Genetic structure of the grey side-gilled sea slug
(Pleurobranchaea maculata) in coastal waters of New Zealand

Yeşerîn Yıldırım1*, Marti J. Anderson1,2, Selina Patel3, Craig D. Millar3 and Paul B. Rainey1,4,5

1New Zealand Institute for Advanced Study, Massey University Albany, Auckland, New Zealand.
2Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand.
3School of Biological Sciences, University of Auckland, New Zealand
4Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön, Germany.
5Chemistry Biology Innovation (CNRS UMR 8231), Ecole supérieure de physique et de chimie industrielles de la Ville de Paris (ESPCI ParisTech), PSL Research University, Paris, France.

*Corresponding author

Correspondence to: Yeşerîn Yıldırım, Center for Ecology and Evolution in Microbial Systems (EEMiS), Department of Biology and Environmental Science, Linnaeus University, Kalmar, Sweden.

E-mail: yeserin.yildirim@lnu.se; Ph: +46 72 5949407.
Abstract

*Pleurobranchaea maculata* is a rarely studied species of the Heterobranchia found throughout the south and western Pacific – and recently recorded in Argentina – whose population genetic structure is unknown. Interest in the species was sparked in New Zealand following a series of dog deaths caused by ingestions of slugs containing high levels of the neurotoxin tetrodotoxin. Here we describe the genetic structure and demographic history of *P. maculata* populations from five principle locations in New Zealand based on extensive analyses of 12 microsatellite loci and the *COI* and *CytB* regions of mitochondrial DNA (mtDNA). Microsatellite data showed significant differentiation between northern and southern populations with population structure being associated with previously described regional variations in tetrodotoxin concentrations. However, mtDNA sequence data did not support such structure, revealing a star-shaped haplotype network with estimates of expansion time suggesting a population expansion in the Pleistocene era. Inclusion of publicly available mtDNA sequence from Argentinian sea slugs did not alter the star-shaped network. We interpret our data as indicative of a single founding population that fragmented following geographical changes that brought about the present day north-south divide in New Zealand waters. Lack of evidence of cryptic species supports data indicating that differences in toxicity of individuals among regions are a consequence of differences in diet.

**Keywords:** *Pleurobranchaea maculata*, tetrodotoxin, microsatellites, cytochrome c oxidase subunit 1, cytochrome b, population structure

**Short Title:** Genetic structure of *Pleurobranchaea maculata*
Introduction

The grey side-gilled sea slug (*Pleurobranchaea maculata*) is an opportunistic carnivore that feeds on invertebrates including sea anemones, marine worms and other molluscs [1]. It is native to New Zealand (NZ), southeastern Australia, China, Sri Lanka and Japan where it is found in habitats ranging from sandy sediments to rocky reefs, and from shallow sub-tidal flats to depths of 300 m [1, 2]. Little is known about the life history of the species but studies of comparative development report the production of planktotrophic veligers that hatch within eight days and remain planktonic for three weeks before juveniles settle [1-3].

In late 2009 this otherwise little-known sea slug attracted attention after it was implicated in dog deaths on beaches in Auckland [4]. Forensic analyses revealed that deaths were a consequence of tetrodotoxin (TTX) poisoning associated with ingestion of *P. maculata* [4]. This was the first time that TTX had been reported in NZ and in a species of the taxonomic clade Heterobranchia [4]. *P. maculata* that have recently invaded coastal waters of Argentina also contain TTX [5, 6].

TTX is a potent neurotoxin found in numerous terrestrial and marine organisms, but neither the origin of TTX nor the causes of variation in TTX levels among species are understood. The structure of TTX suggests a microbial origin [7] and while certain microbes have been implicated in TTX production (reviewed in Magarlamov *et al.*, 2017), all such claims have been refuted [8, 9]. Nonetheless, while not excluding a microbial origin, there is recognition that TTX in animals is often acquired via diet. For example, variability in TTX levels found in puffer fish has been attributed to exposure to toxic food sources (reviewed in Noguchi and Arakawa, 2008). For *P. maculata*, there is mounting evidence that toxin accumulation occurs through feeding [10-13]. An
alternate possibility is that TTX arises from commensal or symbiotic microorganisms that are associated with *P. maculata* [13], but no TTX-producing bacteria have been found [14, 15].

Studies of individual and temporal differences in TTX concentration has established that *P. maculata* populations from northern regions of the North Island (Whangarei, Auckland, Tauranga) have high TTX concentrations (the highest average being 368.7 mg kg$^{-1}$ per individual), while populations from the South Island (Nelson and Kaikoura) have trace amounts of TTX (<0.5 mg kg$^{-1}$) or none at all [2, 4, 10, 13]. A recent study reported TTX concentrations as high as 487 mg kg$^{-1}$ [13]. Significant individual and seasonal variations have also been observed [2]. A single individual obtained from Wellington in the south of the North Island was found to have a low concentration of TTX (2.2 mg kg$^{-1}$) supporting the notion of a geographical cline [2].

New Zealand (NZ) extends from the subtropical Kermadec Islands (29° S) to the sub-Antarctic Campbell Island group (52° S), it comprises ~700 islands affording ~15,000 km of coastline [16, 17]. Subtropical and sub-Antarctic waters converge at the juncture between the two major landmasses (the North and South Islands) creating a complex oceanography of currents and eddies.

Phylogenetic and population genetic studies of NZ marine organisms have shown that different species manifest a range of genetic structures (reviewed in Gardner *et al*, 2010) with a common pattern being a disjunction between northern and southern populations. Such disjunctions are evident in bivalves [19], teleosts [20], polyplacophores [21], echinoderms [22] and arthropods [23].

The genetic structure of *P. maculata* is unknown, but variation in the established differences in toxicity between northern and southern populations suggests that
geographic variability in TTX concentration correlates genetic structure – even the possibility that northern and southern populations define separate species. Here we test this hypothesis using a combination of microsatellite and mitochondrial DNA (mtDNA) sequence markers. Analysis of more rapidly evolving microsatellites showed evidence of a genetic break along the predicted north-south divide. However, no such structure was apparent from analysis of mtDNA data.

**Materials and methods**

**Sampling, DNA extraction and tetrodotoxin assay**

A total of 156 samples were collected from nine regions around New Zealand between 2009 and 2013 (Fig 1 and S1 Table). DNA was extracted as described in Yıldırım et al. [24]. The Tauranga (TR) population included samples from Tauranga Harbour whereas the Auckland (AKL) population included samples from Tamaki Strait and Waitemata Harbour. Some samples were from the studies of Wood *et al.* [2] and Khor *et al.* [10].

![Sampling locations for the Pleurobranchaea maculata individuals.](https://example.com/sampling_map.png)

The numbers within the circles indicate the sampling size of each region. The arrows show magnified maps of Auckland and Tauranga. Populations containing *P. maculata* individuals with high, and low and trace amounts of tetrodotoxin concentrations in red and blue colour, respectively.

At the outset of this study there was limited knowledge of the toxicity of *P. maculata* individuals from Wellington (WL) as only one individual had been previously tested [2]. To obtain a better understanding, the TTX concentration of eight (of eighteen) individuals collected from WL in October 2012 was determined as described in Khor et
al. [10]. The TTX assay was performed at the Cawthron Institute (Nelson) using a liquid chromatography-mass spectrophotometry method that is described in McNabb et al. [4].

Population-level analyses were performed only for five populations which are Ti Point (TP), AKL, TR, Wellington (WL) and Nelson (NL) due to the small sample sizes of the other locations. TP, AKL and TR, which included highly toxic individuals [10] were designated as the “northern cluster”, whereas the WL and NL population, which contained slightly toxic and non-toxic individuals [2, 10, 11] were designated as the “southern cluster”.

Genotyping

Twelve microsatellite markers (Pm01, 02, 07, 08, 09, 10, 11, 13, 17, 19, 20 and 23) [24] were genotyped for 149 samples. PCR amplification and genotyping procedures for the primers were as described in Yıldırım et al. [24] with some modifications (S2 Table). Details regarding amplification and genotyping processes are described in the Supporting Information.

A 1060 bp and 1153 bp region of mitochondrial cytochrome B (CytB) and cytochrome oxidase I (COI) genes, respectively, were amplified and sequenced in all 156 P. maculata individuals. For details regarding the primer pairs and amplification see Supporting Information and S3 Table. Geneious Pro 6.1.6 (Biomatters, New Zealand) was used to trim, assemble, align and concatenate the resulting DNA sequences.

Statistical analysis

Genetic Diversity

Microsatellite genotyping data were tested for scoring errors due to stuttering, null alleles, and large allele dropout using MICRO-CHECKER v.2.2 [25]. Departures from
Hardy-Weinberg equilibrium (HWE) were estimated using Nei’s [26] inbreeding coefficient $G_{IS}$ with 100,000 permutations using GenoDive v2.0b25 [27]. FreeNA [28] was used to estimate null allele frequency as in Dempster \textit{et al.} [29]. Interference of putative null alleles on genetic differentiation between sampling sites was determined by calculating global and pair-wise $F_{ST}$ values [30], either with or without “exclusion of null alleles” [28]. Linkage disequilibrium (LD) between pairs of loci was estimated in FSTAT v2.9.3.2 [31] and the significance of LD was determined by applying a Bonferroni correction. The total number of alleles ($N_a$), allele frequencies, observed heterozygosity ($H_o$), unbiased expected heterozygosity ($H_e$) corrected for a small sampling size [26], and private alleles ($P_A$) per locus and population were calculated using GenAlex [32]. Allelic richness ($A_r$) was calculated using the rarefaction method implemented in ADZE v.1.0 [33]. $H_e$ and $A_r$ were used to compare the amount of genetic diversity among populations from different regions using one-way ANOVA ([http://vassarstats.net](http://vassarstats.net)).

For mtDNA, several estimates of genetic diversity, including the number of singletons ($S_i$), haplotypes ($Hap$) and segregating sites ($S$), the average number of nucleotide differences between sequences ($k$) [34], haplotype ($h$) and nucleotide diversity ($\pi$) [35] were calculated for the \textit{CytB}, \textit{COI} and concatenated sequences for each sampling location using DnaSP 5.10.0.1 [36].

**Population Structure**

For microsatellite data, global differentiation and pairwise differentiation between each pair of populations was investigated using various differentiation estimators, including a log-likelihood ratio (G)-based test [37], fixation index $F_{ST}$ [30], standardised fixation index $G''_{ST}$ [38], and Jost’s [39] differentiation ($Dest$). The statistical power to detect
true population differentiation and α-error probability were assessed in POWSIM v4.1 [40]. STRUCTURE v2.3.4 [41] was used to determine the probable number of distinct populations (K) and individuals were assigned to populations using a Bayesian assignment approach. Parameters were set to 5,000,000 MCMC iterations with a burn-in of 500,000 values of K between one and ten, with a series of ten independent replicates for each K value, assuming an admixture model and correlated allele frequencies across the populations, both with and without introducing a priori sampling location. The most likely value of K was resolved using the ΔK method [42] with the Structure Harvester v0.6.93 [43], and the results were introduced to the CLUMPP v1.1.2 software [44]. Destruct v1.1 [45] was used to visualize the results.

AMOVA [46] was performed in ARLEQUIN to determine the hierarchical genetic structure based on R_{ST} [47] and F_{ST} [30] for microsatellite data. Two separate calculations were performed: one with structure and one without structure. For the former, individuals from all regions were grouped; for the latter, the TP, AKL and TR populations were grouped into a northern cluster and the WL and NL populations were grouped into a southern cluster. This nested design was based on the results of population structure suggested by F-statistics, multivariate and STRUCTURE analyses of microsatellite data.

For the mtDNA sequences, POPART v1 (http://www.leigh.net.nz/software.shtml) was used to create a median joining haplotype network (MJN) [48]. We created an additional MJN for shorter COI sequences (624 bp) in order to accommodate four P. maculata samples obtained from individuals isolated in Argentina (Farias et al., 2016). A saturation test was performed in DAMBE v6.2.9 [49] using the test by Xia et al. [50]. The proportion of invariant sites (P_{inv}) was estimated by Jmodeltest v0.1.1 [51] with the
Akaike information criteria (AIC). The $P_{\text{inv}}$ values (0.844 and 0.789 for $CytB$ and $COI$, respectively) obtained from the most likely models suggested by the software (HKY+I and TrN+I for $CytB$ and $COI$, respectively) and default settings for other parameters were used for the calculations. Haplotype-frequency-based $F_{ST}$ [26] and distance-based $\Theta_{ST}$ [46] were calculated in ARLEQUIN to estimate population differentiation. For $\Theta_{ST}$, the Tamura-Nei mutational model [52] was used for both genes as being the closest models to the ones suggested as most likely to explain mtDNA data by Jmodeltest. For concatenated sequences hierarchical structure was investigated by AMOVA in ARLEQUIN using both F-statistics and $\Phi$-statistics, and with the same structure pattern used for the microsatellite data analysis (TP, AKL and TR were grouped into a northern cluster, and WL and NL were grouped into a southern cluster).

Patterns of differentiation were also analysed using a multivariate approach. For microsatellite data, the Manhattan distance ($DM$), which calculates the mean character differences between individuals, and clonal distances ($DCL$), which assumes a stepwise mutational model (SMM) [46], were used. For mtDNA data, a distance matrix ($D_{SEQ}$) was calculated as a standardized bp difference between every pair of individuals.

Statistical analyses on resulting distance matrices were done using PRIMER v6 [53] with PERMANOVA+ [54]. Patterns of inter-sample distances were visualized using non-metric multi-dimensional scaling ordination (MDS) [55]. Permutational multivariate analysis of variance (PERMANOVA) [56, 57] was used to formally test for differences in genetic structures among different locations and canonical analysis of principal coordinates (CAP) [54, 58] was used to discriminate among specific populations and identify their distinctiveness, using leave-one-out allocation success. PERMDISP was used to test the null hypothesis of homogeneity of within-group dispersions among populations [59]. All permutation tests used 10,000 permutations.
A maximum likelihood (ML) tree using the Tamura-Nei mutational model [34] with default settings was reconstructed for 44 P. maculata individuals from this study and three Argentinian P. maculata individuals using COI sequences [5] (redundant sequences were removed) in MEGA7 [60]. COI sequences of five species from the same family (Pleurobranchidae, Genbank codes in brackets) including Pleurobranchaea meckeli (KU365727.1), Pleurobranchaea nevaezelandiae, Pleurobranchus peronii (KM521745.1), Berthella ocellata (KM521694.1) and Berthellina citrina (KM521694.1) were used as outgroups. The analysis involved 200 informative positions of 616. The phylogeny was tested with 1,000 bootstrap replicates.

**Migration**

The microsatellite data were analysed with GeneClass2 [61] to identify the first-generation migrants using the Bayesian criterion of Rannala and Mountain [62] and the $L_{\text{home}}/L_{\text{max}}$ likelihood, with a threshold $p$-value of 0.01 and a Monte-Carlo resampling algorithm [63].

**Neutrality tests**

BOTTLENECK v1.2 [64] was used to test the possibility of recent population reduction for microsatellite data assuming SMM and two-phase models (TPM) with default settings using a Wilcoxon signed rank test [64]. A possible sign of a recent bottleneck was investigated also by a mode-shift analysis [65].

Deviations from neutrality and demographic changes within and across the populations were calculated with Tajima’s $D$ [66], Fu’s $Fs$ [67] and mismatch distribution analysis in ARLEQUIN for the concatenated mtDNA sequences. The null hypothesis of expansion was statistically tested with the sum of squared deviations (SSD) from the expected values [68] and Harpending’s *raggedness* index [69]. Mismatch distribution
analysis also estimated tau (τ), which is the age of the population expansion. The approximate date of population expansion across all samples was calculated for COI as outlined by Schenekar and Weiss [70] using http://www.unigraz.at/zoowww/mismatchcalc/index.php by converting τ to time since expansion (t) in years using the formula \( t = \frac{\tau}{2\mu k} \), where \( \mu \) is the mutation rate per site per generation, and \( k \) is the sequence length [71]. One year of generation time and both 5.3% divergence/mya (average \( \mu \) for marine invertebrate COI sequences) [72], a compatible mutation rate estimated for a planktotrophic heterobranchia species Costasiella ocellifera (Ellingson and Krug, 2015), were assumed. McDonald and Kreitman’s [73] neutrality test was performed pooling all P. maculata COI sequences (1153 bp) in DnaSP using P. meckeli COI sequences as an outgroup species. Fisher's exact test (two-tailed) was used to identify significant deviations from neutrality.

RESULTS

Tetrodotoxin levels in P. maculata from Wellington

Previous analyses have established that northern WH, AKL, TR, and CR populations have high levels of TTX, marking these populations as “toxic”, while southern populations from NL and KK are recorded as containing either trace, or no TTX [2, 4, 10, 13]. For WL populations, previous measurements existed for only one individual documented as having a low level of TTX (2.2 mg/kg) [2]. For this study, 18 slugs were obtained from WL of which eight randomly chosen individuals were subject to TTX assay. Three contained extremely low concentrations (0.12, 0.16 and 0.5 mg/kg) of TTX. No TTX was detected in the remaining five individuals. Accordingly, the WL, NL and KK samples (the southern cluster) were classified as “non-toxic”.
Genetic diversity

Microsatellite analyses

All loci were highly polymorphic with between five and 23 alleles for each locus (diversity statistics in Table 1 and S4 Table). \( H_e \) across populations ranged from 0.407 to 0.843, with an average of 0.665. Rarefaction curves for \( A_R \) across each locus levelled off for each sampling location indicating that a reasonable portion of the available allelic diversity was sampled at each location (S1 Fig; allele frequencies are reported in S5 Table). Populations did not exhibit significant differences in genetic diversity for either \( A_R (F_{d,146}=0.0048, P=1.000) \) or \( H_e (F_{d,146}=1.102, P=0.358) \). No significant linkage disequilibrium was found after Bonferroni correction (\( P<0.05 \)) (S6 Table). Populations met Hardy-Weinberg expectations (MICROCHECKER, Table 1) with few exceptions (S4 Table). There was a low level of null allele frequency (\( \leq 10\% \)) for all populations at all loci (FreeNA, S7 Table), and there were only slight changes in global and pairwise \( F_{ST} \) values between populations for each locus after accounting for null alleles (S8 Table). Original allele frequencies were therefore used in subsequent analyses.

Table 1. Summary of the genetic diversity statistics at microsatellite loci across five locations.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Na</th>
<th>Size (bp)</th>
<th>( H_0 )</th>
<th>( H_e )</th>
<th>( G_{IS} )</th>
<th>( F_{ST} )</th>
<th>( G''_{ST} )</th>
<th>Dest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pm01</td>
<td>23</td>
<td>108–208</td>
<td>0.842</td>
<td>0.843</td>
<td>0.000</td>
<td>0.057(^c)</td>
<td>0.328(^c)</td>
<td>0.291(^c)</td>
</tr>
<tr>
<td>Pm02</td>
<td>9</td>
<td>105–137</td>
<td>0.671</td>
<td>0.742</td>
<td>0.090(^a)</td>
<td>0.026</td>
<td>0.044</td>
<td>0.033</td>
</tr>
<tr>
<td>Pm07</td>
<td>10</td>
<td>128–164</td>
<td>0.737</td>
<td>0.720</td>
<td>-0.024</td>
<td>0.014</td>
<td>-0.007</td>
<td>-0.005</td>
</tr>
<tr>
<td>Pm08</td>
<td>6</td>
<td>141–165</td>
<td>0.710</td>
<td>0.660</td>
<td>-0.079</td>
<td>0.115(^c)</td>
<td>0.365(^c)</td>
<td>0.274(^c)</td>
</tr>
<tr>
<td>Pm09</td>
<td>16</td>
<td>91–142</td>
<td>0.733</td>
<td>0.736</td>
<td>0.004</td>
<td>0.045(^c)</td>
<td>0.142(^c)</td>
<td>0.108(^c)</td>
</tr>
<tr>
<td>Pm10</td>
<td>6</td>
<td>107–122</td>
<td>0.737</td>
<td>0.699</td>
<td>-0.055</td>
<td>0.071(^c)</td>
<td>0.233(^c)</td>
<td>0.175(^c)</td>
</tr>
<tr>
<td>Pm11</td>
<td>11</td>
<td>157–187</td>
<td>0.838</td>
<td>0.813</td>
<td>-0.032</td>
<td>0.035(^b)</td>
<td>0.137(^b)</td>
<td>0.114(^b)</td>
</tr>
<tr>
<td>Pm13</td>
<td>8</td>
<td>103–136</td>
<td>0.572</td>
<td>0.576</td>
<td>0.007</td>
<td>0.007</td>
<td>-0.024</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>Range</td>
<td>COI</td>
<td>CytB</td>
<td>Ave</td>
<td>SE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pm17</td>
<td>8</td>
<td>155–187</td>
<td>0.457</td>
<td>0.452</td>
<td>-0.011</td>
<td>0.058b</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.097b</td>
<td>0.046b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pm19</td>
<td>5</td>
<td>169–181</td>
<td>0.523</td>
<td>0.519</td>
<td>-0.008</td>
<td>0.041b</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.067b</td>
<td>0.036b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pm20</td>
<td>5</td>
<td>114–132</td>
<td>0.407</td>
<td>0.407</td>
<td>-0.001</td>
<td>0.181c</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.340c</td>
<td>0.174c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pm23</td>
<td>14</td>
<td>156–184</td>
<td>0.703</td>
<td>0.692</td>
<td>-0.016</td>
<td>0.132c</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.467c</td>
<td>0.378c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave</td>
<td>6.667</td>
<td></td>
<td>0.661</td>
<td>0.655</td>
<td>-0.009</td>
<td>0.064c</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.175c</td>
<td>0.122c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>0.428</td>
<td></td>
<td>0.020</td>
<td>0.019</td>
<td></td>
<td>0.046</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0352</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Significant deviation from HWE ($P<0.05$). Significant genetic differentiation: $^{b}$ ($P<0.01$), $^{c}$ ($P<0.001$).

### Mitochondrial DNA analyses

The basic diversity values for COI and CytB sequences are presented in Table 2. The total number of variable sites is 173 (COI: 105; CytB: 68), 98 of which are singleton mutations (COI: 59; CytB: 39); together these defined 130 distinct haplotypes for the concatenated sequences (COI: 103, CytB: 74). In contrast to the range of values obtained for $h$ (0.922-0.989), the corresponding ranges for $k$ (2.9-4.8) and $\pi$ (0.254-0.414%) were low for both genes. The mean values for $k$ and $\pi$ are 7.30 (COI: 3.81±0.018, CytB: 2.75±0.018) and 0.330 % (COI: 0.381%, CytB: 0.275%), respectively for the concatenated sequences. Similar diversity was observed between sampling locations.
Table 2. Summary of the genetic diversity statistics and bottleneck analysis for individuals sampled from five locations.

<table>
<thead>
<tr>
<th>Pop</th>
<th>Microsatellite</th>
<th>COI</th>
<th>CytB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Na</td>
<td>$A_R$</td>
</tr>
<tr>
<td>Ti Point</td>
<td>20</td>
<td>6.08</td>
<td>6.02</td>
</tr>
<tr>
<td>Auckland</td>
<td>30</td>
<td>6.67</td>
<td>5.88</td>
</tr>
<tr>
<td>Tauranga</td>
<td>33</td>
<td>6.88</td>
<td>5.89</td>
</tr>
<tr>
<td>Wellington</td>
<td>18</td>
<td>6.00</td>
<td>6.00</td>
</tr>
<tr>
<td>Nelson</td>
<td>45</td>
<td>7.75</td>
<td>5.93</td>
</tr>
<tr>
<td>Total</td>
<td>146</td>
<td>121</td>
<td>-</td>
</tr>
</tbody>
</table>

$N$: sample size, $Na$: number of alleles, $A_R$: mean allelic richness based on 18 diploid individuals, $Ho$: observed heterozygosity, $He$: unbiased expected heterozygosity, $G_{IS}$: inbreeding coefficient. $S$: number of segregating sites, $S_{in}$: singleton mutations, $H_{ap}$: number of haplotypes detected, $h$: haplotype diversity, $k$: number of pairwise nucleotide differences, $\pi$: nucleotide diversity, SD: standard deviation.
Genetic Structure

Microsatellite analyses

Global genetic differentiation, estimated using $F_{ST}$, $G''_{ST}$ and $Dest$, was low to moderate: 0.064, 0.175 and 0.122, respectively, and highly significant ($P \leq 0.0001$). Differentiation was significant for most loci (Table 2). Location had a significant effect on population structure (PERMANOVA, $DM$: $F_{4,146} = 7.3914$, $DCL$: $F_{4,146} = 9.8256$, $P = 0.0001$). Pairwise comparisons of genic ($\chi^2 = \infty$, d.f. = 24, $P < 0.0001$) and genetic differentiation ($F_{ST}$: 0.074-0.122, $G''_{ST}$: 0.216-0.337, $Dest$: 0.153-0.246) as well as PERMANOVA tests ($t = 2.54-4.69$) showed that southern populations (WL and NL) were significantly differentiated from northern populations (TP, AKL and TR) ($P < 0.0001$, Table 3). Most loci supported this pairwise differentiation pattern among populations ($F_{ST}$ and $Dest$ values in S8 and S9 Tables, respectively). None of the estimators suggested significant differentiation among the northern populations of TP, AKL and TR. $F_{ST}$ and PERMANOVA comparisons suggest weak but significant differentiation between the WL and NL populations ($P < 0.046$). Analysis of statistical power by POWSIM showed a 100% probability of detecting population differentiation at an $F_{ST}$ value of 0.01. The probability of $\alpha$ error was $\sim 0.05$ ($P = 0.04$ and 0.057 based on chi-square and Fisher methods, respectively), suggesting a low risk for Type I error. The differentiation pattern was therefore considered real and reinforced by $F_{ST}$ values between the
significantly differentiated populations being for the most part greater than 0.01. Additionally, PERMDISP analysis showed no significant differences in dispersion for either $DM$ ($F_{4,141}=0.1243, P=0.1243$) or $DCL$ ($F_{4,141}=0.4856, P=0.4856$), meaning that average nucleotide distances from individuals to their own population centroid did not differ among the groups (i.e. that within-group genetic variability did not differ among populations).

Table 3. Pairwise population differentiation estimates and associated tests across five populations.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Microsatellite data</th>
<th>COI data</th>
<th>CytB data*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Genic $F_{ST}$ $G^{\prime\prime}\text{ST}$ Dest $t$-DM $t$-DCL</td>
<td>$t$ $F_{ST}$ $\theta_{ST}$</td>
<td>$F_{ST}$ $\theta_{ST}$</td>
</tr>
<tr>
<td>TP-AKL</td>
<td>17.26 -0.009 -0.027 -0.019 0.699 0.766</td>
<td>1.585$^a$ 0.014$^b$ 0.024</td>
<td>0.505$^c$ 0.028</td>
</tr>
<tr>
<td>TP-TR</td>
<td>29.67 -0.004 -0.012 -0.008 0.727 0.79</td>
<td>1.359 0.017$^b$ 0.018</td>
<td>0.043$^b$ 0.023</td>
</tr>
<tr>
<td>TP-WL</td>
<td>Inf -0.087 0.256$^c$ 0.185$^c$ 2.851$^c$ 2.711$^c$</td>
<td>1.233 0.014$^c$ 0.008</td>
<td>0.033$^c$ -0.026</td>
</tr>
<tr>
<td>TP-NL</td>
<td>Inf 0.118$^c$ 0.330$^c$ 0.241$^c$ 4.160$^c$ 4.083$^c$</td>
<td>1.494 0.030$^b$ 0.029</td>
<td>0.040$^b$ 0.000</td>
</tr>
<tr>
<td>AKL-TR</td>
<td>20.48 -0.004 -0.011 -0.007 0.983 0.655</td>
<td>1.712$^c$ 0.015$^c$ 0.020</td>
<td>0.071$^c$ 0.019</td>
</tr>
<tr>
<td>AKL-WL</td>
<td>Inf 0.095$^c$ 0.274$^c$ 0.198$^c$ 2.548$^c$ 2.958$^c$</td>
<td>1.202 0.012$^a$ 0.014</td>
<td>0.062$^c$ -0.003</td>
</tr>
<tr>
<td>AKL-NL</td>
<td>Inf 0.122$^c$ 0.337$^c$ 0.246$^c$ 3.677$^c$ 4.693$^c$</td>
<td>1.283 0.028$^c$ 0.005</td>
<td>0.068$^c$ -0.001</td>
</tr>
<tr>
<td>TR-WL</td>
<td>Inf 0.074$^c$ 0.216$^c$ 0.153$^c$ 2.541$^c$ 3.035$^c$</td>
<td>1.403 0.016$^a$ 0.017</td>
<td>0.056$^c$ 0.012</td>
</tr>
<tr>
<td>TR-NL</td>
<td>Inf 0.097$^c$ 0.272$^c$ 0.194$^c$ 3.725$^c$ 4.661$^c$</td>
<td>1.076 0.031$^a$ 0.007</td>
<td>0.061$^c$ 0.026$^a$</td>
</tr>
<tr>
<td>WL-NL</td>
<td>53.20$^b$ 0.008$^b$ 0.021 0.013 1.235$^c$ 1.593$^c$</td>
<td>1.326 0.029$^b$ 0.020</td>
<td>0.053$^a$ -0.014</td>
</tr>
</tbody>
</table>

$\chi^2$ values for the homogeneity of allele frequencies in pairwise comparisons tested with the exact G-test (Genic). Pseudo t statistic ($t$) from pairwise PERMANOVA tests on the basis of either Manhattan ($DM$) or Clonal distances ($DCL$) for microsatellite data, or standardized nucleotide differences for mtDNA data. Other estimators used were fixation index $F_{ST}$, standardised fixation index $G^{\prime\prime}\text{ST}$, Jost’s differentiation ($Dest$) and distance-based $\theta_{ST}$.


Significant differentiation: $^a P<0.05$, $^b P<0.0005$, $^c P<0.0001$. *Pairwise comparisons using pseudo t were not done for CytB data because the overall test was not statistically significant.
A north-south differentiation is evident from MDS ordinations of the \textit{DM} and \textit{DCL} matrices (Figs 2A and B). Note that although stress is relatively high (0.23 and 0.24) for the 2-dimensional MDS ordinations, the north-south distinction was also very clearly apparent in the corresponding 3-dimensional MDS ordinations (not shown here), which had lower stress (0.18 in both cases). Discriminant analysis (CAP, S2 Fig) for the two-group north-south split was highly significant ($P<0.001$), with the CAP model showing a leave-one-out allocation success of 97.26\% (142/146 for \textit{DCL} and also for \textit{DM}). In contrast, the five-group CAP models (TP, AKL, TR, WL and NL) were poor, having overall allocation success of only 43.15\% (63/146 for either \textit{DCL} or \textit{DM}). Furthermore, there was no discrimination among the three northern locations (TP, AKL and TR) for either \textit{DCL} or \textit{DM} (CAP, $P>0.76$ in both cases), justifying their pooling into a single group. Similarly, there was no basis for discriminating the two southern populations (WL and NL) using CAP models ($P > 0.11$ for \textit{DCL} and \textit{DM}).

Bayesian clustering of individuals based on allele frequencies as implemented by STRUCTURE showed a $\Delta K$ value and mean log probabilities of data ($\text{LnP (x/K)}$) that were maximal at $K=2$ (Fig 2C), again supporting the same two distinct north-south clusters (Figs 2D and E). This finding was not affected by inclusion of sampling locations as priors (data not shown).
Fig 2. Visualization of genetic structure among *Pleurobranchaea maculata* populations based on microsatellite data.

All individuals from eight sampling locations were included. MDS ordination of pairwise (A) Manhattan (DM) and (B) Clonal (DCL) distances between individuals. Bayesian clustering analysis where the sampling location was not introduced for the calculations, (C) Plot of $\Delta K$ versus $K$ indicating that data are best explained by $K = 2$ clusters, (D) Population structure at $K=2$. Each individual is represented by a vertical line divided into two segments, which indicates proportional membership in the two clusters; (E) Group assignments, indicating proportional membership in $K=2$ clusters.

AMOVA results are shown in S10 Table. Analysis performed assuming no structure revealed that the highest proportion of the total variation stemmed from variation among individuals (88.3%) with a fixation index ($R_{IT}$) value of 0.117. Variation between populations ($R_{ST}$=0.137) constituted 8.02% of the total variation with high significance ($P=0.0000$). Variation between individuals within populations did not contribute to genetic variation (-1.16%) ($R_{IS}$=-0.020, $P=0.700$). AMOVA analysis was repeated by clustering the populations into northern and southern clusters. The results show that the variation between the northern and southern clusters explains 20.0% ($R_{CT}$=0.201) of the total variation with high significance ($P<0.0001$). Variation between populations within clusters explained only 0.1% of the...
total variation ($R_{SC}=0.001, P=0.3418$). AMOVA analysis based of $F_{ST}$ yielded similar results (data not shown). These results show that total variation for the NZ *P. maculata* populations is best explained by differences in the distribution of variance between northern and southern clusters.

**Mitochondrial DNA analyses**

The sequence data for 156 individuals obtained from *COI* and *CytB* were submitted to GenBank (accession numbers: KY094153 - KY094309 for *COI* and KY094310 - KY094466 for *CytB*). MJN analysis of *CytB* (Fig 3) and *COI* (S3 Fig) genes resulted in highly similar patterns. Thus, for simplicity, the *CytB* network was used to infer evolutionary relationships among individuals. The network shows a lack of noticeable geographical structure; common haplotypes are shared across populations. The two most common *CytB* haplotypes are shared by 25 (16.0% of the total dataset) and 22 (14.1% of the total dataset) individuals (frequencies in S1 Table). The network is complex and reticulated with a star-like topology; many private haplotypes descend from central shared nodes with mostly one to two base pair distances [74]. Private haplotypes were detected in all sampling locations. Characteristics of the network showed little change on inclusion of four recently published Argentinian samples (Farias *et al.*, 2015; Farias *et al.*, 2016) (based on a slightly shorter *COI* fragment (see Materials and Methods and S3B Fig). Argentinian samples are closely related to NZ
samples with just a single base pair distance from a commonly shared haplotype. The index of substitution saturation (ISS) was used to test for homoplasy due to multiple substitutions [50]. For both symmetrical and asymmetrical tree topology models and for both genes, the observed ISS values are significantly larger than the critical ISS (ISS.c) values (S12 Table), which indicate that the paired partitions are not saturated, and the degree of homoplasy is low.

**Fig 3. Median joining network of the CytB haplotypes.**

The network is coloured according to the sampling locations. The diameter of the circles reflects relative haplotype frequencies. The hashes indicate the mutational steps between the haplotypes. The black nodes represent the imaginary haplotypes necessary to create a bridge between the present haplotypes.

Analysis based on $F_{ST}$ values showed a moderate but highly significant global genetic differentiation for both COI ($F_{ST}$=0.022, $P$=0.0001) and CytB ($F_{ST}$=0.057, $P$=0.0001). However $\Phi_{ST}$ values were significant for COI ($\Phi_{ST}$ = 0.01552, $P$ = 0.022), but not for CytB ($\Phi_{ST}$=0.00845, $P$=0.12). Pairwise $F_{ST}$ values were low (0.033 – 0.074), yet significant for all the comparisons ($P$< 0.03) (Table 3). Differences in the distribution of allele frequencies were observable for CytB (Fig 4A, COI: S4 Fig) and pairwise $\Theta_{ST}$ values were low for all the comparisons for both COI and CytB (0.007-0.029); a
weak significant differentiation was observed only between TR and NL for
CytB ($\Theta_{ST}=0.026, P=0.0226$, Table 3). MDS ordination of the $D_{SEQ}$ matrices
revealed no observable structure for either gene (Fig 4B). However,
PERMANOVA analysis suggests that sampling location has a significant
effect on the population structure for COI ($F_{4, 145}=1.8791, P=0.0173$), but
not for CytB ($F_{4, 145}=1.3386, P=0.1806$), which was consistent with the
results of $\Phi_{ST}$ analysis. Pairwise PERMANOVA tests showed significant
but weak differentiation between the AKL and TP ($P=0.0276$), and AKL
and TR populations ($P=0.0218$) (Table 3). PERMDISP revealed no
significant differences in dispersion among populations for either COI
($F_{4,141}=0.3793, P=0.847$) or CytB ($F_{4, 141}=0.358, P=0.833$). CAP analysis
was performed for COI but not for CytB because PERMANOVA did not
reveal significant differences between groups for the latter. A CAP analysis
to discriminate the three northern populations was statistically significant
(trace=0.293, $P=0.0038$, S5 Fig), but the populations varied in their degree
of distinctiveness under the CAP model. Specifically, leave-one-out cross-
validation had the highest allocation success for TR (69.7%), followed by
AKL (50%), but there was little or no identifiability of samples from Ti
Point (only 25% correct allocation). Overall, the CAP results were
consistent with the structures revealed by PERMANOVA: among the five
populations, AKL is weakly distant from TP and TR.
Fig 4. Graphical visualization of the results of population structure and demographic analysis for mtDNA data.

(A) The frequencies of CytB haplotypes at each location (N=156). The pie segment represents the relative haplotype frequencies. Each colour corresponds to a different haplotype. The patterned segment represents private haplotypes. The sizes of the circles are proportional to the sample size. (B) Non-metric MDS ordination of distances obtained from the standard nucleotide differences between individuals for the COI data. Each symbol also represents a different sampling location. (C) Mismatch distributions of pairwise base pair differences between the concatenated COI and CytB haplotypes.

PERMANOVA analysis was repeated by contrasting the TP, AKL and TR populations (the northern population cluster) against WL and NL (the southern population cluster) to test whether the north-south differentiation identified with microsatellite data could be observed with the mtDNA data. The results indicate no significant differentiation between the clusters (COI: $F_{1,145}=0.819$, $P=0.5268$; CytB: $F_{1,145}=1.2325$, $P=0.2982$) and do not support the north-south differentiation identified by microsatellite data.

Hierarchical AMOVA analysis of concatenated sequences performed in the absence of population structure showed that differentiation among sampling sites explains only 0.69% and 1.26% of the total variation based on $\Theta_{ST}$ and
\( F_{ST} \) statistics, respectively, with high significance (\( \Theta_{ST} = 0.01260, P = 0.03082; F_{ST} = 0.00686, P = 0.0001 \)), whereas variation within populations explained 98.74\% (\( \Theta_{ST} \)) or 99.31\% (\( F_{ST} \)) of the total variation (S13 Table).

Grouping of populations into northern and southern clusters explained a maximum 0.16\% of the total variation, but it was not significant (\( \Theta_{CT} = -0.0062, P = 0.8994; F_{CT} = 0.00161, P = 0.10029 \)). These analyses show that mtDNA data do not support the north-south differentiation evident in microsatellite data.

A phylogenetic analysis of \( P. \text{maculata} \) shows that \( P. \text{maculata} \) samples from NZ form a single clade (S6 Fig). Inclusion of the Argentinian sequences did not change the topology of the tree. Phylogenetic relationships between \( P. \text{maculata} \) individuals were unsupported (bootstrap values <50\%), whereas there was good bootstrap support (between 74-100\%) for five other species in the Pleurobranchidae family.

**Migration and demographic changes**

**Microsatellite analyses**

Migration analysis with GeneClass2 detected four first generation migrants (\( P = 0.01 \)): one individual sampled from TP is a migrant from AKL, one individual sampled from TR is from AKL, and two individuals sampled from NL are migrants from WL (S14 Table). Individuals from each cluster were more likely to belong to populations from the same cluster. A lack of
first generation migrants between the clusters shows that these clusters are
genetically not well connected. However, CAP and STRUCTURE show
low connectivity between the clusters and admixed individuals in both
northern and southern clusters (Figs 2D and E). The highest
misclassifications between the clusters detected by CAP analysis and
highest admixture proportions detected by STRUCTURE were noted in TR
and WL. This suggests that the TR and WL populations are bridges between
the clusters. Admixture in the WL population may also explain weak
differentiation between the WL and NL populations that was found with $F_{ST}$
and PERMANOVA tests.

The Wilcoxon test did not detect recent bottlenecks in any population under
either TPM or SMM models (Table 4). In addition, analysis of mode-shift in
the distribution of allele frequencies suggests that all the populations exhibit
a normal L-shaped pattern indicating no mode-shift in the frequency
distribution of alleles. Taken together these data suggest that none of the
populations experienced a recent or sudden bottleneck.

Table 4. Results of the neutrality and demographic tests using either
microsatellite or mtDNA data.

<table>
<thead>
<tr>
<th>Population</th>
<th>Bottleneck</th>
<th>Microsatellite</th>
<th>mtDNA – concatenated sequences</th>
<th>Mismatch Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tajima’s $D$</td>
<td>Fu’s $F_s$</td>
<td>$\tau$</td>
</tr>
<tr>
<td>Ti Point</td>
<td>TPM</td>
<td>0.741</td>
<td>0.945</td>
<td>-1.45</td>
</tr>
<tr>
<td>Auckland</td>
<td>SMM</td>
<td>0.715</td>
<td>0.993</td>
<td>-2.14$^b$</td>
</tr>
<tr>
<td>Tauranga</td>
<td>SMM</td>
<td>0.689</td>
<td>0.995</td>
<td>-2.10$^b$</td>
</tr>
<tr>
<td>Wellington</td>
<td>TPM</td>
<td>0.633</td>
<td>0.052</td>
<td>-1.62$^a$</td>
</tr>
<tr>
<td>Nelson</td>
<td>0.954</td>
<td>0.999</td>
<td>-2.20&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-25.08&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Total</td>
<td>-</td>
<td>-</td>
<td>-2.45&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-24.74&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Total values for Tajima’s $D$, Fu’s $F_s$, $\tau$ and the parameters of the mismatch distribution analysis were calculated pooling all available individuals into a single pool.

TPM: Two-phase mutational model, SMM: stepwise mutational model,
$SSD$: sum of squared deviations and statistical significance, $P_{SSD}$: for the validity of the sudden expansion model, $\tau$: time passed since population expansion, $r$: Harpending’s raggedness index and its probability respectively, for the null hypothesis test of goodness-of-fit. $P$-values: <sup>a</sup> <0.05, <sup>b</sup> <0.01, <sup>c</sup> <0.001

### Mitochondrial DNA analyses

Overall, neutrality tests of Tajima’s $D$ and Fu’s $F_s$ revealed significant negative values in pooled samples (Table 4) suggesting a recent population expansion or purifying selection. This was also suggested by the uni-modal mismatch distributions of pairwise base pair differences for $COI$ and $CytB$ haplotypes (Fig 4C and S7 Fig, respectively), non-significant $SSD$ and $raggedness$ patterns (Table 4). Furthermore, the McDonald-Kreitman test found no evidence of positive selection: the ratio of nonsynonymous to synonymous substitutions within $P. maculata$ ($Pn/Ps$=5/136) and between species ($Dn/Ds$=5/100) was statistically similar (neutrality index=0.400, $P$=0.1104).

The demographic parameter tau ($\tau$), which represents the mutational time since expansion, was calculated after pooling all five populations (pooling of populations was justified because all exhibited signs of population expansion): $\tau$ was estimated at 2.703 (CI 95%: 1.488-3.320) and 2.754 (CI 95%: 1.344-3.615) for the $COI$ and $CytB$ data, respectively. The date of
expansion was estimated at 44.2 kya (CI 95%: 24.3-54.3 kya) based on COI data.

**DISCUSSION**

This study marks the first attempt to describe and account for patterns of genetic diversity in *P. maculata*. Overall, we observed marked signals of population structure; however, the population structure suggested by microsatellite versus mtDNA data differed. The nuclear data exhibit patterns of diversity indicative of a north-south disjunction. The northern samples formed one group and southern (WL and NL) samples formed another (with few examples of migration). However this disjunction was not supported by $F_{ST}$ analysis of mtDNA data, which indicated divergence among all populations.

Discordance between results obtained from nuclear versus mitochondrial markers is not uncommon, with explanations ranging from variation in lineage sorting to differences in rates of nuclear versus mitochondrial evolution [reviewed in 75, 76]. Taken together, we interpret our data as indicative of a single founding population that subsequently became fragmented following geographical and oceanographical changes that led to the present north-south divide in NZ waters.
According to this scenario, *P. maculata* was inhabiting the NZ waters before the end of the last glacial maximum when sea levels were low and NZ was a single land mass [77, 78]. Shallow-water species that inhabit intertidal or shallow subtidal areas were primarily affected as decreased sea levels destroyed available habitat [79, 80]. At some later time, most likely following the last glacial maximum (~22,000 years ago), large benthic habitats became available [79], and this may have facilitated population expansion and fragmentation aided by warming temperatures and rising sea levels [81].

In support of this model is the haplotype network based on mtDNA data showing star-like structure and high haplotype diversity, indicative of a population expansion arising from a small initial population [74]. Additional support comes from the unimodal mismatch distribution pattern of pairwise base pair differences for *COI* and *CytB* haplotypes (Fig 4C and S6 Fig, respectively), non-significant SSD and raggedness patterns (Table 4), and significant negative values for Tajima’s *D* and Fu’s *Fs*. The approximate date of population expansion was estimated at 44.2 kya (CI: 24.3–54.3kya). This estimate assumes a 5.3% divergence rate for the *COI* gene, which while an estimate, nonetheless coincides with a periodic of cyclic climatic oscillation defined by the late Pleistocene era (~110–15 kya) [82].

Glaciation has been suggested as the possible cause of demographic changes...
in other NZ marine organisms, including triplefin species [20] and red alga *Bostrychia intricata* [83].

This scenario is also supported by the analysis of more rapidly evolving microsatellites, which are useful for uncovering recent barriers to gene flow [84-86], especially following periods of demographic expansion [84, 86]. As sea levels rose at the end of the last glacial cycle (~13,000 years ago; 78], geographical factors and associated oceanography established barriers to gene flow for marine organisms. In particular, confluence of the East Cape current with the Wairarapa Eddy off the east coast of the North Island (between 37–39°S) created an oceanographic barrier [87]. A barrier was also formed by waters separating North and South Islands (the Cook Strait) [87, 88].

A north-south genetic differentiation has been observed in other marine organisms from NZ (reviewed in Gardner et al, 2010; Ross et al, 2012). Confluence of the East Cape current with the Wairarapa Eddy is thought to be responsible for population differentiation in organisms such as the amphipods *Paracorophium excavatum* and *P. lucasi* [87] and the gastropod *Diloma subrostrata* [89]. The Cook Strait barrier is thought to have underpinned north-south differentiation in organisms such as the green shell mussel (*Perna canaliculus*) [19], blackfoot paua (*Haliotis iris*) [21] and the ornate limpet (*Cellana ornate*) [90].
One additional factor that has likely promoted recent population subdivision in *P. maculata* is the distribution of invasive species [91] that constitute a food source for *P. maculata*. The Asian date mussel, *Arcuatula senhousia* has been established in the Auckland region since the 1970s, forming large transient beds in sub- and inter-tidal areas of the Hauraki Gulf, Manukau Harbour and Whangarei Harbour [92]. Expansion of *A. senhousia* beds in the Hauraki Gulf appears to have preceded increases in the density of *P. maculata* populations. Interestingly, subsequent decline of near-shore beds of *A. senhousia* post 2010 was followed by rapid decline in the density of *P. maculata* populations [93]. Further evidence that range expansion of *P. maculata* may have been facilitated by availability of prey species comes from off-shore mussel farms in Tasman Bay (Nelson, NZ), where culture of the green shell and blue mussels have created new habitats for *P. maculata*, with high-density populations being found beneath mussel farms [93]. Recently, *P. maculata* was identified in Argentinean waters with the species rapidly spreading along the Atlantic coast [5, 6]. The minor difference between mtDNA sequence in Argentinian versus NZ raises the possibility that NZ maybe the source of the recently discovered Argentinian population.

Life history traits such as the nature of egg and larval stages are of understandable importance in shaping population structure of the species. Species with benthic eggs tend to have more structured populations than
ones with pelagic eggs [94, 95] and an inverse relationship between pelagic larval duration (PLD) and genetic structure has been found [96, 97]: in a comparative analysis of NZ pelagic marine species, Ross et al. [96] showed a significant negative correlation between PLD and genetic differentiation. However, in the same meta-study, when NZ-wide sampling regimes were considered, NZ organisms with PLD durations similar to *P. maculata* (2–4 weeks) exhibit structural patterns of diverse types ranging from no structure, to a north-south disjunction, IBD and differentiation within and between sampling locations [96].

Our prediction that the previously recorded cline in TTX might be explained by genetic structure holds only for microsatellite markers. Had this also held for mtDNA markers a case may have been made that *P. maculata* is a complex of two cryptic species, but no such evidence exists. Our phylogenetic analysis indicates that all *P. maculata* populations – including samples from Argentinian waters [5] – are conspecific. Short branches with no or low bootstrap support is also indicative of lack of genetic differentiation among *P. maculata*. Similar lack of differentiation between toxic and non-toxic populations has been shown for *Taricha granulosa* newts from various localities in western North America [98, 99] and the red-spotted newt *Notophthalmus viridescens* [100].
Having called into question substantive genetic differences between north and south populations, differences in TTX levels are thus likely attributable to exogenous factors, such as differences in associated bacteria, exposure, or diet. Work to date is strongly suggestive of diet as the major source of TTX, with *P. maculata* accumulating TTX via feeding [10], while offspring from TTX positive individuals raised in a TTX-free environment become free of TTX [12]. Recent work studying cultured bacteria from *P. maculata* found no evidence for a bacterial origin of the toxin [15], but TTX has been reported in certain prey, including a Platyhelminthes *Stylochoplana* species, that co-occur with TTX-containing *P. maculata* [11].

**ACKNOWLEDGEMENTS**

We acknowledge Susanna Wood and David Taylor (Cawthron Institute); Lauren Salvitti, Dudley Bell and Warrick Powrie (The University of Waikato); Mike McMurtry (Auckland Regional Council); Severine Hannam and Wilma Blom (Auckland Museum); Shane Genage (Victoria University of Wellington); Steve Journee (The Dive Guys, Wellington); Richard Huges, Paul Caiger and Bakhti Patel (The Leigh Marine Laboratory at the University of Auckland); and Don Morrisey, Matthew Smiths and Stephen Brown (The National Institute of Water and Atmospheric Research). We thank Paul McNabb (Cawthron Institute) and Serena Khor (The University of Waikato) for conducting tetrodotoxin assays.
Supporting Information accompanies this paper on the website.

REFERENCES


47. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 1995;139(3):1463-.


59. Anderson MJ. Distance-based tests for homogeneity of multivariate

60. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary

GENECLASS2: a software for genetic assignment and first-

62. Rannala B, Mountain JL. Detecting immigration by using multilocus

63. Paetkau D, Slade R, Burden M, Estoup A. Genetic assignment
methods for the direct, real-time estimation of migration rate: a
simulation-based exploration of accuracy and power. Mol Ecol.

64. Piry S, Luikart G, Cornuet JM. BOTTLENECK: a computer
program for detecting recent reductions in the effective size using

65. Luikart G, Cornuet JM. Empirical evaluation of a test for identifying
recently bottlenecked populations from allele frequency data.

66. Tajima F. Statistical method for testing the neutral mutation


**Supporting Information Captions**

**Supporting Information Text: Supplementary Methodology and Results**

S1 Table. Sampling locations and abbreviations for the New Zealand *Pleurobranchaea maculata* populations.

S2 Table. PCR annealing temperature and elongation time for microsatellite markers.

S1 Table. Primers used to amplify the mtDNA genes.

S2 Table. Summary genetic diversity statistics for the 12 microsatellites and 5 populations of *Pleurobranchaea maculata*.

S3 Table. Allele frequencies at 12 microsatellite loci by population.

S4 Table. Results of the linkage equilibrium test for 12 microsatellite markers. The calculations were performed for each population and across the populations. *P* values were estimated by 6600 permutations. The *P*-value for the 5% nominal level is 0.000152 after Bonferroni correction.

S5 Table. Estimating the null allele frequency using the ENA algorithm.
S6 Table. The estimation of the possible effect of null alleles on population structure analysis. Estimated (A) global $F_{ST}$ of Weir (1996) for each microsatellite locus and (B) pairwise $F_{ST}$ for each pair of populations with and without using the ENA correction [28].

S9 Table. Pairwise population matrix of the $Dest$ values at each locus. $Dest$ values below the diagonal. The $P$ values calculated with 9999 permutations are shown above the diagonal. TP: Ti Point, AKL: Auckland, TR: Tauranga, WL: Wellington, NL: Nelson.

S10 Table. AMOVA results for the Pleurobranchaea maculata populations based on the microsatellite data. Analysis was performed with SMM based on $R_{ST}$-statistics. The hierarchical distribution of variation was determined between the sampling locations without introducing any structure, and nesting the Ti Point-Auckland and Tauranga populations in the northern cluster and the Wellington and Nelson populations in the southern cluster.

S3 Fig. Median joining network of the COI haplotypes. (A) 1153 bp of COI sequences from 156 New Zealand individuals. (B) 624 bp of COI sequences from 156 New Zealand and four Argentinian individuals. The network was coloured according to the sampling locations. The diameters of the circles are proportional to the frequency of the haplotypes. The hashes indicate the mutational steps between the haplotypes. The black nodes
represent the imaginary haplotypes necessary to create a bridge between the present haplotypes.

S11 Table. Haplotype frequencies for the mtDNA sequences by mere counting in populations. The IDs are individual names. ‘Population codes.

*SAMPLE size.

S12 Table. The results of the saturation test for COI and CytB sequences. The results suggest only little saturation both genes while assuming both symmetrical and asymmetrical topology as significant difference was observed between Iss and Iss.c values, and Iss < Iss.c.

S13 Table. AMOVA results of concatenated COI and CytB sequences based on F- and Θ-statistics.

S14 Table. Detection of first generation migrants. Likelihood ratio: L_home/L_max. The number of individuals with a probability below the threshold value is 4. The potential migrants are labelled in red (P < 0.01) and the most likely population in green. ID: ID of the individuals.

S1 Fig. Allele rarefaction curves for the microsatellite data.

S2 Fig. CAP analyses of the multivariate genetic distances between the *P. maculata* individuals obtained from 12 microsatellite markers. (A) DM (the number of axes - m=13) and (B) DCL (m=6) distances. The axes
represent the amount of variation between the populations that is explained by the two axes. Each symbol represents a different sampling location.

**S3 Fig. Median joining network of the COI haplotypes.** (A) 1153 bp of COI sequences from 156 New Zealand individuals. (B) 624 bp of COI sequences from 156 New Zealand and four Argentinian individuals. The network was coloured according to the sampling locations. The diameters of the circles are proportional to the frequency of the haplotypes. The hashes indicate the mutational steps between the haplotypes. The black nodes represent the imaginary haplotypes necessary to create a bridge between the present haplotypes.

**S4 Fig. The graphical representation of haplotypes frequencies for COI data at each sampling location.** The pie segment represents the relative frequencies of the haplotypes. Each colour corresponds to a different haplotype. Private haplotypes specific to each location are represented with a patterned segment.

**S5 Fig. Graphical visualization of the standard nucleotide differences between individuals based on the COI data.** Canonical correlation ordination analysis at \( m=4 \) was performed.

**S6 Fig. Molecular phylogenetic analysis of COI sequences by maximum likelihood method.** The tree with the highest log likelihood (-2685.86 is displayed. Initial tree(s) for the heuristic search were obtained automatically.
by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise
distances estimated using the Maximum Composite Likelihood (MCL)
approach, and then selecting the topology with superior log likelihood
value. The analysis involved 49 nucleotide sequences out of a total of 616.
The nodes are coloured based on the sampling locations. The numbers on
the nodes represent the % bootstrap support (values <50% not presented).

S7 Fig. Mismatch distribution of pairwise base pair differences between
the concatenated COI and CytB haplotypes.