Posted on Authorea 16 Jul 2020 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.158456440.01903363/v2 | This a preprint and has not been peer reviewed. Data may be preliminary.

A layover in Europe: reconstructing the invasion route of asexual lineages of a New Zealand snail to North America

Carina Donne[1], Maurine Neiman[1], James Woodell[1], Martin Haase[2], and Gerlien Verhaegen[3]

[1]The University of Iowa
[2]University of Greifswald
[3]Japan Agency for Marine-Earth Science and Technology

July 16, 2020

Abstract

Non-native species are threatening ecosystems and biodiversity worldwide. High genetic variation is thought to be a critical factor for invasion success. The global invasion of a few clonal lineages of the gastropod *Potamopyrgus antipodarum* is thus both puzzling and has the potential to help illuminate why some invasions succeed while others fail. Here, we used SNP markers and a geographically broad sampling scheme (N = 1617) including native New Zealand populations and invasive North American and European populations to provide the first widescale population genetic assessment of the relationships between and amongst native and invasive *P. antipodarum*. We used a combination of traditional and Bayesian molecular analyses to demonstrate that New Zealand populations harbor very high diversity relative to the invasive populations and are the source of the two main European genetic lineages. One of these two European lineages was in turn the source of at least one of the two main North American genetic clusters of invasive *P. antipodarum*, located in Lake Ontario. The other widespread North American group had a more complex origin that included the other European lineage and two New Zealand clusters. Altogether, our analyses suggest that just a small handful of clonal lineages of *P. antipodarum* were responsible for invasion across continents. Our findings provide critical information for prevention of additional invasions and control of existing invasive populations and are of broader relevance towards understanding the establishment and evolution of asexual populations and the forces driving biological invasion.

Keywords

Approximate Bayesian Computation, SNP, *Potamopyrgus antipodarum* , invasive route, clonal, genetic diversity

Introduction

Biological invasions – non-native species that harm native ecosystems and biodiversity (reviewed in Mallez & McCartney, 2018) – are a worldwide problem. Humans have increased the rate of invasions by facilitating the transport of organisms outside of their native range (Nentwig, Bacher, Kumschick, Pyšek, & Vilà, 2018). Genetic variation within the founding and descendent populations is a critical determinant of whether a potential invasion is successful, with high variation more likely to be associated with success (Estoup & Guillemaud, 2010; Estoup et al., 2016). Invasive populations are predicted to lose most of their genetic variation via founder effects during the colonization process (Dlugosch & Parker, 2008), raising the question

Posted on Authorea 16 Jul 2020 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.158456440.01903363/v2 | This a preprint and has not been peer reviewed. Data may be preliminary.

of how these invasions succeed (reviewed in Estoup & Guillemaud, 2010; Dlugosch et al., 2015; Estoup et al., 2016). One potential solution to the barrier that low genetic variation should pose to successful invasion could be provided by multiple distinct introductions of colonists (Dlugosch & Parker, 2008; Kolbe et al., 2004). While multiple introductions are indeed associated with invasion success, there are also multiple examples of successful invasions in the face of low genetic diversity (reviewed in Estoup & Guillemaud, 2010; Dlugosch et al, 2015).

These seemingly conflicting results highlight the fact that the relationship between genetic diversity and invasion success is not straightforward (recently comprehensively reviewed in Dlugosch et al., 2015; Estoup et al., 2016), with mechanisms like admixture, genotype-by-environment interactions, or purging of harmful alleles coming into play during the invasion process. A recent demonstration of the complicated relationship between diversity and invasion success is provided by the growing body of evidence suggesting that worldwide invasions might often be catalyzed by successful invasive populations rather than directly via the native range ("bridgehead effect"; Lombaert et al., 2010; e.g., Reed et al., 2020, Wylie et al., 2020). Recent advancements in population genomics resources and tools now allow us to reconstruct invasion routes, enabling us to address key questions regarding the identity of source populations and how genetic variation is influenced by the colonization process (Estoup & Guillemaud, 2010; Dlugosch et al., 2015; Estoup et al. 2016). This information can then be applied in turn to characterize whether and how evolution plays a role in the driving successful invasions (Dlugosch & Parker, 2008; Keller & Taylor, 2008; Estoup et al., 2016).

The wide availability of genomic resources overcomes limitations posed by earlier marker types (e.g., microsatellites) with respect to, for example, reconstructing evolutionary relationships among populations (McCartney, Mallez, & Gohl, 2019). In parallel, continued development of analytical methods like Approximate Bayesian Computation (ABC) analysis provides inferences that are useful in untangling complex scenarios often seen in invasive history (Estoup & Guillemaud, 2010). This information offers new insights into the specific mechanisms by which species introductions occur, which can be leveraged towards establishing best practices to prevent and control invasions.

Potamopyrgus antipodarum is a fresh- and brakish water prosobranch snail that has become widely known as a model organism for the evolution of sex, characterized by frequent coexistence between obligately sexual and obligately asexual individuals in its native range in New Zealand (Lively, 1987). Known colloquially as the "New Zealand mud snail" (NZMS) in the invasive range, *P. antipodarum* has successfully invaded every continent except Antarctica and Africa (Alonso & Castro-Díez, 2012; Collado & Collado, 2014). These snails were first discovered in England in 1859 (Ponder, 1988). Over the 19th century, NZMS spread rapidly throughout western and central Europe (Hamada et al., 2013). The New Zealand mud snail is now, $>$ 160 years post-invasion, considered one of the worst alien species in Europe (Nentwig et al., 2018). The introduction of *P. antipodarum* to Australia is estimated to have occurred at a similar time frame as the European invasion, with the first recorded date of presence in Australia in 1872 (Ponder, 1988). *Potamopyrgus antipodarum* was first introduced to North America in 1987 in the western United States (Taylor, 1987) and colonized the Great Lakes in 1991 (Zaranko et al., 1997). In Japan, *P. antipodarum* was first reported in 1990. These Japanese populations seem to represent an invasion that is distinct from European and US invasions (Hamada et al., 2013). Most recently, *P. antipodarum* was reported in central Chile in 2014 (Collado & Collado, 2014), where it has been extending its range (Collado & Fuentealba, 2020).

Invasive *P. antipodarum* populations, in contrast to the native range, are exclusively asexual (Alonso & Castro-Díez, 2012). Despite harboring low genetic variation relative to populations in the native range, invasive *P. antipodarum* have become successful in aquatic ecosystems spanning a wide range of environmental conditions across the globe (Dybdahl & Drown, 2011; Alonso & Castro-Díez, 2012). The high population growth rate that can accompany asexual reproduction is thought to be one of the many reasons that *P. antipodarum* invasions have been successful (Alonso & Castro-Díez, 2008). Apparent adaptive plasticity of shell morphology and life history traits with respect to local environmental conditions also seems to play an important role in invasion success of *P. antipodarum* (Kistner & Dybdahl, 2014; Verhaegen, et al. 2018a, *in press*). Other factors such as ovoviviparity, tolerance for a wide range of chemical and physical conditi-

Posted on Authorea 16 Jul 2020 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.158456440.01903363/v2 | This a preprint and has not been peer reviewed. Data may be preliminary.

ons, and release from coevolving parasites might also contribute to the *P. antipodarum* invasion (Alonso & Castro-Diez, 2012; Verhaegen et al., 2018a).

Documented negative effects of these snails on invaded ecosystems are likely driven in large part by their high population growth rates, translating into extraordinarily high density [e.g., 20,000-50,000 individuals/m^2 (Hall et al., 2003, 2006; Alonso & Castro-Díez, 2012)] and with dramatic consequences for invaded ecosystems. Hall et al. (2003) provided a clear example of the potential impact of invasive *P. antipodarum*, demonstrating that the snails consumed 75% of gross primary productivity and dominated nutrient fluxes in a creek in Wyoming, USA. Hall et al. (2006) went on to show that *P. antipodarum* strongly distorted the distribution of secondary production across invertebrate taxa, sequestering much of the available carbon and altering ecosystem function. Alonso & Castro-Díez (2012) argued that *P. antipodarum*'s ability to radically alter ecosystem nutrient fluxes puts the threat it poses on par with such notorious invaders as the zebra mussel.

The genetic background of these invasive *P. antipodarum* populations has been characterized with mitochondrial DNA (mtDNA) markers, revealing that, as for many other invasive taxa (Estoup et al., 2016), the invasive populations harbor very little diversity relative to the native range (Dybdahl & Drown, 2011; Neiman & Lively, 2004; Verhaegen et al., 2018b). These mtDNA data also indicate that the European invasion is primarily composed of two distinct lineages with haplotypes identical to haplotypes found in New Zealand (Städler et al., 2005; Verhaegen et al., 2018a). Dybdahl and Drown (2011) demonstrated that the picture is broadly similar with respect to the North American invasion and emphasized that this lack of diversity is surprising in light of how rapidly the North American invasive populations have successfully established throughout a wide range of environments. In particular, Dybdahl and Drown (2011) showed that two lineages (US1, US2; defined via a combination of mitochondrial and nuclear markers) predominated in the western and eastern parts of the United States, respectively. The mtDNA haplotypes of lineages US1 and US2 are identical to haplotypes (22 and 37 in Neiman and Lively 2004, respectively) that are also found across Europe (Städler et al. 2005; Verhaegen et al. 2018a) and in New Zealand (Verhaegen et al. 2018b). The other two lineages defined by Dybdahl and Drown (2011), US1a and US3, have genotypes that have to date only been found in the western United States (Dybdahl & Drown, 2011). Dusting (2016) discovered that the dominant western North American lineage US1 and the geographically widespread and relatively common Australian invasive genotype AUS2 share a 22-locus SNP genotype, suggesting a close genetic link between the source populations of these invasions (also see Emblidge Fromme & Dybdahl, 2006). Hershler et al. (2010) used microsatellite genotyping to provide a finer-scale characterization of the genetic diversity of *P. antipodarum* in the western North American Snake River basin population, finding that this population appeared to have been founded by a few rapidly spreading clones.

Together, these studies demonstrate that *P. antipodarum* is a very successful invader even in the absence of genetic variation. What remains a key question is how the worldwide invasive populations are genetically connected. Reconstruction of the invasion route will illuminate whether invasive *P. antipodarum* populations represent a tiny and perhaps preadapted subset of native diversity that has experienced strong selection in the invasive range versus an invasion process driven by chance sampling events such as accidental boat transport. This information regarding the sampling process during invasion and the role of chance vs. deterministic processes in invasion is a central component of characterizing the traits and factors that drive invasion success (Keller & Taylor, 2008; Estoup et al. 2016). We used multilocus nuclear SNP markers to characterize the relationships amongst a large set of native New Zealand populations and invasive North American and European populations in the framework of several different population genetics modeling approaches to provide the first population genetic assessment of the relationships of invasive and native *P. antipodarum* across continents continents [for linguistic simplicity we also refer to New Zealand as continent, and in fact, New Zealand is part of Zealandia, a mostly drowned fragment of Gondwanaland (Mortimer et al. 2017)]. We specifically addressed the key question of whether the successful North American and European *P. antipodarum* invasions represented similar (or even the same) vs. distinct source populations. Discovering that these populations are similar is significant in suggesting that invasions in North America and Europe are closely linked. This result would also be consistent with a scenario where certain native lineages are more invasive than others and where invasion success does not require high genetic variation. Very distinct

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

European and American populations will in contrast suggest different source populations for the North American and European invasions, as expected in a case where genetic variation is an important component of invasion success. Regardless of specific outcome, the answer to the question of how closely these continent-scale invasions are related to one another will provide important new insights into how the invasion of this destructive snail occurred. A deeper understanding of whether and how successful invasive lineages are connected also helps illuminate the factors that contribute to invasion success, species range expansions, and the persistence of populations that are asexual or harbor low genetic variation.

Materials and Methods

Collection, DNA extraction, and genotyping

Freshwater mudsnail (Potamopyrgus antipodarum) estrogen receptor Identification and expression analysis under exposure to (xeno-)hormones

We included a total of 1617 snails from three continents: 1016 snails (51 sites) from the native range in New Zealand, 445 snails (25 sites) from the invaded range in Europe, and 156 snails from 13 invaded sites in North America (**Table 1**). All our sampling locations were freshwater sites with the exception of the European brackish water site DEBIN (Verhaegen et al., 2018a). The genetic data from the New Zealand snails and most of the European (N = 421, 94.6 %) snails were generated in Verhaegen et al. (2018a,b). Snails from North America and from two additional European populations (Mo and Gb) were newly collected in the summers of 2016 and 2017 and genotyped for this study. These snails were collected alive from submerged rocks and vegetation growing in shallow regions of streams and lakes with kick nets and then housed in spring water and fed freshwater fish food until their transport to the University of Iowa. The snails were then transferred to 10-L tanks filled with carbon-filtered tap water at constant 16°C room temperature and a 12 h day/night cycle. Water was changed 2x/week and dried spirulina and chalk were added as food and calcium sources, respectively.

We sacrificed the snails, used forceps to separate the soft tissue from the shell, and flash froze each snail in liquid nitrogen. We then stored the snails at -80°C until DNA extraction. DNA was extracted from the frozen tissue by first lysing the cells in CHAOS buffer [4 M guanidine thiocyanate, 0.1% N-lauroyl sarcosine, 10 mM Tris hydrochloride (Tris-HCl) at pH 8.0, and 0.1 M 2-Mercaptoethanol in 60 mL of double-distilled water (ddH_2O); Fukuma et al., 2004] for 24 h. The buffer was emulsified with a volume of phenol-chloroform extraction buffer [100 mM Tris-HCl at pH 8.0, 10 mM ethylenediaminetetraacetic acid (EDTA), and 0.1% sodium dodecyl sulfate in 80 mL of ddH_2O] equal to that of the tissue slurry, an equal volume of phenol chloroform isoamyl alcohol (PCIA, volumetric ratio of 25:24:1), and 5 µL of RNase A before vortexing. An equal volume of PCIA was added and centrifuged at 10,000 x g for 30 seconds. Another volume of chloroform was added and then vortexed to emulsify before centrifuging again. We pipetted off the aqueous layer, measured its final volume, and added a tenth of this volume of sodium acetate. We then added to this solution twice this final volume of ice-cold 100% ethanol and stored the samples at -20°C overnight. Next, the samples were centrifuged at 13,000 x g at 4°C for 15 minutes and the supernatant pipetted off. We added 1 mL of ice-cold 70% ethanol before centrifuging at 4°C for 10 more minutes. The supernatant was pipetted off and the samples left to dry overnight before being resuspended in 30 µL Tris low-EDTA buffer (10 mM Tris, 0.1 mM EDTA). The extracted DNA was stored on 96-well plates at -80°C until shipment on dry ice to LGC Genomics (Beverly, UK, www.lgcgroup.com/genomics) for KASP^{TM} assay-based genotyping of the same 48 single nucleotide-polymorphism (SNP) markers used in Verhaegen et al. (2018b). Sixteen of these markers were designed by Paczesniak et al. (2013) and 32 by Verhaegen et al. (2018a). The SNP data from the newly genotyped individuals were then merged with the data from Verhaegen et al. (2018ab), for a total N of 1617. Twelve loci were fixed across all individuals and hence discarded. The remaining polymorphic SNPs were the same 36 loci previously found to be variable and in Hardy-Weinberg (H-W) and linkage equilibrium among the New Zealand individuals (Verhaegen et al., 2018b). Forty-five individuals (2.78%)

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

for which genotyping failed for more than five loci were excluded. Genotypes were then assigned by the infinite alleles model distance index with GenoDive v. 2.0b27 (Meirmans & Van Tienderen, 2004) setting the maximum distance threshold for two genotypes to be considered identical to zero (Rogstad et al., 2002). For all subsequent population genetic analyses, we discarded all but one individual per genotype within each unique population or group in order to avoid treating likely clonemates as independent data points.

Genetic distances between sites

We calculated pairwise Slatkin's linearized fixation indices based on F_{ST} (Slatkin, 1995) as implemented in Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010) to evaluate the genetic relationships among sampling sites by building a neighbour-joining (NJ) tree in PAST v.3.25 (Hammer et al., 2001). Two individuals of the closely related *P. estuarinus*, for which the same SNP loci were previously genotyped by Verhaegen et al. (2018b), served as outgroup to root the NJ tree. We used GenGIS v.2.5.3 (Parks et al., 2013) to geographically map the NJ tree.

Clustering

We used clustering methods and discriminant analyses of principal components (DAPC) (Jombart et al., 2010) with the R package *adegenet* v.2.1.1 (Jombart, 2008) to evaluate genetic structure amongst native and invasive *P. antipodarum* populations. We first estimated the number of genetic clusters with a K-means clustering analysis based on the lowest Bayesian information criterion (BIC) (Schwarz, 1978). We then used a DAPC to visualize the relationships among these clusters and to assign the individuals to the clusters *a posteriori*. DAPC first transforms the data using a principal component analysis (PCA) and then identifies clusters through a discriminant analysis (DA). Unlike a STRUCTURE-like approach (Pritchard et al., 2000), a DAPC does not rely on population genetic assumptions and therefore can accommodate populations containing clonal individuals. The geographical distribution of these clusters was then mapped with QGIS v.2.18.24 using the vector maps publicly available from Natural Earth (https://www.naturalearthdata.com/). In order to quantify differences between clusters, we calculated the uncorrected number of mutational changes between pairs of genotypes from DAPC clusters including European and North American individuals with PAUP v.4.0b10 (Swofford, 1998).

ABC analyses

Approximate Bayesian Computation (ABC) analyses implemented in DIYABC v.2.1.0 (Cornuet et al., 2014) were used to retrace the invasion route of *P. antipodarum* to Europe and North America. While DIYABC assumes random mating within populations, we followed the lead of previous successful applications of DIYABC to other clonal (e.g., Goss et al., 2014; Zhang et al., 2014) or mixed sexual/asexual systems (e.g., Barres et al., 2012; Janoušek et al., 2016; Taerum et al., 2017). For the native range, we divided the New Zealand individuals into two groups, one per island (North vs. South Island, labelled **NZN** and **NZS**). For the invaded range, we took into account our clustering results and previously established mtDNA haplotypes (Verhaegen et al., 2018a) to divide the European snails into two groups (**EU14** vs. **EU15**). Based on the clustering results, the North American samples were divided into two groups (**US1** vs. **OnBrF**). The ABC groups are summarised in **Table 2**. Three of our 36 polymorphic SNP markers (SNP IDs: comp148591_c0_seq1, comp149304_c0_seq1, and comp160266_c0_seq4) were discarded from all our ABC analyses because they could not be genotyped for all individuals within at least one group. In order to optimize the computational effort, our ABC analyses were performed in a stepwise fashion as proposed by Lombaert et al. (2014) following the chronological order of the introductions into Europe and the US. In a first step, we clarified the most likely invasion scenario of *P. antipodarum* to Europe. In a second step, we tested all plausible invasion scenarios to the US. For this first step, we tested, in one ABC analysis (ABC1), 12 plausible invasion scenarios of *P. antipodarum* to Europe (**Supplemental Fig. 1**). This included six scenarios without unsampled "ghost" populations (i.e., unsampled populations from anywhere geographically between the sampled New Zealand

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

sites and the invaded regions), four scenarios with one ghost population originating from either **NZN** or **NZS**, and two scenarios with two ghost populations, each originating from a different New Zealand island. As *P. antipodarum* diversity is highest on the North Island (Neiman & Lively, 2004), and considering that *P. antipodarum*'s five most closely related congeners are confined to the North Island (Haase, 2008), we set **NZN** as the most ancestral group in all scenarios. The unit of time of the historical scenarios in DIYABC is set by the generation time of the study organism (Cornuet et al., 2014). Assuming that the generation time of *P. antipodarum* is 6-12 months (Larkin et al., 2016; Verhaegen et al., *in press*; Winterbourn, 1970) and that the introduction of *P. antipodarum* to Europe happened about 180 years ago (Ponder, 1988; Smith, 1889), we set the divergence time of our European groups to a range of 180 to 360 generations. This interval takes into account the uncertainty of the date of the actual introduction and the documented variation in generation time. Posterior probabilities for the scenarios were estimated with a logistic regression on the 1% of simulated datasets that were closest to the observed dataset (Cornuet et al., 2008, 2010). We allowed the software to perform a linear discriminant analysis on the summary statistics before the regression in order to reduce the computation time (Cornuet et al., 2014; Estoup et al., 2012). The most likely scenario was then selected based on these probabilities ensuring that the 95% confidence intervals of competing scenarios did not overlap (Cornuet et al., 2008).

In a second step, we used the most likely scenario of the first step as a base to identify the most likely invasion scenarios of *P. antipodarum* to North America. It was not computationally feasible to analyse all possible scenarios against each other at once, so we divided the 36 scenarios into four different ABC analyses. In a first analysis (ABC2), we tested 16 invasion scenarios without the presence of ghost populations (**Supplemental Fig. 2**). In a second analysis (ABC3), nine scenarios including a ghost population originating from **NZN** and two scenarios with two ghost populations originating from **NZN** and **NZS**, respectively, were tested against each other (**Supplemental Fig. 3**). In ABC4, we tested nine scenarios including a ghost population originating from **NZS** (**Supplemental Fig. 4**). In ABC5, we then tested the most likely scenario of each of the first three analyses (ABC2-4) against one other. In the last analysis, ABC6, based on the clustering results, we tested the most likely scenario of ABC5 against two scenarios with either **US1** originating both from **EU15** and **NZN**, or from **EU15** and **NZS**. As *P. antipodarum* is estimated to have been introduced to North America about more than 30 years before our sampling (Bowler, 1991; Taylor, 1987; Zaranko et al., 1997), we set the divergence time parameter for North American populations to a range from 35 to 70 generations, again considering the uncertainty of the actual arrival of *P. antipodarum* as well as the variation in generation time. As the introduction date of *P. antipodarum* to Australia is estimated to be around the same time as to Europe (Ponder, 1988), we set the divergence time for the ghost populations to be equal or younger than the divergence time of the European populations.

All of our ABC analyses except for ABC2 (671,093 simulations/scenario, over six times the minimum required simulation number) and ABC3 (944,373 simulations/scenario, over nine times the minimum required number of simulations) were performed with one million simulated datasets per scenario, the optimum number recommended for DIYABC. We used fewer than one million simulations for ABC2 and ABC3 because of computational constraints. The default minimum allele frequency criterion for SNP data was used, namely the Hudson's algorithm (Hudson, 2002). We calculated the following summary statistics: 1) per population, the proportion of loci with null gene diversity (i.e., the proportion of monomorphic loci), mean and variance of the gene diversity across polymorphic loci (Nei, 1987), and the mean gene diversity across all loci; 2) per population pair, the proportion of loci with null F_{ST} distance (Weir & Cockerham, 1984), the mean and variance across loci of non-null F_{ST} distances, and the mean across loci of F_{ST} distances. These summary statistics are used to discriminate among scenarios and to compute posterior distributions of parameters. A bottleneck event was included after all divergence events. The effective population size of all tested populations was set to be less than or equal to the size of **NZN** and the effective size of all bottleneck populations was set to be less than or equal to the populations that emerged from these bottlenecks. The divergence time of **NZS** from **NZN** was set to be greater than all the other divergence times present in our scenarios. The goodness of fit of the final chosen scenario was tested through a PCA comparing the prior and posterior distributions of of the summary statistics to the observed data, as implemented in DIYABC

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

("Perform model checking").

Results

Population genetic analyses and NJ tree

We detected a total of 338 SNP genotypes in our dataset; 309 of these genotypes were collected from New Zealand, 20 from Europe, and nine from North America. At 16 of the 89 sites all individuals shared the same genotype (three New Zealand sites, five European sites, and eight North American sites). For sites harbouring more than one genotype, the genetic diversity was low to moderate, with values for Ho varying between 0.016 and 0.486 (0.279 ± 0.038 – mean ± SD), and for Hs between 0.016 and 0.353 (0.213 ± 0.026) (**Table 1**). The NJ tree based on Slatkin's linearized fixation indices between pairs of sites (**Supplemental Table 1**) is given in **Fig. 1** and geographically mapped in **Fig. 2**. European and North American clades were separated and scattered across the entire tree. In only one case - the sample from Ontario- was the closest relative of one of these invasive genotypes not a genotype or clade from New Zealand. This Ontario sample emerged from a European clade.

Clustering

In order to retrace the invasion route of *P. antipodarum*, we first assessed the genetic structure by means of DAPC clustering. With an overall *a posteriori* assignment probability of 93.9% (individual *a posteriori* probabilities are found in **Supplemental Table 2**), individuals were grouped into 15 different clusters (**Fig. 3**). All 15 clusters included individuals from New Zealand (four clusters exclusively from the North Island, seven clusters from the South Island, and the remaining four clusters from both islands), 11 of these clusters contained only New Zealand individuals. Clusters *1* and *2* consisted of New Zealand individuals and five North American individuals from Sn and one from Ml, respectively. Individuals from all US sites except On, European sites DEBIN, Mo, NL1, NL2, and two New Zealand individuals from NZ52 and one from NZ64 were found in cluster *15*. Individuals from the North American site On, one individual from US site BrF, another individual from NZ64, and all European sites except DEBIN formed cluster *14*. The European individuals from cluster *15* were pooled together as group **EU15** for the ABC analyses and the European snails from cluster *14* as **EU14** (**Table 2**). The geographical distribution of these clusters is given in **Fig. 4** and bar plots with the number of individuals per continent within each cluster in **Supplemental Fig. 6**.

The variation within and between the four clusters including individuals from the invaded range was given by the uncorrected number of mutational changes between pairs of their genotypes. Overall, genotypes from these clusters differed from each other by 1 – 36 mutations (**Supplemental Table 3**). Genotypes within clusters *1*, *2*, *14*, and *15* differed from each other by 1 – 20, 1 – 9 , 1 – 23, and 1 – 14 mutations, respectively. All but one individual from the American site On and one individual from BrF shared genotype 324 with one another and were pooled together as the ABC group **OnBrF** (**Table 2**). Genotype 324 was also shared with individuals from **EU14** present at sites Gb and Mo. One remaining individual from site On harbored a genotype that differed only by one mutation from genotype 324 and was therefore included in **OnBrF**. There was relatively high genotypic variation present among American individuals of **US1** (**Table 1**), which harboured seven genotypes from cluster *15* differing from each other by one to four mutations, one genotype (337, at least six mutations from New Zealand genotypes) from cluster *1*, and one genotype (332, at least five mutations from New Zealand individuals) from cluster *2*. The most common and geographically widespread **US1** genotype (325) was also found in one individual from **EU15** (site Mo), and differed by at least eight mutations from European or New Zealand genotypes. The other **US1** genotypes from cluster *15* were at least six mutations away from the two other **EU15** genotypes.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

ABC analyses

All five initial ABC analyses (ABC1-5) had a most likely scenario clearly distinct from competing scenarios with a posterior probability > 97%. The most likely invasion route to Europe (ABC1, scenario 6, **Supplemental Fig. 1**) had **EU14** originating from **NZN** and **EU15** from **NZS** and did not include ghost populations [0.9773 (0.9724,0.9821) – posterior probability (95% confidence interval)]. For the three preliminary ABC analyses (**Supplemental Fig. 2 to 4**) testing the invasion route to North America, scenarios 6 [ABC2, 0.9985 (0.9975, 0.9994)], 11 [ABC3, 0.9996 (0.9994, 0.9998)], and 9 [ABC4, 0.9996 (0.9995, 0.9998)] emerged as the most likely scenarios for their respective analysis. These three most likely scenarios were tested against each other in ABC5 as scenarios 1, 2, and 3, respectively (**Supplemental Fig. 5**). Here, scenario 1 was the most likely invasion scenario [0.9996 (0.9995, 0.9998)], in which **OnBrF** originated from **EU14**, **US1** from **EU15**, and the assumption of ghost populations was not necessary. The probability of scenario 1 was nevertheless low [0.1625 (0.1536,0.1715)] when tested against the two admixture scenarios in the final ABC6, with the most likely scenario as **US1** originating from **EU15** and **NZN** [scenario 2, 0.6193 (0.6069,0.6317)], and the second most likely scenario as **US1** originating from **EU15** and **NZS** [scenario 3, 0.2182 (0.2080,0.2283)] (**Fig. 5**). The final invasion scenario retained by our ABC analyses was summarized and mapped in **Fig. 6**. Posterior probabilities and confidence intervals for all tested scenarios are given in **Supplemental Table 4**. Despite the unambiguity of the posterior probabilities for the scenario choice, the final scenario 2 of ABC6 showed a poor goodness of fit (**Fig. 7**). The posterior distribution of the summary statistics formed a distinct cloud that did not contain the observed data in the PCA plot checking the model. This discrepancy may be due to the limitations that ABC methods face with respect to asexually reproducing organisms.

Discussion

We used genotyping of nuclear SNPs from a large and geographically diverse sample of *P. antipodarum* from invaded sites in North America and Europe and the native range in New Zealand to characterize its invasion route across continents. In particular, the combination of traditional and Bayesian molecular analyses allowed us to resolve outstanding key questions regarding potential direct connections between geographically distinct invasions. Together, these data complement and extend earlier genetic and genomic work in *P. antipodarum* (Städler et al., 2005; Dybdahl & Drown, 2011; Verhaegen et al., 2018a) by demonstrating conclusively that major components of the worldwide NZMS invasion were driven by just a few asexual lineages. We also showed for the first time that the European invasion was seeded by introductions from several different New Zealand populations and that this European invasion in turn is a major source of the invasion of North America. Together, these data provide a qualitative advance forward in suggesting that a bridgehead scenario (Lombaert et al., 2010) is an apt framework for this destructive global invasion. More broadly, these results are of relevance to understanding other invasions of asexual organisms and/or invasions with relatively low genetic variation, and provides critical information on genetic diversity and invasion sources, central in any invasion control strategy.

From where did the European snails originate?

European snails were divided in two distinct genetic groups **EU14** and **EU15**, based on the clustering results of our DAPC (clusters *14* and *15*, respectively). Individuals of **EU14** and the Great Lakes shared the mtDNA haplotype 22 (Dybdahl & Drown, 2011; Neiman & Lively, 2004; Verhaegen et al., 2018a). This haplotype was, however, absent from our New Zealand samples (Verhaegen et al., 2018b), but reported in foregoing studies from both islands [e.g., in Lakes Waikaremoana, Gunn, and Te Anau (Neiman & Lively, 2004; Neiman et al., 2011)]. Using SNP markers, the ABC enabled us to retrace the origin of **EU14** to the North Island. While the only New Zealand-sourced individual from cluster *14* was collected on the South Island (site NZ64), our NJ tree showed that clades from the North Island harboured the closest relatives to the European sites of cluster *14,* indicating the high dispersive potential of *P. antipodarum*. Genotyping New

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Zealand individuals harbouring haplotype 22 from both New Zealand islands will provide an important next step, allowing us to test if the North Island was indeed the origin of **EU14.** As all ABC scenarios testing the origin of **EU14** from unsampled ghost populations were rejected and because no Australian *P. antipodarum* of haplotype 22 has been found (Dusting, 2016), we can state with some confidence that individuals of **EU14** were directly introduced from New Zealand.

By contrast, the ABC analysis indicated that **EU15** originated from New Zealand's South Island. For this analysis, New Zealand individuals of mtDNA haplotype 37 from both the North and South Island were present within our tested dataset (Verhaegen et al., 2018b). A few individuals from DAPC cluster *2* were also found in New Zealand, but only on the South Island (sites NZ52 and NZ64). Invasive scenarios including ghost populations as the origin of **EU15** were rejected by the ABC analysis. Because haplotype 37 is also found in Australia (Dusting, 2016), a useful validation step would come from the inclusion of Australian individuals in future analyses. However, because the introduction of *P. antipodarum* to Australia was reported several decades after the European introduction (Brazier, 1871; Hubendick, 1950; Ponder, 1988), we do not expect Australia to emerge as a stepping stone between New Zealand and Europe in subsequent studies.

What is the source(s) of the North American invasion?

First, the DAPC analyses revealed that the North American snails were represented predominantly by the same two genetic clusters present in Europe (clusters *14* and *15*). This result suggests that *P. antipodarum* of at least two different geographic and/or genetically distinct sources successfully colonized North America. Lake Ontario (ABC group **OnBrF**, all from cluster *14*) *P. antipodarum* were genetically distinct from the other North American sites (ABC group **US1**, predominantly from cluster *15*). Individuals from two other eastern US sites (Mr and PA) were also part of this widespread North American cluster *15*. One individual from group **OnBrF** was found in the Bear River in Idaho (site BrF), more than 2500 km west of the other *P. antipodarum* with the same genotype in Lake Ontario. This result again indicates the high potential of *P. antipodarum* for passive dispersal across large distances (Hubendick, 1950). Additional genotyping of more individuals from the BrF site will be needed to determine whether this western North America snail with an apparently "eastern" genotype is common in this region, providing important information about the connectivity across North American freshwater sites. The general pattern that seems to emerge - that individuals from the Great Lakes and from western US sites constitute two very distinct, different clonal lineages - aligns with the results of earlier analyses of mitochondrial DNA variation (Dybdahl & Drown, 2011; Neiman, et al., 2010, 2011). In a qualitative departure from these earlier studies, we also found that this "western" US haplotype is in fact not limited to western North American sites. Indeed, snails from two eastern North American sites (Mr and PA) shared the same genetic cluster with the majority of the genotypes from the western sites.

We were able to use our data to provide a conclusive answer to the long-standing question of the origin of the *P. antipodarum* invasion of North America. All our analyses also revealed that the Great Lakes group of North American *P. antipodarum* represented by cluster *14* originated in Europe. This result is reinforced by the fact that all individuals from Lake Ontario shared a SNP genotype with individuals from European sites Gb and Mo and differed by one to eight mutations from the other **EU14** genotypes. Individuals of **EU14** (except from the newly sampled and unsequenced sites Mo and Gb) were already known to share the same mtDNA haplotype 22 with individuals from the Great Lakes (Dybdahl & Drown, 2011; Verhaegen et al., 2018a), but so did some New Zealand populations (Neiman & Lively, 2004; Neiman et al., 2011). The use of nuclear SNP markers allowed us to exclude the latter as the source of the North American Great Lakes *P. antipodarum*.

The origin of the North American *P. antipodarum*, represented by ABC group **US1**, is more complex. The ABC suggested that **US1** originated both from Europe [European DAPC cluster *15* (ABC group **EU15**)] and from New Zealand's North Island (ABC group **NZN**). These results are consistent with the DAPC clustering. The NJ tree also showed that New Zealand clades from both islands as well as two European clades (DEBIN and DERUG) are closely related to or ancestral to the **US1** clade. The distribution of **EU15** is much more

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

restricted than **EU14** and included our only brackish-water site, DEBIN. Individuals from **EU15** (except unsequenced site Mo) shared mtDNA haplotype 37 (Verhaegen et al., 2018a), the same haplotype also found in the western USA (Dybdahl & Drown, 2011). The most common genotype of **US1**, genotype 325, was also shared with one individual from the **EU15** site Mo. However, within this same cluster *15*, the other genotypes from **EU15** differed by at least eight mutational steps from those of **US1**. It is difficult to reconcile so many changes accumulating at third-codon positions in protein-coding nuclear DNA sequences with an introduction of asexual snails to western North America only 30 years ago (Bowler, 1991). Accordingly, if **US1** *P. antipodarum* indeed originated from **EU15**, it seems to us more likely that their introduction to North America occurred substantially earlier than reported or that these western North America snails originated from European or New Zealand sites not included in our survey.

The second origin of North American **US1** snails would have been directly from New Zealand, where individuals of haplotype 37 are also present (Neiman & Lively, 2004; Neiman et al., 2010, 2011), including individuals from our dataset (Verhaegen et al., 2018b). A few individuals from **US1** were also part of clusters absent in Europe but present in New Zealand. Twenty-four percent of the individuals from Idaho's Snake River (site Sn) were included in cluster *1*, a cluster otherwise restricted to the North Island of New Zealand. One individual from Idaho's Malad River (Ml) was part of cluster *2*, present on both New Zealand islands. **US1** snails in Idaho harbored relatively high genetic variation (see also BrF discussed above). Because the first reported North American *P. antipodarum* were found in the Snake River (Bowler, 1991), Idaho, it is possible that this variation could have arisen since invasion. If this variation did not arise *in situ*, this assemblage must be due to additional dispersal, possibly through transportation by birds (see Zielske et al., 2017). This latter scenario, however, is in contradiction to the ABC analyses rejecting ghost populations.

To summarise, our North American sites were divided into two distinct genetic groups: one in Lake Ontario, and another spread across the continent. All our analyses pointed towards Europe as the invasion source of *P. antipodarum* from the Great Lakes: some identical SNP genotypes were even shared between some European snails and those of Lake Ontario. The origin of the other widespread North American group is more complex: our analyses pointed towards both Europe and New Zealand as likely sources.

ABC analyses in the context of asexual reproduction

Recent developments in genomic technologies and sophisticated analytical methods allow us to indirectly retrace historical invasion routes of species that would have otherwise remained invisible (Estoup & Guillemaud, 2010). ABC methods provide a qualitative step forward by enabling the testing of invasion scenarios including unsampled "ghost" populations or specific stochastic events (e.g., admixture events). Like most population genetic analysis frameworks, ABC analyses specifically designed for asexually reproducing organisms are yet to be developed. We addressed this limitation of the ABC approach by excluding all but one member of each clone within each ABC group (e.g., Janoušek et al., 2016; Taerum et al., 2017). Even so, the relatively poor goodness of fit of the final invasion scenario that was selected by this analysis suggests that even this modification might fall short in this instance by simulating the genetic genealogies using algorithms that assume recombination (e.g., in our case the Hudson's algorithm, suitable for SNP data, Hudson, 2002) (Cornuet et. al, 2014). Nevertheless, ABC is still often used to test evolutionary scenarios involving systems that are asexual (e.g., wheat aphids, Zhang et al., 2014), mixed sexual/asexual (e.g., plant pathogens, Goss et al., 2014) or that alternate between asexual and sexual reproduction (e.g., trees, Lander et al., 2011; or plant fungi, Barrès et al., 2012; Janoušek et al., 2016; Taerum et al., 2017). While some of these studies also removed all but one member of a clone (e.g., Janoušek et al., 2016; Taerum et al., 2017), even this correction might not be necessary: Barrès et al. (2012) compared the ABC analyses using clone-corrected vs. uncorrected datasets and found no differences in outcome.

Potamopyrgus antipodarum is very similar to many of the invasive taxa focused on in these other studies in that invasive lineages are asexual while native populations harbor both asexually and sexually reproducing individuals, but is distinct in the high genetic diversity of clonal *P. antipodarum* in the native range (e.g., Paczesniak et al., 2013) relative to invasive populations. By contrast, the genetic diversity of invasive clonal

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

populations in these previous studies that did not seem to experience poor model-fitting was high relative to our invasive *P. antipodarum* populations (e.g., Goss et al., 2014; Zhang et al., 2014), with the former possibly mimicking conditions expected in diverse and outcrossing populations. This possibility suggests that a useful next step might be a simulation-based study investigating the robustness and appropriateness of DIYABC under different levels of clonal diversity.

Despite the poor fit of our final ABC model, the ABC approach is still of use in terms of providing distinct and complementary information relative to non-Bayesian analyses. For example, comparisons of the outcomes of these different approaches revealed similar outcomes of the NJ tree and ABC analyses with respect to the invasion route of **OnBrF**, but different outcomes for **US1.** In this latter case, while the NJ tree suggested New Zealand as origin of **US1**, the ABC instead pointed to both Europe and New Zealand as origin. How do we reconcile the different outcomes of these two analyses? We can begin by considering how the analyses are structured: in particular, the NJ tree is generated from pairwise genetic distances between sampling sites, while the ABC uses genotype data. Sampling sites can contain individuals from different genotype clusters, which can result in apparently contradictory inferences when using methods that assess relationships among sampling sites (i.e., NJ tree) vs. individuals (i.e., ABC). Although our data structure likely violates the ABC modeling assumptions, we cannot yet estimate to which extent the results, which appear intuitively reasonable (Fig. 4), are actually misleading. For example, group US1 contained individuals from clusters *1* and *2* (dominating in New Zealand) and Cluster *15*, which probably arrived via Europe and the eastern United States. This pattern suggests that US1 has a more complex origin involving several sources. This in fact turned out to be the most likely scenario. With this in mind, and given that violations of ABC assumptions are not necessarily detrimental (Barres et al. 2012; Goss et al., 2014; Zhang et al., 2014), we should not blindly discard these inferences until simulations have tested the robustness of DIYABC analyses.

Conclusion and perspectives

Here, we used characterization of the genetic connections across invasive populations of a destructive snail and the relationships of these populations to native counterparts to determine the source of the invaders and how the invasion process has affected population genetic diversity. The answers to these questions are a critical component of understanding invasion success (Keller & Taylor, 2008; Dlugosch et al., 2015; Estoup et al., 2016). More broadly, these data also address important unresolved issues involving the evolutionary potential of asexual lineages and populations with low genetic variation.

Our analyses revealed that the native range of New Zealand is the source of the two main European genetic lineages of *P. antipodarum*. One of these European lineages was the unique source of the North American group of the Great Lakes, whereas the other European lineage was the source, in combination with two New Zealand clusters, of the second main widespread Nothern American cluster. In other words, besides demonstrating low genetic variability on a fine genetic scale across two continents, we found that only a few clonal lineages of *P. antipodarum* are responsible for the contential North American and European invasion.

There are multiple and often non-mutually exclusive potential explanations for invasion success in the context of low genetic diversity (reviewed in Estoup et al., 2016). First, a compelling "null" hypothesis is the absence of an adaptive challenge in the native range, which can occur, for example, in situations where the environments in the native and invaded range are similar. The niche modeling work of Loo et al. (2007) does seem to exclude this possibility, demonstrating that invasive *P. antipodarum* thrive in environments found outside New Zealand. Another hypothesis that must be considered is that relatively low variation in molecular markers does not always translate into relatively low variation in ecologically relevant traits (e.g., conversion of non-additive variation during a bottleneck; reviewed in Neiman & Linksvayer, 2005). Recent studies do reveal substantial genetic variation for fitness-relevant traits in invasive *P. antipodarum* (Neiman & Krist, 2016; Levri et al., 2017), indicating that this scenario deserves more attention. It is also possible that these successful invasive lineages of *P. antipodarum* can compensate for low genetic variation through phenotypic plasticity (reviewed in Estoup et al. 2016; e.g., Verhaegen et al., 2018a). Taken together, our new data and these earlier studies hint that relevant and heritable variation - even in the almost total

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

absence of molecular marker-based estimates of genetic diversity - was likely present in the New Zealand and, subsequently, bridgehead European populations that seeded future invasions.

This hypothesis can be evaluated by including further unsampled potential native source populations more recently invaded regions like Japan and Chile (Hamada et al., 2013; Collado, 2014; Collado & Fuentealba, 2020) as well as comparing the phenotypic means and variances of native vs. invasive clones (e.g.,Keller & Taylor, 2008; Neiman & Krist, 2016; Levri et al., 2017). These studies will provide a powerful test of the role - if any - of genetic and phenotypic variation in driving initial vs. later invasions. Further study of *P. antipodarum* from these perspectives will also help illuminate how asexual lineages and populations with low genetic diversity succeed in new environments.

Acknowledgments

We thank the Iowa Science Foundation, the University of Iowa Center for Global & Regional Environmental Research, and the Dewey Stuit Fund for Undergraduate Research for the collection of *P. antipodarum* at the North American sites BrB, BrF, Co, Gr, LC, Md, Ml, Pn, and Sn. We thank Dr. Amy Krist and Dr. Ed Levri for supplying samples from North American sites Pc and Mr, On, and PA, respectively. Several anonymous reviewers provided very useful and constructive feedback. This study was funded by the Research Training Group 2010 RESPONSE grant of the Deutsche Forschungsgemeinschaft (DFG) for the European and New Zealand data, and by the Iowa Science Foundation for the North American data.

References

Alonso, Á., & Castro-Díez, P. (2008). What explains the invading success of the aquatic mud snail *Potamopyrgus antipodarum* (Hydrobiidae, Mollusca)? *Hydrobiologia*, *614*(1), 107–116.

Alonso, Á., & Castro-Díez, P. (2012). The exotic aquatic mud snail *Potamopyrgus antipodarum* (Hydrobiidae, Mollusca): state of the art of a worldwide invasion. *Aquatic Sciences*, *74*(3), 375–383.

Barrès, B., Carlier, J., Seguin, M., Fenouillet, C., Cilas, C., & Ravigné, V. (2012). Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. *Heredity*, *109*(5), 269–279.

Bowler, P. A. (1991). The rapid spread of the freshwater hydrobiid snail *Potamopyrgus antipodarum* (Gray) in the Middle Snake River, southern Idaho. *Proceedings of the Desert Fishes Council*, *21*, 173–182.

Brazier, J. (1871). Descriptions of Seven New Species of the Genus *Helix* and of two Fluviatile Shells from Tasmania. *Proceedings of the Zoological Society of London*, 696–699.

Collado, G. A. (2014). Out of New Zealand: molecular identification of the highly invasive freshwater mollusk *Potamopyrgus antipodarum* (Gray, 1843) in South America. *Zoological Studies*, *53*(1), 1–9.

Collado, G. A., & Collado, G. A. (2014). Out of New Zealand: Molecular identification of the highly invasive freshwater mollusk *Potamopyrgus antipodarum* (Gray, 1843) in South America. *Zoological Studies*, *53*(1), 1–9.

Collado, G. A., & Fuentealba, C. G. (2020). Range extension of the invasive *Potamopyrgus antipodarum* (Gray, 1843)(Gastropoda, Tateidae) in Chile, and a summary of its distribution in the country. *Check List*, *16*(3), 621–626.

Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, *30*(8), 1187–1189.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Cornuet, J. M., Ravigné, V., & Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, *11*. doi: 10.1186/1471-2105-11-401

Cornuet, J. M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J. M., Balding, D. J., ... Estoup, A. (2008). Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics*, *24*(23), 2713–2719.

Dlugosch, K. M., Anderson, S. R., Braasch, J., Cang, F. A., & Gillette, H. D. (2015). The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Molecular ecology*, *24*(9), 2095–2111.

Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, *17*(1), 431–449.

Dusting, A. (2016). *Reconstructing the invasion history of Potamopyrgus antipodarum in Australia* (PhD thesis). University of Canberra.

Dybdahl, M. F., & Drown, D. M. (2011). The absence of genotypic diversity in a successful parthenogenetic invader. *Biological Invasions*, *13*, 1663–1672.

Emblidge Fromme, A., & Dybdahl, M. F. (2006). Resistance in introduced populations of a freshwater snail to native range parasites. *Journal of Evolutionary Biology*, *19*(6), 1948–1955.

Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and so what? *Molecular Ecology*, *19*(19), 4113–4130.

Estoup, A., Lombaert, E., Marin, J. M., Guillemaud, T., Pudlo, P., Robert, C. P., & Cornuet, J. M. (2012). Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discrimimant analysis on summary statistics. *Molecular Ecology Resources*, *12*, 846–855. Retrieved from p 149

Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a genetic paradox of biological invasion? *Annual Review of Ecology, Evolution and Systematics*, *47*, 51–72.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Fukami, H., Budd, A. F., Levitan, D. R., Jara, J., Kersanach, R., & Knowlton, N. (2004). Geographic differences in species boundaries among members of the *Montastraea annularis* complex based on molecular and morphological markers. *Evolution*, *58*(2), 324-337.

Goss, E. M., Tabima, J. F., Cooke, D. E. L., Restrepo, S., Frye, W. E., Forbes, G. A., ... Grünwald, N. J. (2014). The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8791–8796.

Haase, M. (2008). The radiation of hydrobiid gastropods in New Zealand: a revision including the description of new species based on morphology and mtDNA sequence information. *Systematics and Biodiversity*, *6*(01), 99–159.

Hall, R. O., Tank, J. L., & Dybdahl, M. F. (2003). Exotic snails dominate nitrogen and carbon cycling in a highly productive stream. *Frontiers in Ecology and the Environment*, *1*(8), 407–411.

Hall, R., Dybdahl, M. F., & VanderLoop, M. C. (2006). Extremely high secondary production of introduced snails in rivers. *Ecological Applications, 16*, 1121–1131.

Hamada, K., Tatara, Y., & Urabe, M. (2013). Survey of mitochondrial DNA haplotypes of *Potamopyrgus antipodarum* (Caenogastropoda: Hydrobiidae) introduced into Japan. *Limnology*, *14*(3), 223–228.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001). PAST - PAlaeontological STatistics. *Paleontologia Electronica*, (4), 1–9.

Hubendick, B. (1950). The effectiveness of passive dispersal in *Hydrobia jenskinsi. Zoologiska Bidrag Uppsala*, Vol. 28, pp. 493–504.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338.

Ishiwata, S., Saito, K., & Kobayashi, N. (2005). *Benthos in the rivers of Kanagawa Prefecture. Kanagawa environmental Research Center, Hiratsuka (in Japanese)*.

Janoušek, J., Wingfield, M. J., Marmolejo Monsivais, J. G., Jankovský, L., Stauffer, C., Konečný, A., & Barnes, I. (2016). Genetic analyses suggest separate introductions of the pine pathogen *Lecanosticta acicola* into Europe. *Phytopathology*, *106*(11), 1413–1425.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94.

Keller, S. R., & Taylor, D. R. (2008). History, chance and adaptation during biological invasion: Separating stochastic phenotypic evolution from response to selection. *Ecology Letters*, *11*(8), 852–866.

Kistner, E. J., & Dybdahl, M. F. (2014). Parallel variation among populations in the shell morphology between sympatric native and invasive aquatic snails. *Biological Invasions*, *16*(12), 2615–2626.

Kolbe, J. J., Glor, R. E., Schettino, L. R., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, *431*, 177–181.

Lander, T. A., Oddou-Muratorio, S., Prouillet-Leplat, H., & Klein, E. K. (2011). Reconstruction of a beech population bottleneck using archival demographic information and Bayesian analysis of genetic data. *Molecular Ecology*, *20*(24), 5182–5196.

Larkin, K., Tucci, C., & Neiman, M. (2016). Effects of polyploidy and reproductive mode on life history trait expression. *Ecology and Evolution*, *6*(3), 765–778.

Levri, E. P., Landis, S., Smith, B., Colledge, E., Metz, E., & Li, X. (2017). Variation in predator-induced behavioral changes in introduced and native populations of the invasive New Zealand mud snail (*Potamopyrgus antipodarum* Gray, 1843). *Aquatic Invasions*, *12*(4), 499–508.

Lively, C. M. (1987). Maintainence of sex by parasitism. *Nature*, *328*(6), 8–10.

Lombaert, E., Guillemaud, T., Cornuet, J. M., Malausa, T., Facon, B., & Estoup, A. (2010). Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PloS one*, *5*(3), e9743.

Lombaert, E., Guillemaud, T., Lundgren, J., Koch, R., Facon, B., Grez, A., ... Estoup, A. (2014). Complementarity of statistical treatments to reconstruct worldwide routes of invasion: The case of the Asian ladybird *Harmonia axyridis*. *Molecular Ecology*, *23*(24), 5979–5997.

Loo, S. E., Nally, R. M., & Lake, P. S. (2007). Forecasting New Zealand mudsnail invasion range: model comparisons using native and invaded ranges. *Ecological Applications*, *17*(1), 181–189.

Lucas, A. (1959). Les Hydrobia (Bythinellidae) de l'Ouest de la France. *Journal de Conchyliologie*, *99*, 3–14.

Mallez, S., & McCartney, M. (2018). Dispersal mechanisms for zebra mussels: population genetics supports clustered invasions over spread from hub lakes in Minnesota. *Biological Invasions*, *20*(9), 2461–2484.

Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, *4*(4), 792–794.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.

Mortimer, N., Campbell, H. J., Tulloch, A. J., King, P. R., Stagpoole, V. M., Wood, R. A., ... & Seton, M. (2017). Zealandia: Earth's hidden continent. *GSA today*, *27*(3), 27-35.

Neiman, M, & Lively, C. M. (2004). Pleistocene glaciation is implicated in the phylogeographical structure of Potamopyrgus antipodarum, a New Zealand snail. *Molecular Ecology*, *13*, 3085–3098.

Neiman, M., & Linksvayer, T. A. (2006). The conversion of variance and the evolutionary potential of restricted recombination. *Heredity*, *96*(2), 111–121.

Neiman, M, Paczesniak, D., Soper, D. M., Baldwin, A. T., & Hehman, G. (2011). Wide variation in ploidy level and genome size in a new zealand freshwater snail with coexisting sexual and asexual lineages. *Evolution*, *65*(11), 3202–3216.

Neiman, Maurine, Hehman, G., Miller, J. T., Logsdon, J. M., & Taylor, D. R. (2010). Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Molecular Biology and Evolution,* 27*(4), 954–963.*

Neiman, M., & Krist, A. (2016). Sensitivity to dietary phosphorus limitation in native vs. invasive lineages of a New Zealand freshwater snail. *Ecological Applications*, *26*(7), 2218–2224.

Nentwig, W., Bacher, S., Kumschick, S., Pyšek, P., & Vilà, M. (2018). More than "100 worst" alien species in Europe. *Biological Invasions*, *20*(6), 1611–1621.

Paczesniak, D., Jokela, J., Larkin, K., & Neiman, M. (2013). Discordance between nuclear and mitochondrial genomes in sexual and asexual lineages of the freshwater snail *Potamopyrgus antipodarum*. *Molecular Ecology*, *22*(18), 4695–4710.

Parks, D. H., Mankowski, T., Zangooei, S., Porter, M. S., Armanini, D. G., Baird, D. J., ... Beiko, R. G. (2013). GenGIS 2: Geospatial Analysis of Traditional and Genetic Biodiversity, with New Gradient Algorithms and an Extensible Plugin Framework. *PLoS ONE*, *8*(7).

Ponder, W. F. (1988). *Potamopyrgus antipodarum*: a molluscan colonizer of Europe and Australia. *Journal of Molluscan Studies*, *54*(3), 271–286.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Reed, E. M. X., Serr, M. E., Maurer, A. S., & Reiskind, M. B. (2020). Gridlock and beltways: the genetic context of urban invasions. *Oecologia*, 1-14.

Rogstad, S. H., Keane, B., & Beresh, J. (2002). Genetic variation across VNTR loci in central North American *Taraxacum* surveyed at different spatial scales. *Plant Ecology*, *161*(1), 111–121.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457–462.

Smith, E. A. (1889). Notes on British Hydrobiae with a description of a supposed new species. *Journal of Conchology*, *6*, 142–145.

Städler, T., Frye, M., Neiman, M., & Lively, C. M. (2005). Mitochondrial haplotypes and the New Zealand origin of clonal European Potamopyrgus, an invasive aquatic snail. *Molecular Ecology*, *14*, 2465–2473.

Swofford, D. L. (1998). *Phylogenetic Analysis Using Parsimony PAUP* 4.0 beta version* (pp. 1–140). pp. 1–140. Retrieved from http://www.lms.si.edu/%0Ahttp://benedick.rutgers.edu/software-manuals/PAUP4-manual.pdf

Taerum, S. J., Hoareau, T. B., Duong, T. A., de Beer, Z. W., Jankowiak, R., & Wingfield, M. J. (2017). Putative origins of the fungus *Leptographium procerum*. *Fungal Biology*, *121*(1), 82–94.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Taylor, D. W. (1987). Thousand Springs threatened or endangered snails. *Unpublished report submitted to The Nature Conservancy summarizing a 2-5 September 1987 survey.*

Verhaegen, G., McElroy, K. E., Bankers, L., Neiman, M., & Haase, M. (2018a). Adaptive phenotypic plasticity in a clonal invader. *Ecology and Evolution*, *8*(9), 4465–4483.

Verhaegen, G., Neiman, M., & Haase, M. (2018b). Ecomorphology of a generalist freshwater gastropod: complex relations of shell morphology, habitat, and fecundity. *Organisms Diversity & Evolution*, *18*(4), 425–441.

Verhaegen, G., von Jungmeister, K., & Haase, M. (*in press*). Life history variation in space and time: environmental and seasonal responses of a parthenogenetic invasive freshwater snail in northern Germany. *Hydrobiologia*.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358.

Winterbourn, M. J. (1970). Population studies on the New Zealand freshwater gastropod *Potamopyrgus antipodarum* (Gray). *Journal of Molluscan Studies*, *39*(2–3), 139–149.

Wylie, R., Yang, C. C. S., & Tsuji, K. (2020). Invader at the gate: The status of red imported fire ant in Australia and Asia. *Ecological Research*, *35*(1), 6-16.

Zaranko, D.T., Farara, D. G., & Thompson, F. G. (1997). Another exotic mollusc in the Laurentian Great Lakes: the New Zealand native *Potamopyrgus antipodarum* (Gray 1843) (Gastropoda, Hydrobiidae). *Canadian Journal of Fisheries and Aquatic Sciences*, *54*(4), 809–814.

Zaranko, D. T., Farara, D. G., & Thompson, F. G. (1997). Another exotic mollusc in the laurentian great lakes: The new zealand native *Potamopyrgus antipodarum* (Gray 1843) (Gastropoda, Hydrobiidae). *Canadian Journal of Fisheries and Aquatic Sciences*, *54*(4), 809–814.

Zhang, B., Edwards, O., Kang, L., & Fuller, S. (2014). A multi-genome analysis approach enables tracking of the invasion of a single Russian wheat aphid (*Diuraphis noxia*) clone throughout the New World. *Molecular Ecology*, *23*(8), 1940–1951.

Zielske, S., Ponder, W. F., & Haase, M. (2017). The enigmatic pattern of long-distance dispersal of minute freshwater gastropods (Caenogastropoda, Truncatelloidea, Tateidae) across the South Pacific. *Journal of Biogeography*, *44*(1), 195–206.

Data Accessibility Statement

The genetic data used in this study is given in the Supplemental material, as Supplemental Files 1 to 3.

Author contributions

- Manuscript: GV, CD, MN, MH, JDW
- Design, concept: GV, MN, MH
- Fieldwork of new samples: JDW, MN
- DNA extraction of new samples: JDW
- Genetic data of Europe and New Zealand: GV, MH
- Analyses: GV, CD, MH

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Tables

Table 1 Sampling sites and results of the genetic diversity summary statistics as implemented in Genodive analyses. Standard genetic results (Ho and Hs) were not calculated for the asexual populations in the invaded range, neither for New Zealand populations that had all their individuals sharing the same SNP genotype.

Footnote: $^{+}$ABC = Approximate Bayesian Computation, $^{++}$Ho = observed heterozygosity, ^{SS}Hs = expected heterozygosity.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Acronym	Country	Location	Date (month year)	Latitude	Longitude	N	ABC+ group	Ho+	Hs§	Number of unique genotypes
NZ01	New Zealand	Whatawhata, West Waikato	February 2016	S 37° 47' 7.09"	E 175° 4' 7.26"	20	NZN	0.3	0.2	7
NZ03	New Zealand	Whatawhata	February 2016	S 37° 47' 2.3"	E 175° 4' 10.4"	20	NZN	0.34	0.23	8
NZ04	New Zealand	Bridal Veil Falls	February 2016	S 37° 54' 31.17"	E 174° 53' 45.4"	20	NZN	0.23	0.22	8
NZ07	New Zealand	McLaren Falls	February 2016	S 37° 48' 52.57"	E 176° 2' 48.02"	20	NZN	0.2	0.18	7
NZ08	New Zealand	McLaren Falls	February 2016	S 37° 49' 3.02"	E 176° 2' 49.19"	20	NZN	0.32	0.19	2
NZ10	New Zealand	Kaniwhaniwha reserve	February 2016	S 37° 55' 12.79	E 175° 4' 52.88"	20	NZN	0.38	0.28	6
NZ14	New Zealand	Wairere falls	February 2016	S 37° 44' 13.89"	E 175° 51' 52.3"	20	NZN	—	—	1
NZ19	New Zealand	Crown track, Karangahake	February 2016	S 37° 25' 45.77"	E 175° 43' 36.32"	20	NZN	0.08	0.08	19
NZ21	New Zealand	North of Kaimai-Mamaku Forest Park	February 2016	S 37° 28' 20.41"	E 175° 46' 44.67"	20	NZN	0.28	0.21	4
NZ23	New Zealand	North of Kaimai-Mamaku Forest Park	February 2016	S 37° 27' 51.84"	E 175° 46' 46.46"	20	NZN	0.09	0.09	18
NZ24	New Zealand	North of Kaimai-Mamaku Forest Park	February 2016	S 37° 27' 47.09"	E 175° 46' 48.07"	20	NZN	0.32	0.25	2
NZ27	New Zealand	Waitomo	February 2016	S 38° 15' 57.2"	E 175° 4' 37.47"	20	NZN	0.27	0.24	7
NZ29	New Zealand	Te waihou walkway to blue spring	February 2016	S 38° 1' 59.8"	E 175° 49' 31.82"	20	NZN	0.17	0.17	11
NZ31	New Zealand	Wainui falls track, Tata beach	February 2016	S 40° 50' 36.16"	E 172° 56' 9.52"	20	NZS	0.31	0.28	7
NZ32	New Zealand	McShane road, Tata beach	February 2016	S 40° 49' 43.24"	E 172° 56' 44.49"	20	NZS	0.38	0.28	5
NZ37	New Zealand	Collingwood, McDonald Footbridge	February 2016	S 40° 40' 52.41"	E 172° 40' 26.01	20	NZS	0.38	0.22	2
NZ39	New Zealand	Takaka	March 2016	S 41° 4' 49.3"	E 172° 44' 16.0"	20	NZS	0.29	0.24	2
NZ41	New Zealand	Salmon farm, Takaka	March 2016	S 40° 53' 0.73"	E 172° 46' 1.51"	20	NZS	0.36	0.26	7
NZ42	New Zealand	Dry road, Rakopi	March 2016	S 40° 36' 33.84"	E 172° 34' 10.05"	20	NZS	0.39	0.25	5

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Table 2 Groups of individuals used in the Approximate Bayesian Computation analyses

Acronym	Individuals the group contains
NZN	Individuals from New Zealand North Island
NZS	Individuals from New Zealand South Island
EU14	European individuals included in DAPC cluster *14*
EU15	European individuals included in DAPC cluster *15*
OnBrF	American individuals included in DAPC cluster *14*, i.e. all individuals from site On and one from BrF
US1	American individuals included in DAPC clusters *1*, *2* and *15*, i.e. individuals from all North American sites ex

Figure legends

Figure 1 Neighbour-joining tree based on pairwise Slatkin's linearized F_{ST} values and rooted by *Potamopyrgus estuarinus* (= pink). Blue = New Zealand sites; orange = North American sites; green = European sites. Corresponding information for the acronyms can be found in Table 1.

Figure 2 Geographical mapping of the neighbour-joining tree. Blue = New Zealand sites; orange = American sites; green = European sites; pink, *Potamopyrgus estuarinus* (outgroup).

Figure 3 Scatterplot of the first two axes of the discriminant analysis of principal components (DAPC) showing the relationship of 15 clusters. Ellipses represent inertia ellipses. Clusters in dark blue contain exclusively New Zealand individuals; light blue = cluster *1*; pink = cluster *2*; green = cluster *14*; orange = cluster *15*. The DAPC was built by maintaining the first 36 PCs and six discriminant functions.

Figure 4 Geographical distribution of the 15 DAPC clusters. Clusters in dark blue only contained individuals from New Zealand and are therefore not distinguished from each other on this map.

Figure 5 The three competing invasion scenarios of the final Approximate Bayesian Computation (ABC6) analysis with their posterior probabilities and 95% confidence intervals in square brackets. The selected scenario, scenario 2, is highlighted in bold. Bottleneck events represented by dotted lines. The timeline is not to scale. For the constitution of the different ABC groups, see Table 2.

Figure 6 Simplified map-based representation of our current picture of the *P. antipodarum* global invasion. Non-black line colours represent cluster membership for EU14, EU15, OnBrF, and US1. Lines are drawn to roughly represent known shipping routes most likely to have been the source of *P. antipodarum* colonization. The dotted black lines connecting New Zealand to Tasmania and Tasmania to South Australia represent likely invasion routes, though not tested directly here. We have also included the recent Chilean and Japanese invasions on the map, though the genetic data needed to infer invasion routes are not yet available.

Figure 7 Principle component analysis (PCA) for checking the goodness of fit of the most likely scenario 2 of the final Approximate Bayesian Computation analysis ABC5. The PCA compares the prior and posterior distribution of the summary statistics, and the observed data set. The variance explained by each principle component (PC1 and PC2) are given in brackets.

Supplemental figures legends

Supplemental Fig. 1 Invasion scenarios to Europe tested in the Approximate Bayesian Computation analysis ABC1. Bottleneck events are indicated with a lower case "b", ghost populations as "NA".

Supplemental Fig. 2 Invasion scenarios to the USA without presence of ghost populations tested in the Approximate Bayesian Computation analysis ABC2. Bottleneck events are indicated with "b".

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Supplemental Fig. 3 Invasion scenarios to the USA including a ghost population originating from the population **NZN** (New Zealand North Island), and scenarios with two ghost populations originating from **NZN** and **NZS** (New Zealand South Island), respectively, tested in the Approximate Bayesian Computation analysis ABC3. Bottleneck events are indicated with "b"; ghost populations as "NA". The scenarios are given in the historical model parameterization code format of DIYABC v.2.1.0 (Cornuet et al., 2014).

Supplemental Fig. 4 Invasion scenarios to the USA including a ghost population originating from the population **NZS** (New Zealand South Island), tested in the Approximate Bayesian Computation analysis ABC4. Bottleneck events are indicated with "b"; ghost populations as "NA". The scenarios are given in the historical model parameterization code format of DIYABC v.2.1.0 (Cornuet et al., 2014).

Supplemental Fig. 5 The three competing invasion scenarios of the Approximate Bayesian Computation analysis ABC5 with their posterior probabilities and 95% confidence intervals in square brackets. The selected scenario, scenario 1, is highlighted in bold. Unsampled "ghost" populations are indicated in grey; bottleneck events represented by dotted lines. The timeline is not to scale. For the constitution of the different ABC groups, see Table 2.

Supplemental Fig. 6 Distribution of the individuals from New Zealand (blue), Europe (green), and North America (orange) within the 15 clusters. Only one individual per unique SNP genotype was kept per sampling site for the clustering analysis.

Supplemental tables legends

Supplemental Table 1 Pairwise Slatkins linearised F_{ST} values between sampling sites.

Supplemental Table 2 *A priori* and *a posteriori* cluster assignment for all the individuals, as well as *a posteriori* cluster assignment probability.

Supplemental Table 3 Uncorrected number of mutational changes between pairs of genotypes from DAPC clusters including European and American individuals (excel sheet 1); number of individuals for each genotype per continent and DAPC cluster (excel sheet 2)

Supplemental Table 4 Posterior probabilities and 95% confidence intervals for all tested scenarios in the Approximate Bayesian Computation analyses ABC1-6.

Supplemental files legends

Freshwater mudsnail (Potamopyrgus antipodarum) estrogen receptor Identification and expression analysis under exposure to (xeno-)hormones

Supplemental File 1 Raw SNP results for all 1617 collected snails. Excel sheet 1 and 2 = results for the European snails genotyped in Verhaegen et al. (2018a), sheet 3 = results for the New Zealand snails genotyped in Verhaegen et al. (2018b), sheet 4 = results for the US and European snails genotyped in this study, sheet 4 = summary of polymorphic and fixed SNP markers among continents. Corresponding information for the population acronyms can be found in Table 1. More information on the SNP IDs can be found in the Supplementary Table 2 of Verhaegen et al. (2018b).

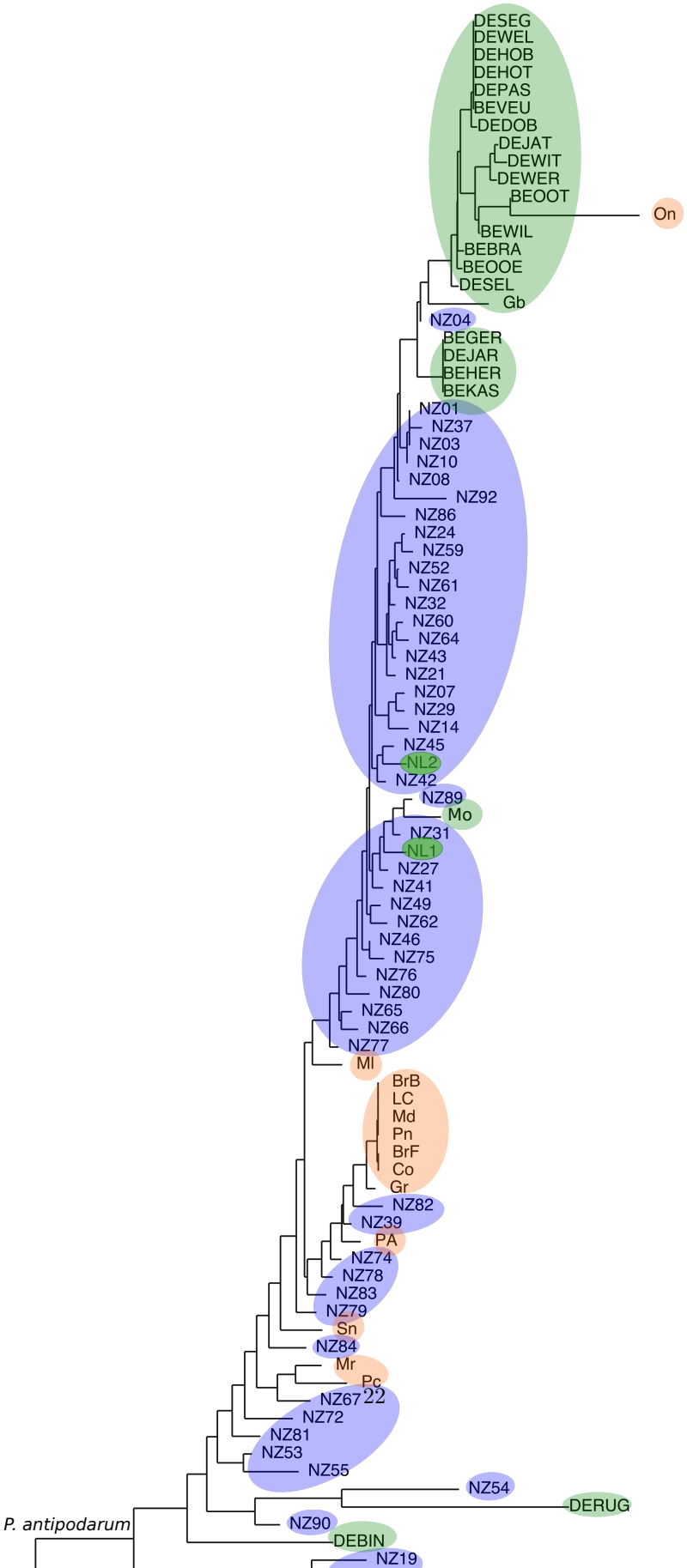
Supplemental File 2 GENEPOP file used for the genetic population analyses, using the same 36 polymorphic SNP loci as in Verhaegen et al. (2018b). All but one individual per genotype within each population were discarded in order to avoid treating likely clonemates as independent data points. Adenine bases were coded as "01", cytosine as "02", guanine as "03", thymine as "04', and unknown bases as "00".

Supplemental File 3 Input file with our SNP data for the ABC analyses ABC2 – 6 conducted with DIYABC v.2.1.0 (Cornuet et al., 2014). Three of our 36 polymorphic SNP markers (SNP IDs: comp148591_c0_seq1,

comp149304_c0_seq1, and comp160266_c0_seq4) were discarded from all our ABC analyses because they could not be genotyped for all individuals within at least one group. All but one individual per genotype within each ABC group were discarded in order to avoid treating likely clonemates as independent data points. For ABC1, only the New Zealand and European groups were used.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.



Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

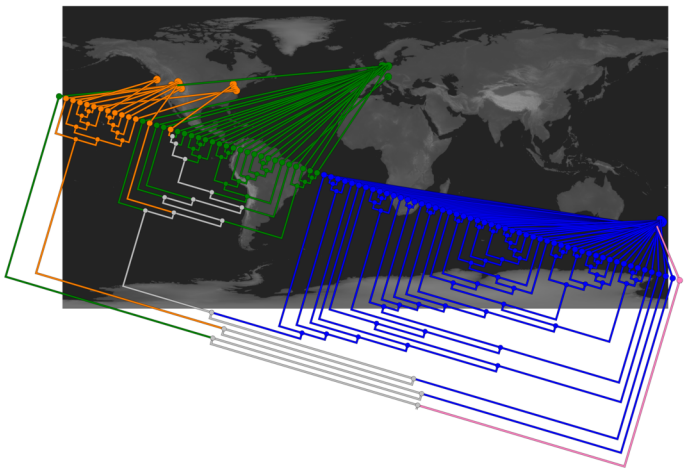


Figure 2: Geographical mapping of the neighbour-joining tree. Blue = New Zealand sites; orange = American sites; green = European sites; pink, *Potamopyrgus estuarinus* (outgroup).

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

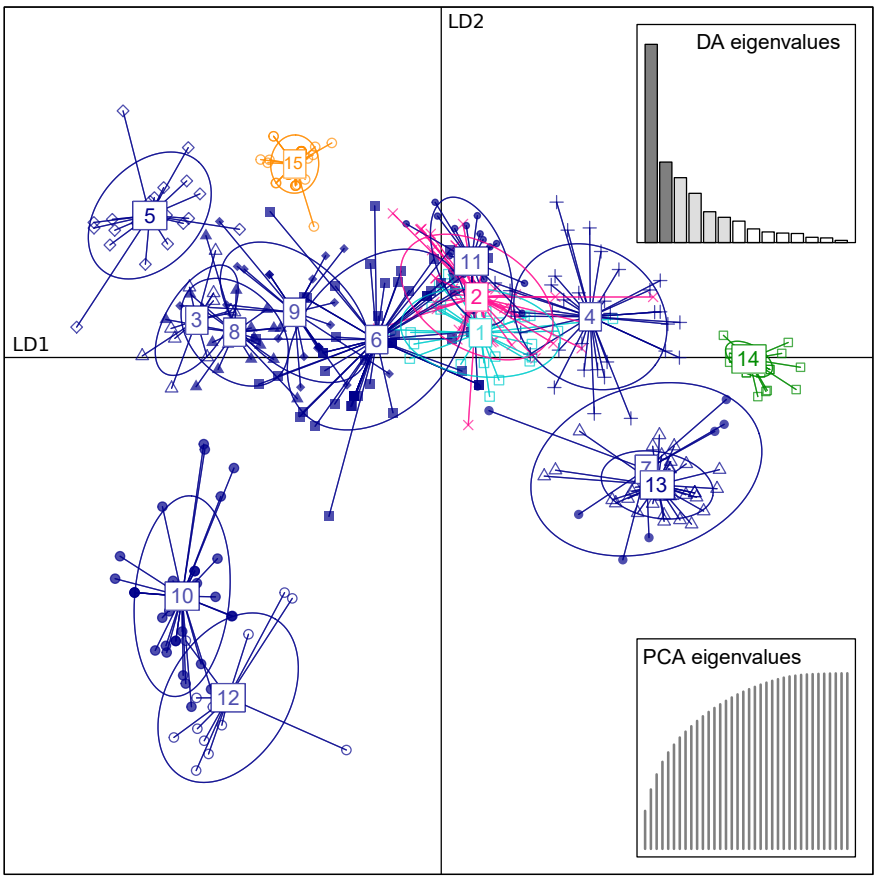


Figure 3: Scatterplot of the first two axes of the discriminant analysis of principal components (DAPC) showing the relationship of 15 clusters. Ellipses represent inertia ellipses. Clusters in dark blue contain exclusively New Zealand individuals; light blue = cluster *1*; pink = cluster *2*; green = cluster *14*; orange = cluster *15*. The DAPC was built by maintaining the first 36 PCs and six discriminant functions.

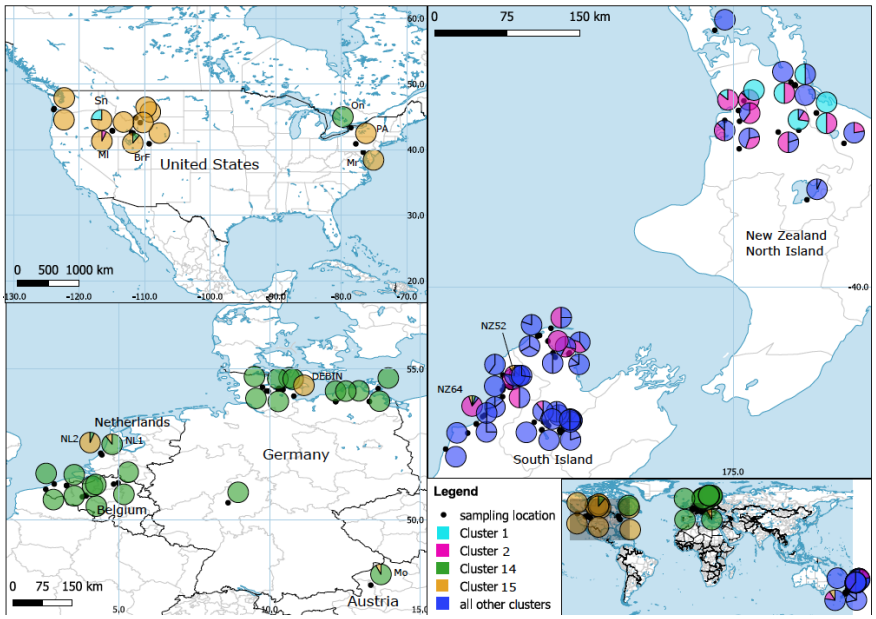


Figure 4: Geographical distribution of the 15 DAPC clusters. Clusters in dark blue only contained individuals from New Zealand and are therefore not distinguished from each other on this map.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

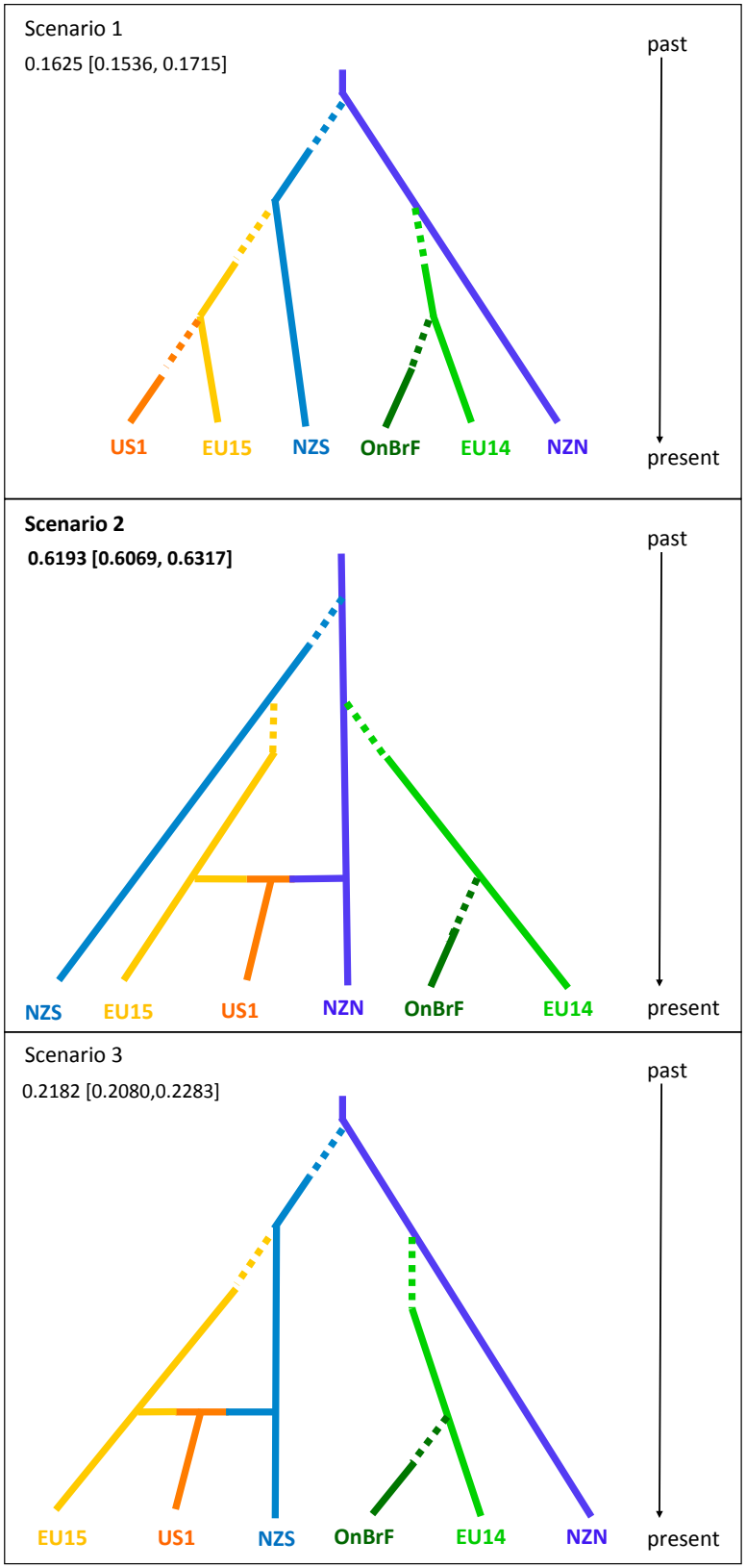


Figure 5: The three competing invasion scenarios of the final Approximate Bayesian Computation (ABC6) analysis with their posterior probabilities and 95% confidence intervals in square brackets. The selected scenario, scenario 2, is highlighted in bold. Bottleneck events represented by dotted lines. The timeline is not to scale. For the constitution of the different ABC groups, see Table 2.

Posted on Authorea 16 Jul 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.158456440.01903363/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

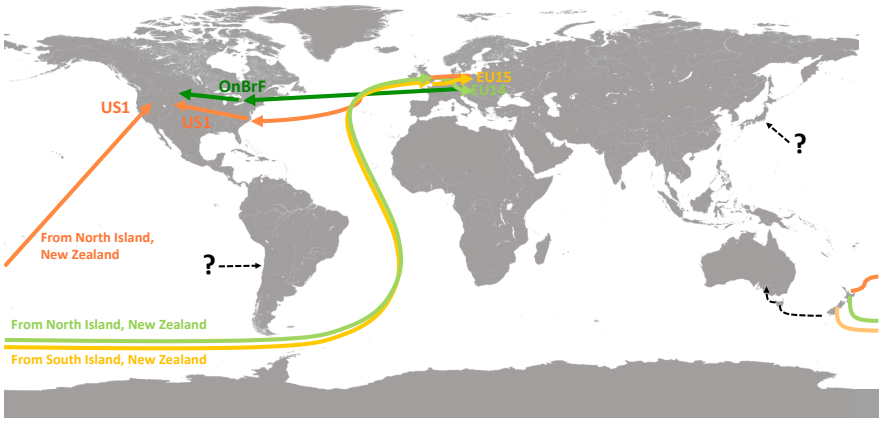


Figure 6: Simplified map-based representation of our current picture of the *P. antipodarum* global invasion. Non-black line colours represent cluster membership for EU14, EU15, OnBrF, and US1. Lines are drawn to roughly represent known shipping routes most likely to have been the source of *P. antipodarum* colonization. The dotted black lines connecting New Zealand to Tasmania and Tasmania to South Australia represent likely invasion routes, though not tested directly here. We have also included the recent Chilean and Japanese invasions on the map, though the genetic data needed to infer invasion routes are not yet available.

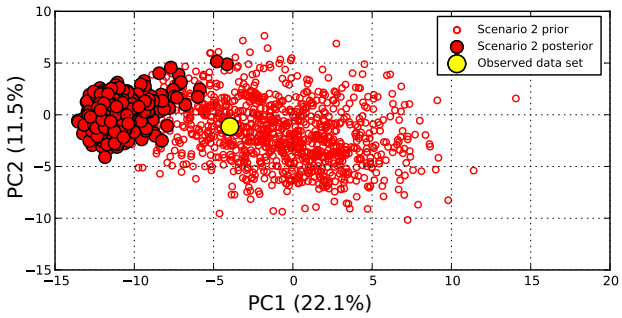


Figure 7: Principle component analysis (PCA) for checking the goodness of fit of the most likely scenario 2 of the final Approximate Bayesian Computation analysis ABC5. The PCA compares the prior and posterior distribution of the summary statistics, and the observed data set. The variance explained by each principle component (PC1 and PC2) are given in brackets.