# Fine-scale seascape genomics of an exploited marine species, the common cockle *Cerastoderma edule*, using a multi-modelling approach

Ilaria Coscia[1*], Sophie B. Wilmes[2], Joseph E. Ironside[3], Alice Goward-Brown[2], Enda O'Dea[4], Shelagh K. Malham[2], Allan D. McDevitt[1], Peter E. Robins[2]

**Running title:** Seascape genomics of the common cockle

**Keywords**: RADseq; Particle tracking; Irish Sea; redundancy analysis; population connectivity; larval dispersal

**Addresses**

[1] Ecosystems and Environment Research Centre, School of Science, Engineering and Environment, University of Salford, Salford, M5 4WT, UK

[2] School of Ocean Sciences, Bangor University, Marine Centre Wales, Menai Bridge, LL59 5AB.

[3] Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Penglais, Aberystwyth, Ceredigion, SY23 3DA, UK.

[4] Met Office, Fitzroy Road, Exeter, EX1 3PB, UK.

**Corresponding author**: I.Coscia@salford.ac.uk

## ABSTRACT

Population dynamics of marine species that are sessile as adults are driven by oceanographic dispersal of larvae from spawning to nursery grounds. This is mediated by life-history traits such as the timing and frequency of spawning, larval behaviour and duration, and settlement success. Here, we use 1725 single nucleotide polymorphisms (SNPs) to study the fine scale spatial genetic structure in the commercially important cockle species *Cerastoderma edule* and compare it to environmental variables and current-mediated larval dispersal within a modelling framework. Hydrodynamic modelling employing the NEMO Atlantic Margin Model (AMM15) was used to simulate larval transport and estimate connectivity between populations during spawning months (April - September), factoring in larval duration and seasonal variability of ocean currents. Results at neutral loci reveal the existence of three separate genetic clusters (mean $F_{ST}$=0.021) within a relatively fine spatial scale in the northwest Atlantic. Environmental association (EA) analysis indicates that oceanographic currents and geographical distance between the populations explain over 20% of the variance observed at neutral loci, while genetic variance (71%) at outlier loci was explained by sea surface temperatures extremes. These results fill an important knowledge gap in the management of a commercially important, overexploited species, and bring us closer to understanding the role of larval dispersal in connecting populations at a fine geographical scale.

## 1. Introduction

The assumption that genetic homogeneity predominates in marine organisms due to the lack of physical barriers and high dispersal potential at all life stages has been challenged in recent years (Allendorf, 2017). The advent of genomic markers generated by scanning the whole genome of an organism has equipped researchers with the necessary statistical power to detect more fine-scale population differentiation within the marine realm (Benestan et al., 2015; Maroso et al., 2018). Furthermore, genome scans have allowed a shift in focus from neutral variation to local adaptation, which is one of the main drivers of structure within marine populations (Araneda, Larraín, Hecht, & Narum, 2016; Nielsen et al., 2012; Woodings et al., 2018) and indicates the potential for resilience to future environmental change (Razgour et al., 2019). Connectivity between marine populations is central to their health and resilience to external pressures such as parasites and pathogens (Rowley et al., 2014), pollution, human exploitation and climate change over ecological and evolutionary timescales (Burgess, Bowring, & Shen, 2014; Cowen & Sponaugle, 2009; Gimenez, 2019). For these reasons, it is vital to distinguish between neutral variation and local adaptation when attempting to understand what drives the observed population structure in marine organisms.

A seascape genomics approach is particularly valuable in this context (Selkoe et al., 2016). Genetic and genomic data can be used in conjunction with environmental variables such as sea water temperature, salinity, water depth, irradiance, turbidity, and sediment type (Viricel & Rosel, 2014) against the neutral differentiation determined by the equilibrium between larval dispersal and effective population size (Coscia, Robins, Porter, Malham, & Ironside, 2013; Young et al., 2015). This can provide new insights for interpreting genetic patchiness in relation to specific environmental features (Benestan et al., 2016; Bernatchez et al., 2019; Truelove et al., 2017).

In particular, seascape genomics has proven to be extremely useful when considering exploited species, as it has the potential to inform management and aid sustainable exploitation by enabling management units to be defined (Bernatchez et al., 2017; Teacher, André, Jonsson, & Merilä, 2013). Despite intense exploitation of many marine species, their management rarely takes into account genomic information (Bernatchez et al., 2017; ICES, 2018) and exploited aquatic invertebrates (shellfish) in particular receive little attention from policy makers and stakeholders in comparison with fish (Elliott & Holden, 2017). In the Irish Sea, shellfish represent 80% of the total landings per year, with an economic value of £46.6 M (Elliott & Holden, 2017). Among these, the common cockle *Cerastoderma edule* fisheries are some of the most valuable fisheries for both Ireland and the United Kingdom (Dare, Bell, Walker, & Bannister, 2003; Hervas, Tully, Hickey, O'Keeffe, & Kelly, 2007) and are of high socio-economic importance, valued at £3.3 M for Wales alone (Elliott & Holden, 2017).

The common cockle has both ecological and commercial importance, providing an important food source for wading birds in addition to employment for coastal communities (Flach & de Bruin, 1994; Hickin, 2019). *C. edule* occurs in intertidal soft sediment regions of the eastern Atlantic, from Norway to Senegal. It can live for six to ten years and is characterized by high fecundity and high dispersal potential due to a pelagic larval phase which lasts for 24 to 35 days following spawning from May to August (Malham, Hutchinson, & Longshaw, 2012). In recent decades, cockle stocks have declined across Europe, with production falling from 108,000 tons in 1987 to less than 25,000 tons in 2008 (Martínez, Mendez, Insua, Arias-Pérez, & Freire, 2013). Cockle declines have been attributed to different factors in different locations, such as overharvesting (Wolff, 2005) and parasitic infections (Longshaw & Malham, 2015; Thieltges, 2006). In the UK, recurrent mass mortalities have occurred at several long-established cockle fisheries, resulting in significant economic losses (Woolmer, 2013). These mortalities have not been attributed to any single environmental

100 factor and interactions between multiple stress factors are suspected (Callaway, Burdon,
101 Deasey, Mazik, & Elliott, 2013; Malham et al., 2012). Sustainable management of the
102 common cockle is hindered by poor understanding of their population connectivity. Analysis
103 of microsatellite and mitochondrial DNA markers suggests weak barriers to gene flow
104 between *C. edule* populations along the North European coast (Coscia et al., 2013; Martínez,
105 Freire, Arias-Pérez, Mendez, & Insua, 2015). However, these markers lack sufficient
106 resolution to investigate connectivity at the finer scales relevant to fisheries management
107 (Bernatchez et al., 2017).

108 Given the major logistical challenges of directly quantifying larval connectivity,
109 efforts have focused on simulating ocean hydrodynamics to deduce larval dispersal
110 probabilities (Cowen, Gawarkiewicz, Pineda, Thorrold, & Werner, 2007; Paris, Cowen, Claro,
111 & Lindeman, 2005; Robins, Neill, Giménez, Jenkins, & Malham, 2013). This approach
112 identifies well-connected population groups as well as weakly-connected, partially-connected
113 or isolated populations. These simulations highlight the importance of local and mesoscale
114 hydrodynamics interacting with species-specific larval behaviours in driving population
115 persistence (Bode et al., 2019; Botsford et al., 2009; North et al., 2008; Robins et al., 2013)
116 and recovery from stock decline (Gimenez, 2019), indicating that the capacity of a population
117 to recover from mass mortalities is contingent on the scale of disturbance relative to the scale
118 of connectivity. Usual circulation patterns and, hence connectivity, can be modulated by
119 severe wind and wave conditions. Previous larval dispersal studies have predicted that given
120 atypical meteorological conditions during spawning events, new connectivity routes can be
121 established (e.g. by reversing the Celtic Sea front circulation; Hartnett, Berry, Tully, &
122 Dabrowski, 2007), or by large distance displacements from expected routes affecting sea
123 turtles: (Monzón-Argüello C. et al., 2012).

In the present study, a seascape genomics approach using single nucleotide polymorphisms (SNPs) is employed, for the first time, to resolve patterns of population structure of the common cockle between estuaries within a commercially active area (the Irish and Celtic Seas). An additional aim is to investigate the influence of environmental factors, including oceanic currents and water temperature upon the cockle's population structure. To do this, larval transport between sites is estimated and connectivity matrices are derived from oceanographic modelling, accounting for variability due to bio-physical parameters, i.e. spawning date and larval duration. Finally, Environmental Association analysis (EA) is used in order to understand the interplay between larval dispersal and abiotic factors in shaping the population connectivity measured with genomic markers.

## 2. Materials and methods

### 2.1. Sampling and DNA extraction

Cockles were collected between 2010 and 2011 from seven locations off the coasts of Ireland (Bannow Bay [BAN] and Flaxfort strand [FLX]) and Britain (Burry Inlet [BUR], Gann Estuary [GAN], Dyfi estuary [DYF], Red Wharf Bay [RWB] and Dee estuary [DEE]) (Fig. 1 and Table 1). Genomic DNA was extracted from ethanol-preserved and frozen tissue using the DNeasy Blood and Tissue Kit with an additional RNaseA step (Qiagen©), as per the manufacturer's protocols. The quality and quantity of the extracted DNA was assessed by gel electrophoresis (1% agarose) as well as Qubit dsDNA HS (high sensitivity, 0.2 to 100 ng) Assay Kit on a Qubit 3.0 fluorometer according to the manufacturer's protocols.

### 2.2 RAD sequencing and bioinformatic analysis

Reduced representation libraries (Baird et al., 2008) were constructed using the restriction enzyme *PstI* (New England Biolabs) for Restriction-Site Associated DNA sequencing (RAD-

149 seq). RAD libraries (each consisting of 24 uniquely barcoded individuals) were produced
150 according to Etter et al. (2011). Each library was quantified using real-time PCR and single-
151 end (100bp target) sequenced on an *Illumina HiSeq2000* at Aberystwyth University, UK.
152 Initial bioinformatic analysis, including quality control, demultiplexing and identification of
153 polymorphisms were performed by *Floragenex* (Eugene, OR; www.floragenex.com), using
154 *Samtools* (Li et al., 2009) and custom scripts, retaining one SNP per tag, with a minimum
155 coverage depth of six for each allele and genotyped in at least 70% of the individuals in the
156 overall dataset. These settings produced an initial dataset of 191 individuals and 4271 single
157 nucleotide polymorphisms (SNPs), which was then further filtered by the authors using the
158 packages *poppr 2.8.1* (Kamvar, Brooks, & Grünwald, 2015; Kamvar et al., 2019; Kamvar,
159 Tabima, & Grünwald, 2014), and *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011; Jombart
160 et al., 2018) in *R* (R Core Team, 2019).
161 In the second round of filtering, markers with data missing in more than 25% of
162 individuals were discarded from the dataset. A second filtering step removed all loci with $F_{IS}$
163 equal to 1, -1 or NA. Three MAF (minimum allele frequency) filters of 0.05, 0.025 and 0.01
164 were then applied, generating three datasets. No significant effect of MAF filter upon
165 heterozygosity, global $F_{ST}$ and population structure was detected, and so a MAF filter of 0.01
166 across all sites was applied (Xuereb et al., 2018). This allowed us to maximize the number of
167 markers and hence the information available at low spatial scale, while reducing the bias that
168 might be introduced by retaining low frequency SNPs (Roesti, Salzburger, & Berner, 2012).
169 All the downstream analyses were thus performed on the MAF=0.01 dataset. Markers in
170 linkage disequilibrium (LD) were identified using the function *pair.ia* of the *poppr* package
171 ($r^2 > 0.7$). Finally, SNPs that were found to deviate from Hardy-Weinberg Equilibrium (HWE;
172 $P$=0.01) in four out of seven populations were removed (Wyngaarden et al., 2018). The final,

filtered dataset included 138 individuals from the seven locations, and 1725 SNPs (Table 1 Supplementary Material).

### 2.3 Neutrality tests and population structure

Genomic markers were tested for neutrality using two complementary methods: *BayeScan* 2.1 (Foll & Gaggiotti, 2008) and the R package *pcadapt.* In order to minimize the detection of false positives, we considered as outliers only those SNPs that were selected by both methods (Coscia, Vogiatzi, Kotoulas, Tsigenopoulos, & Mariani, 2013). *BayeScan* is an $F_{ST}$-based, bayesian method (Beaumont & Balding, 2004), while *pcadapt* is based on a Principal Component Analysis (PCA) of individual genotypes and is known to perform particularly well in presence of weak structure, admixture or range expansions (Luu et al. 2017). To set the most appropriate number of clusters (K), we followed the recommendations of the authors and chose the K before the plateau in a scree plot, which displays in decreasing order the percentage of variance explained by each principal component (PC). Q-values were used to account for false discovery rate and SNPs were considered as significant outliers at alpha values $\leq$0.05. Genetic diversity was estimated on three datasets: overall, neutral and outliers, in order to disentangle the role of demographic processes vs selection. Heterozygosity (expected and observed) was estimated with the R package *hierfstat* (Goudet, 2005; Goudet & Jombart, 2015), while population pairwise $F_{ST}$ (Weir & Cockerham, 1984) and relative 95% confidence interval (1000 bootstraps) was estimated with the R package *assigner* (Gosselin, 2019).

Individual-based population structure was assessed using two different approaches: discriminant analysis of principal components (DAPC) as implemented in *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011; Jombart et al., 2018) and *fastSTRUCTURE,* run with a simple prior (Raj, Stephens, & Pritchard, 2014), with the number of k clusters that best explains the

198 structure in the data chosen using the *chooseK.py* script within the *fastSTRUCTURE* software.
199 The power of the neutral and outlier datasets to discriminate and assign individuals was
200 determined with a genotype accumulation curve (Fig. 2 Supplementary Material), calculated
201 with the function *genotype_curve* of the *poppr* package (Kamvar et al., 2015, 2019, 2014).
202
203 **2.4 Hydrodynamic modelling**
204 Hydrodynamic modelling was used to simulate larval transport and predict connectivity
205 between the seven sampled cockle populations, considering natural variability in larval
206 dispersal caused by the timing of spawning relative to the tide, larval duration and seasonal
207 variability of ocean currents. Simulated 3-D flow fields throughout the Irish and Celtic Seas
208 during the cockle spawning/settlement season (April - September) were obtained from the
209 NEMO Atlantic Margin Model (AMM15). The model has a horizontal resolution of 1.5 km
210 and 51 terrain-following layers. It was developed to resolve key dynamical features of the
211 European north-west shelf including the influence of shelf-break dynamics. Bathymetry was
212 provided from EMODnet (EMODnet Portal, September 2015 release) and bottom friction was
213 controlled through a log layer with a non-linear drag coefficient set at 0.0025. Ocean boundary
214 conditions were taken from the Global Seasonal Forecast System (GLOSEA), which includes
215 assimilation of both satellite and in-situ observations. Atmospheric forcing was driven by the
216 European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis
217 product, ERA-Interim (Dee et al., 2011), while river inputs were based predominantly on daily
218 climatology of gauge data. See Graham et al. (2018) for further details of the model
219 development and validation. The simulated flows from AMM15 were used to drive a Particle
220 Tracking Model (PTM) developed to simulate cockle larval transport from each sampled
221 population. In particular, 3-D velocities during April-Septemberwere output at hourly-
222 averaged temporal resolution.

223

224 **2.5 Particle Tracking Model (PTM)**

225 The PTM simulates the lagrangian movement of individual particles in space and time, based
226 on oceanographic dispersion and individual particle behaviour. The PTM was programmed in
227 Python and all simulations were run in a parellised framework on high-performance
228 computing. Hourly-averaged outputs of 3-D velocities from the ocean model was deemed
229 practical in terms of data storage and PTM computational efficiency, whilst minimising
230 uncertainties in particle dispersal generated from interpolating model output.

231 A previous sensitivity study by Robins et al. (2013) demonstrated that cohorts of
232 10,000 particles released from a range of point sources within the Irish Sea were sufficient to
233 statistically represent dispersal and connectivity patterns. Accordingly, cohorts of 750
234 particles were released from sites 1-7 (Fig. 3) each day (at 12:00) over the initial 16 days of
235 April 2014 (i.e. a total 12,000 particles per site covering a spring-neap tidal cycle). This
236 procedure was repeated from April to September 2014. The dispersal of each particle was
237 tracked for 40 days pelagic larval duration (PLD), reflecting the observed PLD of cockle
238 larvae (Malham et al. 2012). Particles were neutrally buoyant and able to disperse throughout
239 the 3-D flow field. Connectivity between populations was determined from particle
240 trajectories during days 30-40; particles that travelled within 20 km of a settlement site were
241 assumed to have settled there. The 20 km threshold was based on the average tidal excursion
242 for the Irish Sea (Robins et al. 2013), although a range of thresholds are discussed in the
243 Results section.

244

245 **2.6 Modelling uncertainty**

246 The ocean model AMM15 was run in free mode without data-assimilation. When its results
247 were compared with a simulation with data-assimilation for a different year to this study, the

248 inherent biases were small and are not expected to change the advection along the tidal mixing
249 fronts and hence larval transport. For the larval dispersal scenarios considered here, the
250 simulated variability in dispersal distance after 40 days, due to the release day (particles were
251 released daily for 16 days), was of the same order as the variability encountered within each
252 daily cohort of particles. This result demonstrates the importance of 'trickle-spawning' larvae
253 over a full lunar tidal cycle to capture dispersal variability.

254 In the absence of strong evidence for *C. edule* larval swimming behaviour, the effects
255 of larval behaviour were not incorporated into the particle tracking model. The general
256 agreement in population structure between our genomics and modelling analyses could be
257 interpreted as evidence for a larval transport mechanism that is controlled by ocean dispersion
258 rather than swimming behaviour. In effect, this result suggests that the ocean currents are
259 much stronger than larval swimming speeds.

260 In this work, one year of larval dispersal was simulated. Due to interannual variations
261 in weather patterns the timing of the onset of seasonal stratification and its strength can vary
262 between years, variability in dispersal due to atmospheric conditions over one season has been
263 accounted for in our study. As 2014 was an 'average' year with respect to the long-term
264 current and stratification variability, we are confident that this year is representative of the
265 mean transport pathways in the Irish and Celtic Sea region. However, future work could
266 investigate the influence of interannual variations in ocean climate on larval dispersal. since,
267 in the aftermath of severe storm events, ocean circulation patterns have been shown to reverse
268 in UK shelf-seas (Stanev et al., 2019). Although such storms are rare, having only occurred
269 around the UK four times in the last 40 years, changes in ocean circulation during spawning
270 cycles due to rare and severe storms create additional uncertainty in larval trajectories,
271 potentially leading to the establishment of new communities which have no clear connection
272 at other times.

Despite the relatively fine spatial resolution of our model (1.5 km with 51 vertical layers), some coastal currents will inevitably be poorly resolved, adding a potential source of uncertainty for larval transport in the coastal zone. Coastal modelling requires an unstructured grid mapped onto a high-resolution coastal bathymetry and including wetting/drying capabilities within the model and also wave-induced currents. This is a considerable task in terms of validation and application to particle tracking, but should be considered for further studies.

**2.7 Spatial Eigenfunction Analysis (SEA) and Environmental Association (EA) analysis**

To test for association between the oceanographic environment and the cockle genetic structure, a Spatial Eigenfunction Analysis (SEA) (Dray et al., 2012) was performed. This approach was implemented through the *vegan* and *adespatial* packages (Dray et al., 2019; Oksanen et al., 2017) in R. SEA was used to estimate the influence of geographic distance between samples as well as the influence of modelled connectivity on genomic variation. Geographic distance was represented as distance-based Moran's eigenvector maps or dbMEM (Stéphane Dray, Legendre, & Peres-Neto, 2006). dbMEMs were calculated using the *PCNM* function of the *vegan* package on the Euclidean distances (calculated with the function *dist*) in turn estimated from Cartesian-transformed coordinates using the *geoXY* function in the *SoDA* package (Chambers, 2013). *PCNM* transforms spatial distances in rectangular matrices that are suitable for constrained ordination (Legendre & Gallagher, 2001).

For the Environmental Association analysis (EA), modelled connectivity was represented as asymmetric eigenvector maps or AEM (Blanchet, Legendre, & Borcard, 2008; Blanchet, Legendre, Maranger, Monti, & Pepin, 2011). AEM is a spatial eigenfunction approach specifically developed to model multivariate responses to asymmetric and directional processes such as current-driven larval dispersal (Blanchet et al. 2011). The

298 connectivity probability matrix generated by the biophysical model was translated into a
299 nodes-to-edge matrix which records the presence/absence of connectivity links (through
300 *edges*, 19 here) between *nodes* (the 7 sampling locations). Each edge has an associated
301 'weight', based on the simulated probability of connectivity. AEMs were calculated with the
302 *aem* function in R.

303 For each sampling location, the simulated monthly-averaged sea surface temperature
304 (SST) and surface-bottom temperature difference (SBTD - the difference between the
305 temperature at the surface and the bottom of the water column) representing ocean
306 stratification, from April to September, were extracted from the ocean model (Figure 1
307 Supplementary Material). The relative contribution of temperature (SST and SBTD), spatial
308 distribution (dbMEMs) and directional connectivity (AEMs), to genetic variation in both the
309 neutral and the outlier datasets (response variable) was modelled using redundancy analysis
310 (RDA). In particular, the response variable was represented by population-specific minor
311 allele frequencies (MAF) of each SNP calculated in the R package *hierfstat* (function
312 *minorAllele*) (Goudet and Jombart 2015), detrended using the *decostand* function with the
313 *hellinger* transformation available in *vegan* (Oksanen 2018). The most important explanatory
314 variables were chosen by performing the forward selection with 10,000 permutations in *vegan*
315 using the function *ordistep*. Each model's significance was assessed with an analysis of
316 variance (function *anova* in *vegan*) with 1000 permutations, to finally establish which factors
317 were most correlated with genetic variation. Redundancy analysis (RDA) and partial-RDA
318 (corrected for geographic distance between populations) were performed using the *rda*
319 function in *vegan* on the putatively neutral (1711) and outlier (14) SNP datasets, and
320 parsimonious RDAs were carried out using the variables selected (Borcard, Gillet, &
321 Legendre, 2011).
322

# 3. RESULTS

## 3.1 Genetic diversity and population structure

After filtering for missing data, $F_{IS}$, MAF, HWE and LD, 138 individuals and 1725 SNPs were retained (Table 1 Supplementary Material). *Bayescan* detected 28 SNPs as potential outliers, with a False Discovery Rate of 0.05, whereas *PCAdapt* found 62. Only 14 SNPs overlapped between the two approaches. The downstream analyses were then carried out on two datasets: neutral (1711 SNPs) and outliers (14 SNPs).

For the neutral dataset, expected and observed heterozygosity ($H_E$ and $H_O$) were similar across locations, ranging between 0.148-0.157 and 0.103-0.134, respectively (Table 1). Pairwise $F_{ST}$ (Fig. 3 and 4 Supplementary) ranged between 0 (RWB-DEE) and 0.0289 (GAN-DEE). *fastSTRUCTURE*'s *choosek.py* detected three clusters (k=3; Fig. 1D-G), while *DAPC*'s *find.clust* found a maximum of five clusters (k=5; Fig. 1B-C) based on the BIC scores. For the putative outlier dataset, $F_{ST}$ values ranged between 0 (DYF-GAN) and 0.38 (BUR-FLX), but *choosek.py* and *DAPC*'s *find.clust* indicated that one cluster explained the structure in the data (barplots represented in Fig. 1E and 1G for comparison). All population-level neutral $F_{IS}$ values were significant (95% confidence interval) and positive, ranging between 0.08 and 0.27 (Table 1). For the outlier dataset, $H_E$ and $H_O$ varied respectively between 0.21-0.35 and 0.14-0.28 and $F_{IS}$ values were positive and significant for four populations: DEE, RWB, DYF and FLX.

Genotype accumulation curves (Fig. 2 Supplementary Material) for both datasets show that a plateau is reached and variance decreased at 20 neutral and 12 outliers loci respectively, indicating that the datasets have sufficient power to distinguish between individuals (Arnaud-Haond, Duarte, Alberto, & Serrão, 2007).

## 3.2 Seasonal circulation and larval dispersal

348 The typical circulation structure within the Irish/Celtic Seas has, in our study, led to distinct
349 patterns of larval transport and population connectivity, but with spatial variability over
350 seasonal timescales. Whilst the tidal currents in the Irish Sea are characteristically large (e.g.
351 peak flows exceed 1 m/s throughout much of the basin; Robins et al. 2015) and have the
352 potential to advect larvae tens of kilometers during each tidal cycle, these currents are
353 oscillatory, and resulting in small net flows and minimal dispersal over several weeks (Robins
354 et al. 2013). However, the development of thermal stratification and ocean fronts during
355 summer months generates density-driven currents along frontal boundaries (Horsburgh, Hill,
356 & Brown, 1998; Simpson & Hunter, 1974). Although much weaker than tidal currents, these
357 density-driven currents are persistent over time and act as key dispersal pathways for larvae.
358 For example, the Celtic Sea front may facilitate connectivity of cockle populations from South
359 Wales across to Ireland (Coscia et al. 2013).

360 The AMM15 simulation shows the Celtic and Irish Seas during April 2014 to be
361 vertically well mixed with net tidal flows following well-known patterns due to tidal
362 interactions with morphology (Pingree & Griffiths, 1979) (Fig. 2A). These net flows were
363 generally less than 0.1 m/s and directed south/southwest and north/northeast from a parting
364 zone (amphidromic point) south of Dublin. During May-September, the simulation shows the
365 development of thermal stratification leading to density-driven currents (Fig. 2B-F). The
366 Celtic Sea and the western Irish Sea stratified generating persistent east-to-west currents along
367 the Celtic Sea front and an anti-clockwise gyre in the western Irish Sea. For the Celtic Sea
368 front, depth-averaged flows were ~0.15 m/s but strongest at the thermocline at ~30m depth.
369 These density-driven flows persisted but weakened during August and September.

370 The degree of exposure of the spawning grounds influenced the population's ability to
371 connect with other sites (Figs. 3 and 5 Supplementary). This is shown by comparing the
372 simulated particle dispersal from DEE (Site 1) with RWB (Site 2) on the north Wales coast.

373 The majority of particles dispersing from DEE remained <30 km from the release site,
374 resulting in high proportion of self-recruitment (17.4±0.7%), with only small proportions of
375 particles dispersed north to the North Channel and the Scottish coast (Figs. 3A and 5
376 Supplementary). In contrast, particles from RWB were exposed to stronger currents and the
377 site experienced a lower degree of self-recruitment (13.6±3.5%) with 'hot-spots' of higher
378 density particles (up to 0.5%) advected ~100 km northwards (Figs. 3B).

379 Our simulations suggest a high degree of isolation at DYF (Site 3) (Figs. 3C and 5
380 Supplementary), with the majority of particles retained within Cardigan Bay. Due to the weak
381 flows within Cardigan Bay, simulated particles from other sites did not reach DYF within the
382 40-day larval stage. Simulated particles from GAN (Site 4) and BUR (Site 5) were capable of
383 dispersing readily between one another (up to 1% connectivity) and also westwards to the
384 Irish populations, particularly from the more exposed GAN (up to 0.5% connectivity) (Figs.
385 3D-E and 5 Supplementary). Connectivity with DEE was possible but unlikely (<0.01%). It
386 was clear from the simulations that seasonal patterns of dispersal are expected, e.g. from BAN
387 (Site 6) and FLX (Site 7) (Figs. 3F-G and 5 Supplementary). Simulated dispersal from these
388 populations was controlled by the persistent westward residuals around southern Ireland, with
389 the particles generally travelling further west as the residuals strengthened during summer
390 months. For the BAN population, only during April when the Celtic front had not formed,
391 were particles able to disperse northwards into the Irish Sea.
392
393
394 **3.5 Environmental Association analysis**
395 The *ordistep* function found that three variables were linked to the population structure found
396 at neutral loci: July SBTD explained 10% of the genetic variance ($P$=0.002, adj$R^2$=0.10),

dbMEM-1 7% ($P$=0.000, adj$R^2$=0.07) and AEM2 4 %($P$=0.050, adj$R^2$=0.04). Partial RDAs on neutral data using these three variables were not significant.

For the outlier dataset, two environmental factors were found to be highly significant by *ordistep*, explaining 71% of the genetic variance: April min SST ($P$=0.025, adj$R^2$=0.71) and September min SST ($P$=0.004, adj$R^2$=0.71). These SST values were the lowest daily mean SST in each month. The global parsimonious RDA (Fig. 4A) was overall non-significant when including the three selected factors at once ($P$=0.11), although it was significant when including two at a time: dbMEM-1+July mean SBTD ($P$=0.008), AEM2 + dbMEM-1 ($P$=0.022). Partial RDAs were not significant. The parsimonious RDA run on outliers was globally highly significant ($P$=0.002; Fig. 4B).

For the neutral RDA, the first axis explained almost 48% of the total variance, and was mainly driven by dbMEM-1 geographic distance between sites. This axis separates the northern sites of Red Wharf Bay, Dee Estuary and Dyfi Estuary from the southern sites of Flaxfort Strand, Bannow Bay, Gann Estuary and Burry Inlet, describing a north-south divide that is not connected by oceanic currents. The second axis (RDA2) was influenced predominantly by an asymmetrical eigenvector representing modelled connectivity (AEM2) and July sea surface-bottom temperature difference (July mean SBTD), a key determinant of ocean stratification. On this axis, the southernmost population Flaxfort Strand (light green) is strongly related to AEM2, the asymmetrical eigenvector that represents larval dispersal (Fig. 4A). Redundancy analysis carried out on the outliers dataset revealed patterns of local adaptation, with cool sea surface temperatures being the most significant driver (Fig. 4B).

## 4. Discussion

This is the first study to use genomic markers and biophysical larval transport modelling to investigate the drivers of connectivity in a commercially important shellfish in north-western

422 Europe. This research was conducted in the Irish and Celtic Seas where *C. edule* forms
423 valuable shellfisheries for both Ireland and the United Kingdom. A panel of 1725 SNP
424 markers was analysed in relation to important environmental variables such as temperature
425 and oceanographic currents, factors that have been shown to be drivers of population structure
426 in bivalves (Araneda et al., 2016; Gormley et al., 2015; Bernatchez et al., 2019; Lehnert et al.,
427 2019; Xuereb et al., 2018). Using genomic markers, three main genetic groups were identified,
428 which can be considered stocks or management units. Overall, the genetic results fit well with
429 the predictions of the larval transport model, providing a level of empirical validation for both
430 the simulated hydrodynamics and larval behavioural traits. Environmental association
431 analysis revealed that neutral genetic structure was strongly linked to geographical distance
432 between sites and to the strength and direction of the ocean currents acting as corridors for
433 larval dispersal, whereas colder periods (cold SSTs) were identified as the drivers of local
434 adaptation.
435
436
437 **4.1 Connectivity, fine-scale population structure and local adaptation**
438 Neutral genetic diversity was very low across the study area, and compared with previous
439 studies on marine bivalve population genomics (Bernatchez et al., 2019, Lehnert et al., 2019).
440 Populations of cockles in the Irish Sea have been under pressure for at least two decades, with
441 mass mortality events and declines due to overexploitation leading to strict management of
442 most beds across the UK (Woolmer, 2013). In addition, variance in reproductive success is
443 known to occur in bivalves (Hedgecock & Pudovkin, 2011). These events could be responsible
444 for the loss of genetic diversity, as already observed in several marine organisms (Pinsky &
445 Palumbi, 2014). On the other hand, heterozygote deficiency (as indicated by positive $F_{IS}$ in
446 all sampling locations - Table 1) is well known to occur in marine bivalves (Gaffney, 1994),

447 and has already been detected in these populations of common cockles using microsatellites
448 (Coscia et al. 2013). Furthermore, cockles have been shown to undergo boom and bust years
449 (Morgan, O' Riordan, & Culloty, 2013) with the dispersal of cockle larvae and recruitment
450 altering between years and with external parameters such as temperature (Miller, Versace,
451 Matthews, Montgomery, & Bowie, 2013; Morgan et al., 2013).

452 Given the fine spatial scale and the reproductive biology of the study organism
453 (broadcast spawner with a long pelagic larval phase), a lack of, or weak population structure
454 was expected. Previous studies in the same geographic area identified a lack of genetic
455 structure in shellfish species using microsatellite markers (Coscia et al., 2013; Watson et al.,
456 2016; McKeown et al., 2019). In this study, three main genetically distinct units were
457 identified with the neutral marker set. The first group includes the north Wales populations of
458 Red Wharf Bay and Dee Estuary, the second includes Bannow Bay and Flaxfort Strand on the
459 south coast of Ireland and the Gann Estuary in the south-west of Wales, and the final group
460 contains only the Burry Inlet on the south coast of Wales. To our knowledge, this is the first
461 time such a pattern has been detected in *C. edule*, or in any other shellfish resource in the Irish
462 Sea.

463 Geographic proximity certainly favours gene flow, but oceanographic currents aiding
464 larval transport are also major drivers of population structure (Barbut et al., 2019). For
465 example, Red Wharf Bay and the Dee Estuary appear to be genetically homogeneous, due to
466 high levels of gene flow, but these populations are very distinct from those further south (150-
467 350 km away). This concurs with the model's prediction that prevailing currents will mostly
468 disperse larvae from the north coast of Wales northwards. The high levels of gene flow
469 detected between the Gann Estuary on Wales' southwest coast and sites on the southeast coast
470 of Ireland also correspond with the model's prediction of high levels of westward dispersal
471 along the Celtic Sea front (Coscia et al. 2013).

Of particular interest is the genetic make-up of the population in the Burry Inlet. Here, cockles have experienced recurrent mass mortality events for 15 years (https://marinescience.blog.gov.uk/2015/08/14/unusual-cockle-mortalities-burry-inlet/). The relatively high level of genetic differentiation between the neighbouring Burry Inlet and Gann estuary populations indicates that gene flow between these populations is low, seemingly contradicting the larval transport model's prediction of high connectivity. This result suggests that the Burry Inlet population may able to persist through self-recruitment, rather than forming a sink population depending upon immigration from healthier populations elsewhere, and could explain the low genetic diversity and high levels of inbreeding detected there. However, the model showed that larval dispersal is possible from the Gann to Burry estuaries, although we assumed a relatively large settlement zone that extended beyond the mouth of the Burry Inlet (Fig. 6 Supplementary). Furthermore, it must be acknowledged that the Gann estuary cockle population is far smaller than that of the Burry Inlet so larvae dispersing from the Gann into the Burry may be swamped by self-recruitment, or might not survive due to local selection against them. Additionally, the southern coast of Wales contains other large cockle populations, such as the Three Rivers fishery mid-way between the Gann and Burry. These populations were not sampled/modelled in this study but may provide a nearer and greater source of larvae for the Burry Inlet than does the Gann Estuary.

The Spatial Environmental Association analysis identified several environmental factors highly associated with the population structure observed at neutral and outlier genetic markers. This is a strong statistical approach, which has already been successfully employed to study the influence of the environment on the genetic structure of commercially important bivalves, such as the eastern oyster (*Crassostrea virginica*; Bernatchez et al. 2019) and Atlantic deep-sea scallop (*Placopecten magellanicus*; Lehnert et al. 2019). In *C. edule,* neutral genetic structure is strongly dependent on geographic distance between sites (dbMEM-1),

497 indicating that isolation by distance plays an important role in shaping the observed genetic
498 structure in this species, despite its long pelagic larval duration. Nevertheless, it is the interplay
499 between isolation by distance, peak temperature-driven currents (July SBTD) and modelled
500 connectivity (AEM2) that shapes neutral population structure. The summer stratification
501 which strengthens the Celtic Sea front current, directed from south Wales to Ireland (Simpson
502 and Hunter 1974), plays a major role in connecting cockle populations between the south of
503 Wales and Ireland, while separating them from populations further north in the Irish Sea.
504 Considering the 14 outlier loci, the minimum sea surface temperatures recorded in April and
505 September explained the genetic variance observed within this dataset. In particular, the Burry
506 Inlet was strongly associated with the SSTs in September, which are warmer compared to the
507 ones recorded at other locations for the same time (Figure 1 Supplementary). If cockles in the
508 Burry Inlet are indeed adapted to warmer sea surface temperatures than populations at sites
509 further west or north (and especially compared to those at Gann), this may explain the
510 maintenance of genetic differentiation between the Burry Inlet and the Gann Estuary despite
511 their spatial proximity and the potential for high connectivity predicted by the model (Fig 3
512 and Fig. 5 Supplementary Material). Larvae arriving adrift from Gann to the Burry may not
513 be able to survive post-recruitment given selective pressure against them by higher
514 temperatures.
515
516 **4.2 Implications for management**
517 The results of the biophysical modelling are likely to have the greatest relevance to fishery
518 management in terms of the potential seasonal variability in larval supply to cockle grounds.
519 This study demonstrates the existence of three distinct units of cockles using both genomic
520 tools and larval dispersal modelling. As with other recent studies (Lal, Southgate, Jerry,
521 Bosserelle, & Zenger, 2017) these findings have important implications for fishery

522 management (Coscia et al. 2013; Miller et al. 2013) and how fisheries management can be
523 reconciled with conservation and other activities.

524 Given the incidence of recurrent mass mortality events at the Burry Inlet (Callaway et
525 al. 2013), the genetic isolation of this cockle fishery implied by the results of this study should
526 be investigated further. This could be achieved by expanding the sampling coverage of the
527 Burry Inlet to multiple sites and years and assessing its connectivity to other nearby cockle
528 beds that have not been included in this study. Future investigations should be aimed at
529 clarifying the role of local adaptation into the fine-scale population dynamics of the common
530 cockle in this area, in order to improve management, while also assessing the role played by
531 diseases and infections. The results from this study highlight the importance of the use of
532 genomic and hydrodynamic data in assessing population structure and connectivity in an
533 exploited and commercially important marine species and may aid in current and long-term
534 management regimes of this species (Lal et al. 2016; Bernatchez et al. 2017).
535

536 **Data availability**

537 The data that support the findings of this study are openly available in Dryad at
538 http://doi.org/[doi], reference number [reference number].
539

540 **Acknowledgements**

541 The genomic data was generated thanks to an internal grant awarded to IC by Aberystwyth
542 University. The authors also wish to acknowledge the support of the Interreg Ireland-Wales
543 Programme ISPP (http://ispp.bangor.ac.uk/) and Bluefish (bluefishproject.com) projects, the
544 Interreg Atlantic Area Programme Cockles project (https://cockles-project.eu/), and the
545 SUSFISH project. Oceanographic data was provided through the SEACAMS project
546 (www.seacams.ac.uk), funded by the Welsh Government, the Higher Education Funding

Council for Wales, the Welsh European Funding Office, and the European Regional Development Fund Convergence Programme. The model simulations were conducted on the Supercomputing-Wales high performance computing(www.supercomputing.wales) system (a collaboration between Welsh universities and the Welsh Government) , supported by Ade Fewings and Aaron Owen. The authors are grateful to all the people that have helped with sampling, Mandi Knott from the North Western Inshore Fisheries and Conservation Authority and Emer Morgan from University College Cork, as well as in the lab (Matt Hegarty and Elizabeth Harding), and the colleagues Niall McKeown and Hayley V. Watson for continuous help and support. We also are extremely grateful to Paulino Martinez for constructive comments on the manuscript.

## Bibliography

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, *26*(2), 420–430.

Araneda, C., Larraín, M. A., Hecht, B., & Narum, S. (2016). Adaptive genetic variation distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine environments. *Ecology and Evolution*, *6*(11), 3632–3644.

Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, *16*(24), 5115–5139.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLOS ONE*, *3*(10), e3376.

Barbut, L., Crego, C. G., Delerue-Ricard, S., Vandamme, S., Volckaert, F. A. M., & Lacroix, G. (2019). How larval traits of six flatfish species impact connectivity. *Limnology and Oceanography*, *64*(3), 1150–1171.

573 Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among
574 populations from genome scans. *Molecular Ecology*, *13*(4), 969–980.
575 Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD
576 genotyping reveals fine-scale genetic structuring and provides powerful population
577 assignment in a widely distributed marine species, the American lobster (*Homarus*
578 *americanus*). *Molecular Ecology*, *24*(13), 3299–3315.
579 Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., … Bernatchez,
580 L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated
581 population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, *25*(20),
582 5073–5092.
583 Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., …
584 Withler, R. E. (2017). Harnessing the Power of Genomics to Secure the Future of Seafood.
585 *Trends in Ecology & Evolution*, *32*(9), 665–680.
586 Bernatchez, S., Xuereb, A., Laporte, M., Benestan, L., Steeves, R., Laflamme, M., … Bernatchez, L.
587 (2019). Seascape genomics of eastern oyster (Crassostrea virginica) along the Atlantic coast
588 of Canada. *Evolutionary Applications*, *12*.
589 Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Modelling directional spatial processes in
590 ecological data. *Ecological Modelling*, *215*(4), 325–336.
591 Blanchet, F. G., Legendre, P., Maranger, R., Monti, D., & Pepin, P. (2011). Modelling the effect of
592 directional spatial ecological processes at different scales. *Oecologia*, *166*(2), 357–368.
593 Bode, M., Leis, J. M., Mason, L. B., Williamson, D. H., Harrison, H. B., Choukroun, S., & Jones, G.
594 P. (2019). Successful validation of a larval dispersal model using genetic parentage data.
595 *PLOS Biology*, *17*(7), e3000380.
596 Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical Ecology with R*. Springer.
597 Botsford, L. W., White, J. W., Coffroth, M.-A., Paris, C. B., Planes, S., Shearer, T. L., … Jones, G. P.
598 (2009). Connectivity and resilience of coral reef metapopulations in marine protected areas:
599 Matching empirical efforts to predictive needs. *Coral Reefs*, *28*(2), 327–337.

600 Bullimore, B. (2014). Problems and pressures, management and measures in a site of marine
601 conservation importance: Carmarthen Bay and Estuaries. *Estuarine, Coastal and Shelf*
602 *Science*, *150*, 288–298.
603 Burgess, S. D., Bowring, S., & Shen, S. (2014). High-precision timeline for Earth's most severe
604 extinction. *Proceedings of the National Academy of Sciences*, *111*(9), 3316–3321.
605 Callaway, R., Burdon, D., Deasey, A., Mazik, K., & Elliott, M. (2013). The riddle of the sands: How
606 population dynamics explains causes of high bivalve mortality. *Journal of Applied Ecology*,
607 *50*(4), 1050–1059.
608 Chambers, J. M. (2013). SoDA: Functions and Examples for "Software for Data Analysis" (Version
609 1.0-6).
610 Coscia, I, Vogiatzi, E., Kotoulas, G., Tsigenopoulos, C. S., & Mariani, S. (2012). Exploring neutral
611 and adaptive processes in expanding populations of gilthead sea bream, *Sparus aurata* L., in
612 the North-East Atlantic. *Heredity*, *108*(5), 537–546.
613 Coscia, I., Robins, P., Porter, J., Malham, S., & Ironside, J. (2013). Modelled larval dispersal and
614 measured gene flow: Seascape genetics of the common cockle *Cerastoderma edule* in the
615 southern Irish Sea. *Conservation Genetics*, *14*, 451.
616 Cowen, R. K., Gawarkiewicz, G., Pineda, J., Thorrold, S., & E. Werner, F. (2007). Population
617 Connectivity in Marine Systems: An Overview. *Oceanography*, *20*.
618 Cowen, R. K., & Sponaugle, S. (2009). Larval Dispersal and Marine Population Connectivity. *Annual*
619 *Review of Marine Science*, *1*(1), 443–466.
620 Dare, P., Bell, M. C., Walker, P. H., & Bannister, R. A. (2003). *Historical and current status of cockle*
621 *and mussel stocks in The Wash*.
622 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., … Vitart, F. (2011).
623 The ERA-Interim reanalysis: Configuration and performance of the data assimilation system.
624 *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.
625 Dray, S., Pélissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P. R., … Wagner, H. H.
626 (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological*
627 *Monographs*, *82*(3), 257–275.

628 Dray, Stéphane, Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., … Wagner, H. H.
629 (2019). adespatial: Multivariate Multiscale Spatial Analysis (Version 0.3-7).
630 Dray, Stéphane, Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive
631 framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological*
632 *Modelling*, *196*(3), 483–493.
633 Elliott, M., & Holden, J. (2017). *UK Sea Fisheries Statistics 2017* (p. 174). Marine Management
634 Organisation.
635 Etter, P. D., Preston, J. L., Bassham, S., Cresko, W. A., & Johnson, E. A. (2011). Local De Novo
636 Assembly of RAD Paired-End Contigs Using Short Sequencing Reads. *PLOS ONE*, *6*(4),
637 e18561.
638 Fermer, J., Culloty, S. C., Kelly, T. C., & O'Riordan, R. M. (2011). Parasitological survey of the
639 edible cockle *Cerastoderma edule* (Bivalvia) on the south coast of Ireland. *Journal of the*
640 *Marine Biological Association of the United Kingdom*, *91*(4), 923–928.
641 Flach, E. C., & de Bruin, W. (1994). Does the activity of cockles, *Cerastoderma edule* (L.) and
642 lugworms, *Arenicola marina* L., make *Corophium volutator* Pallas more vulnerable to
643 epibenthic predators: A case of interaction modification? *Journal of Experimental Marine*
644 *Biology and Ecology*, *182*(2), 265–285.
645 Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for
646 Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, *180*(2), 977–
647 993.
648 Gaffney, PM. (1994). Heterosis and heterozygote deficiencies in marine bivalves: More light? In
649 *Genetics and Evolution of Aquatic Organisms*. Chapman & Hall.
650 Gimenez, L. (2019). Incorporating the geometry of dispersal and migration to understand spatial
651 patterns of species distributions. *Ecography*, *42*(4), 643–657.
652 Gormley, K., Mackenzie, C., Robins, P., Coscia, I., Cassidy, A., James, J., … Porter, J. (2015).
653 Connectivity and Dispersal Patterns of Protected Biogenic Reefs: Implications for the
654 Conservation of *Modiolus modiolus* (L.) in the Irish Sea. *PLOS ONE*, *10*(12), e0143337.
655 Gosselin, T. (2019). assigner: Assignment Analysis with GBS/RAD Data using R (Version 0.5.6).

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186.

Goudet, Jerome, & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics (Version 0.04-22).

Graham, J. A., O'Dea, E., Holt, J., Polton, J., Hewitt, H. T., Furner, R., … Mayorga Adame, C. G. (2018). AMM15: A new high-resolution NEMO configuration for operational simulation of the European north-west shelf. *Geoscientific Model Development*, *11*(2), 681–696.

Hartnett, M., Berry, A., Tully, O., & Dabrowski, T. (2007). Investigations into the transport and pathways of scallop larvae—the use of numerical models for managing fish stocks. *Journal of Environmental Monitoring*, *9*(5), 403–410.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes Reproductive Success in Highly Fecund Marine Fish and Shellfish: A Review and Commentary. *Bulletin of Marine Science*, *87*(4), 971–1002.

Hervas, A., Tully, O., Hickey, J., O'Keeffe, E., & Kelly, E. (2007). Assessment, Monitoring and Management of the Dundalk Bay and Waterford Estuary Cockle *(Cerastoderma edule) Fisheries in 2007* (p. 44).

Hickin, V. (2019). *Fisheries and Nature Conservation Issues of the Cockle Fishery in the Welsh District of the North Western and North Wales Sea Fisheries Committee*.

Horsburgh, K. J., Hill, A. E., & Brown, J. (1998). A summer jet in the St George's Channel of the Irish Sea. *Estuarine, Coastal and Shelf Science*, *47*(3), 285–294.

ICES. (2018). *Interim Report of the Working Group on the Application of Genetics in Fisheries and Aquaculture (WGAGFA)* (p. 41).

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

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071.

Jombart, T., Kamvar, Z. N., Collins, C., Lustrik, R., Beugin, M.-P., Knaus, B. J., … Ewing, R. J. (2018). adegenet: exploratory analysis of genetic and genomic data (Version 2.1.1).

684 Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide
685 population genetic data with emphasis on clonality. *Frontiers in Genetics*, *6*.
686 Kamvar, Z. N., Tabima, J. F., Everhart, S. E., Brooks, J. C., Krueger-Hadfield, S. A., Sotka, E., …
687 Grünwald, N. J. (2019). poppr: genetic analysis of populations with mixed reproduction
688 (Version 2.8.3). R
689 Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of
690 populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.
691 Lal, M., C. Southgate, P., Jerry, D., Bosserelle, C., & Zenger, K. (2017). Swept away: Ocean currents
692 and seascape features influence genetic structure across the 18,000 Km Indo-Pacific
693 distribution of a marine invertebrate, the black-lip pearl oyster *Pinctada margaritifera*. *BMC*
694 *Genomics*, *18*.
695 Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of
696 species data. *Oecologia*, *129*(2), 271–280.
697 Lehnert, S. J., DiBacco, C., Wyngaarden, M. V., Jeffery, N. W., Lowen, J. B., Sylvester, E. V. A., …
698 Bradbury, I. R. (2019). Fine-scale temperature-associated genetic structure between inshore
699 and offshore populations of sea scallop (*Placopecten magellanicus*). *Heredity*, *122*(1), 69.
700 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The
701 Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.
702 Longshaw, M., & Malham, S. (2015). A review of the infectious agents, parasites, pathogens and
703 commensals of European cockles (*Cerastoderma edule* and *C. glaucum*). *Journal of the*
704 *Marine Biological Association of the United Kingdom*, *93*, 227–247.
705 Malham, S., Hutchinson, T., & Longshaw, M. (2012). A review of the biology of European cockles
706 (Cerastoderma spp.). *Journal of the Marine Biological Association of the United Kingdom*,
707 *92*, 1563–1584.
708 Maroso, F., Hillen, J. E. J., Pardo, B. G., Gkagkavouzis, K., Coscia, I., Hermida, M., … Ogden, R.
709 (2018). Performance and precision of double digestion RAD (ddRAD) genotyping in large
710 multiplexed datasets of marine fish species. *Marine Genomics*, 39, 64–72.

711 Martínez, L, Mendez, J., Insua, A., Arias-Pérez, A., & Freire, R. (2013). Genetic diversity and
712 population differentiation in the cockle Cerastoderma edule estimated by microsatellite
713 markers. *Helgoland Marine Research*, *67*, 179–189.
714 Martínez, Luisa, Freire, R., Arias-Pérez, A., Mendez, J., & Insua, A. (2015). Patterns of genetic
715 variation across the distribution range of the cockle *Cerastoderma edule* inferred from
716 microsatellites and mitochondrial DNA. *Marine Biology*, 162(7), 1393–1406.
717 McKeown, N. J., Watson, H. V., Coscia, I., Wootton, E., & Ironside, J. E. (2019). Genetic variation in
718 Irish Sea brown crab (*Cancer pagurus* L.): Implications for local and regional management.
719 *Journal of the Marine Biological Association of the United Kingdom*, *99*(4), 879–886.
720 Miller, A. D., Versace, V. L., Matthews, T. G., Montgomery, S., & Bowie, K. C. (2013). Ocean
721 currents influence the genetic structure of an intertidal mollusc in southeastern Australia –
722 implications for predicting the movement of passive dispersers across a marine biogeographic
723 barrier. *Ecology and Evolution*, *3*(5), 1248–1261.
724 Monzón-Argüello C., Dell'Amico F., Morinière P., Marco A., López-Jurado L. F., Hays Graeme C.,
725 … Lee Patricia L. M. (2012). Lost at sea: Genetic, oceanographic and meteorological
726 evidence for storm-forced dispersal. *Journal of The Royal Society Interface*, *9*(73), 1725–
727 1732.
728 Morgan, E., O' Riordan, R. M., & Culloty, S. C. (2013). Climate change impacts on potential
729 recruitment in an ecosystem engineer. *Ecology and Evolution*, *3*(3), 581–594.
730 Nielsen, E. E., Cariani, A., Aoidh, E. M., Maes, G. E., Milano, I., Ogden, R., … Carvalho, G. R.
731 (2012). Gene-associated markers provide tools for tackling illegal fishing and false eco-
732 certification. *Nature Communications*, *3*, 851.
733 North, E., Schlag, Z., Hood, R., Li, M., Zhong, L., Gross, T., & Kennedy, V. (2008). Vertical
734 swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle-
735 tracking and hydrodynamic model of Chesapeake Bay. *Marine Ecology Progress Series*, *359*,
736 99–115.
737 Oksanen, F., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., …. (2017). vegan:
738 Community Ecology Package. R package version 2.4-4.

739 Paris, C., K. Cowen, R., Claro, R., & C. Lindeman, K. (2005). Larval transport pathways from Cuban
740 snapper (Lutjanidae) spawning aggregations based on biophysical modeling. *Marine Ecology-*
741 *Progress Series 296*, 93–106.
742 Pingree, R. D., & Griffiths, D. K. (1979). Sand transport paths around the British Isles resulting from
743 M2 and M4 tidal interactions. *Journal of the Marine Biological Association of the United*
744 *Kingdom*, *59*(2), 497–513.
745 Pinsky, M. L., & Palumbi, S. R. (2014). Meta-analysis reveals lower genetic diversity in overfished
746 populations. *Molecular Ecology*, *23*(1), 29–39.
747 R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. https://www.R-
748 project.org/
749 Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of
750 Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589.
751 Razgour, O., Forester, B., Taggart, J. B., Bekaert, M., Juste, J., Ibáñez, C., … Manel, S. (2019).
752 Considering adaptive genetic variation in climate change vulnerability assessment reduces
753 species range loss projections. *Proceedings of the National Academy of Sciences*, *116*(21),
754 10418–10423.
755 Robins, P. E., Neill, S. P., Giménez, L., Jenkins, S. R., & Malham, S. K. (2013). Physical and
756 biological controls on larval dispersal and connectivity in a highly energetic shelf sea.
757 *Limnology and Oceanography*, *58*(2), 505–524.
758 Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans
759 for signatures of selection. *BMC Evolutionary Biology*, *12*(1), 94.
760 Rowley, A. F., Cross, M. E., Culloty, S. C., Lynch, S. A., Mackenzie, C. L., Morgan, E., … Malham,
761 S. K. (2014). The potential impact of climate change on the infectious diseases of
762 commercially important shellfish populations in the Irish Sea—a review. *ICES Journal of*
763 *Marine Science*, *71*(4), 741–759.
764 Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., … Toonen, R. J.
765 (2016). A decade of seascape genetics: Contributions to basic and applied marine
766 connectivity. *Marine Ecology Progress Series*, *554*, 1–19.

767 Simpson, J. H., & Hunter, J. R. (1974). Fronts in the Irish Sea. *Nature*, *250*(5465), 404.

768 Stanev, E. V., Badewien, T. H., Freund, H., Grayek, S., Hahner, F., Meyerjürgens, J., … Zielinski, O.
769 (2019). Extreme westward surface drift in the North Sea: Public reports of stranded drifters
770 and Lagrangian tracking. *Continental Shelf Research*, *177*, 24–32.

771 Teacher, A. G., André, C., Jonsson, P. R., & Merilä, J. (2013). Oceanographic connectivity and
772 environmental correlates of genetic structuring in Atlantic herring in the Baltic Sea.
773 *Evolutionary Applications*, *6*(3), 549–567.

774 Thieltges, D. W. (2006). Parasite-induced summer mortality in the cockle *Cerastoderma edule* by the
775 trematode *Gymnophallus choledochus*. *Hydrobiologia*, *559*(1), 455–461.

776 Truelove, N., J. Box, S., Aiken, K., Blythe-Mallett, A., Boman, E., J. Booker, C., … Stoner, A.
777 (2017). Isolation by oceanic distance and spatial genetic structure in an overharvested
778 international fishery. *Diversity and Distributions*, 23(11), 1292-1300.

779 Viricel, A., & Rosel, P. E. (2014). Hierarchical population structure and habitat differences in a highly
780 mobile marine species: the Atlantic spotted dolphin. *Molecular Ecology*, *23*(20), 5018–5035.

781 Watson, H. V., McKeown, N. J., Coscia, I., Wootton, E., & Ironside, J. E. (2016). Population genetic
782 structure of the European lobster (*Homarus gammarus*) in the Irish Sea and implications for
783 the effectiveness of the first British marine protected area. *Fisheries Research*, *183*, 287–293.

784 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population
785 Structure. *Evolution*, *38*(6), 1358–1370.

786 Wolff, W. J. (2005). The exploitation of living resources in the Dutch Wadden Sea: A historical
787 overview. *Helgoland Marine Research*, *59*(1), 31–38.

788 Woodings, L. N., Murphy, N. P., Doyle, S. R., Hall, N. E., Robinson, A. J., Liggins, G. W., …
789 Strugnell, J. M. (2018). Outlier SNPs detect weak regional structure against a background of
790 genetic homogeneity in the Eastern Rock Lobster, *Sagmariasus verreauxi*. *Marine Biology*,
791 *165*(12), 185.

792 Woolmer, A. (2013). *Review of National Cockle Mortality Issues and Options for Fishery*
793 *Management in Kent and Essex IFCA* (p. 56). Retrieved from https://www.kentandessex-
794 ifca.gov.uk/wp-content/uploads/2015/07/report-of-EU-and-UK-cockle-mortality.pdf

Van Wyngaarden, M., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Zhan, L., … Bradbury, I. R. (2018). Oceanographic variation influences spatial genomic structure in the sea scallop, *Placopecten magellanicus*. *Ecology and Evolution*, *8*(5), 2824–2841.

Xuereb, A., Benestan, L., Normandeau, E., MR Curtis, J., Bernatchez, L., & Fortin, M. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, 27(10), 2347-2364.

Young, E. F., Belchier, M., Hauser, L., Horsburgh, G. J., Meredith, M. P., Murphy, E. J., … Carvalho, G. R. (2015). Oceanography and life history predict contrasting genetic population structure in two Antarctic fish species. *Evolutionary Applications*, *8*(5), 486–509.

821 **Tables and Figures**

822

823 ***Table 1:*** *genetic diversity indices for the two datasets. N, number of individuals remaining*
824 *after filtering for each location;* $H_O$*, observed heterozygosity,* $H_E$*, expected heterozygosity*
825 *and* $F_{IS}$*, the inbreeding coefficient. Values that are significant (95% confidence interval) are*
826 *in bold.*

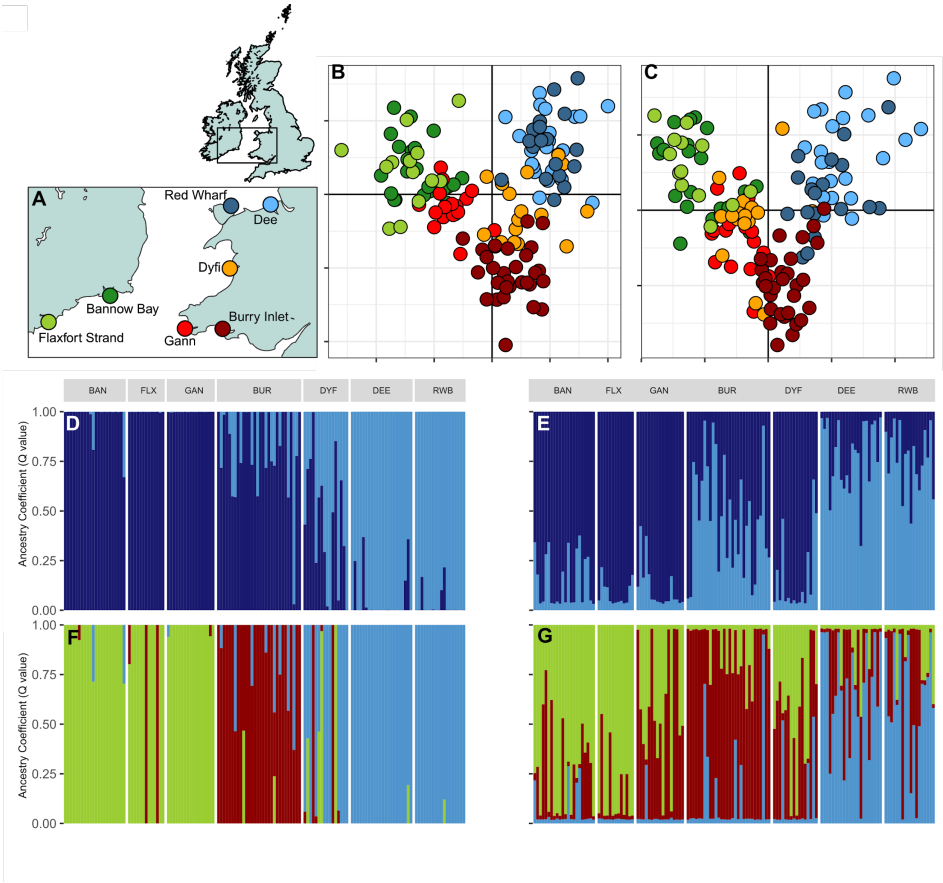
| | | | *neutral* | | | *outlier* | | |
|---|---|---|---|---|---|---|---|---|
| | | *N* | $H_O$ | $H_E$ | $F_{IS}$ | $H_O$ | $H_E$ | $F_{IS}$ |
| Bannow Bay | BAN | 22 | 0.133 | 0.150 | **0.087** | 0.269 | 0.272 | 0.006 |
| Gann | GAN | 17 | 0.129 | 0.147 | **0.094** | 0.246 | 0.240 | 0.032 |
| Dee | DEE | 22 | 0.116 | 0.148 | **0.173** | 0.218 | 0.355 | **0.335** |
| Dyfi | DYF | 11 | 0.103 | 0.149 | **0.238** | 0.146 | 0.234 | **0.394** |
| Flaxfort Strand | FLX | 13 | 0.103 | 0.156 | **0.275** | 0.197 | 0.266 | **0.201** |
| Red Wharf | RWB | 18 | 0.118 | 0.144 | **0.141** | 0.286 | 0.337 | **0.179** |
| Burry Inlet | BUR | 30 | 0.109 | 0.150 | **0.230** | 0.173 | 0.218 | 0.159 |

827

828

829

830




831

832 ***Figure 1:*** *A) Sampling locations; B) DAPC analysis using the neutral dataset; C) DAPC*

833 *analysis using the 14 outliers; below, the barplots generated by fastSTRUCTURE from the*

834 *neutral (D and F) and outlier (E and G) datasets, for both K=2 and K=3.*

835

836

837

838

839

840

841

842

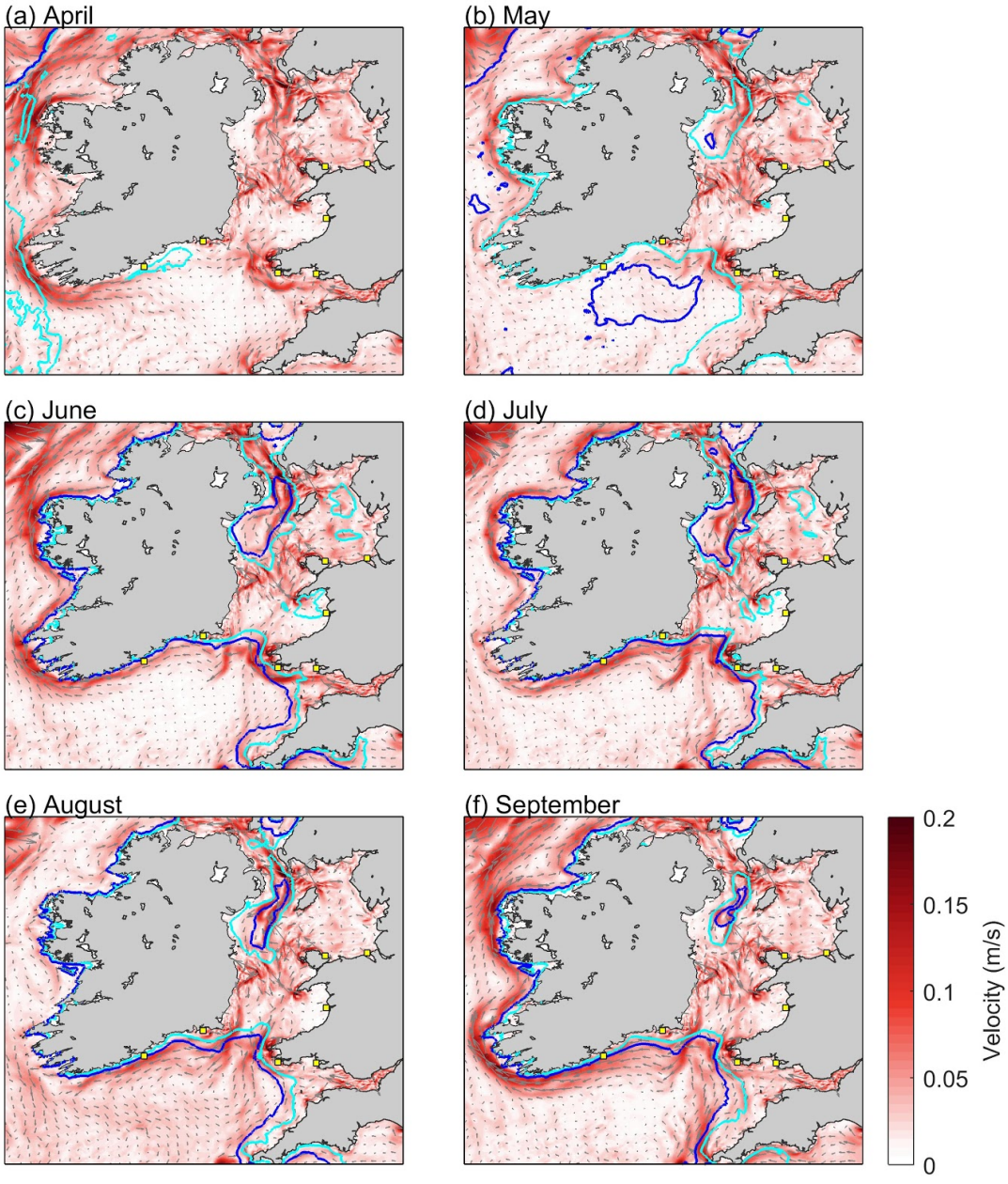
843

844

845

846

847




848

849 ***Figure 2:*** *Simulated development of stratification and density-driven currents in the Irish and*

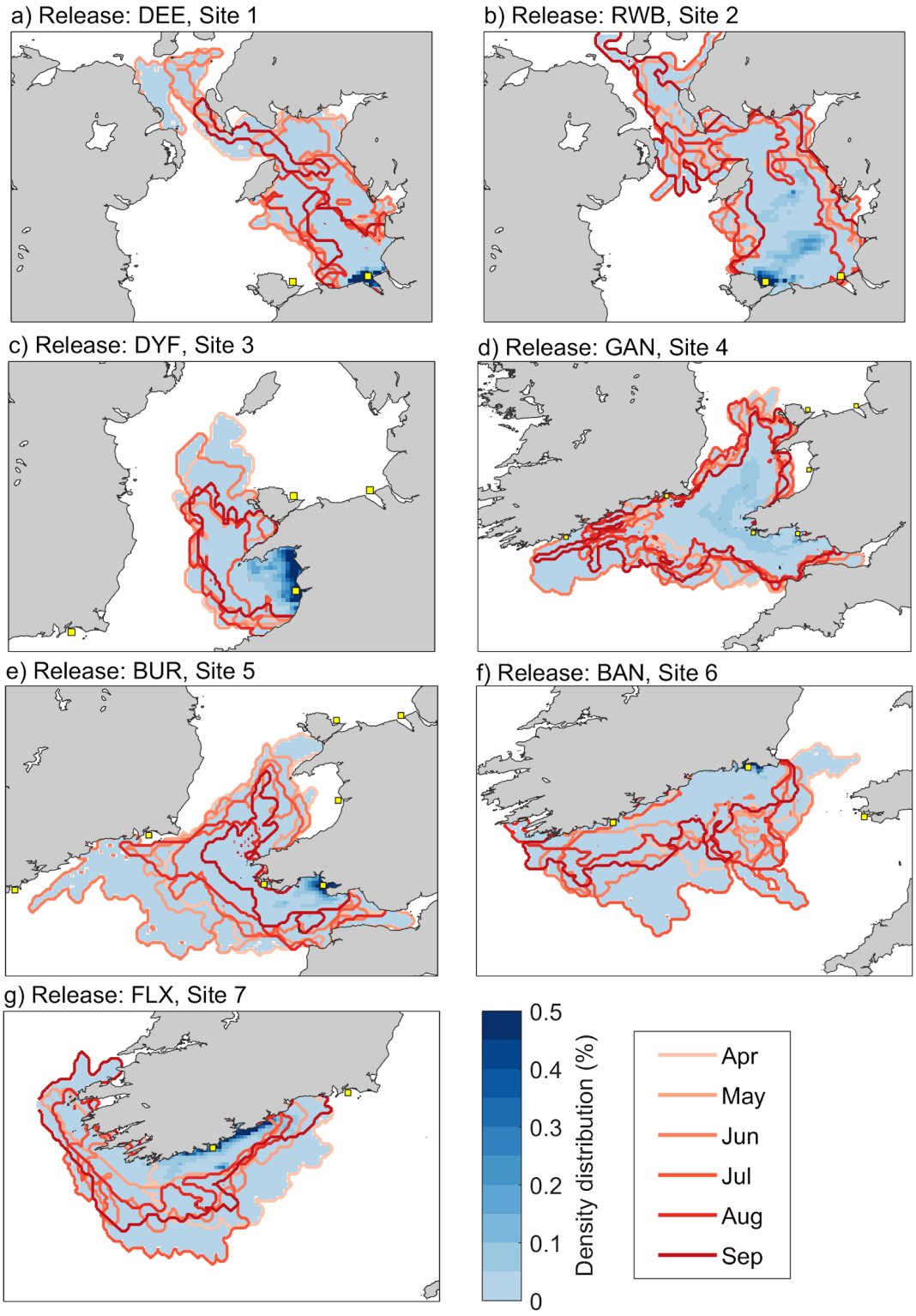
850 *Celtic Seas, during April-September 2014 (a-f). Light and dark blue contours denote surface-*

851 *bottom temperature differences of 1°C and 2.5°C, respectively. Each panel shows monthly-*

852 *averaged and depth-averaged velocity magnitude (red shading) and direction (vectors).*

853 *Sample locations are indicated with yellow squares.*

854



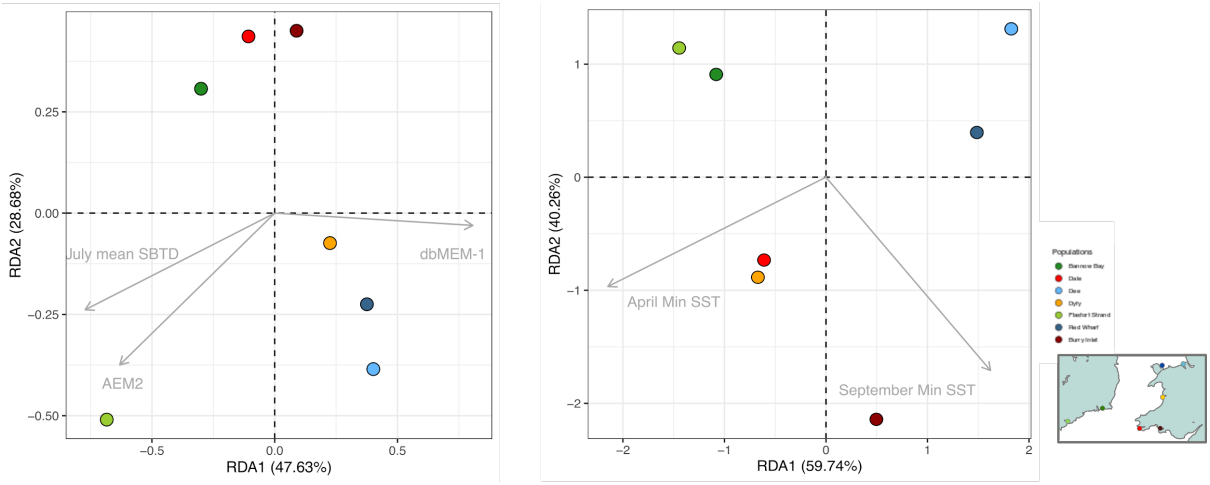

855

856 ***Figure 3:*** *Probability density distribution maps showing simulated dispersal probability from*

857 *release sites 1-7 (yellow squares). Each panel (blue colour scale in a-g) shows dispersal range*

858 *for 11,520,000 particles (12,000 each month × 6 months × 10 settlement days). The*

859 *superimposed contour lines show the maximum spread per month, hence an indication of*
860 *seasonal dispersal variability.*
861
862
863
864
865




866

867 ***Figure 4****: Redundancy analysis (parsimonious RDA) performed for the (A) neutral and (B)*
868 *outlier datasets. Each circle is a sampling location, and each arrow is an environmental*
869 *variable that significantly drives the observed population structure.*
870
871
872
873
874
875
876
877

878 **SUPPLEMENTARY MATERIAL**

879 **Table 1 Supplementary** Filtering steps. *#individuals*, remaining number of individuals after

880 each filtering step; *#SNPs* remaining number of SNPs

881

| | **# individuals** | **# SNPs** |
|---|---|---|
| **Raw dataset** | 191 | 4271 |
| **Missing data** | 159 | 4271 |
| $F_{IS}$ | 138 | 3488 |
| **MAF 1%** | 138 | 1864 |
| **Linkage Disequilibrium** | 138 | 1775 |
| **Hardy-Weinberg eq.** | 138 | 1725 |

882

883

884

885

886

887

888

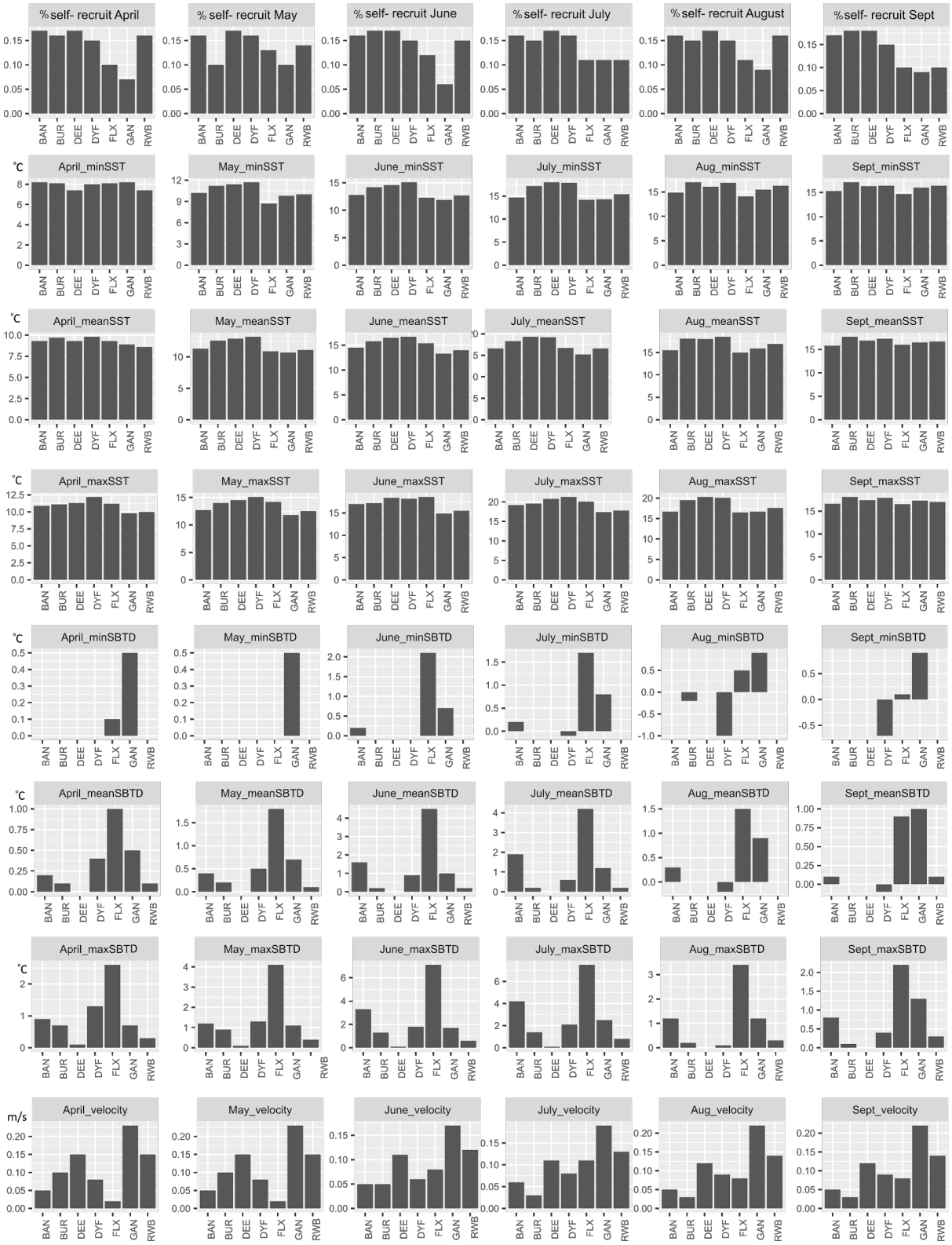



***Figure 1 Supplementary***: Environmental parameters for each location, and each month used in the environmental association analysis.

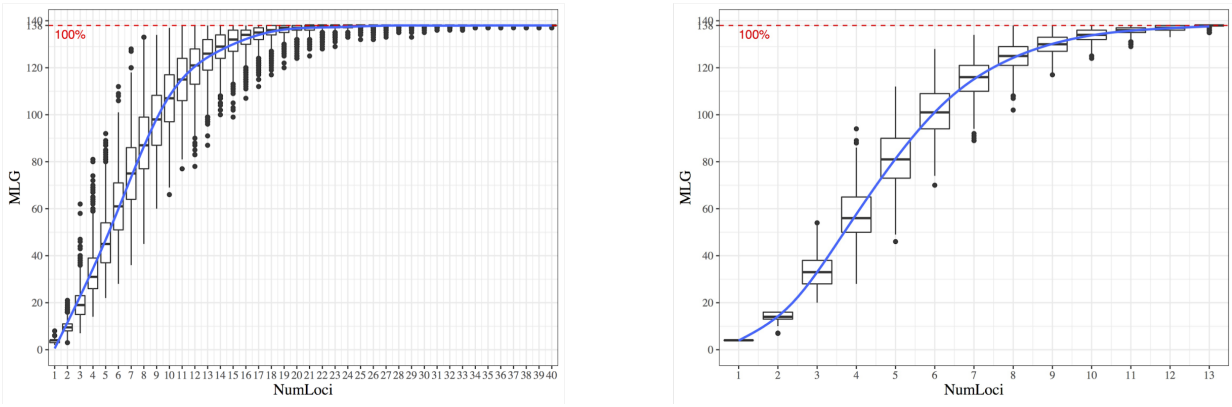
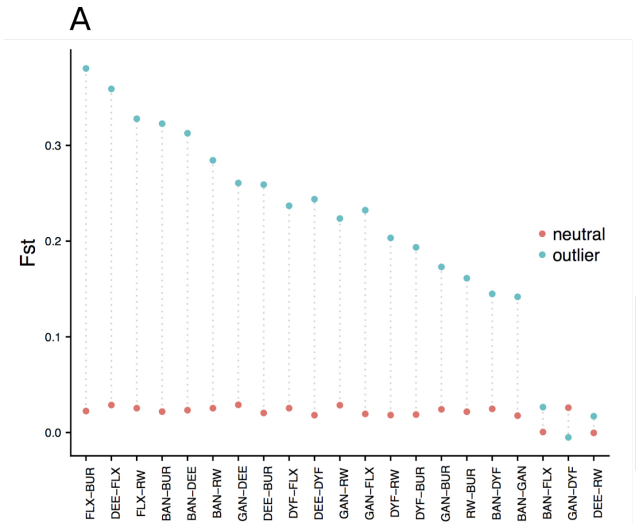


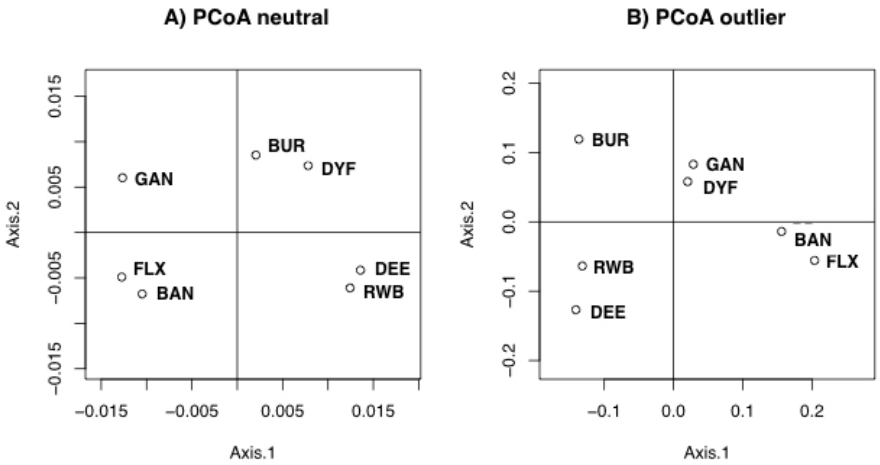

***Figure 2 Supplementary*** *Genotype accumulation curves for neutral (A) and outlier (B) datasets. On the Y axis, the number of multilocus genotypes and on the X axis, the number of loci.*




B

| | BAN | GAN | DEE | DYF | FLX | RWB | BUR |
|---|---|---|---|---|---|---|---|
| BAN | | **0.142** | **0.313** | **0.145** | 0.026 | **0.284** | **0.323** |
| GAN | **0.018** | | **0.261** | -0.005 | **0.232** | **0.224** | **0.173** |
| DEE | **0.023** | **0.029** | | **0.244** | **0.359** | 0.017 | **0.259** |
| DYF | **0.025** | **0.026** | **0.018** | | **0.237** | **0.203** | **0.193** |
| FLX | 0.001 | **0.019** | **0.029** | **0.025** | | **0.328** | **0.380** |
| RWB | **0.025** | **0.029** | 0.000 | **0.018** | **0.025** | | **0.161** |
| BUR | **0.022** | **0.024** | **0.020** | **0.019** | **0.022** | **0.022** | |

***Figure 3 Supplementary:*** *Pairwise Fst calculated for neutral and outlier datasets. The graph (A) shows which pairs have the highest difference between the two. In the table (B) the relative values for neutral (lower diagonal) and outliers (upper diagonal). In bold, values that are significant (95% confidence interval).*




906

907 ***Figure 4 Supplementary*** Principal Coordinates Analysis (MDS plots) based on $F_{ST}$

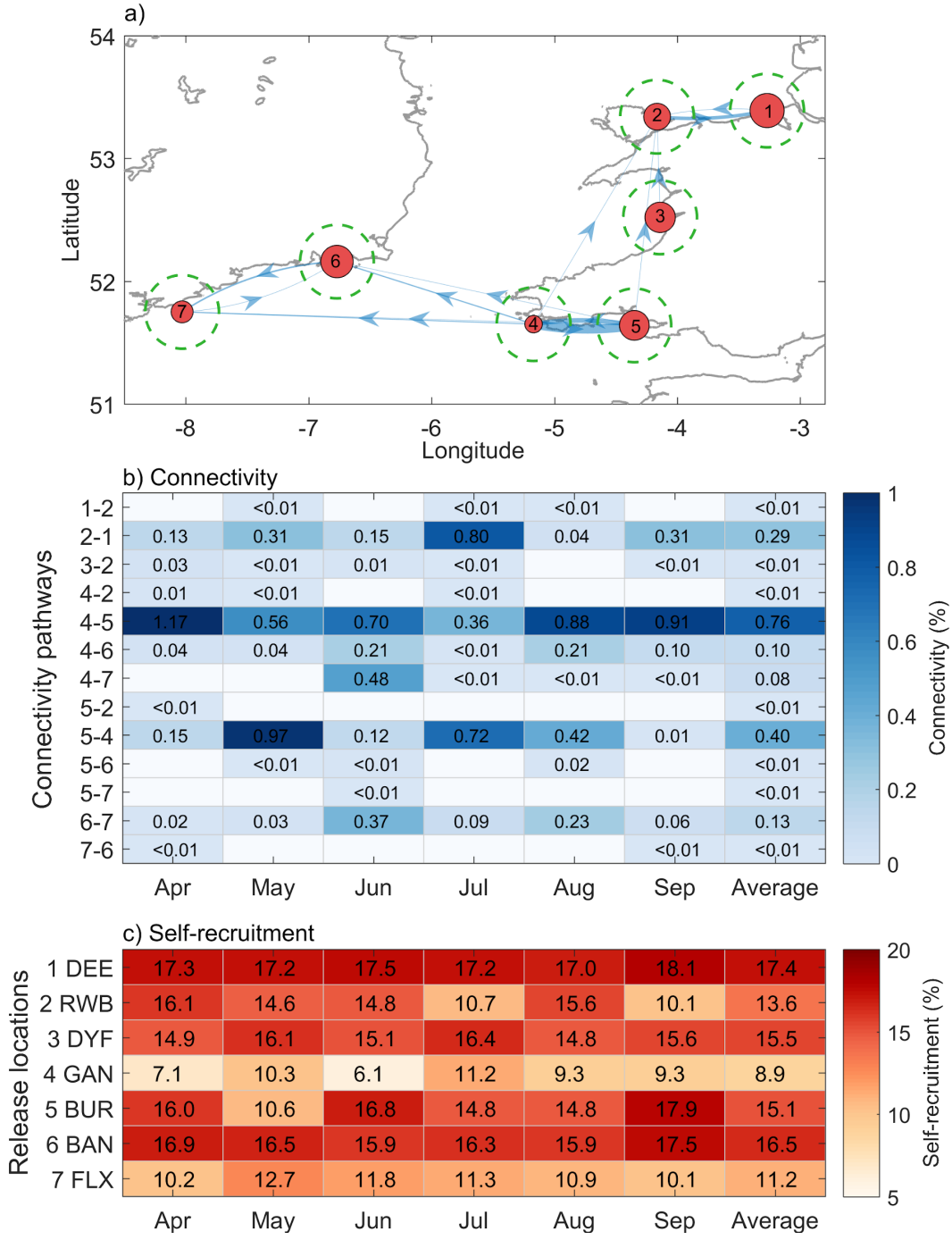



***Figure 5 Supplementary:*** *Simulated connectivity between the seven sampled cockle populations. Seasonally-averaged (2014) connectivity networks are shown geographically in (a). The thickness of the pathways in (a) corresponds to the average values in (b). Self-recruitment in (a) is denoted by the size of the red circles which correspond to the average values in (c). The dashed green circles show the settlement radius used to calculate*

914 *connectivity. Seasonal variability (Apr.-Sep.) is shown for connectivity (b) and self-*
915 *recruitment (c).*
916
917
918
919
920
921
922
923

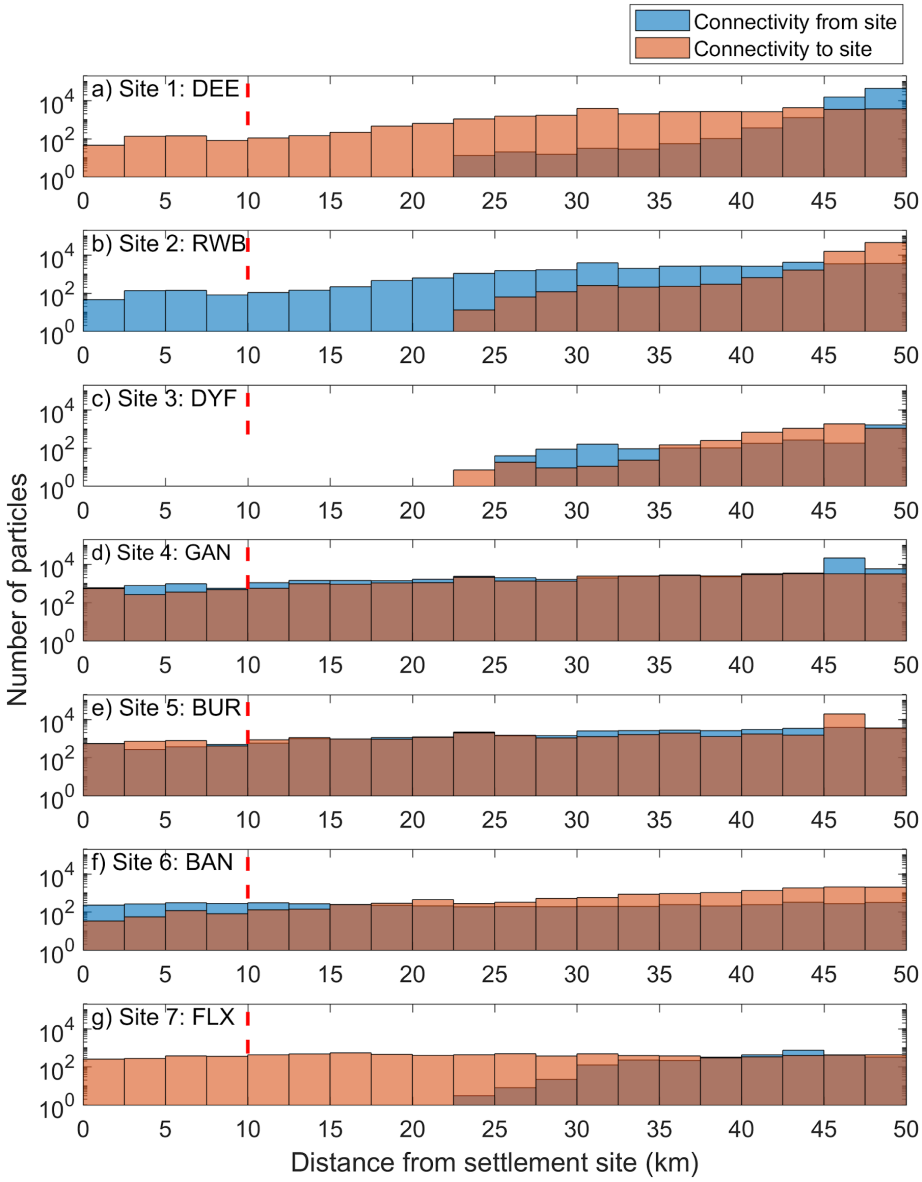



***Figure 6 Supplementary**: Histogram plots showing connectivity as a function of distance. For each particle trajectory, the minimum distance to a settlement site during days 30-40 is presented. Blue bars signify the connectivity potential of source sites 1-7 (a-g). Red bars signify the connectivity potential of settlement sites 1-7 (a-g). For example, (a) shows that particles released from DEE simulated no connectivity within 20 km of a settlement site, whereas particles from elsewhere did connect with DEE. The dashed red line denotes our threshold radius for connectivity: 10 km away from each settlement site.*