A peer-reviewed version of this preprint was published in PeerJ on 3 May 2018.

Islam MR, Schmidt DJ, Crook DA, Hughes JM. 2018. Patterns of genetic structuring at the northern limits of the Australian smelt (*Retropinna semoni*) cryptic species complex. PeerJ 6:e4654
https://doi.org/10.7717/peerj.4654

Patterns of genetic structuring at the northern limits of the Australian smelt (*Retropinna semoni*) cryptic species complex

Md Rakeb-Ul Islam Corresp., [1] , **Daniel J Schmidt** [1] , **David A Crook** [2] , **Jane M Hughes** [1]

[1] Griffith University, Australian Rivers Institute, Brisbane, Australia

[2] Charles Darwin University, Research Institute for Environment and Livelihoods, Darwin, NT, Australia

Corresponding Author: Md Rakeb-Ul Islam
Email address: r.islam@griffithuni.edu.au

Freshwater fishes often exhibit high genetic population structure due to the prevalence of dispersal barriers (e.g., waterfalls) whereas population structure in diadromous fishes tends to be weaker and driven by natal homing behaviour and/or isolation by distance. The Australian smelt (Retropinninae: *Retropinna semoni*) is a facultatively diadromous fish with a broad distribution spanning inland and coastal drainages of south-eastern Australia. Previous studies have demonstrated variability in population genetic structure and movement behaviour (potamodromy, facultative diadromy, estuarine residence) across the southern part of its geographic range. Some of this variability may be explained by the existence of multiple cryptic species. Here, we examined genetic structure of populations at the northern extent of the species' distribution, using ten microsatellite loci and sequences of the mitochondrial cyt b gene. We tested the hypothesis that connectivity among rivers should be low due to a lack of dispersal via the marine environment, but high within rivers due to potamodromous behaviour. We investigated populations corresponding with two putative cryptic species, the South East Queensland (SEQ), and Central East Queensland (CEQ) lineages. In agreement with our hypothesis, highly significant overall F_{ST} values suggested that both groups exhibit very low dispersal among rivers (SEQ F_{ST} = 0.13; CEQ F_{ST} = 0.30). The two putative cryptic species, formed monophyletic clades in the mtDNA gene tree and among river phylogeographic structure was also evident within clades. Microsatellite data indicated that connectivity among sites within rivers was also limited, suggesting potamodromous behaviour does not homogenise populations at the within-river scale. Overall, northern groups in the smelt cryptic species exhibit higher among-river population structure and smaller geographic ranges than southern groups. These properties make northern Australian smelt populations potentially susceptible to future conservation threats, and we define eight genetically distinct management units to guide future conservation management.

Patterns of genetic structuring at the northern limits of the Australian smelt (*Retropinna semoni*) cryptic species complex

Md Rakeb-Ul Islam[1], Daniel J Schmidt[1], David A Crook[2], Jane M Hughes[1]

[1]Australian Rivers Institute, Griffith University, Nathan, 4111, QLD, Australia

[2]Research Institute for Environment and Livelihoods, Charles Darwin University, Darwin, NT, Australia

Corresponding author contact information:

Md Rakeb-Ul Islam

Email: r.islam@griffithuni.edu.au

Phone: +61 0470375409

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3284v1 | CC BY 4.0 Open Access | rec: 26 Sep 2017, publ: 26 Sep 2017

ABSTRACT

Freshwater fishes often exhibit high genetic population structure due to the prevalence of dispersal barriers (e.g., waterfalls) whereas population structure in diadromous fishes tends to be weaker and driven by natal homing behaviour and/or isolation by distance. The Australian smelt (Retropinninae: *Retropinna semoni*) is a facultatively diadromous fish with a broad distribution spanning inland and coastal drainages of south-eastern Australia. Previous studies have demonstrated variability in population genetic structure and movement behaviour (potamodromy, facultative diadromy, estuarine residence) across the southern part of its geographic range. Some of this variability may be explained by the existence of multiple cryptic species. Here, we examined genetic structure of populations at the northern extent of the species' distribution, using ten microsatellite loci and sequences of the mitochondrial cyt b gene. We tested the hypothesis that connectivity among rivers should be low due to a lack of dispersal via the marine environment, but high within rivers due to potamodromous behaviour. We investigated populations corresponding with two putative cryptic species, the South East Queensland (SEQ), and Central East Queensland (CEQ) lineages. In agreement with our hypothesis, highly significant overall F_{ST} values suggested that both groups exhibit very low dispersal among rivers (SEQ $F_{ST} = 0.13$; CEQ $F_{ST} = 0.30$). The two putative cryptic species, formed monophyletic clades in the mtDNA gene tree and among river phylogeographic structure was also evident within clades. Microsatellite data indicated that connectivity among sites within rivers was also limited, suggesting potamodromous behaviour does not homogenise populations at the within-river scale. Overall, northern groups in the smelt cryptic species exhibit higher among-river population structure and smaller geographic ranges than southern groups. These properties make northern Australian smelt populations potentially susceptible to future conservation threats, and we define eight genetically distinct management units to guide future conservation management.

Keywords Dispersal, Population structure, Facultative diadromy, Isolation by distance, Cryptic species

INTRODUCTION

Dispersal refers to the exchange of individuals and genes across the geographical range of a species (Wade & McCauley, 1988). Dispersal allows organisms to escape unsuitable environments, avoid competition and maximise fitness in response to changes in the distributions of temporally and spatially patches resources (Haugen et al., 2006). Maintenance of dispersal pathways is important from a conservation perspective, particularly for species whose natural habitat is fragmented by anthropogenic disturbances. It is often the only mechanism by which organisms can move between populations and thus maintain genetically diverse meta-populations (Clobert et. al., 2012). Dispersal between populations may also reduce local extinction rates through a "rescue effect" (Brown & Kodric-Brown, 1977) by reproduction in the populations into which they disperse, and by increasing genetic diversity. Dispersal also plays a major role in the genetic structuring of natural populations (Slatkin 1987; Waters, Dijkstra & Wallis, 2000; Wong, Keogh & McGlashan, 2004). Highly mobile, free swimming species are likely to exhibit minimal phylogeographic structuring across a broad range, especially where there are no physical barriers (Chapco, Kelln & McFayden, 1992; Wong, Keogh & McGlashan, 2004). In contrast, stronger genetic subdivision among populations is predicted for species with limited dispersal abilities.

Genetic structure in aquatic fauna is strongly influenced by the characteristics of the ambient environment. Freshwater species typically exhibit higher levels of genetic differentiation than those living in estuarine or marine habitats (Ward, Woodwark & Skibinski, 1994; Sharma & Hughes, 2009). Movement by obligate freshwater organisms is limited to the water column and the freshwater environment, preventing inter-catchment dispersal via the sea (Burridge et al., 2008; Hughes, Schmidt & Finn, 2009; Bernays et al., 2015). Within freshwater habitats, a range of other factors also restrict dispersal, including natural topographic barriers, such waterfalls and rapids, and artificial dams and weirs (Alp et al., 2012). As a consequence of the physical

limitations to dispersal in freshwater environments, genetic structure of aquatic organisms is often highly genetically differentiated both among and within catchments (McGlashan & Hughes, 2000; Hughes, 2007; Sharma & Hughes, 2009).

Population fragmentation and subsequent genetic differentiation among populations have resulted in a high incidence of cryptic speciation in freshwater habitats (Adams et al., 2013). Cryptic species are defined as morphologically indistinguishable species that are genetically distinct (Knowlton, 1993; Bickford et al., 2007; Thomas et al., 2014). Australia is considered as one of the top 17 megadiverse countries in the world (Williams et al., 2001) reflecting the species richness and levels of endemism exhibited for many organismal groups (Chapman, 2009; Hammer et al., 2014). However, Australia's freshwater fish fauna has long been described as depauperate compared to that found in other regions of similar size and climatic range (Allen, 1989; Lundberg et al., 2000; Allen, Midgley & Allen, 2003; Adams et al., 2013; Hammer et al., 2014). For instance, 209 freshwater-dependent fish species in Australia were recorded in the most recent field guides (Allen, Midgley & Allen, 2003). In contrast, 713 species were found in continental temperate USA (i.e. excluding Alaska and Hawaii; Page & Burr, 1991; Adams et al., 2013). Most researchers have suggested that these differences are the result of the effect of relative differences in aridity, rainfall reliability, topographic diversity, habitat availability and degree of isolation (Merrick & Schmida, 1984; Williams & Allen, 1987; Allen, Midgley & Allen, 2003; Adams et al., 2013). However, Lundberg et al. (2000) proposed a very different explanation for Australia's low number of species and suggests that it reflects the degree of detailed taxonomic effort devoted to this neglected group. Recent assessments, (Hammer, Adams & Hughes 2013; Hammer et al., 2014) have suggested that there may be twice as many fish species in Australia than previously described.

The Australian smelt (Retropinninae: *Retropinna*) is an abundant fish species distributed throughout the rivers of south-eastern Australia (McDowall, 1996). They reach a maximum length of about 100 mm total length (TL), although adults are usually 50-60 mm TL (Pusey, Kennard & Arthington, 2004). Australian smelts are currently recognised as two formally described species *R. semoni* Weber, and *R. tasmanica* McCulloch, but recent genetic analyses have identified a complex of five or more cryptic species across their geographic range based on

allozymes, microsatellites and mitochondrial DNA data (Hammer et al., 2007; Hughes et al., 2014; Schmidt, Islam & Hughes, 2016). Otolith chemistry studies in the southern part of their distribution have shown that Australian smelt exhibit a range of life history patterns, including freshwater residency, facultative diadromy and estuarine residency (Crook, Macdonald & Raadik, 2008; Hughes et al., 2014). In inland regions of Australia, large numbers of Australian smelt have been observed moving upstream through fishways (e.g., Baumgartner & Harris, 2007) and the species is widely described as potamodromous (i.e., migration within freshwater) (e.g., Rolls, 2011). Nonetheless, Woods et al. (2010) found strong genetic structure among inland populations of Australian smelt and suggested low levels of dispersal in at least some populations.

In most studies to date, diadromous behaviour has been shown to facilitate genetic connectivity among river catchments and typically results in "isolation–by-distance" (IBD) patterns of population genetic structure (Keenan, 1994; Jerry & Baverstock, 1998). In Australian smelt, however, there is strong genetic differentiation among catchments across the southern part of the range - even among populations containing diadromous individuals suggesting high retention of fish within estuaries and a lack of marine dispersal (Hughes et al., 2014). The aim of the current study was to examine patterns of genetic connectivity of populations in the north of the geographic range of Australian smelt, which have not previously been characterised. In light of this, sequence data from mtDNA cytochrome b combined with genotypic data from 10 microsatellite loci were used to test the hypotheses that, i) northern *R. semoni* would display high population structure among rivers similar to southern populations; and ii) that genetic structure within rivers would be low due to potamodromous migration.

MATERIALS AND METHODS

Sampling strategy

A total of 391 individual samples were collected from 15 locations in south-east Queensland, Australia (Fig. 1; Table 1). Samples were collected using a hand - held seine net from an upstream and a downstream site from each river except the Noosa River (downstream only). Where possible, we aimed to collect at least 30 individuals per site. Fin clips or entire individuals were placed in 95% ethanol in the field and stored prior to preparation for analysis.

Molecular methods

Genomic DNA was extracted from fin tissue using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's directions. Microsatellite markers developed for *R. semoni* were amplified and genotyped using primers developed by Islam, Schmidt & Hughes (2017). Ten loci were screened across all individuals. The ten loci were BS18, BS3, BS4, BS5, BS20, BS21, BS22, BS24, BS8 and MS24. All subsequent microsatellite screening was carried out in 10 µl PCR reactions consisting of 0.5 µl of genomic DNA, 0.2 mM reverse primer, 0.05 mM tailed forward primer, 0.2 mM tailed fluorescent tag (either FAM, VIC, NED or PET, Applied Biosystems), 1× PCR buffer (Astral Scientific) and 0.02 units of *taq* polymerase (Astral Scientific). The following basic thermocycler settings for the polymerase chain reaction (PCR) were performed: initial denaturation at 94^0C for 4 min, followed by 35 cycles at 94^0C for 1 min, 57^0C for 30 s, 72^0C for 1 min and a final extension at 72^0C for 7 min. Fluorescently labelled amplified PCR products were pooled and added to 10 µl of Hi-Di™formamide with 0.1 µl of GeneScan™ 500 LIZ size standard. Fragment analysis was conducted on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions. Data were scored using GENEMAPPER version 3.1 software (Applied Biosystems).

Two individuals from each of the 15 populations represented in the microsatellite study were randomly selected for mtDNA analysis. Samples from four additional sites not included in microsatellite analysis were also sequenced – two from Mary River (Booloumba Creek, MBC, 26°41'02.5"S 152°37'10.6"E, n = 9; Yabba creek, MYC, 26°28'09.3"S 152°38'39.5"E, n = 8) and two from the Brisbane River (Bundamba creek, BDC, 27°36'03.9"S 152°48'04.2"E, n = 10; Banks creek, BSV, 27°26'36.9"S 152°40'13.2"E, n = 10) and sample collecting site were also shown in Fig. 1. In total 68 individuals from 19 sites were sequenced. A 666 bp fragment of the cytochrome b region of the mtDNA genome was selected for sequencing analysis. The primers HYPSLA and HYPSHD (Thacker et al., 2007) were used to amplify the region in 10 µL reaction mixtures. PCR conditions were 4 min at 95 ^{0}C, followed by 45 cycles of 30 s at 95 ^{0}C, 45 s at 53 ^{0}C, 45 s at 72 ^{0}C and a final extension cycle of 7 min at 72 ^{0}C. MtDNA sequences were edited and aligned using Geneious version 9.1.5 (Kearse et al., 2012).

Data Analysis

Genetic diversity

Microsatellite genotype frequencies were checked for the presence of null alleles, large allele dropout and stuttering artefacts using Micro-checker v2.2.3 (Van Oosterhout et al., 2004). Tests for linkage disequilibrium (LD) and departures of genotypic proportions expected under Hardy-Weinberg Equilibrium (HWE) were calculated with exact tests for each population and over all loci using default settings in GENEPOP v4 (Rousset, 2008). Probability values were corrected using standard Bonferroni correction (Rice, 1989) whenever multiple testing was performed. Genetic diversity averaged across ten loci within each of the fifteen population samples was calculated from observed and expected heterozygosity using ARLEQUIN v3.5.1.2 (Excoffier & Lischer, 2010). Measures of genetic diversity standardized for sample size including Allelic richness (AR) and private allelic richness (AR_{priv}) were estimated using HP-RARE 1.1 (Kalinowski, 2005). Inbreeding index (F_{IS}) was estimated in FSTAT 2.9.3 (Goudet, 2001).

Population genetic structure

Genetic structure among the 15 populations was quantified by estimating pairwise and global F_{ST} values in ARLEQUIN. These were tested for significant deviation from panmictic expectations by 10,000 permutations of individuals among populations. Population-specific F_{ST} values were calculated using GESTE v2.0 (Foll & Gaggiotti, 2006) to evaluate the contribution of individual population samples to overall F_{ST}.

ARLEQUIN v3.5.1.2 (Excoffier & Lischer, 2010) was used to evaluate the geographic structuring of genetic variation. F_{ST} was calculated for each locus separately and as a weighted average over the ten micosatellite loci. Statistical significance of F_{ST} was determined by 1000 permutations of individuals among populations. Hierarchical structuring of variation was calculated using AMOVA in ARLEQUIN v3.5.1.2 (Excoffier & Lischer, 2010). Two hierarchical arrangements of the 15 populations were analysed where the highest level was either a) two groups, CEQ group (MRD, MRU, NSD, MLD, MLU) and SEQ group (BRD, BRU, LGD, LGU, CMD, CMU, NRD, NRU, CRD, CRU) or b) catchment division, site grouped into 8 rivers according to the connectivity of streams to the upper river reaches. These were: Mary (MRD, MRU), Noosa (NSD), Mooloolah (MLD, MLU), Brisbane (BRD, BRU), Logan (LGD, LGU),

Coomera (CMD, CMU), Nerang (NRD, NRU) and Currumbin (CMD, CMU). Three hierarchical levels of variation were analysed for each arrangement: among groups (F_{CT}), among sites within groups (F_{SC}) and within sites.

Bayesian clustering methods implemented in STRUCTURE v.2.3.1 (Pritchard, Stephens & Donnelly, 2000) were applied to estimate the number of genetically homogeneous clusters (Latch et al., 2006; Hasselman, Ricard & Bentzen, 2013). This programme builds genetic clusters by minimizing linkage disequilibrium and deviations from Hardy-Weinberg equilibrium expectations within clusters. All individuals were assigned to clusters without prior knowledge of their geographic origin using the admixture model with correlated allelic frequencies. Ten independent runs with the number of potential genetic clusters (K) from 1 to 15 were carried out to verify that the estimates of K were consistent across runs. The burn-in length was set at 250,000 iterations followed by a run phase of one million iterations. The generated results were imported into the software STRUCTURE HARVESTER (Earl & vonHoldt, 2012) to calculate the *ad hoc* ΔK statistic (Evanno, Regnaut & Goudet, 2005). The K value, where ΔK had the highest value was identified as the most likely number of clusters.

Analysis of isolation by distance

A test for a positive association between genetic and geographic distances [Isolation by distance (IBD)] based on microsatellite DNA loci was carried out using a Mantel test (10000 permutations) in Arlequin v3.5.2 (Excoffier & Lischer, 2010). Genetic distance was represented as F_{ST}. Stream distances were calculated between river mouths and then sample sites using Google Earth.

Migration and gene flow

BAYESASS v1.3 was used to calculate contemporary migration rates over the past few generations where m_{ji} is the proportion of immigrants in a focal population *i* that arrive from a source population *j* (Wilson & Rannala, 2003). This Bayesian assignment method follows the rule that immigrants and their progeny represent temporary disequilibrium in their microsatellite genotypes relative to the focal population under the assumption that background migration is

comparatively low ($F_{ST} > 0.05$) and that loci are in linkage equilibrium (Faubet, Waples & Gaggiotti, 2007). Analyses were run for 3×10^7 iterations, sampling every 2000 iterations with discarded burn-in of 10^7. Delta values were adjusted to 0.12 to ensure that chain swapping occurred in about 50% of the total iterations as suggested by Wilson & Rannala (2003) and to estimate the accuracy of the results the analysis was repeated three times with different random number of seeds. We also used the Bayesian assignment procedure of Rannala & Mountain (1997), as implemented in GENECLASS 2 (Piry et al., 2004) to estimate whether our samples might contain individuals that were first generation (F_0) immigrants from unsampled populations. Here we used Paetkau et al. (2004) method to compute probabilities from 10,000 simulated genotypes to identify F_0 immigrants.

Analysis of mtDNA sequence data

A Neighbour - joining (NJ) tree analysis was performed using the HKY distance model in Geneious version 9.1.5 with 1000 bootstrap replicates. In addition to the 68 sequences generated from this study, two Genbank accessions were used, one representing *R. tasmanica*: JN232589; and one representing *R. semoni*: JN232588 (Burridge et al., 2012). The *R. semoni* sequence JN232588 lacks locality information but likely belongs to a southern lineage of *R. semoni* which are known to have a closer mtDNA relationship with *R. tasmanica* than to northern lineages (Hughes et al., 2015).

RESULTS

Genetic variability and levels of differentiation

After Bonferroni correction, 3 out of 15 populations exhibited deviations from HWE in only two or three loci. All loci were kept for further analyses since deviations were not consistent across populations. Instances of null alleles estimated using MICRO-CHECKER were rare and not consistently associated with specific loci or populations. We observed little evidence for genotypic linkage disequilibrium between any pair of loci. Among 645 pairwise comparisons, 15 were significant at the $P<0.05$ level after Bonferroni correction. These significant cases were randomly distributed among populations and pairs of loci.

Population genetic diversity indices are shown in Table 1. Microsatellite genetic diversity was high. Mean number of alleles per population ranged from 4.60 (MLD) to 14.70 (CMU). Heterozygosity averaged across loci ranged from 0.566(MLU) to 0.887 (CMD and CMU) and allelic richness averaged across loci ranged from 3.41 to 7.42 when sample sizes were standardized across populations at 6 individuals. Although private alleles were found in all sites, the MRU population had the highest private allelic richness. Most sites exhibited positive F_{IS} values, indicating that most of the populations had slight heterozygote deficit.

Most of the pairwise F_{ST} values between the 15 populations were significant and ranged from -0.018 to 0.404. The CEQ populations were more diverged from one another than the populations in the SEQ group. The lowest pairwise F_{ST} value (F_{ST} = -0.018; $P < 0.05$) was observed between populations NRD and NRU. The highest genetic divergence (F_{ST} = 0.404; $P < 0.05$) was observed between populations NSD and MLU. Out of 105 comparisons, only six comparisons were non-significant ($P > 0.05$) and each of these pairs was from within the same river (Mooloolah; Brisbane; Logan; Coomera; Nerang and Currumbin). Generally F_{ST} comparisons revealed much less divergence among populations within the same river than between populations from different rivers (Table 2).

The STRUCTURE analysis incorporating all individuals suggested that initially the most likely number of clusters was two, one containing all CEQ populations and the other containing all SEQ populations (Fig. 2A). The SEQ group was then further subdivided into two separate groups leaving the Brisbane river populations distinct from all others and remaining populations of SEQ group comprising four distinct clusters (Fig 2B and C). The CEQ group further subdivided into three distinct clusters (Fig. 2C). STRUCTURE analysis revealed that the highest likelihood at K= 8 clusters (Average log probability of data Ln[P(DK)] = -15246.1 ± 1.028753) indicating this as the best estimate of the true number of the genetic clusters. The height of ΔK was used as an indicator of the strength of the signal detected by STRUCTURE (Evanno, Regnaut & Goudet, 2005). ΔK showed the highest peak at K = 8, suggesting eight genetically homogeneous clusters across the sampled populations and negligible immigrations among rivers (Fig. 2C).

Strong population structure was supported by AMOVA with 20.50 % genetic variation exhibited by differences among populations (Table 3A). The AMOVA showed significant genetic

differentiation between the two groups (CEQ and SEQ) (F_{CT} = 0.05), but also among populations within groups (F_{SC} = 0.18) (Table 3B). There were similar pattern between the groups when they were analysed separately, with the F_{CT} value (among rivers) higher than the F_{SC} value (among sites within rivers) in both groups (Table. 3C i and ii). However, the overall F_{ST} values, and each of the other *F* statistics in the hierarchy were higher in the CEQ group than the SEQ group.

Isolation-by-distance

There was a significant correlation between genetic differentiation and stream distance among populations from the SEQ group (R^2 = 0.3687, p = 0.001; BRD, BRU, LGD, LGU, CMD, CMU, NRD, NRU, CRD, and CRU) (Fig. 3A), but not for CEQ group (R^2 = 0.0355, p = 0.302; MRD, NSD, MRU, MLD and MLU) (Fig. 3B).

Contemporary migration

Very little contemporary migration was observed among the coastal river populations. Only six sampled populations contained individuals that were identified as potential immigrants from the BAYESASS analysis. In all cases, the putative source population was the paired site within the same catchment. An average of 19.2 % of individuals at each of the six locations was estimated to be immigrants (range 10 – 33 %, Table 4). In five out of six cases, dispersal was from the upstream to the downstream site. Only individuals from Currumbin creek was estimated to have dispersed in an upstream direction. The highest level of migration was also found in this creek (23%). Only fifteen (< 4%) of 391 individuals across all sites were identified as F_0 migrants using the "detection of first generation migrants" option in GENECLASS2 (Table 5).

MtDNA sequences analysis

The edited alignment for the cyt b gene was 575 bp and included 121 variable positions. All sequences are lodged under GenBank accession numbers XXXXXXX-XXXXXXX. The neighbour - joining tree revealed two strongly supported clades (bootstrap 89% SEQ; 96% CEQ; Fig. 4). Phylogeographic structure was also clearly evident within clades. All individuals from four sites in the Brisbane River formed a distinct clade, and all three rivers sampled for the CEQ lineage formed shallow clades (i.e. Mary, Noosa and Mooloolah rivers; Fig. 4). Genetic distance

was high between northern smelt lineages and the southern smelt sequences used as an outgroup (uncorrected mean nucleotide distance 0.15 - 0.17). The mean nucleotide distance between two northern lineages SEQ and CEQ was 0.04 (SE = 0.007).

DISCUSSION

Population structure and dispersal

Based on previous studies of Australian smelt in south-eastern Australia using mtDNA and microsatellites (Woods et al., 2010; Hughes et al., 2014), we had hypothesized that *R. semoni* in the northern part of their distribution would exhibit limited genetic connectivity among river systems due to a lack of marine dispersal: either because they are non-diadromous or because they are diadromous, but are retained within their natal estuaries (see Hughes et al., 2014). Our findings of strong genetic differentiation among rivers support this hypothesis. In both of the regions (CEQ and SEQ) sampled, there were highly significant F_{ST} values, which indicated that populations were not panmictic within regions. Pairwise F_{ST} values between populations within regions also revealed significant genetic differentiation, suggesting restricted gene flow and limited dispersal among populations of *R. semoni* in both regions. Limited dispersal was supported by our first-generation migrant detection analysis in Geneclass2, which demonstrated that less than 4% of individuals in each population were immigrants.

The sample from Tinana Creek (MRD site), was differentiated from the rest of the populations in the CEQ group (Table 2). This might be the result of a barrier which separates Tinana Creek from the rest of the Mary river system despite their close proximity to one another (Hughes et al., 2015). Tinana Creek runs into the Mary River not far from the mouth, with both drainages having tidal estuarine reaches in the lower sections. The differentiation of the Tinana Creek population from the main stem of the Mary River is also observed in a number of other freshwater species including Mary River Cod, *Maccullochella Mariensis* (Huey, Espinoza & Hughes, 2013), Mary River Turtle, *Elusor macrurus* (Schmidt et al., in press), freshwater crayfish *Cherax disper* (Bentley, Schmidt & Hughes, 2010) and Australian lung fish *Neoceratodus fosteri* (Hughes et al., 2015).

In general, populations in the CEQ group were more highly structured than those in the SEQ group, but fishes in both groups exhibited restricted gene flow. These differences could have several explanations. First, obligate freshwater fish are expected to display greater levels of genetic differentiation and population subdivision than marine species due to the isolating nature of river systems and small effective population size (Ward, Woodwark & Skibinski, 1994; Gyllensten, 1985; McGlashan & Hughes, 2001). The degree of genetic differentiation among populations between drainages was consistent with these expectations, although effective population size is unlikely to be very low, given the high levels of diversity. Another plausible reason is that eustasy may affect the genetic structure of populations through the irregular joining and isolation of drainages along the coastal margin. Long term isolation of populations in separate drainages may lead to extensive genetic differentiation among drainages. Particularly the high genetic structuring might have resulted from limited spatial dispersal patterns of larvae. In addition to genetic drift in pools, genetic differentiation may arise as a result of local extinction/recolonization dynamics because some pools dry out completely during dry seasons and their colonization by a limited number of individuals can result in genetic differentiation due to founder effect (Vrijenhoek, 1979; Vrijenhoek & Lerman, 1982; Barr et al., 2008; Tatarenkov, Healey & Avise, 2010).

An alternative model for stream dwelling species is isolation by distance (IBD). In this model, equilibrium between genetic drift and gene flow may be reached in species where the life time dispersal distance is less than the range. Here, a relationship between stream distance and genetic differentiation should be evident (Wright, 1943). In this study, a strong IBD relationship was identified among the SEQ populations, but not among CEQ populations. This suggests that for SEQ populations, dispersal, when it occurs, is more likely between nearby catchments. Similar IBD relationships have been reported for other coastline restricted species (Keenan, 1994; Jerry & Baverstock, 1998; Shaddick et al., 2011; Schmidt et al., 2014). Lack of IBD for the CEQ group may be attributed to insufficient number of population samples available for comparison and/or the greater degree of population isolation within this group relative to the SEQ group, consistent with the overall higher F_{ST} estimates among CEQ populations. Hughes et al. (2014) observed similarly contrasting patterns of population genetic structure between cryptic species groups of southern Australian smelt. In that study, two informal species groups (MTV

and SEC) with adjacent distributions along the western and eastern coast of southern Victoria had microsatellite-based F_{ST} values of 0.19 and 0.07 respectively (Hughes et al. 2014). Using otolith microchemistry, Hughes et al. (2014) also showed that the more structured western group (MTV) had a greater proportion of nondiadromous populations relative to the weaker structured eastern group (SEC). The similar pattern of contrasting structure observed here between northern groups in the Australian smelt complex (SEQ, CEQ), is probably not due to differences in diadromous behaviour because preliminary evidence from otolith chemistry suggests all of these populations are nondiadromous (R. Islam unpublished data). Higher structuring of the CEQ group could possibly be due to genetic drift if these populations have been established for a longer period of time at the northern-most limit of Australian smelt distribution relative to the SEQ populations.

The complementary pattern of divergence in both microsatellite and mtDNA data between the SEQ and CEQ groups agrees with a putative species-level boundary identified by Hammer et al. (2007) within the taxon currently referred to as *R. semoni*. Mean cyt b divergence of 4% between SEQ and CEQ samples is close to the 3.6% divergence observed for the full mitochondrial molecule reported by Schmidt et al. (2016), and within the range of lineage divergence reported for *R. semoni* in southern Queensland (Page & Hughes, 2014). The level of cyt b divergence between the SEQ and CEQ groups relative to lineages of *R. semoni* from the south of this' species range is very large (15-17%) and adds to previous studies that have highlighted the likely existence of a cryptic species complex within the taxon currently referred to as *R. semoni* (Hammer et al. 2007; Hughes et al. 2014).

Contemporary migration

The Bayesian assignment analysis detected contemporary movement of individuals only between proximate sites within rivers (Table 4). Contemporary dispersal was not observed between rivers. Although, most of the sites that we sampled were within 10-60 km of another sampled site, there was no contemporary dispersal among the majority of those rivers in either group. In addition, this species appears to occur in pools, many of which are isolated from other pools by long stretches of unfavourable habitat. Our data therefore suggest that if local extinctions occur in one or more of these pools within a reach of the river, then recolonization from elsewhere is unlikely

to occur rapidly. However, the evidence of some localised movement within rivers among local populations suggesting potamodromous migration within rivers. This type of migration of Australian smelt was also reported in previous studies where large number of Australian smelt was found to exhibit potamodromous migrations through fishways in perennial lowland rivers (Mallen-Cooper et al., 1995). This contrast with the findings of the southern smelt migration behaviour where contemporary movement among populations is restricted at least to some extent within the catchment (Woods et al., 2010) although this southern smelt exhibited facultative diadromous migration (Hughes et al., 2014).

CONCLUSION

Little conservation attention has been given to the Australian smelt since it has long been considered a common species distributed widely across south-eastern Australia. The findings of the present study and other recent research (Hammer et al., 2007; Crook, Macdonald & Raadik, 2008; Hughes et al., 2014) suggest that Australian smelts are a genetically complex and ecologically diverse taxonomic group. Therefore, proper conservation and management will require appropriate taxonomic treatment to align species names with the clear genetic divisions now recognised across the range of Australian smelt.

In the present study, two major genetic lineages were recognized that are geographically concordant with distinct allozyme groups reported by Hammer et al. (2007) and these lineages can be categorised as Evolutionary Significant Units (ESU) (Moritz, 1994; Bernatchez, 1995; Crandall et al., 2000; Sasaki et al., 2016). The broad genetic divergence implies that these lineages have evolved independently from each other for some time. For long term management the delimitation of ESUs is imperative where conservation strategy should be specified accurately (Moritz 1994; Sasaki et al., 2016). However, in the present study translocation of individuals between lineages is not recommended for short –term management as it may preclude any local adaptation due to mixing of distinct lineages (Tallmon, Luikart & Waples, 2004; Hughes et al., 2015).

Alternatively, eight isolated management units (MUs) were detected in *R. semoni* from the microsatellite dataset (Fig. 2C) demonstrating little to no gene flow between them. These

management units align with individual coastal catchment, which suggests that other genetically distinct populations may exist in coastal rivers not sampled in this study.

ACKNOWLEDGMENTS

We thank Leo Lee, Nathan McIntyre, Dale Bryan Brown, Juan Tao for help in collecting samples. Thanks to Kathryn Real and Jemma Somerville for assistance with genetic laboratory work.

REFERENCES

Adams M, Page TJ, Hurwood DA, Hughes JM. 2013. A molecular assessment of species boundaries and phylogenetic affinities in Mogurnda (Eleotridae): a case study of cryptic biodiversity in the Australian freshwater fishes. *Marine and Freshwater Research* **64:**920-931 DOI 10.1071/MF12237.

Allen GR, Midgley SH, Allen M. 2003. *Field Guide to Freshwater Fishes of Australia*. Western Australian Museum, Perth.

Allen GR. 1989. *Freshwater Fishes of Australia*. T. F. H. Publications: Neptune City, NJ.

Alp M, Keller I, Westram AM, Robinson CT. 2012. How river structure and biological traits influence gene flow: a population genetic study of two stream invertebrates with differing dispersal abilities. *Freshwater Biology* **57**:969–981 DOI 10.1111/j.1365-2427.2012.02758.x.

Barr KR, Lindsay DL, Athrey G, Lance RF, Hayden TJ, Tweddale SA, Leberg PL. 2008. Population structure in an endangered songbird: maintenance of genetic differentiation despite high vagility and significant population recovery. *Molecular Ecology* **17**:3628-3639 DOI 10.1111/j.1365-294X.2008.03868.x.

Baumgartner LJ, Harris JH. 2007. Passage of non-salmonid fish through a Deelder lock on a lowland river. *River Research and Applications* 23(10):1058-1069 DOI 10.1002/rra.1032.

480 Bentley AI, Schmidt DJ, Hughes JM. 2010. Extensive intraspecific genetic diversity of a
481 freshwater crayfish in a biodiversity hotspot. *Freshwater Biology* **55**:1861-1873 DOI
482 10.1111/j.1365-2427.2010.02420.x.
483 Bernatchez L. 1995. A role for molecular systematic in defining evolutionary significant units in
484 fishes. *American Fisheries Society Symposium* **17**:114-132.
485 Bernays SJ, Schmidt DJ, Hurwood DA, Hughes JM. 2015. Phylogeography of two freshwater
486 prawn species from far-northern Queensland. *Marine and Freshwater Research* **66**:256-
487 266 DOI 10.1071/MF14124.
488 Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007.
489 Cryptic species as a window on diversity and conservation. *Trends in Ecology and*
490 *Evolution* **22**(3):148-145 DOI 10.1016/j.tree.2006.11.004.
491 Brown JH, Kodric-Brown A. 1977. Turnover rates in insular biogeography: effect of
492 immigration on extinction. *Ecology* **58**:445-449 DOI 10.2307/1935620.
493 Burridge CP, Craw D, Jack DC, King TM, Waters JM. 2008. Does fish ecology predict dispersal
494 across a river drainage divide? *Evolution* **62**:1484–1499 DOI 10.1111/j.1558-
495 5646.2008.00377.x.
496 Burridge CP, McDowall RM, Craw D, Wilson MVH, Waters JM. 2012. Marine dispersal as a
497 pre-requisite for Gondwanan vicariance among elements of the galaxiid fish fauna.
498 *Journal of Biogeography* **39**:306-321 DOI 10.1111/j.1365-2699.2011.02600.x.
499 Chapco W, Kelln RA, McFayden DA. 1992. Intraspecific mitochondrial DNA variation in the
500 migratory grasshopper, *Melanoplus sanguinipes*. *Heredity* **69**:547-557 DOI
501 10.1038/hdy.1992.170.
502 Chapman AD. 2009. *Number of living species in Australia and the world*. Canberra: Australian
503 Biological Resources Study (ABRS).
504 Clobert J, Baguette M, Benton TG, Bullock JM. 2012. *Dispersal Ecology and Evolution*. Oxford
505 University Press, Oxford, UK.
506 Crandall KL, Bininda-Emonds ORP, Mace GM, Wayne RK. 2000. Considering evolutionary
507 processes in conservation biology. *Trends in Ecology & Evolution* **15**: 290-295 DOI
508 10.1016/S0169-5347(00)01876-0.

509 Crook DA, Macdonald JI, Raadik TA. 2008. Evidence of diadromous movements in a coastal
510 population of southern smelts (Retropinninae: Retropinna) from Victoria, Australia.
511 *Marine and Freshwater Research* **59**:638-646 DOI 10.1071/MF07238.
512 Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for
513 visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
514 *Genetic Resources* **4**:359-361 DOI 10.1007/s12686-011-9548-7.
515 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the
516 software STRUCTURE: a simulation study. *Molecular Ecology* **14**:2611-2620 DOI
517 10.1111/j.1365-294X.2005.02553.x.
518 Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform
519 population genetics analyses under Linux and Windows. *Molecular Ecology Resources*
520 **10**(3):564-567 DOI 10.1111/j.1755-0998.2010.02847.x.
521 Faubet P, Waples RS, Gaggiotti OE. 2007. Evaluating the performance of a multilocus Bayesian
522 method for the estimation of migration rates. *Molecular Ecology* **16**:1149–1166 DOI
523 10.1111/j.1365-294X.2007.03218.x.
524 Foll M, Gaggiotti O. 2006. Indentifying the environmental factors that determine the genetic
525 structure of populations. *Genetics* **174**:875-891 DOI 10.1534/genetics.106.059451.
526 Goudet J. 2001. FSTAT, version 2.9. 3, A program to estimate and test gene diversities and
527 fixation indices. Lausanne University, Lausanne, Switzerland.
528 Gyllensten U. 1985. The genetic structure of fish: differences in the intraspecific distribution of
529 biochemical genetic variation between marine, anadromous and freshwater species.
530 *Journal of Fish Biology* **26**:691-699 DOI 10.1111/j.1095-8649.1985.tb04309.x.
531 Hammer MP, Adams M, Hughes JM. 2013. Evolutionary processes and biodiversity. In: Walker
532 K, Humphreys P, eds. *Ecology of Australian Freshwater Fishes*. CSIRO Publishing:
533 Melbourne, 49-79.
534 Hammer MP, Adams M, Unmack PJ, Walker KF. 2007. A rethink on *Retropinna*: conservation
535 implications of new taxa and significant genetic sub-structure in Australian smelts
536 (Pisces: Retropinnidae). *Marine and Freshwater Research* **58**:327-341 DOI
537 10.1071/MF05258.
538 Hammer MP, Unmack PJ, Adams M, Raadik TA, Johnson JB. 2014. A multigene molecular
539 assessment of cryptic biodiversity in the iconic freshwater blackfishes (Teleostei:

Percichthyidae: *Gadopsis*) of south-eastern Australia. *Biological Journal of the Linnean Society* **111**:521-540 DOI 10.1111/bij.12222.

Hasselman DJ, Ricard D, Bentzen P. 2013. Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular Ecology* **22**:1558-1573 DOI 10.1111/mec.12197.

Haugen TO, Winfield IJ, Vøllestad LA, Fletcher JM, James JB, Stenseth NC. 2006. The ideal free pike: 50 years of fitness-maximizing dispersal in Windermere. *Proceedings of the Royal Society B: Biological Sciences* 273(1604):2917-2924 DOI 10.1098/rspb.2006.3659.

Huey JA, Espinoza T, Hughes JM. 2013. Natural and anthropogenic drivers of genetic structure and low genetic variation in the endangered freshwater cod, *Maccullochella Mariensis*. *Conservation Genetics* **14**:997-1008 DOI 10.1007/s10592-013-0490-y.

Hughes JM, Schmidt DJ, Finn DS. 2009. Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. *BioScience* **59**(7): 573-583 DOI 10.1525/bio.2009.59.7.8.

Hughes JM, Schmidt DJ, Huey JA, Real KM, Espinoza T, McDougall A, Kind PK, Brooks S, Roberts DT. 2015. Extremely low microsatellite diversity but distinct population structure in a long-lived threatened species, the Australian lungfish *Neoceratodus fosteri* (Dipnoi). *PLoS ONE* **10**(4):e0121858 DOI 10.1371/journal.pone.0121858.

Hughes JM, Schmidt DJ, Macdonald JI, Huey JA, Crook DA. 2014. Low interbasin connectivity in a facultatively diadromous fish: evidence from genetics and otolith chemistry. *Molecular Ecology* **23**:1000-1013 DOI 10.1111/mec.12661

Hughes JM. 2007. Constraints on recovery: using molecular methods to study connectivity of aquatic biota in rivers and streams. *Freshwater Biology* **52** (4):616-631 DOI 10.1111/j.1365-2427.2006.01722.x.

Islam MR-U, Schmidt DJ, Hughes JM. 2017. Development and characterization of 21 novel microsatellite markers for the Australian smelt *Retropinna semoni* (Weber, 1895). *Journal of Applied Ichthyology* **33**:824–828 DOI https://doi.org/10.1111/jai.13391.

Jerry DR, Baverstock PR. 1998. Consequences of a catadromous life-strategy for levels of mitochondrial DNA differentiation among populations of the Australian bass, *Macquaria*

novemaculeata. *Molecular Ecology* **7**:1003-1013 DOI 10.1046/j.1365-294x.1998.00418.x.

Kalinowski ST. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5**:187-189. DOI 10.1111/j.1471-8286.2004.00845.x.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.

Keenan CP. 1994. Recent evolution of population structure in Australian barramundi, *Lates calcarifer* (Bloch): an example of isolation by distance in one dimension. *Australian Journal of Marine and Freshwater Research* **45**:1123-1148 DOI 10.1071/MF9941123.

Knowlton N. 1993. Sibling species in the sea. *Annual Review of Ecology and Systematics* **24**: 189-216 DOI 10.1146/annurev.es.24.110193.001201.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE. 2006. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation genetics* **7**:295-302 DOI 10.1007/s10592-005-9098-1.

Lundberg JG, Kottelat M, Smith GR, Stiassny MLJ, Gill AC. 2000. So many fishes, so little time: an overview of recent ichthyological discovery in continental waters. *Annals of the Missouri Botanical Garden* **87**:26–62 DOI 10.2307/2666207.

Mallen-Cooper M., Stuart IG, Hides-Pearson F, Harris JH. 1995. *Migration in the Murray River and assessment of the Torrumbarry Fishway*. NSW Fisheries and CRC for Freshwater Ecology, Cronulla, NSW, Australia.

McDowall RM. 1996. Family Retropinnidae: southern smelts. In: McDowall RM, ed. *Freshwater fishes of South-eastern Australia*. Reed books, Chatswood, Sydney, 92-95.

McGlashan DJ, Hughes JM. 2000. Reconciling patterns of genetic variation with stream structure, earth history and biology in the Australian freshwater fish *Craterocephalusster cusmuscarum* (Athernidae). *Molecular Ecology* **9**:1737-1751 DOI 10.1046/j.1365-294x.2000.01054.x.

McGlashan DJ, Hughes JM. 2001. Low levels of genetic differentiation among populations of the freshwater fish Hypseleotris compressa (Gobiidae: Eleotridinae): implications for its biology, population connectivity and history. *Heretidy* **86**(2):222-233 DOI 10.1046/j.1365-2540.2001.00824.x.

Merrick JR, Schmida GE. 1984. *Australian Freshwater Fishes, Biology and Management*. Griffin Press, Adelaide.

Milton DA, Arthington AH. 1985. Reproductive strategy and growth of the Australian smelt, *Retropinna semoni* (Weber) (Pisces: Retropinnidae), and the olive perchlet, *Ambassis nigripinnis* (de Vis) (Pisces: Ambassidae), in Brisbane, south-eastern Queensland. *Australian Journal of Marine and Freshwater Research* **36**:329-341 DOI 10.1071/MF9850329.

Moritz C. 1994. Defining 'Evolutionary Significant Units' for conservation. *Tree* **9**:373-375 DOI 10.1016/0169-5347(94)90057-4.

Paetkau D, Slade R, Burden M, Estoup A. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation –based exploration of accuracy and power. *Molecular Ecology* **13**:55-65 DOI 10.1046/j.1365-294X.2004.02008.x.

Page LM, Burr BM. 1991. *A Field Guide to Freshwater Fishes (North America north of Mexico)*. Houghton Mifflin Company: Boston, MA.

Page TJ, Hughes JM. 2014. Contrasting insights provided by single and multispecies data in a regional comparative phylogeographic study. *Biological Journal of the Linnean Society* 111:554-569 DOI 10.1111/bij.12231.

Piry S, Alapetite A, Cornuet J-M, Peatkau D, Baudouin L, Estoup A. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**:536-539 DOI 10.1093/jhered/esh074.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945-959.

Pusey BJ, Kennard M, Arthington A. 2004. *Freshwater fishes of north-eastern Australia*. CSIRO Publishing, Collingwood, VIC.

Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences* **94**:9197-9201.

Rice WR. 1989 Analyzing tables of statistical tests. *Evolution* **43**:223-225 DOI 10.2307/2409177.

Rolls RJ. 2011. The role of life-history and location of barriers to migration in the spatial distribution and conservation of fish assemblages in a coastal river system. Biological Conservation 144(1):339-349 DOI 10.1016/j.biocon.2010.09.011

Rousset F. 2008. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources* **8**(1):103–106 DOI 10.1111/j.1471-8286.2007.01931.x

Sasaki M, Hammer MP, Unmack PJ, Adams M, Beheregaray LB. 2016. Population genetics of a widely distributed small freshwater fish with varying conservation concerns: the southern purple-spotted gudgeon, *Mogurnda adspersa*. *Conservation Genetics* **17**: 875-889 DOI 10.1007/s10592-016-0829-2.

Schmidt DJ, Crook DA, Macdonald JI, Huey JA, Zampatti BP, Chilcott S, Raadik TA, Hughes JM. 2014. Migration history and stock structure of two putatively diadromous teleost fishes as determined by genetic and otolith chemistry analyses. *Freshwater Science* **33** (1):193-206 DOI 10.1086/674796.

Schmidt DJ, Espinoza T, Connell M, Hughes JM. (in press). Conservation genetics of the Mary River turtle (*Elusor macrurus*) in natural and captive populations. *Aquatic Conservation: Marine and Freshwater Ecosystems* (accepted July 2017).

Schmidt DJ, Islam MR-U, Hughes JM. 2016. Complete mitogenomes for two lineages of the Australian smelt, *Retropinna semoni* (Osmeriformes: Retropinnidae). *Mitochondrial DNA Part B* **1**(1):615-616 DOI 10.1080/23802359.2016.1209097.

Shaddick K, Gilligan DM, Burridge CP, Jerry DR, Truong K, Beheregaray LB. 2011. Historic divergence with contemporary connectivity in a catadromous fish, the estuary perch (*Macquaria colonorum*). *Canadian Journal of Fisheries and Aquatic Sciences* **68**:304-318 DOI 10.1139/F10-139.

Sharma S, Hughes JM. 2009. Genetic structure and phylogeography of freshwater shrimps (*Macrobrachium australiense* and *Macrobrachium tolmerum*): the role of contemporary and historical events. *Marine and Freshwater Research* **60**:541-553 DOI 10.1071/MF07235.

Slatkin M. 1987. Gene flow and the geographic structure of populations. *Science* **236**:787-792 DOI 10.1126/science.3576198.

Tallmon DA, Luikart G, Waples RS. 2004. The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution* **19**:489-496 DOI 10.1016/j.tree.2004.07.003.

Tatarenkov A, Healey CIM, Avise JC. 2010. Microgeographic population structure of green swordail fish: genetic differentiation despite abundant migration. *Molecular Ecology* **19**:257-268 DOI 10.1111/j.1365-294X.2009.04464.x.

Thacker CE, Unmack PJ, Matsui L, Rifenbark N. 2007. Comparative phylogeography of five sympatric Hypseleotris species (Teleostei: Eleotridae) in south-eastern Australia reveals a complex pattern of drainage basin exchanges with little congruence across species. *Journal of Biogeography* **34**(9):1518-1533 DOI 10.1111/j.1365-2699.2007.01711.x.

Thomas RC Jr, Willette DA, Carpenter KE, Santos MD. 2014. Hidden diversity in sardines: genetic and morphological evidence for cryptic species in the goldstripe sardinella, *Sardinella gibbosa* (Bleeker, 1849). *PLoS ONE* 9(1): e84719 DOI 10.1371/journal.pone.0084719.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**(3):535-538 DOI 10.1111/j.1471-8286.2004.00684.x.

Vrijenhoek RC, Lerman S. 1982. Heterozygosity and developmental stability under sexual and asexual breeding systems. *Evolution* **36**:768-776 DOI 10.1111/j.1558-5646.1982.tb05443.x.

Vrijenhoek RC. 1979. Genetics of a sexually reproducing fish in a highly fluctuating environment. *American Naturalist* **113**:17-29 DOI 10.1086/283362.

Wade MJ, McCauley DE. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* **42**:995-1005 DOI 10.1111/j.1558-5646.1988.tb02518.x.

Ward RD, Woodwark M, Skibinski DOF. 1994. A comparison of genetic diversity levels in marine, freshwater and anadromous fishes. *Journal of Fish Biology* **44**:213-232. DOI 10.1111/j.1095-8649.1994.tb01200.x.

690 Waters JM, Dijkstra LH, Wallis GP. 2000. Biogeography of a southern hemisphere freshwater
691 fish: how important is marine dispersal? *Molecular Ecology* **9**:1815-1821. DOI
692 10.1046/j.1365-294x.2000.01082.x.
693 Williams J, Read C, Norton A, Dovers S, Burgman M, Proctor W, Anderson H. 2001.
694 Biodiversity, Australia state of the environment report 2001. Collingwood: CSIRO
695 Publishing.
696 Williams WD, Allen GR. 1987. Origins and adaptations of the fauna of inland waters. In: Dyne
697 GR, ed. *Fauna of Australia*. Australian Government Printing Service: Canberra, 184-201.
698 Wilson G A, Rannala B. 2003. Bayesian inference of recent migration rates using multilocus
699 genotypes. *Genetics* **163**:1177-1191.
700 Wong BBM, Keogh JS, McGlashan DJ. 2004. Current and historical patterns of drainage
701 connectivity in eastern Australia inferred from population genetic structuring in a
702 widespread freshwater fish *Pseudomugil signifier* (Pseudomugilidae). *Molecular Ecology*
703 **13**:391-401 DOI 10.1046/j.1365-294X.2003.02085.x.
704 Woods RJ, Macdonald JI, Crook DA, Schmidt DJ, Hughes JM. 2010. Contemporary and
705 historical patterns of connectivity among populations of an inland river fish species
706 inferred from genetics and otolith chemistry. *Canadian Journal of Fisheries and Aquatic*
707 *Sciences* **67**:1098-1115 DOI 10.1139/F10-043.
708 Wright S. 1943. Isolation by distance. *Genetics* **28**(2):114.
709
710

Table 1(on next page)

Summary of Sample information and genetic diversity indices for Australian smelt

Number of samples used for genetic analysis (N), mean number of alleles per population (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E), allelic richness (A_R), mean inbreeding index (F_{IS})

1

Group name	Sampling site	Site code	Latitude (E)	Longitude (S)	N	N_A	H_O	H_E	A_R	P_{AR}	F_{IS}	Population specific F_{ST}
CEQ	Tinana	MRD	152°42'57.8"	25°36'04.3"	28	10.20	0.679	0.758	5.59	0.71	0.107	0.164
CEQ	Mary_upper	MRU	152°48'47.9"	26°38'55.5"	29	10.90	0.779	0.826	6.26	0.85	0.059	0.136
CEQ	Noosa_lower	NSD	152°52'21.4"	26°17'05.7"	30	6.10	0.576	0.617	3.99	0.76	0.067	0.340
CEQ	Mooloolah_lower	MLD	153° 0'44.64"	26°46'18.83"	16	4.60	0.670	0.604	3.41	0.18	-0.115	0.390
CEQ	Mooloolah_upper	MLU	152°55'13.1"	26°45'07.9"	32	6.40	0.609	0.566	3.85	0.23	-0.079	0.334
SEQ	Brisbane_lower	BRD	152°55'49.9"	27°30'16.05"	32	9.50	0.754	0.755	5.49	0.4	0.002	0.191
SEQ	Brisbane_upper	BRU	152°35'13.5"	27°58'43.9"	32	8.80	0.760	0.763	5.18	0.39	0.005	0.216
SEQ	Logan/Albert_lower	LGD	152°59'01.8"	28°10'15.6"	8	7.60	0.701	0.847	6.9	0.69	0.184	0.0846
SEQ	Logan/Albert_upper	LGU	152°56'23.6"	28°19'19.7"	24	11.40	0.828	0.845	6.58	0.57	0.021	0.107
SEQ	Coomera_lower	CMD	153° 11'20.9"	28°02'55.5"	24	13.90	0.839	0.887	7.42	0.53	0.054	0.0743
SEQ	Coomera_ upper	CMU	153° 09'13.4"	28°05'01.8"	32	14.70	0.848	0.887	7.4	0.58	0.045	0.0775
SEQ	Nerang_lower	NRD	153° 17'52.0"	28°01'33.7"	8	6.40	0.718	0.798	5.8	0.64	0.106	0.156
SEQ	Nerang_upper	NRU	153° 14'02.8"	28°07'29.2"	32	13.70	0.782	0.853	6.87	0.29	0.084	0.0879
SEQ	Currumbin_lower	CRD	153°25'24.8"	28°10'41.9"	32	11.20	0.771	0.803	5.99	0.33	0.041	0.130
SEQ	Currumbin_upper	CRU	153°23'11.9"	28°12'49.6"	32	10.90	0.769	0.785	6	0.34	0.021	0.135

2

Table 2 (on next page)

Pairwise F_{ST} values among all pairs of populations

Bold values were statistically significant after bonferroni correction

1

	MRD	MRU	NSD	MLD	MLU	BRD	BRU	LGD	LGU	CMD	CMU	NRD	NRU	CRD	CRU
MRD	0.000														
MRU	**0.097**	0.000													
NSD	**0.335**	**0.310**	0.000												
MLD	**0.231**	**0.182**	**0.362**	0.000											
MLU	**0.299**	**0.244**	**0.404**	0.009	0.000										
BRD	**0.161**	**0.131**	**0.323**	**0.203**	**0.276**	0.000									
BRU	**0.159**	**0.136**	**0.320**	**0.218**	**0.295**	0.013	0.000								
LGD	**0.222**	**0.101**	**0.394**	**0.248**	**0.316**	**0.157**	**0.170**	0.000							
LGU	**0.227**	**0.137**	**0.368**	**0.246**	**0.317**	**0.173**	**0.189**	-0.006	0.000						
CMD	**0.194**	**0.151**	**0.319**	**0.186**	**0.259**	**0.130**	**0.146**	**0.126**	**0.161**	0.000					
CMU	**0.184**	**0.141**	**0.293**	**0.165**	**0.236**	**0.124**	**0.139**	**0.126**	**0.164**	0.010	0.000				
NRD	**0.242**	**0.125**	**0.393**	**0.244**	**0.316**	**0.187**	**0.191**	**0.073**	**0.082**	**0.152**	**0.141**	0.000			
NRU	**0.211**	**0.110**	**0.329**	**0.203**	**0.261**	**0.176**	**0.183**	**0.053**	**0.074**	**0.146**	**0.137**	-0.018	0.000		
CRD	**0.244**	**0.152**	**0.353**	**0.222**	**0.277**	**0.195**	**0.211**	**0.125**	**0.147**	**0.166**	**0.156**	**0.076**	**0.062**	0.000	
CRU	**0.287**	**0.178**	**0.397**	**0.270**	**0.320**	**0.248**	**0.256**	**0.120**	**0.150**	**0.215**	**0.201**	**0.073**	**0.059**	0.011	0.000

2

Table 3(on next page)

AMOVA for hierarchical arrangements of the 15 sample sites

*** $P < 0.001$

1

Structure tested		Observed partition		*F*- Statistics
		Variance	% of variation	
A.	All sites			
	Among populations	0.26509 Va	20.50	
	Within populations	1.02804 Vb	79.50	F_{ST} = 0.21***
B.	Based on group (CEQ & SEQ)			
	Among group	0.07121 Va	5.36	F_{CT} = 0.05***
	Among sites within group	0.23035 Vb	17.32	F_{SC} = 0.18***
	Within sites	1.02804 Vc	77.32	F_{ST} = 0.23***
C.	Based on river			
i	Among CEQ group			
	Among rivers	0.10506 Va	22.39	F_{CT} = 0.22***
	Among sites within rivers	0.03605 Vb	7.69	F_{SC} = 0.10***
	Within sites	0.32802 Vc	69.92	F_{ST} = 0.30***
ii	Among SEQ group			
	Among rivers	0.27139 Va	12.49	F_{CT} = 0.13***
	Among sites within rivers	0.01106 Vb	0.51	F_{SC} =0.006***
	Within sites	1.89107 Vc	87.01	F_{ST} = 0.13***

2

Table 4(on next page)

Contemporary gene flow identifying the immigrant

Diagonal values (in italics): proportion of non-migrant Australian smelt. The most relevant migration rates are shown in bold.

1

	MRD	MRU	NSD	MLD	MLU	BRD	BRU	LGD	LGU	CMD	CMU	NRD	NRU	CRD	CRU
MRD	*0.8890*	0.0093	0.0079	0.0076	0.0077	0.0076	0.0080	0.0075	0.0085	0.0078	0.0077	0.0080	0.0078	0.0075	0.0081
MRU	0.0104	*0.8873*	0.0074	0.0075	0.0074	0.0076	0.0122	0.0075	0.0073	0.0076	0.0074	0.0081	0.0076	0.0073	0.0075
NSD	0.0073	0.0068	*0.8969*	0.0074	0.0074	0.0072	0.0073	0.0072	0.0072	0.0070	0.0079	0.0079	0.0074	0.0074	0.0079
MLD	0.0107	0.0109	0.0109	*0.6778*	**0.1789**	0.0108	0.0112	0.0112	0.0111	0.0113	0.0110	0.0109	0.0109	0.0114	0.0108
MLU	0.0071	0.0066	0.0070	0.0070	*0.9005*	0.0073	0.0073	0.0071	0.0073	0.0071	0.0079	0.0072	0.0068	0.0066	0.0069
BRD	0.0076	0.0120	0.0071	0.0079	0.0076	*0.6801*	**0.2181**	0.0076	0.0075	0.0077	0.0072	0.0072	0.0070	0.0077	0.0077
BRU	0.0067	0.0072	0.0068	0.0072	0.0068	0.0070	*0.9021*	0.0070	0.0071	0.0071	0.0073	0.0071	0.0069	0.0069	0.0069
LGD	0.0172	0.0147	0.0146	0.0149	0.0149	0.0160	0.0150	*0.6815*	**0.1205**	0.0158	0.0149	0.0148	0.0155	0.0152	0.0146
LGU	0.0086	0.0089	0.0082	0.0083	0.0085	0.0084	0.0086	0.0086	*0.8803*	0.0081	0.0082	0.0094	0.0083	0.0092	0.0084
CMD	0.0078	0.0082	0.0081	0.0084	0.0082	0.0083	0.0084	0.0078	0.0084	*0.6750*	**0.2179**	0.0087	0.0082	0.0082	0.0084
CMU	0.0071	0.0072	0.0072	0.0074	0.0071	0.0073	0.0067	0.0074	0.0071	0.0072	*0.9000*	0.0070	0.0069	0.0076	0.0068
NRD	0.0144	0.0139	0.0149	0.0144	0.0139	0.0149	0.0140	0.0141	0.0147	0.0141	0.0145	*0.6813*	**0.1325**	0.0145	0.0139
NRU	0.0069	0.0067	0.0071	0.0069	0.0074	0.0076	0.0070	0.0069	0.0075	0.0071	0.0078	0.0070	*0.8986*	0.0089	0.0068
CRD	0.0073	0.0066	0.0073	0.0075	0.0070	0.0072	0.0072	0.0074	0.0071	0.0077	0.0070	0.0076	0.0075	*0.8981*	0.0075
CRU	0.0071	0.0073	0.0070	0.0071	0.0072	0.0070	0.0073	0.0067	0.0075	0.0070	0.0073	0.0070	0.0068	**0.2318**	*0.6759*

2

Table 5(on next page)

Results of the assessment for detecting first-generation migrant performed using GENECLASS2 showing the number of individual migrants ($P < 0.01$) detected per sampling location and results are based on the *Lh/Lmax* statistic

1

Sample	F_0 Migrants From														
	MRD	MRU	NSD	MLD	MLU	BRD	BRU	LGD	LGU	CMD	CMU	NRD	NRU	CRD	CRU
MRD		1	0	0	0	0	0	0	0	0	0	0	1	0	0
MRU	0		0	0	0	0	0	0	0	0	0	0	0	0	0
NSD	0	0		0	0	0	0	0	0	0	0	0	0	0	0
MLD	0	0	0		0	0	0	0	0	0	0	0	0	0	0
MLU	0	0	0	1		0	0	0	0	0	0	0	0	0	0
BRD	0	0	0	0	0		0	0	0	0	0	0	0	0	0
BRU	0	0	0	0	0	1		0	0	0	0	0	0	0	0
LGD	0	0	0	0	0	0	0		1	0	0	0	0	0	0
LGU	0	0	0	0	0	0	0	1		0	0	0	0	0	0
CMD	0	0	0	0	0	0	0	0	0		3	0	0	0	0
CMU	0	0	0	0	0	0	0	0	0	1		0	0	0	0
NRD	0	0	0	0	0	0	0	0	0	0	0		0	0	0
NRU	0	0	0	0	0	0	0	0	0	0	0	2		0	0
CRD	0	0	0	0	0	0	0	0	0	0	0	0	0		2
CRU	0	0	0	0	0	0	0	0	0	0	0	0	0	1	

2

Figure 1

Map of Queensland highlighting the fifteen sampling locations

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3284v1 | CC BY 4.0 Open Access | rec: 26 Sep 2017, publ: 26 Sep 2017

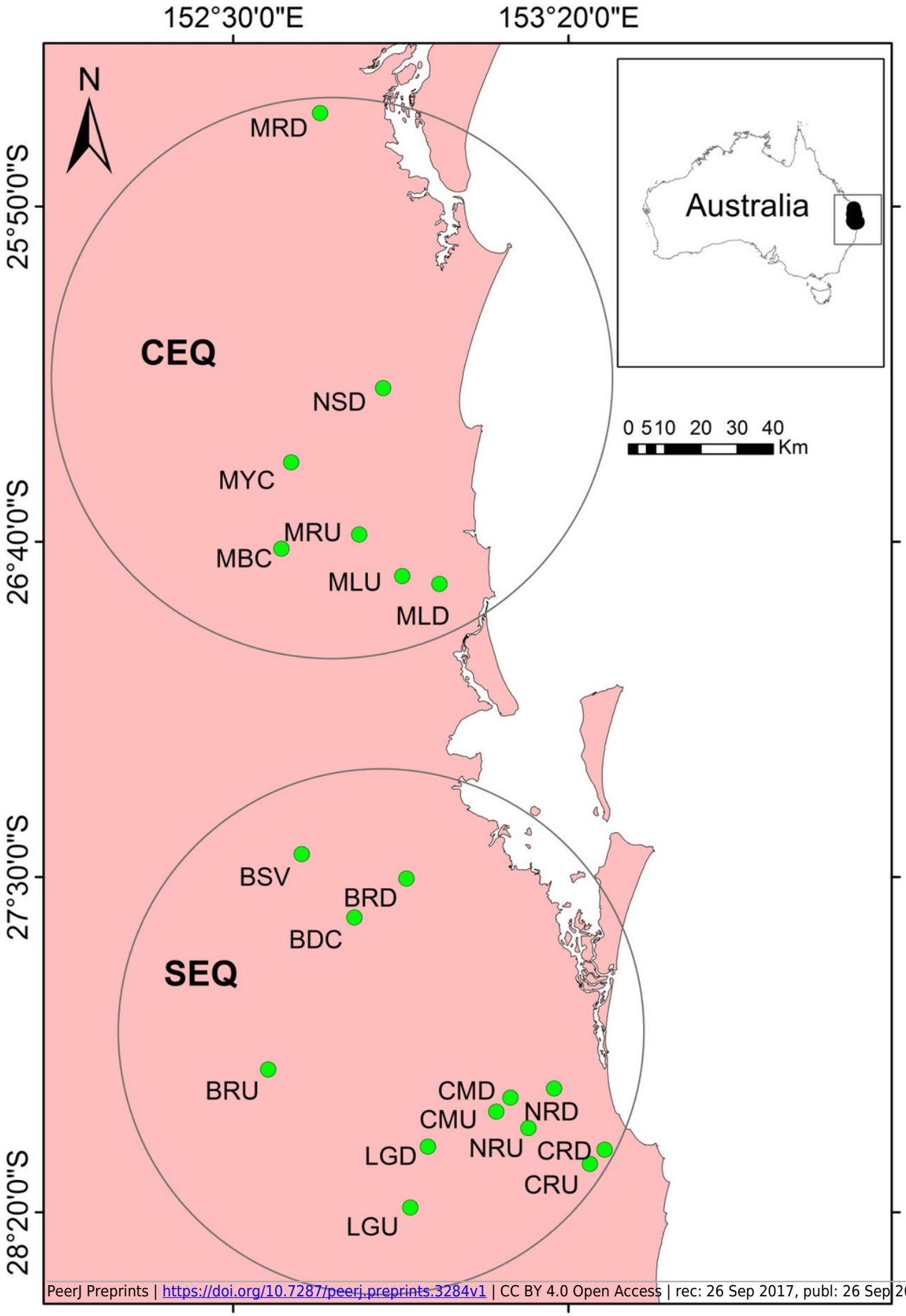

Figure 2

STRUCTURE analysis A) Bar plot of estimated membership of each individual in k = 2 clusters B) Bar plot of estimated membership of each individual in k = 3 clusters C) Bar plot of estimated membership of each individual in k = 8 clusters

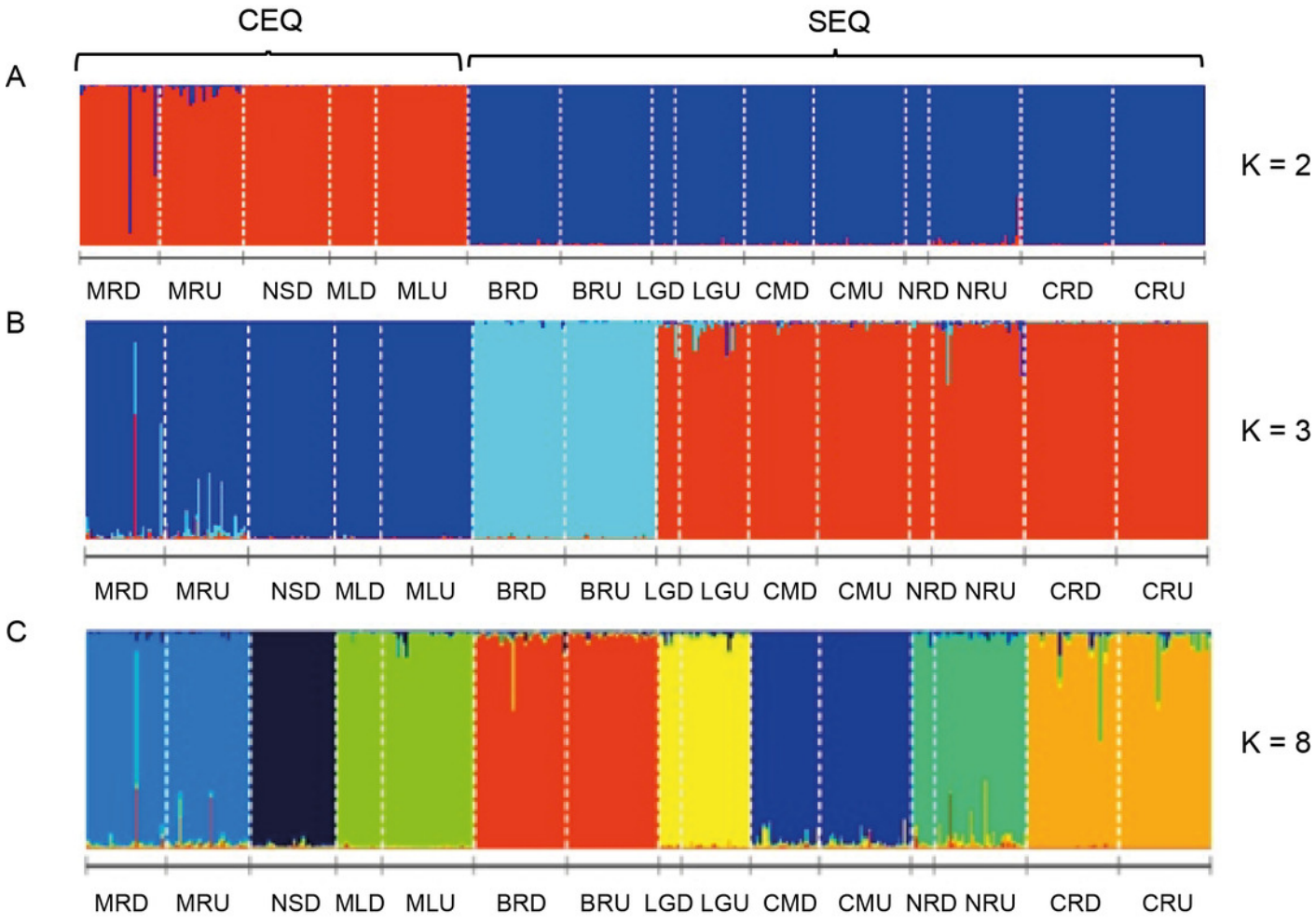

Figure 3

A) Analysis of isolation by distance for SEQ populations B) Analysis of isolation by distance for CEQ populations

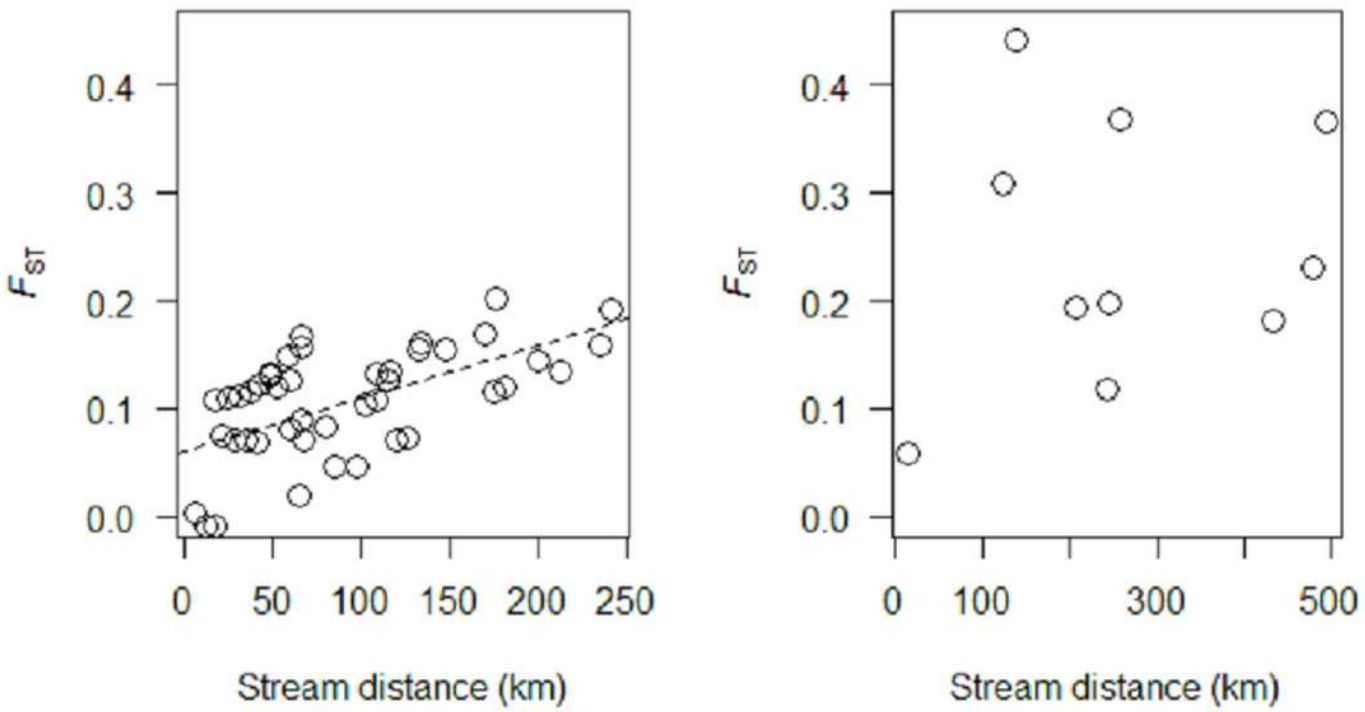

Figure 4

Neighbour-joining tree of the cyt *b* dataset for 68 Australian smelt samples from 19 sampling localities. Individual sample codes coloured according to river. Node values are bootstrap support