Cytonuclear evidence for hybridogenetic reproduction in natural populations of
the Australian carp gudgeon (*Hypseleotris*: Eleotridae).

Daniel J. Schmidt*, Nicholas R. Bond†, Mark Adams‡ and Jane M. Hughes*

*Australian Rivers Institute and eWater CRC, Griffith University, Nathan, 4111, QLD, Australia.
† School of Biological Sciences and eWater CRC, Monash University, 3800, VIC Australia.
‡ Evolutionary Biology Unit, South Australian Museum, Adelaide, 5000, SA Australia.

Correspondence: Daniel J. Schmidt, Fax: +61 (07) 3875 7615; E-mail: d.schmidt@griffith.edu.au

Keywords: asexual reproduction; unisexual; hybridization; hemiclonal;
hybridogenesis; gynogenesis

Running Title: Genetic evidence for asexuality in carp gudgeons.
Abstract

Although most vertebrates reproduce sexually, a small number of fishes, amphibians, and reptiles are known in which reproduction is asexual i.e. without meiotic recombination. In fishes, these so-called “unisexual” lineages usually comprise only females, and utilize co-occurring males of a related sexual species to reproduce via gynogenesis or hybridogenesis. Here we examine patterns of microsatellite and mitochondrial DNA (mtDNA) variation in a widespread group of freshwater fishes (carp gudgeons; Hypseleotris spp.) to investigate a long-standing proposal that this group includes unisexual forms. We show that the mtDNA genome of most carp gudgeons in tributaries of the Goulburn River belong to one of two deeply divided clades (~10% cyt b divergence) and that nuclear variation divides the same individuals into four distinct groups. Group 1 exhibits the genotypic proportions of a random mating population and has a 1:1 sex ratio. Two other groups are extremely sex-biased (98% male, 96% female), exhibit excess heterozygosity at most loci and share at least one allele per locus with group 1. We propose that these two groups represent “unisexual” hybridogenetic lineages, and that both utilize co-occurring group 1 as sexual host. Interestingly, the fourth distinct group appears to represent hybrid offspring of the two putative hybridogenetic lineages. The propagation of clonal haploid genomes by both males and females and the ability of these clones to unite and form sexually mature diploid hybrid offspring may represent a novel mechanism that contributes to the dynamics of coexistence between hybridogenetic lineages and their sexual hosts.

Introduction
Unisexual taxa that reproduce asexually (i.e. without meiotic recombination) are rare and comprise <0.1% of extant vertebrate species (Dawley 1989). About one-third of these asexual vertebrates are freshwater fishes belonging to four families (Poeciliidae, Atherinidae, Cyprinidae and Cobitidae) (Lamatsch & Stöck 2009; Lampert 2009). All unisexual fish employ one of two modes of “sperm-dependent” reproduction, whereby mating with a related sexual species is required to produce their clonal or “hemi-clonal” offspring (Schultz 1969; Vrijenhoek 1994; Lamatsch & Stöck 2009; Lampert 2009). The first of these modes is gynogenesis, in which sperm triggers embryogenesis without incorporation of the paternal genome, resulting in production of clonal all-female offspring (Dawley 1989; Vrijenhoek 1994; Avise 2008). The second asexual mode is hybridogenesis, which involves differential transmission of maternal and paternal genomes. This system typically involves a unisexual female lineage mating with males of a sympatric sexual species to produce offspring that are all-female hybrids, but differs in that eggs are produced without recombination and consist exclusively of the maternal genome, due to a premeiotic cell division that excludes the paternal genome (Schultz 1969; Dawley 1989). Hybridogens can be considered “hemiclones” because one genome is inherited clonally and the other sexually, although only the clonal set of chromosomes (usually maternal) is transmitted across subsequent generations (Dawley 1989; Avise 2008). To further complicate matters, unisexual lineages may be diploid, triploid, or even tetraploid (Lamatsch & Stöck 2009).

Well-known hybridogenetic systems vary in the degree to which “unisexuality” is strictly observed. Natural and laboratory synthesised hybridogenetic lineages of fish
in the Poeciliopsis system are invariably all-female (Schultz 1969; Schultz 1973; Wetherington et al. 1987). Hybridogenetic water frogs of the Pelophylax esculenta complex exhibit a range of sex ratios in natural populations and both sexes may propagate hemiclonal genomes (Graf & Polls Pelaz 1989). This unusual system appears to be partially explained by segregation of the hemiclonal genome with haploid chromosome sets bearing either the X or Y sex chromosome (Graf & Polls Pelaz 1989; Christiansen 2009).

Numerous studies have focussed on understanding the evolutionary origins of asexual lineages as well as ecological and behavioural mechanisms that allow their coexistence with sexual “host” species on which they depend for reproduction (reviewed in: Beukeboom & Vrijenhoek 1998; Schlupp 2005; Avise 2008; Lampert 2009). All known gynogenetic and hybridogenetic vertebrate lineages appear to have arisen spontaneously as a result of interspecies hybridization between two sexually reproducing parental species (Avise 2008). The resulting unisexual lineage may rely on one of the parental species as its sexual host, or switch to a new host (Choleva et al. 2008). Parental species contributing to the formation of a unisexual lineage typically exhibit a moderately high degree of genetic divergence (Avise 2008). For example, cytochrome b divergence between parental species in the Poeciliopsis hybridogenetic system and the Phoxinus gynogenetic system is approximately 10% (Quattro et al. 1992; Angers & Schlosser 2007). Hybridization events between these phylogenetically distant taxa may strike a rare balance between disruption of meiosis in hybrids without compromising their viability or fertility (Wetherington et al. 1987; Avise 2008; Stöck et al. 2010).
Both gynogens and hybridogens may be considered “sexual parasites” that can rapidly outcompete their sperm-donating host due to the two-fold growth rate of asexual lineages relative to sexual ones (Kokko et al. 2008; Heubel et al. 2009). However, while asexual lineages may derive demographic benefits from these short-term advantages, they also render themselves more susceptible to extinction in the long-term, since the lack of recombination allows deleterious mutations to accumulate (Muller 1932; Vrijenhoek 1994). Together with the unusual genetic preconditions required for their formation, this genetic inflexibility might contribute to the rarity of gynogenesis and hybridogenesis which is currently known in only nine genera of freshwater fishes (Lamatsch & Stöck 2009; Stöck et al. 2010). Beukeboom & Vrijenhoek (1998) speculated that additional cases of sperm-dependent parthenogenesis might exist that remain undetected owing to their cryptic characteristics, and that application of molecular markers in natural populations might reveal new cases of hybridogenesis. However, despite the ever growing usage of molecular genetic data, only three new instances of vertebrate unisextuality have been confirmed in the past 20 years, none of which are instances of typical hemiclonal hybridogenesis (Dawley 1992; Adams et al. 2003; Morishima et al. 2008). Combined use of nuclear and mitochondrial markers can be useful for detecting the existence of hemiclonal patterns of inheritance in natural populations (Beukeboom & Vrijenhoek 1998). Non-random patterns of association between mtDNA and nuclear alleles will develop once a unisexual hybrid is formed, due to the coupling of nuclear and cytoplasmic genes in the clonal lineage (Vrijenhoek 1994). Combined analysis of nuclear markers and mtDNA in natural populations has been important in elucidating all of the known unisexual systems in fish. Examples include Poeciliopsis (Mateos &
Although no confirmed cases of sperm-dependent parthenogenesis have been reported from the Australian continent, Bertozzi et al. (2000) suggested that unisexual lineages were a plausible explanation for unusual patterns of genetic variation observed in carp gudgeons (*Hypseleotris*) from the lower Murray River. Their study identified three distinct species with fixed allozyme differences, designated *HA*, *HB*, *HC*. Additionally, three F$_1$ hybrid classes were found (*HA×HB, HB×HX, HA×HX*). Two of the hybrid classes contain the haploid genome designated *HX* which was not sampled in its pure diploid form, even though *HA×HX* and *HB×HX* hybrids comprised 21% of the sample (n =106). Bertozzi et al. (2000) speculated that one or more of the hybrid classes might represent unisexual lineages and that taxon *HX* might be absent (or extinct) in the study area. Under this scenario *HX* could represent a clonal lineage that exists only in hybrid form as either a unisexual (female) gynogen, or as a unisexual hemiclonal hybridogen. However, limited sample sizes and lack of gender information did not permit further exploration of their data. Recent phylogeographic treatment of the carp gudgeon species complex in eastern Australia by Thacker et al. (2007) revealed several divergent mtDNA clades loosely corresponding with some of the informal taxonomic designations used to identify carp gudgeon morphotypes. However, it is currently not possible to reconcile these phylogeographic patterns with the allozyme hybrid classes proposed by Bertozzi et al. (2000) as the two datasets involve different samples.
The aim of this study is to determine whether patterns of genetic variation in natural populations of the carp gudgeon can be attributed to a unisexual mode of reproduction and to characterise the lineages involved in hybridisation. We predict that if carp gudgeons comprise a complex of clonal or hemiclonal lineages and sexual host species, then analysis of mitochondrial and nuclear genetic variation in these populations should support four key expectations. First, well-defined groups of multilocus genotypes should be present, each corresponding to a sexual or hybrid lineage. Second, groups corresponding to hybrid lineages should be unisexual and will commonly exhibit a high degree of mtDNA divergence from the co-distributed sexual species. Third, the apportionment of total genetic variance within individuals relative to their group (\(F_{IS}\)) should be negative in clonal hybrid lineages, owing to fixed heterozygosity (Balloux et al. 2003). Indeed, large negative values of \(F_{IS}\) may be considered as the ultimate signature of clonal reproduction in natural populations of diploid organisms (Halkett et al. 2005b), and all sperm-dependent clonal and hemiclonal lineages show high, fixed heterozygosity relative to their sexual host(s) (e.g. Vrijenhoek 1977; Dawley 1992; Carmona et al. 1997). Fourth, a gynogenetic mode of reproduction may be indicated in natural populations if unisexual hybrid lineages are all-female and if a high proportion of identical multilocus genotypes exist in the population, whereas hybridogenetic reproduction may involve males and/or females, and these lineages should display greater genetic heterogeneity while also sharing at least one allele per locus with their sexual host species.

**Materials and Methods**

**Study species, sampling and DNA extraction**
Carp gudgeons are small freshwater fish (<60mm long) that are abundant and widely distributed throughout inland and coastal waterways of southeastern Australia. They often are a numerically dominant taxon in fish communities, with population sizes varying dramatically in response to stream flow (Perry & Bond 2009). Four carp gudgeon taxa are recognised from the Murray-Darling basin, although only the western carp gudgeon, *H. klunzingeri* (Ogilby) has formal taxonomic status. The validity of three other taxa has not been established; these include Midgley’s carp gudgeon, Murray-Darling carp gudgeon and Lake’s carp gudgeon (Allen *et al.* 2002). All four taxa occur together in the Goulburn River system of central Victoria (Thacker *et al.* 2007), the location of this study. Fertilisation is external and males aggressively guard eggs until hatching (Unmack 2000).

Adult fish were collected from the Granite Creeks system, which comprises a number of small tributaries of the Goulburn River in central Victoria, southeastern Australia (including Castle Creek, Faithful Creek, Honeysuckle Creek, Seven Creeks and Pranjip Creek). The granite creeks all arise in the Strathbogie Ranges and join the Goulburn River within a short distance of one another (~10km; Fig. 1). During periods of low flow these streams contract to isolated waterholes and from 2007 to 2010 the location of all remnant pools was mapped by helicopter during the maximum dry phase in each year. This mapping allowed the identification of key refuge areas throughout the stream network, and was followed by extensive surveys to record the distribution and abundance of fish. Carp gudgeons were found to be patchily distributed but locally abundant, and sufficient numbers for inclusion in genetic analyses were collected from 14 waterholes (Fig. 1; Table 1). Fish were collected
using a combination of Fyke nets set overnight and, where necessary, repeated seine trawls to increase the number of fish collected. The extent of channel drying and seasonal habitat loss coupled with the observed distribution of carp gudgeons across the full array of sites surveyed (~50) suggests that most of the more persistent local populations were represented in the sample.

Adult fish were preserved in 95% ethanol and total genomic DNA was extracted from muscle tissue using a standard phenol-chloroform extraction method (Sambrook *et al.* 1989). The sex of ethanol preserved individuals >25mm total length was determined by examination of the urogenital papilla using a stereomicroscope. To facilitate the cross-referencing of this study with the allozyme study of Bertozzi *et al.* (2000), we genotyped a series of voucher specimens from the lower Murray for the same suite of allozyme loci. This allowed us to incorporate reference samples of known identity, as per the nomenclature of Bertozzi *et al.* (2000), into our mtDNA and microsatellite analyses. These voucher tissues were obtained from the frozen tissue collection maintained at the South Australian Museum (Table S1. Supporting Information).

*mtDNA sequencing and PCR-RFLP assay*

Mitochondrial DNA (mtDNA) was amplified in two overlapping fragments using primers HYPSLA and HYPSHD for the 5’ end of cytochrome *b* (*cyt b*), and HYP510 and PH15938 for the 3’ end of *cyt b* (Thacker *et al.* 2007). PCR conditions were 4 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 45 s at 53 °C, 45 s at 72 °C, and a final extension cycle of 7 min at 72 °C. Initially, mtDNA variation was assessed for 186 samples by sequencing the 5’ *cyt b* fragment using
primer HYPSHD. A subset of samples including vouchers (Table S2. Supporting Information) was also sequenced in both directions for the 3′ cyt b fragment to produce a full-length concatenated cyt b dataset for phylogenetic analysis. Enzyme purification of PCR products was followed by sequencing with the BigDye® Terminator v3.1 Cycle Sequencing Kit. Sequencing was performed on a 3130xl Genetic Analyser (Applied Biosystems) at Griffith University DNA sequencing facility. Sequences were edited and aligned using Sequencher v4.1.2 (GeneCodes Corp.).

Analysis of 186 cyt b sequences revealed two common mtDNA clades (clades D, H sensu Thacker et al 2007) within the study area. A polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) analysis of the 5′ cyt b fragment was developed to distinguish the two clades for an additional 296 samples. PCR product from each sample was digested separately with 2 U of SphI (3 hr incubation at 37 °C) and 2 U of Tsp509I (3 hr incubation at 65 °C). Incubation of PCR product from clade H with Tsp509I produced two bands at 189 and 478 bp, while incubation with SphI produced an uncut fragment of 677 bp. Incubation of PCR product from clade D with SphI produced two bands at 295 and 372 bp and an uncut fragment with Tsp509I. This assay also distinguished the mtDNA of two additional Hypseleotris clades (clades A, L) found in central Victoria by Thacker et al. (2007). Clade L produced 3 bands at 79, 189 and 399 bp with Tsp509I and an uncut fragment with SphI. Clade A produced an uncut fragment of 677 bp with both enzymes. Any individuals with PCR-RFLP results that were ambiguous or different to the profile expected for clades D and H were sequenced to determine their identity.
Phylogenetic analysis of mtDNA data

Bayesian phylogenetic analysis of full-length cyt $b$ sequences was performed using MrBayes version 3.1.2 (Ronquist & Huelsenbeck 2003). An appropriate substitution model (GTR+I+G) was selected using the Akaike information criterion in MrModeltest v2.3 (Nylander 2004). Posterior probabilities of model parameters, topology and branch lengths were estimated using Markov Chain Monte Carlo (MCMC) settings of 2 million generations, sampled every 100 generations with temperature set to 0.1. Two simultaneous runs were used to assess convergence of parameter estimates on the stationary distribution using run diagnostics in the MrBayes output (average standard deviation of split frequencies). Each run was monitored to ensure that convergence was achieved within the first 25% of generations, and parameter estimates made during this period were discarded. Default priors were used for all model parameter estimates and random trees were used to start each MCMC run. Uncorrected pairwise divergence between mtDNA clades was calculated using MEGA v5 (Tamura et al. 2007).

Microsatellite development and genotyping

Genomic DNA from two individuals belonging to mtDNA clades $D$ and $H$ were pooled to develop microsatellite loci. The enrichment protocol used by Real et al. (2009) was followed. Seven loci were polymorphic in individuals from each mtDNA clade and a 20-mer oligonucleotide tail was added to the 5′ end of each forward primer to facilitate fluorescent labelling (Real et al. 2009) (Table S3. Supporting Information). PCR conditions were: $1 \times$ reaction buffer, 1.5 mM MgCl$_2$, 0.1 $\mu$M tailed
forward primer, 0.4 µM reverse primer, 0.4 µM fluorescent labelled tag primer, 0.2 µM dNTP’s, 0.2 U Taq polymerase. Cycling conditions for all loci included 4 min at 95 °C, followed by 30 cycles of 30 s at 95 °C, 30 s at 63 °C, 30 s at 72 °C, and a final extension cycle of 30 min at 72 °C. Fragment analysis was conducted on a 3130 Genetic Analyser (Applied Biosystems) using GeneScan-500 LIZ size standard (Applied Biosystems) and genotypes were scored using GeneMapper v4.0 software (Applied Biosystems). A second round of amplification and genotyping of 30% of individuals showed the mean genotyping error rate per locus was 1.8%.

**Microsatellite analyses**

We tested for the existence of distinct genetic groups in the set of individual multilocus genotypes using a model-based Bayesian clustering method and genetic multivariate analysis. Individuals with mtDNA genomes belonging to clades D and H (n = 478) were treated simultaneously in each analysis. The probability of an admixture model was tested for clusters (K) ranging from one to 14 (total sites) using STRUCTURE 2.3.1 (Pritchard et al. 2000). Models were tested using eight independent MCMC simulations, each consisting of 1×10^6 iterations after a burn-in of 5×10^5 iterations. The most likely number of homogeneous clusters was assessed using the second-order rate of change L’’ (K) following Evanno et al. (2005) using the online application Structure Harvester (http://taylor0.biology.ucla.edu/struct_harvest). The model implemented in STRUCTURE assumes loci are unlinked and in Hardy-Weinberg equilibrium. These conditions are unlikely to be met in a population including clonal or hemiclonal hybrids due to fixed heterozygosity and linkage of multilocus haplotypes. Nevertheless several studies demonstrate that STRUCTURE is
robust to these violations and recovers biologically meaningful groupings of clonal
and partially clonal organisms that are supported by independent analyses (Halkett et
al. 2005a; Dutech et al. 2010; Montarry et al. 2010).

As multivariate analyses of genetic data are free of the assumptions made in the
model-based clustering approach of STRUCTURE (Jombart et al. 2009), we also
used a centred principal component analysis (PCA) to examine clustering of
individuals based on total variation of microsatellite allele frequencies without scaling
of alleles. The PCA was implemented in the package ADEGENET v1.1 (Jombart
project.org). For direct comparison with model-based clustering, individuals were
labelled according to the groupings recovered by STRUCTURE. We compared
microsatellite groups obtained in the present analysis with the allozyme classification
of Bertozzi et al. (2000) by genotyping 27 voucher samples corresponding to five
different allozyme groups (Table S1. Supporting information). A PCA was performed
incorporating the voucher samples and the 95% inertia ellipse for each allozyme
group was contrasted with previously defined clusters.

Distinct genetic groupings of individuals identified with STRUCTURE and PCA were
tested for departure of genotypic proportions from Hardy–Weinberg equilibrium
(HWE) using default settings in GENEPOP v4.0 (Rousset 2008). GENEPOP v4.0 was
also used to calculate fixation indices correlating the total genetic variance within
individuals to their group ($F_{IS}$) and to perform exact tests evaluating the hypothesis of
heterozygote excess in each group with Markov chain settings of 1000 batches and
10000 iterations per batch. The number of different multilocus genotypes in the
dataset was calculated using the software GENECLONE (Arnaud-Haond & Belkhir 2007).

Results

Phylogenetic analysis of mtDNA

A total of 186 individuals were sequenced for the 5′ cyt b fragment, giving an alignment of 582 bp with 90 variable sites and only nine haplotypes resolved in 3 divergent clades. The full-length cyt b alignment of 1140 bp included 24 sequences and comprised four representatives from the Goulburn study area; twelve sequences representing voucher specimens of the Hypseleotris allosyme groupings of Bertozzi et al. (2000); six sequences obtained from GenBank representing the Hypseleotris cyt b clades found by Thacker et al. (2007) in central Victoria; and two sequences obtained from Genbank that were used to root the Hypseleotris cyt b phylogeny of Thacker et al. (2007) and used as outgroups in the present analysis (Table S2. Supporting information).

The mtDNA gene tree recovered from Bayesian analysis of the cyt b alignment shows all ingroup sequences resolved into four well supported clades (Fig. 2). For comparison with the published phylogeny of Thacker et al. (2007), we use their letter designations for each clade. The topology of the tree in Figure 2 shows the mtDNA affinities between the allosyme study of Bertozzi et al. (2000) and the present study. Uncorrected average pairwise distance between clades A, D, H and L in Figure 2 ranged from 8.1 ± 0.8 % to 10.7 ± 0.9 %. Analysis of the frequency and distribution
of mtDNA clades based on the sample of 481 individuals assayed using sequencing and/or PCR-RFLP across 14 sites revealed that clade A was very rare, being present in only three individuals (Table 1), all referable to the species *H. klunzingeri*. Most individuals sampled in the Goulburn study area belong to mtDNA clades D and H, which made up 55.3 and 44.1% of the total sample respectively. The mtDNA distance between these clades is 10.1 ± 0.9 % and individuals bearing these divergent mtDNA genomes were found together in all of the 14 sampled pools (Table 1).

Detection of distinct genetic groups using microsatellite data

Seven microsatellite loci were polymorphic in *Hypseleotris* samples from the Goulburn study area. Observed allelic variation indicates the carp gudgeon sample is diploid as the genotype of all individuals contained either one or two alleles per locus. Further analysis is required to confirm ploidy levels in this system (*e.g.* gene dosage, flow cytometry). Since over 99% of our sample possessed a mtDNA genome belonging to either clade D or H, the following analysis is confined to these 478 individuals. A total of 66 alleles were detected, with five to 15 alleles per locus (Table S3. Supporting Information). No fixed differences at microsatellite loci were found between clades D and H, rather 56 alleles were shared between clades. Bayesian admixture clustering clearly indicated that the posterior distribution of allele frequencies was best explained by three clusters (Fig. S1. Supporting information). Assuming a threshold of q-value of >0.85 for assignment, ~90% of the sample clearly belonged to one of the three clusters and ~10% were admixed with an approximately equal contribution of cluster 2 and 3 (Fig. 3a). All individuals assigned to cluster 1
and 2 belong to mtDNA clade \( D \), while all individuals assigned to cluster 3 and the admixed group belong to mtDNA clade \( H \) (Fig. 3a).

Principal component analysis differentiated four groups of genotypes on the first two axes, which accounted for 25.5 and 11.2% of total genetic variation respectively (Fig. 3b). The PCA groupings reflect the differentiation revealed by Bayesian clustering in STRUCTURE. Individuals assigned to cluster 1 by STRUCTURE appear as a group with positive coordinates on PCA axis 1 (Fig. 3b, boxes) whereas individuals assigned to cluster 2 have positive coordinates on axis 2 (Fig. 3b, circles). All of these individuals (clusters 1 and 2) have clade \( D \) mtDNA. Individuals that form a third group with negative coordinates on PCA axis 2 (Fig. 3b, triangles) were all assigned to cluster 3 by STRUCTURE and have clade \( H \) mtDNA. The remaining fourth group in the PCA (Fig. 3b, diamonds) also has clade \( H \) mtDNA and is comprised of individuals that were assigned as an even admixture between clusters 2 and 3 in the STRUCTURE analysis.

**Association of gender with genetic groups**

The sex ratio was approximately even for individuals assigned to Group 1 (82♂ : 92♀ : 5 indet.) and not significantly different from an expected male:female ratio of 1:1 \((\chi^2 = 0.56, P = 0.448)\). The other groups exhibited highly skewed sex ratios. Group 2 was almost exclusively comprised of males (80♂ : 2♀ : 5 indet.; \( \chi^2 = 74.2, P < 0.0001 \)). Group 3 was heavily biased towards females (7♂ : 152♀ : 9 indet.; \( \chi^2 = 132.24, P < 0.0001 \)), and Group 4 showed a significant but less dramatic bias towards females (9♂ : 29♀ : 6 indet.; \( \chi^2 = 10.53, P = 0.0012 \)).
Analysis of HWE was conducted separately on the four groups of individuals identified by Bayesian clustering and PCA. Significant deviation from genotypic proportions expected under HWE was found for all variable loci in groups 2, 3, and 4 (Table 2). Exact tests for heterozygote excess were significant for all variable loci in these three groups and associated inbreeding coefficients ($F_{IS}$) were negative (Table 2). Figure 4 shows that of the seven loci genotyped, most individuals in groups 2, 3 and 4 were heterozygous at six or seven loci. Of the four loci that did deviate significantly from HWE proportions in group 1, each showed a significant heterozygote deficit. However, exact tests for conformity to HWE genotypic proportions within individual sites showed only two significant deviations ($P < 0.05$) among the 55 locus-population combinations that could be tested. As these two cases involved different loci at different sites (hyp013 in HM and hyp005 in HH), no systematic bias was evident in HWE proportions for marker loci in group 1. Individuals in group 1 therefore exhibit genotypic proportions consistent with a randomly mating sexual population, while individuals in groups 2, 3 and 4 have excessively heterozygous multilocus genotypes consistent with hybridization and/or asexuality. An example of allele profiles illustrating the relationships between the four groups is given in supporting information, Table S4.

From a total of 447 individuals with no missing microsatellite data there were 404 unique multilocus genotypes (MLGs) and 28 cases of repeated MLGs. Most repeated MLGs (i.e. 20) reflected only two individuals, and the maximum number of
individuals displaying the same MLG was four. Repeated MLGs were observed in all
of the four genetic groupings identified by clustering analyses (n = 5, 2, 12, and 9 for
groups 1 to 4 respectively). Of the 28 repeated MLGs, 15 cases involved identical
copies located in different sites and eight cases where at least one of the identical
genotypes was expressed by a male.

Comparison with voucher specimens

Voucher specimens representing allozyme groupings of Bertozzi et al. (2000) have
close mtDNA affinities with clades from the present study area (Fig. 2). Of the four
well supported mtDNA clades (A, D, H, L, Fig. 2), three were shared between the two
studies (clades A, D, H). PCA on the microsatellite data incorporating voucher
samples of allozyme groupings from Bertozzi et al. (2000) shows that Group 1
corresponds with allozyme group HA; Group 2 corresponds with allozyme group
HA×HB; Group 3 corresponds with allozyme group HA×HX and Group 4 corresponds
with allozyme group HB×HX (Figure S2. Supporting information).

Discussion

Evidence for hybridogenesis in natural populations of Hypseleotris.

Criteria established for evidence of asexuality in natural populations included the
presence of distinct nuclear genetic groupings that define sets of unisexual and
bisexual individuals; expected high levels of mtDNA divergence between hybrid
groups and at least some sexual groups; and high or fixed levels of heterozygosity in
unisexual hybrid groups. The carp gudgeon fauna of the Goulburn river system is
dominated by fish that can be divided into four distinct groups based on microsatellite
data. An additional species (H. klunzingeri, not found by Bertozzi et al. 2000 to be
represented in any putative hybrid forms) is also present at very low frequency (<1%).
Two different clustering approaches agreed on the divisions between the four
dominant groups. The genetic results and sex ratio of group 1 (=HA sensu Bertozzi et
al. 2000) indicate this is a randomly mating sexual population. Five cases of identical
MLG’s were found among the 179 individuals belonging to group 1. This appears
unlikely in a sexually reproducing population given the probability of two unrelated
individuals sharing the same multilocus genotype by chance is 0.0002. However, it is
possible that some siblings may be present in the sample and/or that inbreeding within
some pools may explain this anomaly. These individuals are not clones because all
five are discriminated by an additional locus not included in the present study
(hyp025, Schmidt unpub. data).

In contrast, group 2 exhibits excessive heterozygosity and is ~98% male. We propose
that group 2 represents a male hybridogenetic biotype. Individuals in group 2 share at
least one nuclear allele per locus with group 1 in addition to sharing the clade D
mtDNA genome. We predict that this pattern is a consequence of mating between
hemiclonal group 2 males and females of the sexual species group 1. A
hybridogenetic biotype must live sympatrically with a sexual host in order to
reproduce and we found that group 2 individuals co-occurred with their putative
sexual host (group 1), in 9 out of eleven waterholes (Fig 1). Two sites where group 2
was found without group 1 had relatively low sample sizes (SMV, n = 14; SMS n =
7), and thus the absence of group 1 in these cases is consistent with sampling error.
Comparing our data with previous allozyme groupings indicates that our group 2 corresponds to the hybrid class $HA \times HB^*$ of Bertozzi *et al.* (2000). We use an asterisk hereafter to denote the putative clonal genome.

Group 3 carries the divergent clade $H$ mtDNA genome, exhibits fixed heterozygosity at most nuclear loci and is ~96% female. We propose that group 3 represents a female hybridogenetic biotype. All members of group 3 share at least one nuclear allele per locus with group 1 ($= HA$), but belong to a distinct mtDNA clade that is approximately 10% divergent from group 1. We suggest this pattern may be the consequence of mating between hemiclonal group 3 females and males of the sexual species, group 1. We found that group 3 individuals co-occurred with their putative sexual host (group 1) in twelve out of thirteen waterholes (Fig. 1) and cannot rule out low sample size as an explanation for the exception at one site (SMV $n = 14$).

Comparison with allozyme groupings shows that our group 3 corresponds to the hybrid class $HA \times HX^*$ of Bertozzi *et al.* (2000). Until controlled crosses are performed it is difficult to rule out the possibility that gynogenesis might account for the pattern of genetic variation observed in group 3. However gynogenesis appears unlikely because relatively few repeated copies (12) of multilocus genotypes were observed in the 168 individuals belonging to this group. Hybridogenesis is more likely because all of the individuals in group 3 share at least one allele per locus with the putative sexual group and a small proportion of males were found.

Group 4 carries the clade $H$ mtDNA genome in common with group 3 and has excess heterozygosity at most nuclear loci. Individuals of this group were found at a lower frequency than the other groups and the sex ratio (~76% female) was less biased
compared to group 3. Multivariate analysis showed that group 4 is positioned at a
greater distance from the putative sexual species (group1, HA) than either group 2 or
3. Interestingly, the clustering approach implemented using STRUCTURE (Pritchard
et al. 2000) showed that group 4 individuals comprise an even admixture between
group 2 and 3 (Fig. 3a). Accordingly, we propose that group 4 individuals represent
the offspring of a natural “double hybrid cross” between females belonging to group 3
(= HA×HX*) and males of group 2 (= HB*×HA). Under this scenario, the pattern
observed in the PCA is the consequence of each parent contributing its clonally
inherited genome, so that group 4 individuals have a HB*×HX* genetic make-up and
are therefore positioned at a greater distance from the sexual group (group1, HA) in
ordination space. Comparing our data with previous allozyme groupings indicates that
our group 4 does indeed correspond with the hybrid class HB*×HX* identified by
Bertozzi et al. (2000). We infer that the maternal and paternal genomes inherited by
these individuals in our study area are both clonal, as indicated by asterisks. A
summary of interactions inferred between the four groups within the Goulburn study
area is illustrated in Figure 5. All relationships depicted in this model are testable by
controlled breeding experiments. Linkages involving group 4 are left unresolved
because the fertility of this group is unknown (Fig. 5).

Our data suggest that hybrid carp gudgeons in the Goulburn study area are the product
of hybridogenetic reproduction and not due to primary hybridization between sexual
parental taxa. This is because extensive sampling of most refugial pools in the system
did not reveal the presence of any individuals corresponding to a pure HB or a pure
HX group.
Gender association of putative hybridogenetic lineages

We have identified two genetic groups likely to represent hybridogenetic lineages in the Goulburn study area – group 2 (\(=HB^* \times HA\)) is predominantly male, and group 3 (\(=HA \times HX^*\)) is predominantly female. It is possible that the observed gender associations reflect a system of hemiclonal inheritance similar to that seen in the water frog, *Pelophylax esculenta*, where sex determination is an XX-XY system and the hemiclonal genome may be coupled with either X or Y haploid set of chromosomes (Graf & Polls Pelaz 1989). If this were the case in *Hypseleotris*, then our data suggest that the \(HB^*\) haploid genome includes a Y chromosome, and the \(HX^*\) haploid genome similarly includes an X chromosome. A speculative explanation for the presence of a small proportion of females in group 2 and males in group 3 is that the cellular mechanism excluding the non-hemiclonal genome from the germ-line may not be entirely effective. However sex determination may be more complex than male or female heterogamety in unisexual systems (Pala et al. 2009; Lamatsch et al. 2010) and further work is clearly required to explore the interaction between gender and hemiclonal inheritance in *Hypseleotris*.

Our suggestion that group 4 represents the offspring of mating between two independent hybridogenetic lineages (i.e. group 2 and group 3) has not previously been reported in a unisexual fish system. The consequences of genetic interaction between hemiclonal lineages were explored using experimental crosses in the *Poeciliopsis* system (Leslie & Vrijenhoek 1980). In both *Poeciliopsis* and the *P. esculenta* water frog system, matings that unite the same hemiclonal lineage usually
do not produce viable offspring due to the accumulation of deleterious mutations which are exposed in homozygotes (Leslie & Vrijenhoek 1980; Vorburger 2001). However mating between independent hemiclones with different mutations can produce viable offspring (Guex et al. 2002). It is possible that group 4 may constitute an equivalent group of viable “double hybrid” offspring in the carp gudgeon system. If it is assumed that the parental lineages of group 4 are both hemiclonal and both exclude the sexually derived HA genome from their gametes, then group 4 should be all-male. This follows from reasoning (given above) that clonal genome HB* includes a Y chromosome and clonal genome HX* includes an X chromosome. The observed female bias of ~75% in this group is therefore unexpected and again highlights the need for an understanding of sex determination in the carp gudgeon system. We do not yet know whether adult individuals in group 4 are fertile. If they are, then the existence of both sexes in this group may provide these hybridogenetic lineages with the ability to exist independently from a sexual host. The ability to exist independently from parental forms was achieved by polyploidization in the P. esculenta system, where triploids have taken over the role of sexual hosts in some areas (Graf & Polls Pelaz 1989; Arioli et al. 2010). Based on available microsatellite data it appears that the Hypseleotris system reported here is primarily diploid and two putative hybridogenetic lineages broadly co-exist with a single sexual host throughout our study area. Given the two-fold cost of sexual reproduction it is likely that this relationship places significant demographic pressure on the sexual species in this system. Population models show that extinction may be inevitable unless mate discrimination or metapopulation processes are able to counteract the effects of gamete limitation that occur in sexual/asexual systems (Kokko et al. 2008; Heubel et al. 2009). Since carp gudgeons are abundant (Perry & Bond 2009) it is reasonable to
expect that these processes or other counteracting forces are involved in maintaining
carp gudgeon populations. A novel mechanism that may promote coexistence of the
Hypseleotris sexual/sexual system is the potential ability of separate male and female
hybridogenetic lineages to hybridise with each other. If this “double hybrid” group is
fertile it may represent an important contributor to the dynamics of coexistence in this
new hybridogenetic fish complex. Bisexuality within the asexual group may lower the
two-fold demographic advantage normally held by asexuals in other all-female modes
of reproduction.

Acknowledgements

Funding for this work was provided by eWater CRC. All procedures were carried out
in accordance with Australian ethics commission protocol number (Monash AEC
2006-15) and Fisheries permit number RP 914. Assistance in the field was provided
by David Reid, Zoe Squires, Tom Daniel, Matthew Johnson and Damien McMaster.
We thank Stephen Balcombe for expert guidance in the morphological diagnosis of
sex in gudgeons. Kathryn Real performed all DNA extractions. Peter Unmack
provided information to cross-reference Genbank sequences against the published
Hypseleotris tree. Joel Huey, Hanna Kokko, Robert Vrijenhoek and three anonymous
reviewers gave useful advice during the course of this study.

References

australian scincid lizard Menetia greyii: A new instance of widespread
vertebrate parthenogenesis. Evolution 57, 2619-2627.
Allen GR, Midgley SH, Allen M (2002) Field guide to the freshwater fishes of
Australia Western Australian Museum, Perth.


Data Accessibility:

mtDNA sequences: Supporting Information Table S2, Genbank accessions: JF800016 – JF800031.

Microsatellite clone sequences: Supporting Information Table S3, Genbank accessions: JF800032 – JF800038.
Microsatellite genotype data: DRYAD entry doi:10.5061/dryad.3117v
**Figure Captions**

**Figure 1.** Sampling sites for *Hypseleotris* spp. along tributaries of the Goulburn River and composition of each refugial waterhole according to genetic groups defined by clustering analyses of microsatellite data (see Figures 3a,b).

**Figure 2.** Bayesian phylogram based on 1140 bp of mtDNA cytochrome *b* sequence data. Posterior probability of clades calculated from 30,000 post burn-in trees. Clade labels (A, D, H, L) correspond to clade designations in Thacker *et al.* (2007). Data for each terminal include sample code; source of sample (Genbank sequences from Thacker *et al.* 2007, voucher tissue corresponding to Bertozzi *et al.* 2000, this study); a Genbank accession is given for sequences derived from the Thacker study and the taxon nomenclature is given for voucher samples genotyped using the allozyme diagnoses presented in the Bertozzi study.

**Figure 3.** Clustering analyses performed on multilocus genotypes of 478 *Hypseleotris* individuals based on seven microsatellite loci. (a) Bar plot from Bayesian clustering analysis in STRUCTURE showing the admixture proportions of each individual in the three genetic clusters. Four genetic groups of *Hypseleotris* are defined. Group 1 is assigned to cluster 1 (white bars); Group 2 is assigned to cluster 2 (grey bars); Group 3 is assigned to cluster 3 (black bars); Group 4 is composed of an even admixture between clusters 2 and 3. (b) Principle component analysis (PCA), each point represents an individual genotype with symbols corresponding to groups defined by the Bayesian clustering analysis. Group 1, squares; Group 2, circles; Group 3, triangles; Group 4, diamonds. The inertia ellipses indicate dispersion of individuals relative to the mean coordinate of each group. Inset shows the screeplot of
eigenvalues for each principle component and highlights values of the first two components displayed on the principle axes of the plot.

**Figure 4.** Histograms illustrating the observed number of heterozygous genotypes per individual for 7 microsatellite loci. Calculations based on individuals with no missing data. (a) Group 1, n = 163; (b) Group 2, n = 83; (c) Group 3, n = 155; (d) Group 4, n = 41.

**Figure 5.** Diagram illustrating the inferred relationships between four genetic groups of *Hypseleotris* spp. in the Goulburn study area. The nomenclature of Bertozzi *et al.* (2000) is used to denote the three genetic clusters (*HA, HB, HX*) that comprise these four genetic groups. Groups 2, 3, and 4 are hybrid classes as described in the text. Connecting arrows indicate the range of offspring produced by mating between the groups.

**Table headings**

**Table 1.** Information on sampling locations for 14 populations of *Hypseleotris* spp. from tributaries of the Goulburn River, and summary of the PCR-RFLP assay for assignment of individuals to mtDNA clades.

**Table 2.** Heterozygosity estimates and inbreeding coefficients (*F*_IS) for the four genetic groupings of *Hypseleotris* spp. defined by clustering analyses. *H*O, observed heterozygosity; *H*_E expected heterozygosity (*P < 0.05; P < 0.01; P < 0.001); nc, not calculated.
Table 1. Information on sampling locations for 14 populations of *Hypseleotris* spp. from tributaries of the Goulburn River, and summary of the RFLP-PCR assay for assignment of individuals to mtDNA clades.

<table>
<thead>
<tr>
<th>Site Code</th>
<th>Stream</th>
<th>Latitude, Longitude</th>
<th>N</th>
<th>mtDNA clade determined by RFLP-PCR*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>clade D</td>
</tr>
<tr>
<td>HM</td>
<td>Honeysuckle Creek</td>
<td>145.5538, -36.5446</td>
<td>57</td>
<td>31 (10)</td>
</tr>
<tr>
<td>HS</td>
<td>Honeysuckle Creek</td>
<td>145.6051, -36.5792</td>
<td>46</td>
<td>17 (9)</td>
</tr>
<tr>
<td>HSV</td>
<td>Honeysuckle Creek</td>
<td>145.6261, -36.5910</td>
<td>35</td>
<td>16 (8)</td>
</tr>
<tr>
<td>HH</td>
<td>Honeysuckle Creek</td>
<td>145.7438, -36.6553</td>
<td>38</td>
<td>32 (17)</td>
</tr>
<tr>
<td>FDN</td>
<td>Faithful Creek</td>
<td>145.4799, -36.6222</td>
<td>5</td>
<td>1 (1)</td>
</tr>
<tr>
<td>SMV</td>
<td>Sevens Creek</td>
<td>145.4788, -36.6069</td>
<td>15</td>
<td>6 (5)</td>
</tr>
<tr>
<td>SM</td>
<td>Sevens Creek</td>
<td>145.5671, -36.7294</td>
<td>19</td>
<td>4 (1)</td>
</tr>
<tr>
<td>SMS</td>
<td>Sevens Creek</td>
<td>145.4790, -36.6329</td>
<td>7</td>
<td>4 (4)</td>
</tr>
<tr>
<td>CA</td>
<td>Castle Creek</td>
<td>145.3441, -36.5560</td>
<td>46</td>
<td>36 (8)</td>
</tr>
<tr>
<td>CP</td>
<td>Castle Creek</td>
<td>145.5044, -36.7170</td>
<td>47</td>
<td>6 (1)</td>
</tr>
<tr>
<td>CE</td>
<td>Castle Creek</td>
<td>145.5750, -36.7904</td>
<td>56</td>
<td>35 (16)</td>
</tr>
<tr>
<td>CG</td>
<td>Castle Creek</td>
<td>145.5829, -36.8648</td>
<td>56</td>
<td>38 (15)</td>
</tr>
<tr>
<td>CK</td>
<td>Castle Creek</td>
<td>145.5853, -36.9025</td>
<td>13</td>
<td>11 (10)</td>
</tr>
<tr>
<td>PL</td>
<td>Pranjip Creek</td>
<td>145.4542, -36.7993</td>
<td>41</td>
<td>29 (12)</td>
</tr>
</tbody>
</table>

* value in parentheses indicates number sequenced.
Table 2. Heterozygosity estimates and inbreeding coefficients \( (F_{IS}) \) for the four genetic groupings of *Hypseleotris* spp. defined by clustering analyses. \( H_O \), observed heterozygosity; \( H_E \) expected heterozygosity (*P < 0.05; P < 0.01; P < 0.001); nc, not calculated.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Locus</th>
<th>( H_O )</th>
<th>( H_E )</th>
<th>( F_{IS} ) (P-value for heterozygote excess test)</th>
<th>No. alleles shared with Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>179</td>
<td>hyp001</td>
<td>0.452***</td>
<td>0.618</td>
<td>0.427***</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp005</td>
<td>0.619**</td>
<td>0.744</td>
<td>0.171</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp008</td>
<td>0.471</td>
<td>0.445</td>
<td>-0.055</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp009</td>
<td>0</td>
<td>0</td>
<td>nc</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp010</td>
<td>0.439*</td>
<td>0.524</td>
<td>0.167</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp013</td>
<td>0.439***</td>
<td>0.533</td>
<td>0.180</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp021</td>
<td>0.432</td>
<td>0.462</td>
<td>0.068</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>82</td>
<td>hyp001</td>
<td>1.000***</td>
<td>0.774</td>
<td>-0.286***</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp005</td>
<td>0.988***</td>
<td>0.806</td>
<td>-0.220***</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp008</td>
<td>0.963***</td>
<td>0.782</td>
<td>-0.225***</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp009</td>
<td>1.000***</td>
<td>0.576</td>
<td>-0.733***</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp010</td>
<td>0.400***</td>
<td>0.584</td>
<td>0.321</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp013</td>
<td>1.000***</td>
<td>0.663</td>
<td>-0.503***</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp021</td>
<td>0.988***</td>
<td>0.597</td>
<td>-0.650***</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>168</td>
<td>hyp001</td>
<td>0.970***</td>
<td>0.774</td>
<td>-0.251***</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp005</td>
<td>1.000***</td>
<td>0.690</td>
<td>-0.447***</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp008</td>
<td>0.985***</td>
<td>0.695</td>
<td>-0.414***</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp009</td>
<td>1.000***</td>
<td>0.550</td>
<td>-0.815***</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp010</td>
<td>1.000***</td>
<td>0.638</td>
<td>-0.565***</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp013</td>
<td>1.000***</td>
<td>0.642</td>
<td>-0.554***</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp021</td>
<td>0.731***</td>
<td>0.660</td>
<td>-0.105***</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
<td>hyp001</td>
<td>1.000***</td>
<td>0.776</td>
<td>-0.271**</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp005</td>
<td>1.000***</td>
<td>0.641</td>
<td>-0.546***</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp008</td>
<td>0.963***</td>
<td>0.769</td>
<td>-0.235**</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp009</td>
<td>1.000***</td>
<td>0.650</td>
<td>-0.526**</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp010</td>
<td>1.000***</td>
<td>0.611</td>
<td>-0.625**</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp013</td>
<td>0</td>
<td>0</td>
<td>nc</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hyp021</td>
<td>1.000***</td>
<td>0.658</td>
<td>-0.506**</td>
<td>4</td>
</tr>
</tbody>
</table>
(a) Group 1 | Group 2 | Group 3 | Grp 4
mtDNA: clade D | mtDNA: clade H

(b) Group 1 | Group 2 | Group 3 | Group 4

408x420mm (300 x 300 DPI)
Observed number of heterozygous genotypes per individual