

# Stretched to the limit; can a short pelagic larval duration connect adult populations of an Indo-Pacific diadromous fish (*Kuhlia rupestris*)?

P. FEUTRY,\*† A. VERGNES,‡ D. BRODERICK,§ J. LAMBOURDIÈRE,¶ P. KEITH\* and J. R. OVENDEN§

\*Département Milieux et Peuplements Aquatiques, Biologie des Organismes et Écosystèmes Aquatiques, Muséum national d'Histoire naturelle, UMR CNRS-MNHN 7208, CP 26, 57 rue Cuvier, 75231 Paris Cedex 05, France, †Research Institute for Environment & Livelihoods, Charles Darwin University, Darwin 0909, NT, Australia, ‡Département Ecologie et Gestion de la Biodiversité, Muséum national d'Histoire naturelle, UMR CERSP, 55 rue Buffon, 75231 Paris Cedex 05, France, §School of Biological Sciences, University of Queensland Brisbane, Qld 4072, Australia, ¶Département Systématique et Evolution, Muséum national d'Histoire naturelle, UMS 2700 MNHN/CNRS-SSM, CP 26, 57 rue Cuvier, 75231 Paris Cedex 05, France

## Abstract

Freshwater species on tropical islands face localized extinction and the loss of genetic diversity. Their habitats can be ephemeral due to variability in freshwater run-off and erosion. Even worse, anthropogenic effects on these ecosystems are intense. Most of these species are amphidromous or catadromous (i.e. their life cycle includes a marine larval phase), which buffers them against many of these effects. A long pelagic larval duration (PLD) was thought to be critical to ensure the colonization and persistence in tropical islands, but recent findings indicated that several species with short PLDs are successful in those ecosystems. To test the potential of a short PLD in maintaining genetic connectivity and forestalling extirpation, we studied *Kuhlia rupestris*, a catadromous fish species with an extensive distribution in the western Pacific and Indian Oceans. Using a combination of molecular genetic markers (13 microsatellite loci and two gene regions from mtDNA) and modelling of larval dispersal, we show that a short PLD constrains genetic connectivity over a wide geographical range. Molecular markers showed that the short PLD did not prevent genetic divergence through evolutionary time and speciation has occurred or is occurring. Modelling of larvae dispersal suggested limited recent connectivity between genetically homogeneous populations across the Coral Sea. However, a short PLD can maintain connectivity on a subocean basin scale. Conservation and management of tropical diadromous species needs to take into account that population connectivity may be more limited than previously suspected in those species.

**Keywords:** cryptic species, dispersion, microsatellites, mitochondrial markers, ocean-current modelling, tropical islands

Received 4 July 2012; revision received 15 October 2012; accepted 22 October 2012

## Introduction

Island populations are more prone to localized extinction than mainland populations (Vitousek 1988; Case *et al.* 1992) and they generally have less genetic diversity, which is the raw material that allows populations

to evolve in response to environmental changes, whether natural or anthropogenic (Frankham 1997). Tropical freshwater species in insular systems are no exception. In addition, anthropogenic pressures in tropical islands are often particularly intense (Smith *et al.* 2003). Freshwater ecosystems are also affected in the short term by climatic events such as storms, floods, hurricanes and droughts and in the longer term by erosion as mountainous islands are converted to atolls with

Correspondence: Pierre Feutry, Fax: +61889467455; E-mail: pierre.feutry@cdu.edu.au

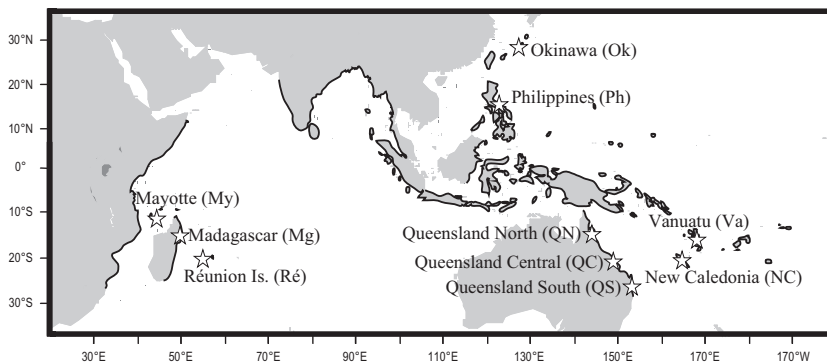
limited running water (Craig 2003). To meet this challenge, the life cycles of most of the freshwater species include a dispersive marine larval phase. This way, exogenous recruits could ensure their persistence if local reproduction fails (MacArthur & Wilson 1963). They are diadromous (i.e. they undergo regular migration between fresh and marine waters), either amphidromous or catadromous. Amphidromous species spawn in fresh water, followed by an almost immediate downstream migration of larvae to the sea; the marine larval phase ends with recruitment into rivers of juveniles for growth and maturation. Catadromous species reproduce at sea, and juveniles recruit into fresh water for growth and maturation (McDowall 1988). Due to these special features, catadromous and amphidromous species can have adult populations in locations separated by thousands of kilometres. For example, the catadromous freshwater eel *Anguilla marmorata* is widely distributed from the western Indian Ocean to the North Pacific and South Pacific (Minegishi *et al.* 2008), as is the amphidromous goby *Sicyopterus lagocephalus* (Keith *et al.* 2005).

A long (i.e. from around 80 to more than 200 days) pelagic larval duration (PLD) was thought to be critical for amphi- and catadromous species to maintain gene flow between widespread populations and to ensure regular recruitment into tropical islands streams (Radtke *et al.* 1988; Keith 2003). However, recent research showed that for some of these species inhabiting tropical islands streams, there is little or no relationship between the geographical extent of the adult populations and PLD. These species include Gobiidae or Kuhliidae with either restricted distributions in Vanuatu (e.g. *Akihito vanuatu*; PLD =  $55.4 \pm 7.5$  days) or New Caledonia (e.g. *Smilosicyopus chloe*; PLD =  $53.6 \pm 5.7$  days), moderate distribution in the West Pacific (e.g. *Sicyopus zosterophorum*; PLD =  $54.6 \pm 5.6$  days) or in the West Indian Ocean (e.g. *Kuhlia sauwagii*; PLD =  $32.3 \pm 3.4$  days) or extensive distribution across the Indo-Pacific region (e.g. *Kuhlia rupestris*; PLD =  $40.6 \pm 6.9$  days; Feutry *et al.* 2012a,c; Taillebois *et al.* 2012). These findings raise the question of the potential of

short PLDs in maintaining connectivity between remote islands.

*Kuhlia rupestris* (jungle perch) is of particular interest as a model species to test this hypothesis. First, this diadromous species (Feutry *et al.* 2012b) has an Indian and Central Pacific distribution from Japan southwards to Australia and from Africa eastwards to American Samoa (Fig. 1). It is unique in being the only diadromous species with a short PLD that occurs in two oceans. Its dispersal relies entirely on its pelagic larval phase, as after the juveniles recruit in fresh water, they stay in fresh or brackish water (Feutry *et al.* 2012b) and reproduction occurs close to the growth habitat, probably in the plume of estuaries (Merrick & Schmida 1984; Hogan & Nicholson 1987). It represents a unique opportunity to examine how well a short PLD facilitates intra-specific gene flow across distances ranging from hundreds to more than 14 000 km. Second, this species has a high conservation status. Within Australia, jungle perch occurs in streams draining the tropical eastern coast and is a popular freshwater recreational sport fish. However, barriers to migration and habitat degradation have been implicated in the decline of southern Queensland stocks since the end of the last century (Hutchison *et al.* 2002). Fishways and restored habitats have been in place since the 2000s that has led to recolonization by many diadromous species like mullet *Mugil cephalus*, Australian bass *Macquaria novemaculeata* and barramundi *Lates calcarifer*, but this has not occurred for jungle perch (Hutchison *et al.* 2002; Hutchison M, personal communication). Two main hypotheses exist: (i) habitat quality is still below the requirements of the species despite restoration, and (ii) recruitment from adjacent populations is nonexistent or too low to facilitate population recovery. Information on connectivity between populations and the dispersal ability of this species would help to distinguish between these hypotheses and to develop measures for this species' conservation and restoration in southern Queensland.

When trying to evaluate the connectivity between populations through marine larval dispersion, several



**Fig. 1** Sampling locations of *Kuhlia rupestris* indicated by white stars. Bold lines show the presumed distributional range of this species.

approaches are available. These approaches include both direct and indirect methods using genetic (e.g. mitochondrial DNA or microsatellites) or geochemical markers (e.g. microchemical signatures in otoliths or shells), as well as the utilization of high-resolution biophysical models (see review by Cowen & Sponaugle 2009). All methods have different strengths and weaknesses. For example, using geochemical markers can be time- and resource-consuming and often limits the observations to a single event. To the contrary, a genetics-based approach provides a view of the connectivity that has taken place over hundreds or thousand of generations through evolutionary time but may not necessarily reflect extant processes (Nielsen & Wakeley 2001; Hedgecock *et al.* 2007). Model-based approaches are promising because advances in computing power and model frameworks allow more complex models to be developed, integrating both biological (i.e. particle behaviour) and physical parameters (i.e. ocean currents). Besides, they allow repeated measures through time and space, thereby capturing the expected environmental variability. Nonetheless, the validation of models requires considerable empirical work and much is required to understand how the interactions between biological and physical processes affect larval dispersal (Cowen & Sponaugle 2009). In addition, modelling approaches predict connectivity over shorter timescales compared to genetics. Combining multiple approaches may help to evaluate accurately larval dispersion (Gilg & Hilbish 2003; Viard *et al.* 2006; Leis *et al.* 2011).

In this study, we developed two different approaches to evaluate the potential of short PLDs in maintaining connectivity between populations of *K. rupestris*. The first investigates connectivity through evolutionary time. It uses molecular markers, including mitochondrial sequences and 15 microsatellite loci to analyse the population genetic structure of jungle perch within its whole Indo-Pacific range. As suggested by Feutry *et al.* (2012c), its short PLD suggests low dispersal abilities and one would expect to find a high genetic structure within the geographical range of the species. The alternative hypothesis is an absence of genetic structure, indicating that despite short PLDs, *K. rupestris* is able to maintain genetic homogeneity through evolutionary time. The second approach focuses on larval dispersal over one generation and uses ocean-current modelling and particle-tracking techniques to predict the actual connectivity between populations in the Coral Sea (southwest Pacific). The combination of these approaches provides new insights into the mechanisms of genetic structuring, speciation and larval dispersal in this species and contributes to the discussion of the colonization and persistence of freshwater fauna in tropical insular streams.

## Material and methods

### Sampling and DNA extraction

A total of 487 *Kuhlia rupestris* were collected from 10 locations throughout its range: Queensland South (QS,  $n = 82$ ), Central (QC,  $n = 55$ ) and North (QN,  $n = 78$ ) in Australia, New Caledonia (NC,  $n = 58$ ), Vanuatu (Efate Is., Pentecost Is., Malékula Is.) (Va,  $n = 40$ ), Japan (Okinawa Is.) (Ok,  $n = 33$ ), Philippines (Ph,  $n = 6$ ), Réunion Is. (Ré,  $n = 57$ ), Mayotte (My,  $n = 73$ ) and Madagascar (Mg,  $n = 5$ ) (Fig. 1, Table 1). Nonlethal fin clips were collected and stored in 90% ethanol.

Laboratory work was carried out in two different institutions, at the Molecular Fisheries Laboratory in Brisbane, Australia (MFL), and at the Muséum national d'Histoire naturelle in Paris, France (MNHN).

Total genomic DNA was extracted using DNeasy Tissue Kits (QIAGEN) at the MFL or using the semi-automated ABI PRISM™ 6100 Nucleic Acid PrepStation (Applied Biosystems) at the MNHN.

### Mitochondrial DNA analysis

We amplified a 795-bp region of the mtDNA ATP gene with the primers ATP8.2 (AAA GCR TYR GCC TTT TAA GC) (Hurwood & Hugues 1998) and PSGR (5' GTGATATGCGTGCGCTTG) (Broderick *et al.* 2011) and a 838-bp region of the mtDNA ND4 gene with the primers H122293-LUE (TTG CAC CAA GAG TTT TTG

**Table 1** Summary of *Kuhlia rupestris* samples that were used in the mitochondrial DNA (ATP, ND4 gene regions) and microsatellite (MS) analyses (13 loci)

Geographical area	Locality	Number of individuals for each analysis		
		ATP	ND4	MS
Southwestern Pacific	QS	24	56	75
	QC	30	42	54
	QN	18	37	78
	NC	53	56	36
	Va	36	40	33
Northwestern Pacific	Ok	30	33	31
	Ph	6	6	6
Southwestern Indian	Ré	28	53	35
	Mg	5	5	—
	My	53	70	37
Total	10	283	398	385

NC, New Caledonia; QC, Queensland Central; QN, Queensland North; QS, Queensland South; Va, Vanuatu; Ok, Okinawa Is.; Mg, Madagascar; My, Mayotte; Ph, Philippines; Ré, Réunion Is.

GTT CCT AAG ACC) (Inoue *et al.* 2001) and ND4r (CAC CTA TGA CTA CCA AAA GCT CAT GTA GAA GC) (Arèvalo *et al.* 1994), in subsets of 281 and 394 samples, respectively. Thermal cycling conditions consisted of an initial denaturation at 94 °C for 1 min 30 s followed by 35 cycles of 94 °C for 5 s, 55 °C for 30 s and 72 °C for 30 s, with a final extension step of 72 °C for 5 min. PCR products were cleaned up prior to sequencing using a Qiagen QIAquick PCR cleanup kit protocol. Approximately 20 ng of DNA was used in standard ABI Dye Terminator sequencing reactions and capillary gel separated on an ABI3130XL (MFL) or an ABI3730XL (MNHN) sequencer. Sequences were verified by forward and reverse comparisons. All sequences for both the ND4 and the ATP genes were manually aligned using BIOEDIT v. 7.1.3 (Ibis Biosciences).

Bayesian and maximum-likelihood (ML) phylogenetic trees were constructed for both ATP and ND4 genes in MrBayes v. 3.1.2 (Huelsenbeck & Ronquist 2001) and TREEFINDER (Jobb *et al.* 2004), respectively. *Kuhlia sauvagii*, which was the most closely related species to *K. rupestris* among the Kuhlidae family (Feutry *et al.* 2013), and *K. marginata* were used as outgroups. Based on the Akaike information criterion in jMODELTEST v. 0.1.1 (Guindon & Gascuel 2003; Posada 2008), the GTR + I +  $\Gamma$  and the HKY + I +  $\Gamma$  models for nucleotide substitution were selected for ATP and ND4 sequences, respectively. Analyses were run using 1st–2nd and 3rd codon positions as two different partitions. MEGA v. 4 (Tamura *et al.* 2007) was used to calculate the mean net divergence (i.e. p-distance) between the major clades found in the haplotype tree.

To investigate the genetic differentiation between *K. rupestris* lineages, we calculated the fixation index ( $\Phi_{ST}$ ) for all pairwise comparisons among different lineages and a permutation test (10 000 permutations) was conducted using ARLEQUIN v 3.5.1.2 (Excoffier & Lischer 2010).  $\Phi_{ST}$  statistic is a molecular analogue of  $F_{ST}$  that includes information on mitochondrial haplotype frequency and genetic distance (here Kimura 2P, Kimura 1980). Different indices, such as number of haplotypes, gene diversity (h) and mean nucleotide diversity ( $\pi$ ), were also calculated in ARLEQUIN to measure the genetic diversity within each lineage.

Recent historical demography in the Coral Sea deserved particular attention in this study. Fu's  $F$  and Tajima's  $D$  indexes were generated with ARLEQUIN for ATP and ND4 genes for each location in this region (i.e. QS, QC, QN, NC and Va) to test for population equilibrium.

#### Microsatellite analysis

A subset of 389 samples was genotyped at 15 microsatellite loci. Primers and PCR conditions are described

elsewhere (Peters *et al.* 2009). PCR products were analysed by electrophoresis on an ABI3130XL (MFL) and an ABI3100 (MNHN) capillary sequencers. Alleles were sized against an internal size standard (GeneScan500 LIZ; Applied Biosystems) and scored using GENEMAPPER v. 3.7 (Applied Biosystems). Sixteen samples were genotyped in both laboratories to confirm allele homology.

Departure from Hardy–Weinberg equilibrium (HWE) within each population by locus and over all loci and linkage disequilibrium among loci was tested using GENEPOP v. 4.0.10 (Raymond & Rousset 1995), using 1000 dememorization steps, 1000 batches and 1000 iterations per batch to reduce the standard errors of  $P$ -values below 0.01. MICRO-CHECKER v. 2.2.3 (Van Oosterhout *et al.* 2004) was employed to check for scoring inconsistencies and to explore whether any deviations from HWE were due to the presence of null alleles.

Genetic differentiation between locations was characterized using  $F$ -statistics theta ( $\theta$ ) (Weir & Cockerham 1984) as implemented in ARLEQUIN v 3.5.1.2 (Excoffier & Lischer 2010). Because microsatellite data sets have maximum fixation index values of  $<1$  (Hedrick 2005), values of population subdivision were standardized to a scale of 0–1 ( $\theta'$ ) using the program RECODEDATA v. 0.1 (Meirmans 2006).

The presence of intraspecific genetic structure was tested using the Bayesian model-based clustering method developed by Pritchard *et al.* (2000), as implemented in STRUCTURE v. 2.3.2. Given a number of clusters ( $K$ ), STRUCTURE assigns individuals to the clusters so as the clusters are in Hardy–Weinberg and linkage equilibrium. Because sample sizes and within-group diversity affect splitting order (Rosenberg *et al.* 2002), we used a 'hierarchical STRUCTURE analysis' as described in the study by Vähä *et al.* (2007) to account for varying levels of population structure among its geographical range. A first round of STRUCTURE was carried out with the entire data set. A second round, and where appropriate a third round, of STRUCTURE analysis was conducted on each of the identified clusters. The admixture model was used with 100 000 iterations and a burn-in period of 100 000 during the first round, but 500 000 iterations and a burn-in of 200 000 were necessary for subsequent rounds to achieve both convergence of statistics ( $\alpha$ ,  $F$ ,  $D_{i,j}$  and the likelihood) and an accurate estimate of the  $[\ln \Pr (X|K)]$  parameter. For all analyses, ten independent runs for each  $K$  (2–6) were performed to evaluate the reliability of the results. The true value of  $K$  was determined from posterior probabilities of  $K$  and the values of  $\Delta K$  as recommended by Evanno *et al.* (2005). To visualize the data, the run of highest value for  $[\ln \Pr (X|K)]$  was selected for the true  $K$  found. Individuals were regarded as correctly assigned to a cluster when their posterior probability of belonging to

that cluster (q) was at least 0.80 (Pritchard *et al.* 2000; Vähä & Primmer 2006). Others were categorized as ambiguous.

### Modelling of larval dispersal in the Coral Sea

We used the web-tool Connie 2 (Condie 2005) to answer questions such as (i) can larvae from healthy populations (i.e. QN, QC, Va and NC) reach the low-density populations in QS? and (ii) where would larvae released in QS most probably go? Connie 2 uses archived current activity from oceanographic models and particle-tracking techniques to predict particle dispersal trajectories from user-specified source regions. Particles are seeded within the user-specified source region at a constant rate of 25 particles per day over a release period specified by the user. Because the calculation of all possible connections over time and space around the Coral Sea would have required far too many simulations, we adopted the following strategy in this study. For all simulations, particles were released each year from the 1 January to the 31 May, which corresponds with the approximate breeding season for *K. rupestris* (Hogan & Nicholson 1987). Our approach assumed that larvae were neutrally buoyant and were passively dispersed with no larval behaviour or mortality. For each location, we considered multiple release points for particles. In Va, particles were released from the western coast of three islands, Santo, Malekula and Efate. In NC, particles were released at each cardinal point of the main island. For both QN and QC, we defined three release points separated by about 100 km. The southern point for QN was at the southern limit of the wet tropics area, just north to Townsville, and the southern point for QC was near Mackay, the southern limit for healthy populations of jungle perch according to Hutchison *et al.* (2002). In QS, we considered four release points, the first one near the border with New South Wales and the other locations to the north about

every 100 km. For all Coral Sea source regions, we considered three depth zones per release point (5, 15 and 25 m) as fish larvae are most concentrated in the upper 30 m of the water column in the Australian region (Smith 2000; Gray & Kingsford 2003). To examine whether larvae from NC, Va, QN or QC can reach QS, the dispersal time used for simulations was 58 days. This value corresponds to the maximum PLD reported for this species (Feutry *et al.* 2012c). To investigate the most probable scenario for the dispersion of larvae released in QS, we used the mean PLD of 44 days reported by Feutry *et al.* (2012c) for the population in the Pacific Ocean. Simulations were run for each year from 1994 to 1999. This period covers a wide range of climatic conditions as it includes two El Niño (1994, 1997) and three La Niña years (1995, 1998, 1999). The results from all years were pooled, and we estimated the dispersal of particles with kernel density estimations at 50%, 90%, 95% and 99% for each depth in each region using HAWTH's tool extension v. 3.27 for ARCGIS v. 9.2 (Beyer 2004).

## Results

### Mitochondrial DNA

Sequence data of 795 nucleotides of the ATP gene from 283 individuals (EMBL accessions HF542116–HF542398) revealed 34 variable positions defining 17 haplotypes. Sequence data of 838 nucleotides of the ND4 gene from 398 individuals (EMBL accessions HF542399–HF542796) revealed 36 variable positions defining 21 haplotypes. A summary of this genetic diversity is given in Table 2.

Both ML and Bayesian phylogenetic trees reconstructed using ATP sequences strongly supported the existence of three monophyletic lineages, but the relationships between these lineages are uncertain (i.e. the bootstrap and posterior probability values were lower than 0.55) (Fig. 2). Lineage 1 consisted of five ATP

**Table 2** Summary of the genetic diversity (ATP, ND4; mitochondrial DNA gene regions and 13 microsatellite loci) of *Kuhlia rupestris* from the Indian and Pacific Oceans

	Marker	<i>N</i>	<i>Na/H</i>	<i>h/H<sub>0</sub></i> ± SD	<i>π</i> ± SD
Lineage 1	ATP	32	5	0.6915 ± 0.0639	0.0013 ± 0.0010
	ND4	38	6	0.5491 ± 0.0878	0.0012 ± 0.0009
	Microsatellite	22	7.2	0.6698 ± 0.2196	—
Lineage 2	ATP	52	2	0.3639 ± 0.0391	0.0013 ± 0.0010
	ND4	86	6	0.2925 ± 0.0619	0.0004 ± 0.0004
	Microsatellite	50	7.2	0.5756 ± 0.2508	—
Lineage 3	ATP	197	10	0.5068 ± 0.0173	0.0009 ± 0.0007
	ND4	270	9	0.3944 ± 0.0354	0.0013 ± 0.0010
	Microsatellite	313	17	0.7159 ± 0.2064	—

*N*, number of individuals; *Na* or *H*, average number of alleles across loci or haplotypes; *h*, haplotype diversity (±standard deviation); *H<sub>0</sub>*, observed heterozygosity (±standard deviation); *π*, nucleotide diversity.

haplotypes, all found exclusively in My. Half of the individuals from this lineage had haplotype Hap06. This lineage was defined by six unique single base pair substitutions (Table S1, Supporting information), and it included three-nucleotide transitions and one-nucleotide transversion. Lineage 2 consisted of only two haplotypes with approximately equal frequencies and differing by two-nucleotide transitions. This lineage was characterized by five single base pair substitutions (Table S1, Supporting information). These haplotypes comprised all samples from Mg and Ré and 22 samples from My. Ten haplotypes were found in lineage 3, including nine-nucleotide transitions and sharing ten single base pair substitutions (Table S1, Supporting information). The most frequent haplotype was Hap01, being shared by all locations in the Pacific Ocean. It was connected to the other eight haplotypes by one or two steps in a 'starlike' haplotype network. These other haplotypes were rare and found only in Va, NC and QN, except Hap14 and 15, which were only found in Ok and the Ph. Hap14 was the most frequent haplotype in these two locations (Fig. 3).

Phylogenetic trees based on ND4 sequences also identified three monophyletic lineages, but only the two lineages from the Indian Ocean had high values of support. However, this gene was more helpful in resolving the phylogenetic relationships between lineages (Fig. S1, Supporting information). Lineage 2 was the most basal, and lineages 1 and 3 were found to be monophyletic. The single base pair substitutions shared by each lineage and the haplotype networks for the ND4 gene are given as supplementary material (Table

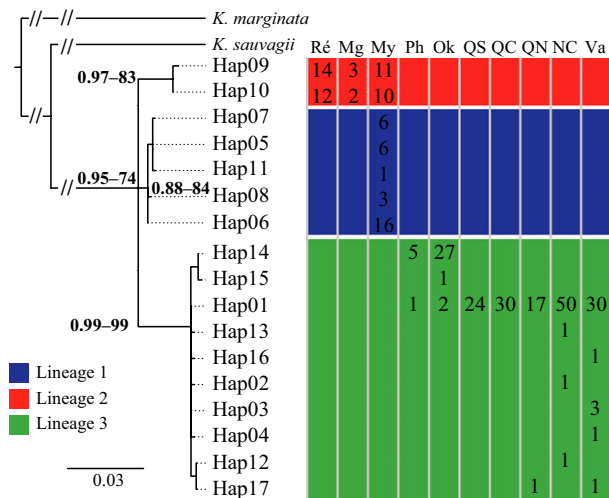


Fig. 2 Phylogenetic tree inferred from 17 ATP sequence haplotypes of *Kuhlia rupestris*. The geographical distribution of haplotypes is given (see Table 1 for location names). The values next to the nodes are posterior probabilities and bootstrap values for Bayesian and maximum likelihood, respectively.

S2 and Fig. S2, Supporting information). The mean inter- and intralineage divergences for the two mitochondrial markers are given in Table 3.

The fixation index showed a significant differentiation among the three lineages ( $\Phi_{ST} = 0.952$ ,  $P < 0.00001$ ). The pairwise  $\Phi_{ST}$  values between lineages 1–2, 1–3 and 2–3 were 0.930 ( $P < 0.00001$ ), 0.957 ( $P < 0.00001$ ) and 0.959 ( $P < 0.00001$ ), respectively.

As expected due to the presence of divergent lineages, the  $\Phi_{ST}$  tests based on the ATP gene showed significant differences in 42 of the 55 lineage–location pairwise comparisons (Table 4). Four groups could be distinguished. Two corresponded to the mtDNA lineages in the Indian Ocean, and two groups were found within the mtDNA lineage 3 in the Pacific. The first one comprised the North Pacific locations, Ok, and the Ph. The second one was made of all other Pacific locations, although a significant difference was found between Va and NC.  $\Phi_{ST}$  tests based on the ND4 gene gave roughly the same results except that the population from Va diverged slightly from QS and QC (Table 4).

Neutrality indices revealed contrasting demographic histories around the Coral Sea. Except ATP-based  $F$  values in QN and ND4-based  $D$  values,  $F$  and  $D$  values for both ATP and ND4 genes were negative and significant

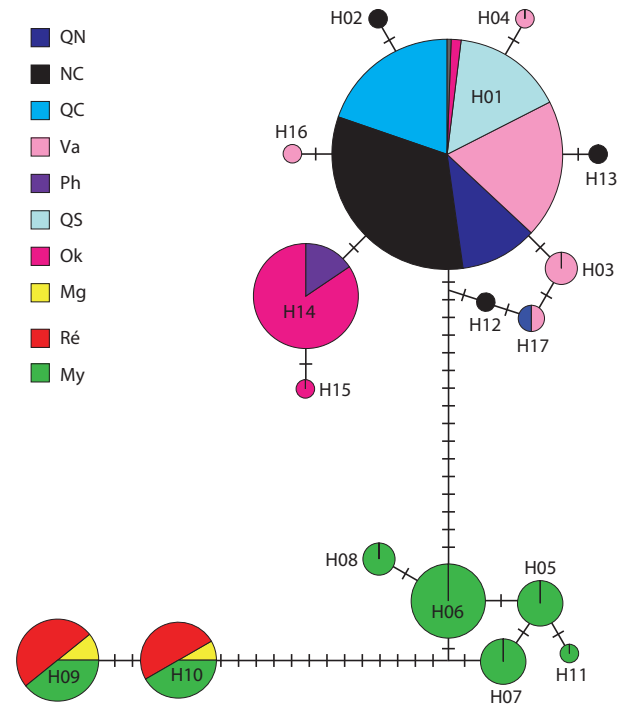


Fig. 3 Network of *Kuhlia rupestris* ATP sequences from ten collection locations in the Pacific and Indian Oceans. Each circle represents a haplotype, whose diameter is proportional to the number of individuals with that haplotype (see Fig. 2). The lengths of the connecting lines are in relation to the number of mutations between haplotypes. See Fig. 1 for location names.

in QN, NC and Va, suggesting historical expansion in these locations (see Table 5). In QC and QS, neither  $F$  nor  $D$  values were significant.

### Microsatellites

A total of 263 alleles were observed across the 15 loci, ranging from six (KRU14) to 44 (KRU03) alleles per locus. All loci were polymorphic except for locus KRU14 in Ré. Significant deviations from HWE were

**Table 3** Intralineage and interlineage mean divergence (percentage of net number of differences) using ATP gene (above) and ND4 gene (below) of mtDNA from *Kuhlia rupestris* sampled from the Indian and Pacific Oceans

		Lineage 1	Lineage 2	Lineage 3
	Intra	<b>0.2</b>	<b>0.3</b>	<b>0.3</b>
L1	<b>0.3</b>		1.9	2.3
L2	<b>0.2</b>	2.6		2.3
L3	<b>0.4</b>	1.8	1.5	

Boldface indicates intralineage mean divergence.

**Table 4** Genetic differentiation between *Kuhlia rupestris* from collection locations (see Fig. 1 for location names) per mtDNA lineage using pairwise  $\Phi_{ST}$  for ATP gene (above, upper value) and ND4 gene (above, lower value) and  $\theta$  (below, upper value) and  $\theta'$  (below, lower value) for microsatellite loci

		Lineage 1				Lineage 2			Lineage 3				
		My	My	Ré	Ma	Ph	Ok	QN	QC	QS	NC	Va	
Localities	N	33	20	26	5	6	30	18	30	24	53	36	
L1	My	22	<b>0.893</b>	<b>0.896</b>	<b>0.879</b>	<b>0.912</b>	<b>0.941</b>	<b>0.927</b>	<b>0.944</b>	<b>0.938</b>	<b>0.955</b>	<b>0.938</b>	
			<b>0.970</b>	<b>0.974</b>	<b>0.960</b>	<b>0.926</b>	<b>0.944</b>	<b>0.961</b>	<b>0.967</b>	<b>0.972</b>	<b>0.968</b>	<b>0.954</b>	
L2	My	15	<b>0.04</b>		−0.046	−0.178	<b>0.947</b>	<b>0.966</b>	<b>0.964</b>	<b>0.977</b>	<b>0.974</b>	<b>0.980</b>	<b>0.967</b>
			<b>0.118</b>		−0.002	−0.104	<b>0.956</b>	<b>0.962</b>	<b>0.975</b>	<b>0.981</b>	<b>0.984</b>	<b>0.979</b>	<b>0.964</b>
	Ré	35	<b>0.091</b>	0.003		−0.126	<b>0.946</b>	<b>0.964</b>	<b>0.961</b>	<b>0.973</b>	<b>0.970</b>	<b>0.977</b>	<b>0.964</b>
			<b>0.249</b>	0.007		−0.036	<b>0.966</b>	<b>0.967</b>	<b>0.978</b>	<b>0.983</b>	<b>0.985</b>	<b>0.981</b>	<b>0.987</b>
	Ma	—	—	—	—	<b>0.950</b>	<b>0.976</b>	<b>0.977</b>	<b>0.992</b>	<b>0.990</b>	<b>0.989</b>	<b>0.975</b>	
			—	—	—	<b>0.912</b>	<b>0.954</b>	<b>0.977</b>	<b>0.987</b>	<b>0.990</b>	<b>0.982</b>	<b>0.959</b>	
L3	Ph	6	<b>0.100</b>	<b>0.112</b>	<b>0.136</b>	—		−0.046	<b>0.808</b>	<b>0.934</b>	<b>0.920</b>	<b>0.895</b>	<b>0.770</b>
			<b>0.364</b>	<b>0.328</b>	<b>0.365</b>			−0.005	<b>0.860</b>	<b>0.901</b>	<b>0.922</b>	<b>0.893</b>	<b>0.804</b>
	Ok	31	<b>0.100</b>	<b>0.102</b>	<b>0.128</b>	—	0.006		<b>0.858</b>	<b>0.915</b>	<b>0.906</b>	<b>0.902</b>	<b>0.833</b>
			<b>0.371</b>	<b>0.321</b>	<b>0.370</b>	—	0.023		<b>0.891</b>	<b>0.913</b>	<b>0.926</b>	<b>0.909</b>	<b>0.860</b>
	QN	78	<b>0.113</b>	<b>0.116</b>	<b>0.127</b>	—	0.005	0.004		0.030	0.016	−0.000	−0.026
			<b>0.406</b>	<b>0.362</b>	<b>0.371</b>	—	0.019	0.015		−0.005	−0.001	−0.015	−0.024
	QC	54	<b>0.100</b>	<b>0.111</b>	<b>0.127</b>	—	0.008	0.003	0.000		0.000	−0.012	0.040
			<b>0.374</b>	<b>0.356</b>	<b>0.378</b>	—	0.031	0.0118	0.000		−0.018	−0.018	<b>0.053</b>
	QS	75	<b>0.116</b>	<b>0.116</b>	<b>0.128</b>	—	0.001	<b>0.005</b>	0.001	0.002		−0.017	0.030
			<b>0.409</b>	<b>0.356</b>	<b>0.368</b>	—	0.004	<b>0.019</b>	0.004	0.007		−0.012	<b>0.064</b>
	NC	36	<b>0.101</b>	<b>0.103</b>	<b>0.116</b>	—	−0.001	0.002	0.001	0.002	0.000		<b>0.043</b>
			<b>0.360</b>	<b>0.314</b>	<b>0.328</b>	—	−0.004	0.008	0.004	0.008	0.000		<b>0.048</b>
	Va	33	<b>0.092</b>	<b>0.098</b>	<b>0.117</b>	—	−0.003	0.006	0.005*	−0.004	0.004	−0.001	
			<b>0.347</b>	<b>0.313</b>	<b>0.343</b>	—	−0.012	0.024	0.019*	−0.016	0.015	−0.004	

Bold  $\Phi_{ST}$ ,  $\theta$  and  $\theta'$  have  $P$ -values  $<0.05$  after false discovery rate (FDR) adjustments (Benjamini & Hochberg 1995).

observed among three of the 135 locus–location combinations (eight loci including KRU22 in My and KRU22 in Ré and Va), and the global test (Fisher's method) across locations for each locus showed significant departure from HWE for loci KRU06 and KRU22 (more likely due to unresolvable scoring problems than the presence of null allele). In addition, linkage disequilibrium was detected between loci KRU06 and KU014 in a global test. Based on these results, loci KRU06 and KRU22 were excluded from further analyses. After these two loci were removed, the global test across loci for each location deviated from HWE for My only (Table S3, Supporting information).

Pairwise  $\theta$  values using microsatellite data between allopatric populations of distinct mtDNA lineages (from 0.091 to 0.136) indicated significant genetic differentiation and were one to two orders of magnitude higher than at the intra-mtDNA lineage level (−0.178 to 0.008). Pairwise  $\theta$  showed genetic homogeneity within each mtDNA lineage after Bonferroni correction. In My, where populations from lineages 1 and 2 are found in sympatry, the genetic variability estimated between those populations was

**Table 5** Fu's  $F$  and Tajima's  $D$  for *Kuhlia rupestris* per sampling location around the Coral Sea

Locations	ND4		ATP	
	$F$	$D$	$F$	$D$
Va	<b>-2.13</b>	-1.08	<b>-3.19</b>	<b>-1.51</b>
NC	<b>-2.84</b>	-1.38	<b>-4.41</b>	<b>-1.69</b>
QldN	<b>-2.53</b>	-1.41	0.07	<b>-1.51</b>
QldC	-0.26	-0.58	0	0
QldS	-0.52	-0.69	0	0

Location names are given in Fig. 1.

Bold values have  $P$ -values  $<0.05$  after FDR adjustments (Benjamini & Hochberg 1995).

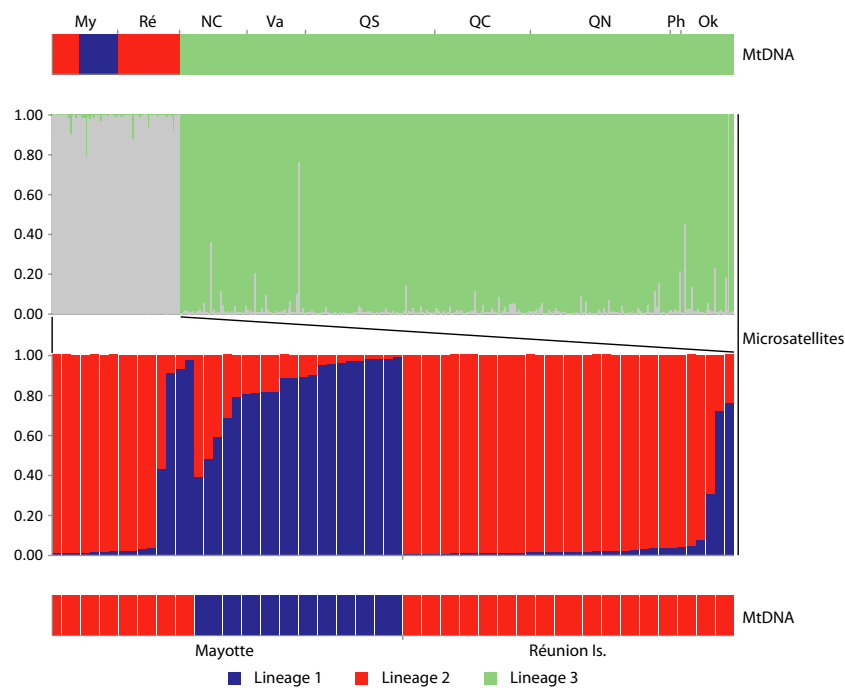
about half of what was reported between allopatric populations of a distinct lineage (Table 4).

The first round of STRUCTURE analyses supported the existence of two clusters within the whole data set based on the  $\Delta K$  method (Evanno *et al.* 2005) (Table S4, Supporting information). The posterior probabilities corresponding to the assignment of individuals ( $q$ ) were all above 0.80 except for seven individuals, one from the Indian Ocean and six from the Pacific (Fig. 4). The first

cluster was exclusively made of individuals from locations in the Indian Ocean (grey bar). The second cluster was exclusively made of individuals from locations in the Pacific Ocean [green bar, hereafter denoted as cluster 3 (as it corresponded to the mtDNA lineage 3)]. During the second round of analysis, no genetic structure was found within the Pacific Ocean samples, but two distinct genetic groups were distinguished in the Indian Ocean. The first Indian group was composed of 20 individuals from My (blue bar, hereafter denoted as cluster 1). The second Indian group was composed of 11 and 32 individuals from My and Ré, respectively (red bar, hereafter denoted as cluster 2). Assignment to either cluster 1 or cluster 2 was unclear for nine individuals (Fig. 4). The third round of analysis did not find any further clustering within any of the Indian or Pacific Ocean groups. Therefore, the finest population structure detected by STRUCTURE consisted of three populations, one in the Pacific Ocean and two in the Indian Ocean.

#### Microsatellites vs. mtDNA

The results of the clustering analyses based on microsatellites were roughly congruent with the mtDNA



**Fig. 4** Correspondence between mtDNA lineages and nuclear STRUCTURE clusters. First and last rows are the distribution of the 3 mtDNA lineages from nine collection locations of *Kuhlia rupestris* from the Indo-Pacific and from collection locations from the Indian Ocean, respectively (see Table 1 for location names). Second row is the result of the first round of the STRUCTURE assignment test based on the microsatellite data. Third row is the result of the second round of assignment test for the samples from the Indian Ocean only. Each cluster ( $K$ ) is designated by a different colour with vertical bars representing individuals and the proportion of a bar assigned to one colour representing the posterior probability that an individual is assigned to that cluster (the colours were attributed according to the mtDNA lineage roughly corresponding to each cluster, grey colour was used for the Indian cluster found in the first round before it was split into clusters 1 and 2 in the second round). The results are only given for  $K = 2$ .



lineages (Fig. 4). Except five individuals classified as ambiguous, all individuals with mtDNA of lineage 1 were assigned to the cluster 1. Moreover, 43 of 50 individuals assigned to the cluster 2 harboured mtDNA representing lineage 2. Among the seven remaining samples, four had ambiguous microsatellite genotypes and three were assigned to the cluster 1. Two different ND4 haplotypes were found among those three individuals (ATP was not sequenced). All 313 individuals harbouring mtDNA from the Pacific lineage were assigned to the cluster 3, and samples with the Pacific lineage mtDNA were not found outside the region.

#### Larval dispersal in the Coral Sea

According to our simulations of larvae dispersal in the Coral Sea, it was clear that assuming a PLD of 58 days and a passive dispersion in the upper 5, 15 and 25 m of the water column, less than 1% (if any) larvae released during the simulation period (from 1994 to 1999) from Va, NC, QN and QC populations (i.e. healthy) will be able to reach the coast of QS. Exceptions are simulations of larvae dispersal at depth 15 m released from QC and at depth 25 m released from QC and QN, for which up

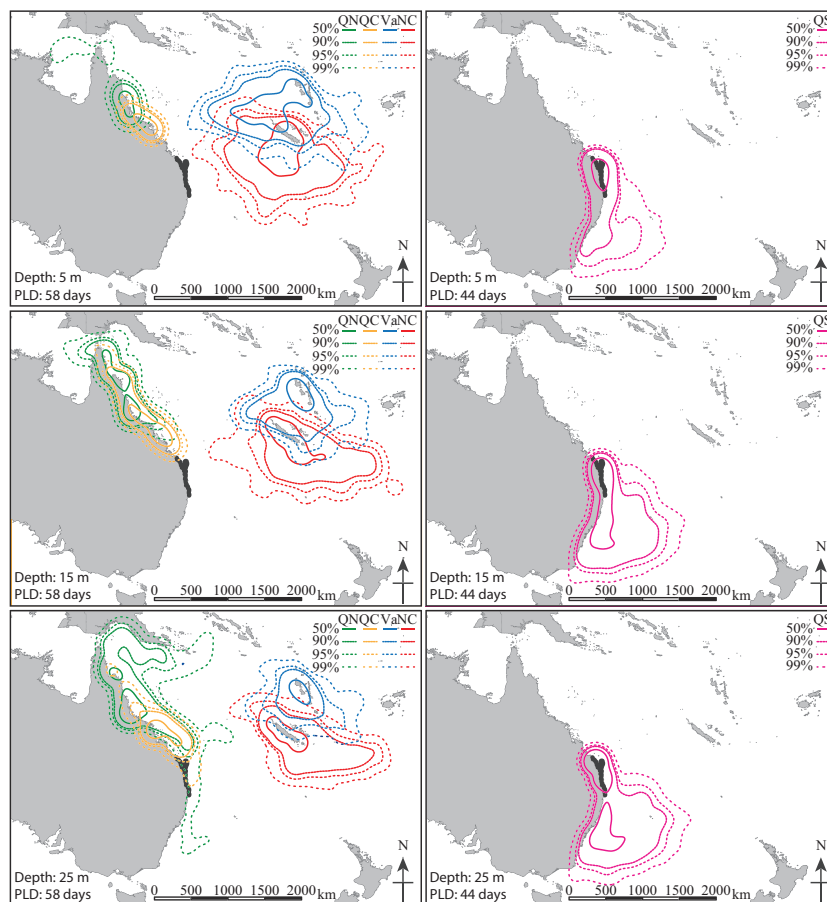
to 5% of larvae per generation may have reach the coast of QS. The extent of larval dispersal along the coast of Queensland seemed to increase with depth of particle dispersion, whereas the connectivity between NC/Va and QS showed the opposite pattern (Fig. 5).

The most probable scenario for larvae released in QS is that they will drift southward or stay along the coast of southern Queensland, and very few will disperse northward and reach QC (Fig. 5).

#### Discussion

##### *Population structure of Kuhlia rupestris and connectivity through evolutionary time*

In the present study, the genetic results suggested that the dispersal capabilities of *K. rupestris* were not sufficient to maintain intraspecific genetic homogeneity through evolutionary time. However, *K. rupestris* populations were homogenous across a fairly large geographical distribution in the Pacific. This species comprises three populations (or stocks), one in the Pacific Ocean, one in the Indian Ocean and a third with a more restricted distribution in My, which is sympatric



**Fig. 5** Kernels of dispersion (50%, 90%, 95% and 99%) for *Kuhlia rupestris* larvae in the Coral Sea using the web-tool Connie 2 according to different depths (5, 15 and 25 m), pelagic larval duration (44 or 58 days) and locations for particle release (Queensland North, Queensland Central, Queensland South, New Caledonia and Vanuatu). Bold line indicates the extent of the QS region with low-density populations.

with the widespread Indian Ocean stock. Both mitochondrial genes and microsatellite loci identify two co-occurring populations at My. It is unlikely that the restricted population only occurs at My because this island is close to the other islands of the Comoros Archipelago. Moreover, sampling in Mg was not exhaustive and several areas within the distribution range of *K. rupestris* were not sampled. For example, the adjacent eastern coast of Africa or Sumatra were not sampled here, which proved to be genetically connected with Western Indian populations in the diadromous giant mottled eel *Anguilla marmorata* (Minegishi *et al.* 2008). The population found in all locations from the Indian Ocean appeared to be fully panmictic, as no significant structure was found between locations, either for mtDNA or for microsatellites. In the Pacific, the differentiation observed between samples from the northern locations (Ok and Ph) and southern locations (QS, QC, QN, Va and NC) based on mtDNA indicated a more subtle structure, which was not readily detected by nuclear markers. Greater population differentiation using mitochondrial than nuclear markers is often observed, as mtDNA drifts to fixation more rapidly than nuclear genes (Zink & Barrowclough 2008).

#### *Larvae dispersal of K. rupestris in the Coral Sea*

Our analyses indicated that the short PLD of *K. rupestris* fails to maintain a homogeneous genetic pool across its whole distributional range, but can connect populations on a subocean basin scale over evolutionary times. Furthermore, the larval dispersal modelling we applied in one of the subocean basins (i.e. Coral Sea) suggested that the actual connectivity of this species might be even more limited over one generation. Particles released in NC, Va, QN or QC may be able to reach the southern coast of Queensland, but this is rare. Also, the PLD was set at 58 days, the maximum value found for *K. rupestris* by Feutry *et al.* (2012c); thus, the real proportion of larvae reaching the coast of QS would be a subfraction of the numbers presented here. Our simulations suggested that given the actual currents and assuming a passive dispersion of larvae in the upper 5, 15 and 25 m of the water column, the southern population in Queensland mostly depends on self-recruitment. Five per cent of larvae released from QN and QC populations may reach QS if the larvae were advected at a depth of 15 or 25 m. However, the model of larval dispersal has limitations as it predicts the presence of adult populations in rivers flowing into the Gulf of Carpentaria. The model predicts that larvae from eastern Queensland would be advected into and beyond the Torres Strait. However, adult populations are not known west of Torres Strait despite the presence of

suitable habitat on the western side of Cape York. Moreover, our simulations suggest a strong connectivity between NC and Va, which is not consistent with the genetic data. Shanks (2009) demonstrated that for PLDs longer than 10 h, both genetic methods and passive larval dispersion models generally overestimate dispersal. He attributed this difference to the behaviour of larvae retarding their movement. Similar retention mechanisms may also occur in *K. rupestris*, and present connectivity is probably even more limited than suggested by our data.

#### *Systematic status of K. rupestris*

The population structure revealed in this study by both nuclear and mitochondrial markers implies that *K. rupestris* may be a complex of separate species or subspecies more than only one widespread species. Randall & Randall (2001) reviewed the systematic of genus *Kuhlia*, but they focused on Pacific species and they only included one specimen from the Indian Ocean (Mauritius) in their analysis of *K. rupestris*.

In the absence of morphological data to identify new species and because the variation in ATP and ND4 genes is similar to that of the cytochrome c oxidase subunit I gene (COI, Feutry *et al.* 2013), one can consider the 'barcoding method' based on the mtDNA distances reported in this study (Hebert *et al.* 2003b). ATP and ND4 interlineage mean divergences (1.9–2.3% and 1.5–2.6%, respectively) are slightly below the 3% threshold suggested by Hebert *et al.* (2003a), but Meyer & Paulay (2005) demonstrated the value would create many false negatives (i.e. rejected real new species) in some taxa. The Kuhliidae family is one of them, for example the mean divergence between *K. marginata* and *K. malo* is as low as 1.5% for the ND4 gene (Feutry P, unpublished data) or 1% for the COI gene (Feutry *et al.* 2013). Another method to screen for novel taxa is to compare inter- and intralineage mean divergences (Hebert *et al.* 2003a). In this study, the ATP and ND4 interlineage divergences exceed intralineage mean divergences (0.2–0.3% and 0.2–0.4%, respectively) by a ratio of 4–13, which falls into the range of values that indicates cryptic species in other groups (Hebert *et al.* 2004; Meyer & Paulay 2005). Another aspect in favour of multiple cryptic species within *K. rupestris* is the ecological difference found between two populations of *K. rupestris*. Feutry *et al.* (2012c) reported that the PLD of individuals from Ré in the Indian Ocean was significantly lower than the PLD of individuals from NC in the Pacific ( $37.3 \pm 4.7$  days in the Indian Ocean and  $44.3 \pm 6.7$  days in the Pacific, Feutry *et al.* 2012c). The congruency between mtDNA, nuclear markers and ecological data strongly suggests *K. rupestris* is a complex

of cryptic species (Dasmahapatra & Mallet 2006), although the matter warrants further investigation including morphological analysis. If two species are sympatric in *My*, further studies should investigate whether secondary contact is occurring after divergence, or whether occurred through the loss of connectivity, or possible ecological speciation is occurring *in situ*.

#### *Survival on self-recruitment*

This study has implications beyond the ecology of *K. rupestris* only. As for many other diadromous species inhabiting mainly tropical insular systems, the marine larval phase of *K. rupestris* provides it with the ability to invade vacant habitats following perturbations. Indeed, in case of perturbations lasting longer than their maximal PLD, the persistence of those species in tropical islands depends on external recruitment. An island has to be on a dispersal route to prevent from extinction. The population expansion revealed by the analysis of demographic history of jungle perch in Va, NC and QN suggested that this sequence of extinction/colonization has occurred naturally in the past. In the wet tropics of northeastern Australia, steep coastal mountain streams are similar to ephemeral tropical insular streams. These streams are faunally more similar to distant Pacific island communities, than to the nearby faunas of large rivers on the Australian mainland (Thuesen *et al.* 2011). In QC and QS, where no signs of population expansion were detected, *K. rupestris* is found in perennial streams. We hypothesize this is why the population was able to persist with low external recruitment. When assessing the potential impact of human-mediated perturbation, we recommend that actual dispersal should be investigated before assuming that recovery will occur easily for diadromous species.

#### **Conclusion**

This study combined population genetic data with an estimation of larvae dispersal using PLD data and current modelling. To the best of our knowledge, this study is the first attempt to combine genetic and modelling estimates of connectivity for a diadromous species. These estimates provided insights into the connectivity of *Kuhlia rupestris* at different timescales and thus were complementary. This study is also an example of how combining these methods can help for the development of management plans for threatened populations.

At an evolutionary timescale (i.e. genetic data), despite a short PLD, *K. rupestris* has been able to colonize remote islands and to extend its distribution to two oceans. However, its short PLD was not sufficient to maintain its genetic homogeneity and speciation is

probably occurring (or has already occurred) across its distribution range. Three different management units (subspecies?) should be considered, possibly four, considering that mtDNA seemed to indicate a barrier to gene flow between the north and the south Pacific. Importantly, genetic data indicated a strong connectivity between the locations in the Coral Sea. Based on these results alone, the best explanation for the absence of recovery in QS is that despite restoration, the habitat quality is still below the species requirements. However, the larval dispersal modelling brought additional insight into the connectivity of *K. rupestris*. At a real timescale, the modelling approach suggested locations across the Coral Sea to be poorly connected over one generation. If this reflects reality, low external recruitment from adjacent populations rather than the quality of restored habitats could explain why recovery has not occurred yet in QS. If managers want to implement measures with 'rapid' effect on populations (i.e. within years or decades), a fine spatial scale may be considered (i.e. each river or watershed). In QS, recovery will probably occur naturally, but it will take a long time. In order to accelerate the recovery process, we would suggest restoring as many habitats as possible along the coast of Queensland in order to maintain a continuum of healthy population and increase the connectivity between locations.

#### **Acknowledgements**

Sampling in Madagascar was part of the DIAMSOI program funded by the 'Institut Français de la Biodiversité' and was partly financed by the BIONECONAL ANR in New Caledonia. We are grateful to Joe Aride from Madagascar National Parks and the Manager of Masoala National Park. We thank the New Caledonian North Province (J. J. Cassan and C. Flouhr) for allowing sampling. We thank Gérard Marquet, Tony Robinet, Roger and staff members of ARDA Henri Grondin, Pierre Valade and Marine Richardson for their help in the field in the Indian Ocean. We are grateful to Laura Taillebois, Clara Lord and Dona Kalfatak, who helped to sample in New Caledonia and Vanuatu. We thank Michael Hutchison for coordinating sample collection in Australia, and the fisheries biologist and scientists who provided fin clips for genetic analyses. We also thank Zohrah Haji Sulaiman and Ken Maeda for the collection of samples from the Philippines and Okinawa, respectively. We are grateful to Raewyn Street for laboratory analyses in Queensland. We thank the three anonymous reviewers for their helpful comments on earlier version of the manuscript. Jenny Ovenden thanks the Queensland Government and the Museum of Natural History, Paris, for support that facilitated this collaboration. Pierre Feutry thanks the North Australia Marine Research Alliance and the Marine Biodiversity Hub, a collaborative partnership supported through funding from the Australian Government's National Environmental Research Program for financial support when finalizing the manuscript.

## References

- Arèvalo E, Davis SK, Sites JW (1994) Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. *Systematic Biology*, **43**, 387–418.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.
- Beyer HL (2004) Hawth's analysis tools for ArcGIS. Available at: <http://www.spatialecology.com/htools> (accessed 15 April 2012).
- Broderick D, Ovenden JR, Buckworth RC, Newman SJ, Lester RJG, Welsh DM (2011) Genetic population structure of grey mackerel (*Scomberomorus semifasciatus* Macleay, 1883) in northern Australia. *Journal of Fish Biology*, **79**, 633–661.
- Case TJ, Bolger DT, Richman AD (1992) Reptilian extinctions: the last ten thousand years. In: *Conservation Biology: The Theory and Practice of Nature Conservation, Preservation and Management* (eds Fiedler PL & Jain SK), pp. 90–125. Chapman and Hall, New York.
- Condie S (2005) Web site on marine connectivity around Australia. *Eos Transactions, American Geophysical Union*, **86**, 246.
- Cowen RK, Sponaugle S (2009) Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, **1**, 443–466.
- Craig DA (2003) Geomorphology, development of running water habitats, and evolution of black flies on Polynesian islands. *BioScience*, **53**, 1079–1093.
- Dasmahapatra KK, Mallet J (2006) Taxonomy: DNA barcodes: recent successes and future prospects. *Heredity*, **97**, 254–255.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Feutry P, Castelin M, Grondin H, Cruaud C, Couloux A, Keith P (2012a) First record of *Kuhlia sauvagii* Regan, 1913 (Perciformes) in Mayotte and Réunion islands, Western Indian Ocean. *Cybius*, **36**, 493–494.
- Feutry P, Tabouret H, Maeda K, Pécheyrac C, Keith P (2012b) Diadromous life cycle and behavioural plasticity in freshwater and estuarine Kuhlidae species (Teleostei) revealed by otolith microchemistry. *Aquatic Biology*, **15**, 195–204.
- Feutry P, Valade P, Ovenden JR, Lopez PJ, Keith P (2012c) Pelagic larval duration of two diadromous species of Kuhlidae (Teleostei: Percoidae) from Indo-Pacific insular systems. *Marine and Freshwater Research*, **63**, 397–402.
- Feutry P, Castelin M, Ovenden JR *et al.* (2013) Evolution of diadromy in fish: insights from a tropical genus (*Kuhlia* spp.). *The American Naturalist*, **181**, 52–63.
- Frankham R (1997) Do island populations have less genetic variation than mainland populations? *Heredity*, **78**, 311–327.
- Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology*, **84**, 2989–2998.
- Gray CA, Kingsford MJ (2003) Variability in thermocline depth and strength, and relationships with vertical distributions of fish larvae and mesozooplankton in dynamic coastal waters. *Marine Ecology Progress Series*, **247**, 211–224.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
- Hebert PDN, Ratnasingham S, de Waard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, S96–S99.
- Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biology*, **2**, e312.
- Hedgecock D, Barber PH, Edmands S (2007) Genetic approaches to measuring connectivity. *Oceanography*, **20**, 70–79.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.
- Hogan AE, Nicholson JC (1987) Sperm motility of sooty grunter *Hephaestus fuliginosus* (Macleay), and jungle perch, *Kuhlia rupestris* (Lacépède), in different salinities. *Australian Journal of Marine and Freshwater Research*, **38**, 523–528.
- Huelsensbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.
- Hurwood DA, Hugues JM (1998) Phylogeography of the freshwater fish, *Mogurnda adspersa*, in streams of northeastern Queensland, Australia: evidence for altered drainage patterns. *Molecular Ecology*, **17**, 1507–1517.
- Hutchison M, Simpson R, Elizur A, Willet D, Collins A (2002) Restoring jungle perch *Kuhlia rupestris* recreational fisheries to South-east Queensland. Internal Report, Department of Primary Industries, Queensland, 10.
- Inoue JG, Miya M, Tsukamoto K, Nishida M (2001) A mitogenomic perspective on the basal Teleostean phylogeny: resolving higher-level relationships with longer DNA sequences. *Molecular Phylogenetics and Evolution*, **20**, 275–285.
- Jobb G, von Haeseler A, Korbinián S (2004) TREFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology*, doi: 10.1186/1471-2148-4-18.
- Keith P (2003) Biology and ecology of amphidromous Gobiidae of the Indo-Pacific and Caribbean regions. *Journal of Fish Biology*, **63**, 831–847.
- Keith P, Galewski T, Cattaneo-Berberi G, Hoareau T, Berberi P (2005) Ubiquity of *Sicyopterus lagocephalus* (Teleostei: Gobiidae) and phylogeography of the genus *Sicyopterus* in the Indo-Pacific area inferred from mitochondrial cytochrome b gene. *Molecular Phylogenetics and Evolution*, **37**, 721–732.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.
- Leis JM, Van Herwerden L, Patterson HM (2011) Estimating connectivity in marine fish populations: what works best? *Oceanography and Marine Biology*, **49**, 205–246.
- MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution*, **17**, 373–387.
- McDowall RM (1988) *Diadromy in Fishes: Migrations Between Freshwater and Marine Environments*. Croom Helm, London Timber Press, Portland.

- Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.
- Merrick JR, Schmida GE (1984) *Australian Freshwater Fishes: Biology and Management*. Merrick J.R., Sydney.
- Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **2**, e422.
- Minegishi Y, Aoyama J, Tsukamoto K (2008) Multiple population structure of the giant mottled eel, *Anguilla marmorata*. *Molecular Ecology*, **17**, 3109–3122.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Peters MB, Ovenden JR, Broderick D, Lance SL, Hagen C, Glenn TC (2009) Fifteen microsatellite loci for the jungle perch, *Kuhlia rupestris*. *Molecular Ecology Resources*, **9**, 1467–1469.
- Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Phylogenetics and Evolution*, **25**, 1253–1256.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Radtke R, Kinzie RA, Folsom S (1988) Age at recruitment of Hawaiian freshwater gobies. *Environmental Biology of Fishes*, **23**, 205–213.
- Randall JE, Randall HA (2001) Review of the fishes of the genus *Kuhlia* (Perciformes: Kuhlidae) of the Central Pacific. *Pacific Science*, **55**, 227–256.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Rosenberg NA, Pritchard JK, Weber JL *et al.* (2002) Genetic structure of human populations. *Science*, **298**, 2342–2343.
- Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. *Biological Bulletin*, **216**, 373–385.
- Smith KA (2000) Active and passive dispersal of *Centroberyx affinis* (Berycidae) and *Gonorynchus greyi* (Gonorynchidae) larvae on the Sydney shelf. *Marine and Freshwater Research*, **51**, 229–234.
- Smith GC, Covich AP, Brasher AMD (2003) An ecological perspective on the biodiversity of tropical island streams. *BioScience*, **53**, 1048–1051.
- Taillebois L, Maeda K, Vigne S, Keith P (2012) Pelagic larval duration of three amphidromous Sicydiinae gobies (Teleostei: Gobioidae) including widespread and endemic species. *Ecology of Freshwater Fish*, **21**, 552–559.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.
- Thuesen PA, Ebner BC, Larson H *et al.* (2011) Amphidromy Links a Newly Documented Fish Community of Continental Australian Streams, to Oceanic Islands of the West Pacific. *PLoS ONE*, **6**, e26685.
- Vähä J-P, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.
- Vähä JP, Erkinaro J, Niemela E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*, **16**, 2638–2654.
- 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**, 535–538.
- Viard F, Ellien C, Dupont L (2006) Dispersal ability and invasion success of *Crepidula fornicata* in a single gulf: insights from genetic markers and larval-dispersal model. *Helgolander Marine Research*, **60**, 144–152.
- Vitousek PM (1988) Diversity and biological invasions of oceanic islands. In: *Biodiversity* (eds Wilson EO & Peters FM), pp. 181–189. National Academy Press, Washington, DC.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

---

P.F., P.K. and J.O. designed the research; P.F., D.B., A.V. and J.L. performed the research; P.F. and J.O. analysed the data; P.F. wrote the manuscript; P.K. and J.O. commented on earlier versions of the manuscript.

---

### Data accessibility

DNA sequences: EMBL accessions HF542116–HF542796.

Final DNA sequence alignments, microsatellite genotypes, raw results from larval dispersal modelling analysis and sampling details for each individual:

DRYAD entry doi: 10.5061/dryad.f81sf

### Supporting information

Additional supporting information may be found in the online version of this article.

**Table S1** Variable positions (numbered) in the 795 base pairs of the mitochondrial ATP segment of *Kuhlia rupestris*.

**Table S2** Variable positions (numbered) in the 838 base pairs of the mitochondrial ND4 segment of *Kuhlia rupestris*.

**Table S3** Genetic variability at 15 microsatellite loci of *Kuhlia rupestris* collected in nine locations from the Indo-Pacific region.

**Table S4** Log probability and  $\Delta K$  (Evanno *et al.* 2005) for each number of clusters ( $K$ ) calculated with the Bayesian based-model method, as implemented in the software STRUCTURE (Pritchard *et al.* 2000).

**Fig. S1** Phylogenetic tree inferred from 21 ND4 sequences haplotypes of *K. rupestris*.

**Fig. S2** Network of *K. rupestris* ND4 sequences from ten locations in the Pacific and Indian Oceans.