

Introgression between ecologically distinct species following increased salinity in the Colorado Delta- Worldwide implications for impacted estuary diversity (#20617)

1

First submission

Please read the **Important notes** below, the **Review guidance** on page 2 and our **Standout reviewing tips** on page 3. When ready [submit online](#). The manuscript starts on page 4.

Important notes

Editor and deadline

Therese Markow / 5 Oct 2017

Files

7 Figure file(s)

8 Table file(s)

6 Raw data file(s)

1 Other file(s)

Please visit the overview page to [download and review](#) the files not included in this review PDF.

Declarations

**One or more DNA sequences were reported.
Involves a field study on animals or plants.**



Please read in full before you begin

How to review

When ready [submit your review online](#). The review form is divided into 5 sections. Please consider these when composing your review:

- 1. BASIC REPORTING**
- 2. EXPERIMENTAL DESIGN**
- 3. VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

 You can also annotate this PDF and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [PeerJ standards](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (see [PeerJ policy](#)).

EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

-  Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
-  Data is robust, statistically sound, & controlled.
-  Conclusions are well stated, linked to original research question & limited to supporting results.
-  Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit <https://peerj.com/about/editorial-criteria/>

7 Standout reviewing tips

3



The best reviewers use these techniques

Tip

Example

Support criticisms with evidence from the text or from other sources

Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.

Give specific suggestions on how to improve the manuscript

Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).

Comment on language and grammar issues

The English language should be improved to ensure that your international audience can clearly understand your text. I suggest that you have a native English speaking colleague review your manuscript. Some examples where the language could be improved include lines 23, 77, 121, 128 - the current phrasing makes comprehension difficult.

Organize by importance of the issues, and number your points

1. Your most important issue
2. The next most important item
3. ...
4. The least important points

Give specific suggestions on how to improve the manuscript

Line 56: Note that experimental data on sprawling animals needs to be updated. Line 66: Please consider exchanging "modern" with "cursorial".

Please provide constructive criticism, and avoid personal opinions

I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC

Comment on strengths (as well as weaknesses) of the manuscript

I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.

Introgression between ecologically distinct species following increased salinity in the Colorado Delta- Worldwide implications for impacted estuary diversity

Clive L.F. Lau¹, David K. Jacobs^{Corresp. 1}

¹ Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, California, United States

Corresponding Author: David K. Jacobs
Email address: djacobs@ucla.edu

We investigate hybridization and introgression between ecologically distinct sister species of silverside fish in the Gulf of California through combined analysis of morphological, sequence, and genotypic data. Water diversions in the past century turned the Colorado River Delta from a normal estuary to a hypersaline inverse estuary, raising concerns for the local fauna, much of which is endangered. Salinity differences are known to generate ecological species pairs and we anticipated that loss of the fresher-water historic salinity regime could alter the adaptive factors maintaining distinction between the broadly distributed Gulf-endemic *Colpichthys regis* and the narrowly restricted Delta-endemic *Colpichthys hubbsi*, the species that experienced dramatic environmental change. In this altered environmental context, these long-isolated species (as revealed by Cytochrome *b* sequences) show genotypic (RAG1, microsatellites) evidence of active hybridization where the species ranges abut, as well as directional introgression from *C. regis* into the range center of *C. hubbsi*. Bayesian group assignment (STRUCTURE) on six microsatellite loci and multivariate analyses (DAPC) on both microsatellites and phenotypic data further support substantial recent admixture between the sister species. Although we find no evidence for recent population decline in *C. hubbsi* based on mitochondrial sequence, introgression may be placing an ancient ecological species at risk of extinction. Such introgressive extinction risk should also pertain to a number of other ecological species historically sustained by the now changing Delta environment. More broadly, salinity gradient associated ecological speciation is evident in silverside species pairs in many estuarine systems around the world. Ecological species pairs among other taxa in such systems are likely poorly understood or cryptic. As water extraction accelerates in river systems worldwide, salinity gradients will necessarily be altered, impacting many more estuary and delta systems. Such alteration of habitats will place biodiversity at risk not only from direct effects of habitat destruction, but also from the potential for the breakdown of ecological species. Thus, evolutionary response to the anthropogenic alteration of salinity gradients in

estuaries merits investigation as the number of impacted systems increases around the globe, permitting parallel study of multiple systems, while also permitting a conservation management response to help preserve this little championed component of biodiversity.

Introgression between ecologically distinct species following increased salinity in the Colorado Delta- Worldwide implications for impacted estuary diversity

Clive L. F. Lau,¹ and David K. Jacobs²

^{1,2} Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, California, USA

Corresponding Author:

David K. Jacobs²

Email address:

djacobs@ucla.edu

ABSTRACT

We investigate hybridization and introgression between ecologically distinct sister species of silverside fish in the Gulf of California through combined analysis of morphological, sequence, and genotypic data. Water diversions in the past century turned the Colorado River Delta from a normal estuary to a hypersaline inverse estuary, raising concerns for the local fauna, much of which is endangered. Salinity differences are known to generate ecological species pairs and we anticipated that loss of the fresher-water historic salinity regime could alter the adaptive factors maintaining distinction between the broadly distributed Gulf-endemic *Colpichthys regis* and the narrowly restricted Delta-endemic *Colpichthis hubbsi*, the species that experienced dramatic environmental change. In this altered environmental context, these long-isolated species (as revealed by Cytochrome *b* sequences) show genotypic (RAG1, microsatellites) evidence of active hybridization where the species ranges abut, as well as directional introgression from *C. regis* into the range center of *C. hubbsi*. Bayesian group assignment (STRUCTURE) on six microsatellite loci and multivariate analyses (DAPC) on both microsatellites and phenotypic data further support substantial recent admixture between the sister species. Although we find no evidence for recent population decline in *C. hubbsi* based on mitochondrial sequence, introgression may be placing an ancient ecological species at risk of extinction. Such introgressive extinction risk should also pertain to a number of other ecological species historically sustained by the now changing Delta environment. More broadly, salinity gradient associated ecological speciation is evident in silverside species pairs in many estuarine systems around the world. Ecological species pairs among other taxa in such systems are likely poorly understood or cryptic. As water extraction accelerates in river systems worldwide, salinity gradients will necessarily be altered, impacting many more estuary and delta systems. Such

24 alteration of habitats will place biodiversity at risk not only from direct effects of habitat
 25 destruction, but also from the potential for the breakdown of ecological species. Thus,
 26 evolutionary response to the anthropogenic alteration of salinity gradients in estuaries merits
 27 investigation as the number of impacted systems increases around the globe, permitting parallel
 28 study of multiple systems, while also permitting a conservation management response to help
 29 preserve this little championed component of biodiversity.

INTRODUCTION

Damming and water extraction in river systems are accelerating around the globe (Nilsson et al., 2005; Liermann et al., 2012), impacting estuarine and deltaic settings downstream (Bonetto et al., 1989; Lu and Siew, 2006; Yang et al., 2005). Physical changes include loss of freshwater, elimination of peak flood flows, loss of sediment supply, and alteration of the nature and position of salinity gradients (Nilsson and Berggren, 2000, Carriquiry and Sánchez, 1999). Such changes necessarily influence the adaptive context under which local species evolve, potentially reducing the selective pressures maintaining species differences and placing ecological species at risk of elimination through introgression. In order to better understand the risks such changes pose, we examine recent introgression in a relatively ancient species pair of silverside fishes of the genus *Colpichthys* in the northern Gulf of California (Gulf hereafter) where the Colorado River enters the Ocean. Following 20th century dam constructions, virtually no river flow reaches the Colorado River Delta (Glenn et al., 1996). As a consequence of this dramatically modified state, this system may serve as a harbinger of evolutionary impacts in the estuaries of the world's major river systems.

Colpichthys, a little studied genus of the silverside family Atherinopsidae, inhabits near-shore settings of the northern part of the Gulf and consists of two endemic species, *Colpichthys regis* (Jenkins and Evermann, 1889) and *Colpichthys hubbsi* (Crabtree, 1989). These are listed on the IUCN Red List as Near Threatened and Endangered, respectively (Findley et al., 2010a; Findley et al., 2010b). The Colorado River enters the Gulf at its northernmost extremity where it forms an extensive delta and estuary system. The Colorado Delta (Delta hereafter) habitat of *C. hubbsi* was historically distinct from the other marine estuaries of the semiarid Gulf where we

recovered *C. regis* samples. Tidal channels in the Delta experience a much greater tidal amplitude of up to 10 meters, and silt as opposed to clay content is much higher in the Delta relative to other estuaries. In the 19th century, the Delta received continuous freshwater flow and extensive flood water during the spring and summer melting of mountain snows in the headwaters. These flows were eliminated through 20th century damming leading to a shift from a typical brackish to a hypersaline inverse estuary (Lavín and Sánchez, 1999). The difference in salinity, sediment supply and food chain necessarily impose significantly different adaptive regimes on taxa in the Delta relative to other Gulf estuaries (Swift et al., 2011); as a result of the shifting environment, these differences may disappear.

Through the study of introgressive process in *Colpichthys* we hope to inform issues of evolution and endangerment of the northern Gulf fauna associated with the potential for introgression in ecological species in the face of habitat change. Direct and climate driven anthropogenic impacts increasingly threaten estuarine faunas globally across a broader suite of river systems. In the remainder of this introduction we consider: similar contexts that have produced ecological species, the potential extirpation of species via introgression, the relevant detail regarding the study taxon *Colpichthys* as well as the suite of data and methods employed in phylogeny reconstruction, and assessment of historic demography and introgression.

Marine taxa that invade fresh water experience strong selection (Lee, 2015). In these cases, advantages in food availability or predator avoidance can overcome the selective costs and rapid evolution of tolerance to fresh water often ensues. Ecological divergence between fresh and marine water commonly occurs in the lower reaches of rivers, and is especially well established in silversides. Ecological diversification and speciation by salinity differences occur in

geographically isolated and taxonomically separate silverside genera from Brazil, Australia, the Mediterranean, and the Eastern US (Bamber and Henderson, 1988; Beheregaray and Sunnucks, 2001; Fluker et al., 2011; Johnson et al., 1994; Klossa-Kilia et al., 2007; Francisco et al., 2006; Olsen et al., 2016; Trabelsi et al., 2002). The strong selection pressures between fresh and salt water, the frequent ecological diversification in silversides, and the historic difference in habitats between the Delta and the Gulf as noted above all support an ecological species interpretation of the *C. regis*/*C. hubbsi* split. As the loss of fresh water input to the Delta disrupts the local salinity gradient, breakdown of the ecological factors supporting distinction of these species seems likely.

Introgressive hybridization involves the transfer of alleles from one species to the gene pool of another species through repeated backcrossing of hybrids (Heiser, 1973; Harrison and Larson, 2014). This process can result from the dissolution of selective regimes that maintain ecological species and has a number of evolutionary consequences. Introgression could allow for the spread of adaptive or deleterious alleles from one species to the other across a hybrid zone. Alternatively, it may promote the homogenization of the two species, and in extreme cases, could lead to the extinction of one or both parental types through effects such as demographic or genetic swamping (Todesco et al., 2016). The hydrologic changes in the Delta make *C. hubbsi* a likely candidate for introgressive hybridization with *C. regis*. *C. regis* has a range rendered disjunct by the Colorado River Delta at the northern end of the Gulf where *C. hubbsi* is locally endemic. *C. regis* and *C. hubbsi* co-occur only at the southwestern border of the Delta where it meets the Baja coast (Fig. 1A). In this study, we test whether hybridization occurs in this region of overlap and whether introgression is occurring between the species.

In the following, we compare the external and internal morphology of specimens recovered from *C. regis* and *C. hubbsi* populations to characterize the variation across both species to help establish baseline for discovering potential hybrids. We also investigate molecular evidence from mitochondrial and nuclear sequences as well as microsatellite genotypes to detect genetic introgression across the species barrier. In addition to investigating the hybridization, with the available genetic data we performed a number of population genetics analyses to infer genetic diversity and demographic histories of the two species.

MATERIALS AND METHODS

Sample Collection

Specimens of the genus *Colpichthys* were collected by seine from estuarine habitats along the coast of the Northern Gulf of California and within the Colorado River Delta in the years 2005, 2006, 2007, and 2011 (Table 1). Collections were carried under Mexican Federal Collecting Permit (Permiso de Pesca de Fomento) DGOPA 14253. 101005.6950, and its extension DGOPA 06435.210606.2640, issued to Findley and Jacobs by the Comisión Nacional de Acuacultura y Pesca of the Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA). Collected specimens were preserved in 95% ethanol in the field and stored at -20°C upon returning to the laboratory for use in both molecular and morphological analyses. For this study, we divided the coastline of the Northern Gulf of California into 4 regions: the Baja Coast, the Delta Edge, the Colorado River Delta, and the Sonora Coast. A full list of collection sites along with the abbreviated designation for each site are shown in Table 1. In addition to the field-collected specimens, formalin-fixed specimens from the Natural History Museum of Los Angeles County (LACM) and Scripps ~~Institute~~ of Oceanography were examined

for their morphology to confirm species identification. A list of these “historical” specimens is included in the supplementary materials (Table 2).

Phenotypic Analyses

The dorsal-lateral scale counts, vertebral counts, hemal spine morphology, and the position of the first dorsal fin were used as species-distinguishing features for this study; these features were chosen based on the species description of *Colpichthys hubbsi* (Crabtree, 1989). The first dorsal fin position is assessed based on the linear length from the snout to the first dorsal fin origin, normalized by the total length. Length measurements were taken with dial calipers. Morphometric and meristic counts were performed according to Hubbs et al. (2004). Radiographs of specimens were taken for examination of their vertebral counts and hemal spine morphology. Six specimens from location NWC and 3 specimens from location YAV were too small to accurately assess their hemal spine morphology; they were thus excluded from the hemal spine data set.

Molecular Procedures

Genomic DNA was isolated from caudal peduncle muscle tissues using the DNeasy Blood and Tissue Kit (Qiagen, Inc., Valencia, CA). The mitochondrial cytochrome *b* gene was amplified by PCR using primers AJG15 and H5 (Akihito et al., 2000) under the following cycling conditions: 95°C for 2 min; 40 cycles of 95°C for 30 sec, 51°C for 30 sec, followed by 72°C for 90 sec; then finally 72°C for 10 min. PCR products were size-checked on 1.5% agarose gel and cleaned with ExoSap before Sanger sequencing with BigDye Terminator v3.1. Samples

were then submitted to DNA Analysis Facility on Science Hill at Yale University for capillary electrophoresis.

Amplification of the nuclear RAG1 gene was carried out using primers RAG1-2533F and RAG1-4090R (López et al., 2004) under the following cycling conditions: 95°C for 2 min; 40 cycles of 95°C for 1 min, 51°C for 1 min with an incremental increase of 0.5°C per cycle until 55°C, followed by 72°C for 90 sec; then finally 72°C for 10 min. Post-PCR processing and sequencing of the RAG1 gene were performed as described above. From the sequences, we were able to identify four polymorphic sites; several individuals display ambiguity in base calling at such sites, and these were tagged as putative heterozygotes. A total of four RAG1 variants were inferred from the sequence data. Homozygous individuals were used to confirm the existence of three RAG1 variants; the remaining RAG1 variant was inferred from a heterozygous individual. To confirm the gametic phase of these RAG1 sequences, we employed allele-specific polymerase chain reaction (ASPCR) as detailed in Wu et al. (1989). Three allele-specific primers (T717F: 5'-CTACAAAATCTTCCAGGAT-3', T770R: 5'-TTTATCTAAGGCTGCCCTCCAGA-3', and A909F: 5'-AACTGGTGCCCTCAGAAGAA-3') were designed to uniquely amplify one of the two diploid copies of RAG1 in the putative heterozygotes. The 3' terminating nucleotide of these primers are designed to specifically anneal at the predetermined polymorphic site, thus enabling unique amplification of alleles and determination of the RAG1 genotypes of the heterozygotes.

Six microsatellite loci (Odont08, Odont09, Odont11, B18, B19, B39) were amplified for genotyping. Primer sequences for loci Odont08, Odont09, and Odont11 were obtained from Beheregaray and Sunnucks (2000). Primer sequences for loci B18, B19, B39 were obtained from Byrne and Avise (2009). Loci which can be easily differentiated by lengths were multiplexed

two at a time using QIAGEN Multiplex PCR kits with 6-FAM labeled M13 primer (Boutin-Ganache et al. 2001). Only the forward primers of each locus are 5' tagged with the M13 sequence, and reactions were performed under the following cycling conditions: 95°C for 15 min; 25 cycles of 94°C for 30 sec, 55°C for 90 sec, 72°C for 1 min; then another 25 cycles of 94°C for 30 sec, 50°C for 90 sec, 72°C for 1 min; and finally 60°C for 30 min. Diluted PCR products were submitted to the UCLA GenoSeq core facility for genotyping on an ABI3730 (Applied Biosystems).

Molecular Analyses

All sequences were aligned, trimmed, and analyzed using Geneious Pro ver. 5.5.6 (Drummond et al., 2011). A Maximum Likelihood tree of Cyt *b* sequence was made with MEGA7 (Kumar et al., 2016) and rooted using *Atherinops affinis* as the outgroup. A time-calibrated tree was inferred using the Reltime method (Tamura et al. 2012) and the Tamura-Nei model (Tamura and Nei, 1993). The timetree was computed using one calibration constraint at the node of the split between *C. regis* and *C. hubbsi*; the calibration point of 4.83Ma was obtained from the age of the ash bed within the Bouse Formation, the rock unit from which two fossilized *C. regis* were documented (Todd, 1976; Spencer et al., 2013).

Microsatellite data were obtained from 101 out of 103 specimens and imported into Geneious for allele scoring and analyzed using STRUCTURE ver. 2.3.4 (Pritchard et al. 2000). STRUCTURE is a program that uses a Bayesian clustering approach to infer population structure from genotypic data. Under an admixture model of K populations or genetic groupings, STRUCTURE computes the q-value, a quantity between 0 and 1, that reflects the proportions of an individual's genome originating from a certain population. Hybridization can then be inferred

based on the computed q-value; for example, first-generation (F1) hybrids of two populations (K=2) are expected to have a q-value close to 0.5 (Vähä and Primmer, 2006). We ran STRUCTURE analyses using an admixture model with 1,000,000 MCMC generations after a burn-in period of 100,000. Sampling locations were incorporated as prior parameters in the *LOCPRIOR* model to improve clustering for our data set; the *LOCPRIOR* model is preferred when there are weak signals of structure due to low number of markers analyzed (Hubisz et al., 2009). Both correlated and independent allele frequency models were used with number of groups set from K=1 to K=4; for each K value, analysis was repeated ten times. Of the two allele frequency models, the correlated model provides greater power in differentiating closely related populations (Falush et al., 2003). The optimal K value was informed by ΔK (Evanno et al., 2005) as computed using Structure Harvester (Earl et al., 2012). Output of the STRUCTURE analyses were summarized in CLUMPP ver. 1.1.2 (Jakobsson and Rosenberg, 2007) and visualized in DISTRICT ver. 1.1 (Rosenberg, 2004).

Multivariate Analysis

We further analyzed our phenotypic and microsatellite data by discriminant analysis of principal components (DAPC) using the R-package *adeigenet* (Jombart et al., 2011). This analysis is used to help describe the phenotypic and genetic clustering of our specimens. Cluster priors for DAPC were identified by *k*-means clustering; based on BIC scores we selected the 2-cluster model to use for the DAPC. To perform DAPC, the program transforms the data into principal components, which is then analyzed through discriminant analysis. Repeated cross-validation was used to evaluate the optimal number of principal components to retain for the discriminant analyses to avoid overfitting of data. For the phenotypic analysis, we used measurements of the 4 species-distinguishing features as our data and retained one PC (91.28% of the cumulative variance) for generating the morphology discriminant function. For the microsatellite analysis, we used the allele frequencies of the 6 loci as our data and retained 20 PCs (66.35% of the cumulative variance) for generating the microsatellite discriminant function.

Demographic Histories

To explore the demographic histories of *Colpichthys*, Cyt *b* data sets were separated by species according to the Cyt *b* gene tree (Fig. 1B). 44 sequences from *C. hubbsi* and 52 sequences from *C. regis* were analyzed for their haplotype diversity. Median-joining haplotype networks were drawn in POPART ver. 1.7 (Leigh and Bryant, 2015). Cyt *b* haplotypes were examined by various neutrality tests – Tajima’s *D* (Tajima, 1989), Fu and Li’s *D** and *F** (Fu and Li, 1993), Fu’s *F_s* (Fu, 1997), and *R₂* (Ramos-onsins and Rozas, 2006) – in DnaSP v. 5.10.01 (Librado and Rozas, 2009). P-values for these test statistics were obtained by coalescent simulations of 1000 replicates, conditional on the number of segregating sites. Pairwise differences were analyzed by mismatch distributions in Arlequin ver. 3.1 (Excoffier and

Schneider, 2005); coalescent simulations were run with 10,000 bootstrap replicates. Past demographic expansions can be inferred from haplotypes based on the mismatch distribution of the pairwise differences; an uneven, multi-modal distribution is expected for populations under demographic equilibrium, whereas a “smooth” uni-modal distribution is expected for populations that had undergone demographic expansion (Rogers and Harpending, 1992). The observed mismatch distributions of the two species were compared to the expected mismatch distribution under a sudden expansion model using the sum of square deviation (SSD) and the raggedness index.

RESULTS

Pure-bred Specimens of C. regis and C. hubbsi are morphologically distinct

Consistent with the findings of Crabtree (1989), our morphological analysis based on both modern and historical specimens, summarized in Fig. 2, shows that *C. regis* (green and gold) differs from *C. hubbsi* (pink) in having higher dorsal-lateral scale counts, higher vertebral counts, and shorter snout to first dorsal fin origin distance. In addition, the symphases of the hemal spines of the anterior-most caudal vertebrae of *C. regis* are modified into an expanded process (Fig. S1), forming a funnel-like structure antero-posteriorly. The number of modified hemal spines within *C. regis* outside of the Delta Edge varies between 4 to 11; the hemal spines of *C. hubbsi* in the Delta appears to show no signs of such modification (Fig. 2D). Outside the putative hybrid zone individuals are relatively easily classified to species based on morphological differences. However, more continuous variation is evident within the hybrid zone.

Introgression of Cytochrome b and RAG1 Haplotypes

Maximum-Likelihood phylogeny reconstruction of Cyt b sequence reveals ~6% divergent clades corresponding to the two species of *Colpichthys* (Fig. 1B). *C. hubbsi* haplotypes were recovered from the Delta region, while Baja and Sonoran Coast samples contained only *C. regis* haplotypes. Both species were expected to cohabit the localities at the southwestern border of the Delta (PRI, SGU, and TRC), but none of these locations contain a mixture of both haplotypes. PRI and SGU fish are all *C. regis* haplotypes, while the more Delta internal TRC sample is exclusively *C. hubbsi* haplotypes albeit with limited sample sizes. Within the Delta, five specimens collected from predominantly *C. hubbsi* populations—ELZ, CHY, and SFG—were found to possess the *C. regis* mitochondrial haplotype; these specimens are morphologically indistinguishable from other *C. hubbsi* specimens.

We identified four RAG1 variants (labeled as H_1 , H_2 , H_3 , and R ; Table 3) based on four polymorphic sites in the RAG1 sequence data (975bp). We find that H_1 , H_2 , and H_3 are common in populations of *C. hubbsi* in the Delta, while R is fixed in all populations of *C. regis* in the Baja Coast and Sonora Coast regions. Thus, the two species are well differentiated by their RAG1 genotypes, with the diagnostic SNPs at positions 717 and 909. Six specimens, heterozygous at both of these positions, are flagged as putative hybrids. Four of the six hybrids occur at the Delta Edge, while the remaining two are found within the Delta proper (Table 4). No RAG1 heterozygotes are detected in any of the Baja and Sonoran collection sites.

As noted above, morphology of the anterior-most caudal hemal spines provides a phenotypic difference between the two species of *Colpichthys*. The non-modified hemal spine phenotype (Fig. S1A) is observed in all specimens possessing any of the three *C. hubbsi* RAG1 alleles (H_1 , H_2 , or H_3).

Bayesian Analyses of Microsatellites Show Evidence of Admixture

Based on the evaluation of ΔK , the STRUCTURE analyses on the six microsatellite loci recovered two genetically distinct clusters corresponding to the two species of *Colpichthys* (Fig. 3; for $K=3$ and $K=4$, see Fig. S2). Under the admixture model, the q -value represents admixture proportion of each individual. We used a q -value threshold to distinguish putative purebred and hybrid individuals; for optimal efficiency, this threshold is set to 0.1 (Vähä and Primmer, 2006). Populations in the Delta proper—ELZ, SFG, and CHY—were identified as purebred *C. hubbsi*. Several populations away from the Delta such as ANI, NWC, and YAV were identified as purebred *C. regis*. All individuals within or near the western Delta Edge—PRI, SGU, and TRC—show substantial degrees of genetic admixture, having q -values between 0.1 and 0.9. Moreover, among these populations there appears to be a gradient in q -values as a function of geographic distance from the mouth of the Colorado River. TRC has an average q -value of 0.17 while PRI, which is farther from the river mouth, has an average q -value of 0.72 (Fig. 3A). A relatively weak signal for admixture in the ADR population beyond the eastern edge of the delta was detected exclusively in the correlated allele frequency model; this signal is absent in the analysis with the independent allele frequency model (Fig. 3B).

Consilience of Phenotypic and Nuclear Data

Results from separate DAPC analyses were pooled to compare phenotypic and microsatellite variation. Two clusters are supported, consequently only one discriminant function was generated for each of the DAPC analyses. These results show two well separated species with congruent cluster assignment from both data sets (Fig. 3C). Delta Edge specimens

293 preferentially situate between the species clusters. Of the six RAG1 heterozygotes (triangles on
 294 Fig. 3C): two recovered in the Delta Edge area fall near the midpoint in morphology between the
 295 two species; two recovered in the Delta proper are near the centroid of *C. hubbsi* and the last two
 296 from the Delta Edge are phenotypically more similar to *C. hubbsi* than *C. regis*.

297

298 *Genetic Diversity and Demographic History*

The Cyt *b* median-joining haplotype networks of the two species are displayed in Fig. 4A and Fig. 4B. Twenty-Eight *C. hubbsi* haplotypes are recovered from the four Delta populations; in contrast, only ten haplotypes are found in *C. regis* over a larger, more disjunct geographic range. *C. regis* haplotype diversity differs substantially between the Baja and the Sonoran Coasts; with the exception of one individual from PRI, all non-major haplotypes are found exclusively in Sonora Coast populations. The major *C. regis* haplotype is present in all locations outside the Delta except the two southernmost Sonora Coast sites, SLD and YAV. All five individuals within the Delta proper with the *C. regis* Cyt *b* sequence possess the major haplotype. Neutrality tests were conducted with the exclusion of the mitochondrial-introgressed individuals; all neutrality tests report significant departure from neutrality. Significantly negative values for Tajima's *D* and Fu's *F_s* in both species are consistent with either population growth or selective sweep (Table 5). The sum of square deviations and raggedness indices for the mismatch distributions are close to zero and not significant (Table 6), thus we are unable to reject the sudden expansion model for both species. Fig. 4C shows the nucleotide mismatch distributions for both species; *C. hubbsi* has a bimodal distribution, while *C. regis* has a unimodal distribution that matches closely with the expected curve under sudden expansion.

DISCUSSION

The Hybrid Zone

Crabtree (1989) recognized that the two *Colpichthys* species co-occur in the region just north of San Felipe, designated the Delta Edge in this study. While external morphology of specimens collected in 1968 confirms the co-occurrence of the two species at this region (Table S1), there is no indication from these samples of historical hybridization. Our study presents the

first evidence for the modern hybridization of the sister species. STRUCTURE and DAPC analyses strongly support our hypothesis that the Delta Edge region currently represents a hybrid zone. STRUCTURE results limited to the correlated allele frequency model raise the possibility that hybridization may also be occurring on the eastern border of the Delta at Bahía Adair (ADR). The fact that weak admixture signals at ADR are not detected using our other approaches suggests that hybridization at this region, if present, is likely infrequent.

Directional Introgression

In addition to hybridization at the Delta Edge, we detected *C. regis* mitochondrial and RAG1 haplotypes near the mouth and upstream along the Colorado Estuary within the Delta. This could be explained by either the dispersal of first generation hybrids from the hybrid zone to the Delta or by multigenerational backcrossing of hybrids with parental *C. hubbsi* in this case far from the Delta Edge (i.e. introgressive hybridization). Our data lend support to the introgressive hybridization interpretation. Unlike the RAG1 hybrids at the Delta Edge, specimens that possess *C. regis* mitochondrial and/or RAG1 haplotypes within the Delta do not exhibit intermediate morphology. Morphological and microsatellite data for these specimens are typical for *C. hubbsi*. Thus, they likely represent the product of multiple backcrosses. Thus *C. regis* haplotypes have been introduced into the center of the *C. hubbsi* range via introgressive hybridization. Conversely, we failed to detect any *C. hubbsi* haplotypes outside of the Delta and Delta Edge regions. Thus, introgression does not appear to be reciprocal, although more sampling south of the Delta Edge would allow a stronger conclusion that *C. hubbsi* haplotypes are absent in predominantly *C. regis* populations.

Parental Sex Asymmetry in Introgression

Phenotypic and microsatellite data (Fig. 3C, Fig. S3) reveal a narrow geographic cline across the 10km wide hybrid zone which spans 3 estuaries at the southwestern Delta border—Estero Primero (PRI), Estero Segundo (SGU), and Estero Tercero (TRC). Yet despite morphological and nuclear evidence of hybridization all thirteen specimens from PRI and SGU possess the *C. regis* mitochondrial haplotype. Given matrilineal mitochondrial inheritance, this pattern may indicate hybridization between female parental *C. regis* and male parental *C. hubbsi*. Little is known about the life history and spawning behavior of *Colpichthys*. However, Crabtree (1989) noted that the ovipositor of *C. hubbsi* is twice the size of that in *C. regis*, suggesting that differences in reproductive organs could contribute to the directionality of introgression. Behavioral or ecological differences may also play a role, but these hypotheses have yet to be assessed.

Timing of Divergence

Due to the lack of available genetic data from historical samples, the onset of hybridization cannot be easily constrained. However, the ~6% mitochondrial sequence divergence between *C. regis* and *C. hubbsi* suggests that the two species have long been separated with hybridization a relatively recent phenomenon. This mitochondrial divergence is comparable to that of another fish sister species pair, *Gillichthys mirabilis* and *Gillichthys detrusus* (Swift et al., 2011), which share distributional patterns with *C. regis* and *C. hubbsi*. Like *C. hubbsi*, *G. detrusus* is narrowly restricted to the silty tidal channels of the Colorado River Delta; its congener, *G. mirabilis*, can be found in the estuaries and lagoons on both coasts of the Gulf. Swift et al. (2011) estimated a ~5Ma divergence time between *G. detrusus* and *G.*

mirabilis. Given that both *C. hubbsi* and *G. detrusus* are Delta endemics, their divergences from their sister species may have been linked to the formation of the Delta estuaries, which occurred after the opening of the Gulf of California. Fossils of *C. regis* have been discovered in the Bouse Formation at the location of Cibola Lake, Arizona (Todd, 1976). The Bouse Formation records the changing depositional environments associated with Late Miocene/Early Pliocene progression of the Colorado River to the Gulf (Poulson and John, 2003; McDougall and Miranda Martinez, 2014; Spencer and Patchett, 1997; Roskowski et al. 2010). Stratigraphic and tectonic reconstructions suggest that the Colorado River reached the Gulf of California by ~4-5Ma (Winker and Kidwell, 1986; Dorsey et al., 2007; Spencer et al., 2013; Crossey et al. 2015; Howard et al., 2015). Based on this chronology, speciation leading to the Delta endemics presumably followed the establishment of the Delta habitats around this time (Swift et al., 2011; Ellingson, 2012).

Selection and Ecological Speciation in Response to Salinity Conditions

The parapatric distribution of *Colpichthys* in the Northern Gulf suggests that *C. regis* may have diverged from *C. hubbsi* due to ecological differentiation reinforced by divergent selection on local adaptations. The divergence is likely linked to the gradient in salinity in the Northern Gulf. While virtually no physiological data exists for *C. hubbsi*, *C. regis* is known to tolerate hypersaline conditions as it is frequently observed in lagoons with salinity values reaching 50 psu (Castro-Aguirre and Espinosa-Pérez, 2006). Prior to the completion of Hoover Dam in 1935, the water output at the Delta was estimated to be 16 to 18 billion cubic meters per year (Stockton and Jacoby 1976), supporting a substantial historic fresh to salt water gradient through much of the Delta. Development of an inverse hypersaline estuary followed in the 20th Century

damming and water removal from the system (Levín and Sánchez 1999). Although we are not aware of any studies of salinity tolerance in captivity for *Colpichthys*, as noted above, silversides have developed ecological species in response to salinity gradients in Brazil and Australia, the Mediterranean and the Eastern US. Thus, it is likely that historic salinity gradients in the Delta played a role in reproductive isolation between the *Colpichthys* sister species. Moreover, selection maintaining these species is prone to breakdown following the loss of freshwater.

Due to the difference in density of fresh and saltwater, salinity is known to influence the morphology of the gas bladder, which maintains neutral buoyancy in teleost fishes. Among its closest relatives in Atherinopsinae, *C. hubbsi* is the only species where the gas bladder does not extend posterior to the visceral cavity (Crabtree, 1989), suggesting that salinity related selection impacts the species. Other potential selective factors may involve the unusual silty conditions and tides in the Delta and the limited depth in the Delta. Relative to other examples of species boundaries particularly maintained by selection and differences in salinity in fishes such as those associated with the transition into the Baltic, which developed during the Holocene (Barrio et al., 2016); or those in Brazilian coastal silversides where divergence of freshwater forms was in the Pleistocene (Beheregaray et al., 2002), our inferred early Pliocene timing of speciation based on sequence differences between *C. regis* and *C. hubbsi* is substantially more ancient.

Evolutionary Implications of introgression for Colpichthys

Introgressive hybridization potentially has a number of distinct causes, including anthropogenic introduction and habitat modification. It also has a number of different evolutionary outcomes ranging from sharing of advantageous alleles between species (Martin et al., 2006; Whitney et al., 2010; Pardo-Diaz et al., 2012; Hedrick, 2003) to extinction (Rhymer

and Simberloff, 1996; Todesco et al., 2016). Of particular interest here are examples of habitat alteration that lead to "speciation reversal", or the introgressive loss of ecological speciation. Eutrophication, for example, has led to numerous examples of species loss in fishes (Vonlanthen et al., 2012; Seehausen et al., 1997). Given that salinity is associated with ecological species in a number of silversides, and the loss of freshwater in the Colorado Delta, introgression into the more freshwater adapted *C. hubbsi* from the marine adapted *C. regis* would be expected. Ultimately, introgression of *C. regis* alleles adapted for higher salinity could leave *C. hubbsi* at risk of genetic swamping. In the long run, the risk of extinction by speciation reversal for *C. hubbsi* will depend on the extent to which the loss of salinity gradient has eliminated the habitat distinction in the Northern Gulf. While this study only identified backcrossed hybrids based on the limited genetic loci sampled, it is possible that many of the morphological "pure bred" *C. hubbsi* are in fact "cryptic" hybrids. Depending on the frequency and relative abundance of such "cryptic" hybrids, interspecific gene flow may be more pervasive than suggested by our data. To assess whether "speciation reversal" could pose a threat to *C. hubbsi*, future whole-genome analyses will be required to quantify the amount of genomic introgression as well as identify loci that may remain under divergent selection in the face of gene flow.

Demographic expansions

Given the geographic restriction of *C. hubbsi*, one immediate concern is whether the reduced Colorado River outflow has directly contributed to a population decline. The rise in salinity at the Delta has been implicated in the decline of the endemic Delta clams, *Mulinia coloradoensis* (Rodriguez et al., 2001). Based on our results from neutrality tests and mismatch analyses, we find significant departure from neutrality, but no evidence to support a recent

437 bottleneck in *C. hubbsi*. On the other hand, contrary to initial expectation, *C. regis* exhibits
 438 surprisingly low genetic diversity compared to its geographically restricted congener. This
 439 pattern is seen in all three types of molecular markers (Cyt *b*: Fig. 4; RAG1: Table 3;
 440 microsatellites: Table S2). The results from this study are consistent with demographic
 441 expansions in the recent past. Other factors, such as selective sweep or fine-scaled population
 442 structure (Ptak and Przeworski, 2002), could produce similar mtDNA patterns, but these
 443 mechanisms seem unlikely to have generated the similar patterns of diversity in all 3 data sets.
 444 We offer the following scenario as a possible explanation for the observed data. In the mismatch
 445 analysis, the model parameter τ estimates the age of demographic expansion. Our results (Table
 446 6) indicate that *C. hubbsi* may have experienced a more ancient expansion relative to *C. regis*
 447 and potentially attained larger populations in the large Delta estuarine habitat. One possible
 448 explanation for this could come from Pleistocene climate and sea-level fluctuations. Recent work
 449 (Dolby et al. 2016) demonstrated that interactions between coastal geomorphology and
 450 Pleistocene sea-level fluctuations limited estuary habitats to refugia during glacial sea-level
 451 minima, impacting modern population structures of estuarine fish. In addition, during the
 452 Pleistocene, wetter climates increased Colorado River outflow into the Gulf, while the gentle
 453 sloping bathymetry of the Northern Gulf together with the lowering of sea-level likely expanded
 454 habitats for *C. hubbsi* much further south,  allowing for population growth. Meanwhile, the
 455 steeper coastal shelves in the lower portion of the Gulf could have restricted suitable refugial
 456 habitats for *C. regis* during sea-level lowstand. The lack of genetic diversity in *C. regis*
 457 populations may represent rapid range expansion from these limited refugia following sea-level
 458 rise at the end of the Pleistocene. Crucial to this hypothesis is the precise timing of the
 459 demographic and range expansions. More complete sampling of *C. regis* within each of the

modern localities or genomic approaches are needed in future studies to further explore this issue.

Potential implications for Delta endemics

Of the numerous Gulf endemics (Palacios-Salgado et al., 2012), some are confined to the northernmost Gulf, and a smaller subset are very locally endemic to the Delta itself, which is an area protected as a Biosphere Reserve (Reservade la Biosfera del Alto Golfo de California y Delta del Rio Colorado). Of these northernmost endemics, some are relicts of broader distributions, while others appear to be ecological species that evolved in association with the Colorado Delta (Jacobs et al. 2004). Apparent relictual species include the critically endangered Vaquita, the world's smallest cetacean, whose closest relative is found on the coasts of South America south of Peru (Munguia-Vega et al., 2007). The totoaba (*Totoaba macdonaldi*) was historically an important fishery resource that depended on freshwater input for rapid growth (Rowell et al., 2008). This taxon appears to be similarly relictual; although the genus is known from fossils in the Miocene of California Central Valley (Huddleston & Takeuchi, 2007), its closest relatives are in the Atlantic and IndoPacific (Lo et al., 2015). In addition to these relicts which may have been directly impacted by salinity changes or other anthropogenic activities such as overfishing, three taxa in the region appear to have evolved ecological species pairs associated with the Delta. Apart from the fish genera *Colpichthys* and *Gillichthys* as discussed above, the fiddler Crab *Uca monilifera* and *U. princeps* conform to the same pattern (Brusca, 1980). Thus, out of a suite of five northernmost Gulf endemic estuarine species, three are potentially at risk of introgression due to hydrologic modification. While a number of studies have argued that the northern Gulf fauna has been directly impacted by salinity (Rodriguez et al.,

2001; Rowell et al., 2008), others have emphasized the minimal change in Northern Gulf biological productivity associated with the anthropogenic loss of freshwater input from the Colorado River (Brusca et al., 2017). However, there may nevertheless be strong impacts on the endemic diversity of this biologically unique region, especially with regards to introgression in endemics specific to the Colorado Delta proper.

Global Change Context

Early 20th Century water removal tended to impact larger river systems primarily in semi-arid settings, where freshwater is at a premium and where economic development is well advanced; this is well exemplified by the Southwestern US. With increasing population growth and economic development in a wider range of geographic settings, damming and diversion efforts have continued to expand from rivers that traverse arid regions, impacting the Deltas estuaries of rivers such as the Nile (Stanley & Warne 1998) and the Euphrates (Isaev & Mikhailova 2009), to many of the largest river systems of Asia, such as the Mekong (Kondolf et al. 20014). Despite this, the impact on estuarine diversity, especially as regards to the potential loss of diversity due to introgression between ecological species, has not been widely discussed or assessed. In the case of the Colorado River Delta, we have been able to establish that introgression is ongoing in the context of the ecologically altered Delta environment. This is the first step in establishing the risk of introgressive species loss in river estuaries that have suffered from water extraction and loss of freshwater input. This is yet another factor in a long list of ecological and biodiversity impacts consequent to altering riverine hydrologic processes, and should be systematically considered in a global context so that the full ecological consequences and diversity costs of such development can be properly understood.

CONCLUSION

We document the Pliocene separation and recent introgression between a sister species pair in the northern Gulf of California. Our evidence supports hybridization and introgression into the IUCN endangered Delta endemic *Colpichthys hubbsi* from the more widespread Gulf species *C. regis* following 20th century environmental alteration through water removal and salinity change in the Delta system. The work combines multiple lines of phenotypic and genetic evidence to demonstrate this directional introgressive hybridization. We find no evidence for

population decline in *C. hubbsi*, yet continued introgression may jeopardize the genetic integrity of this taxon. Genetic diversity of *C. regis* is revealed to be lower than expected for a healthy, widely distributed species so it likely merits its IUCN threatened designation. Our work highlights the need for closer monitoring and further demographic investigations of *Colpichthys* populations not just within the Delta but also throughout the Northern Gulf. More generally, our work implies that other Colorado Delta-specific endemics such as the goby *Gillichthys detrusus* and the fiddler crab *Uca monilifera* may also be ecological species at risk of loss through introgression from their proximally distributed congeners. More broadly, a general suite of ecological species around the world that evolved in response to salinity gradients in major river systems are likely at increasing risk of introgressive extinction as water extraction and climate change impacts accelerate in river systems globally.

ACKNOWLEDGEMENTS

The authors would like to thank Kirk Lohmueller, Blaire Van Valkenburgh, and Jonathan Richmond for feedback and comments; Rick Feeney of the Los Angeles County Natural History Museum for access to an X-ray machine and the specimens in the ichthyology collection; H.J. Walker and Phil Hastings of the Scripps Institute of Oceanography for providing the historical specimens; Greer Dolby and Ryan Ellingson for advice on data analysis; Bruno Passerelli for assistance in data collection; Lloyd Findley for facilitating permitting and collection. Collections were under Mexican federal collecting permit (Permiso de Pesca de Fomento) DGOPA 14253. 101005.6950, and its extension DGOPA 06435.210606.2640, issued to Findley and Jacobs by the Comisión Nacional de Acuacultura y Pesca of the Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA). Assistance with field sampling was

provided by T. Baumiller, V. Cassano, R. Ellingson, R. Hechinger, F. Hertel, J. Lorda, and D. Yuan.

LITERATURE CITED

Akihito, Iwata A, Kobayashi T *et al.* (2000) Evolutionary aspects of Gobioid fishes based upon a phylogenetic analysis of mitochondrial cytochrome b genes. *Gene*, **259**, 5–15.

Bamber R, Henderson P (1988) Pre-adaptive plasticity in atherinids and the estuarine seat of teleost evolution. *Journal of Fish Biology*, **33**, 17–23.

Barrio AM, Lamichhaney S, Fan G *et al.* (2016) The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *Elife*, **5**, e12081.

Beheregaray LB, Sunnucks P (2000) Microsatellite loci isolated from *Odontesthes argentinensis* and the *O. perugiae* species group and their use in other South American silverside fish. *Molecular Ecology*, **9**, 629–631.

Beheregaray LB, Sunnucks P (2001) Fine-scale genetic structure, estuarine colonization and incipient speciation in the marine silverside fish *Odontesthes argentinensis*. *Molecular Ecology*, **10**, 2849–2866.

Beheregaray LB, Sunnucks P, Briscoe DA (2002) A rapid fish radiation associated with the last sea-level changes in southern Brazil: The silverside *Odontesthes perugiae* complex. *Proceedings of the Royal Society of London B: Biological Sciences*, **269**, 65–73.

Bonetto A, Wais J, Castello H (1989) The increasing damming of the Paraná basin and its effects on the lower reaches. *River Research and Applications*, **4**, 333–346.

Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques*, **31**, 24–6.

Brusca RC (1980) *Common intertidal invertebrates of the Gulf of California*. Univ of Arizona Pr.

Brusca RC, Álvarez-Borrego S, Hastings PA, Findley LT (2017) Colorado river flow and biological productivity in the northern Gulf of California, Mexico. *Earth-Science Reviews*, **164**, 1–30.

Byrne RJ, Avise JC (2009) Multiple paternity and extra-group fertilizations in a natural population of California grunion (*Leuresthes tenuis*), a beach-spawning marine fish. *Marine Biology*, **156**, 1681–1690.

Carriquiry J, Sánchez A (1999) Sedimentation in the Colorado River Delta and upper Gulf of California after nearly a century of discharge loss. *Marine Geology*, **158**, 125–145.

- 570 Castro-Aguirre JL, Espinosa Pérez H (2006) Los peces de la familia Atherinopsidae (Teleostei:
571 Atheriniformes) de las lagunas costeras neutras e hipersalinas de México. *Hidrobiológica*, **16**,
572 89–101.
- 573 Crabtree CB (1989) A New Silverside of the Genus Colpichthys (Atheriniformes: Atherinidae)
574 from the Gulf of California, Mexico. *Copeia*, **1989**, 558–568.
- 575 Crossey LC, Karlstrom KE, Dorsey R *et al.* (2015) Importance of groundwater in propagating
576 downward integration of the 6-5 Ma Colorado River system: Geochemistry of springs,
577 travertines, and lacustrine carbonates of the Grand Canyon region over the past 12 Ma.
578 *Geosphere*, **11**, 660–682.
- 579 Dolby GA, Hechinger R, Ellingson RA *et al.* (2016) Sea-level driven glacial-age refugia and
580 post-glacial mixing on subtropical coasts, a palaeohabitat and genetic study. *Proceedings of the*
581 *Royal Society of London B: Biological Sciences*, **283**.
- 582 Dorsey RJ, Fluette A, McDougall K *et al.* (2007) Chronology of Miocene–Pliocene deposits at
583 Split Mountain Gorge, Southern California: A record of regional tectonics and Colorado River
584 evolution. *Geology*, **35**, 57.
- 585 Drummond A, Ashton B, Buxton S *et al.* (2011) Geneious pro v5. 5.6. available from
586 <http://www.geneious.com/>.
- 587 Earl DA, others (2012) STRUCTURE harvester: A website and program for visualizing structure
588 output and implementing the Evanno method. *Conservation genetics resources*, **4**, 359–361.
- 589 Ellingson R (2012) *Phylogenetics and phylogeography of north Pacific bay gobies: Adaptive*
590 *convergence, relictual endemism, and climate-driven population structure*. University of
591 California, Los Angeles.
- 592 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the
593 software structure: A simulation study. *Molecular ecology*, **14**, 2611–2620.
- 594 Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package
595 for population genetics data analysis. *Evolutionary bioinformatics*, **1**.
- 596 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus
597 genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- 598 Findley L, Collette B, Espinosa H (2010a) Colpichthys hubbsi. *The IUCN Red List of Threatened*
599 *Species 2010: e.T183457A8116697*.
- 600 Findley L, Collette B, Espinosa H (2010b) Colpichthys regis. *The IUCN Red List of Threatened*
601 *Species 2010: e.T183357A8099025*.
- 602 Fluker BL, Pezold F, Minton RL (2011) Molecular and morphological divergence in the inland
603 silverside (Menidia beryllina) along a freshwater-estuarine interface. *Environmental biology of*
604 *fishes*, **91**, 311.

- 605 Francisco SM, Cabral H, Vieira MN, Almada VC (2006) Contrasts in genetic structure and
606 historical demography of marine and riverine populations of *Atherina* at similar geographical
607 scales. *Estuarine, Coastal and Shelf Science*, **69**, 655–661.
- 608 Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking
609 and background selection. *Genetics*, **147**, 915–925.
- 610 Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.
- 611 Glenn EP, Lee C, Felger R, Zengel S (1996) Effects of Water Management on the Wetlands of
612 the Colorado River Delta , Mexico., **10**, 1175–1186.
- 613 Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species
614 boundaries. *Journal of Heredity*, **105**, 795–809.
- 615 Hedrick PW (2013) Adaptive introgression in animals: Examples and comparison to new
616 mutation and standing variation as sources of adaptive variation. *Molecular ecology*, **22**, 4606–
617 4618.
- 618 Heiser CB (1973) Introgression re-examined. *The Botanical Review*, **39**, 347–366.
- 619 Howard KA, House PK, Dorsey RJ, Pearthree PA (2015) River-evolution and tectonic
620 implications of a major Pliocene aggradation on the lower Colorado river: The bullhead
621 alluvium. *Geosphere*, **11**, 1–30.
- 622 Hubbs CL, Lagler KF, Smith GR, others (2004) *Fishes of the great lakes region*. University of
623 Michigan.
- 624 Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with
625 the assistance of sample group information. *Molecular ecology resources*, **9**, 1322–32.
- 626 Huddleston RW, Takeuchi GT (2007) First fossil record of *Totoaba villamar* 1980 (Teleostei:
627 Sciaenidae) based upon early Miocene otoliths from California with comments on the ontogeny
628 of the saccular otolith. *Bulletin, Southern California Academy of Sciences*, **106**, 1–15.
- 629 Isaev VA, Mikhailova MV. (2009) The Hydrography, Evolution, and Hydrological Regime of
630 the Mouth Area of the Shatt Al-Arab River. *Water Resources*, **36**, 380–395.
- 631 Jacobs DK, Haney TA, Louie KD (2004) Genes, diversity, and geologic process on the Pacific
632 coast. *Annu. Rev. Earth Planet. Sci.*, **32**, 601–652.
- 633 Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for
634 dealing with label switching and multimodality in analysis of population structure.
635 *Bioinformatics*, **23**, 1801–1806.
- 636 Jenkins OP, Evermann BW (1889) Description of eighteen new species of fishes from the Gulf
637 of California.

- Johnson M, Watts R, Black R (1994) High levels of genetic subdivision in peripherally isolated populations of the Atherinid fish *Craterocephalus capreoli* in the Houtman Abrolhos islands, western Australia. *Marine Biology*, **119**, 179–184.
- Jombart T, Ahmed I (2011) Adegnet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
- Klossa-Kilia E, Papasotiropoulos V, Tryfonopoulos G, Alahiotis S, Kiliass G (2007) Phylogenetic relationships of *Atherina hepsetus* and *Atherina boyeri* (pisces: Atherinidae) populations from Greece, based on mtDNA sequences. *Biological journal of the Linnean Society*, **92**, 151–161.
- Kondolf, G. M., Rubin, Z. K., & Minear, J. T. (2014). Dams on the Mekong: cumulative sediment starvation. *Water Resources Research*, *50*(6), 5158-5169.
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, **33**, 1870–1874.
- Lavin M, Sánchez S (1999) On how the Colorado River affected the hydrography of the upper Gulf of California. *Continental Shelf Research*, **19**, 1545–1560.
- Lee CE (2016) Evolutionary mechanisms of habitat invasions, using the copepod *Eurytemora affinis* as a model system. *Evolutionary applications*, **9**, 248–270.
- Leigh JW, Bryant D (2015) Popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110–1116.
- Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Liermann CR, Nilsson C, Robertson J, Ng RY (2012) Implications of dam obstruction for global freshwater fish diversity. *BioScience*, **62**, 539–548.
- Lo P-C, Liu S-H, Chao NL *et al.* (2015) A multi-gene dataset reveals a tropical new world origin and early Miocene diversification of croakers (Perciformes: Sciaenidae). *Molecular phylogenetics and evolution*, **88**, 132–143.
- Lu XX, Siew RY (2006) Water discharge and sediment flux changes over the past decades in the Lower Mekong River: possible impacts of the Chinese dams. *Hydrology and Earth System Sciences Discussions*, **10**, 181–195.
- Martin NH, Bouck AC, Arnold ML (2006) Detecting Adaptive Trait Introgression Between *Iris fulva* and *I. brevicaulis* in Highly Selective Field Conditions. *Genetics*, **172**, 2481–2489.
- McDougall K, Martínez AYM (2014) Evidence for a marine incursion along the lower Colorado River corridor. *Geosphere*, **10**, 842–869.
- Munguia-Vega A, Esquer-Garrigos Y, Rojas-Bracho L *et al.* (2007) Genetic drift vs. natural selection in a long-term small isolated population: Major histocompatibility complex class ii variation in the Gulf of California endemic porpoise (*Phocoena sinus*). *Molecular ecology*, **16**, 4051–4065.

- 676 Nilsson C, Berggren K (2000) Alterations of riparian ecosystems caused by river regulation: dam
677 operations have caused global-scale ecological changes in riparian ecosystems. how to protect
678 river environments and human needs of rivers remains one of the most important questions of
679 our time. *BioScience*, **50**, 783–792.
- 680 Nilsson C, Reidy CA, Dynesius M, Revenga C (2005) Fragmentation and flow regulation of the
681 world's large river systems. *Science*, **308**, 405–408.
- 682 Olsen Z, Anderson J, McDonald D (2016) Morphological and molecular variation among
683 populations of tidewater (*Menidia peninsulæ*) and inland (*M. beryllina*) silversides: Insight into
684 drivers of adaptation and speciation of silverside fishes. *Environmental Biology of Fishes*, **99**,
685 857–871.
- 686 Palacios-Salgado DS, Burnes-Romo LA, Tavera JJ, Ramírez-Valdez A (2012) Endemic fishes of
687 the Cortez biogeographic province (Eastern Pacific Ocean). *Acta Ichthyologica et Piscatoria*, **42**,
688 153–164.
- 689 Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive Introgression across Species
690 Boundaries in *Heliconius* Butterflies (M R. Kronforst, Ed.). *PLoS Genetics*, **8**, e1002752.
- 691 Poulson SR, John BE (2003) Stable isotope and trace element geochemistry of the basal Bouse
692 Formation carbonate, southwestern United States: Implications for the Pliocene uplift history of
693 the Colorado Plateau. *Bulletin of the Geological Society of America*, **115**, 434–444.
- 694 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus
695 genotype data. *Genetics*, **155**, 945–959.
- 696 Ptak SE, Przeworski M (2002) Evidence for population growth in humans is confounded by fine-
697 scale population structure. *Trends in Genetics*, **18**, 559–563.
- 698 Ramos-onsins SE, Rozas J (2006) Statistical Properties of New Neutrality Tests Against
699 Population Growth. *Molecular Biology and Evolution*, **23**, 1642–1642.
- 700 Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review*
701 *of Ecology and Systematics*, **27**, 83–109.
- 702 Rodriguez C, Flessa K, Dettman D (2001) Effects of upstream diversion of Colorado River water
703 on the estuarine bivalve mollusc *Mulinia coloradoensis*. *Conservation Biology*, **15**, 249–258.
- 704 Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise
705 genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.
- 706 Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure.
707 *Molecular Ecology Notes*, **4**, 137–138.
- 708 Roskowski JA, Patchett PJ, Spencer JE *et al.* (2010) A late Miocene-early Pliocene chain of
709 lakes fed by the Colorado River: Evidence from Sr, C, and O isotopes of the Bouse Formation
710 and related units between Grand Canyon and the Gulf of California. *Geological Society of*
711 *America Bulletin*, **122**, 1625–1636.

Rowell K, Flessa KW, Dettman DL *et al.* (2008) Diverting the Colorado River leads to a dramatic life history shift in an endangered marine fish. *Biological Conservation*, **141**, 1138–1148.

Seehausen O, Van Alphen JJ, Witte F (1997) Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, **277**, 1808–1811.

Spencer JE, Patchett PJ (1997) Sr isotope evidence for a lacustrine origin for the upper Miocene to Pliocene Bouse formation, lower Colorado River trough, and implications for timing of Colorado Plateau uplift. *Geological Society of America Bulletin*, **109**, 767–778.

Spencer JE, Patchett PJ, Pearthree PA *et al.* (2013) Review and analysis of the age and origin of the Pliocene Bouse Formation, lower Colorado River Valley, southwestern USA. *Geosphere*, **9**, 444–459.

Stanley DJ, Warne AG (1998). Nile Delta in its destruction phase. *Journal of Coastal Research*, 1998, 795-825

Stockton CW, Jacoby Jr GC (1976) Long-term surface-water supply and streamflow trends in the upper Colorado River basin based on tree-ring analyses. *Natl. Sci. Found. Lake Powell Res. Proj. Bull*, **18**, 70.

Swift CC, Findley LT, Ellingson RA, Flessa KW, Jacobs DK (2011) The delta mudsucker, *Gillichthys detrusus*, a valid species (Teleostei: Gobiidae) endemic to the Colorado River Delta, northernmost Gulf of California, México. *Copeia*, **2011**, 93–102.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, **10**, 512–526.

Tamura K, Battistuzzi FU, Billings-Ross P *et al.* (2012) Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences*, **109**, 19333–19338.

Todd TN (1976) Pliocene occurrence of the recent Atherinid fish *Colpichthys regis* in Arizona. *Journal of Paleontology*, **50(3)**, 462–466.

Todesco M, Pascual MA, Owens GL *et al.* (2016) Hybridization and extinction. *Evolutionary applications*, **9**, 892–908.

Trabelsi M, Gilles A, Fleury C *et al.* (2002) *Atherina punctata* and *Atherina lagunae* (Pisces, Atherinidae), new species found in the Mediterranean Sea. 2. molecular investigations of three atherinid species. *Comptes rendus biologiques*, **325**, 1119–1128.

Vähä JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

- 749 Vonlanthen P, Bittner D, Hudson AG *et al.* (2012) Eutrophication causes speciation reversal in
750 whitefish adaptive radiations. *Nature*, **482**, 357–362.
- 751 Whitney KD, Randell RA, Rieseberg LH (2010) Adaptive introgression of abiotic tolerance traits
752 in the sunflower *Helianthus annuus*. *New Phytologist*, **187**, 230–239.
- 753 Winker CD, Kidwell SM (1986) Paleocurrent evidence for lateral displacement of the Pliocene
754 Colorado River delta by the San Andreas fault system, southeastern California. *Geology*, **14**, 788.
- 755 Wu DY, Ugozzoli L, Pal BK, Wallace RB (1989) Allele-specific enzymatic amplification of
756 beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proceedings of the National*
757 *Academy of Sciences of the United States of America*, **86**, 2757–2760.
- 758 Yang SL, Zhang J, Zhu J *et al.* (2005) Impact of dams on Yangtze river sediment supply to the
759 sea and delta intertidal wetland response. *Journal of Geophysical Research: Earth Surface*, **110**.

Figure 1(on next page)

Map and Phylogeography of *Colpichthys*

Fig. 1 (A) Map of the coastline of the Gulf of California with sampling localities, color coded by species distribution. *C. hubbsi* native populations are colored blue; *C. regis* native populations are colored red; and populations where both species co-occur are colored magenta. Location codes are summarized in Table 1 and used in text throughout. (B) Time calibrated Maximum-Likelihood (ML) tree of cytochrome *b* (872bp). Branch lengths (substitutions per site) are indicated below the branches. One specimen of *Atherinops affinis* was used as outgroup to root the tree. Branch labels show branch lengths (substitutions/site).

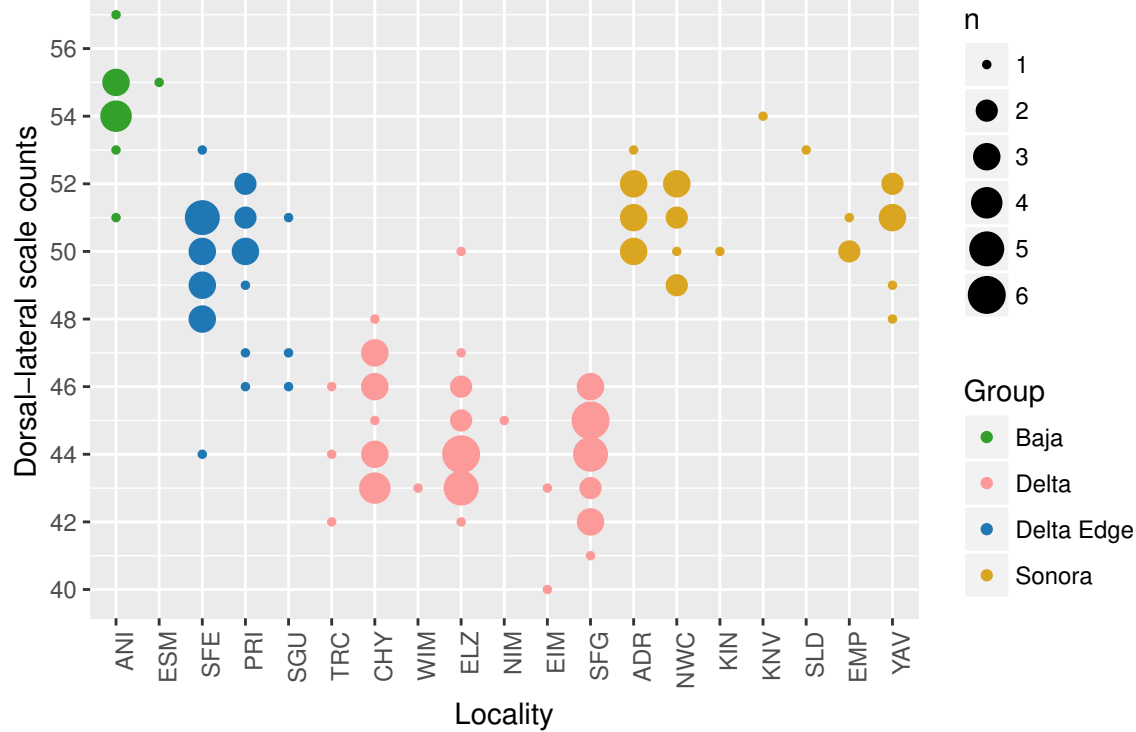
Figure 2 (on next page)

Dot plots of morphological data

Fig. 2 Dot plots showing external (A and C) and vertebral (B and D) morphological data; size of dots is proportional to the number of specimens (n). “Delta Edge” is the hypothesized hybrid zone; “Delta” is the territory for *C. hubbsi*; “Baja” and “Sonora” are the territories for *C. regis*. Refer to Table 1 and Table 2 for location codes of modern and historical specimens, respectively. Six specimens from NWC and three specimens from YAV were excluded from the hemal spine data set (D) as these specimens were too small to reliably identify the presence/absence of the expanded process on the hemal spines (See Fig. S1).

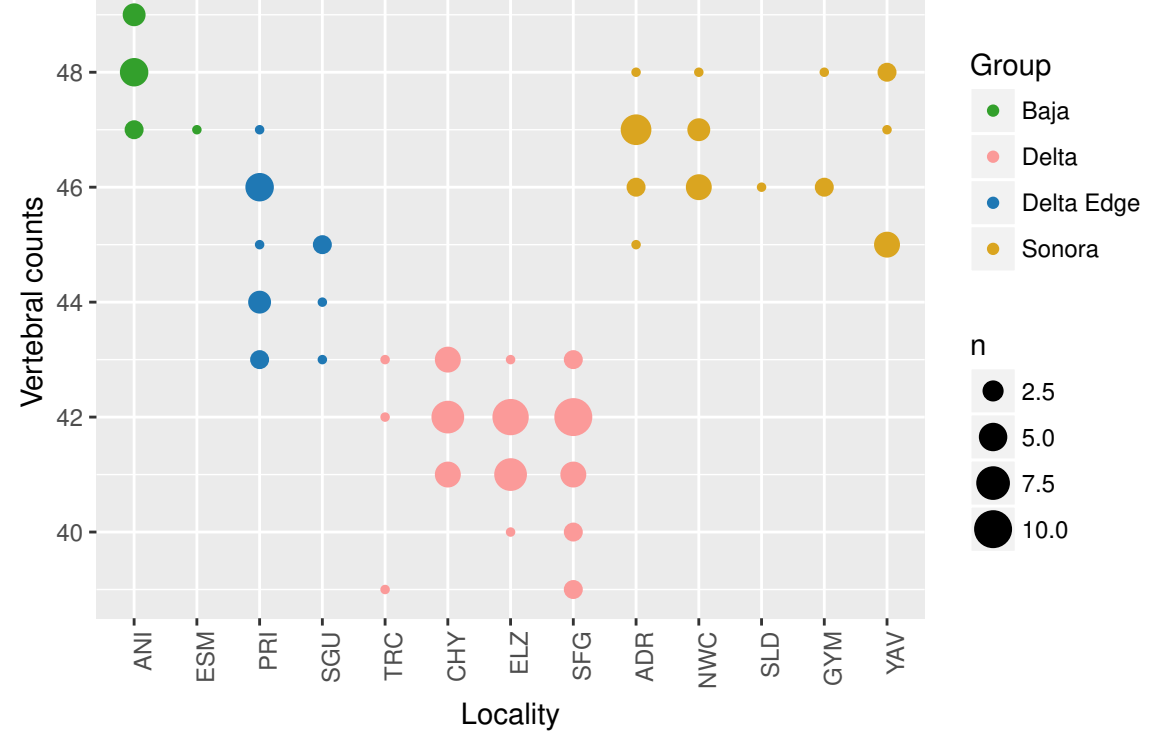
A.

PeerJ

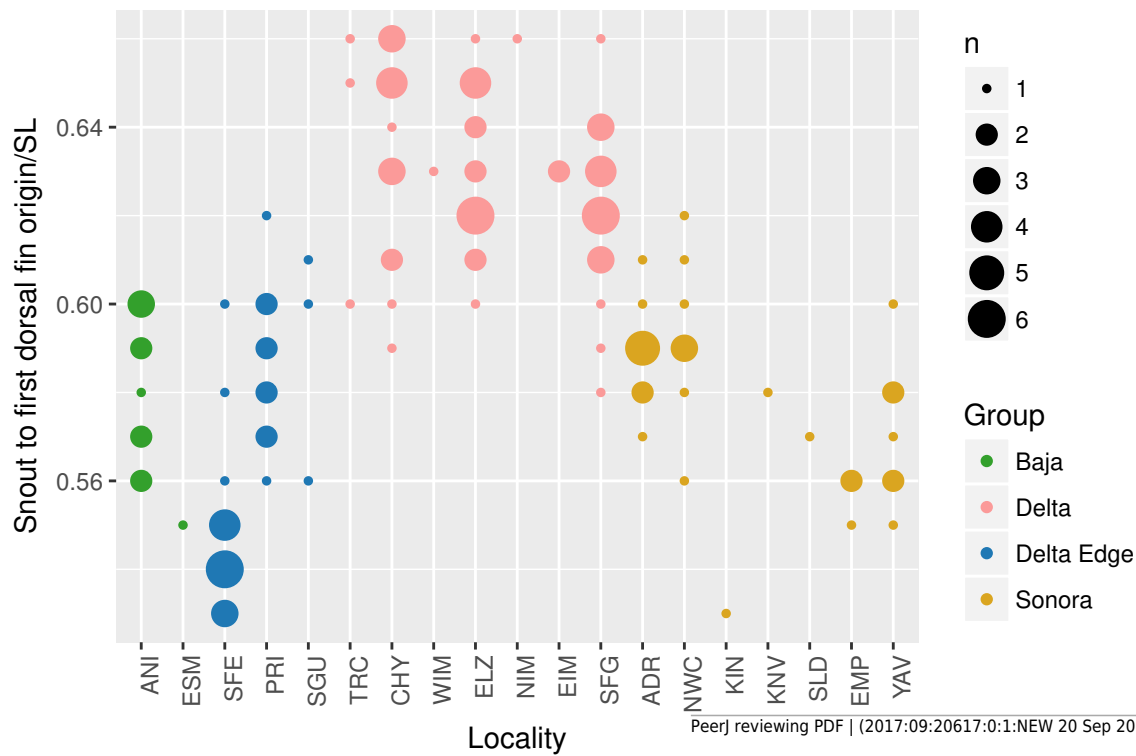


B.

Manuscript to be reviewed



C.



D.

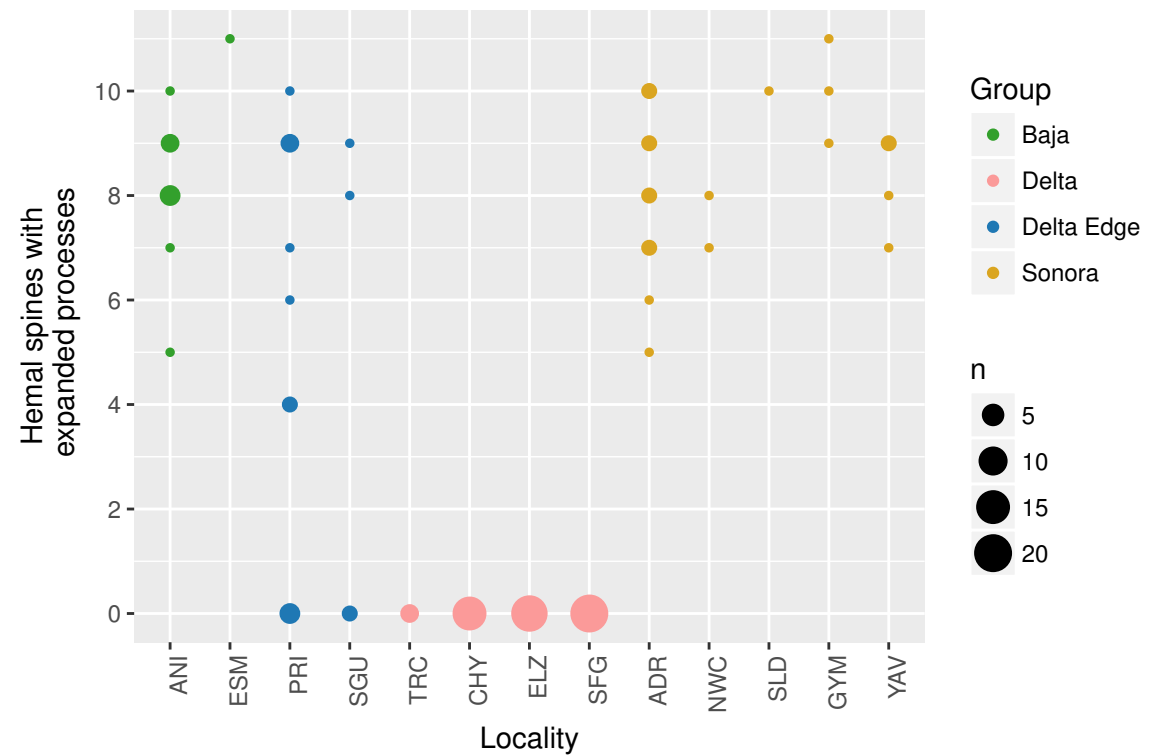


Figure 3(on next page)

STRUCTURE assignment of individuals to groups

Fig. 3 STRUCTURE bar plots showing q-values of each individual in each sampling locale for K=2. Color for each bar represents admixture proportions of *C. regis* (red) and *C. hubbsi* (blue). Correlated allele frequency model (A). Independent allele frequency model (B).

Scatter plot of the first discriminant functions in the DAPC analyses based on phenotypic and microsatellite data (C). Histograms on the top and the right side show the distributions for each axis.

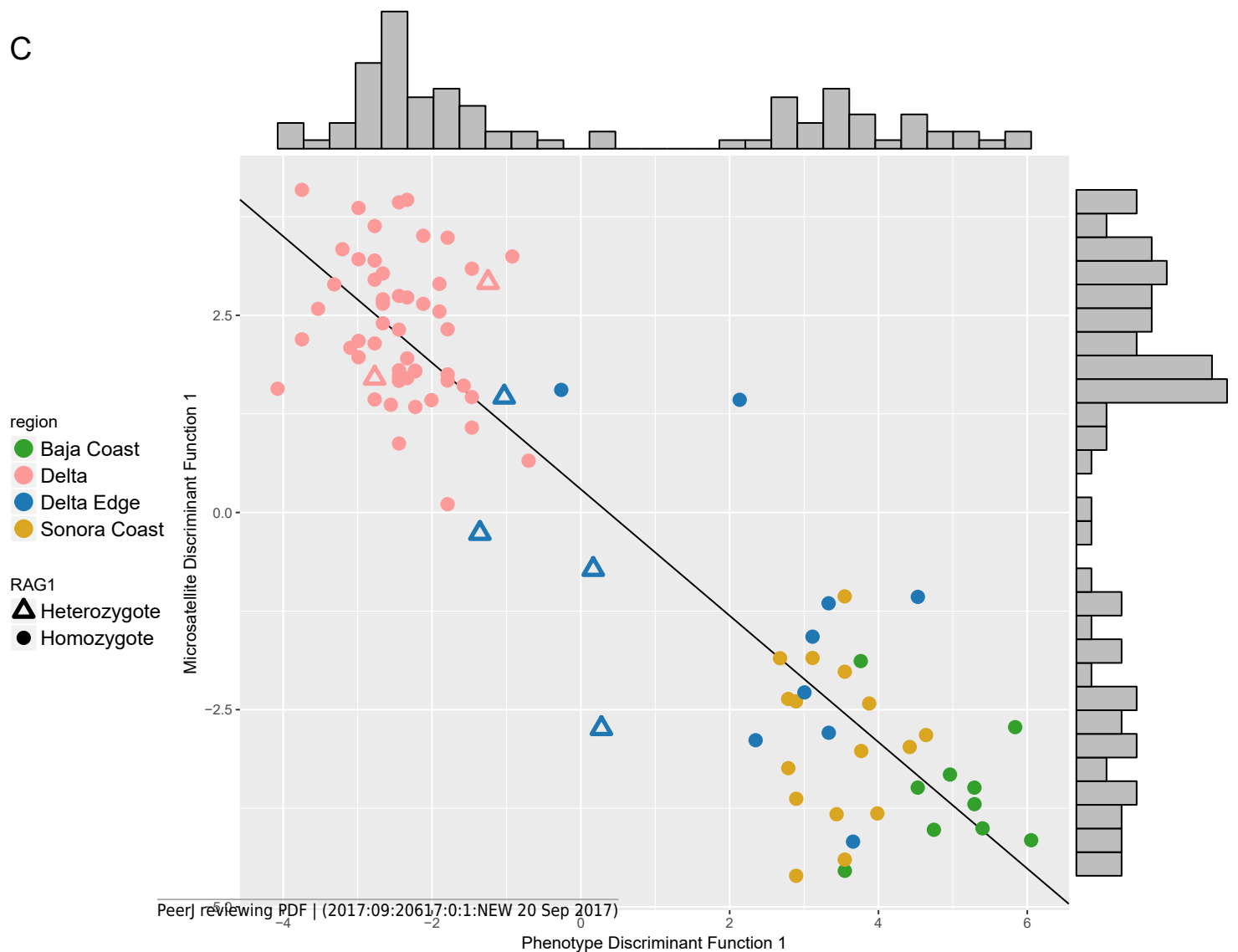
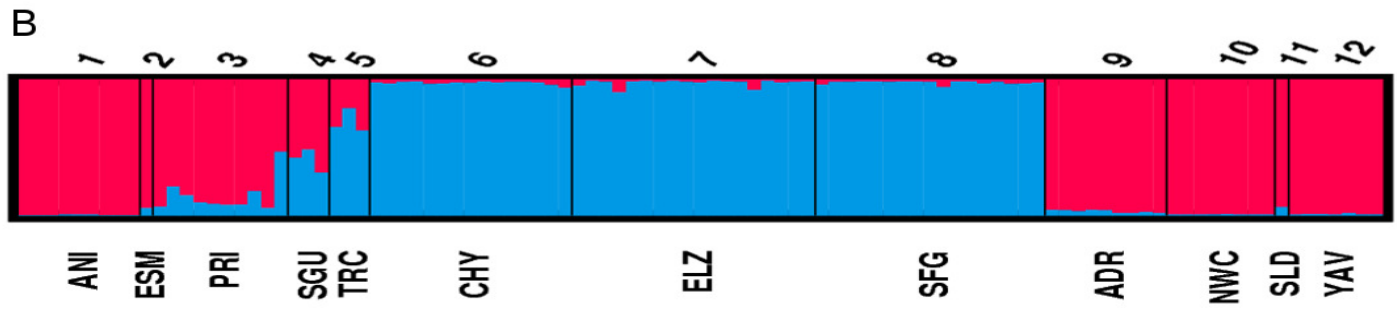
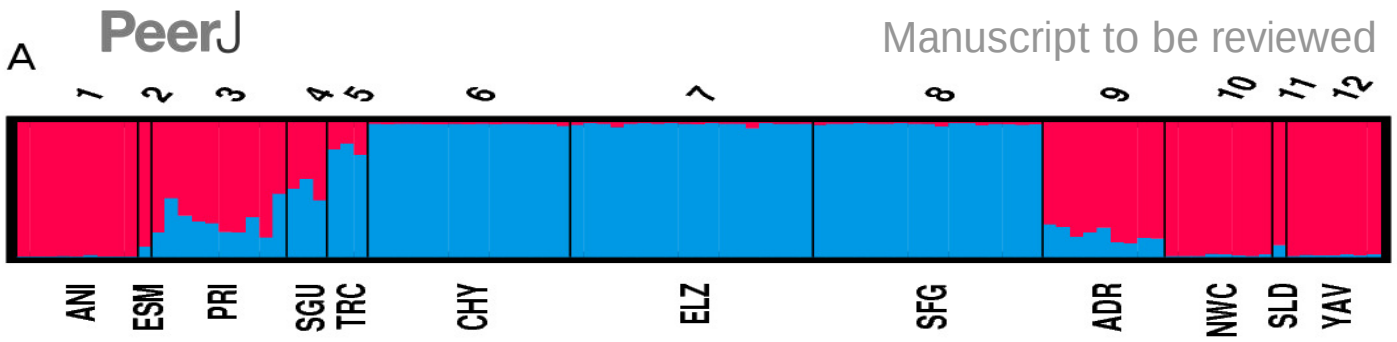
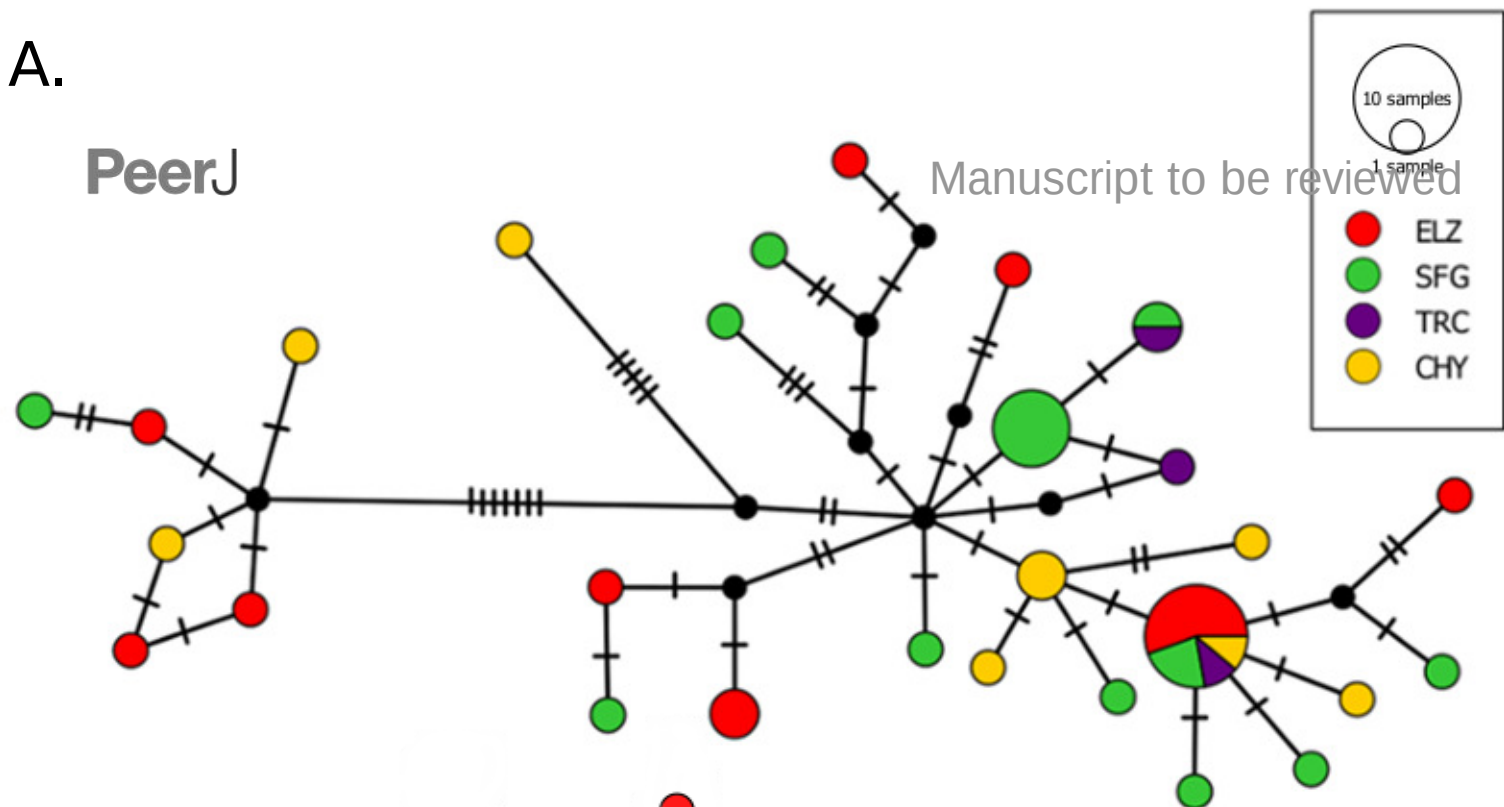


Figure 4(on next page)

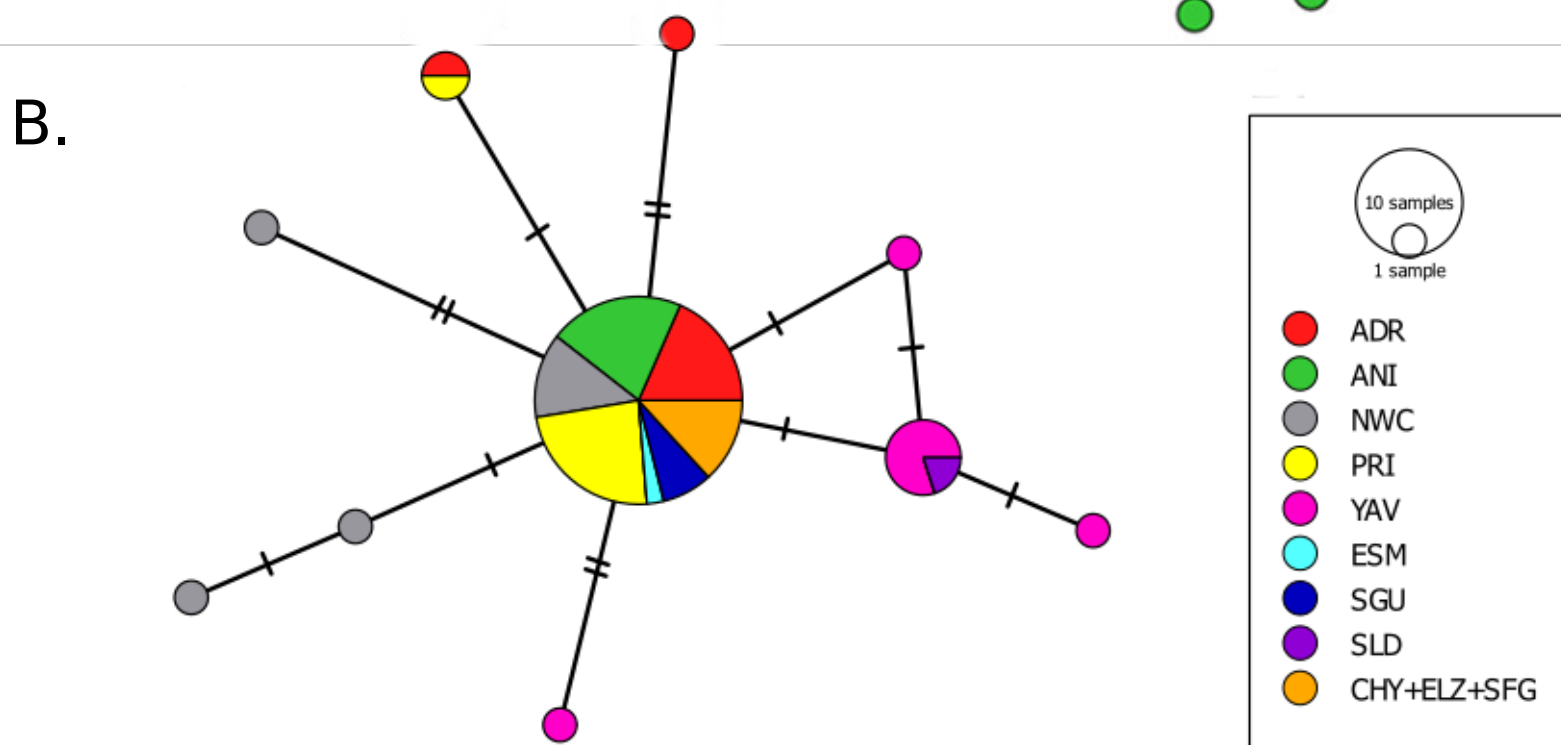
Cytochrome b haplotype networks

Fig. 4 Cyt *b* median-joining haplotype networks of *C. hubbsi* (A) and *C. regis* (B). Each circle represents a unique haplotype; size of circles indicates number of specimens with that haplotype. Each tick mark represents a nucleotide difference; black circles represent inferred missing haplotypes. Mismatch distributions for *C. hubbsi* and *C. regis* (C). Histograms represent the observed pairwise differences; black curves represent the expected pairwise differences under the sudden expansion model.

A.



B.



C.

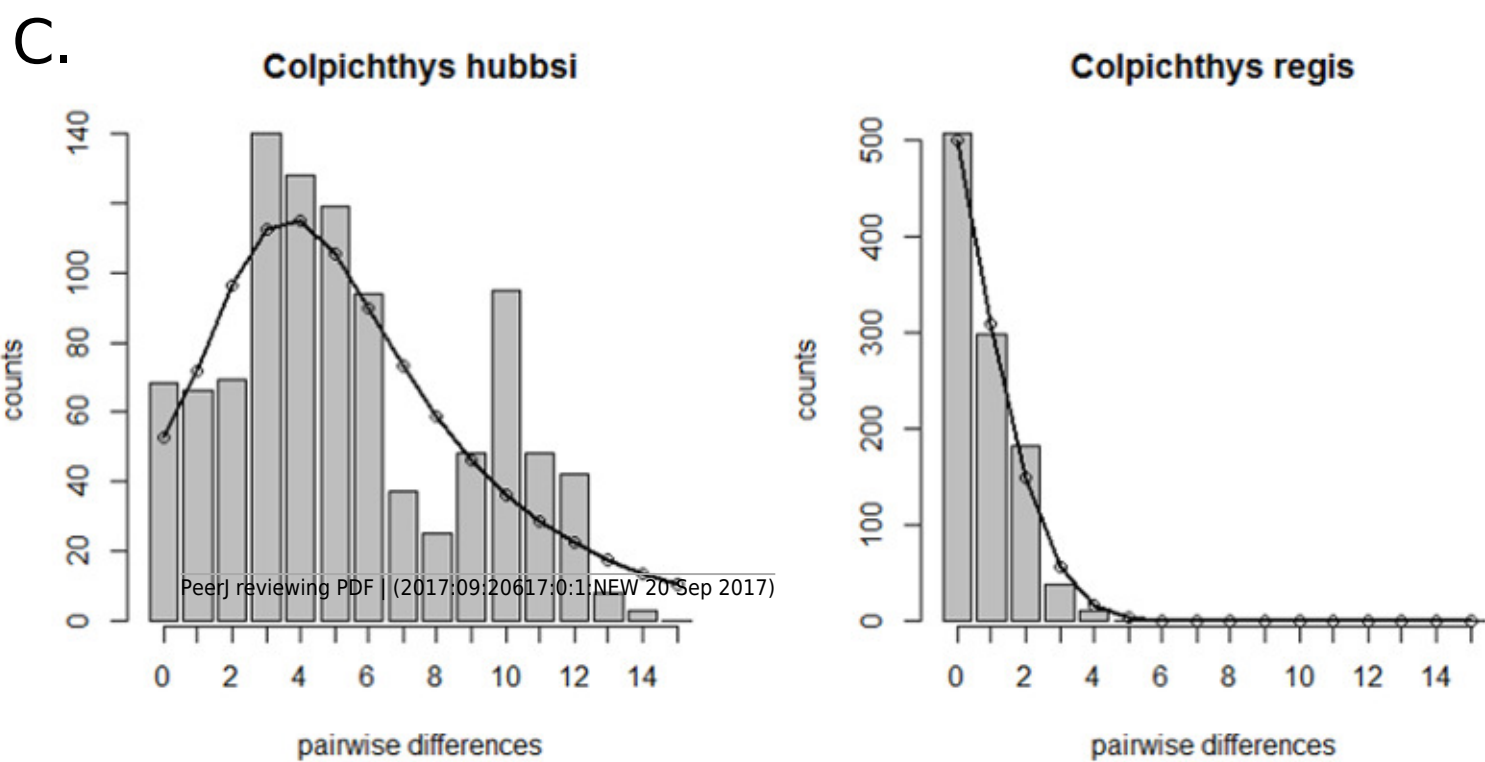


Table 1 (on next page)

Collection localities for the modern specimens.

Table 1 List of collection localities for the modern specimens.

<u>Locality</u>	<u>Code</u>	<u>Region</u>	<u>Sampling Date</u>	<u>Coordinates</u>	<u>N =</u>
Las Animas	ANI	Baja Coast	4 May 2007	28° 47.855N - 113° 20.894W	9
Estero Santa María	ESM	Baja Coast	10 Nov 2005	30° 44.73' N - 114° 42.01W	1
Estero Primero	PRI	Delta Edge	22 Sep 2006	31° 11.903N - 114° 53.437W	10
Estero Segundo	SGU	Delta Edge	24 Sep 2006	31° 15.355N - 114° 53.011W	3
Estero Tercero	TRC	Colorado River Delta	16 Jun 2011	31° 17.354N - 114° 54.831W	3
Estero Chayo	CHY	Colorado River Delta	3 Dec 2005	31° 40.119N - 114° 41.529W	15
“Port Elizabeth”	ELZ	Colorado River Delta	3 Dec 2005	31° 49.405N - 114° 49.566W	18
“Shrimp Farm” El Golfo	SFG/ SFGb	Colorado River Delta	2 Dec 2005/ 15 Jun 2011	31° 46.480N - 114° 34.931W	7 12
Bahía Adair	ADR	Sonora Coast	14 Jun 2011	31° 32.244N - 113° 58.910W	9
Northwest of Cholla	NWC	Sonora Coast	14 Jun 2011	31° 27.822N - 113° 37.898W	8
Estero del Soldado	SLD	Sonora Coast	6 Oct 2006	27°57'26"N - 110°58'48"W	1
Yavaros	YAV	Sonora Coast	5 Mar 2006	26° 40' 42"N - 109° 29' 36"W	7
Total=					103

1 **Table 1** List of collection localities for the modern specimens.

Table 2(on next page)

List of museum specimens

Table 2 List of “historical” specimens from the Natural History Museum of Los Angeles County collection (LACM) and Scripps Institute of Oceanography collection (SIO).

Collection ID	Collection Date	Locality	Locality Code	Coordinates	Species	Feature Examined	N =
SIO 81-158	19 Jan 1973	East of Isla Montague	EIM	31°46.0' N - 114°42.5' W	<i>C. hubbsi</i>	External morphology	2
SIO 81-161	15 Apr 1973	North of Isla Montague	NIM	31°49.0' N - 114°48.5' W	<i>C. hubbsi</i>	External morphology	1
SIO 81-156	17 Dec 1972	West of Isla Montague	WIM	31°44.0' N - 114°48.0' W	<i>C. hubbsi</i>	External morphology	1
SIO 68-46	20 Jan 1968	North of San Felipe	SFE	31°0.0' N - 114°52.0' W	<i>C. hubbsi</i> *	External morphology	16
SIO 60-484	5 Dec 1960	West from Empalme	EMP	27°55.0' N - 110°55.0' W	<i>C. regis</i>	External morphology	3
SIO 63-532	17 Aug 1963	Kino Bay	KIN	28°51.5' N - 112°1.5' W	<i>C. regis</i>	External morphology	1
SIO 63-531	17 Aug 1963	Laguna La Cruz	KNV	28°48.0' N - 111°54.5' W	<i>C. regis</i>	External morphology	1
LACM 35730	23 May 1974	Guaymas	GYM	27° 55' N - 110° 57' W	<i>C. regis</i>	Vertebral morphology	3

* Specimens were classified as *C. hubbsi* on label, but all except one specimen from this locality display the external morphology of *C. regis*. (For details, see Table S1.)

Table 2 List of “historical” specimens from the Natural History Museum of Los Angeles County collection (LACM) and Scripps Institute of Oceanography collection (SIO).

Table 3 (on next page)

RAG1 haplotypes

Table 3 Variants of RAG 1. H_1 , H_2 , H_3 are common in populations of *C. hubbsi* in the Delta; R is fixed in all populations of *C. regis*. SNPs at positions 717 and 909 (bolded) are species-specific and heterozygotes at these positions are used to identify putative hybrids.

<i>RAG1</i> Variants	<i>Position 495</i>	<i>Position 717</i>	<i>Position 770</i>	<i>Position 909</i>
<i>H₁</i>	G	C	G	G
<i>H₂</i>	G	C	T	G
<i>H₃</i>	A	C	T	G
<i>R</i>	G	T	G	A

1 **Table 3** Variants of RAG 1. *H₁*, *H₂*, *H₃* are common in populations of *C. hubbsi* in the Delta; *R* is
2 fixed in all populations of *C. regis*. SNPs at positions 717 and 909 (bolded) are species-specific
3 and heterozygotes at these positions are used to identify putative hybrids.

Table 4(on next page)

Distribution of RAG1 heterozygotes

Table 4 Counts of RAG1 heterozygotes and non-heterozygotes at positions 717 and 909 at each of the Delta and Delta Edge collection sites.

		heterozygotes	non-heterozygotes	het./non-het. ratio
Delta	Port Elizabeth (ELZ)	0	18	0
	Shrimp Farm (SFG)	1	15	0.067
	Estero Chayo (CHY)	1	12	0.083
	Estero Tercero (TRC)	0	3	0
Delta Edge	Estero Segundo (SGU)	1	2	0.500
	Estero Primero (PRI)	3	7	0.429

Table 4 Counts of RAG1 heterozygotes and non-heterozygotes at positions 717 and 909 at each of the Delta and Delta Edge collection sites.

Table 5(on next page)

Neutrality tests of Cytochrome b

Table 5 Neutrality test results. N = number of samples; S = number of segregating sites.

(Bold: $p < 0.05$; *: $p \leq 0.02$; **: $p \leq 0.005$).

	Neutrality Tests						
	N	S	Tajima's D	Fu & Li's D^*	Fu & Li's F^*	Fu's F_s	R_2
<i>C. regis</i>	47	11	-2.1410**	-3.3113*	-3.4502**	-7.1724*	0.0514
<i>C. hubbsi</i>	44	40	-1.5645	-2.2374	-2.3783*	-16.1789**	0.0596

Table 5 Neutrality test results. N = number of samples; S = number of segregating sites. (Bold: $p < 0.05$; *: $p \leq 0.02$; **: $p \leq 0.005$).

Table 6(on next page)

Mismatch analysis of Cytochrome b

Table 6 Mismatch analysis results. SSD = sum of square deviation; CI = confidence interval.

1

	Mismatch Analysis (sudden expansion)				
	SSD	raggedness	θ_0 (99% CI)	θ_1 (99% CI)	τ (99% CI)
<i>C. regis</i>	0.0015	0.0779	0.018 (0 – 0.965)	66842 (0 – 99999)	1.110 (0 - 4.605)
<i>C. hubbsi</i>	0.0092	0.0158	1.332 (0 – 13.558)	17339 (6 – 99999)	5.827 (0.105 – 27.619)

2 **Table 6** Mismatch analysis results. SSD = sum of square deviation; CI = confidence interval.