Villacorta-Rath, Cecilia, Ilyushkina, Irina, Strugnell, Jan M., Green, Bridget S., Murphy, Nicholas P., Doyle, Stephen R., Hall, Nathan E., Robinson, Andrew J., and Bell, James J. (2016) ***Outlier SNPs enable food traceability of the southern rock lobster, Jasus edwardsii.*** **Marine Biology, 163 (11) pp. 1-11.**

Authors

Cecilia Villacorta-Rath[1*], Irina Ilyushkina[2], Jan M. Strugnell[3], Bridget S. Green[1], Nicholas P. Murphy[3], Stephen R. Doyle[4], Nathan E. Hall[5,6], Andrew J. Robinson[5,6], James J. Bell[2]

Title

Outlier SNPs enable food traceability of the southern rock lobster, *Jasus edwardsii*

Author Affiliations

1. Institute for Marine and Antarctic Studies, University of Tasmania, Tasmania 7001, Australia

2. School of Biological Sciences, Victoria University of Wellington, Wellington 6140, New Zealand

3. Department of Ecology, Environment and Evolution, School of Life Sciences, La Trobe University, Victoria 3086, Australia

4. Department of Animal, Plant and Soil Sciences, School of Life Sciences, La Trobe University, Victoria 3086, Australia

5. Genomics Research Platform, La Trobe University, Victoria 3086, Australia

6. Life Sciences Computation Centre, Victorian Life Sciences Computation Initiative, Victoria 3053, Australia

* Corresponding author: Cecilia.Villacorta@utas.edu.au; Phone: +61 448 452 083

Abstract Recent advances in next generation sequencing have enhanced the resolution of population genetic studies of non-model organisms through increased

marker generation and sample throughput. Using double digest restriction site-associated DNA sequencing (ddRADseq), we investigated the population structure of the commercially important southern rock lobster, *Jasus edwardsii,* in Australia and New Zealand with the aim of identifying a panel of SNP markers that could be used to trace country of origin. Four ddRADseq libraries comprising a total of 91 individuals were sequenced on the Illumina MiSeq platform and demultiplexed reads were used to create a reference catalog of loci. Individual reads were then mapped to the reference catalog and variant calling was performed. We have characterized two single nucleotide polymorphism (SNP) panels comprised in total of 656 SNPs. The first panel contained 535 neutral SNPs, and the second, 121 outlier SNPs that were characteristic of being putatively under selection. Both neutral and outlier SNP panels showed significant differentiation between the two countries, with the outlier loci demonstrating much larger F_{ST} values (F_{ST} outlier SNP panel = 0.134, $P < 0.0001$; F_{ST} neutral SNP panel = 0.022, $P < 0.0001$). Assignment tests performed with the outlier SNP panel allocated 100% of the individuals to country of origin, demonstrating the usefulness of these markers for food traceability of *J. edwardsii*.

Acknowledgements We would like to thank Gary Carlos (University of Tasmania), Colin Fry (University of Tasmania), Daniel Ierodiaconou (Deakin University), Andrew Kent and Kent Way for field assistance and sample collection in Australia. Thanks to Daryl Sykes (New Zealand Rock Lobster Industry Council) for organizing all sample collections in New Zealand and Don Nelson (New Zealand Rock Lobster Industry Council) and Dr. Debbie Freeman (Department of Conservation, New Zealand) for collecting samples in New Zealand. Thanks for lab assistance to Mel Best, Adam Smolenski and Cecilia Carrea (University of Tasmania). We also thank

Michael Amor and Laura Woodings (La Trobe University) who helped developing the ddRADseq protocol and the rad-loci pipeline, respectively. Special thanks to Karen J Miller who contributed to the original project idea. We would like to thank the editor and two reviewers for their constructive suggestions.

55 **Introduction**

56 Marine benthic invertebrates typically exhibit a pelagic larval phase that
57 serves as a mechanism of dispersal and maintains connectivity between sub-
58 populations (Eckman 1996). It is widely accepted that larvae with a short pelagic
59 larval duration (PLD) are more prone to settling close to their parents, generating
60 genetic structure at broad spatial scales (Palumbi 1994). In contrast, long PLDs can
61 potentially lead to an absence of, or low population structure due to dispersal of larvae
62 over large geographical areas (Shanks et al. 2003). However, most larval transport is
63 largely determined by hydrodynamic features, which can cause strong genetic
64 differentiation, even in species with a relatively long PLD (Palumbi 1994). An
65 increasing number of studies using genetic markers have concluded that larval
66 duration cannot be directly used as a predictor of genetic structure (Shanks 2009; Wei
67 et al. 2013; Teske et al. 2015).

68 The southern rock lobster, *Jasus edwardsii*, is distributed from southern
69 Australia and the Tasman Sea to all coasts of New Zealand. *J. edwardsii* is a
70 commercially important species in both countries and fisheries management is carried
71 out independently in Australia and New Zealand. This resource represents a
72 substantial income for economies of both countries, providing annual revenue of
73 approximately US$292 million to Australia (ABARE-BRS 2010) and US$204 million
74 to New Zealand (Statistics New Zealand 2016). The main export market for both
75 countries is Asia, where lobsters are mostly exported live (ABARE-BRS 2010; Jeffs
76 et al. 2013).

77 Despite the protracted pelagic larval duration of up to 24 months (Booth and
78 Phillips 1994), modeling simulations of larval trajectories have estimated that only
79 8% of larvae released from Australia have the potential to reach New Zealand (Bruce

et al. 2007). An earlier genetic study that characterized the structure of six rock lobster populations from around New Zealand and two populations in Australia found evidence for restricted gene flow across the species range (Thomas and Bell 2013). The hypothesis of a panmictic *J. edwardsii* population throughout its geographical range was rejected based on significant F_{ST} (F_{ST} = 0.011) from nine microsatellite markers (Thomas and Bell 2013). However, the authors suggested the possibility of larval migration from Tasmania (Australia) to central New Zealand. In support of this hypothesis, a subsequent study conducted with eight microsatellite markers using lobsters collected from Tasmania and the southern zone of the South Island of New Zealand also revealed population structure between countries (Morgan et al. 2013). The presence of genetic structure between Australia and New Zealand populations detected using microsatellites (Morgan et al. 2013; Thomas and Bell 2013) highlights the potential for using genetic markers to assign location of origin to lobsters, which may be useful for fisheries management purposes.

The use of genome-wide SNP markers, in contrast to microsatellite markers, has the potential to improve resolution in the estimation of population structure, migration rates, dispersal and population assignment (Morin et al. 2004; Benestan et al. 2015), as well as the ability to explore genomic regions under selection. Recent studies have identified high levels of population structure when analyzing small numbers of outlier markers in marine fish (Corander et al. 2013; Milano et al. 2014; Candy et al. 2015). For example, 299 neutral SNPs identified large-scale population subdivision of the widespread European hake, *Merluccius merluccius*, between the Atlantic and Mediterranean Seas, but significantly finer scale resolution was found when analyzing just 7 and 19 outlier SNPs within the Atlantic and Mediterranean basins, respectively (Milano et al. 2014). Similarly, fine scale population structure of

105 eulachon, *Thaleichthys pacificus*, in North America was distinguished through 193
106 outlier SNPs, in comparison to lower genetic differentiation detected when analyzing
107 3911 neutral SNPs (Candy et al. 2015). Both studies attributed the high level of
108 genetic variation in outlier SNPs to local adaptation.
109 The high levels of population differentiation detected with markers under
110 selection makes them appropriate for traceability of commercial fisheries species
111 (Araneda et al. 2016). Traceability of fish products is being increasingly used for
112 consumer protection and for regulatory enforcement, especially in unreported and
113 unregulated fishing (Ogden 2008). Although the *J. edwardsii* fisheries in Australia
114 and New Zealand are managed sustainably, both countries export lobster to the Asian
115 market. In the past, China has restricted Australian imports due to public health
116 concerns and economic reasons. Therefore, efficient assignment of commercialized *J.*
117 *edwardsii* to country of origin could prevent any conflict between Australia and New
118 Zealand if any further bans are imposed in the future.
119 The aim of this study was to identify a panel of SNP markers that would
120 enable high population assignment success and therefore could be used to trace
121 country of origin for *J. edwardsii* to either New Zealand or Australia. We used a
122 double digest restriction site-associated DNA (ddRADseq) approach (Peterson et al.
123 2012) to explore genetic structure of *J. edwardsii* using both neutral markers and
124 markers putatively under selection. The high level of genetic differentiation exhibited
125 by the outlier SNP panel allowed us to successfully assign individuals to population
126 of origin.
127
128 **Materials and methods**
129 Sample collection

A total of 40 individuals from five sites (corresponding to three regions) in Australia and 48 individuals from four sites (corresponding to four regions) in New Zealand were collected for the present study (Fig. 1) between 2011 and 2014 (Table 1). Adult lobsters on West Tasmania (AUS), East Tasmania (AUS), the Hauraki Gulf (NZ) and the Chatham Islands (NZ) were caught using commercial baited lobster pots. For the Australian samples, a pleopod clip was taken from each lobster and preserved in 90% ethanol. For the New Zealand samples, lobster legs were removed from live specimens and frozen. Downstream analyses suggest that differences in sample preservation did not produce a batch effect, since samples from the Hauraki Gulf and Chatham islands were assigned into the same cluster as the rest of New Zealand sampling sites (see Results). In the case of Stewart Island (NZ), Tonga Island (NZ) and Merri Marine Sanctuary (AUS), adult lobsters were collected by divers and legs (New Zealand specimens) or pleopod clips (Australian specimens) were taken from each lobster and immediately preserved in 90% ethanol.

DNA extractions

DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. The DNA concentration of each sample was determined using a Qubit® 2.0 Fluorometer (Life Technologies). DNA integrity was determined through gel electrophoresis and samples with predominantly high molecular weight DNA (corresponding to a band 1000 base pairs, bp, or higher) were preferentially selected for ddRADseq library preparation.

ddRADseq library preparation and sequencing

A modified version of the ddRADseq protocol developed by Peterson et al. (2012) was used to make multiplexed sequencing libraries (https://molecularbiodiversity.wordpress.com/home/protocols/). Briefly, 250 ng of genomic DNA was digested using the restriction enzyme EcoRI (GAATTC, infrequent cutter) and AciI (CCGC, frequent cutter) at 37°C for 16 h. Subsequently, sequencing adapters containing in-line barcodes were ligated to the sheared DNA fragments, after which low molecular weight DNA fragments as well as non-ligated adapters were removed using a double size selection protocol employing Agencourt AMPure XP magnetic beads (Beckman Coulter) (Lennon et al. 2010). Index sequences (based on TruSeq LT) and flow-cell attachment regions were incorporated by primer extension PCR. PCR products were cleaned using AMPure XP (Beckman Coulter), after which the DNA concentration was standardized, pooled and a gel size selection was performed to obtain a DNA fragments between 400 and 500 bp. DNA was extracted from the gel using the Wizard SV Gel and PCR Clean-Up System (Promega) and the concentration of the final ddRADseq library was determined using Qubit 2.0 fluorometer. All ddRADseq libraries were spiked with 10% PhiX Control v3 and sequenced on the Illumina MiSeq next generation sequencing platform using v2 2x250 bp kits.

The number of samples to be sequenced in a single run was determined through two pilot sequencing runs. For the first pilot run, 12 individuals from New Zealand were sequenced and for the second pilot run 13 specimens from Australia were sequenced. We determined the number of polymorphic loci and the number of reads and sequencing depth per individual and concluded that ddRADseq libraries consisting of 45 individuals would yield sufficient coverage and depth of loci to be

sequenced in a single run. Subsequently, two more ddRADseq libraries were sequenced, for a total of four ddRADseq datasets.

Technical replicates

Technical replicates were included in the second, third and fourth sequencing libraries prepared. This was particularly important given that our ddRADseq libraries were not prepared and sequenced at the same time, which may have had introduced technical- and sequencing-derived differences between libraries (Mastretta-Yanes et al. 2015). A Principal Component Analysis (PCA) was used to visualize the spatial distribution of replicates and identify whether there was a batch effect due to library preparation and sequencing that could bias results. The PCA was performed using the R package Adegenet v.1.4-1 (Jombart and Ahmed 2011). A summary of the distribution of technical replicates among libraries is given in Table S1.

Preliminary analyses of raw sequencing data

Quality of the reads was initially examined using the FastQC v.0.10.1 quality control tool (Babraham Bioinformatics). Subsequently, uniquely indexed and barcoded samples were demultiplexed using the *"*process_radtags" protocol from Stacks v.1.29 (Catchen et al. 2011). Based on the FastQC report, sequences were trimmed to 75 bp to assure that the Phred Quality Score (Q) of all reads were above 30. Trimmed reads were assessed for bacterial and viral contamination using Kraken v.3.5.0 (Wood and Salzberg 2014). This software compares sequence reads against a database to identify reads that match the taxonomic groups present in the database. Reads that do not match those of the database are output as "unclassified", or non-bacterial or viral reads, which were extracted from the raw trimmed data for further

203 analysis. In the absence of a reference genome, removal of bacterial or viral reads is
204 important since it is impossible to determine whether a sequence belongs to the study
205 organism or to a contaminant, providing biased results in downstream analyses (see
206 Merchant et al. 2014).
207
208 Reference catalog building, alignment and variant calling
209 Since *J. edwardsii* is a non-model species, we identified a catalog of the most
210 frequently sequenced loci using the "rad-loci" pipeline
211 (https://github.com/molecularbiodiversity/rad-loci) developed at La Trobe University.
212 Sequence data from all individuals was first pooled and then clustered using VSearch
213 v.1.1.3 to identify putative loci. Initially only clusters of reads with a depth of at least
214 103 (and therefore likely to appear at least once in most of the samples) were retained.
215 This means that the number of raw reads required to form a cluster was 103.
216 Sequences were considered to be sufficiently related if they shared at least 94%
217 identity (4 bp maximum difference in a 75 bp read, allowing single nucleotide
218 variations and indels). Therefore, the allowed number of mismatches between two
219 clusters was set to 5bp. Assuming that each member of the cluster was an allele, only
220 clusters that were composed by a minimum of two members and a maximum of 16
221 members were kept. The minimum number was based on the fact that we wanted to
222 obtain bi-allelic data. A second round of clustering of the remaining reads at 94%
223 identity was performed, followed by another filtering of clusters that were not
224 comprised by a minimum of two and a maximum of 16 members. After re-filtering,
225 individual samples were mapped back to the filtered clusters, each cluster was now
226 called a "locus" and it consisted of one representative sequence and up to 16
227 variations. Finally, samples were mapped back to the identified loci allowing for a

maximum of 40% missing data across samples in each locus. The total number of "reference" loci identified by this pipeline was 1,054 (Table S2).

Individual reads were subsequently mapped to the reference loci catalog using the software Bowtie2 v.0.7.12 (Langmead and Salzberg 2012). Finally, variant calling of mapped loci was performed using the Genome Analysis Toolkit (GATK) v.3.3_0 (McKenna et al. 2010), yielding a total of 947 SNPs (Table S2).

SNP filtering

The putative RAD loci were filtered to ensure that known confounding variables, such as non bi-allelic loci, missing loci, allele dropout (ADO) and linkage disequilibrium (LD) (Henning et al. 2014), were minimized prior to population level analyses using VCFtools v.0.1.13 (Danecek et al. 2011). Only bi-allelic loci were retained, using the options --min-alleles 2 --max-alleles 2. Additionally we selected a single SNP per locus with the option --thin 75, given that loci were 75 bp long. Since paired-end sequencing was performed and therefore pairs of loci of 75 bp length could potentially be in LD, we set a pairwise LD measure threshold of $r^2 > 0.8$ to remove potentially linked loci using the option --min-r2. Average coverage was 44, allowing for a minimum coverage of 5x to minimize ADO. Rare alleles were also removed by setting a minor allele frequency of 0.1 with the option --maf. Finally, both loci and individuals with more than 20% missing data were excluded from the analysis, yielding a final total number of 656 SNPs (Table S2).

SNP characterization

Detection of neutral loci and loci putatively under selection was performed using Lositan (Beaumont and Nichols 1996; Antao et al. 2008) using 100,000

simulations, a confidence interval of 0.99, and a false discovery rate of 0.1 (Jacobsen et al. 2014). Lositan uses an F_{ST}-outlier approach that identifies loci as outliers when their FST values are too high or too low compared to neutral expectations (Antao et al. 2008). Loci found to be under balancing and positive selection will be hereafter referred as outlier loci. All downstream analyses were performed and are reported for each SNP panel separately.

Finally, to examine whether sequence reads aligned to protein coding regions, demultiplexed untrimmed reads were screened through tBLASTx v.2.2.29+ (Altschul et al. 1997). This program searches a translated nucleotide database with putative translated nucleotide queries. Raw reads (150 - bp) were used rather than the 75 bp reference loci to improve the BLAST alignment length and hence specificity. Subsequently, queries with statistically significant e-values ($E < 0.01$, Karlin and Altschul 1990) were screened against the reference loci using BLASTn v.2.2.29+ (Altschul et al. 1997) to identify any loci that were contained on those reads and therefore linked to those genes.

Analyses of genetic diversity

The level of observed (H_o) and expected heterozygosity (H_E) in each population, as well as F statistics for both SNP panels, were calculated using the R packages Adegenet v.1.4-1 (Jombart and Ahmed 2011) and Pegas v.0.8-2 (Paradis 2010). A two-sample Wilcoxon test was used to detect whether mean H_o differed significantly from mean H_E for both SNP panels.

Effective population size estimation

277 Effective population size (Ne) was estimated by the software NeEstimator
278 v.2.01 (Do et al. 2014) using the linkage disequilibrium model based on allele
279 frequencies of all 656 loci. This model provides the most precise estimation of Ne
280 among other single-sample methods (Waples and Do 2010; Do et al. 2014).
281
282 Analysis of population structure
283 Population structure was investigated through Discriminant Analysis of
284 Principal Components (DAPC). DAPC was performed and results were plotted using
285 the R package Adegenet v.1.4-1 (Jombart and Ahmed 2011). This analysis assigns
286 individuals to clusters and selects the best number of clusters based on Bayesian
287 Information Criterion (BIC) (Jombart et al. 2010). The results are comparable to those
288 obtained by STRUCTURE (Pritchard et al. 2000), with the advantage that DAPC
289 explores genetic structure without making assumptions about the genetic model of the
290 study population (Jombart et al. 2010).
291
292 Assignment of individuals to country of origin
293 Assignment tests were performed using the program Geneclass2.0 (Piry et al.
294 2004) to test the effectiveness of the outlier SNP panel to assign individuals to
295 country of origin. We simulated $10x10^3$ multilocus genotypes using the algorithm
296 described by Paetkau et al. (2004). We used the Bayesian allele frequency estimation
297 method (Rannala and Mountain (1997) whereby each individual is removed from the
298 baseline and assigned to the most likely population.
299
300 **Results**
301 Variant calling and SNP filtering

ddRADseq sequencing libraries were prepared from seven sub-populations of the southern red rock lobster from Australia and New Zealand and sequenced using an Illumina MiSeq. A "reference" catalog of loci was developed from a total of 88 individuals using an average of 580,000 reads per individual. After variant calling and SNP filtering, our final dataset comprised 75 individuals, including 30 Australian and 45 New Zealand samples. From this data, a total of 954 SNPs were identified, and after filtering (MAF = 0.1, LD < 0.8 between loci, 1 SNP per loci, < 20% missing data of loci and individuals), a total of 656 SNPs were obtained. Using the software Lositan to discriminate between putatively neutral variants and variants characteristic of being under selection, two panels of SNPs were identified consisting of 535 neutral and 121 outlier SNPs respectively.

Consistency in assignment of technical replicates

Technical replicates were included in three ddRADseq libraries. The PCA showed consistency in assigning technical replicates from the same individual close to each other and within the cluster of each country. This means there was no batch effect produced by different library preparation and sequencing runs (Fig. S1).

Genetic diversity and effective population size

Mean expected and observed heterozygosity for both countries were higher for the neutral SNP panel than for the outlier panel and the New Zealand population exhibited higher heterozygosity than the Australian population (Table 2). The negative values of the inbreeding coefficient (F_{IS}) estimated for both SNP panels and both countries are indicative of heterozygote excess. The two-sample Wilcoxon test indicated that H_O was significantly higher than H_E in both SNP panels for the New

327 Zealand population ($P < 2.2e-16$ and $P = 0.0199$, respectively for the neutral and
328 outlier SNP panels). For the Australian population, significant differences between H_O
329 and H_E were only found for the neutral SNP panel ($P < 2.2e-16$ and $P = 0.07222$,
330 respectively for the neutral and outlier SNP panels).
331 Estimation of effective population size in Australia and New Zealand gave
332 infinite values, with confidence intervals of 1334.6 - ∞ for Australia and infinite for
333 the New Zealand *J. edwardsii* population. Infinite values of estimated Ne and
334 confidence intervals may be due to larger than expected sampling error (Do et al.
335 2014).
336
337 Population structure
338 While both panels detected population structure between Australia and New
339 Zealand, the outlier loci detected greater differentiation between countries (Table 2).
340 DAPC analysis using both SNP panels detected two clusters, each of them
341 representing each of the two countries (Fig. 2). However, the higher genetic
342 divergence given by the outlier SNP panel was detectable in the distance between
343 both clusters, which was one order of magnitude higher than that of the neutral SNP
344 panel (Fig. 2c, a, respectively). To determine the outlier SNPs causing most of the
345 differentiation between countries we inspected the associated allele loadings (Jombart
346 et al. 2010), showing that 30 SNPs contribute to the discrimination between Australia
347 and New Zealand (Fig. S2).
348
349 Assignment of individuals to country of origin
350 The outlier SNP panel correctly assigned 100% of individuals to population
351 of origin, with a quality index of 87.72% (Table S3).

352

353 Characterization of catalog loci

354 An analysis of the ddRAD loci to identify coding sequence variants using
355 tBLASTx did not produce any significant alignment with loci containing neutral or
356 outlier SNPs.

357

358 **Discussion**

359 This study represents the first genome-wide population genetic analysis of a
360 rock lobster species. We produced a panel of outlier and neutral loci to investigate
361 population structure between Australia and New Zealand and the potential for
362 assignment of individuals to population of origin. The southern red rock lobster, *J.*
363 *edwardsii*, is a commercially important species in Australia and New Zealand, and,
364 given reported evidence of differing degrees of genetic connectivity (Morgan et al.
365 2013; Thomas and Bell 2013), further fine scale genetic mapping of populations from
366 both countries is required to understand the extent of connectivity between countries.
367 As seen in other marine species, our outlier SNP panel showed greater genetic
368 differentiation between Australia and New Zealand than the neutral panel. Therefore
369 we propose that the outlier SNP panel has potential as an effective method for
370 determining population of origin in *J. edwardsii*.

371

372 Evidence for a population bottleneck

373 Significant differences between mean H_O and H_E were detected for the
374 neutral SNP panel in both countries and the outlier SNP panel in New Zealand, and
375 the negative values of the fixation index indicates heterozygote excess. Heterozygote
376 excess can be a result of low depth of coverage, since in the presence of low depth

377 data homozygote genotypes could be erroneously called heterozygous (Nielsen et al.
378 2012b). However, we do not consider this to be the case in the present study since our
379 average depth of coverage is 43 and it has been suggested that a minimum of 5x
380 coverage gives reliable estimates of allele frequencies (Ellegren 2014).
381 Heterozygote excess may also be the result of a small effective population
382 size caused by a past population bottleneck many generations ago (Luikart and
383 Cornuet 1998) or as a result of a relatively small number of individuals contributing to
384 each generation (Hedgecock et al. 2011). However, heterozygote excess caused by a
385 population bottleneck can be a transient state (Luikart and Cornuet 1998). Thomas
386 and Bell (2013) estimated an effective population size of 2,717 in Australia and 1,189
387 in New Zealand. These estimates are low and therefore it is possible that the
388 populations have undergone a bottleneck. The infinite estimates of Ne in the present
389 study do not provide any evidence of genetic drift due to small Ne (Do et al. 2014).
390 However, as discussed above, our infinite values of estimated Ne and confidence
391 intervals may be due to larger than expected sampling error. When Ne is large or if
392 there is limited data available, the estimate of Ne will be negative and the biological
393 interpretation is that Ne = ∞ (see Waples and Do 2010).
394
395 Performance of the neutral SNP panel for detecting population structure
396 Studies on population structure of marine organisms based on neutral
397 markers typically demonstrate significant but weak genetic differentiation (Nielsen et
398 al. 2009; Milano et al. 2014). Neutral markers are affected by demography and the
399 evolutionary history of populations; therefore they evolve as a result of genetic drift
400 and migration (Luikart et al. 2003). For this reason, even low levels of migrant
401 exchange can maintain genetic homogeneity between populations over long periods of

402 time (Cano et al. 2008; Allendorf et al. 2010). The neutral SNP panel used in the
403 present study provided evidence for significant but weak genetic differentiation
404 between countries, in accordance with the findings of microsatellite markers, which
405 displayed similar levels of genetic differentiation between Tasmania and New
406 Zealand (F_{ST} = 0.029) (Morgan et al. 2013).
407
408 Performance of the outlier SNP panel for detecting population structure and
409 implications for fisheries management
410 In the present study, the much higher and statistically significant F_{ST} value
411 exhibited by the outlier SNP panel (F_{ST} = 0.134) demonstrated that loci putatively
412 under selection have high power for detecting genetic structure between Australia and
413 New Zealand and therefore could be used for determining country of origin in *J.*
414 *edwardsii*.
415 Molecular techniques for seafood authentication are increasingly used to
416 monitor fish stocks that are still commercially viable but are becoming threatened by
417 overfishing, and to protect the consumer from fraudulent practices (Ogden 2008;
418 Sorenson et al. 2013; Larraín et al. 2014). DNA barcoding is one of the preferred
419 techniques due to its applicability to degraded material, low DNA requirement, simple
420 protocol, time efficiency and reproducibility (Wong and Hanner 2008). However,
421 DNA barcoding is more effective for inter-specific differentiation, since it targets the
422 mitochondrial cytochrome *c* oxidase I (COI) gene which can be highly conserved
423 between subpopulations of the same species (Ogden 2008). In contrast, techniques
424 that target the nuclear genome provide the potential for intra-species assignment to
425 population of origin (Nielsen et al. 2012a).

426 Trials using microsatellite markers for food traceability have been carried out
427 with limited success in other commercial marine species. For example, a panel of nine
428 microsatellites produced up to 50% of correct assignments to country of origin of the
429 mussel *Mytilus chilensis* in the presence of a global F_{ST} lower than 0.042 (Larraín et
430 al. 2014). These authors suggested that having more informative loci and using SNP
431 markers could improve assignment success. In contrast, higher success (92%) was
432 obtained when combining 13 microsatellite markers with data from the mitochondrial
433 control region to accurately assign individuals to ocean of origin in Atlantic and
434 Pacific blue marlin stocks with low genetic differentiation ($F_{ST} < 0.01$) (Sorenson et
435 al. 2013). However, the suitability of loci putatively under selection for more reliable
436 assignment to population of origin has been recognized (Martinsohn and Ogden 2009;
437 Nielsen et al. 2012a) and recent studies have successfully demonstrated the power of
438 outlier SNPs for this purpose (Araneda et al. 2016).
439 The outlier SNP panel obtained in the present study demonstrated strong
440 genetic differentiation of *J. edwardsii* between Australia and New Zealand ($F_{ST} =$
441 0.134). Based on these markers, complimentary high throughput technologies that
442 rely on prior sequence information, such as target capture or loop-mediated isothermal
443 amplification (LAMP assay) (Tomita et al. 2008) could be used to differentiate
444 lobsters caught in New Zealand from those caught in Australian waters in order to
445 avoid mislabeling of country of origin. In particular, LAMP assay amplifies specific
446 regions of the DNA with high specificity, efficiency, rapidity and low cost for
447 preparation and visualization of results (Tomita et al. 2008). This assay is being
448 increasingly used for clinical diagnosis of infectious diseases in developing countries
449 since it does not require expensive laboratory equipment (i.e. a thermal cycler) and it
450 can be performed in 1 hour (Parida et al. 2008). For the particular case of *J.*

451 *edwardsii*, primers could be specifically designed to target regions of the genome
452 containing the outlier SNPs with highest loadings identified in Fig. S2 that account for
453 most of the divergence between Australia and New Zealand (Martinsohn and Ogden
454 2009).

455 Genetic structure detected using outlier loci could suggest differences at
456 regions of the genome putatively subject to selection. This divergence could be due to
457 local adaptation to environmental conditions or to post-settlement mortality of
458 unsuited genotypes (Holt and Gaines 1992; Caley et al. 1996; Marshall et al. 2010).
459 Local adaptation to environmental conditions will result in genetic divergence
460 between populations in the presence of high self-recruitment (Holt and Gaines 1992;
461 Sanford and Kelly 2011). Herein, we demonstrated very high levels of self-
462 recruitment within each country using outlier loci. Also, post-settlement mortality of
463 unsuited genotypes will preserve the local genetic pool by removing migrants that are
464 not fit to survive under particular environmental conditions (DeWitt et al. 1998;
465 Marshall et al. 2010). Currently it is impossible to unravel the degree at which self-
466 recruitment and post-settlement mortality are contributing to genetic differentiation of
467 *J. edwardsii* between Australia and New Zealand and this is beyond the scope of this
468 study.

469 Based on the high genetic divergence resulting from the outlier SNP panel,
470 we hypothesize that local conditions may have helped shape patterns of genetic
471 diversity within adaptive regions of the genome in these populations, as shown in
472 other studies (see Corander et al. 2013; Fraser et al. 2014), however in the absence of
473 any reference genes or transcriptome for *J. edwardsii* or closely related species it is
474 impossible to ascertain this. We can only speculate that differences in environmental
475 conditions between sampling sites may be driving differences at the outlier loci.

476 Empirical evidence suggests that *J. edwardsii* are adapted to specific local conditions,
477 but there is also extensive evidence of very high phenotypic plasticity. For example,
478 growth rates in this species are highly variable and are mainly determined by
479 temperature, density and food availability (Annala and Bycroft 1985; Jeffs and James
480 2001). Site-specific differences in carapace coloration and growth rates of this species
481 have been reported in Australia (Punt et al. 1997; McGarvey et al. 1999). Individuals
482 inhabiting deep waters are white-colored due to a diet with low concentration of
483 carotenoid pigments and lower nutritional value, which can also impact growth
484 negatively (McGarvey et al. 1999). Translocation experiments of white-colored
485 lobsters into shallow areas demonstrated a change in coloration, growth rates and
486 body condition after 12 months of translocation (Chandrapavan et al. 2009;
487 Chandrapavan et al. 2010; Green et al. 2010; Chandrapavan et al. 2011). Therefore,
488 even when self-recruitment could help retain locally adapted genotypes, phenotypic
489 plasticity can also act to promote growth of *J. edwardsii* in certain environments.
490 Greater density of SNPs, together with improved but as yet unavailable
491 genetic resources for *J. edwardsii* or closely related species (such as reference
492 genome and transcriptome datasets), would provide further insight into potential
493 genetic evidence of adaption to local environments. Seascape genetics could also help
494 coupling local environmental conditions to genetic distance in order to explain
495 patterns of genetic divergence between populations (Giles et al. 2015; Saenz-Agudelo
496 et al. 2015).
497
498 **Conclusions**
499 In this study, we used ddRADseq and Illumina MiSeq next generation
500 sequencing to explore the genetic connectivity of southern rock lobster populations

501 from Australia and New Zealand. Data from two SNP panels are presented, from
502 which a panel of 121 outlier markers have allowed us to identify clear genetic
503 structure in *J. edwardsii* populations between the two countries as well as high levels
504 of self-recruitment. In addition, highly significant F_{ST} values estimated from the
505 outlier SNP panel could be indicative of local adaptation driving the genetic
506 differentiation between countries. This is particularly important in a commercial
507 species managed by different agencies. Therefore, the outlier SNP panel developed in
508 the present study could be used to differentiate New Zealand from Australian lobsters
509 and therefore be useful for food traceability. We believe that more extensive
510 sampling, including sites along the whole distribution of the species, could identify
511 source and sink regions within each country precisely, which would also help
512 management decision-making in Australia and New Zealand. Finally, continued
513 development of genomics resources, such as transcriptome sequencing, gene
514 characterization and quantitative trait locus discovery is needed in order to explore the
515 link between genotype, phenotype and the environment.
516
517 Data accessibility: Reference loci sequences available through Dryad,
518 doi:10.5061/dryad.5c960.
519
520 Funding: Funding for this research was provided by an Australian Research Council
521 Linkage Project grant (Project No. LP120200164) from B.S.G., an Australian
522 Research Council Discovery Project grant (Project No. DP150101491) awarded to
523 JMS NPM, BG and JJB, Fisheries Research and Development Corporation grant
524 2015-025 as well as the Tasmanian Rock Lobster Fisherman's Association, the
525 Department of Primary Industries, Park Water and Environment (Tasmania,

526 Australia), Seafood Innovations Limited (Wellington, New Zealand) and the New
527 Zealand Rock Lobster Industry Council.

528

529 All Authors A declare that they have no conflict of interest.

530

531 Ethical approval: All applicable international, national, and/or institutional guidelines
532 for the care and use of animals were followed.

533

534 **References**

535 ABARE–BRS (2010) Australian fisheries statistics 2009. Canberra
536 Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of
537 conservation genetics. Nat Rev Genet 11:697-709. doi: 10.1038/nrg2844
538 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ
539 (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database
540 search programs. Nucleic Acids Res 25:3389-3402
541 Annala JH, Bycroft BL (1985) Growth rate of juvenile rock lobsters (*Jasus edwardsii*)
542 at Stewart Island, New Zealand. New Zeal J Mar Fresh 19:445-455
543 Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A
544 workbench to detect molecular adaptation based on a F(st)-outlier method.
545 BMC Bioinform 9:323. doi: 10.1186/1471-2105-9-323
546 Araneda C, Larraín MA, Hecht B, Narum S (2016) Adaptive genetic variation
547 distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine
548 environments. Ecol Evol. doi:10.1002/ece3.2110
549 Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of
550 population structure. Proc R Soc B 263:1619-1626. doi:
551 10.1098/Rspb.1996.0237
552 Benestan L, Gosselin T, Perrier C, Sainte-Marie B, Rochette R, Bernatchez L (2015)
553 RAD genotyping reveals fine-scale genetic structuring and provides powerful
554 population assignment in a widely distributed marine species, the American
555 lobster (*Homarus americanus*). Mol Ecol 24:3299–3315.
556 doi:10.1111/mec.13245

557 Booth JD, Phillips BF (1994) Early life history of spiny lobster. Crustaceana 66:271-
558 294. doi: 10.1163/156854094x00035
559 Bruce B, Griffin DA, Bradford R (2007) Larval transport and recruitment processes of
560 southern rock lobster. CSIRO, Hobart
561 Caley MJ, Carr MH, Hixon MA, Hughes TP, Jones GP, Menge BA (1996)
562 Recruitment and the local dynamics of open marine populations. Annu Rev
563 Ecol Syst 27:477-500
564 Candy JR, Campbell NR, Grinnell MH, Beacham TD, Larson WA, Narum SR (2015)
565 Population differentiation determined from putative neutral and divergent
566 adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an
567 anadromous Pacific smelt. Mol Ecol Res 15(6):1421-1434. doi: 10.1111/1755-
568 0998.12400
569 Cano JM, Shikano T, Kuparinen A, Merliä J (2008) Genetic differentiation, effective
570 population size and gene flow in marine fishes: implications for stock
571 management. J Integr Field Sci 5:1-10
572 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH, De Koning DJ
573 (2011) Stacks: building and genotyping loci de novo from short-read
574 sequences. G3 1:171-182. doi: 10.1534/g3.111.000240
575 Chandrapavan A, Gardner C, Linnane A, Hobday D (2009) Colour variation in the
576 southern rock lobster *Jasus edwardsii* and its economic impact on the
577 commercial industry. New Zeal J Mar Fresh 43:537-545
578 Chandrapavan A, Gardner C, Green BS (2010) Growth rate of adult rock lobsters
579 *Jasus edwardsii* increased through translocation. Fish Res 105:244-247
580 Chandrapavan A, Gardner C, Green BS (2011) Haemolymph condition of deep-water
581 southern rock lobsters (*Jasus edwardsii*) translocated to inshore reefs. Mar
582 Freshw Behav Phy 44:21-32
583 Corander J, Majander KK, Cheng L, Merila J (2013) High degree of cryptic
584 population differentiation in the Baltic Sea herring *Clupea harengus*. Mol Ecol
585 22:2931-2940. doi: 10.1111/mec.12174
586 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,
587 Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011)
588 The variant call format and VCFtools. Bioinformatics 27:2156-2158. doi:
589 10.1093/bioinformatics/btr330

590 DeWitt TJ, Sih A, Wilson DS (1998) Costs and limits of phenotypic plasticity. Trends
591 Ecol Evol 13:77-81. doi: 10.1016/S0169-5347(97)01274-3
592 Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator
593 v2: re-implementation of software for the estimation of contemporary
594 effective population size (Ne) from genetic data. Mol Ecol Res 14:209-214.
595 doi: 10.1111/1755-0998.12157
596 Eckman JE (1996) Closing the larval loop: linking larval ecology to the population
597 dynamics of marine benthic invertebrates. J Exp Mar Biol Ecol 200:207-237.
598 doi: 10.1016/s0022-0981(96)02644-5
599 Ellegren H (2014) Genome sequencing and population genomics in non-model
600 organisms. Trends Ecol Evol 29(1):51-63. doi: 10.1016/j.tree.2013.09.008
601 Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a
602 multilocus Bayesian method for the estimation of migration rates. Mol Ecol
603 16:1149-1166. doi: 10.1111/j.1365-294X.2007.03218.x
604 Fraser DJ, Debes PV, Bernatchez L, hutchings JA (2014) Population size, habitat
605 fragmentation, and the number of adaptive variation in a stream fish. Proc R
606 Soc B 281:20140370. doi:10.1098/rspb.2014.0370
607 Giles EC, Saenz-Agudelo P, Hussey NE, Ravasi T, Berumen ML (2015) Exploring
608 seascape genetics and kinship in the reef sponge *Stylissa carteri* in the Red
609 Sea. Ecol Evol 5:2487-2502. doi: 10.1002/ece3.1511
610 Green BS, Gardner C, Linnane A, Hawthorne PJ (2010) The good, the bad and the
611 recovery in an assisted migration. PLoS ONE 5:e14160. doi:
612 10.1371/journal.pone.0014160
613 Hedgecock D, Pudovkin AI (2011) Sweepstakes reproductive success in highly
614 fecund marine fish and shellfish: a review and commentary. Bull Mar Sci
615 87(4):971-1002. doi: 10.5343/bms.2010.1051
616 Henning F, Lee HJ, Franchini P, Meyer A (2014) Genetic mapping of horizontal
617 stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD
618 markers for dense linkage mapping. Mol Ecol 23:5224-5240. doi:
619 10.1111/mec.12860
620 Holt RD, Gaines MS (1992) Analysis of adaptation in heterogeneous landscapes -
621 implications for the evolution of fundamental niches. Evol Ecol 6:433-447.
622 doi: 10.1007/Bf02270702

623 Jacobsen MW, Pujolar JM, Bernatchez L, Munch K, Jian J, Niu Y, Hansen MM
624 (2014) Genomic footprints of speciation in Atlantic eels (*Anguilla anguilla*
625 and *A. rostrata*). Mol Ecol 23:4785-4798. doi: 10.1111/mec.12896
626 Jeffs AG, James PJ (2001) Sea-cage culture of the spiny lobster *Jasus edwardsii* in
627 New Zealand. Mar Freshwater Res 52(8):1419-1424. doi:10.1071/MF01064
628 Jeffs AJ, Gardner C, Cockcroft A (2013) *Jasus* and *Sagmariasus* species. In: Phillips
629 B (ed) Lobsters: Biology, management, aquaculture and fisheries, 2nd edn.
630 Blackwell Scientific Publications, Oxford, pp 259-288
631 Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal
632 components: a new method for the analysis of genetically structured
633 populations. BMC Genet 11:94. doi: 10.1186/1471-2156-11-94
634 Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-
635 wide SNP data. Bioinformatics 27:3070-3071. doi:
636 10.1093/bioinformatics/btr521
637 Karlin S, Altschul SF (1990) Methods for assessing the statistical significance of
638 molecular sequence features by using general scoring schemes. P Natl Acad
639 Sci USA 87:2264-2268
640 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature
641 Methods 9:357-U354. doi: 10.1038/NMETH.1923
642 Larraín MA, Díaz NF, Lamas C, Uribe C, Araneda C (2014) Traceability of mussel
643 (*Mytilus chilensis*) in southern Chile using microsatellite molecular markers
644 and assignment algorithms. Exploratory survey. Food Res Int 62:104-110. doi:
645 10.1016/j.foodres.2014.02.016
646 Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W, Daza R,
647 Erlich RL, Giannoukos G, Green L, Hollinger A, Hoover CA, Jaffe DB, Juhn
648 F, McCarthy D, Perrin D, Ponchner K, Powers TL, Rizzolo K, Robbins D,
649 Ryan E, Russ C, Sparrow T, Stalker J, Steelman S, Weiand M, Zimmer A,
650 Henn MR, Nusbaum C, Nicol R (2010) A scalable, fully automated process
651 for construction of sequence-ready barcoded libraries for 454. Genome Biol
652 11:R15. doi: 10.1186/gb-2010-11-2-r15
691 Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently
692 bottlenecked populations from allele frequency data. Conserv Biol 12:228-
693 237. doi: 10.1111/j.1523-1739.1998.96388.x/epdf

694 Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and
695 promise of population genomics: from genotyping to genome typing. Nat Rev
696 Genet 4:981-994. doi: 10.1038/nrg1226
697 Marshall DJ, Monro K, Bode M, Keough MJ, Swearer S (2010) Phenotype-
698 environment mismatches reduce connectivity in the sea. Ecol Lett 13:128-140.
699 doi: 10.1111/j.1461-0248.2009.01408.x
700 Martinsohn JT, Ogden R (2009) FishPopTrace – developing SNP-based population
701 genetic assignment methods to investigate illegal fishing. Forensic Sci Int:
702 Genetics Suppl Ser 2:294-296. doi:10.1016/j.fsigss.2009.08.108
703 Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC
704 (2015) Restriction site-associated DNA sequencing, genotyping error
705 estimation and de novo assembly optimization for popualtion genetic
706 inference. Mol Ecol Resour 15:28-41. doi: 10.1111/1755-0998.12291
707 McGarvey R, Ferguson GJ, Prescott JH (1999) Spatial variation in mean growth rates
708 at size of southern rock lobster, *Jasus edwardsii*, in South Australian waters.
709 Mar Freshwater Res 50:333-342. doi: 10.1071/Mf97172
710 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella
711 K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome
712 Analysis Toolkit: A MapReduce framework for analyzing next-generation
713 DNA sequencing data. Genome Res 20:1297-1303. doi:
714 10.1101/gr.107524.110
715 Merchant S, Wood DE, Salzberg SL (2014) Unexpected cross-species contamination
716 in genome sequencing projects. Peer J 2:e675. doi: 10.7717/peerj.675
717 Milano I, Babbucci M, Cariani A, Atanassova M, Bekkevold D, Carvalho GR,
718 Espiñeira M, Fiorentino F, Garofalo G, Geffen AJ, Hansen JH, Helyar SJ,
719 Nielsen EE, Ogden R, Patarnello T, Stagioni M, Tinti F, Bargelloni L (2014)
720 Outlier SNP markers reveal fine-scale genetic structuring across European
721 hake populations (*Merluccius merluccius*). Mol Ecol 23:118-135. doi:
722 10.1111/mec.12568
723 Morgan EMJ, Green BS, Murphy NP, Strugnell JM (2013) Investigation of genetic
724 structure between deep and shallow populations of the Southern rock lobster,
725 *Jasus edwardsii* in Tasmania, Australia. PLoS ONE 8:e77978. doi:
726 10.1371/journal.pone.0077978

727 Morin PA, Luikart G, Wayne RK, the SNP Workshop Group (2004) SNPs in ecology,
728 evolution and conservation. Trends Ecol Evol 19:208-216. doi:
729 10.1016/j.tree.2004.01.009
730 Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics
731 of marine fishes: identifying adaptive variation in space and time. Mol Ecol
732 18:3128-3150. doi: 10.1111/j.1365-294X.2009.04272.x
733 Nielsen EE, Cariani A, Mac Aoidh E, Maes GE, Milano I, Ogden R, Taylor M,
734 Hemmer-Hansen J, Babbucci M, Bargelloni L, Bekkevold D, Diopere E,
735 Grenfell L, Helyar S, Limborg MT, Martinsohn JT, McEwing R, Panitz F,
736 Patarnello T, Tinti F, Van Houdt JKJ, Volckaert FAM, Waples RS, Carvalho
737 GR, Albin JEJ, Baptista JMV, Barmintsev V, Bautista JM, Bendixen C, Berge
738 JP, Blohm D, Cardazzo B, Diez A, Espineira M, Geffen AJ, Gonzalez E,
739 Gonzalez-Lavin N, Guarniero I, Jerome M, Kochzius M, Krey G, Mouchel O,
740 Negrisolo E, Piccinetti C, Puyet A, Rastorguev S, Smith JP, Trentini M,
741 Verrez-Bagnis V, Volkov A, Zanzi A, Consortium F (2012a) Gene-associated
742 markers provide tools for tackling illegal fishing and false eco-certification.
743 Nat Commun 3:851. doi: 10.1038/Ncomms1845
744 Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012b) SNP calling,
745 genotype calling, and sample allese frequency estimation from new-generation
746 sequencing data. PLoS ONE 7(7):e37558. doi: 10.1371/journal.pone.0037558
747 Norman-López A, Pascoe S, Thébaud O, van Putten I, Innes J, Jennings S, Hobday A,
748 Green B, Plaganyi E (2014) Price integration in the Australian rock lobster
749 industry: implications for management and climate change adaptation. Aust J
750 Agr Resour Ec 58:43–59. doi: 10.1111/1467-8489.12020
751 Ogden R (2008) Fisheries forensics: the use of DNA tools for improving compliance,
752 traceability and enforcement in the fishing industry. Fish Fish 9:462-472. doi:
753 10.1111/j.1467-2979.2008.00305.x
754 Paetkau D, Slade R, Burden M, Estoup A (2004) Direct, real-time estimation of
755 migration rate using assignment methods: a simulation-based exploration of
756 accuracy and power. Mol Ecol 13:55-65
757 Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation.
758 Annu Rev Ecol Syst 25:547-572. doi: 10.1146/Annurev.Ecolsys.25.1.547

759 Paradis E (2010) pegas: an R package for population genetics with an integrated-
760 modular approach. Bioinformatics 26:419-420. doi:
761 10.1093/bioinformatics/btp696
762 Parida M, Sannarangaiah S, Dash PK, Rao PVL, Morita K (2008) Loop mediated
763 isothermal amplification (LAMP): a new generation of innovative gene
764 amplification technique; perspectives in clinical diagnosis of infectious
765 diseases. Rev Med Virol 18:407-421. doi: 10.1002/rmv.593
766 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest
767 RADseq: an inexpensive method for *De Novo* SNP discovery and genotyping
768 in model and non-model species. PLoS ONE 7:e37135. doi:
769 10.1371/journal.pone.0037135
770 Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004)
771 GeneClass2: a software for genetic assignment and first-generation migrant
772 detection. J Hered 95:536-539
773 Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using
774 multilocus genotype data. Genetics 155:945-959
775 Punt AE, Kennedy RB, Frusher SD (1997) Estimating the size-transition matrix for
776 Tasmanian rock lobster, *Jasus edwardsii*. Mar Freshwater Res 48:981-992.
777 doi: 10.1071/Mf97017
778 Rannala B, Mountain JL (1997) Detecting immigration by using multilocus
779 genotypes. Proc Natl Acad Sci USA 94:9197-920
780 Saenz-Agudelo P, Dibattista JP, Piatek MJ, Gaither MR, Harrison HB, Nanninga GB,
781 Berumen ML (2015) Seascape genetics along environmental gradients in the
782 Arabian Peninsula: insights from ddRAD sequencing of anemonefishes. Mol
783 Ecol 24:6241-6255. doi: 10.1111/mec.13471
784 Sanford E, Kelly MW (2011) Local adaptation in marine invertebrates. Ann Rev Mar
785 Scie 3:509-535. doi: 10.1146/annurev-marine-120709-142756
786 Shanks AL, Grantham BA, Carr MH (2003) Propagule dispersal distance and the size
787 and spacing of marine reserves. Ecol Appl 13:S159-S169
788 Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. Biol Bull-
789 Us 216:373-385
790 Sorenson L, McDowell JR, Knott T, Graves JE (2013) Assignment test method using
791 hypervariable markers for blue marlin (*Makaira nigricans*) stock

identification. Conserv Genet Resour 5:293-297. doi: 10.1007/s12686-012-9747-x

Statistics New Zealand and licensed by Statistics NZ for re-use under the Creative Commons Attribution 4.0 International license. http://www.stats.govt.nz

Teske PR, Sandoval-Castillo J, van Sebille E, Waters J, Beheregaray LB (2015) On-shelf larval retention limits population connectivity in a coastal broadcast spawner. Mar Ecol Prog Ser 532:1-12. doi: 10.3354/meps11362

Thomas L, Bell JJ (2013) Testing the consistency of connectivity patterns for a widely dispersing marine species. Heredity 111:345-354. doi: 10.1038/hdy.2013.58

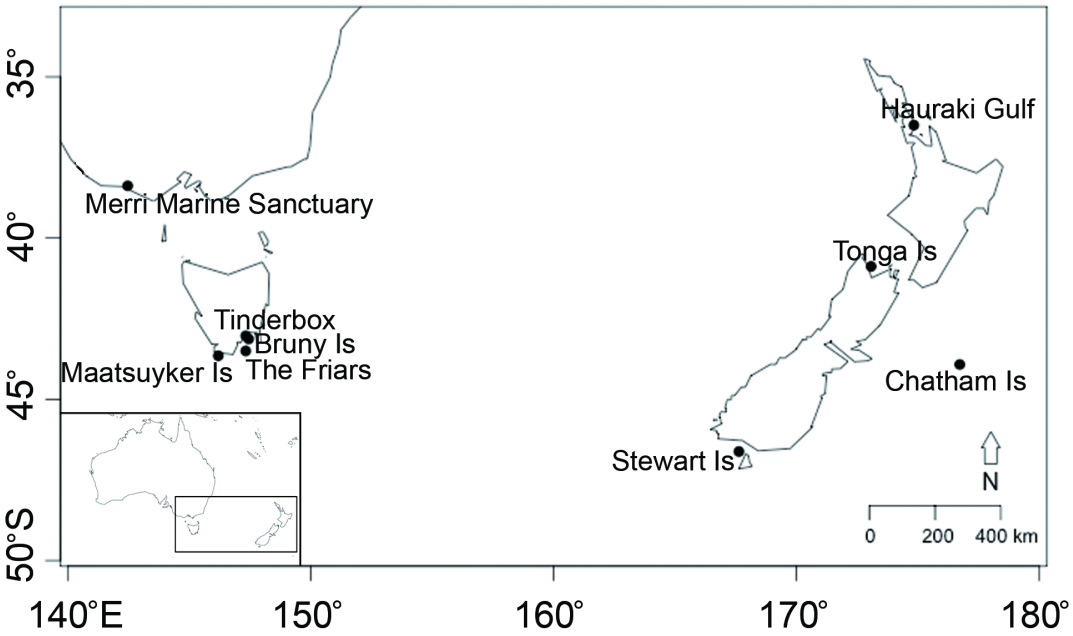
Tomita N, Mori Y, Kanda H, Notomi T (2008) Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. Nat Protoc 3:877-882. doi: 10.1038/nprot.2008.57

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol Appl 3:244-262. doi: 10.1111/j.1752-4571.2009.00104.x

Wei KJ, Wood AR, Gardner JPA (2013) Population genetic variation in the New Zealand greenshell mussel: locus-dependent conflicting signals of weak structure and high gene flow balanced against pronounced structure and high self-recruitment. Mar Biol 160:931-949. doi: 10.1007/s00227-012-2145-9

Wong EHK, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood. Food Res Int 41:828-837. doi: 10.1016/j.foodres.2008.07.005

Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46. doi: 10.1186/gb-2014-15-3-r46



825
826 **Fig. 1** Sampling sites in Australia and New Zealand

827
828
829
830
831
832
833
834
835
836
837

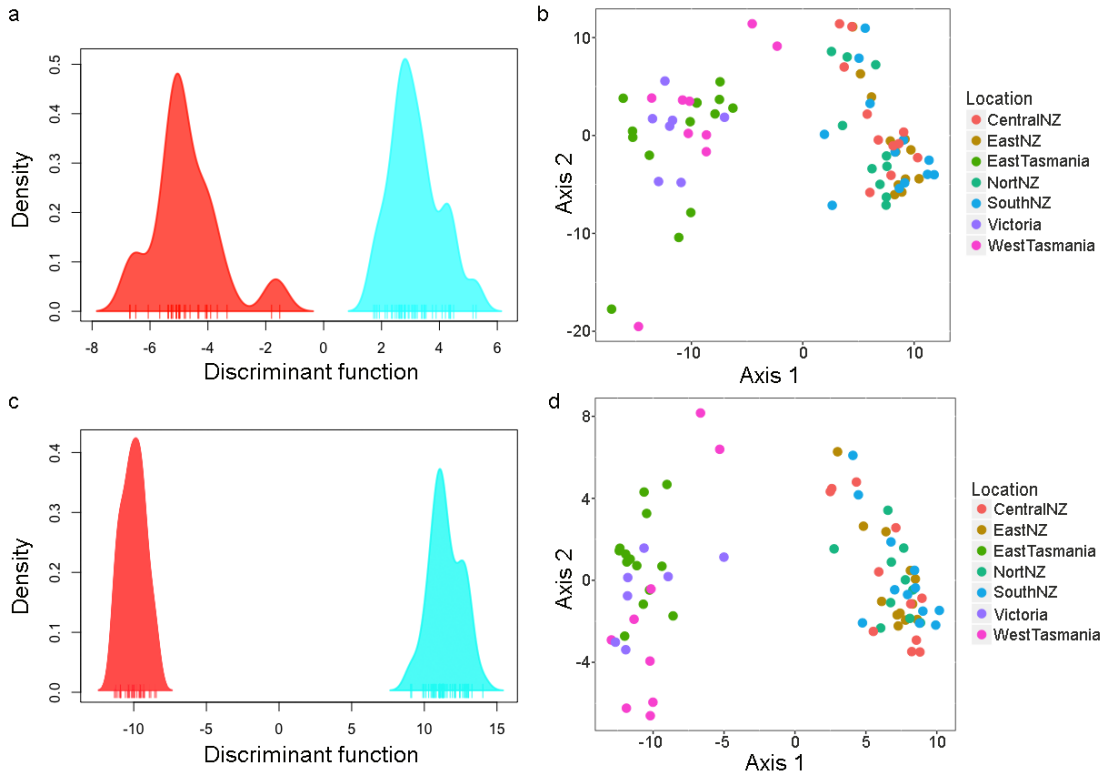


Fig. 2 First principal component resulting from the Discriminant Analysis of Principal Components (DAPC) using (a) the neutral SNP panel and (c) the outlier SNP panel. The red group represents 30 Australian individuals while the blue group represents 45 New Zealand individuals. Principal component 1 and 2 resulting from a Principal Component Analysis (PCA) using (b) the neutral SNP panel and (d) the outlier SNP panel. Sampling regions are represented by colors

Table 1 Sampling sites for the *J. edwardsii* collected in Australia and New Zealand. Number of individuals sequenced (n) and final number of individuals (n_f) after filtering for missing data and removal of replicates are reported for each sampling site

Country	Region	Sampling site	Sampling year	Stage	Latitude	Longitude	n	n_f
Australia	Victoria	Merri Marine Sanctuary (MMS)	2013	Adult	38°23'S	142°28'E	8	7
	West Tasmania	Maatsuyker Island (MAA)	2014	Adult	43°39'S	146°12'E	11	9
	East Tasmania	The Friars (FSX)	2014	Adult	43°30'S	147°20'E	8	4
	East Tasmania	Bruny Island (BRU)	2013	Adult	43°08'S	147°27'E	4	3
	East Tasmania	Tinderbox (TXX)	2013	Adult	43°02'S	147°20'E	9	7
New Zealand	North NZ	Hauraki Gulf (HGU)	2011	Adult	36°30'S	174°50'E	13	10
	Central NZ	Tonga Island (TIS)	2013	Adult	40°53'S	173°04'E	11	11
	East NZ	Chatham Islands (CHI)	2013	Adult	43°55'S	176°43'E	12	12
	South NZ	Stewart Island (SIS)	2013	Adult	46°38'S	167°37'E	12	12

Table 2 Descriptive statistics for Australia and New Zealand given by the neutral and outlier SNP panels

	N	N_{loci}	H_O (± SD)	H_E (± SD)	F_{IS}	F_{ST}	*P*
Neutral SNP panel							
New Zealand	45	535	0.633 (± 0.279)	0.399 (± 0.118)	-0.588		
Australia	30	535	0.518 (± 0.268)	0.369 (± 0.126)	-0.401	0.022	<0.0001
Outlier SNP panel							
New Zealand	45	121	0.386 (± 0.319)	0.264 (± 0.176)	-0.463		
Australia	30	121	0.355 (± 0.271)	0.273 (± 0.169)	-0.297	0.134	<0.0001

N sample size, N_{loci} number of SNPs, H_O mean observed level of heterozygosity, H_E mean expected level of heterozygosity, F_{IS} fixation index (inbreeding coefficient), F_{ST} pairwise fixation index between Australia and New Zealand

Table S1 Number of inter- and intralibrary technical replicates

	Intralibrary technical replicates	Interlibrary technical replicates
Library 1	NA	NA
Library 2	NA	1
Library 3	15	5
Library 4	NA	4

Supplemental Materials and methods

Migration rates

Recent migration rates between Australia and New Zealand were inferred with a subset of 86 neutral SNPs that exhibited the highest F_{ST} values ($F_{ST} > 0.05$) in the program BayesAss v.3.0.4 (Wilson and Rannala 2003) using a Markov chain Monte Carlo (MCMC) for 10 x 10^6 iterations with a burn-in length of 10^6 iterations.

Supplemental Results

Migration rates

The subset of 86 neutral SNPs with F_{ST} values larger than 0.05 revealed 98% self–recruitment in Australia and 67% in New Zealand. According to these results, approximately 33% of the New Zealand populations are migrants from Australia as well as approximately 0.02% of individuals from the Australian population originated in New Zealand (Table S4).

Table S4 Migration rates between Australia and New Zealand resulting from 86 neutral SNPs (±SD) with $F_{ST} > 0.05$. Rows are source populations and columns are sink populations

	Australia	New Zealand
Australia	0.9792(±0.0140)	0.0208(±0.0140)
New Zealand	0.3262(±0.0070)	0.6738(±0.0070)

925
926

927 **Supplemental References**

928 Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177-1191

929

930