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

View the peer-reviewed version (peerj.com/articles/4654), which is the preferred citable publication unless you specifically need to cite this preprint.

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*\textsubscript{ST} values suggested that both groups exhibit very low dispersal among rivers (SEQ *F*\textsubscript{ST} = 0.13; CEQ *F*\textsubscript{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 Islam¹, Daniel J Schmidt¹, David A Crook², Jane M Hughes¹

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

²Research 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
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°C for 4 min, followed by 35 cycles at 94°C for 1 min, 57°C for 30 s, 72°C for 1 min and a final extension at 72°C 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 °C, followed by 45 cycles of 30 s at 95 °C, 45 s at 53 °C, 45 s at 72 °C and a final extension cycle of 7 min at 72 °C. 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 microsatellite 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 $\Delta K$ statistic (Evanno, Regnaut & Goudet, 2005). The K value, where $\Delta K$ 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 \(3 \times 10^7\) 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 $\text{Ln}[P(D|K)] = -15246.1 \pm 1.028753$) indicating this as the best estimate of the true number of the genetic clusters. The height of $\Delta K$ was used as an indicator of the strength of the signal detected by STRUCTURE (Evanno, Regnaut & Goudet, 2005). $\Delta K$ 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 patterns 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 \( (R^2 = 0.3687, p = 0.001; \text{BRD, BRU, LGD, LGU, CMD, CMU, NRD, NRU, CRU}) \) (Fig. 3A), but not for CEQ group \( (R^2 = 0.0355, p = 0.302; \text{MRD, NSD, MRU, MLD and MLU}) \) (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-XXXXXXXX. 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
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.
Australian Museum, Perth.
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.
Population structure in an endangered songbird: maintenance of genetic differentiation
despite high vagility and significant population recovery. Molecular Ecology 17:3628-
Baumgartner LJ, Harris JH. 2007. Passage of non-salmonid fish through a Deelder lock on a


Goudet J. 2001. FSTAT, version 2.9. 3, A program to estimate and test gene diversities and fixation indices. Lausanne University, Lausanne, Switzerland.


Hammer MP, Unmack PJ, Adams M, Raadik TA, Johnson JB. 2014. A multigene molecular assessment of cryptic biodiversity in the iconic freshwater blackfishes (Teleostei:


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


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 \( (F_{IS}) \)
<table>
<thead>
<tr>
<th>Group name</th>
<th>Sampling site</th>
<th>Site code</th>
<th>Latitude (E)</th>
<th>Longitude (S)</th>
<th>N</th>
<th>Nₛ</th>
<th>Hₒ</th>
<th>Hₑ</th>
<th>Aᵦ</th>
<th>Pᵦₑ</th>
<th>Fₛ</th>
<th>Population specific Fₛₜ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEQ</td>
<td>Tinana</td>
<td>MRD</td>
<td>152°42'57.8&quot;</td>
<td>25°36'04.3&quot;</td>
<td>28</td>
<td>10.20</td>
<td>0.679</td>
<td>0.758</td>
<td>5.59</td>
<td>0.71</td>
<td>0.107</td>
<td>0.164</td>
</tr>
<tr>
<td>CEQ</td>
<td>Mary upper</td>
<td>MRU</td>
<td>152°48'47.9&quot;</td>
<td>26°38'55.5&quot;</td>
<td>29</td>
<td>10.90</td>
<td>0.779</td>
<td>0.826</td>
<td>6.26</td>
<td>0.85</td>
<td>0.059</td>
<td>0.136</td>
</tr>
<tr>
<td>CEQ</td>
<td>Noosa lower</td>
<td>NSD</td>
<td>152°52'21.4&quot;</td>
<td>26°17'05.7&quot;</td>
<td>30</td>
<td>6.10</td>
<td>0.576</td>
<td>0.617</td>
<td>3.99</td>
<td>0.76</td>
<td>0.067</td>
<td>0.340</td>
</tr>
<tr>
<td>CEQ</td>
<td>Mooloolah lower</td>
<td>MLD</td>
<td>153°0'44.64&quot;</td>
<td>26°46'18.83&quot;</td>
<td>16</td>
<td>4.60</td>
<td>0.670</td>
<td>0.604</td>
<td>3.41</td>
<td>0.18</td>
<td>-0.115</td>
<td>0.390</td>
</tr>
<tr>
<td>CEQ</td>
<td>Mooloolah upper</td>
<td>MLU</td>
<td>152°55'13.1&quot;</td>
<td>26°45'07.9&quot;</td>
<td>32</td>
<td>6.40</td>
<td>0.609</td>
<td>0.566</td>
<td>3.85</td>
<td>0.23</td>
<td>-0.079</td>
<td>0.334</td>
</tr>
<tr>
<td>SEQ</td>
<td>Brisbane lower</td>
<td>BRD</td>
<td>152°55'49.9&quot;</td>
<td>27°30'16.05&quot;</td>
<td>32</td>
<td>9.50</td>
<td>0.754</td>
<td>0.755</td>
<td>5.49</td>
<td>0.4</td>
<td>0.002</td>
<td>0.191</td>
</tr>
<tr>
<td>SEQ</td>
<td>Brisbane upper</td>
<td>BRU</td>
<td>152°35'13.5&quot;</td>
<td>27°58'43.9&quot;</td>
<td>32</td>
<td>8.80</td>
<td>0.760</td>
<td>0.763</td>
<td>5.18</td>
<td>0.39</td>
<td>0.005</td>
<td>0.216</td>
</tr>
<tr>
<td>SEQ</td>
<td>Logan/Albert lower</td>
<td>LGD</td>
<td>152°59'01.8&quot;</td>
<td>28°10'15.6&quot;</td>
<td>8</td>
<td>7.60</td>
<td>0.701</td>
<td>0.847</td>
<td>6.9</td>
<td>0.69</td>
<td>0.184</td>
<td>0.0846</td>
</tr>
<tr>
<td>SEQ</td>
<td>Logan/Albert upper</td>
<td>LGU</td>
<td>152°56'23.6&quot;</td>
<td>28°19'19.7&quot;</td>
<td>24</td>
<td>11.40</td>
<td>0.828</td>
<td>0.845</td>
<td>6.58</td>
<td>0.57</td>
<td>0.021</td>
<td>0.107</td>
</tr>
<tr>
<td>SEQ</td>
<td>Coomera lower</td>
<td>CMD</td>
<td>153°11'20.9&quot;</td>
<td>28°02'55.5&quot;</td>
<td>24</td>
<td>13.90</td>
<td>0.839</td>
<td>0.887</td>
<td>7.4</td>
<td>0.53</td>
<td>0.054</td>
<td>0.0743</td>
</tr>
<tr>
<td>SEQ</td>
<td>Coomera upper</td>
<td>CMU</td>
<td>153°09'13.4&quot;</td>
<td>28°05'01.8&quot;</td>
<td>32</td>
<td>14.70</td>
<td>0.848</td>
<td>0.887</td>
<td>7.4</td>
<td>0.58</td>
<td>0.045</td>
<td>0.0775</td>
</tr>
<tr>
<td>SEQ</td>
<td>Nerang lower</td>
<td>NRD</td>
<td>153°17'52.0&quot;</td>
<td>28°01'33.7&quot;</td>
<td>8</td>
<td>6.40</td>
<td>0.718</td>
<td>0.798</td>
<td>5.8</td>
<td>0.64</td>
<td>0.106</td>
<td>0.156</td>
</tr>
<tr>
<td>SEQ</td>
<td>Nerang upper</td>
<td>NRU</td>
<td>153°14'02.8&quot;</td>
<td>28°07'29.2&quot;</td>
<td>32</td>
<td>13.70</td>
<td>0.782</td>
<td>0.853</td>
<td>6.87</td>
<td>0.29</td>
<td>0.084</td>
<td>0.0879</td>
</tr>
<tr>
<td>SEQ</td>
<td>Currumbin lower</td>
<td>CRD</td>
<td>153°25'24.8&quot;</td>
<td>28°10'41.9&quot;</td>
<td>32</td>
<td>11.20</td>
<td>0.771</td>
<td>0.803</td>
<td>5.99</td>
<td>0.33</td>
<td>0.041</td>
<td>0.130</td>
</tr>
<tr>
<td>SEQ</td>
<td>Currumbin upper</td>
<td>CRU</td>
<td>153°23'11.9&quot;</td>
<td>28°12'49.6&quot;</td>
<td>32</td>
<td>10.90</td>
<td>0.769</td>
<td>0.785</td>
<td>6</td>
<td>0.34</td>
<td>0.021</td>
<td>0.135</td>
</tr>
</tbody>
</table>
**Table 2** *(on next page)*

Pairwise $F_{st}$ values among all pairs of populations

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

1

2
Table 3 (on next page)

AMOVA for hierarchical arrangements of the 15 sample sites

*** P < 0.001
<table>
<thead>
<tr>
<th>Structure tested</th>
<th>Observed partition</th>
<th>F- Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variance</td>
<td>% of variation</td>
</tr>
<tr>
<td>A. All sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>0.26509 V&lt;sub&gt;a&lt;/sub&gt;</td>
<td>20.50</td>
</tr>
<tr>
<td>Within populations</td>
<td>1.02804 V&lt;sub&gt;b&lt;/sub&gt;</td>
<td>79.50</td>
</tr>
<tr>
<td>B. Based on group (CEQ &amp; SEQ)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among group</td>
<td>0.07121 V&lt;sub&gt;a&lt;/sub&gt;</td>
<td>5.36</td>
</tr>
<tr>
<td>Among sites within group</td>
<td>0.23035 V&lt;sub&gt;b&lt;/sub&gt;</td>
<td>17.32</td>
</tr>
<tr>
<td>Within sites</td>
<td>1.02804 V&lt;sub&gt;c&lt;/sub&gt;</td>
<td>77.32</td>
</tr>
<tr>
<td>C. Based on river</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i Among CEQ group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among rivers</td>
<td>0.10506 V&lt;sub&gt;a&lt;/sub&gt;</td>
<td>22.39</td>
</tr>
<tr>
<td>Among sites within rivers</td>
<td>0.03605 V&lt;sub&gt;b&lt;/sub&gt;</td>
<td>7.69</td>
</tr>
<tr>
<td>Within sites</td>
<td>0.32802 V&lt;sub&gt;c&lt;/sub&gt;</td>
<td>69.92</td>
</tr>
<tr>
<td>ii Among SEQ group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among rivers</td>
<td>0.27139 V&lt;sub&gt;a&lt;/sub&gt;</td>
<td>12.49</td>
</tr>
<tr>
<td>Among sites within rivers</td>
<td>0.01106 V&lt;sub&gt;b&lt;/sub&gt;</td>
<td>0.51</td>
</tr>
<tr>
<td>Within sites</td>
<td>1.89107 V&lt;sub&gt;c&lt;/sub&gt;</td>
<td>87.01</td>
</tr>
</tbody>
</table>
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.
<table>
<thead>
<tr>
<th></th>
<th>MRD</th>
<th>MRU</th>
<th>NSD</th>
<th>MLD</th>
<th>MLU</th>
<th>BRD</th>
<th>BRU</th>
<th>LGD</th>
<th>LGU</th>
<th>CMD</th>
<th>CMU</th>
<th>NRD</th>
<th>NRU</th>
<th>CRD</th>
<th>CRU</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRD</td>
<td>0.8890</td>
<td>0.0093</td>
<td>0.0079</td>
<td>0.0076</td>
<td>0.0077</td>
<td>0.0076</td>
<td>0.0080</td>
<td>0.0075</td>
<td>0.0085</td>
<td>0.0078</td>
<td>0.0077</td>
<td>0.0080</td>
<td>0.0078</td>
<td>0.0075</td>
<td>0.0081</td>
</tr>
<tr>
<td>MRU</td>
<td>0.0104</td>
<td>0.8873</td>
<td>0.0074</td>
<td>0.0075</td>
<td>0.0074</td>
<td>0.0076</td>
<td>0.0122</td>
<td>0.0075</td>
<td>0.0073</td>
<td>0.0076</td>
<td>0.0074</td>
<td>0.0081</td>
<td>0.0076</td>
<td>0.0073</td>
<td>0.0075</td>
</tr>
<tr>
<td>NSD</td>
<td>0.0073</td>
<td>0.0068</td>
<td>0.8969</td>
<td>0.0074</td>
<td>0.0074</td>
<td>0.0076</td>
<td>0.0072</td>
<td>0.0072</td>
<td>0.0072</td>
<td>0.0070</td>
<td>0.0079</td>
<td>0.0079</td>
<td>0.0074</td>
<td>0.0074</td>
<td>0.0079</td>
</tr>
<tr>
<td>MLD</td>
<td>0.0107</td>
<td>0.0109</td>
<td>0.0109</td>
<td>0.6778</td>
<td><strong>0.1789</strong></td>
<td>0.0108</td>
<td>0.0112</td>
<td>0.0112</td>
<td>0.0111</td>
<td>0.0113</td>
<td>0.0110</td>
<td>0.0109</td>
<td>0.0114</td>
<td>0.0108</td>
<td></td>
</tr>
<tr>
<td>MLU</td>
<td>0.0071</td>
<td>0.0066</td>
<td>0.0070</td>
<td>0.9005</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.0073</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.0079</td>
<td>0.0079</td>
<td>0.0072</td>
<td>0.0068</td>
<td>0.0066</td>
<td>0.0069</td>
</tr>
<tr>
<td>BRD</td>
<td>0.0076</td>
<td>0.0120</td>
<td>0.0071</td>
<td>0.0079</td>
<td>0.0076</td>
<td>0.6801</td>
<td><strong>0.2181</strong></td>
<td>0.0076</td>
<td>0.0075</td>
<td>0.0077</td>
<td>0.0072</td>
<td>0.0070</td>
<td>0.0077</td>
<td>0.0077</td>
<td></td>
</tr>
<tr>
<td>BRU</td>
<td>0.0067</td>
<td>0.0072</td>
<td>0.0068</td>
<td>0.0072</td>
<td>0.0068</td>
<td>0.0070</td>
<td>0.9021</td>
<td>0.0070</td>
<td>0.0071</td>
<td>0.0071</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.0069</td>
<td>0.0069</td>
<td></td>
</tr>
<tr>
<td>LGD</td>
<td>0.0172</td>
<td>0.0147</td>
<td>0.0146</td>
<td>0.0149</td>
<td>0.0149</td>
<td>0.0160</td>
<td>0.0150</td>
<td><strong>0.6815</strong></td>
<td><strong>0.1205</strong></td>
<td>0.0158</td>
<td>0.0149</td>
<td>0.0149</td>
<td>0.0155</td>
<td>0.0152</td>
<td>0.0146</td>
</tr>
<tr>
<td>LGU</td>
<td>0.0086</td>
<td>0.0089</td>
<td>0.0082</td>
<td>0.0083</td>
<td>0.0085</td>
<td>0.0084</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.8803</td>
<td>0.0081</td>
<td>0.0082</td>
<td>0.0094</td>
<td>0.0083</td>
<td>0.0092</td>
<td>0.0084</td>
</tr>
<tr>
<td>CMD</td>
<td>0.0078</td>
<td>0.0082</td>
<td>0.0081</td>
<td>0.0084</td>
<td>0.0082</td>
<td>0.0083</td>
<td>0.0084</td>
<td>0.0084</td>
<td>0.0078</td>
<td>0.0084</td>
<td>0.6750</td>
<td><strong>0.2179</strong></td>
<td>0.0087</td>
<td>0.0082</td>
<td>0.0084</td>
</tr>
<tr>
<td>CMU</td>
<td>0.0071</td>
<td>0.0072</td>
<td>0.0072</td>
<td>0.0074</td>
<td>0.0071</td>
<td>0.0073</td>
<td>0.0067</td>
<td>0.0074</td>
<td>0.0071</td>
<td>0.0072</td>
<td>0.9000</td>
<td>0.0070</td>
<td>0.0069</td>
<td>0.0076</td>
<td>0.0068</td>
</tr>
<tr>
<td>NRD</td>
<td>0.0144</td>
<td>0.0139</td>
<td>0.0149</td>
<td>0.0144</td>
<td>0.0139</td>
<td>0.0149</td>
<td>0.0140</td>
<td>0.0141</td>
<td>0.0147</td>
<td>0.0141</td>
<td>0.0145</td>
<td>0.6813</td>
<td><strong>0.1325</strong></td>
<td>0.0145</td>
<td>0.0139</td>
</tr>
<tr>
<td>NRU</td>
<td>0.0069</td>
<td>0.0067</td>
<td>0.0071</td>
<td>0.0069</td>
<td>0.0074</td>
<td>0.0076</td>
<td>0.0070</td>
<td>0.0069</td>
<td>0.0075</td>
<td>0.0071</td>
<td>0.0078</td>
<td>0.0070</td>
<td>0.8986</td>
<td>0.0089</td>
<td>0.0068</td>
</tr>
<tr>
<td>CRD</td>
<td>0.0073</td>
<td>0.0066</td>
<td>0.0073</td>
<td>0.0075</td>
<td>0.0070</td>
<td>0.0072</td>
<td>0.0072</td>
<td>0.0074</td>
<td><strong>0.0071</strong></td>
<td>0.0070</td>
<td>0.0076</td>
<td>0.0075</td>
<td>0.8981</td>
<td>0.0075</td>
<td></td>
</tr>
<tr>
<td>CRU</td>
<td>0.0071</td>
<td>0.0073</td>
<td>0.0070</td>
<td>0.0071</td>
<td>0.0072</td>
<td>0.0070</td>
<td>0.0073</td>
<td>0.0067</td>
<td>0.0075</td>
<td>0.0070</td>
<td>0.0073</td>
<td>0.0070</td>
<td>0.0068</td>
<td><strong>0.2318</strong></td>
<td><strong>0.6759</strong></td>
</tr>
</tbody>
</table>
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.

**Table 5 (on next page)**
<table>
<thead>
<tr>
<th>Sample</th>
<th>Fg Migrants From</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MRD</td>
</tr>
<tr>
<td>MRD</td>
<td>1</td>
</tr>
<tr>
<td>MRU</td>
<td>0</td>
</tr>
<tr>
<td>NSD</td>
<td>0</td>
</tr>
<tr>
<td>MLD</td>
<td>0</td>
</tr>
<tr>
<td>MLU</td>
<td>0</td>
</tr>
<tr>
<td>BRD</td>
<td>0</td>
</tr>
<tr>
<td>BRU</td>
<td>0</td>
</tr>
<tr>
<td>LGD</td>
<td>0</td>
</tr>
<tr>
<td>LGU</td>
<td>0</td>
</tr>
<tr>
<td>CMD</td>
<td>0</td>
</tr>
<tr>
<td>CMU</td>
<td>0</td>
</tr>
<tr>
<td>NRD</td>
<td>0</td>
</tr>
<tr>
<td>NRU</td>
<td>0</td>
</tr>
<tr>
<td>CRD</td>
<td>0</td>
</tr>
<tr>
<td>CRU</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 1

Map of Queensland highlighting the fifteen sampling locations
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.