PNAS
www.pnas.org

Main Manuscript for

Genomics-informed models reveal extensive stretches of coastline under threat by an ecologically dominant invasive species

Jamie Hudson[1,*], Juan Carlos Castilla[2], Peter R. Teske[3], Luciano B. Beheregaray[4], Ivan D. Haigh[1], Christopher D. McQuaid[5], Marc Rius[1,3]

[1] School of Ocean and Earth Science, University of Southampton, National Oceanography Centre, Southampton, SO14 3ZH, United Kingdom
[2] Departamento de Ecología and Estación Costera de Investigaciones Marinas, Las Cruces, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Alameda Libertador B. O´Higgins 340, Santiago, Chile
[3] Centre for Ecological Genomics and Wildlife Conservation, Department of Zoology, University of Johannesburg, Auckland Park 2006, South Africa
[4] Molecular Ecology Lab, Flinders University, Bedford Park, 5042, South Australia, Australia
[5] Department of Zoology and Entomology, Rhodes University, Grahamstown 6140, South Africa

* Corresponding author: Jamie Hudson

Email: J.Hudson@soton.ac.uk

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological sciences, Evolution

Keywords: Ascidians, climate change, intertidal, invasion biology, population genomics, range expansion, thermal biology.

This PDF file includes:

Main Text
Figures 1 to 3

Abstract

Explaining why some species are widespread whilst others are not is fundamental to biogeography, ecology and evolutionary biology. A unique way to study evolutionary and ecological mechanisms that either limit species' spread or facilitate range expansions is to conduct research on species that have restricted distributions. Non-indigenous species, particularly those that are highly invasive but have not yet spread beyond the introduced site, represent ideal systems to study range size changes. Here, we used species distribution modelling and genomic data to study the restricted range of a highly invasive Australian marine species, the ascidian *Pyura praeputialis*. This species is an aggressive space occupier in its introduced range (Chile), where it has fundamentally altered the coastal community. We found high genomic diversity in both native and introduced populations, indicating high adaptive potential in Chile. In addition, genomic data clearly showed that a single region from Australia was the only donor of genotypes to the introduced range. We identified over 3,500 km of suitable habitat adjacent to its current introduced range that has so far not been occupied, though species distribution models were only accurate when genomic data were considered. Our results suggest that a slight change in currents, or a change in shipping routes, may lead to an expansion of the species' introduced range that will encompass a vast portion of the South American coast. Our study shows how combining species distribution modelling and population genomic data can unravel mechanisms shaping range sizes and forecast future range shifts of invasive species.

Significance Statement

Species with narrow distributions provide unique opportunities for understanding the mechanisms that limit their spread. We studied a marine invader that exhibits ecological dominance within its range, and has the capacity to fundamentally alter the coastal habitat when introduced to new locations via artificial transport. We found evidence of the species' potential to establish itself far beyond its present introduced range from both genomic and species distribution modelling data. Therefore, minor oceanographic changes (due to for example contemporary climate change) or alteration to human-mediated dispersal of species may trigger a large-scale invasion along vast stretches of coastlines leading to alterations of community structure. Our work provides a holistic framework to assess potential changes in the distribution of invasive species.

Main Text

Introduction

Fundamental to biogeography, ecology and evolutionary biology is understanding why some species are widespread whereas other are not (1). Studies focussing on species ranges often use predictive modelling to infer the potential spatial spread of species (2), with a growing number of studies comparing fundamental niches (i.e., the entire set of conditions that a species can tolerate) and realised niches [the actual set of conditions under which a species is found (3–5)]. In recent times, our understanding of range sizes has advanced considerably thanks to integrative studies (6, 7) that have provided key insights into how local adaptation (8), physiological tolerance (9), and propagule dispersal (10) shape population persistence and spread potential. Despite all this progress, our understanding of how ecological and evolutionary mechanisms shape range sizes and niche occupancy remains limited (11–13).

Non-indigenous species (NIS) offer unique opportunities to study range sizes, especially when they establish and spread into new geographic areas. Most studies to date have focused on highly invasive species that have already had significant ecological and economic impacts (14, 15), and that have spread over large areas of their introduced range (16). In turn, relatively little research has been afforded to invasive species with restricted introduced ranges, or naturalised species [i.e., species that establish self-sustaining populations beyond their native range, but that have not yet expanded from their point of introduction (17–19)]. These species represent unique systems to study mechanisms mediating range sizes (18, 20), such as biotic resistance (e.g. 21, 22) and genetic bottlenecks (23). These study systems can thus be key for understanding mechanisms responsible for limiting NIS spread and/or facilitating range expansions (24).

A rarely used approach to study NIS is the combined use of species distribution models (SDMs) and population genomics (e.g. 25). SDM has become a widely used method to both identify regions of suitable habitat across landscapes (26) and to predict areas at risk of future range shifts (27)(28) . In addition, recently developed analytical tools in population genomics offer powerful ways of studying demographic history, fine-scale population structure, adaptive divergence (29) and eco-evolutionary processes associated with NIS' range shifts (30, 31). Therefore the use of both genomic data and SDM therefore has the potential of improving our ability to characterise the mechanisms that shape range sizes. This includes allowing high resolution spatial delineation of population structure (32), identification of landscape elements that drive ecological and evolutionary patterns (33), and substantially improving our ability to predict future range shifts (25, 34).

Here we used population genomics and SDMs to investigate ecological and evolutionary mechanisms that shape the distribution of a species that has proven to be an aggressive invader, but that has a geographically constrained introduced range. In addition, we evaluated the potential for the spread of this species to adjacent areas. We specifically aimed to: i) understand the nature of recent range size changes through the reconstruction of the species' invasion history, ii) determine the occupancy levels of suitable habitat within and beyond the current native and introduced ranges, and iii) evaluate if our data can explain the current constrained distribution of the study species and predict future spread. We predicted that only a sub-set of the genomic diversity from the native range would be present in the introduced range, potentially indicating the presence of a genetic bottleneck that limits the spread potential of this NIS. Additionally, we predicted that our genomics-informed SDM would reveal suitable habitat along adjacent coastlines of the current introduced range, but that lack of adaptive capacity, as revealed by low genomic diversity in introduced populations, would limit its spread.

Results

Processing of raw GBS data

We retained a total of 1,205 putatively unlinked single nucleotide polymorphisms (SNPs), with the final dataset comprising 164 individuals from 13 sampling sites (see *SI Appendix* and Table S1). A total of 49 candidate adaptive loci were identified by bayenv2 and 30 by Redundancy Analysis (RDA) (Fig. S1), with 14 loci shared between the two methods (for a full description of the genotype-environment association (GEA) analyses, see *SI Appendix*). We therefore treated the 65 loci that were retained by either method as candidate adaptive loci (hereby called the "candidate dataset") and generated a "neutral dataset" using the remaining 1,140 putatively neutral SNPs.

Population structure and reconstruction of invasion routes

ADMIXTURE, discriminant analysis of principal components (DAPC), and F_{ST} (Figs. S2, S3, and S4) analyses based on both the neutral and candidate datasets all supported the result of two highly differentiated groups of populations within Australia. The individuals from Antofagasta Bay were recovered in the same cluster as those from the eastern Australian sites (Fig. 1). When south-eastern Australian populations were removed from the DAPC, the populations from Antofagasta Bay and eastern Australia still clustered together (Figs. S3C, S3D). In contrast, there was no fine-scale genomic structure within the introduced range (Figs. S3E, S3F). This was also the case when only candidate loci associated with SST (as inferred from the RDA) were used (Fig. S5). Similarly, Approximate Bayesian Computation (ABC) analyses showed that Antofagasta Bay was most likely founded by individuals from eastern Australia (Probability P=1.000, 95% CI=1.000, 1.000; Table S2 and Fig. S6), rather than from south-eastern Australia or from an admixture of the two Australian lineages (Table S2). In addition, these simulations suggested that the effective population size introduced to Chile from eastern Australia was of the order of thousands (Fig. S7).

Species distribution modelling

When models were produced using the appropriate genetic lineage of *P. praeputialis* (i.e. genomics-informed, see Fig. 2), coastlines adjacent to and far beyond the introduced range of Antofagasta Bay were found to be suitable habitat for this species. These models showed a dominating effect of the variable 'distance to shore', with 'maximum sea surface temperature' being the second most important variable (Table S3). Upon removing the 'distance to shore' variable, we observed models where the variables 'maximum current velocity' and 'maximum sea surface temperature' were the most dominant (Table S4, Fig. S8). For comparison, SDMs produced excluding the introduced range and built using native lineages including one lineage that did not source the introduction (i.e. not genomics-informed, see Fig. 3) did not recover vast stretches of coastline as suitable for *P. praeputialis*.

Discussion

Our study shows how combining SDMs and population genomic approaches can refine both our understanding of mechanisms responsible for range size changes and our predictions of NIS spread potential. By utilising genomics-informed SDMs, we unravelled the spread potential of a regionally constrained introduced marine invasive speces. We first found genomic evidence of considerable adaptive potential in the highly restricted introduced range, suggesting great potential for spread. In addition, our analyses revealed large population size in the introduced range, which is in line with historic high levels of artificial transport between the native and introduced ranges. Although only one of two lineages contributed to the successful introduction of this species, no footprint of a genetic bottleneck could be found. With this information, our SDM showed that over 3,500km of coastline along the eastern Pacific is potentially at risk of invasion.

The aggressive invasive behaviour shown by this bioengineering NIS within its introduced range suggests that if this species expands its range, it could potentially threaten marine ecosystems along thousands of kilometres of coastline (including the coastline of northern Chile, as well as the entire coastline of Peru and much of the coast of Ecuador).

Our genomic dataset confirmed the presence of two native lineages of *P. praeputialis* and no evidence of finer-scale cryptic population structure, as previously suggested using a limited number of genetic markers (35, 36). We found that south-eastern Australian populations did not contribute to the invasion of *P. praeputialis* in Chile, confirming a single-lineage introduction to Chile (i.e. the lineage found on the east coast of Australia, Fig. 1C) (35, 37). The south-eastern Australian lineage of *P. praeputialis* also inhabits a region with busy ports (e.g. Melbourne), and one would expect that this region would also have been linked to South America through historical shipping. Species that occur in multiple biogeographic zones are often subdivided into distinct evolutionary lineages that are adapted to regional conditions (38), and movement to other biogeographic zones can disrupt growth, reproduction, development, and survival of the migrants (39, 40). Such effects may have limited the ability of the south-eastern Australian lineage to colonise South America, an interpretation supported by the limited suitability of the Chilean coastline for this lineage inferred by our SDMs (Fig. 3E). The first introduction of *P. praeputialis* to Chile would most likely have occurred during the late 19th century when maritime traffic was not only considerable between Australia and Chile, but when fouling organisms were rarely removed from the wooden hulls of ships (41).

High standing genomic diversity is expected to be the primary cause of adaptive potential in NIS (42) within both terrestrial (43, 44) and aquatic (45) environments. We found similar levels of genomic diversity across the range of the study species (Table S5), and thus such observed levels of genomic diversity within the introduced range were not consistent with our initial hypothesis of a genetic bottleneck. Instead, these showed evidence of multiple introductions (46). Our results therefore suggest that if the species was to spread beyond Antofagasta Bay, it may have sufficient genomic variation to adapt to the variety of conditions found along the South American coastline.

Research on local retention of marine organisms has shown that both active and passive mechanisms can promote such retention, including nutrient composition (47), odour cues (48), and hydrodynamic processes (49). Previous research has shown that idiosyncratic characteristics of both Antofagasta Bay and *P. praeputialis* facilitate retention of this ascidian within the bay. For example, an 'upwelling shadow' (sensu 50) is present within Antofagasta Bay, leading to stratification due to a shallow thermocline, cyclonic circulation and high retention of water, with the existence of a persistent warm-water patch (51–53). This patch is generally found within the bay immediately downwind of an upwelling centre, with temperatures on average 2-3°C warmer inside the bay than outside (52, 53). Such phenomena are not limited to Antofagasta Bay, with similar features exhibited in the California current system (e.g. Monterey Bay and the Gulf of the Farallones) where areas of larval retention lead to distinct zooplankton assemblages over distances of just a few kilometres (52). In the particular case of Antofagasta Bay, the geometry of the coastline further aids in trapping surface water within the northern portion of the bay, creating an 'upwelling trap' (52) where the surface waters are retained for several days. This upwelling trap has been proposed as a key retention mechanism for the planktonic larvae of *P. praeputialis*, which remain pelagic for less than three hours (54). In contrast, other gregarious intertidal organisms found within Antofagasta Bay, such as the mussel *Perumytilus purpuratus*, have longer pelagic durations and extensive distributions along the west coast of South America (55). In addition to the short pelagic life-history stage of *P. praeputialis*, gametes and larvae can be retained by bio-foam produced by *P. praeputialis* adults (56), which further limits the dispersal of this species along the coastline outside of Antofagasta Bay (52, 56).

Improved predictive power of SDMs requires input from both native and introduced ranges (57), but this is not feasible for NIS that are either cryptic, or occupy limited areas in the introduced range and thus remain unsampled. Genetic identification of source populations has previously been used to inform climatic niche shifts (25), and our results showed the need for knowledge on cryptic genetic diversity to accurately predict potential range expansion. The results of SDMs using occurrence data from both the native and introduced ranges of the genomics-informed source lineage indicate that the species has the potential to inhabit a much more extensive area along the south-eastern Pacific than is currently observed (Fig. 2). However, using only occurrence data from the eastern Australian lineage (i.e. excluding the introduced range), Antofagasta Bay was considered unsuitable for *P. praeputialis* (Fig. 3H). Furthermore, when the information on the exact source of the invasion was not considered, but the species' complete native range or only its south-eastern Australian range (i.e. non-source native lineage) were used without occurrence data from the introduced range, the Chilean coast was not identified as suitable habitat (Fig. 3B).

Although our SDMs incorporated a wide variety of environmental variables, they did not incorporate other important variables such as microhabitat characteristics and biotic interactions (57). Interestingly, despite the lack of evidence of naturalisation or spread of *P. praeputialis* to sites outside of Antofagasta Bay, work conducted by Castilla *et al*. (58) has shown that juvenile *P. praeputialis* can be transplanted outside of Antofagasta, and continue to survive and grow. Additionally, these authors concluded that there is no obvious biotic rationale, such as predator intensification or competitive exclusion, for the lack of expansion from Antofagasta bay (58). This is in contrast to other NIS, which show limited introduced ranges due to predation (21), genetic bottlenecks (23), or population divergence (45). Successful transplantation of species outside of their ranges, overcoming niche constraints, is not rare (59) and shows that dispersal is an important explanatory variable of range limits. Therefore, it may be possible that Allee effects (60), the positive relationship between mean fitness and population density (i.e. mate limitation in broadcast spawners (61)), limit the ability of *P. praeputialis* to colonise naturally outside of Antofagasta Bay. Our genomic data suggested limited structuring within Antofagasta Bay, with sites closest to the mouth of the bay (C2 and C6, Table S1) genetically separated in the DAPC results using loci associated with SST (Fig. S5). The warmest site within Antofagasta (C1) is also the region with the largest individuals and highest biomass, whilst the site furthest to the south of the bay (C6) has the lowest population density (53, 62). This in combination with larval retention mechanisms suggests that, should the abiotic conditions present in Antofagasta Bay (upwelling trap, sub-surface water bringing in water rich in chlorophyll, bay retention mechanisms, etc.) change it has great potential for range expansion and subsequently alter biodiversity along the eastern Pacific coast from Chile to Ecuador.

Studying failed introductions is inherently difficult as there is often no footprint left of the introduction event (63). It would be unreasonable to assume that *P. praeputialis* exclusively attached itself to ships that travelled between Australia and Antofagasta Bay (for more information see text in *SI Appendix*). Despite this, we found no evidence behind an abiotic rationale limiting the colonisation of additional regions along the eastern Pacific coastline. This suggests that previously failed introductions may have been due to variables not included in our analyses, changes in environmental conditions since the 19th century, or through limited propagule pressure or opportunity (64). The jump between introduction to a limited geographic area and widespread invasiveness has previously been linked to increased residence time (65). Indeed, time since invasion has been reported as the best predictor of range sizes in marine invertebrates (66), presumably enabling species to overcome the characteristic lag period of biological invasions (67). Whilst *P. praeputialis* may be an exception to the lag phase rule of biological invasions, as there has not been expansion along the coast in more than 100 years, we nonetheless urge caution against complacency when monitoring NIS with restricted distributions. Recreational boating with poorly maintained hulls or sea-chests could spread adult individuals around, providing an opportunity for *P. praeputialis* to escape the unique oceanographic conditions of Antofagasta Bay.

In summary, we showed how the study of species with restricted introduced ranges provides key insights into the mechanisms shaping range sizes. We found that the highly invasive *P. praeputialis* has great potential for spread beyond its point of introduction and thus threatens coastal biodiversity along a large stretch of South American coastline. Thus, further monitoring of this and others NIS with restricted introduced ranges is strongly recommended. Finally, future studies should consider both habitat suitability and genomic data to holistically assess the potential for spread of NIS.

Materials and Methods

Study species and Pacific-wide distribution

Pyura praeputialis (Heller, 1887), previously known as *P. stolonifera* or *P. stolonifera praeputialis* (e.g. 68), is a solitary ascidian (Tunicata, Ascidiacea) that forms densely packed sessile aggregates that can monopolise the intertidal and subtidal along extensive stretches of coastline. Along its native Australian habitat, *P. praeputialis* ranges from Cape Otway in Victoria (south-east coast) to southern Queensland (north-east coast) (Fig. 1), with a biogeographic barrier at Wilson's Promontory separating two genetically-differentiated populations of the species (35, 36). *Pyura praeputialis* is also present in South America as a single, isolated population in Antofagasta Bay, northern Chile (Fig. 1) (35, 37). Here, *P. praeputialis* achieves ecological dominance and the highest biomass per unit area ever reported for any intertidal species (69), reducing survival of native species by growing over them (58). It has been proposed that *P. praeputialis* was transported to Antofagasta Bay by shipping in the mid-late 19th century (62, 70) during the onset of the nitrate trade between Chile and Australia (71, 72). Accordingly, genetic studies show a close relation between Chilean and eastern Australian populations (35, 37).

Field sampling

A total of 190 individuals of *P. praeputialis* were collected along the coastline of Antofagasta Bay and from several locations throughout the species' Australian range (Fig. 1, Table S1). Details on tissue dissection are found in *SI Appendix*.

Genotyping-by-sequencing (GBS) libraries were prepared at the University of Wisconsin Biotechnology Centre following Elshire *et al.* (73) using the *ApeKI* restriction enzyme. The GBS libraries were subsequently sequenced on an Illumina NovaSeq6000 sequencer. The GBS reads were assembled using ipyrad v. 0.7.30 (74) with parameters recommended for paired-end GBS data (http://ipyrad.readthedocs.io/). As no published genome of any species for the family Pyuridae is available, we used *de novo* assembly methods, as described in the *SI Appendix*.

These data were used to create two single-nucleotide polymorphism (SNP) datasets. First, a dataset of candidate loci under environmental selection was generated using two GEA analyses, bayenv2 (75) and RDA (76). The GEA methods were used to identify putative adaptive SNPs based on hypothesised associations between genotype and thirteen environmental variables related to temperature, salinity, dissolved oxygen, and pH. These environmental variables were chosen as they have been shown to influence distributions of ascidians (77, 78). Details on the parameters used for the analyses are in *SI Appendix*. Secondly, a dataset of putatively neutral loci was generated using loci not recovered by either GEA analysis.

Comparative population genomics

For each sampling site, population genomic statistics for the neutral dataset were calculated using the R package "diveRsity" v.1.9.0 (79). These included observed heterozygosity (H_O),expected heterozygosity (H_E) and inbreeding coefficient (F_{IS}) (80). Significance in F_{IS} values were inferred using bootstrapping over 10,000 permutations using the neutral dataset.

Population structure was assessed using two methods that assign individuals to pre-defined clusters (ADMIXTURE and DAPC), and by estimating F_{ST} for pairs of sites using both the neutral and candidate datasets. ADMIXTURE v.1.3 (81) was used to estimate the likelihood that an individual comes from one of a pre-defined number of putative sample populations (*K*), and a DAPC (82) was performed using the R package “adegenet” v.2.1.1. We ran the DAPC with *a priori* knowledge of individual populations. Pairwise population genetic differentiation was examined by calculating F_{ST} values following Weir & Cockerham (80) in the R package “hierfstat” v.0.04-30 (83), with significance assessed by running 10,000 permutations after correcting for multiple comparisons using Benjamini-Yekutieli false discovery rate correction (84).

In order to reconstruct the invasion history of *P. praeputialis*, ABC analyses were conducted using DIYABC v.2.1.0 (85). Two sets of scenarios were tested. The aim of the first was to infer the colonisation history of *P. praeputialis* from Australia to South America, and the second was used to estimate the effective population size of individuals that founded the population in Antofagasta Bay (see *SI Appendix* for a detailed explanation of these scenarios and other methodological considerations).

Species distribution modelling

To determine the extent of suitable habitat for *P. praeputialis* across both sides of the Pacific, we used the maximum entropy method implemented in Maxent v.3.4.1 (86). We gathered spatial records by combining occurrence data from the Global Biodiversity Information Facility (www.gbif.com), the Ocean Biodiversity Information System (www.obis.org), and sample sites from the present study plus three previous studies (35, 36, 87). We extracted 22 ecologically relevant environmental datasets from Bio-ORACLE (88) and MARSPEC (89). To account for redundancy and the effects of collinearity amongst variables, we removed environmental variables that were highly correlated ($|r| > 0.7$) (90), retaining a total of eight variables (Table S6). We produced models using all occurrence data and also models using only the occurrence points relevant to the invasion (i.e. genomics-informed). We used the R package ENMeval to evaluate model predictive ability and avoid overfitting, a frequently overlooked aspect of SDMs (91). Additional details are given in *SI Appendix*.

Acknowledgments

We thank Mark de Bruyn, Rhiarn Hoban and Katie Dafforn for their help in providing samples of *Pyura praeputialis* muscle tissue from Australia. MR received funds from ASSEMBLE (an EU FP7 research infrastructure initiative comprising a network of marine research stations) to visit the Estación Costera de Investigaciones Marinas de Las Cruces and the Pontificia Universidad Católica de Chile. PRT was supported by the University of Johannesburg (FRC/URC grant), and LBB by the Australian Research Council. JCC acknowledges financial support from the Nucleo Milenio en Conservación Marina. JCC sincerely acknowledges long-term financial support (PUC-Escondida Project) from Minera Escondida Ltda, the last 10 years support from the Academic Chair “Arauco-UC”, and the important and constant assistance and advise from Dr Patricio Manríquez. CDM received funding from the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation (Grant number 64801). The authors thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing advice on their sequencing facilities and services.

References

1. J. H. Brown, G. C. Stevens, D. M. Kaufman, The Geographic Range: Size, shape, boundaries, and internal structure. *Annu. Rev. Ecol. Evol. Syst.* **27**, 597–623 (1996).
2. A. Hastings, *et al.*, The spatial spread of invasions: new developments in theory and evidence. *Ecol. Lett.* **8**, 91–101 (2005).
3. G. E. Hutchinson, Concluding Remarks. *Cold Spring Harb. Symp. Quant. Biol.* **22**, 415–427 (1957).

4. L. Gallien, T. Münkemüller, C. H. Albert, I. Boulangeat, W. Thuiller, Predicting potential distributions of invasive species: where to go from here? *Divers. Distrib.* **16**, 331–342 (2010).
5. A. F. A. de Andrade, S. J. E. Velazco, P. De Marco, Niche mismatches can impair our ability to predict potential invasions. *Biol. Invasions* **21**, 3135–3150 (2019).
6. R. Massatti, L. L. Knowles, Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Mol. Ecol.* **25**, 3974–3986 (2016).
7. E. W. Malone, *et al.*, Which species, how many, and from where: Integrating habitat suitability, population genomics, and abundance estimates into species reintroduction planning. *Glob. Change Biol.* **24**, 3729–3748 (2018).
8. K. E. Atkins, J. M. J. Travis, Local adaptation and the evolution of species' ranges under climate change. *J. Theor. Biol.* **266**, 449–457 (2010).
9. P. Calosi, D. T. Bilton, J. I. Spicer, S. C. Votier, A. Atfield, What determines a species' geographical range? Thermal biology and latitudinal range size relationships in European diving beetles (Coleoptera: Dytiscidae). *J. Anim. Ecol.* **79**, 194–204 (2010).
10. J. P. Sexton, S. B. Hangartner, A. A. Hoffmann, Genetic isolation by environment or distance: which pattern of gene flow Is most common? *Evolution* **68**, 1–15 (2014).
11. R. Tingley, M. Vallinoto, F. Sequeira, M. R. Kearney, Realized niche shift during a global biological invasion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10233–10238 (2014).
12. N. Battini, N. Farías, C. B. Giachetti, E. Schwindt, A. Bortolus, Staying ahead of invaders: using species distribution modeling to predict alien species' potential niche shifts. *Mar. Ecol. Prog. Ser.* **612**, 127–140 (2019).
13. L. Sales, L. Culot, M. M. Pires, Climate niche mismatch and the collapse of primate seed dispersal services in the Amazon. *Biol. Conserv.* **247**, 108628 (2020).
14. N. Bax, A. Williamson, M. Aguero, E. Gonzalez, W. Geeves, Marine invasive alien species: a threat to global biodiversity. *Mar. Policy* **27**, 313–323 (2003).
15. J. L. Molnar, R. L. Gamboa, C. Revenga, M. D. Spalding, Assessing the global threat of invasive species to marine biodiversity. *Front. Ecol. Environ.* **6**, 485–492 (2008).
16. E. V. Moran, J. M. Alexander, Evolutionary responses to global change: lessons from invasive species. *Ecol. Lett.* **17**, 637–649 (2014).
17. D. M. Richardson, *et al.*, Naturalization and invasion of alien plants: concepts and definitions. *Divers. Distrib.* **6**, 93–107 (2000).
18. M. F. Richardson, C. D. H. Sherman, R. Lee, N. Bott, A. J. Hirst, Multiple dispersal vectors drive range expansion in an invasive marine species. *Mol. Ecol.* **25**, 5001–5014 (2016).
19. T. M. Blackburn, *et al.*, A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339 (2011).
20. J. T. Carlton, J. K. Thompson, L. E. Schemel, F. H. Nichols, Remarkable invasion of San Francisco Bay (California, USA) by the Asian clam *Potamocorbula amurensis*. I. Introduction and dispersal. *Mar. Ecol. Prog. Ser.* **66**, 81–94 (1990).
21. B. C. Harvey, J. L. White, R. J. Nakamoto, An emergent multiple predator effect may enhance biotic resistance in a stream fish assemblage. *Ecology* **85**, 127–133 (2004).
22. M. Rius, E. E. Potter, J. D. Aguirre, J. J. Stachowicz, Mechanisms of biotic resistance across complex life cycles. *J. Anim. Ecol.* **83**, 296–305 (2014).
23. A. P. Kinziger, R. J. Nakamoto, E. C. Anderson, B. C. Harvey, Small founding number and low genetic diversity in an introduced species exhibiting limited invasion success (speckled dace, *Rhinichthys osculus*). *Ecol. Evol.* **1**, 73–84 (2011).
24. M. Szűcs, *et al.*, Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13501–13506 (2017).
25. E. E. Sotka, *et al.*, Combining niche shift and population genetic analyses predicts rapid phenotypic evolution during invasion. *Evol. Appl.* **11**, 781–793 (2018).
26. J. Elith, J. R. Leathwick, Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* **40**, 677–697 (2009).
27. V. Srivastava, V. Lafond, V. C. Griess, Species distribution models (SDM): applications, benefits and challenges in invasive species management. *CAB Reviews* **14**, 1–13 (2019).

28. A. Guisan, B. Petitpierre, O. Broennimann, C. Daehler, C. Kueffer, Unifying niche shift studies: insights from biological invasions. *Trends Ecol. Evol.* **29**, 260–269 (2014).
29. J. A. Grummer, *et al.*, Aquatic landscape genomics and environmental effects on genetic variation. *Trends Ecol. Evol.* **34**, 641–654 (2019).
30. M. Rius, S. Bourne, H. G. Hornsby, M. A. Chapman, Applications of next-generation sequencing to the study of biological invasions. *Curr. Zool.* **61**, 488–504 (2015).
31. F. Viard, P. David, J. A. Darling, Marine invasions enter the genomic era: three lessons from the past, and the way forward. *Curr. Zool.* **62**, 629–642 (2016).
32. A. Marcer, B. Méndez-Vigo, C. Alonso-Blanco, F. X. Picó, Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecol Evol* **6**, 2084–2097 (2016).
33. O. Razgour, *et al.*, Scale-dependent effects of landscape variables on gene flow and population structure in bats. *Divers. Distrib.* **20**, 1173–1185 (2014).
34. D. H. Ikeda, *et al.*, Genetically informed ecological niche models improve climate change predictions. *Glob. Change Biol.* **23**, 164–176 (2017).
35. P. R. Teske, *et al.*, “Nested” cryptic diversity in a widespread marine ecosystem engineer: a challenge for detecting biological invasions. *BMC Evol. Biol.* **11**, 176 (2011).
36. M. Rius, P. R. Teske, Cryptic diversity in coastal Australasia: a morphological and mitonuclear genetic analysis of habitat-forming sibling species. *Zool. J. Linn. Soc.* **168**, 597–611 (2013).
37. J. C. Castilla, A. G. Collins, C. P. Myer, R. Guiñez, D. R. Lindberg, Recent introduction of the dominant tunicate, *Pyura praeputialis* (Urochordata, Pyuridae) to Antofagasta, Chile. *Mol. Ecol.* **11**, 1579–1584 (2002).
38. P. R. Teske, *et al.*, Thermal selection as a driver of marine ecological speciation. *Proc. R. Soc. B* **286**, 20182023 (2019).
39. P. R. Teske, *et al.*, Oceanic dispersal barriers, adaptation and larval retention: an interdisciplinary assessment of potential factors maintaining a phylogeographic break between sister lineages of an African prawn. *BMC Evol. Biol.* **8**, 341 (2008).
40. I. Papadopoulos, P. R. Teske, Larval development reflects biogeography in two formerly synonymised southern African coastal crabs. *Afr. J. Aquat. Sci.* **39**, 347–350 (2014).
41. J. T. Carlton, “The scale and ecological consequences of biological invasions in the world’s oceans” in *Invasive Species and Biodiversity Management*, O. T. Sandlund, P. J. Schei, Å. Viken, Eds. (Kluwer Academic Publishers, 1999).
42. C. K. Tepolt, Adaptation in marine invasion: a genetic perspective. *Biol. Invasions* **17**, 887–903 (2015).
43. S. Lavergne, J. Molofsky, Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3883–3888 (2007).
44. P. J. Prentis, J. R. U. Wilson, E. E. Dormontt, D. M. Richardson, A. J. Lowe, Adaptive evolution in invasive species. *Trends Plant Sci.* **13**, 288–294 (2008).
45. K. W. Wellband, H. Pettitt-Wade, A. T. Fisk, D. D. Heath, Differential invasion success in aquatic invasive species: the role of within- and among-population genetic diversity. *Biol. Invasions* **19**, 2609–2621 (2017).
46. J. Roman, Diluting the founder effect: cryptic invasions expand a marine invader’s range. *Proc. R. Soc. B* **273**, 2453–2459 (2006).
47. S. E. Swearer, J. E. Caselle, D. W. Lea, R. R. Warner, Larval retention and recruitment in an island population of a coral-reef fish. *Nature* **402**, 799–802 (1999).
48. G. Gerlach, J. Atema, M. J. Kingsford, K. P. Black, V. Miller-Sims, Smelling home can prevent dispersal of reef fish larvae. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 858–863 (2007).
49. P. R. Teske, J. Sandoval-Castillo, E. van Sebille, J. Waters, L. B. Beheregaray, Oceanography promotes self-recruitment in a planktonic larval disperser. *Scientific Reports* **6**, 34205 (2016).
50. W. M. Graham, J. L. Largier, Upwelling shadows as nearshore retention sites: the example of northern Monterey Bay. *Cont. Shelf Res.* **17**, 509–532 (1997).

51. R. Escribano, P. Hidalgo, Spatial distribution of copepods in the north of the Humboldt Current region off Chile during coastal upwelling. *J. Mar. Biol. Assoc. U.K.* **80**, 283–290 (2000).
52. J. C. Castilla, N. A. Lagos, R. Guiñez, J. L. Largier, "Embayments and nearshore retention of plankton: the Antofagasta Bay and other examples" in *The Oceanography and Ecology of the Nearshore and Bays in Chile*, J. C. Castilla, J. L. Largier, Eds. (Ediciones Universidad Católica de Chile, Santiago, 2002), pp. 179–203.
53. A. Piñones, J. C. Castilla, R. Guiñez, J. L. Largier, Nearshore surface temperatures in Antofagasta Bay (Chile) and adjacent upwelling centers. *Cienc. Mar.* **33**, 37–48 (2007).
54. M. Clarke, V. Ortiz, J. C. Castilla, Does early development of the Chilean tunicate *Pyura praeputialis* (Heller, 1878) explain the restricted distribution of the species? *Bull. Mar. Sci.* **65**, 745–754 (1999).
55. R. Guiñez, J. C. Castilla, A tridimensional self-thinning model for multilayered intertidal mussels. *Am. Nat.* **154**, 341–357 (1999).
56. J. C. Castilla, *et al.*, Bio-foam enhances larval retention in a free-spawning marine tunicate. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18120–18122 (2007).
57. H. Verbruggen, *et al.*, Improving transferability of introduced species' distribution models: new tools to forecast the spread of a highly invasive seaweed. *PLOS ONE* **8**, e68337 (2013).
58. J. C. Castilla, R. Guinez, A. U. Caro, V. Ortiz, Invasion of a rocky intertidal shore by the tunicate *Pyura praeputialis* in the Bay of Antofagasta, Chile. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8517–8524 (2004).
59. A. L. Hargreaves, K. E. Samis, C. G. Eckert, Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. *Am. Nat.* **183**, 157–173 (2014).
60. P. A. Stephens, W. J. Sutherland, R. P. Freckleton, What Is the Allee effect? *Oikos* **87**, 185–190 (1999).
61. J. Gascoigne, L. Berec, S. Gregory, F. Courchamp, Dangerously few liaisons: a review of mate-finding Allee effects. *Popul. Ecol.* **51**, 355–372 (2009).
62. J. C. Castilla, R. Guiñez, Disjoint geographical distribution of intertidal and nearshore benthic invertebrates in the Southern Hemisphere. *Rev. Chil. Hist. Nat.* **73**, 585–603 (2000).
63. R. D. Zenni, M. A. Nuñez, The elephant in the room: the role of failed invasions in understanding invasion biology. *Oikos* **122**, 801–815 (2013).
64. A. V. Suarez, D. A. Holway, P. S. Ward, The role of opportunity in the unintentional introduction of nonnative ants. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17032–17035 (2005).
65. M. L. Phillips, B. R. Murray, M. R. Leishman, R. Ingram, The naturalization to invasion transition: Are there introduction-history correlates of invasiveness in exotic plants of Australia? *Austral Ecol.* **35**, 695–703 (2010).
66. J. E. Byers, *et al.*, Invasion Expansion: Time since introduction best predicts global ranges of marine invaders. *Sci. Rep.* **5**, 12436 (2015).
67. J. A. Crooks, M. E. Soulé, Lag times in population explosions of invasive species: causes and implications. *Invasive species and biodiversity management. Based on papers presented at the Norway/United Nations (UN) Conference on Alien Species, 2nd Trondheim Conference on Biodiversity, Trondheim, Norway, 1-5 July 1996.*, 103–125 (1999).
68. P. Kott, The Australian Ascidiacea. Part 1, Phlebobranchia and Stolidobranchia. *Mem. Queensl. Mus.* **23**, 1–439 (1985).
69. J. C. Castilla, R. Guiñez, J. L. Alvarado, C. Pacheco, M. Varas, Distribution, population structure, population biomass and morphological characteristics of the tunicate *Pyura stolonifera* in the Bay of Antofagasta, Chile. *Mar. Ecol.* **21**, 161–174 (2000).
70. M. Rius, P. R. Teske, A revision of the *Pyura stolonifera* species complex (Tunicata, Ascidiacea), with a description of a new species from Australia. *Zootaxa* **2754**, 27–40 (2011).
71. V. Maino, *Islas oceánicas chilenas y el desarrollo de las rutas de comunicación en el Pacífico suroriental* (Editorial Antártica S. A., 1985).
72. I. Arce, *Narraciones Históricas de Antofagasta* (Lama Industrial S.A., 1997).

73. R. J. Elshire, *et al.*, A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLOS ONE* **6**, e19379 (2011).
74. D. A. R. Eaton, I. Overcast, ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**, 2592–2594 (2020).
75. T. Günther, G. Coop, Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205–220 (2013).
76. B. R. Forester, J. R. Lasky, H. H. Wagner, D. L. Urban, Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.* **27**, 2215–2233 (2018).
77. N. Shenkar, B. J. Swalla, Global diversity of Ascidiacea. *PLOS ONE* **6**, e20657 (2011).
78. D. M. Lins, P. de Marco, A. F. A. Andrade, R. M. Rocha, Predicting global ascidian invasions. *Divers. Distrib.* **24**, 692–704 (2018).
79. K. Keenan, P. McGinnity, T. F. Cross, W. W. Crozier, P. A. Prodöhl, diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**, 782–788 (2013).
80. B. S. Weir, C. C. Cockerham, Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
81. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
82. T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94 (2010).
83. J. Goudet, hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
84. D. Yekutieli, Y. Benjamini, Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Stat. Plan. Inference* **82**, 171–196 (1999).
85. J.-M. Cornuet, *et al.*, 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**, 1187–1189 (2014).
86. S. J. Phillips, R. P. Anderson, M. Dudík, R. E. Schapire, M. E. Blair, Opening the black box: an open-source release of Maxent. *Ecography* **40**, 887–893 (2017).
87. A. R. Davis, M. Becerro, X. Turon, Living on the edge: Early life history phases as determinants of distribution in *Pyura praeputialis* (Heller, 1878), a rocky shore ecosystem engineer. *Mar. Environ. Res.* **142**, 40–47 (2018).
88. J. Assis, *et al.*, Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Glob. Ecol. Biogeogr.* **27**, 277–284 (2018).
89. E. J. Sbrocco, P. H. Barber, MARSPEC: ocean climate layers for marine spatial ecology. *Ecology* **94**, 979–979 (2013).
90. C. F. Dormann, *et al.*, Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
91. N. S. Morales, I. C. Fernández, V. Baca-González, MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ* **5** (2017).

Figures

Fig. 1. Sample sites along the coast of (A) Australia and (B) within Antofagasta on the Chilean coastline. Barplots in (A) and (B) represent the results of the ADMIXTURE clustering analysis inferred with neutral loci at $K = 2$ (see full details in *SI Appendix*). (C) The most likely scenario, as revealed using Approximate Bayesian Computation (see full details in *SI Appendix*), depicts the invasion route that *P. praeputialis* followed from Australia to Antofagasta Bay.

Fig. 2. Maps illustrating the results of genomics-informed Maxent modelling. The figures show habitat suitability for *Pyura praeputialis* across: (A) its native range, (B) the coastline adjacent to Antofagasta Bay (inset, Antofagasta Bay indicated with an arrow) and the western coastline of South America. This model includes distance to shore as a variable, explaining the observed narrow regions of suitable habitat. For full list of variables used in model creation see Table S3. The scale bar represents Maxent's logistic output representing habitat suitability, with "yellow" indicating high habitat suitability. Crosses in (A) represent the Australian sample locations of *P. praeputialis*. Note that the Maxent's logistic output only considers the genomic-informed relevant sites (i.e. sites from Chile and the eastern coastline of Australia, see details in main text and *SI Appendix*).

Fig. 3. Maps illustrating habitat suitability for *Pyura praeputialis* when genomic data are not incorporated into the model. The figures show the output of Maxent modelling using occurrence data from: both native lineages (A-C), only the south-eastern Australia lineage (D-F), and only the eastern Australia lineage (G-I). Maps depict: native range (A,D,G); adjacent coastlines of Antofagasta Bay (bay represented by arrow) (B,E,H); and extensive coastlines along the western coast of South America continent (C,F,I). The scale bar represents Maxent's logistic output on habitat suitability (see details in *SI Appendix*).