MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1566–1580 doi: 10.1111/mec.13581

New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod *Centropages typicus*

L. BLANCO-BERCIAL*† and A. BUCKLIN*

*Department of Marine Sciences, University of Connecticut, 1080 Shennecossett Rd, Groton, CT 06340, USA, †Bermuda Institute of Ocean Sciences, 17 Biological Station, St. George's GE 01, Bermuda

Abstract

Detection of population genetic structure of zooplankton at medium-to-small spatial scales in the absence of physical barriers has remained challenging and controversial. The large population sizes and high rates of gene flow characteristic of zooplankton have made resolution of geographical differentiation very difficult, especially when using few genetic markers and assuming equilibrium conditions. Next-generation sequencing now allows simultaneous sampling of hundreds to thousands of genetic markers; new analytical approaches allow studies under nonequilibrium conditions and directional migration. Samples of the North Atlantic Ocean planktonic copepod, *Centropages typicus*, were analysed using restriction site-associated DNA (RAD) sequencing on a PROTON platform. Although prior studies revealed no genetic differentiation of populations across the geographical range of the species, analysis of RAD tags showed significant structure across the North Atlantic Ocean. We also compared the likelihood for models of connectivity among NW Atlantic populations under various directional flow scenarios that replicate oceanographic conditions of the sampled domain. High-density marker sampling with RAD sequencing markedly outperformed other technical and analytical approaches in detection of population genetic structure and characterization of connectivity of this high geneflow zooplankton species.

Keywords: *Centropages typicus*, connectivity, population genetics, RAD tag, Zooplankton

Received 28 January 2015; revision received 24 January 2016; accepted 1 February 2016

Introduction

Migration is a fundamental property that largely determines population connectivity, genetic differentiation and demographic history, with significant impacts on the ecology and evolution of populations and species. Furthermore, dispersal of individuals is key to controlling a population's ability to adapt to changes and colonize suitable areas (Kokko & López-Sepulcre 2006). Holozooplankton have shown significant genetic differentiation over a wide range of geographical scales (Cowen *et al.* 2007; Goetze 2005; Hellberg 2009), including within the North Atlantic Ocean (Blanco-Bercial *et al.* 2014; Patarnello *et al.* 2010; Peijnenburg *et al.* 2006; Unal & Bucklin 2010; Yebra *et al.* 2011). However, most (if not all) of the findings of significant genetic structuring have involved the presence of hydrographic or geological barriers (e.g. straits, biogeographical boundaries).

Population genetic studies of zooplankton to date have in general used mitochondrial DNA (but see, e.g., Provan *et al.* 2009; Unal & Bucklin 2010). More recently, next-generation sequencing (NGS) has allowed rapid development of novel population genetic markers with no need of previous knowledge on the genomics of the organism (Baird *et al.* 2008). These advances have allowed studies of nonmodel organisms (Helyar *et al.* 2011) and improved our understanding of marine ecosystems with studies that were not possible with classical methods (Reitzel *et al.* 2013). Analysis of hundreds to several thousands of markers per individual

Correspondence: L. Blanco-Bercial, Fax: +1 441 297 8143; E-mail: leocadio@bios.edu

has also allowed the use of more sophisticated tools to address questions of connectivity between populations. Due to nearly universal application in population genetic studies, hierarchical analysis of variance using Wright's *F*-statistics relatives (Excoffier *et al.* 1992) provides useful benchmarks for comparisons among species, regions and environments. However, there are assumptions for *F*-statistics that are surely not met for zooplankton, including genetic equilibrium conditions, symmetrical migration and stable populations (Hellberg 2009). The usefulness of *F*-statistics is further limited by the very large population sizes of many holozooplankton species, which result in relatively larger confidence intervals for the resultant very small *F* values (Waples 1998) and lack of statistical significance for high gene-flow species (see Waples *et al.* 2008; especially when applied on a single locus or on a limited number loci. The most commonly used methods for estimating migration are based on the consequences of past dispersal events under existing environmental conditions, which thus provide indirect estimations of real dispersion and demographic connectivity (Broquet & Petit 2009). More recently developed methods focus on the direct measurement of demographic connectivity by assigning individuals to their immediate population of origin and parentage (see review by Broquet & Petit 2009). These genetic analyses account for all dispersed individuals, not their effect on the frequencies of the genotypes of the population, and they are more comparable to the concept of demographic connectivity. These analyses, however, are not suitable – or have very limited power – when only one or few markers are used (Broquet & Petit 2009; Wilson & Rannala 2003).

The copepod *Centropages typicus* Krøyer 1849 is one of the key species of zooplankton communities on the continental shelf in the temperate North Atlantic Ocean (Beaugrand *et al.* 2007; Bonnet *et al.* 2007; Carlotti & Harris 2007). The species' distributional range includes the Mediterranean Sea, NE Atlantic Ocean from Gibraltar Strait to the Norwegian Sea, and NW Atlantic Ocean from the Mid-Atlantic Bight to the Gulf of Saint Lawrence and the Grand Banks (Beaugrand *et al.* 2007; Johnson *et al.* 2010; Pepin *et al.* 2011; Continuous Plankton Recorder Survey Team 2004). In addition, *C. typicus* has been reported outside these regions as far as the SE Pacific Ocean (Razouls *et al.* 2005-2014). Across this distributional range, there are noticeable differences in the phenology of the species, with a latitudinal shift in the timing of maximum abundance. During winter, the species distribution is limited to the southernmost edges of the range (south of the North Sea, Bay of Biscay in the NE Atlantic; Mid-Atlantic Bight and Southern New England shelf in the NW Atlantic). Population densities increase in spring and are highest in summer in the NE Atlantic; in the NW Atlantic, populations multiply in late spring and summer, with maximum abundances recorded during the fall (Beaugrand *et al.* 2007; Bonnet *et al.* 2007; Durbin & Kane 2007; Grant 1988). Despite the apparent isolation between the NE and NW continental shelf ecosystems, *C. typicus* can also be found in low numbers in oceanic areas, likely expatriated from the shelf edge (Beaugrand *et al.* 2007). This finding suggests the possibility of contact between the various coastal regions, although the consequences for genetic connectivity are unclear. *Centropages typicus* has shown apparent panmixia based on mitochondrial markers, while nuclear ribosomal data suggest some differentiation between the NE and the NW Atlantic shelf ecosystems (Castellani *et al.* 2012). These apparently contradictory results could be due to the different nature of these markers (e.g. the maternal-only inheritance and lack of recombination of mitochondrial loci), or simply to the use of single loci, since discrepancies in results using different single genetic markers are not uncommon (Toews & Brelsford 2012). In contrast to the single-locus or few-loci approaches, RAD sequencing can provide thousands of independent markers across the entire genome (Baird *et al.* 2008; Davey & Blaxter 2010).

Centropages typicus shows a complex pattern of population responses to variability in environmental conditions, including both short- and long-term forcings. The abundance of this species within a locality appears to be correlated with mesoscale temporal and spatial conditions, especially food availability (Carlotti *et al.* 2007, 2014; Durbin & Kane 2007; Stegert *et al.* 2012) and likely under a top-down (mortality) control (Ji *et al.* 2013). Along a latitudinal gradient, populations show signs of local adaptation to temperature (see reviews by Carlotti *et al.* 2007; and Gaudy & Thibault-Botha 2007), which is likely a strong factor influencing the evolution of populations (Stegert *et al.* 2012). Population models of the NW Atlantic shelf ecosystem suggest that *C. typicus* populations will likely be favoured in the ongoing and projected increase in temperatures in the region (Stegert *et al.* 2012).

Significant long-term changes in *C. typicus* populations may have a considerable effect on the marine ecosystem of the temperate North Atlantic, since *C. typicus* plays a key role in these shelf ecosystems. This species is able to exploit a wide spectrum of prey, ranging from small phytoplankton to fish larvae (Calbet *et al.* 2007; Turner *et al.* 1985). This may allow the species to flourish under postspring bloom conditions, and avoid competition with other ecologically important and abundant copepods (e.g. species of *Calanus* or *Pseudocalanus*) linked to the spring bloom. This ability to exploit a wide range of prey might also facilitate the

survival of C. *typicus* in the open ocean, where phytoplankton-based resources are limited (Calbet *et al.* 2007) and allow potential connection across the North Atlantic Ocean. The projected increase in *C. typicus* populations linked to increasing temperatures (Stegert *et al.* 2012) could also have a significant impact on ecosystems and fisheries, especially if combined with a decrease in the abundances of other copepod species, such as *Pseudocalanus* spp. (Erikson *et al.* 2014; Kane 2014), which are the preferred prey of these economically important fish species (Petrik *et al.* 2009). *Centropages typicus* is not, however, a preferred prey item of haddock and cod larvae (Petrik *et al.* 2009).

In this study, the population genetic structure and connectivity of *C. typicus* is revisited using a restriction site-associated DNA (RAD) tag technique. The specific objectives of this study include analysis of population genetic structure across the North Atlantic and population connectivity along the NW Atlantic continental shelf ecosystem, from the Mid-Atlantic Bight to the Gulf of Maine. Our purpose was to test the power of RAD markers to reveal new understanding of marine ecosystem dynamics and population connectivity of zooplankton species characterized by high rates of gene flow and individual dispersal, lack of apparent hydrographical or geological barriers, and large population sizes, which have posed challenges for the resolution of population genetic structure and accurate estimation of migration.

Material and methods

Zooplankton samples containing *Centropages typicus* were collected during November 2011 onboard the RV *Delaware,* as part of the NOAA Northeast Continental Shelf Ecosystem Monitoring (EcoMon) Program, with sampling locations in the Mid-Atlantic Bight (MAB), Southern New England Shelf (SNE), Georges Bank (GB) and Gulf of Maine (GoM) regions. The NE Atlantic was represented by samples from the North Sea (NS), where specimens were obtained in October 2010 at the Helgoland Roads station, and from the Bay of Biscay (BB), where a sample was taken during the RADIALES programme in September 2012 (Fig. 1; Table 1). Samples were preserved in 95% undenatured ethanol and stored at room temperature or 4 °C.

DNA extraction and library preparation

Total DNA was extracted from individual specimens using the E.Z.N.A.® Mollusc DNA Kit (OMEGA, Norcross, Georgia), following the procedure indicated for arthropods, with the modification of a 14-min centrifugation at 10 000 ***g*** during the phase separation step. Although contamination from prey DNA contents is not possible to be absolutely discounted, the possibility that this influenced the results is deemed to be extremely unlikely. Apart from the fast digestion rates, only prey items simultaneously present in significant amounts in almost all individuals from all populations (see genotyping filtering) would have had any impact. For final elution, two consecutive elution steps were carried out, with each step using 50 μL of Ambion® nuclease-free water (NFW) preheated to 70 °C, and columns were incubated at 65 °C for 5 min before centrifugation. DNA concentration was measured using the Qubit® dsDNA HS Assay Kit (Invitrogen™). DNA extractions were transferred to 0.2-mL vials and dried in a vacuum using a Savant SpeedVac system. The dry DNA was then resuspended overnight in 8 μL of NFW at 4 °C.

For the 2b-RAD (Wang *et al.* 2012) library preparation, we followed the two-PCR method therein, with modifications to adjust it to the small DNA yield per individual (Fig. 2). No verifications of aliquot DNA concentrations were carried out at intermediate steps, to avoid DNA loss.

The type IIB restriction enzyme chosen for the experiment was BsaXI (New England BioLabs), which produces fragments of 27 bp with 3-base 3′ overhangs (https://www.neb.com/products/r0609-bsaxi). Digestion reactions were carried out in a total volume of 12 μL, using 8 μL resuspended DNA, 4 units BsaXI and 1.2 μL 10× NEBuffer 4 for 3 h. After 3 h, 13 μL ligation

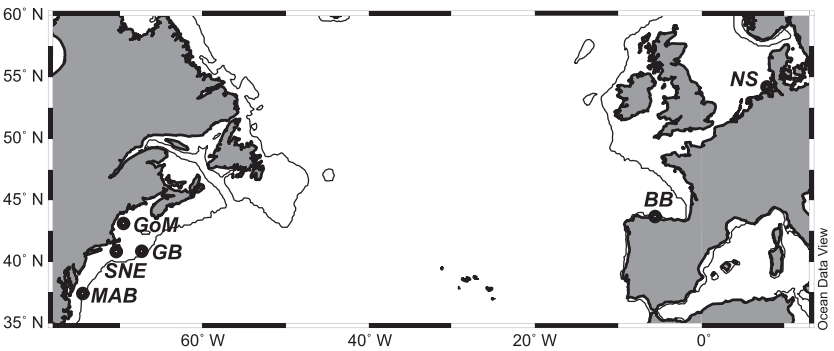

Fig. 1 Map with the stations used for this study. See text and Table 1 for abbreviations and details.

Table 1 Sample names and locations, number of individuals per sample and cruise and station data. *N* indicates numbers of individuals analysed; project names are explained in the text

Region and Sample	*N*	Cruise / Project	Date	Latitude	Longitude
NW Atlantic					
Mid-Atlantic Bight	12	DE1109 / EcoMon	8-Nov-2011	37°33.8′N	74°47.2′W
Southern New England	12	DE1109 / EcoMon	12-Nov-2011	40°48.9′N	70°21.5′W
Georges Bank	9	DE1109 / EcoMon	13-Nov-2011	40°50.8′N	67°17.4′W
Gulf of Maine	12	DE1109 / EcoMon	17-Nov-2011	43°06.6′N	42°58.2′W
NE Atlantic					
Bay of Biscay	6	RADIALES	March-2012	43°40.5′N	5°34.8′W
North Sea	10	Helgoland Roads	18-Oct-2010	54°11.3′N	7°54.0′E

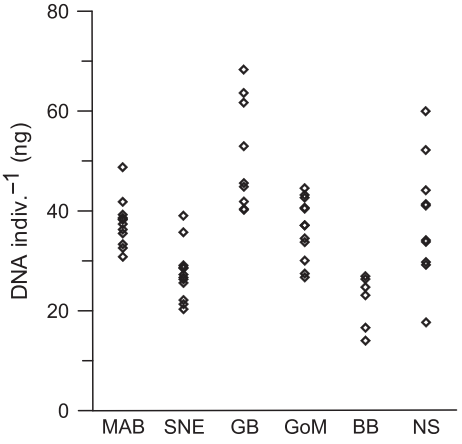

Fig. 2 Extracted DNA per individual. Individuals from different stations were mixed across the different extraction batches to minimize batch-related artefacts. The observed differences between stations are therefore likely due to between-sample heterogeneity in DNA content per individual.

Table 2 Adaptors and PCR primer sets for the library preparation. In the Ion PA-Key Adapter, bold indicates the *key*, and the barcode (BC) includes from 10 to 12 unique nucleotides

Name	Sequence
Adaptors (5′ - 3′)	
Ion-rtr-P1-BsaXI	CTCTCTATGGGCAGTCGGTGATNNG (invT)ATGGGCAGTCGGTGAT
Ion-rtr-PA-BsaXI	CTGCTGTACGGCCAAGGCGAANNG (invT)ACGGCCAAGGCGAA
PCR primer sets (5′ - 3′)	
1st PCR	
Ion trP1	CCTCTCTATGGGCAGTCGGTGAT
Ion PA PCR	CTGCTGTACGGCCAAGGCGAA
2nd PCR	
Ion trP1	CCTCTCTATGGGCAGTCGGTGAT
Ion PA-Key Adapter	CCATCTCATCCCTGCGTGTCTCCGA C**TCAG**(BC)GATCTGCTGTACGGCCA AGGCGAA

master mix containing 0.5 μL 10 mM ATP (TEKNOVA), 800 U T4 DNA ligase (New England BioLabs), 2.5 μL of each adaptor (4 μM; Table 2) and 3.5 μL NFW was added to 12 μL digested DNA and incubated at 4 °C for 16 min. The size of the genome for *C. typicus* is unknown, and the known genome size for several other calanoid copepods from the same superfamily (Centropagoidea) is highly variable, with C-values ranging from 0.63 to 5.71 (Gregory 2013). The predicted density of BsaXI sites (assuming a Poisson distribution and an idealized 50% GC content) would produce 244 cuts per Mb. To reduce the explored density to 1/16th, both adaptors included less degenerate cohesive ends 5′-NNG-3′. This reduced the sites to be sequenced to only those with a 'C' on the corresponding end, for a reduction to ¼th per adaptor, or 1/16th overall. This reduced the risk of sampling too many sites, which can result in insufficient depth of coverage at each site (Wang *et al.* 2012).

Ligation products were amplified in 50 μL PCRs, each containing 25 μL of the ligation product, 1.5 μL of each primer (1st PCR; Table 2) at a concentration of 10 μM, 6.25 μL of 2.5 mM dNTPs, 0.5 μL (1 U) of Phusion® high-fidelity DNA polymerase (New England Biolabs), 10 μL of Phusion® HF buffer (5X) and 5.25 μL of NFW. The PCR was carried out in a GeneAmp® PCR System 9700 (Applied Biosystems) and entailed 22 cycles of 98 °C for 5 s, 60 °C for 20 s and 72 °C for 10 s, followed by a final extension of 72 °C for 10 min. The potential impact of PCR duplicates on the 2b-RAD protocol is minimized due to the identical and short length of the fragments (33 bp). This minimizes the bias due to the differences in CG/AT content between alleles; there is no bias due to differences in length between fragments and alleles. Furthermore, priming sites are identical. These three facts together would minimize PCR duplicates bias (Puritz *et al.* 2014), although complete avoidance cannot be guaranteed. The PCR products were run on 2% agarose gels in TBE; bands of 77 bp

were cut and gel-purified using the QIAquick Gel Extraction Kit (QIAGEN) following the manufacturer's instructions, with an elution volume of 35 μL. Individual barcodes were introduced by means of five additional PCR cycles, using 30 μL of the purified DNA under the same PCR conditions, replacing the Ion PA PCR primer for the barcode-bearing Ion Proton adaptors (Table 2). The resultant PCR products were 120–122 bp and were purified following the same procedure using the QIAquick Gel Extraction Kit in 50 μL. An additional purification step using the Agencourt® AMPure® XP system (Beckman Coulter) was carried out following the manufacturer instructions, with elution in 40 μL TE buffer.

Size, quantity and quality of each individual library were measured on the Agilent 2100 Bioanalyzer (Agilent Technologies) using the Agilent DNA 1000 Kit. The concentration of the PCR product of interest was calculated for each library using the AGILENT 2100 EXPERT software (Agilent), and 100 μL of a 26 pM equimolar mixture of the fragment was prepared. The pooled library was sequenced on an Ion Proton™ (Life Technologies) platform using an Ion PI™ Chip Kit ver. 2; sequencing was carried out at the Center for Genome Innovation (http://cgi.uconn.edu/). Ion sphere particle (ISP) emulsion PCR amplification was carried out on an Ion OneTouch™ 2 using the Ion PI Template OT2 200 Kit (Life Technologies). After quality control, 75% of the ISPs carried a template, so the emulsion PCR was repeated with 1:3 diluted template, which reduced the bead loading to <30% (closer to the optimum range of 10–25%). To control the quality of the final output, a test fragment TF_C (Life Technologies) was added to the pooled library before the emulsion PCR.

RAD tag genotyping

After the removal of polyclonal, primer–dimer and low-quality reads using the TORRENT SUITE Ver. 3.4.2 (with default parameters), the sequences were analysed using BsaXI-adapted Perl scripts from Wang *et al.* (2012) and the RAD-specific software STACKS (Catchen *et al.* 2013, 2011). First, we extracted the fragments of interest for each individual using the 2b_Extract.pl (from Meyer Lab http://people.oregonstate.edu/~meyere/index.html). Since the BsaXI recognition site is not palindromic, a second extraction was performed on the reverse complement of the sequences, and both resulting files containing the RAD tags were concatenated at the individual level. The number of selected fragments for each extraction was compared to confirm and validate the expected subequal distribution. Adaptor ligation sites were excluded from each read to eliminate errors during the ligation and first PCR steps; therefore, all RAD tags were 27 bp long. This preselection removed a large number of sequences that (although incomplete) could have increased the depth of coverage of some of the markers, but could also have reduced the reliability and quality of the sequences.

Files containing all RAD tags for all individuals were then analysed in STACKS Ver. 1.02 to 1.08 (Catchen *et al.* 2013) for the *de novo* assembly. Initially, all sequences were processed in *ustacks*, which takes as input a set of short-read sequences from a single individual and aligns them into exactly matching stacks. Comparing the stacks, a set of loci is formed and SNPs are detected at each locus using a maximum-likelihood framework (Hohenlohe *et al.* 2010). The minimum depth of coverage to create a stack was set at three sequences. To account for the short nature of our sequences, which would severely limit the expected number of differences within a single locus, the maximum distance allowed between stacks was two nucleotides, and the maximum number of stacks allowed per *de novo* locus was three. The Deleveraging algorithm, which resolves overmerged tags, and the Removal algorithm, which drops highly repetitive stacks and nearby errors, were enabled. The alpha value for the SNP model was set at 0.05. *Cstacks* was used to build a catalog of consensus loci containing all the stacks (loci) from all the individuals and merging all alleles together. Then, each individual genotype was compared against the catalog using *sstacks.* Finally, the *Populations* program was used to obtain the loci that were present in all samples in at least 80% of the individuals from each sample, with at least five RAD tags per allele at each locus (5X coverage per allele). To avoid linkage bias for the SNP calling, only the first SNP per locus was included in the final analysis. Additional analyses were carried out to explore and evaluate the effect of higher minimum threshold for coverage (10X), which also decreased the number of markers. These results are reported in the Supplementary Data.

Population genetic analysis

File conversion from the Stacks output files to the desired formats was carried out with PGDSPIDER Ver. 2.0.4.0 (Lischer & Excoffier 2012).

Prior to population genetic analysis, markers under selection were identified using two different analyses: BAYESCAN Ver. 2.1 (Foll & Gaggiotti 2008) and the hierarchical island model described by Excoffier *et al.* (2009) as implemented in ARLEQUIN Ver. 3.5 (Excoffier & Lischer 2010). BAYESCAN uses a Bayesian approach to estimate the probability that each locus is under selection, taking into account populations and incorporating uncertainties in allele frequencies due to small sample

sizes. This analysis has been shown to be more reliable than alternative methods in detecting outliers and reducing the number of false positives (Narum & Hess 2011; Pérez-Figueroa *et al.* 2010). BAYESCAN was run under default settings, except for a modification to increase the number of pilot runs to 40. The False discovery rate (FDR) was set at 0.1. The hierarchical island model method implemented in ARLEQUIN accounts for hierarchical structure of the populations, in which migration is higher within than between groups. The groups were the same ones established for the AMOVA analyses (see below). We explored the impacts on the results of different *P*-values (0.005 and 0.001) and the starting parameters (number of simulations 150 000–250 000; number of groups 10–20; demes per group was fixed at 100). The loci that consistently appeared to be under selection under the various starting parameters and iterations were reported. A complete summary of the results can be found in the Supplementary Data.

Another method, LOSITAN (Antao *et al.* 2008), uses the F_{ST}-outlier method to detect loci under selection, which show higher divergence than the expected under a neutral model (Beaumont & Nichols 1996). The analysis was run for 1 000 000 simulations under the infinite sites model. However, this method gave inconsistent results after repeated analyses and identified many tens to several hundreds of loci as subject to both balancing and positive selection; in many instances, the same locus changed from one category to another in different runs. Inconsistencies using this method have been noted by other authors (e.g. Poelstra *et al.* 2013); these might be due to the low level of differentiation between our samples (average $F_{ST} = 0.022$). Results from this approach are thus not reported, and analyses are based on the two previous methods.

Loci under positive selection were removed from the analyses; loci identified as subject to balancing selection were included and treated as neutral, due to published reports of the unreliability of the methods used in this circumstance (Beaumont & Balding 2004; Lotterhos & Whitlock 2014; Narum & Hess 2011).

ARLEQUIN Ver. 3.5.1.2 (Excoffier & Lischer 2010) was used to determine F_{ST} distances and characterize genetic structure and variance within and between the NW and NE Atlantic subgroups by hierarchical locus-by-locus analysis of molecular variance (AMOVA; Excoffier *et al.* 1992). Significance was tested after 1000 permutations. Genetic differentiation was also estimated with G_{ST} (Nei 1973) and D_{est} (Jost 2008) using GENALEX Ver. 6.5 (Peakall & Smouse 2012); significance was tested after 999 permutations. Total and population-level heterozygosity were calculated in ARLEQUIN Ver. 3.5.1.2 (Excoffier & Lischer 2010).

The potential number of genetic clusters and the membership of each individual were estimated using STRUCTURE Ver. 2.3.4 (Pritchard *et al.* 2000). The software uses Markov chain Monte Carlo (MCMC) simulations to estimate those parameters, with the number of clusters to be tested (*K*) specified by the user. The MCMC simulation was run for 500 000 repetitions, after a burn-in period of 500 000. No admixture model was assumed, and the location information was included as a prior, since these data help detect the true structure, despite weak signal and in small-to-medium data sets (Hubisz *et al.* 2009). Fifteen replicates of *K* values from 1 to 10 were evaluated. The optimal *K* value was selected based on the *L(K)* values, the individual assignment patterns, using STRUCTURE HARVESTER Ver. 0.6.93 (Earl & vonHoldt 2012), which assesses the likelihood value at each *K* and selects the optimal value using an ad hoc statistic ΔK, which is based on the rate of change in the log probability of data between successive *K* values (Evanno *et al.* 2005). The results from 15 replicates of the selected *K* values were summarized into a single result and were then aligned and analysed in CLUMPP Ver. 1.1.2 (Jakobsson & Rosenberg 2007) using the Greedy algorithm and 1000 repeats. The results were then visualized in DISTRUCT Ver. 1.1 (Rosenberg 2004). Test runs using other priors (admixture model or without location information) failed to detect any structure.

NW Atlantic shelf population connectivity

Population connectivity patterns were estimated as migration rates between populations in MIGRATE-N Ver. 3.6.4 (Beerli 2006). Since the disparity between the temporal and spatial scales of connectivity across the N Atlantic basin vs. within the NW Atlantic shelf ecosystem would limit the power of hypothesis testing of migration pathways, we have focused our analysis on the sampled NW Atlantic regions. The whole RAD tag sequences at all loci (27 bp) were used for a multilocus analysis. The inclusion of the entire 27-bp sequence reduced the possible ascertainment bias associated with the selection of a single nucleotide (i.e. SNP-based approaches). Also, the evolutionary model implemented in MIGRATE-N is better suited to DNA sequences than to SNP data (Beerli 2012). To balance between an adequate sampling and the computational effort needed, 200 random tags (either containing or not containing a SNP site; see Supplementary data) were selected from the total pool. For each analysis, initial runs (with prior parameters based on estimated F_{ST}) were used to establish the priors for the mutation-scaled migration (M) and mutation-scaled population size (θ) parameters. After the test runs, the settings for Migrate-N were established as follows: maximum θ value = 1, maxi-

mum $M = 40\,000$, and metropolis sampling every 1000 steps for 10 million generations (10 000 recorded steps), with a burn-in period of 20 million steps. Four parallel chains were run under a static heating scheme (using the default heating scheme) and a swapping interval of 1. Each analysis (including the preliminary ones) was run twice to ensure consistency.

Patterns of migration were calculated following the procedure described by Beerli & Palczewski (2010). In short, this procedure compares and ranks the marginal likelihood between different models of gene flow among the populations and calculates their specific probability based on the Bayes factor. This procedure has the advantage of being very robust to circumstances of incomplete sampling of populations and non-normality (Beerli 2004; Beerli & Palczewski 2010). The tested connectivity models were chosen based on the hydrography of the region (Fig. 3). In all, six models were evaluated: (i) panmixia, with all samples belonging to a single population; (ii) full migration, with all migration paths open (asymmetric gene flow allowed); (iii) north to south, with gene flow following the main current flow on the system; (iv) south to north, with gene flow opposed to the main current – this model was added to evaluate the effect of reducing the number of parameters to estimate and to ensure that an improvement on the likelihood would not be due to spurious reasons; (v) adjacent, with migration paths between adjacent zones only (asymmetric gene flow allowed); and (vi) Gulf Stream, with northward displacement from MAB to SNE and GB, based on published particle modelling indicating the suitability of this route (Zhang *et al.* 2015).

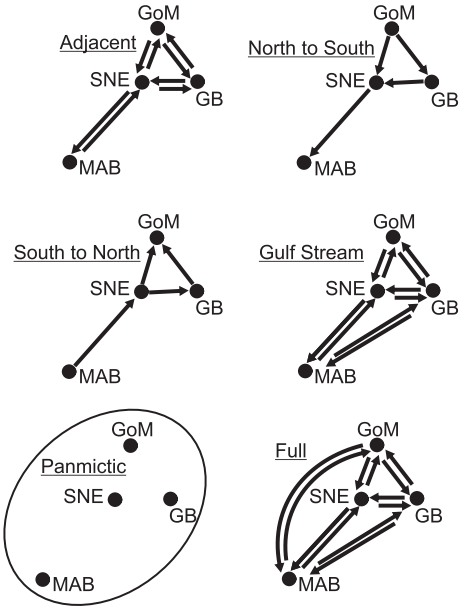

Fig. 3 Geneflow models compared in MIGRATE-N. The likelihood of each model was compared based on Bayes factors (ratios of marginal likelihood). See text for detailed description of methods used.

Results

RAD tag genotyping

There were differences among the samples analysed in terms of DNA obtained per individual (Fig. 2); however, this did not affect the success of the procedure. The pooled libraries from the 61 individuals produced 88.4 million RAD tags (68% of the capacity of the Ion PI™ Chip Kit Ver. 2); 48 million reads carried perfect barcodes and passed the quality control filter. The mean number of reads per individual was 787 614 (SD = 735 763; range 234 798–4 746 599). Of these, an average of 34% (SD = 7%; range 21–50%) carried the complete 27-bp RAD tag. Since the adaptor ligation was nondirectional and the enzyme target was not palindromic, we verified that the amount of extracted fragments from each direction of the target fragment was not statistically different from a 50/50 proportion for each individual (paired *t*-test, $P > 0.05$).

The average number of stacks per individual with a minimum 3X coverage was 10 594 (SD = 4974; range 5049–31 516), with mean coverage per stack ranging 10–21X per individual. The final catalog, after merging all stacks from each individual, contained 77 533 loci, of which 25 102 contained at least one SNP (90.6% biallelic, 7.6% triallelic, 1.8% tetrallelic). A total of 675 loci with at least one SNP met the criterion of being present in at least 80% of the individuals for all populations at >5X coverage per allele; these were included in the population analyses.

There was a decrease in the values and significance levels of indices of genetic differentiation when more restrictive allele coverage conditions were considered (10X), most likely due to the marked decrease in the number of markers available (184), although there was no effect on the general pattern of results (Supplementary File S1). Further analyses focused on the impact of the number of SNPs are included in the Supplementary File S1.

Population genetic analysis

Four SNP loci were identified as candidates for diversifying selection in BAYESCAN among the 675 loci selected using FDR of 0.1 (Fig. 4); the equivalent analysis in ARLEQUIN flagged 5 to 18 loci as candidates for positive selection, depending upon the chosen value of α (0.001 and 0.005). All four markers detected by BAYESCAN were also identified by ARLEQUIN. Additional markers identi-

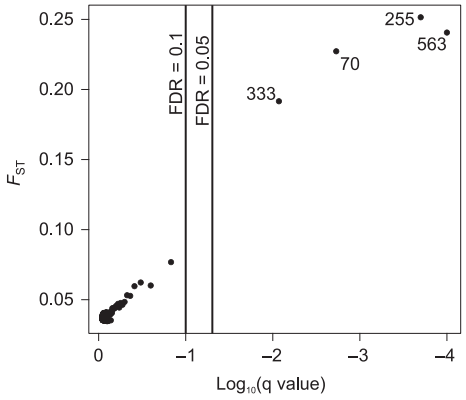

Fig. 4 BAYESCAN plot of the 675 SNP loci selected for the analyses. For each locus, F_{ST} value is plotted against the $\log_{10}$ of its q-value (the FDR analogue of the *P*-value). The vertical bars indicate the thresholds for the FDR = 0.1 and FDR = 0.05 values used in our study to identify outlier markers (on the right side). These markers had a posterior probability of selection of PP > 0.97. All the other markers showed a PP < 0.30.

fied by ARLEQUIN were not consistent between replicate runs, even under the same starting parameters, with variation in both the number and the identity of the markers flagged. Removing all flagged markers caused a general decrease in the significance of the analyses, although the results were consistent with those obtained after removing only the four markers detected in both analyses, despite the reduction in statistical power (Supplementary File S1). The definitive results shown are from analyses carried out after removing the four markers identified by both BAYESCAN and ARLEQUIN.

Pairwise genetic distances between populations showed agreement among the three statistical approaches employed, with all significant differences between pairs of populations resulting from comparisons between samples from different sides of the North Atlantic Ocean (Table 3). No significant differences were found between samples within the NW Atlantic, or between the Bay of Biscay and the North Sea. Mean heterozygosity levels were very similar in all populations (range 0.087–0.114) and in total (0.094); therefore, local events that could have caused deviations from Hardy–Weinberg equilibrium (e.g. local population bottlenecks) are unlikely.

Analysis of molecular variance comparing the NW Atlantic vs. the NE Atlantic revealed that a significant fraction of the total variance was due to between-group variance (3%; F_{CT} = 0.030; P < 0.0001); most of the variance was due to variation within populations (97%, F_{ST} = 0.029; P < 0.0001). This difference between the NE and NW Atlantic was corroborated by Bayesian clustering, which reflected the presence of three clusters (K = 3; Fig. 5) with a clear divide between the NW and NE Atlantic (Fig. 5). The isolation between the two groups was not complete, and some individuals from both groups showed mixed origin. The proportional contributions of each of the dominant genetic pools to the other domain were 0.0089 (NE into the NW) and 0.0028 (NW into the NE).

Table 3 F_{ST}, G_{ST} and D_{est} distances between samples (below diagonal) and corresponding p-value (above diagonal). In bold, significant distances after Bonferroni correction. See text for the explanation of abbreviations for each region

	MAB	SNE	GB	GoM	BB	NS
F_{ST}						
MAB		0.280	0.088	0.554	<0.001	<0.001
SNE	−0.002		0.386	0.885	<0.001	<0.002
GB	0.001	−0.004		0.378	<0.001	<0.001
GoM	−0.007	−0.011	−0.006		<0.001	0.002
BB	**0.039**	**0.035**	**0.037**	**0.029**		0.551
NS	**0.027**	**0.019**	**0.030**	**0.017**	−0.007	
G_{ST}						
MAB		0.019	0.200	0.192	0.001	0.001
SNE	0.006		0.537	0.922	0.002	0.002
GB	0.003	0.001		0.573	0.001	0.001
GoM	0.004	−0.003	0.001		0.001	0.001
BB	**0.025**	**0.022**	**0.025**	**0.018**		0.167
NS	**0.017**	**0.014**	**0.019**	**0.012**	0.006	
D_{est}						
MAB		0.021	0.201	0.193	0.001	0.001
SNE	0.001		0.537	0.922	0.002	0.002
GB	0.001	0.000		0.573	0.001	0.001
GoM	0.001	0.000	0.000		0.001	0.001
BB	**0.005**	**0.005**	**0.005**	**0.004**		0.159
NS	**0.003**	**0.003**	**0.003**	**0.002**	0.001	

NW Atlantic shelf population connectivity

The full migration model showed the highest support among the considered models (Table 4). The north to south model, which might seem consistent with a passive particle drift model following the main currents, was the second best model, but had a much worse fit. The relative contribution of migrants (M) to each region is indicated in Table 5. Mutation-escalated population sizes (θ) were slightly large for MAB and SNE than for GB and GoM (Table 5). These values (both M and θ) should, however, be considered with caution, due to the violation of assumptions with impact on the accurate quantification of migrants (Beerli 2010).

Discussion

The primary conclusion that can be inferred from the results of this study is the existence of population genetic structure of *Centropages typicus* across the North

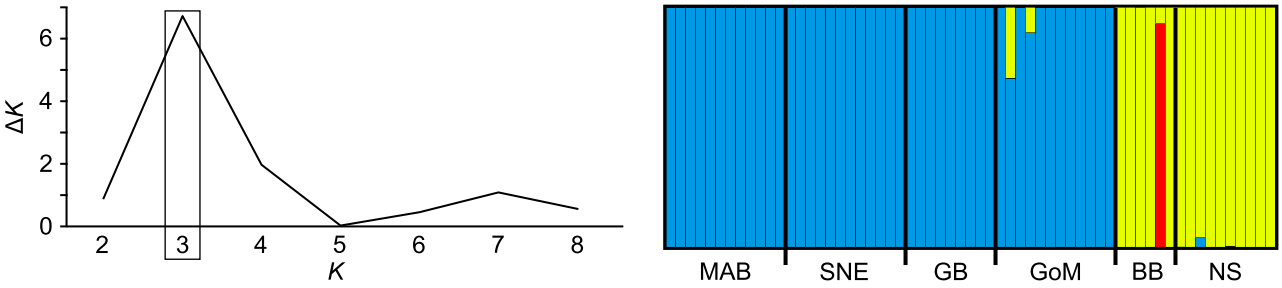

Fig. 5 Genetic clustering analysis for the whole data set (671 SNPs). On the left, graph of ΔK = mean(|L''(*K*)|) / sd(L(*K*)) (Evanno *et al.* 2005) as a function of *K* (potential number of genetic clusters). The most likely number of clusters is indicated by the modal value, in this case $K = 3$. On the right, genetic clustering graph for the number of clusters $K = 3$. Each colour represents a different genetic cluster. Bar graphs show average probability of membership (*y*-axis) of each individual. Regions are defined by thick vertical lines.

Table 4 Log probability of the data given the model (marginal likelihood, based on the Bezier approximation score) and corresponding Bayes factors for the tested connectivity models (see methods). The most probable model considers bidirectional connectivity between all regions

	Migration model					
	Panmixia	Full	Adjacent	N to S	S to N	Gulf Stream
Bezier lLm	−8574.76	−8198.41	−8703.44	−8275.81	−8288.87	−8577.36
Log Bayes Factor	−752.7	0	−1010.06	−154.8	−180.92	−757/9
Choice	5	1	4	2	3	6

Atlantic Ocean, including significant differences between continental shelf populations of the NE and NW Atlantic. Even more noteworthy is the capability of analysing the directional migration and testing of hypotheses of patterns of dispersal in this dynamic and complex current system.

These results agree with the conclusions inferred from morphological characteristics by Castellani *et al.* (2012), but contradict the molecular results from the same study, which indicated lack of genetic structure across the entire distributional range of the species (Castellani *et al.* 2012). In contrast, our finding of significant genetic structure reflects the power and potential of next-generation technologies to resolve highly sensitive and specific molecular genetic markers, such as RAD tags (Reitzel *et al.* 2013), which can discern subtle population genetic structure, despite the marked challenges associated with analysis of marine holoplanktonic species. Analysis of only a single or a few genetic markers results in relatively larger confidence intervals for very small *F* values (Waples 1998) and thus lacking statistical significance for species showing high gene flow and/or very large population sizes (Bowman 1953; Brinton 1962; Waples *et al.* 2008), as is typical of marine holoplanktonic species. Another advantage of using numerous independent markers that are randomly distributed throughout the genome (which can thus be assumed to be in linkage equilibrium) for population genetic analysis is that this significantly reduces the minimum number of individuals per population required for adequate sampling. Importantly, high

Table 5 95% confidence interval and mode for mutation-escalated population sizes (θ) and migrations (M) for each population and migration pathway, respectively

	2.50%	Mode	97.50%
Population			
MAB θ	0.001	0.018	0.034
SNE θ	0.001	0.017	0.033
GB θ	0	0.016	0.032
GoM θ	0	0.016	0.032
Pathway			
SNE to MAB	0	173.3	826.7
MAB to SNE	0	466.7	1066.7
GB to MAB	853.3	1666.7	2240
MAB to GB	0	93.3	746.7
GoM to MAB	0	386.7	1013.3
MAB to GoM	0	173.3	1466.7
GB to SNE	0	200	826.7
SNE to GB	0	200	1120
GoM to SNE	0	146.7	800
SNE to GoM	0	173.3	800
GoM to GB	0	173.3	800
GB to GoM	0	120	773.3

genetic marker densities allow the use of highly specific analytical approaches, which can be applied for the direct and separate measurement of population genetic structure (e.g. *F*-statistics) and connectivity (understood as the exchange of individuals) between those populations, without resorting to indirect estimates of migration rates based on *F*-statistics, which are unlikely to be reliable in our study system (Whitlock & McCauley 1999).

Centropages typicus population connectivity between the NW and NE Atlantic Ocean

The genetic pool of *C. typicus* dominating the NE Atlantic continental shelf was found to be distinct from that making up the majority of the NW Atlantic population. Clearly, planktonic species can and do show population genetic structure at ocean basin scales, despite high dispersal potential. The differential contribution of each of the two main genetic pools to the populations on either side of the Atlantic was detected and discriminated, allowing quantification of the exchange of individuals across the ocean basin. Large-scale, macroecological and morphological analyses of *C. typicus* across its distributional range have shown the effects of apparent demographic isolation between the neritic, overwintering populations occurring in the NE and NW Atlantic (Beaugrand *et al.* 2007; Castellani *et al.* 2012). The expatriation of individuals to the open ocean and the limited connection between the two sides of the North Atlantic Ocean during the seasonal maximum of the species (Beaugrand *et al.* 2007) may possibly result in the reduced genetic exchange observed and associated population differentiation. Given such high levels of gene flow, analyses using a single or few markers will likely fail to resolve population genetic structure, due to statistical limitations. Even low levels of gene flow across the North Atlantic Ocean (even on multigenerational timescales) could be biologically significant enough to minimize the effects of the genetic drift, resulting in genetic differences that are undetectable based on a single or few loci (Waples 1998). Our ability to detect population genetic structure is thus directly dependent on our capacity to resolve variation at the timescale/space scale that are relevant to life cycle processes (including dispersal) and to the responses of the species to environmental conditions. For this North Atlantic planktonic copepod species, these timescale/space scale are likely weeks to months and hundreds to several thousands of km (Beaugrand *et al.* 2007; Carlotti *et al.* 2014; Durbin & Kane 2007; Mazzocchi *et al.* 2007).

The detection of significant population genetic structure can allow further studies of deeper complexity, such as population-specific responses to predicted changes in the marine environment (Kokko & López-Sepulcre 2007). Not only has previously published research pointed to the potential existence of structured populations within the distribution of the species, but the physiology and biology of *C. typicus* have been shown to be geographically adapted to local conditions at regional scales (Bonnet *et al.* 2007; Carlotti *et al.* 2007; Gaudy & Thibault-Botha 2007). For species with populations distributed along a latitudinal or environmental gradient and/or exposed to contrasting local conditions, differences in physiological responses could reflect phenotypic or physiological plasticity, genetic adaptation to local conditions or a combination of both (Logan *et al.* 2012; Somero 2010, 2012; Stillman & Tagmount 2009; Whitehead 2012). Population connectivity would play a prominent role in the global resilience of the species – both at a given location and throughout its distributional range (Pauls *et al.* 2013), since individuals from other populations with genetic traits that are adapted to the newly dominant conditions may prevent local extinctions of the species through dispersal, recolonization and/or favourable selection (Morecroft *et al.* 2012; O'Connor *et al.* 2012).

NW Atlantic shelf population connectivity

Data analyses with all the SNP markers showed no significant genetic differentiation of *C. typicus* samples collected from different NW Atlantic regions. The NW Atlantic continental shelf is characterized by dominant currents that flow north to south most of the time (Aikman *et al.* 1988; Beardsley *et al.* 1985), but the migration flow patterns (obtained from all markers) indicated significant south-to-north flow across the sampled domain. The rates obtained from MIGRATE-N showed stronger northward flow from the MAB to the SNE than the reverse, although southward pathways appeared in general to be stronger. Overall, there was apparent equilibrium among SNE, GB and GoM. However, calculations of recent or ongoing migration rates obtained from MIGRATE-N should be viewed with caution, since these violate key assumptions of the method, especially when considering species with large effective population sizes (Beerli 2010).

The apparently equilibrium between the southern and northern regions of the NW Atlantic continental shelf – despite the dominant southern flow – could be explained by a number of reasons. First, population size and stability may be important when interpreting rates and flow directions. The southern regions (MAB and SNE) sustain large densities of *C. typicus* throughout the year, with densities $>10^5$ m^{-2}. In contrast, the species is nearly undetectable during winter and early spring in northern regions (GB, GoM) (Durbin & Kane

2007; Grant 1988; Meise-Munns 1990; Pershing *et al.* 2005, 2010). During unfavourable times in northern regions, any transport from the dense southern populations (e.g. linked to short-term hydrographic features like eddies, brief reversals of the circulation pattern) might have a significant impact on the genetic signature of northern populations. In fact, despite the southward mean flow (Beardsley *et al.* 1985; Chapman & Beardsley 1989), occasional northward transport with reversal of flow or partial weakening of the southward intensity is known to occur (Bi *et al.* 2014; Dong & Kelly 2003). Since migration rates refer to the proportion of immigrants compared to the total population, the impact of these events might be much larger than expected, given the low densities of northern *C. typicus* populations. In contrast, despite the dominant flow, the impact of the southward migration path would be relatively contained, since the high numbers of resident individuals at the receiving populations would reduce the proportion of migrants. An alternative hypothesis (untestable with the available data) is that persistent year-round populations in the northern Gulf of Saint Lawrence (Johnson *et al.* 2010) may seed populations throughout the entire length of the NW Atlantic continental shelf.

More detailed study is needed on the hydrographic processes and environmental conditions around the time of sampling for this study. *Centropages typicus* abundances in the two most northern regions are related to complex hydrographic phenomena associated with the Gulf Stream and the Gulf Stream North Wall Index (GSNWI; Taylor *et al.* 1992) (Bi *et al.* 2014). A northward displacement (positive GSNWI) would favour a northward alongshore transport, with the consequent displacement of individuals from the southern areas. Particle tracking models have showed the plausibility of this northward displacement from the MAB into GB following the Gulf Stream route, resulting from passive drifting for less of the equivalent of *C. typicus* generation time (Zhang *et al.* 2015). Such northward migration could be the signal identified by our analyses. During 2006-2010, the alongshore current and the GSNWI favoured northward displacements (Bi *et al.* 2014), although during 2011 showed a low yearly average but with very positive indices for multiple months, including November (data available at http://pml-gulfstream.org.uk/data.htm). On the other hand, the MIGRATE-N method might be expected to be robust to such punctual events and reflect the long-term effect of migration patterns on the populations. Other methods more suitable for detecting short-term events (like BIMr; Faubet & Gaggiotti 2008) showed no consistent results (data not shown) and would require a denser sampling of individuals (Faubet & Gaggiotti 2008; Faubet *et al.* 2007).

Mitochondrial vs. nuclear markers

With the advent of technically accessible and economically affordable methods to detect multiple nuclear markers (e.g. SNPs, microsatellite DNA) at the individual level, recent research efforts have revisited previous population genetic studies using mitochondrial DNA markers. There are multiple advantages of approaches using nuclear markers, such as RAD tags. The most obvious is the increased genetic signal due to the multiplicity of markers. Equally important is the possibility of applying approaches that are independent of *F*-statistics (see review by Broquet & Petit 2009; Falush *et al.* 2003; Pritchard *et al.* 2000). In many cases, the results from studies using multiple nuclear markers contradict the previous studies and find evidence of significant population genetic structure (e.g. Unal & Bucklin 2010; this study), where mitochondrial DNA markers did not (e.g. Castellani *et al.* 2012; Provan *et al.* 2009). This is a common result from comparison of many and diverse studies (Toews & Brelsford 2012), not only for high-dispersal species. The difference between population genetic results using nuclear vs. mitochondrial markers does not imply that either approach is more correct. Apart from particular issues, for example hybridization or introgression (see Toews & Brelsford 2012), the presence of detectable population genetic structure based on mitochondrial markers might indicate stronger genetic isolation between populations. In contrast, numerous, independent nuclear markers may in general detect lower levels of divergence, with some exceptions (e.g. impact of population sizes). Accurate and cautious interpretation of the results is thus required, with full consideration of the impacts of the selection of genetic markers for the measurement of both genetic structure and connectivity among populations.

Acknowledgements

We acknowledge the contributions of Astrid Cornils (Alfred Wegener Institute, Germany), Sara Ceballos (Spanish Institute of Oceanography, Spain) and Jon Hare and David R. Richardson (Northeast Fisheries Science Center, Narragansett RI, USA). Some samples for this study were provided by the NOAA/NMFS/NEFSC Ecosystem Monitoring (EcoMon) Survey Program. We gratefully acknowledge technical assistance and support from Rachel J. O'Neill and Craig Obergfell (Center for Genome Innovation, University of Connecticut). Support to LB-B was provided by the College of Liberal Arts & Sciences, University of Connecticut. The Bioinformatics Facility of the University of Connecticut Biotechnology and Bioservices Center (BBC) provided computing resources for some analyses performed for this study. This study with the contribution number 3005 is of the Bermuda Institute of Ocean Sciences.

References

Aikman F, Ou HW, Houghton RW (1988) Current variability across the New England continental shelf-break and slope. *Continental Shelf Research*, **8**, 625–651.

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

Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Beardsley RC, Chapman DC, Brink KH, Ramp SR, Schlitz R (1985) The Nantucket Shoals Flux Experiment (NSFE79). Part I: a basic description of the current and temperature variability. *Journal of Physical Oceanography*, **15**, 713–748.

Beaugrand G, Lindley JA, Helaouet P, Bonnet D (2007) Macroecological study of *Centropages typicus* in the North Atlantic Ocean. *Progress in Oceanography*, **72**, 259–273.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

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

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, **13**, 827–836.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Beerli P (2010) *Violation of assumptions, or are your migration estimates wrong when the populations split in the recent past?* http://popgen.sc.fsu.edu/Migrate/Blog/Entries/2010/8/15_Violation_of_assumptions%2C_or_are_your_migration_estimates_wrong_when_the_populations_split_in_the_recent_past.html

Beerli P (2012) *Migrate Documentation, Version 3.2.1*. Department of Scientific Computing, Florida State University, Tallahassee, FL.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.

Bi H, Ji R, Liu H, Jo Y-H, Hare JA (2014) Decadal changes in zooplankton of the Northeast U.S. Continental Shelf. *PLoS ONE*, **9**, e87720.

Blanco-Bercial L, Cornils A, Copley N, Bucklin A (2014) DNA barcoding of marine copepods: assessment of analytical approaches to species identification. *PLOS Currents Tree of Life*, **1**, 1–22.

Bonnet D, Harris R, Lopez-Urrutia A, *et al.* (2007) Comparative seasonal dynamics of *Centropages typicus* at seven coastal monitoring stations in the North Sea, English Channel and Bay of Biscay. *Progress in Oceanography*, **72**, 233–248.

Bowman TE (1953) *The Systematics and Distribution of Pelagic Amphipods of the Families Vibiliidae, Paraphronimidae, Hyperiidae, Dairellidae and Phrosinidae from the Northeastern Pacific*. University of California, Los Angeles, California.

Brinton E (1962) The distribution of Pacific euphausiids. *Bulletin of the Scripps Institution of oceanography, University of California*, **8**, 21–270.

Broquet T, Petit EJ (2009) Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 193–216.

Calbet A, Carlotti F, Gaudy R (2007) The feeding ecology of the copepod *Centropages typicus* (Kröyer). *Progress in Oceanography*, **72**, 137–150.

Carlotti F, Harris R (2007) The biology and ecology of *Centropages typicus*: an introduction. *Progress in Oceanography*, **72**, 117–120.

Carlotti F, Bonnet D, Halsband-Lenk C (2007) Development and growth rates of *Centropages typicus*. *Progress in Oceanography*, **72**, 164–194.

Carlotti F, Eisenhauer L, Campbell R, Diaz F (2014) Modeling the spatial and temporal population dynamics of the copepod *Centropages typicus* in the northwestern Mediterranean Sea during the year 2001 using a 3D ecosystem model. *Journal of Marine Systems*, **135**, 97–116.

Castellani C, Lindley AJ, Wootton M, Lee CM, Kirby RR (2012) Morphological and genetic variation in the North Atlantic copepod, *Centropages typicus*. *Journal of the Marine Biological Association of the United Kingdom*, **92**, 99–106.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Chapman DC, Beardsley RC (1989) On the origin of shelf water in the Middle Atlantic Bight. *Journal of Physical Oceanography*, **19**, 384–391.

Continuous Plankton Recorder Survey Team (2004) Continuous plankton records: Plankton atlas of the North Atlantic Ocean (1958–1999). II. Biogeographical charts. *Marine Ecology Progress Series, Supplement*, **2004**, 11–75.

Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE (2007) Population connectivity in marine systems: an overview. *Oceanography*, **20**, 14–21.

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Dong S, Kelly KA (2003) Seasonal and interannual variations in geostrophic velocity in the Middle Atlantic Bight. *Journal of Geophysical Research: Oceans*, **108**, 3172.

Durbin E, Kane J (2007) Seasonal and spatial dynamics of *Centropages typicus* and *C. hamatus* in the western North Atlantic. *Progress in Oceanography*, **72**, 249–258.

Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Erikson K, Blanco-Bercial L, Richardson D, Hare J, Bucklin A (2014) Watching time fly: visualization of zooplankton population dynamics 1977 - 2013 from NOAA-NEFSC ecosystem monitoring of the NW Atlantic continental shelf. In: Joint Aquatic Sciences Meeting, Portland, OR.

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, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Faubet P, Gaggiotti OE (2008) A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics*, **178**, 1491–1504.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology*, **16**, 1149–1166.

Foll M, Gaggiotti O (2008) A Genome-Scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Gaudy R, Thibault-Botha D (2007) Metabolism of *Centropages* species in the Mediterranean Sea and the North Atlantic Ocean. *Progress in Oceanography*, **72**, 151–163.

Goetze E (2005) Global population genetic structure and biogeography of the oceanic copepods *Eucalanus hyalinus* and *E. spinifer*. *Evolution*, **59**, 2378–2398.

Grant GC (1988) Seasonal occurrence and dominance of *Centropages* congeners in the Middle Atlantic Bight, USA. *Hydrobiologia*, **167–168**, 227–237.

Gregory TR (2013) *Animal Genome Size Database*. http://www.genomesize.com

Hellberg ME (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 291–310.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, **11**, 123–136.

Hohenlohe PA, Bassham S, Etter PD, *et al.* (2010) Population genomics of parallel adaptation in Threespine Stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

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.

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.

Ji R, Stegert C, Davis CS (2013) Sensitivity of copepod populations to bottom-up and top-down forcing: a modeling study in the Gulf of Maine region. *Journal of Plankton Research*, **35**, 66–79.

Johnson C, Curtis A, Pepin P, Runge J (2010) Spatial patterns in zooplankton communities and their seasonal variability in the northwest Atlantic. *AZMP Bulletin PMZA*, **9**, 27–32.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kane J (2014) Decadal distribution and abundance trends for the late stage copepodites of *Pseudocalanus* spp. (Copepoda: Calanoida) in the US Northeast Continental Shelf Ecosystem. *Journal of Northwest Atlantic Fishery Science*, **46**, 1–13.

Kokko H, López-Sepulcre A (2006) From individual dispersal to species ranges: perspectives for a changing world. *Science*, **313**, 789–791.

Kokko H, López-Sepulcre A (2007) The ecogenetic link between demography and evolution: can we bridge the gap between theory and data? *Ecology Letters*, **10**, 773–782.

Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Logan C, Kost L, Somero G (2012) Latitudinal differences in *Mytilus californianus* thermal physiology. *Marine Ecology Progress Series*, **450**, 93–105.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Mazzocchi MG, Christou ED, Capua ID, *et al.* (2007) Temporal variability of *Centropages typicus* in the Mediterranean Sea over seasonal-to-decadal scales. *Progress in Oceanography*, **72**, 214–232.

Meise-Munns C (1990) Interannual variability in the copepod populations of Georges Bank and the Western Gulf of Maine. *Marine Ecology Progress Series*, **65**, 225–232.

Morecroft MD, Crick HQP, Duffield SJ, Macgregor NA (2012) Resilience to climate change: translating principles into practice. *Journal of Applied Ecology*, **49**, 547–551.

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.

O'Connor MI, Selig ER, Pinsky ML, Altermatt F (2012) Toward a conceptual synthesis for climate change responses. *Global Ecology and Biogeography*, **21**, 693–703.

Patarnello T, Papetti C, Zane L (2010) Chapter two – genetics of Northern Krill (*Meganyctiphanes norvegica* Sars). In: *Advances in Marine Biology* (ed. Geraint AT), vol. 37, pp. 41–57. Academic Press, London, UK.

Pauls SU, Nowak C, Bálint M, Pfenninger M (2013) The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, **22**, 925–946.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research —an update. *Bioinformatics*, **28**, 2537–2539.

Peijnenburg KTCA, Fauvelot C, Breeuwer JAJ, Menken SBJ (2006) Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology*, **15**, 3319–3338.

Pepin P, Colbourne E, Maillet G (2011) Seasonal patterns in zooplankton community structure on the Newfoundland and Labrador Shelf. *Progress in Oceanography*, **91**, 273–285.

Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Pershing AJ, Greene CH, Jossi JW, *et al.* (2005) Interdecadal variability in the Gulf of Maine zooplankton community,

with potential impacts on fish recruitment. *ICES Journal of Marine Science: Journal du Conseil*, **62**, 1511–1523.
Pershing AJ, Head EHJ, Greene CH, Jossi JW (2010) Pattern and scale of variability among Northwest Atlantic Shelf plankton communities. *Journal of Plankton Research*, **32**, 1661–1674.
Petrik CM, Kristiansen T, Lough RG, Davis CS (2009) Prey selection by larval haddock and cod on copepods with species-specific behavior: an individual-based model analysis. *Marine Ecology Progress Series*, **396**, 123–143.
Poelstra JW, Ellegren H, Wolf JBW (2013) An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Heredity*, **111**, 467–473.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
Provan J, Beatty GE, Keating SL, Maggs CA, Savidge G (2009) High dispersal potential has maintained long-term population stability in the North Atlantic copepod *Calanus finmarchicus*. *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 301–307.
Puritz JB, Matz MV, Toonen RJ, *et al.* (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937–5942.
Razouls C, de Bovée F, Kouwenberg J, Desreumaux N (2005-2014) *Diversity and geographic distribution of marine planktonic copepods*. http://copepodes.obs-banyuls.fr/en.
Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Molecular Ecology*, **22**, 2953–2970.
Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
Somero GN (2010) The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. *The Journal of Experimental Biology*, **213**, 912–920.
Somero GN (2012) The physiology of global change: linking patterns to mechanisms. *Annual Review of Marine Science*, **4**, 39–61.
Stegert C, Ji R, Li N, Davis CS (2012) Processes controlling seasonality and spatial distribution of *Centropages typicus*: a modeling study in the Gulf of Maine/Georges Bank region. *Journal of Plankton Research*, **34**, 18–35.
Stillman JH, Tagmount A (2009) Seasonal and latitudinal acclimatization of cardiac transcriptome responses to thermal stress in porcelain crabs, *Petrolisthes cinctipes*. *Molecular Ecology*, **18**, 4206–4226.
Taylor AH, Colebrook JM, Stephens JA, Baker NG (1992) Latitudinal displacements of the Gulf Stream and the abundance of plankton in the North-East Atlantic. *Journal of the Marine Biological Association of the United Kingdom*, **72**, 919–921.
Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.
Turner J, Tester P, Hettler W (1985) Zooplankton feeding ecology. *Marine Biology*, **90**, 1–8.
Unal E, Bucklin A (2010) Basin-scale population genetic structure of the planktonic copepod *Calanus finmarchicus* in the North Atlantic Ocean. *Progress in Oceanography*, **87**, 175–185.
Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, **9**, 808–810.
Waples R (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 438–450.
Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries*, **9**, 423–449.
Whitehead A (2012) Comparative genomics in ecological physiology: toward a more nuanced understanding of acclimation and adaptation. *The Journal of Experimental Biology*, **215**, 884–891.
Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity*, **82**, 117–125.
Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.
Yebra L, Bonnet D, Harris R, Lindeque P, Peijnenburg K (2011) Barriers in the pelagic: population structuring of *Calanus helgolandicus* and *C. euxinus* in European waters. *Marine Ecology Progress Series*, **428**, 135–149.
Zhang X, Haidvogel D, Munroe D, *et al.* (2015) Modeling larval connectivity of the Atlantic surfclams within the Middle Atlantic Bight: model development, larval dispersal and metapopulation connectivity. *Estuarine, Coastal and Shelf Science*, **153**, 38–53.

L.B.B. and A.B. conceived and designed the study. L.B.B. performed the experiments and analyzed the data. A.B. contributed reagents and materials. L.B.B. and A.B. wrote the manuscript.

Data accessibility

Raw sequencing data are available at the BioProject PRJNA265130 http://www.ncbi.nlm.nih.gov/bioproject/265130. Final SNP data set as obtained from Stacks uploaded as online supplemental material. All infiles uploaded as online supplemental material. Sampling locations included in the text (Table 1).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 BAYESCAN plot of the 185 SNP loci selected for the analyses.

Fig. S2 Genetic clustering analysis for the whole dataset with 10× minimum coverage per allele (184 SNPs).

Fig. S3 Total F_{ST}, G_{ST} and D_{est} estimates (horizontal tick) and 95 % CI (vertical bar) for the whole dataset (653 SNPs) and for random subsets of 500, 400, 300, 200, 100 and 50 SNPs (five subreplicates/replicates each).

Table S1 F_{ST} distances between samples (below diagonal) and corresponding *P*-value (above diagonal).

Table S2 Locus by locus AMOVA analysis, considering NW vs. NE Atlantic groups. Asterisks (*) indicate significance $P < 0.05$.

Table S3 Loci under selection from genome scans based on *F*-statistics as implemented in ARLEQUIN.

Table S4 F_{ST} distances between samples after removing all candidate loci for positive selection at $P < 0.001$ (A) and $P < 0.005$ (B); and after removing all candidates for both positive and balancing selection at $P < 0.001$ (C) and $P < 0.005$ (D).

Table S5 Locus by locus AMOVA analysis, considering NW vs. NE Atlantic groups after removing all candidate loci for positive selection at $P < 0.001$ (A) and $P < 0.005$ (B); and after removing all candidates for both positive and balancing selection at $P < 0.001$ (C) and $P < 0.005$ (D). Asterisks (*) indicate significance $P < 0.05$.