopsis. *Science*, *157*(3796), 1564–1567. https://doi.org/10.1126/science.157.3796.1564

Schultz, R. J. (1969). Hybridization, unisexuality, and polyploidy in the teleost Poeciliopsis (Poeciliidae) and other vertebrates. *The American Naturalist*, *103*(934), 605–619. https://doi.org/10.1086/282629

Schultz, R. J. (1977). Evolution and Ecology of Unisexual Fishes. In *Evolutionary Biology* (pp. 227–331). https://doi.org/doi.org/10.1007/978-1-4615-6953-4_6

Schultz, R. Jack. (1966). Hybridization Experiments with an All-Female Fish of the

Genus Poeciliopsis. *The Biological Bulletin*, *130*(3), 415–429. https://doi.org/10.2307/1539747

Semagn, K., Babu, R., Hearne, S., & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. *Molecular Breeding*, *33*(1), 1–14. https://doi.org/10.1007/s11032-013-9917-x

Sharbrough, J., Luse, M., Boore, J. L., Logsdon, J. M., & Neiman, M. (2018). Radical amino acid mutations persist longer in the absence of sex. *Evolution*, *72*(4), 808–824. https://doi.org/10.1111/evo.13465

Sharpley, M. S., Marciniak, C., Eckel-Mahan, K., McManus, M., Crimi, M., Waymire, K., Lin, C. S., Masubuchi, S., Friend, N., Koike, M., Chalkia, D., MacGregor, G., Sassone-Corsi, P., & Wallace, D. C. (2012). Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. *Cell*, *151*(2), 333–343. https://doi.org/10.1016/j.cell.2012.09.004

Smith, J. M. (1978). The evolution of sex. In *Cambridge: Cambridge University Press* (Vol. 4). https://doi.org/10.1017/S0021937100024837

Stewart, J. B., & Chinnery, P. F. (2015). The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease. *Nature Reviews Genetics*, *16*(9), 530–542. https://doi.org/10.1038/nrg3966

Stöck, M., Croll, D., Dumas, Z., Biollay, S., Wang, J., & Perrin, N. (2011). A cryptic heterogametic transition revealed by sex-linked DNA markers in Palearctic green toads. *Journal of Evolutionary Biology*, *24*(5), 1064–1070.

https://doi.org/10.1111/j.1420-9101.2011.02239.x

Stöck, Matthias, Kratochvíl, L., Kuhl, H., Rovatsos, M., Evans, B., Suh, A., Valenzuela, N., Veyrunes, F., Zhou, Q., Gamble, T., Capel, B., Schartl, M., Guiguen, Y., Authors, C., & Authors, C. (2021). A brief review of vertebrate sex evolution with a pledge for integrative research-towards "sexomics." *Philosophical Transactions of the Royal Society B*, *376*(1832), 20200426. http://mc.manuscriptcentral.com/issue-ptrsb

Stöck, Matthias, Lampert, K. P., Möller, D., Schlupp, I., & Schartl, M. (2010). Monophyletic origin of multiple clonal lineages in an asexual fish (Poecilia formosa). *Molecular Ecology*, *19*(23), 5204–5215. https://doi.org/10.1111/j.1365-294X.2010.04869.x

Sullivan, J. A., & Schultz, R. J. (1986). Genetic and Environmental Basis of Variable Sex Ratios in Laboratory Strains of Poeciliopsis lucida. *Evolution*, *40*(1), 152. https://doi.org/10.2307/2408612

Suzuki, S., Miyake, S., Arai, K., & Munehara, H. (2020). Unisexual hybrids break through an evolutionary dead end by two-way backcrossing. *Evolution*, *74*(2), 392–403. https://doi.org/10.1111/evo.13903

Swofford, D. L., & Sullivan, J. (2012). Phylogeny inference based on parsimony and other methods using PAUP. *The Phylogenetic Handbook*, 267–312. https://doi.org/10.1017/cbo9780511819049.010

Syvänen, A. C., Aalto-Setälä, K., Harju, L., Kontula, K., & Söderlund, H. (1990). A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E.

Genomics, *8*(4), 684–692. https://doi.org/10.1016/0888-7543(90)90255-S

Taylor, S., Scott, R., Kurtz, R., Fisher, C., Patel, V., & Bizouarn, F. (2010). A practical guide to high resolution melt analysis genotyping. *Bio-Rad Bull*, 1–9.

Tejedo, M., Semlitsch, R. D., & Hotz, H. (2000). Differential Morphology and Jumping Performance of Newly Metamorphosed Frogs of the Hybridogenetic Rana esculenta Complex Author (s): Miguel Tejedo , Raymond D . Semlitsch and Hansjürg Hotz Published by : *Journal of Herpetology*, *34*(2), 201–210. https://doi.org/10.2307/1565416

Tobler, M., Wahli, T., & Schlupp, I. (2005). Comparison of parasite communities in native and introduced populations of sexual and asexual mollies of the genus Poecilia. *Journal of Fish Biology*, *67*(4), 1072–1082. https://doi.org/10.1111/j.0022-1112.2005.00810.x

Trifinopoulos, J., Nguyen, L. T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*(W1), W232–W235. https://doi.org/10.1093/nar/gkw256

Volff, J. N., & Schartl, M. (2001). Variability of genetic sex determination in poeciliid fishes. *Genetica*, *111*(1–3), 101–110. https://doi.org/10.1023/A:1013795415808

Volff, J. N., & Schartl, M. (2002). Sex determination and sex chromosome evolution in the medaka, Oryzias latipes, and the platyfish, Xiphophorus maculatus. *Cytogenetic and Genome Research*, *99*(1–4), 170–177. https://doi.org/10.1159/000071590

Vrijenhoek, R. C. (1972). Genetic relationships of unisexual-hybrid fishes to their progenitors using lactate dehydrogenase isozymes as gene markers (Poeciliopsis,

Poeciliidae). *The American Naturalist*, *106*(952), 754–766. https://doi.org/10.1086/282810

Vrijenhoek, R. C. (1975). Gene dosage in diploid and triploid unisexual fishes (Poeciliopsis, Poeciliidae). In *Isozymes* (pp. 463–475). Academic Press. https://doi.org/10.1016/b978-0-12-472704-5.50034-3

Vrijenhoek, R. C. (1979). Genetics of a Sexually Reproducing Fish in a Highly Fluctuating Environment. *The American Naturalist*, *113*(1), 17–29. https://doi.org/10.1086/283362

Vrijenhoek, R. C., & Schultz, R. J. (1974). Evolution of a Trihybrid Unisexual Fish (Poeciliopsis, Poeciliidae). *Evolution*, *28*(2), 306–319. https://doi.org/10.1111/j.1558-5646.1974.tb00750.x

Vrijenhoek, R C. (1989). Genetic and ecological constraints on the origins and establishment of unisexual vertebrates. *Evolution and Ecology of Unisexual Vertebrates*, *466*, 24–31.

Vrijenhoek, Robert C. (1979). Factors affecting clonal diversity and coexistence. *Integrative and Comparative Biology*, *19*(3), 787–797. https://doi.org/10.1093/icb/19.3.787

Vrijenhoek, Robert C. (1994). Unisexual fish: model systems for studying ecology and evolution. *Annual Review of Ecology and Systematics*, *25*(1), 71–96. https://doi.org/10.1146/annurev.es.25.110194.000443

Vrijenhoek, Robert C., Angus, R. A., & Schultz, R. J. (1977). Variation and Heterozygosity in Sexually vs. Clonally Reproducing Populations of Poeciliopsis.

Evolution, *31*(4), 767. https://doi.org/10.2307/2407438

Vrijenhoek, Robert C., Angus, R. A., & Schultz, R. J. (1978). Variation and Clonal Structure in a Unisexual Fish. *The American Naturalist*, *112*(983), 41–55. https://doi.org/10.1086/283251

Vrijenhoek, Robert C., Pfeiler, E., & Wetherington, J. D. (1992). Balancing Selection in a Desert Stream-Dwelling Fish, Poeciliopsis monacha. *Evolution*, *46*(6), 1642. https://doi.org/10.2307/2410021

Williams, G. C. (2020). *Sex and Evolution. ({MPB}-8), Volume 8*. Princeton University Press.

Woodward, J. (2014). Bi-allelic SNP genotyping using the TaqMan® assay. In *Methods in Molecular Biology* (Vol. 1145, pp. 67–74). Springer New York. https://doi.org/10.1007/978-1-4939-0446-4_6

Xu, B., & Yang, Z. (2013). PamlX: A graphical user interface for PAML. *Molecular Biology and Evolution*, *30*(12), 2723–2724. https://doi.org/10.1093/molbev/mst179

Zeyl, C., Mizesko, M., & Visser, J. A. G. M. de. (2001). Mutational Meltdown in Laboratory Yeast Populations. *Evolution*, *55*(5), 909–917. https://doi.org/10.1554/0014-3820(2001)055[0909:mmilyp]2.0.co;2

Zhang, J. (2000). Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *Journal of Molecular Evolution*, *50*(1), 56–68. https://doi.org/10.1007/s002399910007

APPENDIX

Table S1. Statistics of raw sequence reads of the 31 samples.

Voucher	Species	Number of raw reads	Number reads after Trimmomatic QC	Number of reads mapped to reference mt genome	Mean coverage
Cx-B S68-5	*P. monacha-monacha-lucida*	55,317,848	52,025,189	214,241	1,030
MVH99-12#3	*P. monacha-lucida*	49,144,945	45920,073	573,627	4,936
MVH99-18#3	*P. monacha-lucida*	49,054,209	46110,956	438,379	3,759
ML IV Cz M61-35_Hap_C	*P. monacha-lucida*	60,605,687	57271,235	169,422	1,446
ML Cx I MVH99-16#5_Hap_B	*P. monacha-lucida*	46,166,556	43576,813	581,516	4,900
ML Cx II MVH99-16#48_Hap_A	*P. monacha-lucida*	40,046,780	37776,127	238,913	2,017
Cx SV73-7s_Hap_A	*P. monacha-lucida*	50,062,240	47,273,773	208,360	1,782
MVH99-18#1	*P. monacha-lucida*	49,099,037	45,824,131	383,673	3,269
MVH99-6#7	*P. monacha*	54,388,044	50,885,454	522,305	4,419
MVH99-15#10	*P. monacha-latidens*	53,636,275	50,573,930	135,963	1,149
MV00-6A#10_Hap_X	*P. monacha-latidens*	50,779,101	47,879,413	101,006	853
MVH99-17#72	*P. monacha-latidens*	66,875,143	62,568,383	494,879	4,091
MVH99-16#59	*P. monacha-latidens*	53,991,830	50,579,546	148,760	1,243
MVH99-17#61	*P. monacha-latidens*	53,764,398	50,000,890	378,129	3,140
MV00-10#45	*P. monacha-latidens*	46,949,385	44,132,421	454,079	3,676
M65-23#6	*P. viriosa*	70,453,048	67,353,113	247,343	2,204
Cw S68-4	*P. monacha-lucida*	62,515,238	59,764,568	379,966	3,290
Cx T70-3#M1	*P. monacha-lucida*	79,633,519	76,050,011	160,617	1,376
MV00-6A#11	*P. monacha-latidens*	75,479,988	72,187,665	155,093	1,330
Cx-B S68-4#M1	*P. monacha-monacha-lucida*	79,111,546	75,560,621	323,457	2,806
Cy M61-31#M1	*P. monacha-lucida-lucida*	81,286,385	77,628,497	211,193	1,823
Cy SV73-4	*P. monacha-lucida-lucida*	69,186,395	66,010,657	335,607	2,900
MVH99-12#1	*P. monacha-lucida*	88,060,759	84,213,982	810,066	6,909
Cx_Sv73-7s	*P. monacha-lucida*	84,196,070	80,541,960	413,232	3,576
Cx M65-24	*P. monacha-lucida*	68,245,879	65,141,877	315,797	2,667

Voucher	Species	Number of raw reads	Number reads after Trimmomatic QC	Number of reads mapped to reference mt genome	Mean coverage
MV00-9#1	*P. monacha-lucida-viriosa*	73,510,772	70,276,298	523,735	4,493
MV00-10#3	*P. monacha-lucida-viriosa*	87,834,834	83,670,937	1,131,980	9,067
MV00-8#17	*P. monacha-lucida-viriosa*	110,873,393	104,320,307	6,289,808	55,003
MVH99-16#22	*P. monacha-lucida-viriosa*	90,200,350	86,285,654	430,103	3,761
MVH99-16#9	*P. monacha-lucida-viriosa*	759,12,238	72,534,143	463,804	4,009
MV00-9#5	*P. monacha-lucida-viriosa*	105,348,899	101,134,943	7,775,749	68,138
Cx_II_MV00-10#18	*P. monacha-lucida*	65,799,046	63,561,879	539,095	4,691
Cx_I_MV00-10#33	*P. monacha-lucida*	99,651,633	95,309,811	7,666,676	6,679

Table S1 Continued

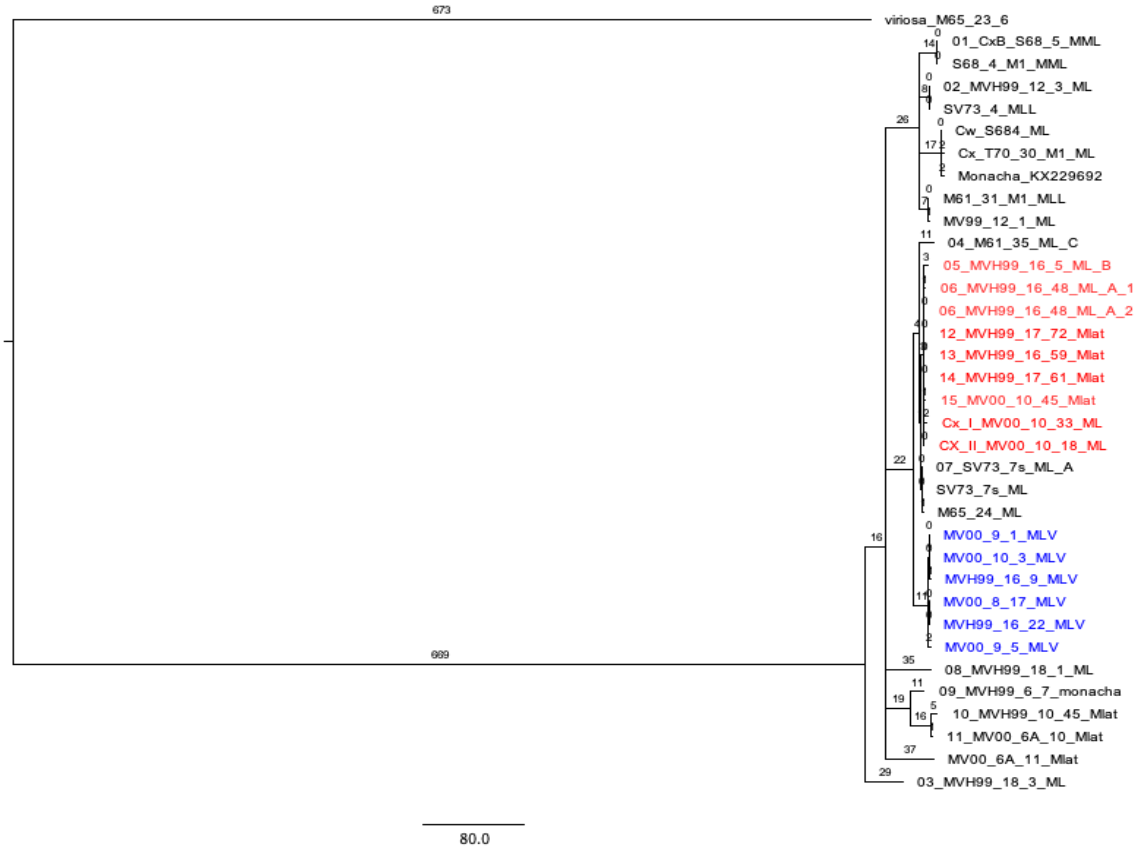

Figure S1. The strict consensus tree generated from 7 most parsimonious trees.

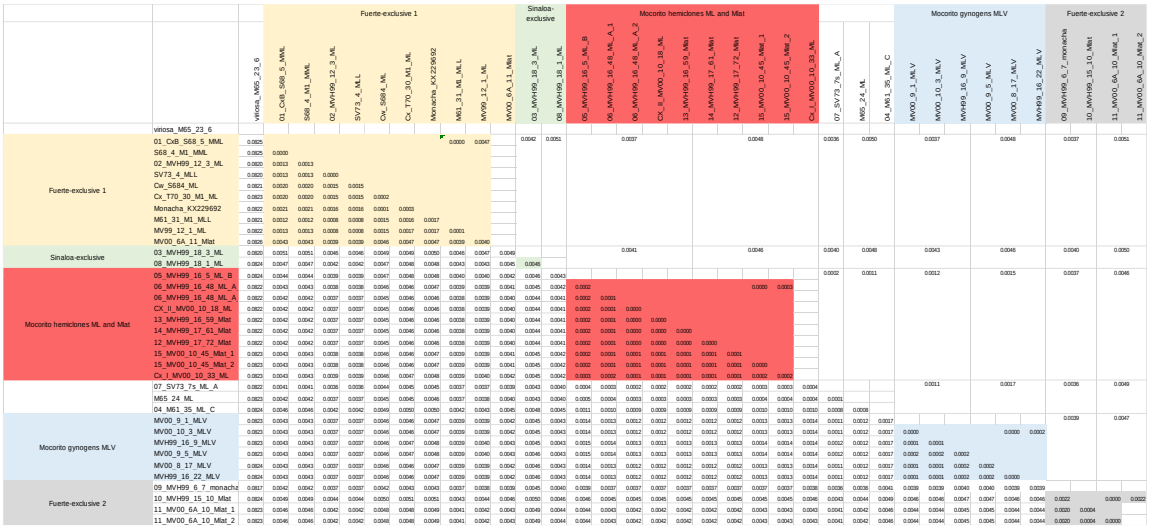

Figure S2. The maximum and minimum whole mitochondrial divergence (uncorrected pairwise p distance) within and between the 5 clades identified in Figure 8. The number of base differences per site from between sequences are shown.

		Kc	Kr	Kn	Ks	Kn+Ks	Kc/Ks	Kr/Ks	Kn/Ks	Kr/Kc	Kr/Kn	Kn>Ks	Kr>Kc	Kn>Ks	Kr>Kc
	P. monacha	2	0	2	6	8	0.33	0.00	0.33	0.00	0.00	no	no		
Fuerte-exclusive 1	P. monacha-	1	1	2	2	4	0.50	0.50	1.00	1.00	0.50	no	no		
	P. monacha-	1	0	1	0	1	1/0	1/0	1/0	0.00	0.00	yes	no	1>0	
	P. monacha-	2	1	3	10	13	0.20	0.10	0.30	0.50	0.33	no	no		
	P. monacha-	0	0	0	1	1	0.00	0.00	0.00	0/0	0/0	no	no		
	P. monacha	0	0	0	2	2	0.00	0.00	0.00	0/0	0/0	no	no		
Fuerte-exclusive 2	P. monacha-	0	0	0	1	1	0.00	0.00	0.00	0/0	0/0	no	no		
	P. monacha-	5	1	6	20	26	0.25	0.05	0.30	0.20	0.17	no	no		
Sinaloa-exclusive	P. monacha-	1	1	2	28	30	0.04	0.04	0.07	1.00	0.50	no	no		
	P. monacha-	5	1	6	18	24	0.28	0.06	0.33	0.20	0.17	no	no		
	P. monacha-	3	0	3	0	3	3/0	0/0	3/0	0.00	0.00	yes	no	3>0	
Mocorito hemiclones	P. monacha-	0	1	1	0	1	0/0	1/0	1/0	1/0	1.00	yes	yes	1>0	1>0
	P. monacha-	0	0	0	1	1	0.00	0.00	0.00	0/0	0/0	no	no		
ML and Mlat	P. monacha-	1	0	1	1	2	1.00	0.00	1.00	0.00	0.00	no	no		
	49-48	0	1	1	0	1	0/0	1/0	1/0	1/0	1.00	yes	yes	1>0	1>0
	P. monacha-	0	0	0	1	1	0.00	0.00	0.00	0/0	0/0	no	no		
	P. monacha-	1	1	2	1	3	1.00	1.00	2.00	1.00	0.50	yes	no	2>1	
gynogens MLV	P. monacha-	1	0	1	0	1	1/0	0/0	1/0	0.00	0.00	yes	no	1>0	
	P. monacha-	1	0	1	1	2	1.00	0.00	1.00	0.00	0.00	no	no		
	46-54 (MLV	2	3	5	4	9	0.50	0.75	1.25	1.50	0.60	yes	yes	5>4	3>2
	55-55	0	0	0	1	1	0.00	0.00	0.00	0/0	0/0	no	no		
	54-58	0	0	0	1	1	0.00	0.00	0.00	0/0	0/0	no	no		
	60-01	2	0	2	6	8	0.33	0.00	0.33	0.00	0.00	no	no		
	40-45	3	0	3	1	4	3.00	0.00	3.00	0.00	0.00	yes	no	3>1	
	40-42	2	0	2	9	11	0.22	0.00	0.22	0.00	0.00	no	no		
	40-44	2	0	2	2	4	1.00	0.00	1.00	0.00	0.00	no	no		

Figure S3. Numbers of radical changes (Kr), conservative changes (Kc), synonymous changes (Ks), and nonsynonymous changes (Kn). Kc/Ks, Kr/Ks, Kn/Ks and Kr/Kn were also calculated.

Table S2. Mitochondrial genes where radical changes are and the corresponding ETS complex. The positions of radical changes, ancestral and derived amino acid, CRI values were identified by PAMLX. ETS complex was identified based on Sazanov (2015).

Clade	Branch	Site	Ancestral AA	Derived AA	CRI	Mtgene	ETS Complex
39-40	3	643	Q	R	0.71	ND2	I
	3	3739	N	D	0.71	ND6	I
40-41	4	1308	D	N	0.71	COX2	IV
39-31 MV00-6A#11_Mlat	18	595	K	N	0.86	ND2	I
47-11 M61_35_ML_C	21	1413	S	L	0.71	COX2	IV
48-49	23	3657	Y	F	0.57	ND6	I
49-51	32	326	T	I	0.71	ND2	I
46-54	41	7	F	S	0.71	ND1	I
	41	373	H	Y	0.71	ND2	I
	41	2232	F	S	0.71	ND4	I
59-32 MVH99-18#1_ML	53	2661	Y	H	0.71	ND5	I
59-37 MVH99-18#3_ML	54	2708	I	T	0.71	ND5	I
60-61	57	118	T	M	0.71	ND1	I
61-34 MVH99-15#10_Mlat	58	689	T	I	0.71	COX1	I