Biol Invasions (2018) 20:1577–1595
https://doi.org/10.1007/s10530-017-1649-0

ORIGINAL PAPER

A genetic reconstruction of the invasion of the calanoid copepod *Pseudodiaptomus inopinus* across the North American Pacific Coast

Eric Dexter · Stephen M. Bollens · Jeffery Cordell · Ho Young Soh · Gretchen Rollwagen-Bollens · Susanne P. Pfeifer · Jérôme Goudet · Séverine Vuilleumier

Received: 28 April 2017 / Accepted: 5 December 2017 / Published online: 12 December 2017
© Springer International Publishing AG, part of Springer Nature 2017

Abstract The rate of aquatic invasions by planktonic organisms has increased considerably in recent decades. In order to effectively direct funding and resources to control the spread of such invasions, a methodological framework for identifying high-risk transport vectors, as well as ruling out vectors of lesser concern will be necessary. A number of estuarine ecosystems on the North American Pacific Northwest coast have experienced a series of high impact planktonic invasions that have slowly unfolded across the region in recent decades, most notably, that of the planktonic copepod crustacean *Pseudodiaptomus inopinus*. Although introduction of *P. inopinus* to the United States almost certainly occurred through the discharge of ballast water from commercial vessels originating in Asia (the species' native range), the mechanisms and patterns of subsequent spread remain unknown. In order to elucidate the migration events shaping this invasion, we sampled the genomes of copepods from seven invasive and two native populations using restriction-site associated DNA sequencing. This genetic data was evaluated against spatially-explicit genetic simulation models to evaluate competing scenarios of invasion spread. Our results

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10530-017-1649-0) contains supplementary material, which is available to authorized users.

E. Dexter (✉) · S. M. Bollens · G. Rollwagen-Bollens
Washington State University, 14204 NE Salmon Creek Avenue, Vancouver, WA 98686-9600, USA
e-mail: eric.dexter@wsu.edu

J. Cordell
School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

H. Y. Soh
Chonnam National University, 50 Daehagro, Yeosu, Jellanam-do 59626, Republic of Korea

S. P. Pfeifer
École Polytechnique Fédérale de Lausanne, EPFL SV IBI-SV UPJENSEN, AAB 0 45 (Batiment AAB), Station 15, 1015 Lausanne, Switzerland

S. P. Pfeifer
School of Life Sciences, Arizona State University, 427 E Tyler Mall, Tempe, AZ 85287, USA

J. Goudet
University of Lausanne, Quartier UNIL-Sorge, Bâtiment Biophore Bureau : 4326, 1015 Lausanne, Switzerland

S. Vuilleumier
École Polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland

S. Vuilleumier
La Source, University of Applied Sciences and Arts, Western Switzerland (HES-SO) Health Science, Lausanne, Switzerland

indicate that invasive populations of *P. inopinus* exhibit a geographically unstructured genetic composition, likely arising from infrequent and large migration events. This pattern of genetic patchiness was unexpected given the linear geographic structure of the sampled populations, and strongly contrasts with the clear invasion corridors observed in many aquatic systems.

Keywords RADseq · Zooplankton · Aquatic invasions · Migration/colonization pattern · ABC · Genetic simulation · Copepod

Introduction

Aquatic invasive species

Aquatic invasive species are nearly ubiquitous in urbanized bodies of water, and in some cases, catalyze dramatic ecosystem shifts. Notorious examples include the invasion of zebra and quagga mussels to the North American Great Lakes (Schloesser et al. 1996; Vanderploeg et al. 2002; Cuhel and Aguilar 2013) as well as the introduction of the ctenophore *Mnemiopsis leidyi* to the Black Sea, which led to the collapse of several fisheries (Shiganova 1998; Knowler 2005). Besides dramatic ecological effects, aquatic invaders can also inflict heavy damages on water supplies and power generation facilities (Connelly et al. 2007), interfere with recreation on lakes and rivers (Simberloff 1997), and promote blooms of harmful algal species (Vanderploeg et al. 2001). Economic impacts of such invasions are estimated at $120 billion per year in the USA alone (Pimentel et al. 2005).

In the Pacific Northwest region of North America, several major estuaries have experienced high impact planktonic invasions in recent decades (Bollens et al. 2002; Cordell et al. 2008; Winder and Jassby 2011). Successful planktonic invaders in this region include four species of copepods (*Sinocalanus doerrii*, *Limnoithona tetraspina*, *Pseudodiaptomus inopinus*, and *Pseudodiaptomus forbesi*), a bosminid cladoceran (*Bosmina coregoni*), and larval Asian clams (*Corbicula fluminea*) (Bollens et al. 2012; Breckenridge et al. 2015; Dexter et al. 2015). The calanoid copepod *Pseudodiaptomus inopinus* is of particular interest, as it has emerged as the dominant invasive zooplankton in at least nine major estuaries in the Pacific Northwest (Bollens et al. 2002).

Dispersal of planktonic copepods

Copepods are one of the most abundant groups of multicellular animals, and are the dominant primary consumers in many bodies of water (Humes 1994). Copepod populations show strong intra-annual variation in the temperate latitudes, with peak abundances in large bodies of water numbering in the billions or trillions of individuals, followed by precipitous declines—typically during winter months (Bron et al. 2011; Emerson et al. 2015; Dexter et al. 2015). Most well-studied copepod taxa are marine copepods, which passively disperse through the action of oceanic and coastal currents (Palumbi 1994), thus possessing enormous potential for rapid dispersal across vast geographic distances (Tatebe et al. 2010). Marine copepod populations likely experience very high rates of migration—thus forming well-mixed populations with little genetic structure (Palumbi 1994; Hellberg et al. 2002). Indeed, significant genetic structure is typically detected only across major geographic barriers such as oceanic basins, gyres, or continental landmasses (Goetze 2005; Goetze et al. 2015)—with some notable exceptions (Lee 2000; Blanco-Bercial and Bucklin 2016).

In contrast, many of the lesser-studied copepod species are associated with discrete patches of habitat, such as coastal rock pools (Ganz and Burton 1995; Van Wormhoudt 2015) and inland bodies of water (Boxshall and Defaye 2008), constraining the exchange of migrants. Indeed, one of these, the aforementioned *P. inopinus*, appears to be restricted to estuarine waters of intermediate salinity (Cordell et al. 2010). Relative to marine species, the demographic patterns of copepods with a more constrained capacity for dispersal (i.e., those inhabiting rivers, lakes, ephemeral ponds, and estuaries) have been investigated in a relatively limited fashion (Bron et al. 2011).

Such geographically isolated species of zooplankton require transport vectors to facilitate migration between discontinuous patches of habitat (Havel and Shurin 2004). For estuarine species tolerating fully saline waters, this vector can be an outgoing tide or coastal current—a mechanism quite common among

many estuarine crustaceans (Christy and Stancyk 1982). Salt-intolerant species or those inhabiting inland waters may exchange migrants using transport mechanisms that range from plumage of aquatic birds (Frisch et al. 2007), transport of desiccation-resistant eggs via airborne currents (Cáceres and Soluk 2002), the movement of recreational boats (Havel and Stelzleni-Schwent 2001), or the discharge of ballast water from commercial cargo vessels (Carlton and Geller 1993; Ruiz et al. 2000; Drake et al. 2014). Although many vectors have been identified as potential agents of migration, the degree to which these mechanisms actually shape the genetic structure of copepod populations in relatively unknown. Given the rapidly increasing economic and ecological costs of planktonic invasions (Pimentel et al. 2005), this question deserves timely elucidation.

Genetic reconstruction of demographic patterns

The genetic structure of a species is inextricably linked to its demographic history, including its past patterns of migration (Avise 2000). Planktonic organisms, by definition, rely primarily on passive transport mechanisms for dispersal, and thus have demographic histories tightly correlated with those of their transport vector(s). Because different migration patterns are associated with certain genetic signatures (Excoffier 2004; Vuilleumier et al. 2010), the genomes of invasive populations may provide important insight into the histories of those invasions (Estoup and Guillemaud 2010).

Traditionally, demographic reconstruction has been based on a small number of well-characterized genetic markers. For example, the analysis of several nuclear and mitochondrial markers in Mediterranean populations of the highly invasive ctenophore *Mnemiopsis leidyi* confirmed that introduction occurred via Black Sea populations, and also identified a previously unknown secondary introduction via ballast water originating from the Gulf of Mexico (Ghabooli et al. 2013). Similarly, a microsatellite-based approach was successfully employed to reconstruct the invasion pathway of the cladoceran *Daphnia lumholtzi* across eastern North America—an invasion that originated in North Africa, with subsequent introductions from Asia (Frisch et al. 2013).

Unfortunately, studies relying on a relatively small number of genetic markers often yield unclear or even contradictory demographic results, especially when those markers are drawn from both the nuclear and mitochondrial genome (Toews and Brelsford 2012). Recently developed next-generation sequencing methods allow for relatively low-cost sequencing of a large number of genetic markers, typically in excess of 10,000 per sample, conducted across many individuals in parallel (McCormack et al. 2013). Thus, these methods impart much more statistical power than those employed in the studies mentioned above (Davey and Blaxter 2010). This power is especially salient in relatively fine-scaled phylogeographic studies with potentially complicated patterns of gene flow (Davey and Blaxter 2010). In particular, Restriction-enzyme Associated DNA sequencing (RADseq) and its subsequent modifications (Baird et al. 2008; Peterson et al. 2012) have been employed in several phylogeographic studies to great effect (Emerson et al. 2010; Vandepitte et al. 2013). For example, a RADseq-based study of the calanoid copepod *Centropages typicus* uncovered significant population structure in North Atlantic populations, which was undetectable using traditional genetic markers (Blanco-Bercial and Bucklin 2016). RADseq or similar approaches involving large genomic datasets have also been employed to elucidate population structure in a number of other crustacean species as well [e.g. American lobster (*Homarus americanus*)—Benestan et al. 2015, and European green crab (*Carcinus maenus*)—Tepolt and Palumbi 2015; Jeffery et al. 2017].

The invasion of *Pseudodiaptomus inopinus*

Pseudodiaptomus inopinus is an abundant native member of planktonic communities across the temperate estuaries of Japan, China, and Korea (Sakaguchi and Ueda 2010, 2011; Ueda et al. 2010; Lin et al. 2011; Soh et al. 2012; Park et al. 2013) (Fig. 1). This species was first observed in North America at the lower Columbia River in 1990 (Cordell et al. 1992), and was subsequently found in at least ten other Pacific Northwest estuaries (Bollens et al. 2002). In affected bodies of water, there are marked seasonal losses of zooplankton community diversity. Indeed, plankton surveys undertaken during late summer and early autumn have shown that, while non-invaded rivers exhibit a diverse assemblage of copepod species (> 8 commonly abundant species), invaded rivers tend to

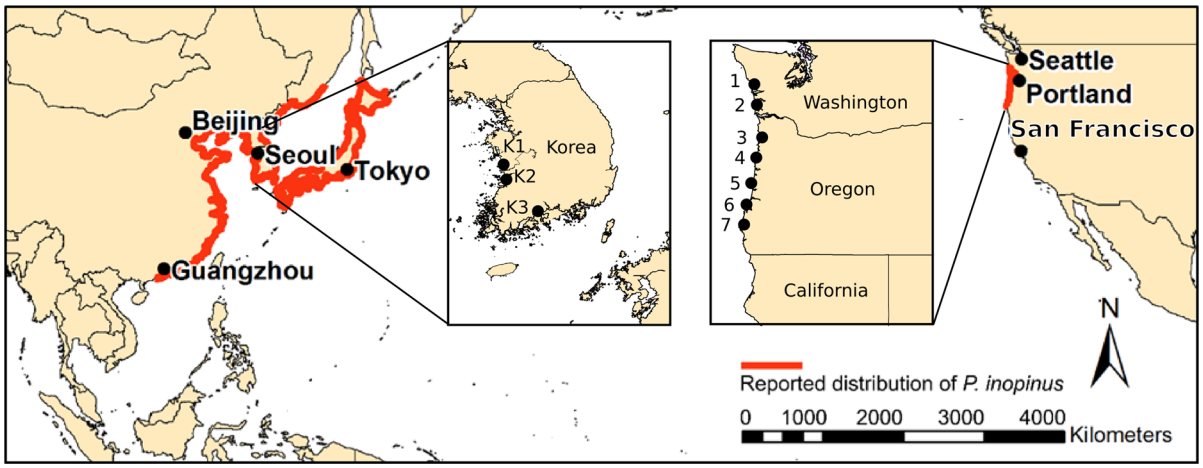

Fig. 1 Presumed geographic distribution of *P. inopinus* based on observations from the literature (Sakaguchi and Ueda 2010, 2011; Ueda et al. 2010; Lin et al. 2011; Soh et al. 2012; Park et al. 2013). This figure does not reflect the recent discovery of several cryptic species formerly identified as *P. inopinus* in Korea and Japan. Sampled estuaries shown on the North American inset are: Chehalis (*1*), Willapa (*2*), Tillamook (*3*), Yaquina (*4*), Umpqua (*5*), Coos (*6*), and Coquille (*7*). Sampled estuaries shown on the Korean inset are: Geum (K1) and Mankyung (K2), and Beolgyo (K3)

be completely dominated by *P. inopinus* (Bollens et al. 2002; Cordell et al. 2008). It remains unclear if *P. inopinus* is the causal agent of these declines in diversity, or if the presence of a relatively depauperate community facilitated the establishment of *P. inopinus* at these locations. Curiously, *P. inopinus* appears to have been extirpated in the Columbia River, the presumed site of first introduction, potentially due to the establishment of a latter-arriving congeneric copepod, *P. forbesi* (Cordell et al. 2008; Bollens et al. 2012; Dexter et al. 2015).

The initial introduction of *P. inopinus* into Pacific Northwest estuaries probably occurred through the transfer of ballast water from the native Asian range via commercial vessels (Cordell et al. 2009; Lawrence and Cordell 2010). The Port of Portland is a likely candidate for the site of initial introduction, as it is situated on the lower Columbia River (the site of first detection), and is by far the largest international port within the region invaded by *P. inopinus* (http://www.pmanet.org/port-locations-stats). However, the mechanisms facilitating the subsequent spread of *P. inopinus* across the Northeast Pacific ocean are unknown. Multiple direct introductions from the native range are improbable given most of these estuaries' inaccessibility to international shipping vessels (Port Locations & Stats | Pacific Maritime Association n.d.). Likewise, secondary spread among estuaries via coastal currents is unlikely given the apparent limited physiological tolerance of *P. inopinus* to cold saline waters (Cordell et al. 2010). As the secondary spread of *P. inopinus* largely occurred prior to the most recent observational record (Lawrence and Cordell 2010), the historical record of first observations of *P. inopinus* cannot be employed to reconstruct the progression of this invasion.

Here, we reconstructed the invasion history of *P. inopinus* across the Pacific Northwest region of North America using a genetic-based approach coupled to spatially-explicit simulation models. These models simulate the patterns of genetic structure expected to arise under various biologically-plausible scenarios of *P. inopinus* migration. The resultant simulated genetic patterns were evaluated against genetic data under an approximate Bayesian computation framework. Our methodological approach of sequencing large pools of whole specimens illustrates a promising new avenue for the study of the invasion process in planktonic organisms.

Materials and methods

Specimen collection, DNA extraction, and library preparation

Pseudodiaptomus inopinus samples were collected from seven populations in the North American range

and from two putatively native populations in South Korea, which is one of the major maritime trading partners with the USA (4th largest trading partner by value and 9th largest by tonnage http://www.aapa-ports.org). A number of major international ports (notably those centered upon the city of Busan—Yeo et al. 2008) exists within the native Korean range of *P. inopinus*.

North American samples were collected from the estuaries of the Chehalis, Willapa, Tillamook, Yaquina, Umpqua, Coos, and Coquille rivers (Fig. 1) in August and September of 2012. Specimens were collected via horizontal near-surface net tows of a 75 micron mesh, 0.5 m diameter net towed slowly upriver (approximately 1 m s^{-1}), and preserved in a 70% ethanol solution. Samples of native *P. inopinus* populations were collected from the Geum River and Mankyung River, located in western South Korea. Both rivers empty into Yellow sea at locations approximately 15 km apart. Additionally, samples of the congeneric *P. koreanus* were collected from southern Korea as an outgroup for the genetic analysis. *P. koreanus* is a recently erected taxon which has been recognized as distinct from *P. inopinus* based upon morphological details related to reproductive structures and evidence from several genetic markers (Soh et al. 2012). Genetic divergence between *P. inopinus* and *P. koreanus* has been measured as 12–14% at the ribosomal internal transcribed spacer 1 (ITS1) gene and 14–22% at mitochondrial cytochrome oxidase subunit I (mtCOI) gene (Soh et al. 2012). *P. koreanus* samples were collected from the Beolgyo River, which empties into the Korea Strait at an approximate distance (traced along the coastline) of 350 km from the other Korean sampling locations. Samples from Korean estuaries were collected via vertical tows using a 200-micron mesh net in April of 2015 and preserved in a 99% ethanol solution. Geographic coordinates and pairwise distances between sampling locations are provided in supplemental materials (Tables S1 and S2).

Due to the small size of *P. inopinus* individuals (typical adult length < 1.5 mm) and the small yield of genetic material per specimen, copepods from each population were sequenced as pools of individuals. Two replicate pools of approximately 40 copepods (adult males and females) and copepodites (stage I–V juveniles) were sequenced for each population. The sampling scheme differed somewhat for the two most northern North American populations (Willapa and Chehalis), which had not yet entered the annual period of population boom at these higher-latitude locations by the time of collection (September). At the Chehalis River, only seven overwintering adult females were collected, which were sequenced as a single unreplicated pool. At the Willapa River, only small copepodites were collected and pools of approximately 150 individuals were required to yield enough DNA for the RADseq protocol. An accounting of the contents of each pool is given in supplemental materials (Table S1). In order to account for the pooled sequencing of specimens, genetic summary statistics were calculated using Popoolation2 v1.2, a software tool designed for the comparison of pooled RADseq data (see genetic data analysis below).

Because adult-stage *P. inopinus* were uncommon in most samples, copepodites were included in all pools of sequenced individuals. Although adult copepods are typically preferred for reliable taxonomic identification, *P. inopinus* copepodites can be readily distinguished from co-occurring copepods in the North American range via a characteristic arrangement of its caudal setae visible in all copepodite stages (Dexter, personal observation). Though recent studies have identified several cryptic species formerly identified as *P. inopinus* (Sakaguchi and Ueda 2010; Soh et al. 2012), morphological examination of specimens (i.e. reproductive structure of both sexes) against *P. inopinus* from the native range confirmed the sampling of *P. inopinus* sensu stricto in this study.

DNA from pooled copepod samples was extracted using a QIAamp DNA micro kit (Qiagen products) according to the supplied manufacturer protocol for tissue-based extraction. *P. inopinus* samples were genotyped using RADseq (Baird et al. 2008)—a next-generation sequencing method ideally suited to population-level genetic studies (Davey and Blaxter 2010). The RADseq libraries were prepared according to the protocol of Parchman et al. (2012), with modifications (e.g. digestion with restriction enzymes Sbf1 and MseI) as suggested Brelsford et al. (2016). Prepared libraries were sequenced via paired-end sequencing on a single lane of an Illumina Hiseq 2500 by the Lausanne Genomic Technologies Facility at the University of Lausanne Center for Integrative Genomics.

Genetic data analysis

Raw sequence data were demultiplexed and quality filtered using the 'process_radtags' script in Stacks v1.34 (Catchen et al. 2013), with forward and reverse reads treated as unpaired. Any reads with an uncalled base were filtered out, as were reads which showed a decreasing quality level along the read length as measured in a sliding window encompassing 15% of the read (Stacks optional parameters "-c" and "-q" enabled). Replicate pools from each site were merged for all downstream analyses. In order to remove sequence data from non-copepod contaminants (e.g., copepod gut contents, bacterial microfilms, and common laboratory contaminants), all sequence data were filtered through Kraken v0.10.5 (Wood and Salzberg 2014) using a greatly expanded custom genome database, including bacterial, fungal, protozoan, viral, plasmid, and mitochondrial genomes as well as a panel of eukaryotic genomes from the RefSeq representative genomes collection (Supplemental materials: Table S3). Kraken operates by partitioning sequence reads into a set of k-mers, which are queried against this custom database of genomes. The lowest common ancestor associated with the set of matching k-mers is returned, which can be at any taxonomic level. Any sequence reads classified as non-crustacean (subphylum Crustacea) in origin via the Kraken alignment algorithm (operating under default parameters) were removed from downstream analyses.

Reads that passed all filtering steps were assembled into loci using the 'denovo_map' pipeline in Stacks v1.34, operating under default parameters. Loci with a stack depth less than 7 or minor allele frequency less than 0.1 were not retained in the final catalog, nor were loci present in fewer than four populations. This relatively low cutoff for stack depth was necessary to compensate for the large amount of contaminants introduced by the sequencing of entire specimen. Some replicate samples from Mankyung and Beolgyo rivers were poorly amplified during library preparation and thus were excluded from downstream analyses.

The following genetic summary statistics were calculated from each population: the total number of genomic sites sequenced, the number of private alleles within the population, the number of polymorphic alleles within the population, genome-wide expected heterozygosity and nucleotide diversity. The number of polymorphic sites were given both as raw values and as the ratio of polymorphic sites to covered sites within each population. Additionally, pairwise F_{ST} values were calculated for all possible pairs of populations. As the Stacks pipeline is optimized for calculation of genetic summary statistics from individual specimens (as opposed to pooled samples), single nucleotide polymorphism (SNP) calling and the calculation of genetic summary statistics were conducted using Popoolation2 v1.2 (Kofler et al. 2011). To achieve this, the catalog of loci from Stacks was exported as a synthetic genome, to which the previously filtered sequence data was aligned using BWA v0.7.12 (Li 2013). SNPs were then called against this aligned sequence data in Popoolation2, with the number of organisms in each pool provided as an optional input. Region-wide differences in genetic diversity (evaluated by the numbers of polymorphic loci and values of nucleotide diversity) were assessed via a two-tailed *t* test, adjusted for multiple comparisons with a Bonferroni correction (Bland and Altman 1995). Additionally, genetic distances between populations were visualized via PCA ordination using the R package "adegenet" (Jombart and Ahmed 2011) with the strength of correlation between straight-line geographic distance and genetic distance assessed for North American populations via a non-parametric Mantel test. Discriminant analysis of principal components (DAPC, Jombart et al. 2010) was also employed via the adegenet package in order to assess population clustering and the extent of admixture between populations. All statistical tests were conducted in R version 3.2.2 (R Core Team 2015).

Pseudodiaptomus inopinus migration models

Three migration models were formulated to evaluate competing scenarios of the *P. inopinus* invasion history (Fig. 2). Model 1, *the stepping-stone model*, features a stepping-stone migration pattern, in which migrants are exchanged continuously between adjacent populations, representing transport by vectors such as coastal currents, waterfowl, and/or air-borne eggs. This migration pattern is frequently associated with the invasion process, as is its resultant isolation-by-distance genetic signature (Hellberg et al. 2002).

Model 2, *the mixed-migration model* is similar to that of model 1, except that some portion of migrants (ranging from 0 to 100% of the migration pool) travel via long-distance dispersal to any other population.

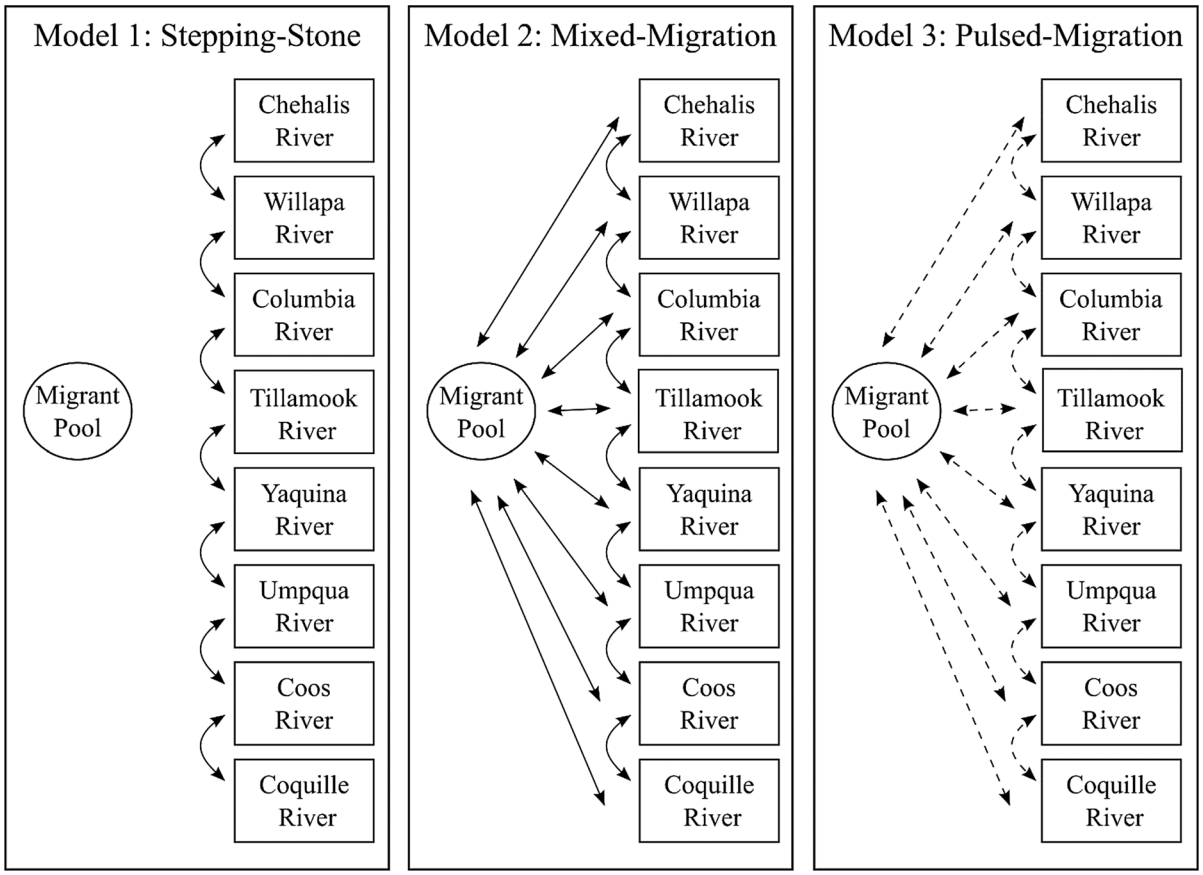

Fig. 2 Schematic view of the three simulated dispersal models. Dispersal in model 1 (stepping-stone migration) occurs every generation, but only between adjacent patches. In model 2 (mixed-migration), a randomly drawn percentage of all migrants enter a long-distance migrant pool every generation. Model 3 (pulsed-migration) progresses as model 2 except that migration only occurs in a small number of pulsed events (indicated by dashed arrows)

Depending on the value of this mixture parameter, this could represent exclusive secondary spread via the transport of ballast water (when values of the mixture parameter are high) or through a combination of ballast water transfer and natural processes (e.g., coastal currents, migrating waterfowl, and airborne eggs).

Model 3, *the pulsed-migration model*, is similar to model 2, except that migration events (both near and far) only occur during a limited number of pulses ($n = 2\ldots6$) over the duration of the simulation. In this third model, invasion occurs via a small number of chance migration events, and thus is expected to progress in a highly stochastic manner. This model represents infrequent dispersal between populations, or alternatively, frequent dispersal with low rates of survival and reproduction.

Genetic simulation

The *P. inopinus* invasion models were simulated using quantiNEMO2 v0.9.0 (Neuenschwander et al. 2008), an individual-based, genetically-explicit stochastic simulation program. Each simulation began with all habitat patches representing North America estuarine ecosystems empty and a single habitat patch (representing the Asian source population) filled to carrying capacity (1500–2000 individuals) with a population of diploid individuals possessing 1000 randomly generated, selectively-neutral SNPs at mutation/drift equilibrium. The range of simulated population size corresponds to the small effective population sizes (relative to large census sizes) observed in copepod species (Goetze 2005; Winkler et al. 2008). For example, effective population sizes for the pelagic

copepods *Calanus finmarchicus* and *Nannocalanus minor* have been observed to be seven to ten orders of magnitude smaller than census size (Bucklin and Wiebe 1998) and copepods inhabiting inland waters, such as *Eudiaptomus graciloides and E. gracilis*, have been estimated at effective population sizes of only 10^2–10^3 individuals in some lake populations (Zeller et al. 2008).

Simulations proceeded as follows: At each time-step (each representing a generation), offspring dispersed to other habitat patches according to the model-specific migration pattern. Following migration, population size within each patch was culled to carrying capacity if necessary. Populations which were under carrying capacity were not changed. At the midpoint of each simulation, the carrying capacity of the Columbia River population was reduced to zero to reflect the historical extirpation of *P. inopinus* from that system (Cordell et al. 2008; Bollens et al. 2012). At the conclusion of each simulation, individuals were randomly sampled (matching the sample size of the RADseq pool from each location) and the following summary statistics were recorded: F_{ST} (pairwise and global), expected heterozygosity (local and global), and nucleotide diversity (local and global). 100,000 independent iterations were conducted for each model with biological parameters such as migration rate, carrying capacity, generation time, and population growth rate stochastically drawn from a uniform distribution at the start of each iteration. In addition, a new set of initial genotype configurations were generated at the start of each simulation using the R package 'hierfstat' version 0.04–22 (Goudet 2005).

The range of values for the stochastic model parameters were drawn from literature specific to the invasion of *P. inopinus* when possible, and from the broader literature of copepod biology when not. The number of generations spanning the period from the introduction of *P. inopinus* to the Columbia River and the collection of genetic samples in 2012 was estimated at 30–300. The lower bound of this range represents a scenario in which only a single generation has been produced each year (Emerson et al. 2015) since establishment of *P. inopinus* in the Columbia River—at a point between 1980 and 1990 (Cordell et al. 1992). The upper limit of the range was estimated using temperature dependent copepod growth rates (Huntley and Lopez 1992) evaluated against average monthly water temperatures for the Columbia River. Values for intrinsic growth rates encompass estimates derived across a number of copepod species (Allan 1976). Simulated population size values represent estimates of effective population size and are similar to the range of values utilized in genetic studies for the simulation of demographic patterns in the copepod *Calanus finmarchicus* (Provan et al. 2009). The full range of parameter values associated with each migration model is shown in Table 1.

Model selection

The best-fit model of *P. inopinus* secondary spread was selected using an approximate Bayesian computation (ABC) approach (Beaumont et al. 2002; Csilléry et al. 2010; Lopes and Beaumont 2010). This approach evaluates summary statistics obtained from the simulation of competing scenarios across large parameter spaces, and compares them to the ones obtained from observed data. ABC selects the scenario for which simulated statistics match most closely the observed ones, and provide corresponding parameter values for the fits. These results may be summarized using Bayes factors, which are simple ratios of the number of iterations from each model that pass the selection algorithm to those that do not—evaluated pairwise for each possible combination of models. The best-fit model of *P. inopinus* secondary spread was determined using Bayes factors and a suite of genetic summary statistics (i.e., goodness-of-fit) associated with each model.

In order to validate that the chosen suite of summary statistics could reliably distinguish simulations produced from different models, a leave-one-out cross-validation procedure was performed using a simple rejection algorithm set to a tolerance of 0.05 (Arlot and Celisse 2010). This process was repeated 1000 times for each model, with a different pseudo-observed data set (the output from one simulation) randomly selected for each iteration of the validation procedure. The posterior probabilities of the three migration models given the observed genetic data were calculated as the proportion of simulations accepted under a simple rejection algorithm set to a tolerance of 0.05.

In order to assess how well simulations passing the rejection algorithm represented the overall performance of each model across parameter space, a distribution of summary statistics was estimated for

Table 1 The range of values utilized in the genetic simulation for the stochastic parameters associated with each migration model

Model	Carrying capacity	Number of generations	Population growth rate	Emigration rate	# of migration pulses	% long distance migration
1	1500–2000	30–300	0.1–0.5	0.0–0.5	1 per generation	0
2	1500–2000	30–300	0.1–0.5	0.0–0.5	1 per generation	0–100
3	1500–2000	30–300	0.1–0.5	0.0–0.5	2–6 total	0–100

each model, using a subset of 1000 randomly selected simulations per model. A goodness-of-fit test statistic was then found by calculating the median distance between this null distribution of summary statistics and the observed genetic data (Lemaire et al. 2016). All ABC analyses were conducted in R v3.2.2 (R Core Team 2015) using the "abc" package (Csilléry et al. 2012).

Results

Quality and contaminant filtering of sequence data

A total of 233,312,056 reads were produced during sequencing, of which 40,086,380 (17%) were discarded due to ambiguous genetic barcodes. Another 6,085,788 (2.6%) reads were removed due to low quality read scores and 17,370,996 (7.4%) due to ambiguous restriction enzyme cut sites, leaving 169,768,892 reads (73%) for further analyses. By sampling location, 22.3–97.8% of these reads passed the custom non-target DNA filter (constructed with the Kraken software). Upon assembling these reads into stacks of loci, the mean depth of coverage was 28.7X across an estimated 1.3 Mb of the *P. inopinus* genome.

Summary statistics and Mantel test

The genetic summary statistics for the *P. inopinus* sequence data are shown by population and by region in Table 2. A 2-tailed *t* test for unequal levels of site coverage between regions indicates that sampling effort was commensurate between native and invasive populations ($p > 0.05$). Genetic diversity, as gauged by the raw number of polymorphic sites (either adjusted for the total number of sites sequenced or presented as raw values), and mean genome-wide levels of nucleotide diversity, was significantly lower in the North American populations. The number of private alleles did not significantly differ between regions. The mean F_{ST} value obtained across all North American populations was 0.06 (SD = 0.02). The collection of all possible pairwise comparisons ranged from 0.03 to 0.09. Mantel test results indicated that genetic distances between American populations are not correlated to geographic distances between sampling locations ($p > 0.05$; Fig. 3).

PCA ordinations and DAPC clustering

A PCA ordination of the population-level genetic variance among all sites is shown in Fig. 4. The first principal component (shown on axis 1) explains 44% of total variance, and the second principal component (shown on axis 2) explains 22%. The scree plot shown in the Fig. 4 inset indicates that higher dimensional principal components (3+) do not explain much variance. Invasive North American populations of *P. inopinus* are tightly clustered together, while the two native Asian populations are widely separated across this genetic space, at points roughly equidistant from the cluster of North American populations (Fig. 4). Additionally, the ordination indicates that the *P. koreanus* outgroup is the most distantly related sample in the dataset.

A second PCA was performed with only the North American populations (Fig. 5). Here, the first principal component represents 35% of total variance (axis 1) and the second principal component (axis 2) represents 26%. The spatial configuration of samples in this ordination does not suggest a clear pattern of structure within the data, nor does such a pattern emerge when the ordination is viewed in a higher number of dimensions (results not shown). Thus, no clear pattern of structure in the North American populations could be detected among sample localities.

Table 2 Genetic summary statistics calculated from each population and averaged within regions

Population	Copepods sequenced	Covered sites	Polymorphic sites (raw)	Polymorphic sites (adjusted %)	Private alleles	Expected heterozygosity	Nucleotide diversity
Chehalis	7	650,551	1195	0.1837	562	0.0009	0.0018
Willapa	303	253,063	316	0.1249	186	0.0006	0.0011
Tillamook	80	576,463	1402	0.2432	887	0.0011	0.0019
Yaquina	71	729,205	1126	0.1544	641	0.0007	0.0014
Umpqua	80	541,018	1301	0.2405	869	0.0011	0.0018
Coos	80	415,481	654	0.1574	280	0.0007	0.0014
Coquille	80	524,875	1299	0.2475	663	0.0012	0.0021
Geum (Korea 1)	80	606,527	3229	0.5324	1196	0.0024	0.0040
Mankyung (Korea 2)	80	536,961	2297	0.4278	665	0.0021	0.0043
P. koreanus (Korea 3)	73	615,475	2167	0.3521	1155	0.0016	0.0030
Region mean							
North America	100	527,237	1042*	0.1931*	584	0.0009*	0.0016*
Korea	80	571,744	2763*	0.4801*	931	0.0023*	0.0038*

Statistics were calculated post-construction of RAD stacks. The total number of genomic sites sequenced are indicated as covered sites. Polymorphic sites (raw) indicates the total number of polymorphic sites within each population. Polymorphic sites (adjusted) are the ratio of polymorphic sites to covered sites within each population—expressed as a percentage. Private alleles are the number of alleles unique to each population. Expected heterozygosity and nucleotide diversity are the mean values considered across all genomic sites. Values marked with an asterisk are significantly different (2-tailed *t* test $p < 0.05$) between regions. *p* values have been adjusted for multiple comparisons via Bonferroni correction. Sequence data has been downsampled to account for unequal sample sizes as shown by the approximately equivalent number of covered sites across most populations

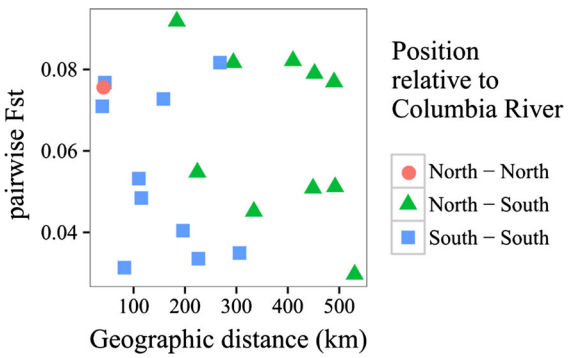

Fig. 3 Mean pairwise F_{ST} values calculated across all pairs of North American sites and plotted as function of geographic distance. Mantel test results indicate that genetic distances are not correlated with geographic distances ($p > 0.05$)

Discriminant analysis of principal components (DAPC) conducted upon the full set of populations revealed a hierarchical nesting of populations at regional scales (Fig. 6). At all values of k, the *P. koreanus* population remained distinct from all *P. inopinus* populations. Membership probabilities at k = 3 support a regional distinction between native and invasive *P. inopinus* populations, with a potential degree of greater genetic similarity between North American populations and those collected at site Korea 2 (Mankyung River). However, evidence for admixture between Korean and North American populations was not observed when a finer degree of clustering was permitted at k = 4.

Approximate Bayesian computation

Cross validation of 1000 randomly selected simulations from each model returned the posterior probability confusion matrix shown in Table 3. This confusion matrix shows the frequency with which simulations are incorrectly classified (i.e., confused) respective to the model that generated the data. The matrix indicates that on average, a randomly selected set of summary statistics would be correctly attributed to the model from which those statistics originated in 85% of model 1 simulations, 86% of model 2 simulations, and 90% of model 3 simulations. These results were robust to changes in the ABC rejection algorithm tolerance level. For the observed genetic

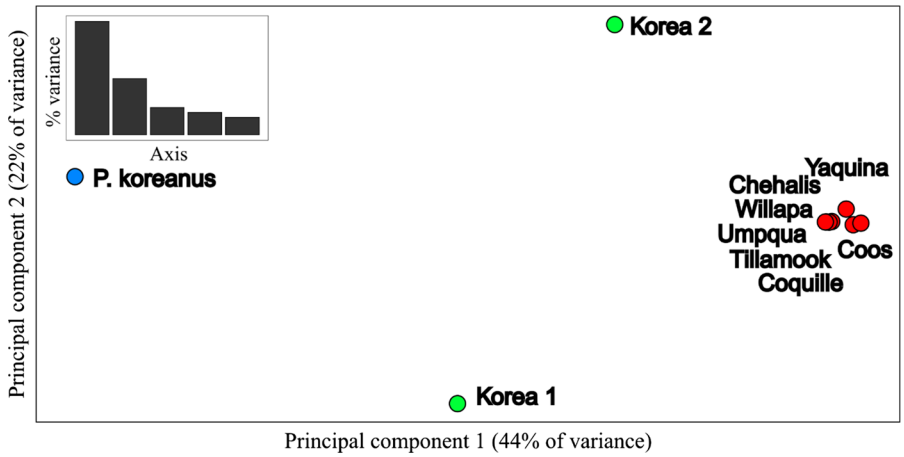

Fig. 4 PCA ordination of population-level genetic similarity among all studied populations. The length of each axis is scaled to the percentage of variation explained by the corresponding principle component. The inset scree plot shows the contribution of the first five principal components to the total variance explained. Populations from North American sites are shown in red, Korean in green, and *P. koreanus* outgroup in blue

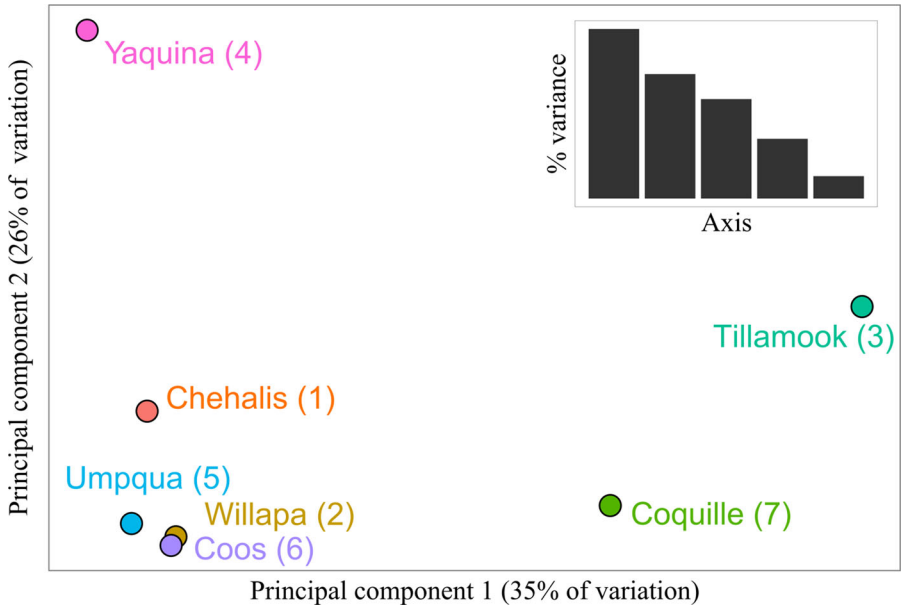

Fig. 5 PCA ordination of genetic similarity between North American populations. The length of each axis is scaled to the variation explained by the corresponding principle component. Scree plot is shown as figure inset. Populations are numbered in descending order from north to south

data, model selection rates (generated from posterior probabilities) were 44.3% for model 1, 3.2% for model 2, and 52.5% for model 3. The Bayes factor (posterior probability of model X/model Y) for each pairwise model comparison is shown in Table 4. A large Bayes factor indicates that a model (model X) provides a much better fit to the observed genetic data than the model to which it is being compared (model Y). Models 1 and 3 each produce a large Bayes factor when evaluated against model 2 (13.53 and 12.15, respectively), but are relatively equivalent when compared against each other—producing a Bayes factor of 1.11 (model 1/model 3) or 0.90 (model 3/model 1).

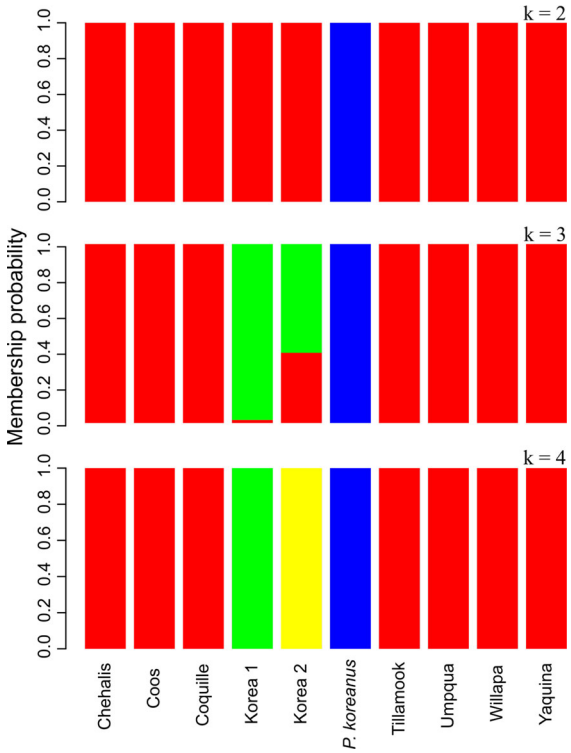

Fig. 6 DAPC generated probabilities of group membership based upon k = 2:4 genetic clusters. At all values of k, the *P. koreanus* population remains clearly distinct from all *P. inopinus* populations. Membership probabilities at k = 3 support a regional distinction between native and invasive *P. inopinus* populations, with a potential degree of greater genetic similarity between North American populations and those collected at site Korea 2 (Mankyung River). However each Korean population appears quite genetically distinct from all other populations when a finer degree of clustering is permitted at k = 4

A null distribution of summary statistics was estimated for each simulation model, and a goodness-of-fit test statistic was then calculated as the median distance between this null distribution and the observed genetic data. This distance from the distribution mean (a dimensionless test statistic) was 29.5 for model 1, 17.1 for model 2, and 2.0 for model 3. The distribution of distances associated with each model (Fig. 7) show that model 3 ("Pulsed-Migration") clearly provides a better fit to the observed genetic data than either of the two competing models.

Table 4 ABC Bayes factor for each pairwise comparison of the three models

	Model 1 (Y)	Model 2 (Y)	Model 3 (Y)
Model 1 (X)	1.00	13.53	1.11
Model 2 (X)	0.08	1.00	0.08
Model 3 (X)	0.90	12.15	1.00

Values in each cell represent the ratio of simulations passing the ABC rejection algorithm for each possible pairwise model comparison (i.e., the ratio of model X over model Y)

Discussion

Observed patterns of *P. inopinus* population structure in North America

The pattern of genetic connectivity most frequently associated with the invasion process is that of 'stepping-stone' gene flow (aka isolation-by-distance), which is defined by a correlation between genetic differentiation and geographic distance among samples (Hellberg et al. 2002), although this pattern can be obscured by complicated invasion dynamics (Frisch et al. 2013; Bayha et al. 2014). In the simplest case of a set of serial invasions, each new population arises from a subsample of the genetic diversity found in a source population. Such demographic patterns are thought to be quite common in aquatic invasions (Floerl et al. 2009) but to date only limited molecular

Table 3 Migration model cross-validation confusion matrix

	Classified as 1	Classified as 2	Classified as 3
Generated by 1	0.85	0.10	0.05
Generated by 2	0.12	0.86	0.02
Generated by 3	0.06	0.03	0.90

Generated by calculating the posterior probability of each migration model from 3000 randomly selected pseudo-observed datasets. Rows indicate which model generated the data, and columns indicate the model to which the data was classified. Values in off-diagonal cells indicate misclassification. The results indicate that on average, data were correctly attributed to the model from which they were generated for 85% of model 1 simulations, 86% of model 2 simulations, and 90% of model 3 simulations

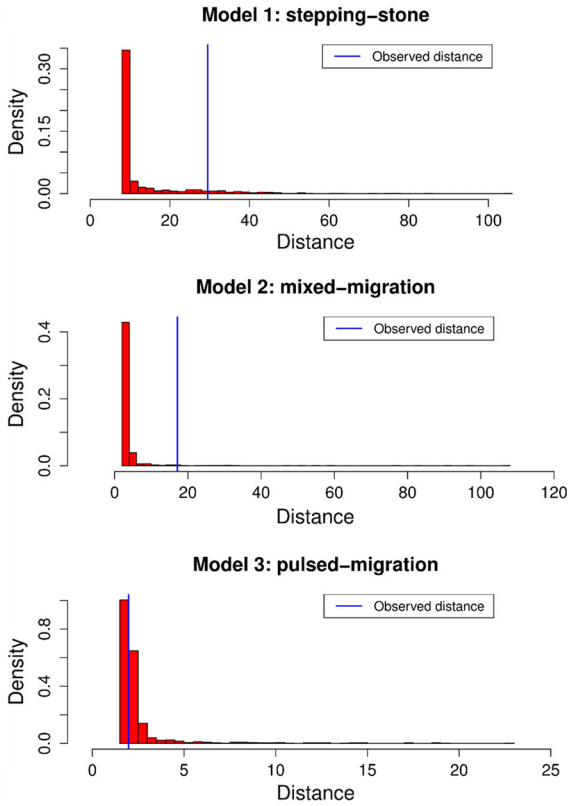

Fig. 7 Goodness-of-fit test statistic for observed data (blue) plotted against null distribution (red) of summary statistics generated by randomly selected simulations from each migration model. Scale length varies on both axes according to the distribution of values produced by each model. Note the substantially smaller range of distance values for model 3

evidence of such events exists (Lee 2000; Makino et al. 2010).

Our results do not support a stepping-stone model of population structure in North American populations of *P. inopinus*. This conclusion is supported by two lines of evidence. First, genetic and geographic distances among North American sites have little if any correlation (Fig. 3). Second, PCA ordination of the North American samples does not indicate any clear geographic structure (Fig. 5). Although this lack of coherent geographic structure may have arisen from multiple, non-mutually exclusive processes, the rejection of a stepping-stone model also dismisses some dispersal mechanisms. For example, transport in coastal currents is likely not a major factor, given the complete lack of support for an isolation-by-distance genetic structure (Fig. 3) in our *P. inopinus* data. This conclusion is supported by prior investigations of the ecological constraints of *P. inopinus*, suggesting that they may be intolerant of salinity levels found in oceanic waters (Cordell et al. 2010). Similarly, we can dismiss other diffusive processes (e.g., transport of eggs via air currents or waterfowl, or subsurface transport in groundwater) as weak agents of dispersal at best. These results mirror the geographically discordant genetic structure observed in populations of the ascidian (sea squirt) *Ciona intestinalis*, which are more heavily structured by long-distance anthropogenic transport than innate biological processes (i.e. short distance dispersal of pelagic larvae), even within the species' native range (Hudson et al. 2016).

Our results instead suggest that the North American invasion of *P. inopinus* has been structured via a series of infrequent and stochastic dispersal events. This conclusion is strongly supported by our genetic simulations and corresponding ABC analysis. At first glance, the stepping-stone model (model 1) appears to produce a slightly better fit for the observed genetic data than the pulsed-migration (model 3), with a Bayes factor (a pairwise comparison of probability) of 1.11. However, support falls strongly in favor of model 3 when overall goodness-of-fit is considered. Examination of the goodness-of-fit metric, and the underlying distribution of values from which the metric is drawn, shows that models 1 and 2 typically produce patterns of genetic structure quite unlike the patterns observed in the *P. inopinus* sequence data (Fig. 7). The simulations generated by models 1 and 2 that are retained by the ABC rejection algorithm are likely unrepresentative of those models in general, instead representing a very limited range of parameter space (Supplemental materials: Table S4). The best fit simulations generated under models 1 and 2 show extremely low levels of migration, an attribute which reflects the severely restricted pulsed migration pattern of model 3. When considered alongside the PCA ordinations and results of the Mantel tests, these findings suggest that migration by *P. inopinus* between adjacent patches occurs only in a very limited fashion, if at all.

It is worth noting that we have not identified the single "correct" model of invasion spread, but rather, identified the best fit candidate from a pool of biologically plausible models. Certain model assumptions, such as an initial introduction of *P. inopinus* to the Columbia River, have been unavoidable, although

we have attempted to evaluate model parameters across the widest biologically plausible range of values. We expect that sampling of additional North American populations would not improve our ability to resolve the invasion history of *P. inopinus*. This conclusion is supported by three lines of evidence: (1) The invasive sites showed broadly similar values for the suite of summary statistics that were examined, (2) region-specific patterns were not discernable when pairwise Fst values were examined in subsets, nor through PCA ordination, and (3) the sampled North American sites encompassed the majority of known invasive P. *inopinus* populations, and spanned the complete geographic extent of the North American invasive range. Genetic data from native *P. inopinus* populations were not included in the ABC analysis, and thus the inclusion of additional samples from the native range would not influence the analysis outcome. However, additional sampling of *P. inopinus* from a larger fraction of its native range could provide a greater context to the levels of genetic diversity found among invasive North American populations, and to the degree of genetic differentiation among putatively native populations.

Relationship of North American populations to putatively native populations

The tight clustering of the North American samples on the PCA ordination (Fig. 4) suggests that these invasive populations arose from a common source population. Given the wide ordination distance separating the two Korean populations versus the tightly clustered North American populations, it seems unlikely that the latter have disparate origins. Alternatively, the North American populations may have arisen from multiple source populations with subsequent admixture (and thus loss) of distinct genetic signatures occurring within the invaded range. However, this alternative interpretation is lacking in support given the evidence for limited gene flow within the invaded range. This single-origin scenario strongly contrasts with the pattern of invasion via multiple source populations that has been observed in a number of other crustaceans which have invaded North American waters—notably the green crab, *Carcinus maenas* (Tepolt and Palumbi 2015; Jeffery et al. 2017), the Eurasian spiny water flea, *Bythotrephes longimanus* (Colautti et al. 2005) and the cladoceran, *Daphnia lumholtzi* (Frisch et al. 2013).

The high degree of genetic similarity among samples collected from the North American range confirms morphological evidence (Soh, personal observation) that North American populations comprise only a single species. The wide distance separating North American samples from native *P. inopinus* samples on the PCA ordination raises the possibility that they represent distinct species, but morphological comparison of North American specimens against *P. inopinus* sensu stricto from the Korean range does not support this hypothesis. That the *P. koreanus* outgroup is the most distantly placed sample on the PCA ordination lends (somewhat tangential) support to recent taxonomic revision to the *Pseudodiaptomus* genus (Sakaguchi and Ueda 2010, 2011; Soh et al. 2012; Park et al. 2013), which has split *P. koreanus* from the *P. inopinus* species complex based upon morphological characteristics and genetic data (Soh et al. 2012).

Genome-wide levels of nucleotide diversity ranged from 0.0011 to 0.0043 across the populations studied, with significantly larger values observed in Korean sites. The levels of genetic diversity among the studied populations are consistent with those reported in a large number of pelagic copepod species: *Eucalanus hyalinus* (0.0026–0.1467) and *Eucalanus spinifer* (0.000–0.004) (Goetze 2005), *Calanus finmarchicus* (0–0.046) and *Nannocalanus minor* (0–0.017) (Bucklin and Wiebe 1998), and *Haloptilus longicornis* (0.0033–0.0077) (Goetze et al. 2015). As expected by the invasion process, nucleotide diversity had significantly lower values among invasive North American populations than observed in Korean populations—a clear signature of a founder event. However, the range of values observed do not reflect the magnitude of genetic bottleneck that is traditionally associated with the invasion process, and accords with findings from the field of virology that large census population sizes, skewed offspring distributions (e.g. unequal reproductive success in copepods), and multiple genetic bottlenecks may result in unexpected genetic structures (Irwin et al. 2016). As the values of nucleotide diversity observed in the North American populations are unlikely to have arisen from recent accumulation of mutations (i.e., during the last $\sim$ 30 generations which have elapsed since population founding), this result contributes to the growing line of evidence that

the established founder population was relatively large. This pattern indicates that the transport vectors mediating the introduction of *P. inopinus* to North America (i.e., ballast water) had effectively transported large numbers of individuals, and in accordance with our ABC analysis, most likely as a series of high-density pulses.

Future directions

This study raises new questions regarding the genetic structure of native *P. inopinus* populations, and illustrates several important technical points. First, the amount of non-target DNA recovered as a result of whole-organism sequencing was particularly high. We employed a bioinformatics-based approach for removal of this non-target DNA but at significant computational expense and accompanied by concomitant reduction of sequencing depth. Even at the relatively low sequencing depth cutoff of 7X, nearly all populations exhibited some missing loci due to insufficient coverage. This low coverage undoubtedly increased the degree of error in our calculation of allele frequencies, especially given the higher depths of coverage that are recommended for the sequencing of pooled samples (Schlötterer et al. 2014).

Fortunately, the genetic signature which we recovered from the North American invasion of *P. inopinus* unambiguously supported a single migration model and did not force draconian measures such as dropping lower coverage populations in order to raise the overall coverage threshold. Nonetheless, this study highlights the importance of bioinformatic and/or laboratory based methods to reduce contamination when sequencing microscopic organisms in their entirety. In such situations, an ideal solution would be alignment of sequence data against a reference genome, although overall levels of coverage will still suffer from the amplification and sequencing of non-target DNA. A promising complementary and/or alternative strategy is based upon the removal of contaminants pre-sequencing via the purging of gut contents and exterior bacterial microfilm sensu the protocol developed by I. Colson, J. Routtu and M. Dukic at the laboratory of Dr. Deiter Ebert (available at http://evolution.unibas.ch/ebert/lab/daphnia_dna.htm).

In respect to further study of *P. inopinus* specifically, our results suggest that considerable population structure occurs amongst native populations of *P. inopinus*, even across relatively short geographic distances. This result would suggest that anthropogenic transport is an effective agent of dispersal to new habitats for *P. inopinus*, but that historical barriers to dispersal remain in place within the native range. A geographically distributed sampling of populations within the native range would help clarify this apparent paradox, while simultaneously providing new opportunities to identify the source population of the invasive North American populations. Given the high risk of future exchange of aquatic invasions between these regions, the identification of historically active invasion routes should remain a major (if not key) objective of future studies. Finally, although non-selective events (e.g. founder effects, periods of low effective population size, and gene surfing at the invasion front) may obscure signatures of selection within these invasive populations, it would still be of great interest to examine these populations for enrichment of particular alleles which may be associated with positive selection during the invasion process or other known biological functions of interest.

Conclusion

Aquatic invasions have increased considerably in recent decades (Pimentel et al. 2005). Given the large number of potentially invasive species that are regularly detected in the ballast water of commercial cargo vessels (Carlton and Geller 1993; Ruiz et al. 2000; Lawrence and Cordell 2010), the introduction of exotic organisms to estuaries and inland waters will be a continued global occurrence. In order to efficiently direct funding and resources to combat the spread of aquatic invasive species, it is crucial to develop a methodological framework indicating high risk transport vectors, and to rule out vectors which are of lesser concern.

Our research incorporates two recent methodological advances (i.e., RADseq and ABC) to reconstruct the invasion history of the Asian calanoid copepod *P. inopinus* across estuaries of the North American Pacific Northwest coast. Our results indicate that invasive *P. inopinus* populations have likely arisen from a common origin, and that migration between invasive populations occurs infrequently and via

stochastic processes. Evidence from genetic simulation also supports this conclusion, as our best-fit demographic model features a low number of geographically unstructured migration pulses.

As the dispersal of *P. inopinus* has likely occurred through non-selective vectors that move bulk water, it is highly likely that similar migration processes are employed by other aquatic invaders, even globally. The stochastic pattern of long-range dispersal observed in *P. inopinus* suggests that planktonic invaders may spread across estuarine systems in a highly unpredictable manner. This pattern strongly contrasts with the diffusive spread that is characteristic of invaders such as zebra and quagga mussels (Griffiths et al. 1991; Johnson and Padilla 1996; Schneider et al. 1998). Given the significant economic and ecological importance of estuarine ecosystems, and the susceptibility of such systems to large-scale disruption by planktonic invaders, we recommend increased application of molecular-based methods to further understand the unique patterns of invasion spread in estuarine systems.

Acknowledgements The authors would like to thank Kristen Irwin for helpful comments and suggestions, as well as Alan Brelsford, Catherine Bernay, and Roberto Semier for their assistance with RADseq library preparation. Funding for this research was provided in part by the U.S. Department of State Fulbright program and the Swiss Government Excellence Scholarships for Foreign Scholars and Artists, as well as a U.S. Environmental Protection Agency STAR graduate fellowship (grant #FP91780901-0) awarded to E. Dexter and S. Bollens. Additional funding was provided by the Swiss National Science Foundation (grant #PMPDP3_158381) and a University of Lausanne FBM Interdisciplinary Grant to S. Vuilleumier, and by Washington Sea Grant for zooplankton sample collection (to S. Bollens and J. Cordell) and manuscript preparation (to G. Rollwagen-Bollens and S. Bollens). All computations were performed at the Vital-IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics (http://www.vital-it.ch) and on the University of Lausanne's GenpopBD high-performance computing cluster.

Author contributions ED conceived the research question, collected North American samples, performed DNA extraction and RADseq library preparation, conducted bioinformatics and statistical analysis, programmed genetic simulations and led composition of the manuscript. JG guided selection of molecular protocols, helped design genetic simulations, and provided access to laboratory space, high performance computing facilities and sequencing resources. JG also contributed to data analysis and composition of the manuscript. SMB helped conceive of the initial problem and research question; supervised the North American field collections; and edited and co-wrote all versions of the manuscript, from initial outline to final submission. JC supervised the North American field collections and contributed to composition of the manuscript. HYS collected samples from the Republic of Korea, provided taxonomic and morphological expertise, and contributed to the composition of the manuscript. GRB supervised the North American field collections and contributed to composition of the manuscript. SPP contributed to bioinformatics analysis and contributed to the composition of the manuscript. SV helped conceive, design, and coordinate the project, select laboratory protocols, design genetic simulations, and arrange for laboratory space and specialized training provided to ED at the University of Lausanne. SV also contributed to bioinformatics and statistical analysis as well as composition of the manuscript.

References

Allan JD (1976) Life history patterns in zooplankton. Am Nat 110:165–180

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Stat Surv 4:40–79. https://doi.org/10.1214/09-SS054

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge

Baird NA, Etter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376. https://doi.org/10.1371/journal.pone.0003376

Bayha KM, Chang MH, Mariani CL et al (2014) Worldwide phylogeography of the invasive ctenophore Mnemiopsis leidyi (Ctenophora) based on nuclear and mitochondrial DNA data. Biol Invasions. https://doi.org/10.1007/s10530-014-0770-6

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population. Genetics 162:2025–2035

Benestan L, Gosselin T, Perrier C, Sainte-Marie B, Rochette R, Bernatchez L (2015) RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). Mol Ecol 24(13):3299–3315

Blanco-Bercial L, Bucklin A (2016) New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod Centropages typicus. Mol Ecol 25:1566–1580. https://doi.org/10.1111/mec.13581

Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. BMJ 310:170. https://doi.org/10.1136/bmj.310.6973.170

Bollens SM, Cordell JR, Avent S, Hooff R (2002) Zooplankton invasions: a brief review, plus two case studies from the northeast Pacific Ocean. Hydrobiologia 480:87–110. https://doi.org/10.1023/A:1021233018533

Bollens SM, Breckenridge JK, Cordell JR et al (2012) Invasive copepods in the Lower Columbia River Estuary: seasonal abundance, co-occurrence and potential competition with native copepods. Aquat Invasions 7:101–109. https://doi.org/10.3391/ai.2012.7.1.011

Boxshall GA, Defaye D (2008) Global diversity of copepods (Crustacea: Copepoda) in freshwater. Hydrobiologia 595:195–207. https://doi.org/10.1007/s10750-007-9014-4

Breckenridge JK, Bollens SM, Rollwagen-Bollens G, Roegner GC (2015) Plankton assemblage variability in a river-dominated temperate estuary during late spring (high-flow) and late summer (low-flow) periods. Estuaries Coasts 38:93–103. https://doi.org/10.1007/s12237-014-9820-7

Brelsford A, Dufresnes C, Perrin N (2016) High-density sex-specific linkage maps of a European tree frog (Hyla arborea) identify the sex chromosome without information on offspring sex. Heredity 116:177–181. https://doi.org/10.1038/hdy.2015.83

Bron JE, Frisch D, Goetze E et al (2011) Observing copepods through a genomic lens. Front Zool 8:22. https://doi.org/10.1186/1742-9994-8-22

Bucklin A, Wiebe PH (1998) Low mitochondrial diversity and small effective population sizes of the copepods Calanus finmarchicus and Nannocalanus minor: possible impact of climatic variation during recent glaciation. J Hered 89:383–392. https://doi.org/10.1093/jhered/89.5.383

Cáceres C, Soluk D (2002) Blowing in the wind: a field test of overland dispersal and colonization by aquatic invertebrates. Oecologia 131:402–408. https://doi.org/10.1007/s00442-002-0897-5

Carlton JT, Geller JB (1993) Ecological roulette: the global transport of nonindigenous marine organisms. Science 261:78–82. https://doi.org/10.1126/science.261.5117.78

Catchen J, Hohenlohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Christy J, Stancyk S (1982) Timing of larval prodcution and flux of invertebrate larvae in a well-mixed estuary. In: Estuarine comparisons: proceedings of the Sixth Biennial International Estuarine Research Conference, Gleneden Beach, Oregon, November 1–6, 1981. Academic Press, p 489–503

Colautti RI, Manca M, Viljanen M, Ketelaars HA, Buergi H, Macisaac HJ, Heath DD (2005) Invasion genetics of the Eurasian spiny waterflea: evidence for bottlenecks and gene flow using microsatellites. Mol Ecol 14(7):1869–1879

Connelly NA, O'Neill CR, Knuth BA, Brown TL (2007) Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. Environ Manage 40:105–112. https://doi.org/10.1007/s00267-006-0296-5

Cordell JR, Morgan CA, Simenstad CA (1992) Occurrence of the Asian Calanoid Copepod Pseudodiaptomus inopinus in the Zooplankton of the Columbia River Estuary. J Crustac Biol 12:260–269. https://doi.org/10.2307/1549079

Cordell JR, Bollens SM, Draheim R, Sytsma M (2008) Asian copepods on the move: recent invasions in the Columbia-Snake River system, USA. ICES J Mar Sci J Cons 65:753–758. https://doi.org/10.1093/icesjms/fsm195

Cordell JR, Lawrence DJ, Ferm NC et al (2009) Factors influencing densities of non-indigenous species in the ballast water of ships arriving at ports in Puget Sound, Washington, United States. Aquat Conserv Mar Freshw Ecosyst 19:322–343. https://doi.org/10.1002/aqc.986

Cordell JR, Tear LM, Bollens SM (2010) Modelling physico-chemical factors affecting occurrences of a non-indigenous planktonic copepod in northeast Pacific estuaries. Biol Invasions 12:1427–1445. https://doi.org/10.1007/s10530-009-9558-5

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol 25:410–418. https://doi.org/10.1016/j.tree.2010.04.001

Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). Methods Ecol Evol 3:475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x

Cuhel RL, Aguilar C (2013) Ecosystem Transformations of the Laurentian Great Lake Michigan by nonindigenous biological invaders. Annu Rev Mar Sci 5:289–320. https://doi.org/10.1146/annurev-marine-120710-100952

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. Brief Funct Genomics 9:416–423. https://doi.org/10.1093/bfgp/elq031

Dexter E, Bollens SM, Rollwagen-Bollens G et al (2015) Persistent vs. ephemeral invasions: 8.5 years of zooplankton community dynamics in the Columbia River. Limnol Oceanogr 60:527–539. https://doi.org/10.1002/lno.10034

Drake DAR, Chan FT, Briski E et al (2014) Assemblage structure: an overlooked component of human-mediated species movements among freshwater ecosystems. J Limnol. https://doi.org/10.4081/jlimnol.2014.802

Emerson KJ, Merz CR, Catchen JM et al (2010) Resolving postglacial phylogeography using high-throughput sequencing. Proc Natl Acad Sci 107:16196–16200. https://doi.org/10.1073/pnas.1006538107

Emerson JE, Bollens SM, Counihan TD (2015) Seasonal dynamics of zooplankton in Columbia-Snake River reservoirs, with special emphasis on the invasive copepod Pseudodiaptomus forbesi. Aquat Invasions 10:25–40. https://doi.org/10.3391/ai.2015.10.1.03

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? Mol Ecol 19:4113–4130. https://doi.org/10.1111/j.1365-294X.2010.04773.x

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Mol Ecol 13:853–864. https://doi.org/10.1046/j.1365-294X.2003.02004.x

Floerl O, Inglis GJ, Dey K, Smith A (2009) The importance of transport hubs in stepping-stone invasions. J Appl Ecol 46:37–45. https://doi.org/10.1111/j.1365-2664.2008.01540.x

Frisch D, Green AJ, Figuerola J (2007) High dispersal capacity of a broad spectrum of aquatic invertebrates via waterbirds. Aquat Sci 69:568–574. https://doi.org/10.1007/s00027-007-0915-0

Frisch D, Havel JE, Weider LJ (2013) The invasion history of the exotic freshwater zooplankter Daphnia lumholtzi (Cladocera, Crustacea) in North America: a genetic analysis. Biol Invasions 15:817–828. https://doi.org/10.1007/s10530-012-0329-3

Ganz HH, Burton RS (1995) Genetic differentiation and reproductive incompatibility among Baja California populations of the copepod Tigriopus californicus. Mar Biol 123:821–827. https://doi.org/10.1007/BF00349126

Ghabooli S, Shiganova TA, Briski E et al (2013) Invasion pathway of the ctenophore Mnemiopsis leidyi in the Mediterranean Sea. PLoS ONE 8:e81067. https://doi.org/10.1371/journal.pone.0081067

Goetze E (2005) Global populatiion genetic structure and biogeography of the oceanic copepods Eucalanus Hyalinus and E. Spinifer. Evolution 59:2378–2398. https://doi.org/10.1111/j.0014-3820.2005.tb00948.x

Goetze E, Andrews KR, Peijnenburg KTCA et al (2015) Temporal stability of genetic structure in a mesopelagic copepod. PLoS ONE 10:e0136087. https://doi.org/10.1371/journal.pone.0136087

Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Griffiths RW, Schloesser DW, Leach JH, Kovalak WP (1991) Distribution and dispersal of the zebra mussel (Dreissena polymorpha) in the Great Lakes region. Can J Fish Aquat Sci 48:1381–1388. https://doi.org/10.1139/f91-165

Havel JE, Stelzleni-Schwent J (2001) Zooplankton community structure: the role of dispersal. Int Ver Theor Angew Limnol Verhandlungen 27(5):3264–3268

Havel JE, Shurin JB (2004) Mechanisms, effects, and scales of dispersal in freshwater zooplankton. Limnol Oceanogr 49:1229–1238. https://doi.org/10.4319/lo.2004.49.4_part_2.1229

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. Bull Mar Sci 70:273–290

Hudson J, Viard F, Roby C, Rius M (2016) Anthropogenic transport of species across native ranges: unpredictable genetic and evolutionary consequences. Biol Lett 12(10):20160620

Humes A (1994) How many copepods. Hydrobiologia 293:1–7. https://doi.org/10.1007/BF00229916

Huntley ME, Lopez MDG (1992) Temperature-dependent production of marine copepods: a global synthesis. Am Nat 140:201–242

Irwin KK, Laurent S, Matuszewski S, Vuilleumier S, Ormond L, Shim H, Bank C, Jensen JD (2016) On the importance of skewed offspring distributions and background selection in viral population genetics. bioRxiv 048975. https://doi.org/10.1101/048975

Jeffery NW, DiBacco C, Van Wyngaarden M, Hamilton LC, Stanley RR, Bernier R, FitzGerald J, Matheson K, McKenzie CH, Nadukkalam Ravindran P, Beiko R (2017) RAD sequencing reveals genomewide divergence between independent invasions of the European green crab (Carcinus maenas) in the Northwest Atlantic. Ecol Evol 7(8):2513–2524

Johnson LE, Padilla DK (1996) Geographic spread of exotic species: ecological lessons and opportunities from the invasion of the zebra mussel Dreissena polymorpha. Biol Conserv 78:23–33. https://doi.org/10.1016/0006-3207(96)00015-8

Jombart T, Ahmed I (2011) Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94

Knowler D (2005) Reassessing the costs of biological invasion: Mnemiopsis leidyi in the Black sea. Ecol Econ 52:187–199. https://doi.org/10.1016/j.ecolecon.2004.06.013

Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Lawrence DJ, Cordell JR (2010) Relative contributions of domestic and foreign sourced ballast water to propagule pressure in Puget Sound, Washington, USA. Biol Conserv 143:700–709. https://doi.org/10.1016/j.biocon.2009.12.008

Lee CE (2000) Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate "populations". Evolution 54:2014–2027. https://doi.org/10.1111/j.0014-3820.2000.tb01245.x

Lemaire L, Jay F, Lee I, Csilléry K, Blum MG (2016) Goodness-of-fit statistics for approximate Bayesian computation. arXiv preprint arXiv:1601.04096

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997

Lin D, Li X, Fang H et al (2011) Calanoid copepods assemblages in Pearl River Estuary of China in summer: relationships between species distribution and environmental variables. Estuar Coast Shelf Sci 93:259–267. https://doi.org/10.1016/j.ecss.2011.03.008

Lopes JS, Beaumont MA (2010) ABC: a useful Bayesian tool for the analysis of population data. Infect Genet Evol 10:825–832. https://doi.org/10.1016/j.meegid.2009.10.010

Makino W, Knox MA, Duggan IC (2010) Invasion, genetic variation and species identity of the calanoid copepod Sinodiaptomus valkanovi. Freshw Biol 55:375–386. https://doi.org/10.1111/j.1365-2427.2009.02287.x

McCormack JE, Hird SM, Zellmer AJ et al (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. Mol Phylogenet Evol 66:526–538. https://doi.org/10.1016/j.ympev.2011.12.007

Neuenschwander S, Hospital F, Guillaume F, Goudet J (2008) quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. Bioinformatics 24:1552–1553. https://doi.org/10.1093/bioinformatics/btn219

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. Annu Rev Ecol Syst 25:547–572

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle C (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol ecol 21(12):2991–3005

Park E-O, Cordell JR, Soh HY (2013) Occurrence characteristics of two sibling species, Pseudodiaptomus inopinus and Pseudodiaptomus poplesia (Copepoda, Calanoida, Pseudodiaptomidae), in the Mankyung River estuary, South Korea. Zool Stud 52:7. https://doi.org/10.1186/1810-522X-52-7

Peterson BK, Weber JN, Kay EH et al (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecol Econ 52:273–288. https://doi.org/10.1016/j.ecolecon.2004.10.002

Port Locations & Stats | PMA Pacific Maritime Association (n.d.). Retrieved 11 Dec 2017, from http://www.pmanet.org/port-locations-stats

Provan J, Beatty GE, Keating SL et al (2009) High dispersal potential has maintained long-term population stability in the North Atlantic copepod Calanus finmarchicus. Proc R Soc Lond B Biol Sci 276:301–307. https://doi.org/10.1098/rspb.2008.1062

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Ruiz GM, Fofonoff PW, Carlton JT et al (2000) Invasion of coastal marine communities in North America: apparent patterns, processes, and biases. Annu Rev Ecol Syst 31:481–531

Sakaguchi SO, Ueda H (2010) A new species of Pseudodiaptomus (Copepoda: Calanoida) from Japan, with notes on the closely related P. inopinus Burckhardt, 1913 from Kyushu Island. Zootaxa 2623:52–68

Sakaguchi SO, Ueda H (2011) Morphological divergence of *Pseudodiaptomus inopinus* Burckhardt, 1913 (Copepoda: Calanoida) between the Japan Sea and Pacific coasts of western Japan. Plankton Benthos Res 6:124–128. https://doi.org/10.3800/pbr.6.124

Schloesser DW, Nalepa TF, Mackie GL (1996) Zebra mussel infestation of unionid bivalves (Unionidae) in North America. Am Zool 36:300–310. https://doi.org/10.1093/icb/36.3.300

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals–mining genome-wide polymorphism data without big funding. Nat Rev Genet 15(11):749

Schneider DW, Ellis CD, Cummings KS (1998) A transportation model assessment of the risk to native mussel communities from zebra mussel spread. Conserv Biol 12:788–800. https://doi.org/10.1111/j.1523-1739.1998.97042.x

Shiganova TA (1998) Invasion of the Black Sea by the ctenophore Mnemiopsis leidyi and recent changes in pelagic community structure. Fish Oceanogr 7:305–310

Simberloff D (1997) Strangers in paradise: impact and management of nonindigenous species in Florida. Island Press, Washington, D.C.

Soh HY, Kwon SW, Lee W, Yoon YH (2012) A new Pseudodiaptomus (Copepoda, Calanoida) from Korea supported by molecular data. Zootaxa 3368:229–244

Tatebe H, Yasuda I, Saito H, Shimizu Y (2010) Horizontal transport of the calanoid copepod Neocalanus in the North Pacific: the influences of the current system and the life history. Deep Sea Res Part Oceanogr Res Pap 57:409–419. https://doi.org/10.1016/j.dsr.2009.11.009

Tepolt CK, Palumbi SR (2015) Transcriptome sequencing reveals both neutral and adaptive genome dynamics in a marine invader. Mol Ecol 24(16):4145–4158

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21:3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Ueda H, Kuwatani M, Suzuki KW (2010) Tidal vertical migration of two estuarine copepods: naupliar migration and position-dependent migration. J Plankton Res 32:1557–1572. https://doi.org/10.1093/plankt/fbq078

Van Wormhoudt A (2015) Seasonal and cyclical changes in genetic composition of the marine intertidal rock pool copepod tigriopus brevicornis. Biochem Genet 53:79–92. https://doi.org/10.1007/s10528-015-9674-0

Vandepitte K, Honnay O, Mergeay J et al (2013) SNP discovery using Paired-End RAD-tag sequencing on pooled genomic DNA of Sisymbrium austriacum (Brassicaceae). Mol Ecol Resour 13:269–275. https://doi.org/10.1111/1755-0998.12039

Vanderploeg HA, Liebig JR, Carmichael WW et al (2001) Zebra mussel (Dreissena polymorpha) selective filtration promoted toxic Microcystis blooms in Saginaw Bay (Lake Huron) and Lake Erie. Can J Fish Aquat Sci 58:1208–1221. https://doi.org/10.1139/f01-066

Vanderploeg HA, Nalepa TF, Jude DJ et al (2002) Dispersal and emerging ecological impacts of Ponto-Caspian species in the Laurentian Great Lakes. Can J Fish Aquat Sci 59:1209–1228. https://doi.org/10.1139/f02-087

Vuilleumier S, Goudet J, Perrin N (2010) Evolution in heterogeneous populations: from migration models to fixation probabilities. Theor Popul Biol 78:250–258. https://doi.org/10.1016/j.tpb.2010.08.004

Winder M, Jassby AD (2011) Shifts in zooplankton community structure: implications for food web processes in the upper San Francisco estuary. Estuaries Coasts 34:675–690. https://doi.org/10.1007/s12237-010-9342-x

Winkler G, Dodson JJ, Lee CE (2008) Heterogeneity within the native range: population genetic analyses of sympatric invasive and noninvasive clades of the freshwater invading copepod Eurytemora affinis. Mol Ecol 17:415–430. https://doi.org/10.1111/j.1365-294X.2007.03480.x

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

Yeo GT, Roe M, Dinwoodie J (2008) Evaluating the competitiveness of container ports in Korea and China. Transp Res Part A Policy Pract 42(6):910–921

Zeller M, Reusch TBH, Lampert W (2008) Small effective population sizes in two planktonic freshwater copepod species (Eudiaptomus) with apparently large census sizes. J Evol Biol 21:1755–1762. https://doi.org/10.1111/j.1420-9101.2008.01589.x