MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 4145–4158 doi: 10.1111/mec.13294

Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader

C. K. TEPOLT and S. R. PALUMBI
Hopkins Marine Station of Stanford University, 120 Ocean View Boulevard, Pacific Grove, CA 93950, USA

Abstract

Species invasions cause significant ecological and economic damage, and genetic information is important to understanding and managing invasive species. In the ocean, many invasive species have high dispersal and gene flow, lowering the discriminatory power of traditional genetic approaches. High-throughput sequencing holds tremendous promise for increasing resolution and illuminating the relative contributions of selection and drift in marine invasion, but has not yet been used to compare the diversity and dynamics of a high-dispersal invader in its native and invaded ranges. We test a transcriptome-based approach in the European green crab (*Carcinus maenas*), a widespread invasive species with high gene flow and a well-known invasion history, in two native and five invasive populations. A panel of 10 809 transcriptome-derived nuclear SNPs identified significant population structure among highly bottlenecked invasive populations that were previously undifferentiated with traditional markers. Comparing the full data set and a subset of 9246 putatively neutral SNPs strongly suggested that non-neutral processes are the primary driver of population structure within the species' native range, while neutral processes appear to dominate in the invaded range. Non-neutral native range structure coincides with significant differences in intraspecific thermal tolerance, suggesting temperature as a potential selective agent. These results underline the importance of adaptation in shaping intraspecific differences even in high geneflow marine invasive species. They also demonstrate that high-throughput approaches have broad utility in determining neutral structure in recent invasions of such species. Together, neutral and non-neutral data derived from high-throughput approaches may increase the understanding of invasion dynamics in high-dispersal species.

Keywords: adaptation, invasive species, natural selection and contemporary evolution, neutral processes, population genetics – empirical, transcriptomics

Received 6 March 2015; revision received 21 June 2015; accepted 24 June 2015

Introduction

Modern biological invasions present a significant threat to ecosystem function and are recognized as a major component of global ecological change (Ricciardi 2007). Invasive species deliver a heavy economic burden in ecosystems – estimated at upwards of $120 billion annually in the United States alone (Pimentel *et al.* 2005) – and can transform ecosystems. Because of these impacts, managing and preventing species invasions is a conservation priority, and genetic approaches play a growing role in this process (Lodge *et al.* 2006; Rollins *et al.* 2006). Simple demographic data on invasions – such as their source and levels of genetic diversity – have traditionally been the goals of genetic research (Wilson *et al.* 2009). More recently, attention has turned to the role of adaptation in species invasions (Prentis 2008). Local adaptation within a species' native range has been suggested to increase invasion risk, with certain source populations 'pre-adapted' for success in a specific environment, while other native populations

Correspondence: Carolyn K. Tepolt, Smithsonian Environmental Research Center, 647 Contees Wharf Road, Edgewater, MD 21037, USA, Fax: 443 482 2380;
E-mail: carolyn.tepolt@gmail.com

may be less well suited to their new habitats (Fausch 2008). There is also evidence that invasive species may evolve rapidly in their new ranges (Huey & Pascual 2009; Hodgins *et al.* 2012) and that an ability to quickly adapt to local conditions may facilitate invasion (Vandepitte *et al.* 2014).

Marine invasive species have posed a particular challenge for traditional population genetics approaches. Many of these species have very high dispersal potential, which both facilitates invasion through transport in currents or in ships' ballast water (Carlton & Geller 1993; Byers & Pringle 2006) and complicates the process of tracing invasions because of low population genetic differentiation due to high gene flow (Waples 1998; Palumbi & Pinsky 2013). A recent review suggests that the least structured marine invasive species seem to make the best invaders: low genetic differentiation in the native range is correlated with greater continuous geographic spread in the invasive range (Gaither *et al.* 2013). This, coupled with the short time frame and frequent bottlenecking of contemporary invasions, makes distinguishing invasive populations difficult and can hinder the determination of population structure and dynamics. However, the same traits that reduce resolution with traditional markers also make marine invasions of particular interest to adaptation research. With high dispersal and few small-scale barriers, it has been hypothesized that selection may be the primary contributor to genetic structure in some marine systems (Bierne *et al.* 2003; Moura *et al.* 2014).

Over the past decade, high-throughput sequencing approaches have revolutionized the study of population genetics in nonmodel organisms, in part by providing a much finer resolution view of the genome than traditional sequencing (Luikart *et al.* 2003; Ekblom & Galindo 2011). This higher genomic coverage has frequently been leveraged in scans for selection to identify potentially adaptive differences between populations (e.g. Namroud *et al.* 2008; De Wit & Palumbi 2012; White *et al.* 2013). However, potentially selected loci reflect a different demographic and evolutionary process than differentiation of many loci by drift and migration (Wright 1978). A focus on F_{ST} outliers – loci that are more diverged between populations than expected from neutral processes – does not by itself reveal patterns of gene flow or dispersal because such loci are also influenced by selection (Sotka & Palumbi 2006). As a result, a focus on demographic separation of populations requires information on differentiation of neutral loci. Distinguishing loci likely to be under selection from a large panel of markers allows a test of whether the putatively neutral remaining loci show similar population structure (Hess *et al.* 2013; Moore *et al.* 2014), and reveals the relative contributions of neutral and adaptive processes to differentiation between populations (Lemay & Russello 2015).

Here, we use high-throughput sequencing to examine genetic structure and adaptation in the globally invasive European green crab, *Carcinus maenas*, examining populations spanning the native and invasive ranges. *C. maenas* is native to the northeast Atlantic from Morocco to Iceland and has successfully invaded every continent except Antarctica (Carlton & Cohen 2003; Darling *et al.* 2008; Blakeslee *et al.* 2010). The first documented introduction was to the East Coast of North America *circa* 1817, spreading south to the Chesapeake Bay and north to Halifax, Nova Scotia, by 1951 (Carlton & Cohen 2003). Starting in the 1980s, the species' range expanded to encompass the Canadian Maritime provinces and Newfoundland, concurrent with a second introduction from northern Europe (Roman 2006; Blakeslee *et al.* 2010). Green crabs are also invasive on the West Coast of North America, where they were first detected in San Francisco Bay in 1989, subsequently spreading north to Vancouver Island in <10 years (Cohen *et al.* 1995; Behrens Yamada & Hunt 2000). While *C. maenas* is a global invader, this study focuses on the native and North American invasive ranges, which represent many of the complex dynamics typical of invasion biology: multiple introductions, serial invasion and bottlenecking and a range of population ages.

Global population genetic analysis has been previously carried out on *C. maenas* using mitochondrial COI sequencing and microsatellites, leading to an overview of the structure and invasion pathways of the species (Roman & Palumbi 2004; Roman 2006; Darling *et al.* 2008; Tepolt *et al.* 2009; Blakeslee *et al.* 2010). However, traditional sequencing approaches have raised or left open important questions for the species. First, in much of the North American invasive range, bottlenecking has caused genetic diversity to drop to the point where prior studies were unable to distinguish between populations (Darling *et al.* 2008; Tepolt *et al.* 2009). Second, the role of selection in these invasions has been suggested by several lines of evidence, but has not been rigorously tested. For example, in the native range, mitochondrial COI shows distinct north–south differentiation, while putatively neutral microsatellites demonstrated very little structure within mainland Europe (Roman & Palumbi 2004; Darling *et al.* 2008). In addition, a physiological study of green crabs shows a pattern of local adaptation in thermal tolerance among populations, particularly in the native range (Tepolt & Somero 2014). Taken together, these data suggest that different introductions from differentially adapted native populations may have contributed to the explosive spread of the species on the East Coast, a fascinating proposition that cannot be addressed by the existing genetic data.

Here, we sequenced the cardiac transcriptomes of 12 individuals from each of seven populations in the native and invasive ranges and identified a panel of 10 809 high-quality SNPs for population genetic analysis. To explore the neutral components and possible impact of selection on shaping this structure, we extracted a subset of 9246 putatively neutral SNPs to contrast with the full SNP set. We used these data in two ways. First, we used the full SNP panel to explore patterns in population structure and genetic diversity, with a focus on increasing resolution in the genetically depauperate invasive range. Second, we compared the full and neutral SNP panels to investigate the relative contribution of neutral and non-neutral processes in shaping differentiation within and between the native and invasive ranges of the species. We show that these analyses illuminate strong patterns of population differentiation despite the recent invasion of these populations, and demonstrate that the balance between neutral and non-neutral processes appears to differ between the native and invasive ranges.

Materials and methods

Sample collection

European green crabs were sampled from seven sites across the native and invasive range: Skitnepollen, Askøy, Norway, and Seixal, Portugal, in Europe; North Harbour, NL, Canada, Walpole, ME, USA, and Tuckerton, NJ, USA, on the East Coast of North America; and Toquart Bay, BC, Canada, and Seadrift Lagoon, CA, USA, on the West Coast of North America (Fig. 1; Table 1). At each site, 12 adult crabs were collected by trapping from shore.

At each site, six animals were acclimated to 5 °C for 4 weeks and six were acclimated to 25 °C for 3 weeks in the laboratory before tissues were collected for sequencing. These parameters were chosen for a complementary study of cardiac thermal physiology in *Carcinus maenas* and are described in more detail in Tepolt & Somero (2014). After acclimation, animals were sacrificed by severing the ventral nerve cord and whole hearts were placed in an RNA stabilizing reagent. Hearts were stored at 4 °C for 1–5 days to allow the reagent to penetrate the tissues, moved to −20 °C for 1–3 months and then archived at −80 °C. Like acclimation parameters, the use of cardiac tissue was chosen to complement simultaneous work on cardiac thermal physiology (Tepolt & Somero 2014).

Extraction and sequencing

RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) following the manufacturer's suggested protocol (Simms *et al.* 1993). For each extraction, a piece of cardiac tissue approximately 1 × 1 mm was excised using a razor blade, and the tissue was broken up either manually or by flash-freezing in liquid nitrogen and powdering. At the end of the extraction, RNA pellets were eluted in 25 μL RNase-free ultrapure water and frozen at −80 °C for 1–7 days prior to cDNA preparation.

Total RNA was quantified using the RNA BR assay on a Qubit 2.0 fluorometer (Invitrogen), and approximately 4 μg of RNA was used to construct a cDNA

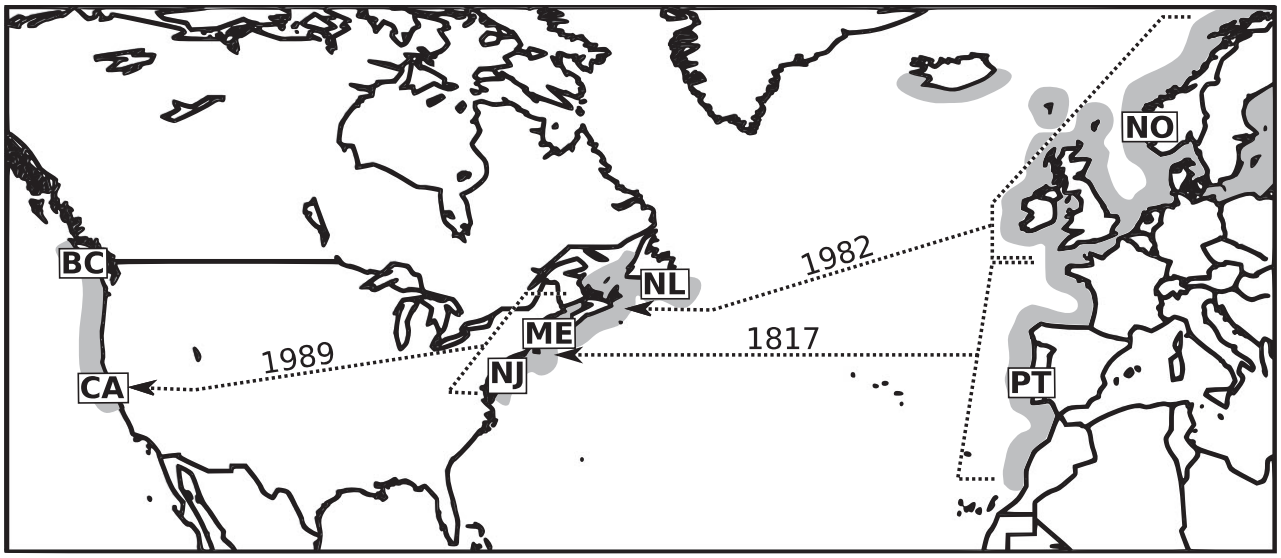

Fig. 1 Map showing collection locations (boxed codes per Table 1) and invasion history. Current range shown in grey. Approximate source populations are shown in dashed brackets, with arrows leading to initial invasion sites. Dates indicate the first record of each invasion.

Table 1 Details of collection locations

Site	Code	*N*	Lat	Long	Date
Toquart Bay, BC, Canada	BC	12	49.0203	−125.3586	Aug 2010
Seadrift Lagoon, CA, USA	CA	12	37.9079	−122.6726	Apr 2011
North Harbour, NL, Canada	NL	12	47.8459	−54.0969	Sep 2011
Walpole, ME, USA	ME	12	43.9351	−69.5809	Aug 2011
Tuckerton, NJ, USA	NJ	12	39.5095	−74.3248	Jul 2011
Skitnepollen, Askøy, Norway	NO	12	60.5061	5.0226	Aug 2012
Seixal, Portugal	PT	12	38.6455	−9.1036	Jul 2012

N, number of individuals; Lat, latitude; Long, longitude; Date, time of sampling.

library for each sample using the TruSeq RNA Sample Prep Kit from Illumina (San Diego, CA, USA). No in-line control reagents were used, and samples were tagged with one of 12 different adapter index sequences from Illumina for multiplexing.

Libraries were then sent to the University of Utah's Microarray and Genomics Core Facility, where they were quantified using a Bioanalyzer (Agilent, Santa Clara, CA, USA) and pooled into seven lanes of 12 samples each. All lanes contained a mix of samples from 5 to 7 sites. After pooling, samples were sequenced on an Illumina HiSeq 2000 sequencer. One lane was sequenced using 101-bp paired-end reads to aid with transcriptome assembly, while the remaining six lanes were sequenced using 50-bp single-end reads.

Sequence processing and transcriptome assembly

Processing of raw reads was performed using the SFG pipeline and scripts, available at http://sfg.stanford.edu (De Wit *et al.* 2012). Reads were initially processed using the FASTX toolkit v0.0.13; after processing, reads were imported into CLC GENOMICS WORKBENCH v6.0.4 (CLC bio, Cambridge, MA, USA) and used to assemble a *de novo* transcriptome. The parameters for this assembly included a mismatch cost of 1, insertion and deletion costs of 2, similarity fraction of 0.9, length fraction of 0.5 and minimum contig length of 200 bp.

Contigs were annotated with UniProt's Swiss-Prot and TrEMBL databases and GenBank's nr protein database using the BLASTX algorithm and NCBI's default parameters. Matches were considered significant at e-values of $\leq 1 \times 10^{-3}$. BLASTing was performed using the University of Oslo's Bioportal (https://www.bioportal.uio.no/).

MEGAN v4.70.4 was used to identify potential noncrab sequences (Huson *et al.* 2011); contigs were removed if they were identified as deriving from Bacteria, Archaea, Fungi or Plantae. Because of the dearth of genomic resources for this species, contigs that could not be identified taxonomically were retained in the assembly. Sequences identified as mitochondrial were also removed from the nuclear assembly used for SNP detection.

Mapping and SNP detection

A modified and updated version of the SFG protocol and scripts were used to map reads and detect SNPs. Initially, all reads were mapped back onto the *de novo* assembly using the BURROWS-WHEELER ALIGNER (BWA) v0.7.4 (http://bio-bwa.sourceforge.net/) and default parameters, except for $n = 0.004$ and $k = 5$ (Li & Durbin 2009). The GENOME ANALYSIS TOOLKIT v2.4 was used to detect high-quality nuclear SNPs as described in the SFG protocol (De Wit *et al.* 2012), updated to reflect the GATK v2.4 syntax and recommendations (McKenna *et al.* 2010; DePristo *et al.* 2011).

High-quality SNPs detected by this pipeline (Phred ≥ 30) were further filtered with a custom python script. First, we removed SNPs with >2 alleles and SNPs for which the minor allele frequency was <1.5%. At the individual genotype level, we removed low-quality (Phred < 20) and low-coverage (<10×) genotypes, and heterozygous genotypes where the ratio of one allele to the other was >4:1. We retained only SNPs that passed all of these filters and had high-quality genotypes in ≥10 individuals per population. Finally, we filtered SNPs to remove likely paralogs by removing SNPs where half of populations were significantly out of Hardy–Weinberg equilibrium (HWE; $P < 0.01$).

Identification of neutral SNPs

In order to reduce our SNP panel to a neutral subset, we used an outlier detection approach with extremely loose criteria for outliers in order to identify and exclude any loci identified as potentially non-neutral. While this approach undoubtedly discarded some truly neutral loci (i.e. false positives), it minimized the retention of weakly selected loci which would bias the neutral SNP panel (i.e. false negatives). Potentially non-neutral SNPs were identified using the *fdist* method for outlier detection (Beaumont & Nichols 1996), as implemented in *LOSITAN* (Antao *et al.* 2008). We ran *LOSITAN* 10 separate times, performing 500 000 simulations for each run with the 'Neutral' mean F_{ST}' and 'Force mean F_{ST}' options. Any SNP having an outlier probability ≥0.9 ($P \leq 0.1$) in any of these 10 runs was flagged as

potentially under selection and removed from the neutral SNP panel, similar to the approach of Hess *et al.* (2013). These cut-offs were chosen to be particularly conservative in order to build a panel of truly neutral SNPs.

Genetic structure

Principal component analysis (PCA) was performed on both the full and the neutral SNP panels, using the EIGENSOFT package (Patterson *et al.* 2006). The full SNP panel was also analysed with STRUCTURE v2.3.4 (Pritchard *et al.* 2000; Hubisz *et al.* 2009), using a burn-in period of 50 000 replicates followed by 150 000 MCMC replicates. Optimal K was inferred using the approach of Evanno *et al.* (2005) and implemented in STRUCTURE HARVESTER v0.6.93 (Earl & vonHoldt 2011), with 10 replicates each at $K = 1$–6. CLUMPP v1.1.2 was used to create composite data from the 10 replicates for each K (Jakobsson & Rosenberg 2007), and results were visualized using DISTRUCT v1.1 (Rosenberg 2003). To pick up on more subtle structure, this process was repeated on two subsets of the data: Europe (NO, PT) and populations derived from the initial invasion to the East Coast of North America (CA, BC, NJ, ME).

Pairwise F_{ST} was calculated using ARLEQUIN v3.1.5.3 (Excoffier & Lischer 2010), using 10 000 permutations and a significance level of 0.05. This analysis was performed twice, once on the full data set and once on the neutral SNP panel. The difference in F_{ST} between these approaches was calculated for each population pair by dividing pairwise F_{ST} from the neutral data set by pairwise F_{ST} from the full data set, and multiplying by 100 to get a percentage.

Genetic diversity

For the full data set, observed and expected heterozygosity (as 1-Q_{intra} and 1-Q_{inter}) and F_{IS} were calculated for each population using GENEPOP v4.2.1 (Rousset 2008). The significance of F_{IS} was assessed using a score test for heterozygote excess implemented in GENEPOP. The proportion of all SNPs that were polymorphic in each population (Pn) was calculated in ARLEQUIN. Allelic richness and private allelic richness (Ar and pAr) across all SNPs were calculated in ADZE v1.0 (Szpiech *et al.* 2008). The significance of differences in Ar and pAr between locations was calculated from the mean and standard error for these values (from ADZE) and sample size ($N = 12$ in all cases) using unpaired t-tests implemented in GRAPHPAD QUICKCALCS online (http://graphpad.com/quickcalcs/ttest2/).

All plotting was carried out using the R package GGPLOT2 (Wickham 2009; R Core Team 2012).

Results

Sequencing, transcriptome assembly, and SNP genotyping

mRNA sequencing resulted in an average of 15.2 million (m) single-end sequenced reads (11.8 m–21.0 m) and 14.0 m paired reads (12.0 m–17.3 m) per individual, for a total of 1.43 billion reads. After trimming and clipping of low-quality bases, and removal of duplicate reads, 1.35 billion reads were used for transcriptome assembly and genotyping. The initial assembly comprised 117 189 contigs ranging from 160 to 21 325 bp with an N50 of 1016 bp. A total of 948 contigs were identified as noncrab contamination and were removed. The final working assembly consisted of 116 241 contigs, of which 32 913 (28%) were annotated. Thirty-nine of these contigs were identified as mitochondrial and were excluded from nuclear SNP detection analyses.

Initial SNP detection and screening resulted in 531 675 high-quality nuclear SNPs with Phred ≥30. We further filtered these SNPs to retain only those that were biallelic with a minor allele frequency of >1.5%, and which had high coverage (≥10×) and high-quality genotypes (Phred ≥20; allele ratio <4:1 for heterozygotes) for ten or more individuals per population; this screening resulted in 10 811 SNPs. Two of these were identified as potential paralogs due to being out of HWE in three or more populations and were also removed. The resulting full SNP panel comprised 10 809 nuclear SNPs in 1673 contigs, of which 1299 (78%) were annotated.

Neutral subset

The subset of neutral SNPs were identified by conservatively removing all SNPs that showed higher levels of F_{ST} than expected by chance using the FDIST method (Beaumont & Nichols 1996). We ran *LOSITAN* (Antao *et al.* 2008) 10 times and removed all SNPs that showed an outlier probability ≥0.9 ($P \leq 0.1$) in any of these 10 runs. Although these methods undoubtedly remove many neutral SNPs, these cut-offs were chosen to be particularly conservative to identify and remove as many non-neutral SNPs as possible. A total of 1563 of our SNPs fell inside the 10% threshold for non-neutrality and were removed to make a putatively neutral panel of 9246 SNPs.

Overall population structure

For the full SNP panel, all pairwise population comparisons showed significant genetic differentiation (Table 2), in contrast to previous studies. Within-region

Table 2 Pairwise F_{ST} (below diagonal) and significance (above diagonal) for all 10 809 SNPs

	NO	PT	NL	ME	NJ	BC	CA
NO	—	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001
PT	0.049	—	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001
NL	0.062	0.083	—	<0.00001	<0.00001	<0.00001	<0.00001
ME	0.100	0.107	0.039	—	0.02	<0.00001	<0.00001
NJ	0.100	0.093	0.043	0.003	—	<0.00001	<0.00001
BC	0.115	0.112	0.054	0.015	0.015	—	<0.00001
CA	0.134	0.133	0.072	0.031	0.032	0.021	—

Sites as in Table 1, Fig. 1.

differentiation was strongest in the native European range, with a pairwise F_{ST} between locations of 0.049. Differentiation in the East Coast was highest between Newfoundland and both other sites (F_{ST} with ME: 0.039; with NJ: 0.043) and lowest between Maine and New Jersey (F_{ST} = 0.003; P = 0.02). On the West Coast, locations in British Columbia and California were significantly differentiated (F_{ST} = 0.021, P < 0.0001). Of the two West Coast sites, California was more differentiated from all other populations than was British Columbia, despite their recent derivation from the same source (Table 2).

The PCA shows similar features and also reflects differences between regions. Our sampled sites between Europe (NO, PT) are distinct from one another and are the most separate from the rest of the samples (Fig. 2). The North American populations form a cluster, except for Newfoundland, which appears intermediate

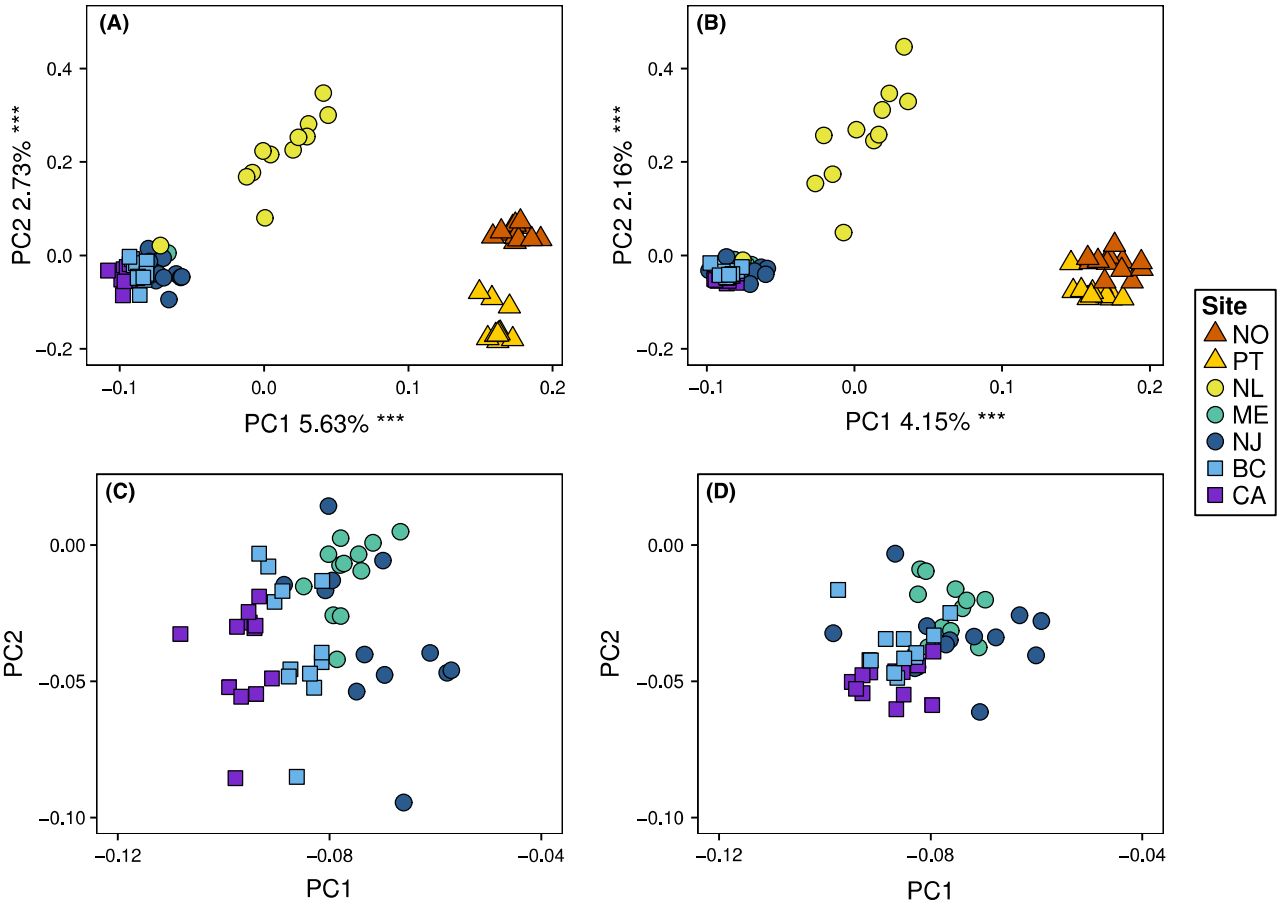

Fig. 2 Principal component analysis showing genetic structure for (A) all sites, all SNPs, and (B) all sites, neutral SNPs only. C and D show structure in Maine, New Jersey, California and British Columbia, expanded for clarity; C shows all SNPs, while D shows neutral SNPs only. Asterisks indicate highly significant PC axes (P < 0.001).

between the rest of North America and Europe (Fig. 2). Analysis with STRUCTURE also showed a split between Europe and the East Coast, with Newfoundland appearing to be the product of extensive introgression between these clusters (Fig. 3). When STRUCTURE was rerun on the European and East Coast clusters alone, each group was split into two subclusters. In Europe, locations were neatly separated, while on the East Coast, there was a less obvious East–West Coast split, with California differentiated from both East Coast sites while individuals from British Columbia were given equal assignment probability to both groups (Fig. 3).

Neutral population structure

The putatively neutral subset of SNPs (9246 of 10 809) showed a lower magnitude of population differentiation than the full SNP panel, as expected given that high-F_{ST} SNPs were removed from this data set. Yet, pairwise

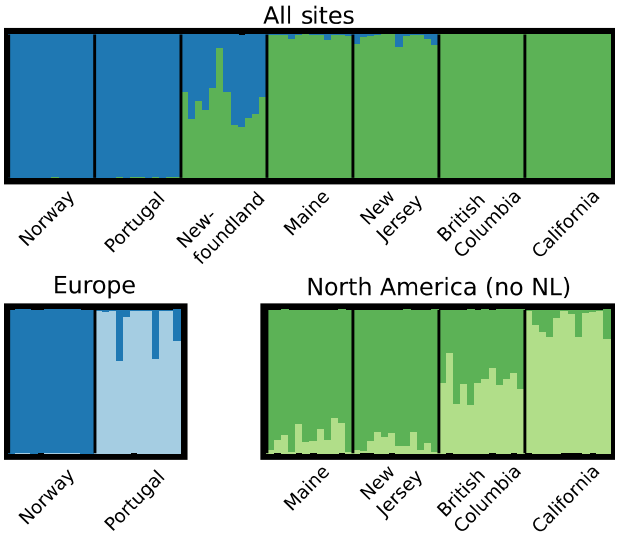

Fig. 3 STRUCTURE plot of all sites, all SNPs with $K = 2$ (optimal), and plots of Europe only ($K = 2$, optimal) and sites derived from the initial invasion to North America ($K = 2$, optimal $K = 3$ but did not provide additional information).

F_{ST} remained significant for all population pairs except Maine–New Jersey ($P = 0.5$; Table 3). Within-region differentiation was strongest between Newfoundland and the other two sites in the invasive East Coast range, with pairwise F_{ST} between locations of 0.024 (ME; $P < 0.0001$) and 0.026 (NJ; $P < 0.0001$). Differentiation in the West Coast was somewhat lower, yet still significant, with an F_{ST} of 0.014 ($P < 0.0001$). Populations in Europe showed a particularly low neutral F_{ST} of 0.009 ($P < 0.0001$), in striking contrast to an F_{ST} of 0.049 in the full SNP set. Of the two West Coast sites, California remained more differentiated from all other populations than did British Columbia (Table 3).

The magnitude of difference in pairwise F_{ST} calculated with the neutral and full data sets showed that, for most population pairs, neutral SNPs demonstrated approximately 50–70% as much differentiation as did all SNPs (Fig. 4). The exceptions were the Maine–New Jersey and Norway–Portugal comparisons. In the case of Maine and New Jersey, F_{ST} was very low with the full SNP set and dropped effectively to 0 with neutral SNPs (Tables 2 and 3). Within Europe, the F_{ST} calculated with neutral markers was only 18% of the F_{ST} calculated using the full data set, a much larger proportional drop than any other pairwise comparison except Maine–New Jersey (Tables 2 and 3; Fig. 4).

The neutral PCA reflects these differences: there is an overall lower level of pairwise differentiation, particularly along axis 2, with less separation evident between sites and individuals (Fig. 2B). This compression is especially obvious between Portugal and Norway in Europe. While each of these sites still forms a distinct cluster, these two European groups are nearly overlapping, without the clear separation evident in the full data set (Fig. 2A, B). This decreasing differentiation between sites is also evident, albeit less extreme, for the cluster made up of Maine, New Jersey, British Columbia and California (Fig. 2D). While the overall position of the loose Newfoundland cluster is unchanged, individuals from Newfoundland appear more dispersed along axis 2 when considering only neutral SNPs (Fig. 2A, B).

Table 3 Pairwise F_{ST} (below diagonal) and significance (above diagonal) for 9246 putatively neutral SNPs

	NO	PT	NL	ME	NJ	BC	CA
NO	—	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001
PT	0.009	—	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001
NL	0.038	0.041	—	<0.00001	<0.00001	<0.00001	<0.00001
ME	0.058	0.056	0.024	—	0.5	0.0001	<0.00001
NJ	0.056	0.053	0.026	−0.001	—	<0.00001	<0.00001
BC	0.065	0.065	0.033	0.010	0.010	—	<0.00001
CA	0.076	0.073	0.045	0.023	0.023	0.014	—

Sites as in Table 1, Fig. 1.

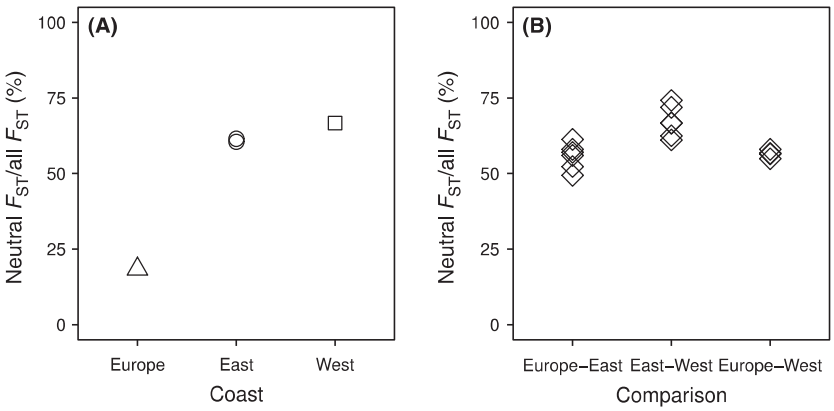

Fig. 4 Percentage change in pairwise F_{ST} using the neutral vs. full SNP set, between (A) sites within a region and (B) sites between regions. In panel A, the ME–NJ comparison has been omitted for clarity.

Genetic diversity

By all measures, genetic diversity was highest in the European native range and lowest in the West Coast (Table 4; Fig. 5). Allelic richness (Ar) ranged from 1.43 in CA to 1.55 in NO, with sites derived from the initial North American introduction all below 1.49 (1.43 – 1.48) and those in Europe both at 1.55 (Because this study employed only biallelic SNPs, the maximum possible Ar is 2.). As befits its hybrid origin, NL was more diverse than any of the other North American sites (Ar = 1.53), but not as diverse as populations in the native range (Table 4). Private allelic richness followed a similar but more extreme trend (Fig. 5), ranging from <0.011 in North American sites (0.007–0.010) through 0.049 in NL to >0.072 in Europe (0.073–0.076). Likewise, the proportion of global SNPs that were variable in a given population ranged from just over half (50.2%) in CA to more than two-thirds (68.5%) in PT (Table 4).

Discussion

Our study represents the first application of high-throughput sequencing to an invasive marine species in both its native and its invasive ranges. The increased resolution of genomewide SNP markers uncovered previously undetected population structure in the highly bottlenecked invasive range, confirming significant divergence in as little as 20 years. By isolating a subset of putatively neutral SNPs, we were able to contrast overall population structure with neutral structure. These comparisons strongly suggest that non-neutral forces are the primary driver of genetic structure in the native range, with native range F_{ST} dropping to <20% of its overall value after the removal of potentially non-neutral SNPs. In contrast, F_{ST} between invasive populations was much more robust to the removal of non-neutral SNPs. These data suggest that neutral forces such as drift predominantly shape the species in its invasive range, while selection may drive intraspecific differences in the older and more stable native range. The genetic data presented here are consistent with significant differences in thermal physiology between the same populations, especially in the native range, suggesting that temperature may be a major driver of selection in the native range. Further study is required to determine whether these overall patterns reflect selection on specific genomic regions, and to more thoroughly explore the potential role of temperature as a selective force. Our study illustrates the utility of high-throughput sequencing for understanding the complex population dynamics characteristic of invasion and highlights the importance of selection in shaping structure in high-dispersal marine invasive species.

Table 4 Genetic diversity by location

Site	Pn	Ar	pAr	H_o	H_E	F_{IS}
NO	0.683	1.551	0.076	0.173	0.174	0.004
PT	0.685	1.550	0.073	0.174	0.174	−0.000
NL	0.642	1.533	0.049	0.175	0.175	−0.002
ME	0.560	1.474	0.009	0.162	0.162	0.002
NJ	0.568	1.483	0.010	0.170	0.166	−0.025***
BC	0.532	1.456	0.007	0.159	0.157	−0.009
CA	0.502	1.432	0.007	0.152	0.151	−0.006

Pn, proportion of polymorphic loci; Ar, allelic richness; pAr, private allelic richness; H_o, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, Wright's inbreeding coefficient [asterisks indicate a significant heterozygote excess ($P < 0.001$)].

Population structure

This SNP panel represents a considerably larger portion of the genome than previous approaches, covering at least 1673 regions. These data showed significant genetic structure between all locations, even between invasive populations that appeared undifferentiated when using mtDNA and microsatellites (Darling *et al.*

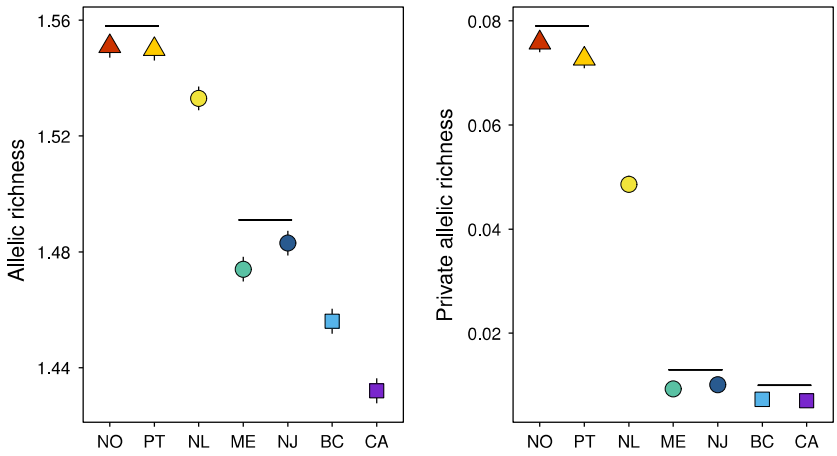

Fig. 5 Allelic richness and private allelic richness by location (mean ± standard error). Horizontal lines link locations that are not significantly different ($P \geq 0.05$).

2008). In the native European range, the SNP panel shows that the northern and southern populations are significantly different. This result is consistent with a distinct mtDNA cline concurrent with a wide-scale biogeographic break between northern and southern Europe (Roman & Palumbi 2004), characterized by differences in temperature regime and species composition (Adey & Steneck 2001). Populations along the East Coast of North America were also highly structured, reflecting a history of multiple introductions. Newfoundland was diverged from both Maine and New Jersey and showed extensive admixture between the 200-year-old initial introduction and a recently arrived introduction from the native range (Fig. 3). In contrast, Maine and New Jersey were the least differentiated of all the population pairs we studied, and despite the detection of northern mtDNA as far south as New York (Blakeslee *et al.* 2010), it appears that our site in Maine has little to no nuclear genetic influence from the second introduction.

In marked contrast to previous work (Darling *et al.* 2008), we demonstrate clear and significant differentiation between the East and West Coast invasive populations. The spread of *C. maenas* to the U.S. West Coast in the 1980s is a classic example of a secondary introduction (Darling *et al.* 2008), and genetic diversity levels in the SNP panel reflected that serial bottlenecking: high allelic richness in native range populations, a significant drop in diversity after the initial introduction to the East Coast and another significant drop after secondary introduction to the West Coast.

We also found significant structure within the West Coast invasive range, despite its very recent founding *circa* 1989. The California population, which was colonized early and is in the 'core' of the West Coast range, appears more differentiated from all other locations than does the more recently founded British Columbia population, located at the periphery of the species' West Coast range. This increased divergence of the California population is surprising because the dynamics associated with range expansion frequently favour higher differentiation of peripheral, recently colonized sites relative to more established sites in the centre of a species' range (Excoffier *et al.* 2009). In the case of West Coast *C. maenas*, this discrepancy is potentially resolved by considering the specifics of the sample locations. While the British Columbia population was sampled in a near-pristine natural environment in Barkley Sound, the California population was sampled in Seadrift Lagoon, a small artificial lagoon created in the 1960s by the dredging out of a wide sandbar running in front of the estuarine Bolinas Lagoon (Ritter 1970). This lagoon is shallow, heavily modified and relatively isolated from the open ocean (C.K. Tepolt, personal observation), suggesting that its population may experience idiosyncratic conditions and low genetic exchange with neighbouring embayments. The California population also had significantly lower allelic richness than its northern counterpart in British Columbia, supporting the hypothesis of increased isolation.

Genetic diversity was highest in both native range sites, dropping sharply in Maine and New Jersey. In Newfoundland, we observed an increase in diversity caused by new alleles from the second introduction from northern Europe. However, this bump was not enough to bring Newfoundland up to levels seen in the native range. Intriguingly, Newfoundland remains intermediate between Europe and the East Coast when private allelic richness is concerned, indicating that Newfoundland possesses a number of SNPs not found in either of our native range populations. This suggests that the second introduction to the East Coast derived from an unsampled European source containing private alleles, and hints at the possibility that, with additional sampling, SNPs may be able to detect much more fine-grained population structure in the native range.

Selection and drift in invasion

Population structure due to selection in the native range is quite relevant to the pace and pattern of species invasions. Local adaptation may influence the probability of invasion success, dependent on the quality of matching between source and invaded environment. Once a species has invaded, further adaptation in the invasive range may enhance a species' ability to persist and thrive by rapidly shifting its ecological niche to better match its new home (Sultan *et al.* 2013; Vandepitte *et al.* 2014). In several well-documented examples in insects and plants, native range latitudinal clines were rapidly recapitulated postinvasion (Balanyà *et al.* 2003; Maron *et al.* 2004; Balanyà *et al.* 2006). This dynamic has also been documented at range edges, where rapid adaptation may facilitate the spread of species beyond their previous limits (Phillips *et al.* 2006; Preisser *et al.* 2008; Krehenwinkel & Tautz 2013).

In many high geneflow marine species, where structure may be subtle and largely driven by adaptive differences (Sanford & Kelly 2011), SNPs are beginning to highlight patterns that are not apparent using traditional sequencing approaches such as microsatellites (Lamichhaney *et al.* 2012). Several recent genome-scale studies have suggested that intraspecific adaptation may be widespread in the sea, and include species such as Atlantic cod (Nielsen *et al.* 2009), purple sea urchin (Pespeni & Palumbi 2013) and Atlantic herring (Limborg *et al.* 2012). Studies suggesting adaptation in invasive marine species are much more limited, but suggest that adaptation may play as important a role in the sea as it does on land (Tepolt 2015). Studies within the native range of the invasive snail *Littorina saxatilis* strongly suggest adaptation in response to environmental factors related to tidal height (Grahame *et al.* 2006; Wood *et al.* 2008; Galindo *et al.* 2009, 2010). An AFLP- and microsatellite-based outlier scan of the slipper shell *Crepidula fornicata* suggested local adaptation within the native range, but no evidence for outliers within the invasive range or between the native and invasive ranges, although resolution was limited by a relatively small number of markers (Riquet *et al.* 2013). A similar study in the Pacific oyster *Crassostrea gigas* suggested local adaptation in the invasive range, although this result was confounded by invasion history (Rohfritsch *et al.* 2013).

In *C. maenas*, it has been hypothesized that the second introduction to the East Coast may have introduced cold-adapted genotypes to the range, allowing the species to expand into waters from which they had previously been thermally excluded (Roman 2006). While this pattern of expansion is also consistent with current-mediated dispersal (Pringle *et al.* 2011), physiological work suggests that crabs in Norway and in Newfoundland are more cold tolerant than their southern counterparts (Tepolt & Somero 2014). Our work supports this by showing that much of the differentiation between Norway and Portugal is driven by non-neutral processes (Tables 2 and 3; Fig. 4). Thus, much of the difference between Norway and Portugal may be driven by local adaptation, operating across a low level of differentiation maintained due to neutral processes such as dispersal limitation.

Further, the difference in between neutral SNPs and all SNPs was much higher in Europe than in any other pairwise comparison, suggesting that the native range harbours more adaptive alleles remaining in place than do recently invaded populations. While this study sampled only two native populations, these populations were chosen because they span much of the species' latitudinal range, and possess significantly different mitochondrial lineages and thermal physiology (Roman & Palumbi 2004; Tepolt & Somero 2014). Thus, they are likely to capture extremes of adaptation in the native range. Sampling from additional, intermediate native range populations may be able to determine whether adaptation plays as important a role in shaping structure across smaller distances.

In contrast to the native range, neutral processes appear to dominate in shaping invasive range structure. In the invasive range, Maine and New Jersey were not significantly different when only the neutral SNP panel was considered (Table 2). It is possible that this difference between neutral markers and all markers reflects some adaptation between these sites, located nearly 5° apart in latitude, but given the low level of initial differentiation, this loss of significance could also be due to the effect of removing high-F_{ST} loci from the data set. However, it is important to note that this apparent dominance of neutral processes does not rule out adaptation. Our estimates of structure are based on F_{ST}, which assumes that populations are in mutation-drift equilibrium (Wright 1978). While the assumption of equilibrium may be sound for the native range populations, it is probably violated in the invasive populations, which have had no more than 200 years (ME–NJ) and as few as 15 (CA–BC) to reach equilibrium (Whitlock & McCauley 1999; Grosberg & Cunningham 2001). Thus, our invasive populations may indeed be experiencing selection, but at a level that is swamped by the signal of recent founding and rapid population expansion, as opposed to the more extensive adaptation that appears to be operating between the 'stable' native populations.

On the West Coast, the neutral SNP panel recapitulated the pattern of California being more differentiated than British Columbia from all other sites (Tables 2 and 3; Fig. 2C, D). As discussed above, this surprising drop

in the core of the range may be related to idiosyncratic conditions in the specific study site. Interestingly, this population also had a dramatic loss of heat tolerance relative to all other studied populations (Tepolt & Somero 2014). While in no way conclusive, it is possible that significant bottlenecking may have negatively influenced the potential future fitness of this population. The fact that California remains more differentiated than British Columbia with both the full and neutral SNP panels supports this hypothesis, suggesting that neutral processes – such as bottlenecking in this enclosed environment – may primarily contribute to divergence.

Comparing neutral and overall structure was able to shed light on the dynamics of invasion in *C. maenas*, and in doing so, it improves our understanding of marine invasions more generally. In the native range, adaptation appeared to be the primary driver of structure. This fits with previous data, which showed no structure for putatively neutral microsatellites, but did indicate a clear break in potentially adaptive mitochondrial lineages. If populations of invasive species are locally adapted, this suggests that there may be a benefit to preventing future introductions to already-invaded areas, in an effort to curb further spread. In the oldest invasive population pair, Maine and New Jersey, weak overall structure was not captured by neutral markers, suggesting high gene flow but potentially incipient adaptive differences. In contrast, neutral structure remained strong between the West Coast sites, reinforcing the potential isolation and divergent physiology of the California site. Because these newly founded populations are not likely to have reached mutation-drift equilibrium, we expect that these patterns – and the dominance of neutral structure – are likely to shift over time. Further studies will focus on identifying specific genes underlying these overarching patterns, and their relationship with potential environmental drivers such as temperature.

Conclusions

Understanding genetic structure plays a crucial role in many conservation and management programmes. In the case of many invasive species, invasion history and pathways are poorly understood, yet this information can aid in designing effective management strategies to prevent further introduction and spread (Estoup & Guillemaud 2010). While these panels are more typically used to identify specific candidate genes, large panels of genomewide markers delivered by high-throughput sequencing permit researchers to begin untangling the relative importance of neutral and adaptive structure in shaping and facilitating marine invasions (Savolainen *et al.* 2013). In the case of our study, these data support the hypothesis that local adaptation in the native range may have played an important role in permitting the spread of the invasion via multiple introductions. At the same time, neutral forces of bottlenecks and genetic drift appear to differentiate many of the populations within the invasive North American range, suggesting that even a highly bottlenecked population may retain an exceptional capacity to thrive and spread without invoking extensive adaptation. We suggest that the ability of population genomics to elucidate fine-scale population structure, and the relative contributions of neutral and selective forces on shaping that structure, holds great promise for understanding the complicated dynamics of marine invasions.

Acknowledgements

We sincerely thank John Lee; the Bamfield Marine Centre; Greg Ruiz and the Smithsonian Environmental Research Center; the Rutgers University Marine Field Station; the Darling Marine Center; Iain McGaw and the Ocean Sciences Centre; João Canning-Clode and the Laboratório Marítimo da Guia; and Henrik Glenner and the Espegrend Marine Biological Station. We are grateful for support from the Partnership for the Interdisciplinary Study of Coastal Oceans, the Myers Trust, the Explorer's Club Exploration Fund, the Lerner Gray Memorial Fund of the American Museum of Natural History, the Vice Provost for Graduate Education at Stanford, the Eugene C. and Aileen E. Haderlie Memorial Fund, and a National Science Foundation Doctoral Dissertation Improvement Grant (1210057). CKT was supported on this project by a National Defense Science and Engineering Grant, a Stanford Graduate Fellowship, and a Stanford Center for Computational, Evolutionary, and Human Genomics Fellowship; she is currently supported by a Smithsonian Institution Biodiversity Genomics Fellowship. We also thank two anonymous reviewers for their thoughtful comments which significantly improved this manuscript.

References

Adey WH, Steneck RS (2001) Thermogeography over time creates biogeographic regions: temperature/space/time-integrated model and an abundance-weighted test for benthic marine algae. *Journal of Phycology*, **37**, 677–698.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{st}-outlier method. *BMC Bioinformatics*, **9**, 323.

Balanyà J, Oller JM, Huey RB, Gilchrist GW, Serra L (2006) Global genetic change tracks global climate warming in *Drosophila subobscura*. *Science*, **313**, 1773–1775.

Balanyà J, Serra L, Gilchrist GW *et al.* (2003) Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series. *Evolution*, **57**, 1837–1845.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Behrens Yamada S, Hunt C (2000) The arrival and spread of the European green crab, *Carcinus maenas*, in the Pacific Northwest. *Dreissena*, **11**, 1–7.

Bierne N, Bonhomme F, David P (2003) Habitat preference and the marine-speciation paradox. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1399–1406.

Blakeslee AMH, McKenzie CH, Darling JA *et al.* (2010) A hitchhiker's guide to the Maritimes: anthropogenic transport facilitates long-distance dispersal of an invasive marine crab to Newfoundland. *Diversity and Distributions*, **16**, 879–891.

Byers J, Pringle J (2006) Going against the flow: retention, range limits and invasions in advective environments. *Marine Ecology Progress Series*, **313**, 27–41.

Carlton JT, Cohen AN (2003) Episodic global dispersal in shallow water marine organisms: the case history of the European shore crabs *Carcinus maenas* and *Carcinus aestuarii*. *Journal of Biogeography*, **30**, 1809–1820.

Carlton JT, Geller JB (1993) Ecological roulette: the global transport of nonindigenous marine organisms. *Science*, **261**, 78–82.

Cohen AN, Carlton JT, Fountain MC (1995) Introduction, dispersal and potential impacts of the green crab *Carcinus maenas* in San Francisco Bay, California. *Marine Biology*, **122**, 225–237.

Darling JA, Bagley MJ, Roman J, Tepolt CK, Geller JB (2008) Genetic patterns across multiple introductions of the globally invasive crab genus *Carcinus*. *Molecular Ecology*, **17**, 4992–5007.

De Wit P, Palumbi SR (2012) Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Molecular Ecology*, **22**, 2884–2897.

De Wit P, Pespeni MH, Ladner JT *et al.* (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources*, **12**, 1058–1067.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

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

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

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

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Fausch KD (2008) A paradox of trout invasions in North America. *Biological Invasions*, **10**, 685–701.

Gaither MR, Bowen BW, Toonen RJ (2013) Population structure in the native range predicts the spread of introduced marine species. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20130409.

Galindo J, Morán P, Rolán-Alvarez E (2009) Comparing geographical genetic differentiation between candidate and non-candidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology*, **18**, 919–930.

Galindo J, Grahame JW, Butlin RK (2010) An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *Journal of Evolutionary Biology*, **23**, 2004–2016.

Grahame JW, Wilding CS, Butlin RK (2006) Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution*, **60**, 268–278.

Grosberg R, Cunningham CW (2001) Genetic structure in the sea: from populations to communities. In: *Marine Community Ecology* (eds Bertness MD, Gaines SD, Hay ME), pp. 61–84. Sinauer Associates, Sunderland, Massachusetts.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2013) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, **22**, 2898–2916.

Hodgins KA, Lai Z, Nurkowski K, Huang J, Rieseberg LH (2012) The molecular basis of invasiveness: differences in gene expression of native and introduced common ragweed (*Ambrosia artemisiifolia*) in stressful and benign environments. *Molecular Ecology*, **22**, 2496–2510.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Huey RB, Pascual M (2009) Partial thermoregulatory compensation by a rapidly evolving invasive species along a latitudinal cline. *Ecology*, **90**, 1715–1720.

Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, **21**, 1552–1560.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Krehenwinkel H, Tautz D (2013) Northern range expansion of European populations of the wasp spider *Argiope bruennichi* is associated with global warming-correlated genetic admixture and population-specific temperature adaptations. *Molecular Ecology*, **22**, 2232–2248.

Lamichhaney S, Martinez Barrio A, Rafati N *et al.* (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences, USA*, **109**, 19345–19350.

Lemay MA, Russello MA (2015) Genetic evidence for ecological divergence in kokanee salmon. *Molecular Ecology*, **24**, 798–811.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Limborg MT, Helyar SJ, De Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, **21**, 3686–3703.

Lodge DM, Williams S, MacIsaac HJ *et al.* (2006) Biological invasions: recommendations for US policy and management. *Ecological Applications*, **16**, 2035–2054.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Maron JL, Vilà M, Bommarco R, Elmendorf S, Beardsley P (2004) Rapid evolution of an invasive plant. *Ecological Monographs*, **74**, 261–280.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Moore J-S, Bourret V, Dionne M *et al.* (2014) Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology*, **23**, 5680–5697.

Moura AE, Kenny JG, Chaudhuri R *et al.* (2014) Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Molecular Ecology*, **23**, 5179–5192.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, **17**, 3599–3613.

Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.

Palumbi SR, Pinsky ML (2013) Marine dispersal, ecology and conservation. In: *Marine Community Ecology*, 2nd edn (eds Bertness M, Bruno J, Silliman B, Stachowicz J), pp. 57–83. Sinauer Associates, Sunderland, Massachusetts.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.

Pespeni MH, Palumbi SR (2013) Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. *Molecular Ecology*, **22**, 3580–3597.

Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed in toads. *Nature*, **439**, 803.

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Preisser EL, Elkinton JS, Abell K (2008) Evolution of increased cold tolerance during range expansion of the elongate hemlock scale *Fiorinia externa* Ferris (Hemiptera: Diaspididae). *Ecological Entomology*, **33**, 709–715.

Prentis P (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, **13**, 288–294.

Pringle JM, Blakeslee AMH, Byers JE, Roman J (2011) Asymmetric dispersal allows an upstream region to control population structure throughout a species' range. *Proceedings of the National Academy of Sciences, USA*, **108**, 15288–15293.

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

R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Ricciardi A (2007) Are modern biological invasions an unprecedented form of global change? *Conservation Biology*, **21**, 329–336.

Riquet F, Daguin-Thiébaut C, Ballenghien M, Bierne N, Viard F (2013) Contrasting patterns of genome-wide polymorphism in the native and invasive range of the marine mollusc *Crepidula fornicata*. *Molecular Ecology*, **22**, 1003–1018.

Ritter JR (1970) A summary of preliminary studies of sedimentation and hydrology in Bolinas Lagoon, Marin County, California. *Geological Survey Circular*, **627**, 28.

Rohfritsch A, Bierne N, Boudry P *et al.* (2013) Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas*. *Evolutionary Applications*, **6**, 1064–1078.

Rollins LA, Woolnough AP, Sherwin WB (2006) Population genetic tools for pest management: a review. *Wildlife Research*, **33**, 251–261.

Roman J (2006) Diluting the founder effect: cryptic invasions expand a marine invader's range. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2453–2459.

Roman J, Palumbi SR (2004) A global invader at home: population structure of the green crab, *Carcinus maenas*, in Europe. *Molecular Ecology*, **13**, 2891–2898.

Rosenberg NA (2003) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Sanford E, Kelly MW (2011) Local adaptation in marine invertebrates. *Annual Review of Marine Science*, **3**, 509–535.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.

Simms D, Cizdziel PE, Chomczynski P (1993) Trizol: a new reagent for optimal single-step isolation of RNA. *Focus Journal*, **15**, 99–102.

Sotka EE, Palumbi SR (2006) The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology*, **87**, 1094–1103.

Sultan SE, Horgan-Kobelski T, Nichols LM, Riggs CE, Waples RK (2013) A resurrection study reveals rapid adaptive evolution within populations of an invasive plant. *Evolutionary Applications*, **6**, 266–278.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.

Tepolt CK (2015) Adaptation in marine invasion: a genetic perspective. *Biological Invasions*, **17**, 887–903.

Tepolt CK, Somero GN (2014) Master of all trades: thermal acclimation and adaptation of cardiac function in a broadly-distributed marine invasive species, the European green crab, *Carcinus maenas*. *Journal of Experimental Biology*, **217**, 1129–1138.

Tepolt CK, Darling JA, Bagley MJ *et al.* (2009) European green crabs (*Carcinus maenas*) in the northeastern Pacific: genetic evidence for high population connectivity and current-medi-

ated expansion from a single introduced source population. *Diversity and Distributions*, **15**, 997–1009.
Vandepitte K, de Meyer T, Helsen K *et al.* (2014) Rapid genetic adaptation precedes the spread of an exotic plant species. *Molecular Ecology*, **23**, 2157–2164.
Waples R (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.
White TA, Perkins SE, Heckel G, Searle JB (2013) Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, **22**, 2971–2985.
Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. *Heredity*, **82**, 117–125.
Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.
Wilson JRU, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution*, **24**, 136–144.
Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123–3135.
Wright S (1978) *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations*, pp. 580. University of Chicago Press, Chicago, Illinois.

C.K.T. was responsible for project planning, tissue collection, sample preparation, data analysis and manuscript writing. S.R.P. was responsible for project planning, manuscript writing and partially funding the sequencing.

Data accessibility

The *de novo Carcinus maenas* cardiac transcriptome, SNP positions and individual genotypes have been deposited in Dryad at doi:10.5061/dryad.g8b96. The raw sequencing reads have been archived at NCBI's Sequence Read Archive under BioProject ID PRJNA283611, BioSample numbers SAMN03653390–SAMN03653473.