

# SCIENTIFIC REPORTS

OPEN

## Genomic DNA variation confirmed *Seriola lalandi* comprises three different populations in the Pacific, but with recent divergence

H. K. A. Premachandra<sup>1</sup>, Fabiola Lafarga-De la Cruz<sup>2</sup>, Yutaka Takeuchi<sup>3</sup>, Adam Miller<sup>4</sup>, Stewart Fielder<sup>5</sup>, Wayne O'Connor<sup>5</sup>, Celine H. Frère<sup>1</sup>, Nguyen Hong Nguyen<sup>1</sup>, Ido Bar<sup>1</sup> & Wayne Knibb<sup>1</sup>

Captive breeding programs and aquaculture production have commenced worldwide for the globally distributed yellowtail kingfish (*Seriola lalandi*), and captive bred fingerlings are being shipped from the Southern Hemisphere to be farmed in the Northern Hemisphere. It was recently proposed that Pacific *S. lalandi* comprise at least three distinct species that diverged more than 2 million years ago. Here, we tested the hypothesis of different “species” in the Pacific using novel genomic data (namely single nucleotide polymorphisms and diversity array technology markers), as well as mtDNA and DNA microsatellite variation. These new data support the hypothesis of population subdivision between the Northeast Pacific, Northwest Pacific and South Pacific, and genetic divergence indicates restriction to the gene flow between hemispheres. However, our estimates of maximum mtDNA and nuclear DNA divergences of 2.43% and 0.67%, respectively, were within the ranges more commonly observed for populations within species than species within genera. Accordingly our data support the more traditional view that *S. lalandi* in the Pacific comprises three distinct populations rather than the subdivisions into several species.

Commercial aquaculture of yellowtail kingfish (*Seriola lalandi* Valenciennes, 1833, Carangidae) has recently commenced around the world. Global aquaculture production of the *Seriola* species exceeds 150,000 tonnes per year<sup>1</sup>, where yellowtail kingfish contributes to the second highest production within the *Seriola* species<sup>2</sup>. At present yellowtail kingfish aquaculture is practised either in sea cages or in land-based facilities and in many different countries including Japan<sup>3</sup>, Australia<sup>4,5</sup>, New Zealand<sup>6</sup>, USA<sup>7</sup>, Mexico<sup>8</sup>, Chile<sup>9</sup>, Netherlands<sup>10</sup> and South Africa<sup>11</sup>. A recent study reported the potential expansion to further countries, such as China and Spain<sup>2</sup>. Recent population genetic studies have proposed that globally distributed *S. lalandi* is a complex of at least three closely related species with a restriction of gene flow across the equator, and with a divergence of about two millions years<sup>8,12</sup>. If so, it could be of concern that there is international trade of yellowtail kingfish fingerlings and eggs across the geographical boundaries including across the equator from the Southern to Northern Hemisphere<sup>13,14</sup>. For example Australian yellowtail kingfish fingerlings have been exported commercially from the Southern Hemisphere to the Northern Hemisphere (Adam Miller, personal communication) and Chilean yellowtail kingfish fingerlings have been exported across the hemispheres to Mexico for commercial aquaculture in 2013<sup>15</sup>. These exchanges or movements of potentially different breeding material have the potential to impact not only local natural wild populations adapted to local regions<sup>16–18</sup> but also have the potential produce genetic incompatibilities from aquaculture crosses<sup>19,20</sup>. Indeed, a significant reduction of fitness and viability in F<sub>2</sub> generation as a consequences of hybrid breakdown of cichlid fish was reported by Stelkens *et al.*<sup>21</sup>, and others reported a significant decrease of

<sup>1</sup>Genecology Research Centre, Faculty of Science, Health, Education, and Engineering, University of the Sunshine Coast, Maroochydore, DC, QLD 4558, Australia. <sup>2</sup>Ensenada Center for Scientific Research and Higher Education (CICESE), Ensenada, B.C., Mexico. <sup>3</sup>Division of Fisheries Resource Sciences, Faculty of Fisheries, Kagoshima University, Shimoarata 4-50-20, Kagoshima city, 890-0056, Japan. <sup>4</sup>Clean Seas Tuna Limited, 7 North Quay Boulevard, Port Lincoln, SA, 5606, Australia. <sup>5</sup>Department of Primary Industries, Port Stephens Fisheries Institute, Taylors Beach Road, Taylors Beach, NSW 2316, Australia. Correspondence and requests for materials should be addressed to W.K. (email: [wknibb@usc.edu.au](mailto:wknibb@usc.edu.au))

Variable site				6	9	24	45	51	81	117	126	129	138	225	261	267	288	348	366	384	387	390	459	468	495	498	501	508	516	525	582
Region	Haplo-type	Fre-quency	GenBank accession no.	T	A	A	T	A	A	A	G	T	C	T	A	G	C	C	A	A	T	T	T	A	A	A	T	T	C	A	G
				NW	Hap_1	4.92%	MF167270	.	.	.	.	.	G	.	.	.	T	.	.	A	.	.	.	G	.	C	C	T	G	.	C
	Hap_2	1.23%	MF167271	.	.	G	.	.	G	.	.	.	T	.	.	A	.	.	.	G	.	C	C	T	G	.	C	C	T	.	A
NE	Hap_3	29.92%	MF167272	.	.	.	.	<u>G</u>	.	.	.	.	T	.	G	.	.	<u>T</u>	.	G	.	.	C	T	G	<u>G</u>	.	C	T	G	.
	Hap_4	9.02%	MF167273	.	.	.	.	<u>G</u>	.	.	.	.	.	.	G	.	.	<u>T</u>	.	G	.	.	C	T	G	<u>G</u>	.	C	T	G	.
	Hap_5	1.23%	MF167274	.	.	.	.	<u>G</u>	.	.	.	.	T	C	.	.	.	<u>T</u>	.	G	.	.	C	T	.	<u>G</u>	.	C	T	G	.
	Hap_6	1.64%	MF167275	.	.	.	.	<u>G</u>	.	.	.	.	T	.	.	.	.	<u>T</u>	.	G	.	.	C	T	G	<u>G</u>	.	C	T	G	.
SP	Hap_7	0.82%	MF167276	.	G	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
	Hap_8	44.67%	MF167277	.	G	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	Hap_9	1.64%	MF167278	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	Hap_10	0.41%	MF167279	.	G	.	.	.	.	G	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
	Hap_11	2.46%	MF167280	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	Hap_12	0.41%	MF167281	.	G	.	.	.	.	G	.	C	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	Hap_13	0.41%	MF167282	.	G	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
	Hap_14	0.41%	MF167283	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
Hap_15	0.82%	MF167284	C	.	.	C	.	.	G	.	.	T	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	

**Table 1.** Variable sites of haplotypes based on mtDNA COI sequence. Unique nucleotide positions which separate the Northern and Southern Hemispheres, the Northwest Pacific from the others and the Northeast Pacific from others were represented in bold, italic and underlined text, respectively (NW = Northwest Pacific; NE = Northeast Pacific; SP = South Pacific).

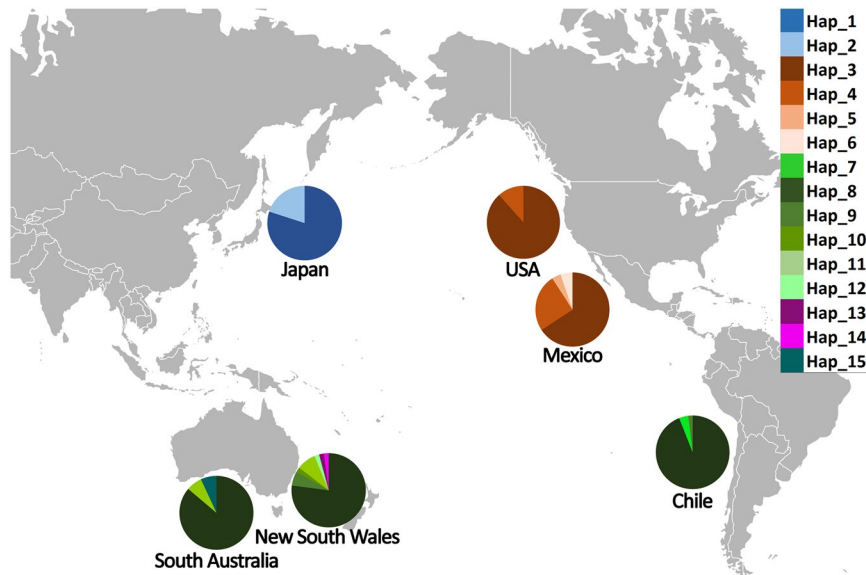
fitness with increasing genetic distance<sup>22</sup> and significant  $F_2$  hybrid breakdown<sup>23</sup> in the copepod *Tigriopus californicus*. Given the worldwide distribution of yellowtail kingfish and commercial movement of fingerlings around the world it is important to determine with accuracy the genetic relatedness and consider the possible genetic “compatibility” of different populations around the world.

Most previous population genetic studies of *Seriola* populations applied microsatellite markers and mitochondrial DNA (mtDNA) sequence analysis, or combination of mtDNA and a few nuclear genes for the assessment of genetic diversity and species delimitation<sup>8, 12, 14, 24</sup>. However, the effectiveness of mtDNA and microsatellite loci data in population genetic analysis is questionable due to their low representation of genomic information. Moreover, mtDNA represents only maternal inheritance and may provide a different image of population structure than that provided by biparentally inherited nuclear DNA, as evident in many sea turtle species and white sharks *Carcharodon carcharias*<sup>25, 26</sup>. While other potential limitations of mtDNA relate to clonal inheritance, neutrality and erratic evolutionary rates suggested that mtDNA may no longer be considered as an ideal marker for evolutionary studies<sup>27</sup>, it still represents a powerful tool for phylogenetic studies, particularly when relating new research to previously published data. The use of DNA microsatellite data has also been challenged for genetic analyses on the grounds of its extremely high mutation rate, size homoplasy, irregularities in mutation process and degradation over time<sup>28, 29</sup>; it is questionable how well evolutionary rates for DNA microsatellite sequences reflect those across the whole genome<sup>30</sup>. Likewise, it is of concern how well just a few nuclear genes reflect whole genome evolution and divergence due to the possibility of locus specific anomalous evolutionary rates<sup>31, 32</sup>. Therefore, to address these concerns arising from the use of traditional markers to resolve phylogenetics, we carried out partial genome sequencing of *S. lalandi* from different Pacific regions to reassess the genetic structure of Pacific *S. lalandi* populations. This allowed comparison of diversity estimates obtained by traditional methods (mtDNA sequences and DNA microsatellite alleles from new data of this report and previously published mtDNA data) with new estimates, provided for the first time, from partial genomic sequencing.

## Results

**Mitochondrial DNA diversity.** Analysis of the mtDNA cytochrome oxidase subunit I (COI) fragments sequenced in this study identified 15 haplotypes containing a total of 28 polymorphic sites (GenBank accession numbers: MF167270 – MF167284; Table 1). From the previously published data, there were 29 haplotypes containing a total of 38 polymorphic sites. The combined new and published data recorded 31 haplotypes containing a total of 41 polymorphic sites (Supplementary Table 1).

There appeared to be marked differences in presence of particular mtDNA haplotypes among the regions of the Southern Hemisphere, Northeast (NE) Pacific and Northwest (NW) Pacific (Fig. 1), but with broad similarities within regions. No haplotypes were shared between regions. Each region (NE Pacific, NW Pacific and South Pacific) shared one major common haplotype and also shared less common haplotypes. Thus one haplotypes were shared between South Pacific locations (Chile, New South Wales: NSW and South Australia: SA); similarly two haplotypes were shared between USA and Mexico (Fig. 1). These general patterns remained when the published data were pooled with data of the present study (Supplementary Fig. 1), except the most common haplotype in South Africa was not the most common haplotype in the other Southern Hemisphere samples.



**Figure 1.** mtDNA haplotype distributions using data of the present study (Hap\_1 to Hap\_15) among the sample locations. Maps available from Wikipedia Common web page ([https://commons.wikimedia.org/wiki/File:Blank\\_Map\\_Pacific\\_World.svg#filelinks](https://commons.wikimedia.org/wiki/File:Blank_Map_Pacific_World.svg#filelinks)) were modified under the Creative Commons public license (<https://creativecommons.org/licenses/by-sa/3.0/deed.en>) using Adobe Photoshop CS6 software.

Population	Sample size	mtDNA COI			Microsatellite loci				
		$n_h$	$\pi$ (%)	$h$ ( $\pm$ SD)	$g$ ( $\pm$ SD)	$N_a$	$H_o$	$H_e$	$F_{IS}$
Japan	15	2	0.060	$0.343 \pm 0.02$	$0.800 \pm 0.43$	9.778	0.726	0.800	0.092
USA	26	2	0.040	$0.212 \pm 0.09$	$0.756 \pm 0.40$	8.778	0.859	0.763	-0.125
Mexico	76	4	0.120	$0.507 \pm 0.05$	$0.800 \pm 0.42$	15.222	0.825	0.816	-0.011
Chile	50	3	0.020	$0.117 \pm 0.06$	$0.695 \pm 0.37$	8.222	0.700	0.706	0.008
NSW	48	7	0.120	$0.402 \pm 0.08$	$0.710 \pm 0.37$	10.333	0.759	0.723	-0.050
SA	29	3	0.150	$0.256 \pm 0.10$	$0.692 \pm 0.37$	8.444	0.719	0.712	-0.010

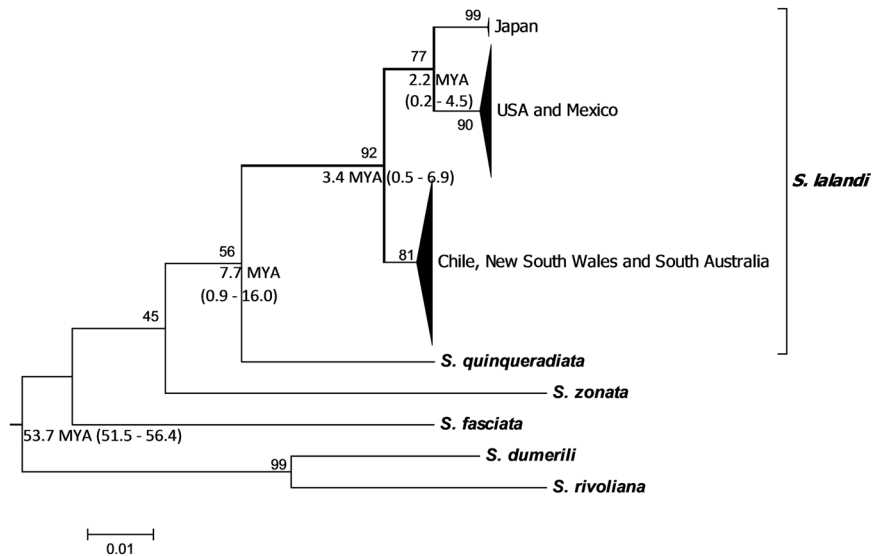
**Table 2.** Estimates of genetic diversity indices from mtDNA and microsatellite data. Abbreviations: Number of haplotypes ( $n_h$ ), percent nucleotide diversity ( $\pi$ ), haplotype diversity ( $h$ ), gene diversity ( $g$ ), mean number of alleles per locus ( $N_a$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), inbreeding coefficient ( $F_{IS}$ ), standard deviation (SD). NSW: New South Wales; SA: South Australia.

	Japan	USA	Mexico	Chile	NSW	SA
Japan		0.134*	0.094*	0.202*	0.194*	0.204*
USA	0.737*		0.059*	0.227*	0.219*	0.226*
Mexico	0.543*	0.056*		0.200*	0.192*	0.198*
Chile	0.823*	0.849*	0.661*		0.001	0.012*
NSW	0.617*	0.673*	0.539*	0.056*		0.009*
SA	0.711*	0.765*	0.584*	0.025*	0.001	

**Table 3.** Pairwise population matrix of  $F_{ST}$  based on mtDNA (below diagonal) and microsatellite (above diagonal) data. \* $P < 0.05$ . NSW: New South Wales; SA: South Australia.

Multiple sequence alignment confirmed there was no deletions or insertions within the analysed COI sequence fragment. There were five nucleotide positions (represented in bold text, Table 1), in which all Northern Hemisphere samples shared a common nucleotide and which was different from that found in the Southern Hemisphere samples. Similarly there were nucleotide differences at five positions in NW Pacific samples that were categorically different from all others samples (represented in italic text, Table 1); likewise there were three categorical nucleotide differences that separated the NE Pacific from all others (represented in underlined text, Table 1). Again the pooled data present plus published showed the same trends (Supplementary Table 1).

Estimates of genetic diversity from COI (nucleotide diversity:  $\pi$  and haplotype diversity:  $h$ ) indicated relative low values in Chile and higher values in Mexico (Table 2). Analysis of molecular variance (AMOVA) indicated



**Figure 2.** The maximum-likelihood (ML) phylogeny considering mtDNA COI sequences of the *Seriola lalandi* from the present report and sequences from other *Seriola* species. The ML bootstrap percentages are shown above the branches. Mean divergence time estimated from Bayesian inferences in BEAST, with 95% highest posterior density (HPD) for importance nodes are indicated below the nodes.

	Japan	USA	Mexico	Chile	NSW	SA
Japan		1.76%	1.78%	2.42%	2.39%	2.42%
USA	na		0.09%	2.38%	2.34%	2.38%
Mexico	0.27%	na		2.33%	2.30%	2.33%
Chile	0.58%	na	0.66%		0.09%	0.01%
NSW	0.57%	na	0.66%	0.22%		0.07%
SA	0.58%	na	0.67%	0.23%	0.23%	

**Table 4.** Estimates of pairwise sequence divergence for SNP-based sequences (below diagonal) and mtDNA COI sequences (above diagonal) between populations. na – data not available. NSW: New South Wales; SA: South Australia.

significant population differentiation ( $F_{ST} = 0.961$ ,  $P < 0.001$ ) among regions where 83% of the total variation were observed among regions (NW Pacific, NE Pacific and South Pacific). Pairwise comparison of  $F_{ST}$  indicated significant differences among populations, except among the Australian populations (Table 3).  $F_{ST}$  values were highest when comparing between hemispheres, between the NW and NE Pacific locations, but were close to zero when comparing among South Pacific locations. Genetic diversity and AMOVA analyses for the pooled data (present plus published) showed the same trends (Supplementary Table 2 and 3), noting the high diversity in the South African samples.

All phylogenetic tree analyses (Neighbour-Joining, Maximum Parsimony, Maximum Likelihood) using data from the present study formed a single major group of the *S. lalandi* sequences, separate from other *Seriola* species, with very high bootstrap support for all the models tested (Fig. 2). However, the single major *S. lalandi* group was split into three subgroups corresponding to geographic location (Fig. 2); subgroups were supported with high bootstrap values ( $>81$ ), where Northern (Japan, USA and Mexico) and Southern (Chile, NSW and SA) Hemisphere locations were separated from each other. Although fish from Japan were placed within the Northern Hemisphere group, they were separated from USA and Mexico (Fig. 2). The same clustering was evident after pooling the published data with the current data (Supplementary Fig. 2), noting that China clustered with Japan and South Africa and New Zealand clustered with Australia and Chile.

Generalized Mixed Yule Coalescent (GMYC) analyses, a likelihood method for delimiting species, indicated that while the Northern and Southern Hemisphere groups have genetic differences, all these lineages, assessed by clade support values, are from a single species group with four clades (single threshold test: null  $\ln L = 220.83$ , max.  $\ln L = 229.41$ , likelihood ratio = 17.16,  $P < 0.001$ ; using pooled data from the current study and published).

Based on molecular dating conducted using Bayesian inferences in BEAST, for the node between Southern and Northern *S. lalandi* populations were estimated to be split 3.4 million years ago (MYA) with 95% highest posterior density (HPD) of 0.5–6.9 MYA again using pooled data from the current study and published (Fig. 2 and Supplementary Fig. 3).

	Japan	Mexico	Chile	NSW	SA
Japan		0.219*	0.272*	0.254*	0.217*
Mexico	0.474*		0.562*	0.456*	0.441*
Chile	0.711*	0.752*		0.017*	0.048*
NSW	0.708*	0.743*	0.006*		0.045*
SA	0.702*	0.741*	0.026*	0.022*	

**Table 5.** Pairwise population matrix of  $F_{ST}$  analysis of SNPs (below diagonal) and DArT markers (above diagonal) data. \* $P < 0.001$ . NSW: New South Wales; SA: South Australia.

Simply considering the percent DNA difference, the maximum COI sequence divergence detected between sampled populations was 2.42% (between Chile and Japan), although samples from the same hemisphere were more closely related (eg. Chile and NSW diverged only by 0.09%, Table 4). The current data of this report and the combined current and published data gave similar results (Supplementary Table 4), noting that the maximum difference for all *S. lalandi* samples was 2.43% which is equal to a divergence of 1.2 MYA between hemispheres using the method of Brown *et al.*<sup>33</sup>, whereas the minimum difference between established but related species, eg. *S. dumerili* and *S. rivoliiana* was 6.49%.

**Microsatellite DNA variation.** Most measures of DNA microsatellite diversity (for example gene diversity, average number of alleles, observed and expected heterozygosity) were broadly similar among the six locations (Table 2). The  $F_{IS}$  index did not deviate significantly from zero in any population, suggesting that Hardy-Weinberg equilibrium (HWE) could be assumed in all populations. However, when considering the differences in terms of allele type and frequencies, there were significant differences among populations (AMOVA test  $F_{ST} = 0.195$ ,  $P < 0.001$ ) and 15% of the molecular variance occurred among populations. Pairwise  $F_{ST}$  comparisons between populations varied from 0.001 to 0.227 for microsatellite data and were statistically significant for most pair combinations ( $P < 0.05$ , Table 3). Higher  $F_{ST}$  values were observed between pairs of populations from different hemisphere locations rather than within hemisphere locations.

Analysis of microsatellite data using “STRUCTURE” software indicated the presence of two genetic clusters ( $K = 2$ ) with a high probability of membership coefficient (Supplementary Fig. 4a). Fish from Northern Hemisphere (Japan, USA and Mexico) were clearly separated from Southern Hemisphere (Chile, NSW and SA) fish populations.

**SNP and DArT markers.** A large proportion of the total variation (73%) of single nucleotide polymorphisms (SNPs) was attributed to among regions (South Pacific, Japan, Mexico; AMOVA test,  $F_{ST} = 0.732$ ,  $P < 0.001$ ), whereas less variation was observed within populations. Pairwise comparison of  $F_{ST}$  values also indicated significant differences among hemispheres and regions but not among populations within Southern Hemisphere where values were close to zero (Table 5).

Principal component analysis (PCA) of SNPs from five sampling locations were sorted into three clusters where Southern Hemisphere locations (Chile, NSW & SA) made one tight cluster and the Mexican population and Japanese population made another two separate clusters, which represent the separation of three regions (Supplementary Fig. 5a). STRUCTURE analysis indicated that the Northern and Southern Hemisphere locations were clearly separated (Supplementary Fig. 4b). DNA divergences for sequence reads were higher (0.57–0.67%) between hemispheres than within hemispheres (0.22–0.27%, Table 4).

AMOVA test for DArT markers (presence/absence of restriction fragments) indicated significant population differentiation ( $F_{ST} = 0.369$ ,  $P < 0.001$ ) among regions. Pairwise comparison of  $F_{ST}$  among populations indicated significant and higher values among the locations from different hemispheres and regions, whereas values were lower (close to zero) among the populations within Southern Hemisphere (Table 5).

PCA of DArT markers indicated presence of two clusters where Southern Hemisphere populations (Chile, NSW & SA) grouped into one cluster and the Mexican and Japanese populations grouped into another (separated by an Eigen vector which contributes to close to 23.5% of the variability in the data; Supplementary Fig. 5b). Similar to the PCA analyses, the STRUCTURE analysis of DArT markers also revealed two clusters ( $K = 2$ ) with high probability of membership coefficient based on hemisphere locations (Supplementary Fig. 4c).

Three DArT markers (considering either their presence or absence) were present only in Japan and not in other populations, similarly, another three markers distinguished the Mexican population. Conversely, no single South Pacific population had markers restricted only to that population, i.e. all markers from a given South Pacific population were found in each of the other South Pacific populations. When combining the populations in the South Pacific and comparing them to the combined Northern Hemisphere (Japan and Mexico), 20 markers were identified in the Northern Hemisphere which were not present in the South Pacific. Whereas, only two markers were identified in the South Pacific that were not present in the Northern Hemisphere (Supplementary Table 5).

## Discussion

The present mtDNA and DNA microsatellite data indicate the presence of diverged genetic lineages of *S. lalandi* from the NW Pacific (Japan and China), the NE Pacific (USA & Mexico), the South Pacific (Chile, NSW & SA) Oceans and South Africa. This interpretation aligns with previous reports of global genetic differences in the *S. lalandi* species complex based on mtDNA and microsatellite loci in the NE Pacific, NW Pacific and South Pacific



populations<sup>8,12</sup>. *S. lalandi* is known to be a long distance migratory fish: a study on kingfish migration reported that they were able to move >2,000 km from the tagged site, but were recaptured within the same hemispheric locations<sup>34</sup>. Mobility of *S. lalandi* was also suggested by genetic differences among years in Chile<sup>35</sup>. However the present data suggest very limited mixing between hemispheres; this may be due to the geographical barriers related to crossing the equator; factors such as the surface gradient of water temperature and salinity that are not preferred by the *S. lalandi*, as discussed in previous reports<sup>8,12</sup>.

Even though the South American and Australian populations are separated by the whole of the Pacific Ocean, there appears to be close similarity among the Southern Hemisphere populations. Purcell *et al.*<sup>8</sup> report that *S. lalandi* from Chile is not significantly different from the *S. lalandi* in New Zealand waters. Nugroho *et al.*<sup>24</sup> and Miller *et al.*<sup>14</sup> also reported that the *S. lalandi* population in Australian waters is not significantly different from that in New Zealand. Martinez-Takeshita *et al.*<sup>12</sup> considered fish from Chile, New Zealand and South Africa to be one species. Results from the present study revealed that there is no evidence of major genetic differentiation among two Australian populations (NSW & SA), whether considering mtDNA or nuclear DNA data. Furthermore, the current study observed that there is no significant differentiation among the South and East coast Australian, New Zealand and Chilean populations, suggesting *S. lalandi* from the South Pacific Ocean is from one genetic lineage. The GMYC analyses did indicate the South African population was a separate genetic lineage from South Pacific samples, even though some haplotypes were shared between South Pacific locations and South Africa. However there were only ten samples available from South Africa, and more samples are needed to confirm the differences between South Africa and other Southern Hemisphere populations.

This picture for the Southern Hemisphere contrasts that from the Northern Hemisphere where mtDNA and nuclear DNA data revealed fish from Japan as a somewhat separate genetic lineage from those in the NE Pacific<sup>8,12</sup>. Although the geographical distance between Japan (NW Pacific) and USA (NE Pacific) is less than the distance between Australia and Chile, there may be a limitation/barrier to gene flow between NW and NE Pacific locations. Rohde and Hayward<sup>36</sup> reported that Eastern Pacific Barrier is a 100% effective barrier to dispersal of species of *Scomberomorus*, *Grammatocynus* and their parasites. Martinez-Takeshita *et al.*<sup>12</sup> considered the South Pacific offers many island stepping stones for this inshore pelagic species which facilitate migration, stepping stones which are less abundant in the North Pacific. Moreover, Sepulveda and Gonzalez<sup>35</sup> provided evidence of the migration of genetically different populations over the time and space in South East Pacific Ocean.

Our partial genome sequencing data comprised SNP data and absence/presence of restriction fragments (i.e. the DArT data) were available for Mexico, Japan, Chile and two Australian populations. Broadly, the genome data analysed by a variety of tests supported patterns evident from the mtDNA and DNA microsatellite data, namely major differences between hemispheres but little variation within the Southern Hemisphere. There were some grouping differences regarding Japan which probably reflects the small number of Japanese samples. In general, the SNPs and DArT markers provided larger  $F_{ST}$  values than those estimated from the nine microsatellite loci used in the present study and also, the  $F_{ST}$  values observed from SNPs were greater than the values observed for DArT markers. Likewise AMOVA analyses for SNPs and DArT data gave much greater  $F_{ST}$  estimates, at least two fold of between region variation than those obtained from microsatellite loci. Again, as for  $F_{ST}$  values, greater between region variance using AMOVA was evident for SNPs than DArT. A set of DArT markers could be identified that categorically distinguished between the Mexican and Japanese populations and among the Northern and Southern Hemisphere populations.

Recently, Martinez-Takeshita *et al.*<sup>12</sup> and others<sup>8</sup> have proposed a revision to the taxonomy of *S. lalandi*, specifically that it should be split into three species, namely, resurrection of *Seriola aureovittata* for NW Pacific, *Seriola dorsalis* for NE Pacific, and leaving *S. lalandi* as the Southern Hemisphere species. The most straight forward tests for species division are those based just on percent DNA sequence homologies. Considering many different fish species, the recorded average genetic distance between species (within genera) for mtDNA COI gene for marine fish was at least above 9.54%<sup>37,38</sup>, and at least 5.71% for freshwater species<sup>39</sup>. However, the maximum COI sequence divergence observed between the populations in the present study was just 2.43% (present and published data pooled), which is prima facie evidence that all the populations studied in the present study are from a single species. The maximum nuclear DNA divergence was 0.67%, which is comparable to that between humans and Neanderthals<sup>40</sup> and there is some debate as to whether humans and Neanderthals are one or separate species<sup>41,42</sup>. According to the conventional molecular clock of 2% sequence divergence per million years for mtDNA<sup>33,43</sup>, *S. lalandi* populations seem to have diverged about 1.2 MYA between the two hemispheres using our data; this estimate of 1.2 MYA suggests a younger divergence, by about 50%, than previously reported<sup>12</sup>.

More involved tests for species delimitation are available, but do require a greater number of assumptions. Generalized mixed Yule Coalescent analyses indicated all populations belonged to one species group. On the other hand, a divergence time of 0.5 to 6.9 MYA between Hemispheres was estimated from Bayesian inferences in BEAST.

Overall, most of the analyses, both simple and complex point to relatively recent divergence and relatively little genetic difference (apropos that which normally accompanies speciation). The converse does not seem to be the case, that there is consistent and categorical evidence for major genetic change consistent with speciation. For major taxonomic revisions, consistency of the data are probably required and until this evidence is available, the revision proposed by Martinez-Takeshita *et al.*<sup>12</sup> may be premature.

Nuclear genome sequences were reported to have a lower rate of nucleotide substitution than the mtDNA, about 10-fold lower in primates<sup>33,44,45</sup>. However the rate of nucleotide substitution in mtDNA observed in the current study was only about 4 times greater than the nuclear DNA in *S. lalandi*. This apparent inconsistency between kingfish and primates indicates a need for more work to properly calibrate the evolutionary rate of kingfish mtDNA sequence divergence over time, and may be needed prior to relying on kingfish mtDNA for major taxonomic revision.

As to whether genetic incompatibilities could arise from mixing Southern and Northern Hemisphere stocks, on one hand our revised estimates of divergence, would lessen this risk as previously considered. On the other hand, our finding using genomic data of many categorical differences could indicate adaptation to different local conditions in the different hemispheres, or they could just reflect genetic drift. Perhaps direct empirical testing of  $F_2$  breakdown would help resolve this matter, and until then any conclusions regarding  $F_2$  breakdown would only be of a speculative nature.

## Methods and Materials

**Specimen collection and DNA extraction.** Fin tissue samples of yellowtail kingfish were collected from six different populations representing New South Wales, Australia (NSW,  $n = 48$ ), South Australia (SA,  $n = 29$ ), Chile ( $n = 50$ ), Japan ( $n = 15$ ), Mexico ( $n = 76$ ) and USA ( $n = 26$ ). Tissue samples were collected into tubes containing 80% ethanol and were stored at 4 °C. Genomic DNA was extracted and purified using the DNeasy Blood and Tissue kits (Qiagen, Hilden, Germany) and the NucleoMag<sup>®</sup> 96 Tissue kit (MACHEREY-NAGEL GmbH & Co. KG, Germany) following the manufacturers' protocol and extracted DNA was stored at -18 °C for future analysis.

**Mitochondrial DNA cytochrome oxidase subunit I (COI) gene amplification and sequencing.** A length of 655 bp fragment from the COI gene was amplified by polymerase chain reaction (PCR) using gene-specific primers, sense primer FishF1 5'-TCAACCAACCACAAAGACATTGGCAC-3' and anti-sense primer FishR1 5'-TAGACTTCTGGGTGGCCAAAGAATCA-3' designed by Ward *et al.*<sup>37</sup>. PCRs were carried out in 25  $\mu$ L reaction volumes containing 0.5  $\mu$ L of DNA (50 ng/ $\mu$ L), 12.5  $\mu$ L of 2 $\times$  AmpliTaq Gold 360 master mix, 1.0  $\mu$ L of each gene-specific primer (10 pmol/ $\mu$ L) and 10  $\mu$ L of PCR-grade H<sub>2</sub>O. Amplifications were performed using an Eppendorf Mastercycler<sup>®</sup> nexus (Hamburg, Germany). The PCR cycle program consisted of an initial denaturation step at 95 °C for 10 min, followed by 35 cycles of 95 °C for 30 s, 50 °C for 30 s, and 72 °C for 30 s, and a final extension of 72 °C for 10 min. PCR products were subjected to standard DNA sanger sequencing (Macrogen Inc., Korea) from both directions using the gene-specific PCR primer pair mentioned above (FishF1 and FishR1).

**Microsatellite DNA genotyping.** Nine established microsatellite markers were used to genotype the 231 specimens. Six markers (Sel001, Sel002, Sel008, Sel011, Sel017 and Sel019) were selected from *S. lalandi* (see Supplementary Table 6) transcriptome sequences<sup>5</sup>, and another three loci (Sdu21, Sdu32 and Sdu46) were selected from twenty five published microsatellite primer pairs<sup>46,47</sup>. Microsatellite primers were combined into two multiplex PCR sets and DNA amplification was achieved using the Qiagen Multiplex PCR PLUS Kit (Qiagen, Germany) in 13.5  $\mu$ L reactions, each containing 1.25  $\mu$ L of 10 $\times$  primer mix, 6.25  $\mu$ L of Multiplex PCR Master Mix, 2.75  $\mu$ L of RNase free water, 1.25  $\mu$ L of Q-Solution and 2.0  $\mu$ L of approximately 20 ng template DNA. Amplification was performed using an Eppendorf Mastercycler<sup>®</sup> nexus (Hamburg, Germany) with cycling conditions as follows: initial denaturation at 95 °C for 15 min, followed by 35 cycles of 95 °C for 30 s, 57 °C for 90 s, and 72 °C for 30 s; with a final extension at 68 °C for 30 min. PCR products were separated by capillary electrophoresis on an AB3500 Genetic Analyser (Applied Biosystems) at the University of the Sunshine Coast, Australia.

**DARtseq genotyping.** Genomic DNA from five populations were selected for genome-wide markers development; Japan ( $n = 2$ ), Mexico ( $n = 38$ ), Chile ( $n = 20$ ), New South Wales ( $n = 48$ ) and South Australia ( $n = 44$ ). DNA aliquots were sent to DARtseq<sup>™</sup> (Diversity Arrays Technology Pty. Ltd, Canberra, Australia) for markers development. Four methods of complexity reduction were tested (data not presented) and the *PstI-SphI* method was selected. DNA samples were processed in digestion/ligation reactions principally as per Kilian *et al.*<sup>48</sup> but replacing a single *PstI*-compatible adaptor with two different adaptors corresponding to two different restriction enzyme overhangs. The *PstI*-compatible adaptor was designed to include Illumina's flowcell attachment sequence, sequencing primer sequence and "staggered", varying length barcode region, similar to the sequence reported by Elshire *et al.*<sup>49</sup>. Reverse adaptor contained flowcell attachment region and *SphI*-compatible overhang sequence. Only "mixed fragments" (*PstI-SphI*) were effectively amplified in 30 cycles of PCR using the following reaction conditions: an initial denaturation step at 94 °C for 1 min, followed by 30 cycles of 94 °C for 20 s, 58 °C for 30 s, and 72 °C for 45 s; with a final extension at 72 °C for 7 min.

After PCR equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to c-Bot (Illumina) bridge PCR followed by sequencing on an Illumina HiSeq. 2500. The sequencing (single read) was run for 77 cycles, producing 77 bp reads. Sequences generated from each lane were processed using proprietary DARt analytical pipelines. In the primary pipeline the fastq files were first processed to filter away poor quality sequences, applying more stringent selection criteria to the barcode region compared to the rest of the sequence. In that way the assignments of the sequences to specific samples carried in the "barcode split" step were very reliable. Approximately 2,500,000 sequences per barcode/sample were identified and used in marker calling. Finally, identical sequences were collapsed into "fastqcoll files". The fastqcoll files were "groomed" using DARt PLs proprietary algorithm which corrects low quality base from singleton tag into a correct base using collapsed tags with multiple members as a template. The "groomed" fastqcoll files were used in the secondary pipeline for DARt PLs proprietary SNP and silicoDARt (presence/absence of restriction fragments in representation; DARt markers) calling algorithms (DARtsoft14). For SNP calling all tags from all libraries included in the DARtsoft14 analysis are clustered using DARt PLs C++ algorithm at the threshold distance of 3, followed by parsing of the clusters into separate SNP loci using a range of technical parameters, especially the balance of read counts for the allelic pairs. Additional selection criteria were added to the algorithm based on analysis of approximately 1,000 controlled cross populations. Testing for Mendelian distribution of alleles in these populations facilitated selection of technical parameters discriminating well true allelic variants from paralogous sequences. In addition multiple samples were processed from DNA to allelic calls as technical replicates

and scoring consistency was used as the main selection criteria for high quality/low error rate markers. Calling quality was assured by high average read depth per locus (Average across all markers was over 30 reads/locus).

**Data analysis. Mitochondrial DNA COI fragment.** Two hundred and forty-four mtDNA sequences generated in this study were trimmed to 604 bp (see Table 1 for GenBank accession numbers) and were analysed using the BLAST algorithm from NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). COI region sequences were evaluated and aligned using ClustalW multiple sequence alignment program in BioEdit sequence alignment editor V.7.2.5<sup>50</sup>. Published sequences were also collated by directly acquiring homologous sequences from GenBank (Genbank accession numbers: EF609460<sup>38</sup>; HM007727-HM007730<sup>51</sup>; JF494499; JF494500; JQ738432; JQ738434; KM877615-KM877656<sup>12</sup>; KU312946-KU312964<sup>35</sup>; KX781877); sequences were trimmed to 604 bp to correspond to the sequences generated in the present study. Also we regenerated DNA sequences from published haplotype frequencies as found in Sepúlveda and González<sup>35</sup>; for example, the sequence of the “H1” haplotype was reported along with its frequency (which was 80%) from a total of 291; thus we inferred there were a total count of 233 “H1” haplotypes.

Analysis of molecular variance (AMOVA) test and pairwise genetic differences ( $F_{ST}$ ) were estimated as indicators of genetic differentiation between all samples groups using Arlequin 3.5.2 and significance was assessed using 1,000 permutations<sup>52</sup>. DNA polymorphism/divergence analysis were conducted to estimate the average number of nucleotide differences ( $k$ ), nucleotide diversity ( $\pi$ ) and haplotype diversity ( $h$ ) using DnaSP v. 5.0<sup>53</sup>.

Prior to the construction of phylogenetic trees, an assessment was made using MEGA 7.0<sup>54</sup> to find the best model of sequence evolution. The best model for COI was determined to be the “K2 + G: Kimura 2-parameter + Gamma” model which was selected based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values<sup>55</sup>. Phylogenetic tree analyses were conducted with aligned sequences from the present study (244 sequences) and previously published (341 sequences) COI gene sequences from *S. lalandi* and from related *Seriola* species as outgroups (*S. dumerili*: KC501452; *S. rivoliiana*: JN021317; *S. zonata*: KP930431; *S. fasciata*: KF930429; *S. quinqueradiata*: KU168712) using Neighbour-Joining (NJ)<sup>56</sup>, Maximum Parsimony (MP)<sup>57</sup> and Maximum Likelihood (ML)<sup>58</sup> methods, with 1,000 bootstrap replicates in MEGA 7.0<sup>54</sup>.

To test how many different species were evident from the COI sequences, a species delimitation analysis was carried out using the Generalized Mixed Yule Coalescent (GMYC) approach<sup>59</sup>. Bayesian phylogenetic inferences were characterized with HKY + G (Hasegawa-Kishino-Yano + Gamma) model, assuming coalescent constant population priors with relaxed lognormal clock<sup>60</sup> in BEAST v2.4.6<sup>61</sup>. Single representatives of given COI haplotypes sequences (to avoid many zero-length branches) together with five outgroup sequences (from different *Seriola* species used in ML analysis) were used in phylogeny analysis. BEAST was run for 10 million of Markov-Chain Monte Carlo (MCMC) chain length (sampled every 1000 generations; burn-in 10%), which reached convergence with effective sample size (ESS) > 200 for all parameters. Independent GMYC analyses for the trees generated from BEAST were conducted using “rnc1”<sup>62</sup> and “splits”<sup>59</sup> packages in R<sup>63</sup> to discover species.

Divergence time estimates were assessed using two approaches. First, molecular dating for COI haplotype sequences were conducted using Bayesian inferences in BEAST v2.4.6<sup>61</sup>, setting birth-death tree prior, since this is well suited and used for multi species data with varying degrees of lineage divergence<sup>64</sup>. The uncorrelated log-normal relaxed molecular clock was selected for node ages estimates. Monophyly constraints were enforced for all clades represented by calibration points. Calibration points were set to 55.0 MYA (SD = 1.0) to calibrate the ancestral node of the genus *Seriola* from the fossil record<sup>65,66</sup> that has already been used in time-calibration studies<sup>67</sup> and 52.0 MYA (SD = 1.0) for the split between *Seriolina nigrofasciata* and *Elagatis bipinnulata* derived from phylogenetic data<sup>68</sup> and TIMETREE database<sup>69,70</sup> and thus here included *S. nigrofasciata* and *E. bipinnulata* (GenBank accession No. HQ560985 and JF493409, respectively) as outgroups. Three independent runs were implemented with 10 million MCMC steps and sampled every 1000 generations. A burn-in of 10% was used and the convergence of all parameters was assessed (ESS > 200) using the TRACER v1.6<sup>71</sup>. Subsequently, TreeAnnotator v2.4.6<sup>61</sup> was used to generate the consensus tree with median node age and FigTree v1.4.3<sup>72</sup> was used to visualise the annotated tree.

The second approach to estimate the divergence time among populations was to calculate pairwise sequence divergence percentages (1-sequence identity) using BioEdit v7.2.5<sup>50</sup>. The number of base differences per site was averaged over all corresponding sequence pairs for the full COI data set and calibrated according to Brown *et al.*<sup>33</sup>.

**Microsatellite DNA markers.** Molecular diversity indices including gene diversity ( $g$ ), mean number of alleles per locus ( $N_a$ ), observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_E$ ) were estimated with 10,000 permutations and Hardy-Weinberg equilibrium (HWE) tests were performed with 1,000,000 Markov Chain steps and 100,000 dememorization steps using Arlequin V. 3.5.2<sup>52</sup>. AMOVA test and population level of inbreeding ( $F_{IS}$ ) were performed using Genetic Analysis in Excel (GenAlEx) V. 6.5<sup>73,74</sup>. Microsatellite data were used to conduct Bayesian cluster analysis using STRUCTURE software V. 2.3 program<sup>75</sup>, assessing individuals to genetic clusters without a priori population definition and using ‘admixture’ ancestry model with correlated allele frequencies. The number of assumed genetic clusters ( $K$ ) was set to vary between 1 to 6 and the program setup for five independent runs with burn-in period of 100,000 iterations and 100,000 MCMC (Markov Chain Monte Carlo) iterations. The most likely number of  $K$  was determined using the STRUCTURE HARVESTER program<sup>76</sup>.

**SNP and DA<sub>r</sub>T markers.** SNP and DA<sub>r</sub>T markers (presence/absence data) data were filtered to remove any loci which had missing marker information for more than 5% of the samples within each population, as well as those with multiple markers in the same locus, to allow assumption of HWE and single-loci markers in downstream population structure and sequence homology analyses. In addition, samples with missing data in more than 100 markers were filtered out. Markers unique to each population were determined as those present in at least 95% of the individuals in the population, while present in not more than 5% of the individuals of all other populations.



Data filtration and analysis was performed in the R statistical programming language v3.2.1<sup>63</sup> environment, using the 'adegenet' package<sup>77</sup>.

Both DARt and SNPs data were subjected to AMOVA test using Genetic Analysis in Excel (GenAlEx) V. 6.5<sup>73,74</sup> and Arlequin V. 3.5.2<sup>52</sup> for population differentiation. In addition, a Bayesian model-based cluster analysis was performed for three independent runs with K from 1 to 5 and each run with a burn-in period of 100,000 iterations followed by 100,000 MCMC iterations, assuming an admixture ancestry model and correlated allele frequencies using STRUCTURE software V. 2.3 program<sup>75</sup>. The results of STRUCTURE were subsequently analysed to determine the most likely number of K using the STRUCTURE HARVESTER program<sup>76</sup>.

Since each single-loci DARtseq™ marker comprised of a single polymorphic base (SNP) in a consensus sequence of fixed tag length (70 bp in this case), the sequence homology (in percentage) across one allele of a polymorphic marker can be defined as  $(1 - 1/70 \text{ bp}) \times 100$ . Therefore, the overall sequence dissimilarity between samples can be calculated as the total number of varying SNPs ( $S_v$ ) across both alleles of single-loci markers, divided by the total number of alleles in the markers ( $2n$ ) multiplied by the tag length ( $l$ ), as presented in Equation 1. The overall sequence homology within and between populations was then calculated as the mean of homologies from each pairwise comparison of individual samples from a single or compared populations, respectively.

Equation 1. SNP-based sequence homology calculation

$$H_s = 1 - \frac{\sum_{i=0}^{2n} S_v}{2n \times l}$$

## References

1. FAO. *Global Aquaculture Production 1950–2014*, <http://www.fao.org/fishery/statistics/en> (2016).
2. Sicuro, B. & Luzzana, U. The state of *Seriola* spp. other than Yellowtail (*S. quinqueradiata*) farming in the world. *Reviews in Fisheries Science & Aquaculture* **24**, 314–325 (2016).
3. Nakada, M. In *Capture-based aquaculture. Global overview* (eds A. Lovatelli & P. F. Holthus) Ch. Capture-based aquaculture of yellowtail, 199–215 (FAO, 2008).
4. Hutson, K. S., Ernst, I., Mooney, A. J. & Whittington, I. D. Metazoan parasite assemblages of wild *Seriola lalandi* (Carangidae) from eastern and southern Australia. *Parasitol Int* **56**, 95–105 (2007).
5. Whatmore, P. *et al.* Genetic parameters for economically important traits in yellowtail kingfish *Seriola lalandi*. *Aquaculture* **400–401**, 77–84 (2013).
6. Kolkovski, S. & Sakakura, Y. In *Avances en Nutrición Acuícola VII* (eds L. E. Cruz Suárez *et al.*) 109–125 (Memorias del VII Simposium Internacional de Nutrición Acuícola, 2004).
7. Buentello, A., Jirsa, D., Barrows, F. T. & Drawbridge, M. Minimizing fishmeal use in juvenile California yellowtail, *Seriola lalandi*, diets using non-GM soybeans selectively bred for aquafeeds. *Aquaculture* **435**, 403–411 (2015).
8. Purcell, C. M. *et al.* Developing a genetic baseline for the yellowtail amberjack species complex, *Seriola lalandi* sensu lato, to assess and preserve variation in wild populations of these globally important aquaculture species. *Conservation Genetics* **16**, 1475–1488 (2015).
9. Orellana, J., Waller, U. & Wecker, B. Culture of yellowtail kingfish (*Seriola lalandi*) in a marine recirculating aquaculture system (RAS) with artificial seawater. *Aquacultural Engineering* **58**, 20–28 (2014).
10. Abbinck, W. *et al.* The effect of temperature and pH on the growth and physiological response of juvenile yellowtail kingfish *Seriola lalandi* in recirculating aquaculture systems. *Aquaculture* **330–333**, 130–135 (2011).
11. O'Neill, B., Le Roux, A. & Hoffman, L. C. Comparative study of the nutritional composition of wild versus farmed yellowtail (*Seriola lalandi*). *Aquaculture* **448**, 169–175 (2015).
12. Martínez-Takeshita, N. *et al.* A Tale of Three Tails: Cryptic Speciation in a Globally Distributed Marine Fish of the Genus *Seriola*. *Copeia* **103**, 357–368 (2015).
13. Yeoman, R. In *FISH Vol. 21* (ed. Horvat, P) 24–25 (Fisheries Research And Development Corporation (FRDC), Deakin, ACT, Australia, 2013).
14. Miller, P. A., Fitch, A. J., Gardner, M., Hutson, K. S. & Mair, G. Genetic population structure of Yellowtail Kingfish (*Seriola lalandi*) in temperate Australasian waters inferred from microsatellite markers and mitochondrial DNA. *Aquaculture* **319**, 328–336 (2011).
15. Scarratt, D. Yellowtail: First sapplings for new hatchery in Mexico. *Hatchery International* **15**, 12–13 (2014).
16. Danancher, D. & García-Vázquez, E. Genetic population structure in flatfishes and potential impact of aquaculture and stock enhancement on wild populations in Europe. *Reviews in Fish Biology and Fisheries* **21**, 441–462 (2011).
17. Gross, M. R. One species with two biologies: Atlantic salmon (*Salmo salar*) in the wild and in aquaculture. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 131–144 (1998).
18. Naylor, R. *et al.* Fugitive salmon: Assessing the risks of escaped fish from net-pen aquaculture. *Bioscience* **55**, 427–437 (2005).
19. Orr, H. A. The Population-Genetics of Speciation - the Evolution of Hybrid Incompatibilities. *Genetics* **139**, 1805–1813 (1995).
20. Maheshwari, S. & Barbash, D. A. The Genetics of Hybrid Incompatibilities. *Annu Rev Genet* **45**, 331–355 (2011).
21. Stelkens, R. B., Schmid, C. & Seehausen, O. Hybrid Breakdown in Cichlid Fish. *Plos One* **10**, e0127207 (2015).
22. Edmands, S. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution* **53**, 1757–1768 (1999).
23. Pritchard, V. L. & Edmands, S. The Genomic Trajectory of Hybrid Swarms: Outcomes of Repeated Crosses between Populations of *Tigriopus Californicus*. *Evolution* **67**, 774–791 (2013).
24. Nugroho, E., Ferrell, D. J., Smith, P. & Taniguchi, N. Genetic divergence of kingfish from Japan, Australia and New Zealand inferred by microsatellite DNA and mitochondrial DNA control region markers. *Fisheries science* **67**, 843–850 (2001).
25. Bowen, B. W. & Karl, S. A. Population genetics and phylogeography of sea turtles. *Molecular Ecology* **16**, 4886–4907 (2007).
26. Pardini, A. T. *et al.* Sex-biased dispersal of great white sharks - In some respects, these sharks behave more like whales and dolphins than other fish. *Nature* **412**, 139–140 (2001).
27. Galtier, N., Nabholz, B., Glemin, S. & Hurst, G. D. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol* **18**, 4541–4550 (2009).
28. Goldstein, D. B. & Pollock, D. D. Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 335–342 (1997).
29. Estoup, A., Jarne, P. & Cornuet, J. M. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**, 1591–1604 (2002).
30. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* **5**, 435–445 (2004).
31. Patron, N. J., Rogers, M. B. & Keeling, P. J. Comparative rates of evolution in endosymbiotic nuclear genomes. *Bmc Evolutionary Biology* **6**, Art 46 (2006).

32. Hedges, S. B. & Kumar, S. Genomic clocks and evolutionary timescales. *Trends in Genetics* **19**, 200–206 (2003).
33. Brown, W. M., George, M. & Wilson, A. C. Rapid Evolution of Animal Mitochondrial-DNA. *P Natl Acad Sci USA* **76**, 1967–1971 (1979).
34. Gillanders, B. M., Ferrell, D. J. & Andrew, N. L. Estimates of movement and life-history parameters of yellowtail kingfish (*Seriola lalandi*): how useful are data from a cooperative tagging programme? *Marine and Freshwater Research* **52**, 179–192 (2001).
35. Sepulveda, F. A. & Gonzalez, M. T. Spatio-temporal patterns of genetic variations in populations of yellowtail kingfish *Seriola lalandi* from the south-eastern Pacific Ocean and potential implications for its fishery management. *Journal of Fish Biology* **90**, 249–264 (2017).
36. Rohde, K. & Hayward, C. J. Oceanic barriers as indicated by scombrid fishes and their parasites. *International Journal for Parasitology* **30**, 579–583 (2000).
37. Ward, R. D., Zemlak, T. S., Innes, B. H. & Last, P. R. DNA Barcoding Australia's Fish Species. *Philosophical Transactions of the Royal Society* **360**, 1847–1857 (2005).
38. Ward, R. D. & Holmes, B. H. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. *Molecular Ecology Notes* **7**, 899–907 (2007).
39. April, J., Mayden, R. L., Hanner, R. H. & Bernatchez, L. Genetic calibration of species diversity among North America's freshwater fishes. *P Natl Acad Sci USA* **108**, 10602–10607 (2011).
40. Noonan, J. P. *et al.* Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113–1118 (2006).
41. Phillips, D. *Neanderthals Are Still Human!* <http://www.icr.org/article/neanderthals-are-still-human/> (visited on 2017-06-20).
42. Powlledge, T. M. *Were Neanderthals a different species?* <https://geneticliteracyproject.org/2014/02/04/were-neanderthals-a-different-species/> (visited on 2017-06-20).
43. Bowen, B. W., Bass, A. L., Rocha, L. A., Grant, W. S. & Robertson, D. R. Phylogeography of the trumpETFishes (Aulostomus): Ring species complex on a global scale. *Evolution* **55**, (1029–1039 (2001).
44. Crawford, A. J. Relative rates of nucleotide substitution in frogs. *Journal of Molecular Evolution* **57**, 636–641 (2003).
45. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. Mitochondrial-DNA Sequences of Primates - Tempo and Mode of Evolution. *Journal of Molecular Evolution* **18**, 225–239 (1982).
46. Renshaw, M. A., Patton, J. C., Rexroad, C. E. & Gold, J. R. Isolation and characterization of dinucleotide microsatellites in greater amberjack. *Seriola dumerili*. *Conservation Genetics* **8**, 1009–1011 (2007).
47. Renshaw, M. A., Patton, J. C., Rexroad, C. E. & Gold, J. R. PCR primers for trinucleotide and tetranucleotide microsatellites in greater amberjack. *Seriola dumerili*. *Molecular Ecology Notes* **6**, 1162–1164 (2006).
48. Kilian, A. *et al.* In *Data Production and Analysis in Population Genomics: Methods and Protocols* (eds François Pompanon & Aurélie Bonin) 67–89 (Humana Press, 2012).
49. Elshire, R. J. *et al.* A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One* **6**, ARTN e19379 (2011).
50. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98 (1999).
51. Cawthorn, D. M., Steinman, H. A. & Witthuhn, R. C. Establishment of a mitochondrial DNA sequence database for the identification of fish species commercially available in South Africa. *Molecular Ecology Resources* **11**, 979–991 (2011).
52. Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50 (2005).
53. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
54. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870–1874 (2016).
55. Posada, D. Selecting models of evolution. *The phylogenetic handbook. A practical approach to DNA and protein phylogeny*. (pp. 256–282. Cambridge University Press, Cambridge, 2003).
56. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425 (1987).
57. Tamura, K. *et al.* MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731–2739 (2011).
58. Tamura, K. & Nei, M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees. *Molecular Biology and Evolution* **10**, 512–526 (1993).
59. Fujisawa, T. & Barraclough, T. G. Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. *Syst Biol* **62**, 707–724 (2013).
60. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **4**, e88 (2006).
61. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**, ArtN 214 (2007).
62. Michonneau, F. *rnc1: A new package to import NEXUS and Newick trees in R* <https://CRAN.R-project.org/package=rnc1> (visited on 2017-06-12).
63. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299–314 (1996).
64. Bagley, J. C. *et al.* Assessing Species Boundaries Using Multilocus Species Delimitation in a Morphologically Conserved Group of Neotropical Freshwater Fishes, the Poecilia sphenops Species Complex (Poeciliidae). *Plos One* **10**, UNSP e0121139 (2015).
65. Patterson, C. & Owen, H. G. Indian Isolation or Contact? A Response to Briggs. *Syst Biol* **40**, 96–100 (1991).
66. Berg, L. S. Classification of Fishes Both Recent and Fossil (J. W. Edwards, 1947).
67. Swart, B. L. & von der Heyden, S. Bester-van der Merwe, A. & Roodt-Wilding, R. Molecular systematics and biogeography of the circumglobally distributed genus *Seriola* (Pisces: Carangidae). *Mol Phylogenet Evol* **93**, 274–280 (2015).
68. Santini, F. & Carnevale, G. First multilocus and densely sampled timetree of trevallies, pompanos and allies (Carangoidei, Percomorpha) suggests a Cretaceous origin and Eocene radiation of a major clade of piscivores. *Mol Phylogenet Evol* **83**, 33–39 (2015).
69. Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution* **32**, 835–845 (2015).
70. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* **34**, 1812–1819 (2017).
71. Rambaut, A., Drummond, A. J. & Suchard, M. *Tracer v1.6* <http://beast.bio.ed.ac.uk/Tracer> (visited on 2017-06-12).
72. Rambaut, A. *FigTree v1.4.3* <http://tree.bio.ed.ac.uk/software/figtree/> (visited on 2017-06-12).
73. Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539 (2012).
74. Peakall, R. & Smouse, P. E. genalex 6: genetic analysis in Excel. *Population genetic software for teaching and research. Molecular Ecology Notes* **6**, 288–295 (2006).
75. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).

76. Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* **4**, 359–361 (2012).
77. Jombart, T. & Ahmed, I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).

### Acknowledgements

This study funded by the Australian Seafood Cooperative Research Centre (CRC Project No. 2010/768) and the University of the Sunshine Coast, Queensland, Australia. We would like to thank Jane Quinn, Daniel Powell, Sankar Subramanian and Philip Brady (at Genecology, University of the Sunshine Coast) for their support and technical advice.

### Author Contributions

H.K.A.P. wrote the manuscript and conducted the laboratory work. H.K.A.P., F.L., Y.T., A.M., S.F., W.O., N.H.N. and W.K. contributed to data and tissue sample collection. H.K.A.P., C.H.F., N.H.N. and I.B. performed data analysis. H.K.A.P and W.K. conceived and designed the experiment and W.K. oversaw the study assisted with the revisions. All authors read and approved the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-07419-x](https://doi.org/10.1038/s41598-017-07419-x)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017