Genetic evidence supports recolonisation by Mya arenaria of western Europe from North America

Cross, M. E., Bradley, C. R., Cross, T. F., Culloty, S., Lynch, S., McGinnity, P., ... Prodöhl, P. A. (2016). Genetic evidence supports recolonisation by Mya arenaria of western Europe from North America. Marine Ecology: Progress Series, 549, 99-112. DOI: 10.3354/meps11672

Published in:
Marine Ecology: Progress Series

Document Version:
Peer reviewed version

Accepted Manuscript

Genetic evidence supports recolonisation by *Mya arenaria* (L.) of western Europe from North America

M. E. Cross, C. R. Bradley, T. F. Cross, S. Culloty, S. Lynch, P. McGinnity, R. M. O'Riordan, S. Vartia and P. A. Prodöhl

Published in MEPS Vol. 549:99-112 (2016).

DOI:10.3354/meps11672

Accepted date: 25 February 2016

Online publication date: May 10, 2016

Print ISSN: 0171-8630; Online ISSN: 1616-1599

Genetic evidence supports recolonisation by *Mya arenaria* (L.) of western Europe from North America.

M. E. Cross[1*], C. R. Bradley[2], T. F. Cross[1], S. Culloty[1], S. Lynch[1], P. McGinnity[1], R. M. O'Riordan[1], S. Vartia[3] and P. A. Prodöhl[2]

1. Aquaculture and Fisheries Development Centre, School of Biological, Earth and Environmental Sciences, University College Cork, Ireland.

2. Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road Belfast BT9 7BL, Northern Ireland – UK.

3. Carna Research Station, Ryan Institute, National University of Ireland, Galway, Carna, Connemara, Co. Galway, Ireland.

*To whom correspondence should be addressed: maud.cross@ucc.ie

Running title: Genetics of *Mya arenaria.*

Abstract

Mya arenaria (L.), the softshell clam, is currently widespread on the east and west coasts of North America. This bivalve also occurs on western European shores where the post Pleistocene origin of the species, whether introduced or relict, has been debated. In this study 320 *M. arenaria* from eight locations in Europe and North America were collected. Eighty-four clams from seven of the locations were examined for mitochondrial DNA variation by sequencing a section of the cytochrome oxidase I (*COX*I) gene. These were analysed together with 212 sequences, sourced from Genbank, from the same gene from 12 additional locations, chiefly from eastern North America but one site each from western North America and one from western Europe. Ten microsatellite loci were also investigated in all 320 clams. Nuclear markers showed reduced levels of variation in certain European samples. The same common *COX*I haplotypes and microsatellite alleles were present throughout the range of *M. arenaria*, though significant differences were identified in haplotypic and allelic composition between many samples, particularly those from the two continents, Europe and North America. These findings support the hypothesis of post Pleistocene colonisation of European shores from eastern North America (and the recorded human transfer of clams from the east to the west coast of North America in the 19th Century).

Keywords: *COX*I, Europe, microsatellite loci, mitochondrial DNA, *Mya arenaria*, North America.

INTRODUCTION

Following the retreat of the glaciers at the end of the Pleistocene some 12,000 years ago, recolonisation of northern Europe, in particular for most terrestrial species, is presumed to have originated from one, two or three refugia in southern Europe i.e. Iberia, Italy and the Balkans (Hewitt 1999). In the case of shallow water and inter-tidal marine species, the situation was different in that post-glacial colonists from the south (contemporary Iberia, western North Africa and the Mediterranean) had to follow coastlines, which experienced rapid changes in sea level (Maggs et al. 2008). The fact that many marine species have pelagic larvae is thought to increase the potential for rapid colonisation (Luttikhuizen et al. 2003, Dupont et al. 2007, Arias-Perez et al. 2012), despite other factors (e.g. ocean currents, salinity and temperature levels) which can prevent gene flow. In more recent times, planktonic larvae of marine species could also be distributed anthropogenically in ballast or bilge water of ships (Briski et al. 2012, Ruiz et al. 1997) or as adults, particularly gastropod or bivalve mollusc species, which, when retained as fresh food items, can survive long sea journeys, if held in low temperature sea water.

Notwithstanding the European refugial hypothesis (see Maggs et al. 2008), certain marine mollusc species currently occupying North Atlantic coastal regions appear to have had a different route of colonisation which involved trans-oceanic movement. The edible winkle *Littorina littorea* (Linnaeus 1758) (Prosobranchia, Gastropoda) is thought to be one such species (Chapman 2007, Chapman et al. 2007, Chapman et al. 2008). Arguably not present on eastern North American shores after the Pleistocene glaciations (Chapman 2007, Chapman et al. 2008, Blakeslee et al. 2008, Blakeslee & Byres, 2008), the suggestion is that it was introduced as a food source to Newfoundland, or through rock ballast to Nova Scotia (Chapman et al. 2007). It was first recorded in the latter area in 1840 (Brawley et al. 2009) and thereafter spread south (Beebee & Rowe 2008, Harley et al. 2013). The opposite (colonisation of European shores from eastern North America (Strasser 1999)) is believed to be the case for the infaunal bivalve species, the softshell clam *Mya arenaria* (Linnaeus 1758), considered in the present study. Now recognised as an important member of the intertidal infaunal community in the North East Atlantic (Strasser 1999), the species occurs over a wide geographical range from Iceland, the White Sea and northern Norway to Portugal (Strasser 1999, Conde et al. 2010), including the Barents and Black Sea, with recent reports of its presence also in the Mediterranean Sea (Zaitsev & Mamaev 1997, Weston & Buttner 2010,

Crocetta & Turolla 2011). Archaeological evidence suggests that *M. arenaria* was present in this region in the late Pliocene (Strauch 1972), and that it disappeared from European shores during the Pleistocene glaciations. *M. arenaria* is then thought to have reappeared, post-glacially, on European shores. The reintroduction may have been either natural or anthropogenic, with previous suggestions that the Vikings were responsible (MacNeil 1965, Strauch 1972, Petersen et al. 1992).

Molecular genetics, as used in the present study, provides a mechanism for testing the hypothesis of trans-Atlantic colonisation. If trans-Atlantic colonisation was exclusively or predominately the case, similar genetic composition would be expected in *M. arenaria* from the coasts of the eastern and western Atlantic. This constitutes the hypothesis explored here i.e. that colonisation of western European shores was from eastern North America. In contrast, different genetic composition would be expected in softshell clams from either side of the ocean if post-Pleistocene northern European colonisation had involved southern or peri-glacial European refugia. In this case, genetic differences would have evolved as a consequence of forces such as mutation, genetic drift and selection, while the two assemblages were separated throughout the Pleistocene.

In addition, a single or limited number of colonisation events, limited to a small number of founders, would result in substantially lower genetic variability in northern European populations. Therefore, *M. arenaria* occurring south of the glaciated areas in eastern North America, would be more genetically variable than those in Europe. In this way, another feature of the colonisation process can be elucidated.

Previous genetic studies of *M. arenaria* concentrated on eastern North American coasts and used markers such as allozymes (Levinton 1973, Morgan et al. 1978). Where European samples were included, little inter-continental variation was observed throughout the contemporary range (Lasota et al. 2004). As *M. arenaria* is known to be a successful coloniser within European waters (Bologa et al. 1995, Zaitsev & Mamaev, 1997), and similar genetic composition was observed on each side of the Atlantic, a tentative agreement with the hypothesis of trans-Atlantic colonisation was formed (Lasota et al. 2004). Similarly, investigations of mtDNA using both the RFLP method and sequencing methodologies (Caporale et al. 1997, Sparagano et al. 2002) supported the trans-Atlantic colonisation concept, but discerned little or no population structure. These findings were interpreted as

evidence of high levels of gene flow between areas. In their study of the *COX*1 mtDNA gene of *M. arenaria*, based mainly in North America, Strasser & Barber (2009) observed a unique common haplotype in a single small sample from the Wadden Sea (Sylt, Germany – NSE). While this observation could have been linked to a possible existence of European refugia, it was dismissed on the basis of the relatively small number of individuals within samples, resulting in low discovery of rarer haplotypes (Strasser & Barber 2009).

In the current study sequences from a large section of the *COX*1 mitochondrial gene were considered, together with sequences previously published by Strasser & Barber (2009). Genetic data from ten nuclear microsatellite DNA loci were also investigated. These molecular data were analysed to investigate several eastern North American samples, a single western North American and five European samples (the latter being the most intensive coverage to date of European samples). Both mitochondrial and nuclear data were included to test for correspondence in evolutionary inference between these two genomic systems. Recently published microsatellite loci were included (St-Onge et al. 2011, Krapal et al. 2012), as their intensive polymorphism and high evolutionary rate make this marker type ideal for determining routes of colonisation, quantifying relative numbers of founders and determining contemporary population structure in donor and colonised areas.

In summary, the aims of the current study were to investigate genetic composition at the *COX*1 gene and ten microsatellite loci in samples of *M. arenaria* from eastern and western North America and from north Western Europe, and to test the hypothesis about the origin of European colonization.

Materials and Methods

Sampling

Specimens of *Mya arenaria* (n=320) were sampled from eight intertidal locations throughout the contemporary species range, in Ireland, North Wales, the Netherlands, eastern Canada and three locations in the United States (Figure 1). Specimens were collected by digging in the lower intertidal, or with a benthic hydraulic escalator dredge, at each site. Gill tissue samples from each individual were preserved in 90% molecular grade ethanol. Details of sampling locations are shown in Figure 1 and Table 1.

Mitochondrial DNA

Ten to 16 *M. arenaria* specimens representing the sampling locations covered in the current study (Table 1) were randomly chosen for mtDNA sequencing analysis. Genomic DNA was extracted from gill tissue using QIAamp DNA mini Kit (QIAGEN). DNA quality and concentration was assessed by gel electrophoresis (0.8% 0.5X TBE agarose gel). For each of these samples, a 533 base pair (bp) region of the cytochrome oxidase subunit I mitochondrial DNA gene (mtDNA *COX*I) was amplified using the universal PCR primers HCO-2198 and LCO-1490 (Folmer et al. 1994). PCR amplifications were carried out in 50 µl volumes, containing 5 µl 10x Invitrogen Buffer, 3.5 µl $MgCl_2$ (50mM), 2 µl dNTPs (5mM), 1 µl of each primer (10pM/ µl), 0.5 µl BSA (10 mg/ ml), 0.8 µl (4 U) Ampli-*Taq* polymerase (Invitrogen), 34.2 µl dH_2O and 2 µl (~50ng/ µl) template DNA. PCR amplifications were carried out on a Techne TC-Plus thermocycler as follows: 95 °C for 15 min, 5x (95 °C for 60 sec, 50 °C for 60 sec, 72 °C for 60 sec), 30x (95 °C for 60 sec, 52 °C for 60 sec, 72 °C for 60 sec), and 72 °C for 5 minutes.

Amplified products were checked for quality by gel electrophoresis (1.5% 0.5x TBE agarose gel) and purified using EXOSAP-IT (Affymetrix, USB Products) purification kits. Purified products were bi-directionally sequenced using HCO-2198 and LCO-1490 primers and the Big Dye Terminator (V3.1) sequencing chemistry (Applied Biosystems) following manufacturer's recommendation. Cycle sequencing reactions were carried out on a Techne TC-Plus thermocycler as follows: 94 °C for 3 min, 25x (94 °C for 10 sec, 50 °C for 5 sec, 60 °C for 4 min), and 8 °C for 10 minutes. Resulting sequencing products were purified by ethanol precipitation (125mM EDTA and 100% molecular grade ethanol). Sequencing was carried out using an ABI 3730XL 96 capillary system DNA analyser (Applied Biosystems).

The sequencing analysis of 212 *M. arenaria* specimens from 12 additional sites (Figure 1 and Table 2) were included in the current analysis. This allowed the integration of novel mtDNA sequencing data generated within this investigation to existing lineages identified in a previous study by Strasser & Barber (2009).

Microsatellite DNA

For microsatellite analysis, genomic DNA was extracted from all *M. arenaria* specimens (n=320) using CHELEX-100 resin (Bio-Rad) extraction method. All samples were

screened for ten microsatellite marker loci described by St-Onge et al. (2011) (*Mar*01, *Mar*04, *Mar*06, *Mar*07, *Mar*08) and Krapal *et al.* (2012) (Ma02, Ma06, Ma11, Ma14, Ma15), which were designed from *M. arenaria* microsatellite clone sequences from Genbank (accession numbers: JN850609.1, JN850610.1, JN850611.1, JN850612.1, JN850613.1, JN850614.1, JN850615.1, JN850616.1, JN850617.1) (Krapal et al. 2012). Primers were designed using Primer3Plus (Rozen & Skaletsky 2000, Untergasser et al. 2007) with optimal primer length as 20 bp and optimal T_m at 60°C. The forward primers were tailed with one of four universal dye-labelled tails in their 5' end: T3: PET: 5'-AATTAACCCTCACTAAAGGG-3', M13 Reverse: NED: 5'-GGATAACAATTTCACACAGG-3' (Diniz et al. 2007), Hill: 6FAM: 5'-TGACCGGCAGCAAAATTG-3' (Tozaki et al. 2001) and Neomycin rev: VIC: 5'-AGGTGAGATGACAGGAGATC-3'. PIG-tails were added to the 5' end of all the reverse primers (Brownstein et al. 1996). Primers were combined into two multiplex PCR reactions with Multiplex Manager 1.0 (Holleley & Geerts 2009).

Microsatellite screening was performed using multiplex PCR amplification carried out in reaction volumes of 5 µl, including 1 µl of DNA (approximately 50 ng), 2.5 µl of Qiagen 2x Multiplex Mix ™, 0.5 µl 10X primer mix (outlined below) and 1 µl dH_2O. The forward primer in each case incorporated a tail, to which a labelled tail would anneal during PCR. Final reaction primer concentrations were 0.2 µM of forward (tailed) primer, 0.2 µM of reverse primer (which was PIG-tailed) and 0.05 µM of labelled tail primer. Four different labelled tails were used (one for each channel, FAM, NED, VIC and PET). Tailing methods for each locus are outlined in SI Table 1. Amplifications were performed on a Techne TC-Plus thermocycler as follows: 15 min at 95 °C, 30 cycles of 30 sec at 94 °C, 90 sec at 56 °C, and 60 sec at 72 °C, 8 cycles of 30 sec at 94 °C, 90 sec at 53 °C, and 60 sec at 72 °C, with a final step of 30 min at 60 °C. To minimise genotyping error, PCR amplification of each individual was repeated at least three times and three known genotype individuals were run on each plate for control purposes. All PCR products were run on an ABI 3730XL Genetic Analyser 16 capillary system (Applied Biosystems) and sized with internal lane standard (GENESCAN 500 LIZ, Applied Biosystems) using the program GENEMARKER v 1.6 (softgenetics).

Data analysis

Mitochondrial sequences

Resulting sequences were checked for base call quality and ambiguity using CHROMASPRO 150 (version 1.7.6) (Technelysium Pty Ltd.) and subsequently aligned using 'CLUSTAL W' implemented in BIOEDIT (version 7.1.9) (Hall 1999), alongside sequences representing all haplotypes described by Strasser & Barber (2009). Summary sample diversity statistics including haplotype (h) and nucleotide diversity (π), and average number of nucleotide differences (k) between haplotypes were estimated using DNASP v5.1 (Rozas et al. 2003). To assess the level of genetic structuring both on macro- (i.e. trans-Atlantic) and micro- (North America and Europe) geographical levels, sample pairwise Φ_{st} estimators and associated *P* values were calculated in ARLEQUIN 3.01. Patterns of genetic subdivision both at macro- and micro- geographical scales were evaluated using analysis of molecular variance (AMOVA), also using ARLEQUIN 3.01. Statistical significance of the AMOVA was tested by permutation (10,000). To examine the relationship among resulting mtDNA haplotypes and to provide additional insights into the phylogeography of *M. arenaria* on both sides of the Atlantic, a haplotypic network was constructed using the median-joining method (Bandelt et al. 1999), implemented in Network 4.5.0.2 (fluxus-engineering.com, Fluxus Technology Ltd. 2004). For this analysis, sequencing data from Strasser & Barber (2009) was also included for comparison.

Microsatellites

Within-sample statistics, including total number of alleles and mean number of alleles, expected and observed heterozygosity and allele richness (Kalinowski, 2004) were estimated using diveRsity (Keenan et al. 2013). The statistical significance of observed differences in the measures of allelic richness between samples (pairwise tests) was assessed using a sign test (Wilcoxon signed-rank tests) across loci as suggested by Kalinowski (2005).

Deviations from Hardy-Weinberg equilibrium (HWE) were assessed using exact probability tests implemented in GENEPOP 3.4 (Raymond & Rousset, 1995). Significance levels for multiple comparisons were adjusted using standard Bonferroni correction (Rice 1989). Because null alleles can result in an underestimation of within-population genetic variation,

loci that deviated from HWE, with heterozygote deficiencies, were evaluated for the existence of null alleles following approach implemented in MICRO-CHECKER 2.2.3 (van Oosterhout et al. 2004).

Genetic divergence among samples was compared using F_{ST} (Weir & Cockerham 1984), estimated using diveRsity (Keenan et al. 2013), and statistical significance was assessed by bootstrapping loci (5,000). Patterns of nuclear genetic subdivision observed both at macro- and micro- geographical scales were evaluated using analysis of molecular variance (AMOVA) also using ARLEQUIN 3.01 with 10,000 permutations.

To further investigate the presence and patterns of population sub-structuring within the sample, an unrooted neighbour joining (NJ) tree was created based on Nei et al. (1983) genetic distances. Genetic distances and bootstrap values (10,000) were calculated using Populations 1.2.3.1. A tree was then drawn using Figtree (http://tree.bio.ed.ac.uk/software/figtree/).

As an additional approach to examine the patterns of population genetic within the data, the Bayesian analytical framework implemented in the programme STRUCTURE (Pritchard et al. 2000) was also used. This framework allows for the identification of genetic partitioning (i.e. genetic sub-structuring) within a sample for which no *a priori* hypothesis of population structuring exists. Thus, given a sample set consisting of genotypic data from a number of specimens, the analytical framework estimates the minimum number of Mendelian populations (i.e. independent genetic entities) that best explains the data. Here, the STRUCTURE analysis was carried out using a hierarchical approach, which was primarily intended to identify major genetic evolutionary lineages within the data and subsequently refining these to the population level. This analysis was carried out as follows: 1) using the the complete data set, STRUCTURE runs were set for *k* varying from 1 to 10 (20 interactions for each *k* value), with 100,000 burn-in iteractions, followed by 100,000 MCMC iteractions for each independent run. STRUCTURE analysis was carried out using the admixture model with correlated allele frequencies. Group matching among independent run interactions was carried out using CLUMP v.1.1.2b (Jakobsoh & Rosenberg 2007) employing the "greedy" algorithm with random input orders and 20,000 repeats. In each case, the optimal *k* was determined using the *ad-hoc* method of Evanno et al. (2005). 2).Following this initial run, in order to identify further genetic sub-structuring within the data (i.e. sub-groups within the main groups), STRUCTURE was independently run with genotypic data from each of the main groups separately. In each instance, the same approach/parameters adopted for the

whole data set were used. If additional substructuring was identified, STRUCTURE was again independently run against samples representing each sub-group using same approach/parameters. This iterative/hierarchical approach was repeated until no further sub-structuring was evident from the data. Results of this hierarchical analysis were summarsed in graphical format using STRUCTURE plots. Finally, as an alternative non-parametric approach to assess the existence of population sub-structuring within the data, a Factorial Component Analysis (FCA) was carried out using GENETIX 4.05 (Belkhir et al. 1996-2004). The results of this analysis were summarised using a three dimensional plot generated in MATLAB 2.04 (The Mathworks Inc.) using the scatter three-function.

Results

*Mt DNA COX*I *Region*

The alignment of all *COX*I gene sequences (i.e. 84 generated in the present study in addition to the 212 sequences from Strasser & Barber (2009) revealed 34 haplotypes of which 27 were previously described by Strasser & Barber (2009) and seven were novel (Table 3) (KU720416, KU720417, KU720418, KU720419, KU720420, KU720421, KU720422, KU720423, KU720424, KU720425, KU720426, KU720427, KU720428, KU720429, KU720430). Haplotype A was the only one to be shared among all samples on both sides of the Atlantic and in the Pacific. It was predominant in all except for the samples from Ireland (BB & FF). As has been reported by Strasser & Barber (2009), with few exceptions, most haplotypes are restricted to particular sampling locations, occurring once only. Among the notable exceptions was haplotype B, which was relatively common throughout the sampling range but appears to be particularly common in Europe, and haplotype E which, with one exception (Canada - C), was restricted to European samples. Of the seven novel haplotypes described in the present study, five were restricted to North America and two (AB and AC) were restricted to Europe.

The median-join network depicting the genetic relationship among haplotypes, revealed a star-shaped phylogeny (Figure 2). In most cases, haplotypes differed from each other by a single mutational step stemming from haplotype A.

Within sample variability

293 Haplotype diversity (*h*) ranged from a zero value in Wales (W) and Nova Scotia (NS),
294 which were fixed for haplotype A to 0.800 and 0.894 in Maryland (MLMR) and Bannow Bay
295 (BB) respectively with an average of 0.422 (Table 2). Nucleotide diversity (π) was relatively
296 low for all localities, ranging from 0.00023 in the Canadian sample (C) to 0.00245 in the
297 Miles River (MLMR) group. The values for the average number of nucleotide differences (*k*)
298 ranged from zero in Wales (W) and 0.125 in Canada (C) to 1.303 in the Miles River (MLMR)
299 sample (Table 2).

300 *Among sample variability*

301 After standard Bonferroni correction, pairwise inter-population Φ_{st} comparisons
302 revealed certain samples of *M. arenaria* to be statistically significantly different from each
303 other (Table 4), including the Irish samples (BB & FF) with Welsh (W) and Maryland
304 (MLEB & MLMR) *M. arenaria*, and the Canadian sample (C) with the Flaxfort (FF) and
305 Miles River (MLMR) Maryland individuals. The Netherlands (N) sample was not shown to
306 be statistically significantly different from any other areas using this method. In the
307 hierarchical AMOVA, 90% of mtDNA *COX*1 variation was found within samples, whereas
308 7.9% and 2.1% of total haplotype frequency variation (both values statistically significant
309 from zero) was represented by variation between European and North American samples and
310 among samples within regions respectively (SI Table 2).

311 *Microsatellites*

312 *Intra sample parameters*

313 Three hundred and twenty *M. arenaria* from eight geographical sites across the
314 species range were genotyped at ten microsatellite loci (SI Table 6). No linkage
315 disequilibrium was observed between locus pairs and all ten loci were polymorphic. The
316 three loci that deviated from HWE with heterozygote deficiencies (Ma 02, Ma 06 and Ma 15)
317 were positively identified as possessing null alleles in MICRO-CHECKER 2.2.3 (van
318 Oosterhout et al. 2004). Since the apparent presence of null alleles at these three loci had a
319 negligible effect on the estimation of overall F_{ST} (uncorrected value 0.058 and corrected
320 value 0.057), all ten loci were used in calculation of pairwise F_{ST} values (Table 5). The
321 average number of alleles per sample varied from 7.2 in the Welsh sample (W), to 24.8 in the
322 Miles River, Maryland sample (MLMR) (Table 1, SI Table 3), with the total average number
323 of alleles per locus of 15, and a total of 355 alleles being detected overall. Across all loci

sampled, alleles unique to an area were present in small numbers in all sites except the Welsh sample (W) (SI Table 4).

Observed and expected within-sample heterozygosity values varied from 0.678 (Flaxfort - FF) to 0.914 (Oregon - NEP), and from 0.706 (Wales - W) to 0.901 (Maryland samples – MLEB & MLMR) (Table 1, SI Table 3). The highest percentage of unique alleles present in an area was found in the Oregon sample (NEP) of *M. arenaria*, and the lowest in the Welsh sample (W) (SI Table 4). Genetic variability as measured by allelic richness (R_s) ranged from 11.1 in the two Maryland sites (MLEB & MLMR), to 5.6 in the Welsh sample (W), with a mean allelic richness of 8.41. Results from the Wilcoxon signed-rank tests for differences in allele diversity (i.e. allele richness) among samples indicates significant statistical outcomes between all pair-wise comparison involving samples from North America and Europe (P<0.01 in all instances). Thus, samples from North America consistently displayed significantly higher levels of allele diversity in comparison to European samples. There were no significant differences in allele richness among pairwise comparison involving North American samples only. Among the European samples, however, the Netherlands sample was found to display significantly higher levels of allele diversity (P<0.01) in comparison to all other European samples, which in turn, displayed similar levels of allele diversity (i.e. not statistically significant).

Inter sample comparisons

Genetic differentiation over the eight softshell clam samples was significant at an overall value of 0.0224 (95% CI 0.0140 - 0.0341). Pairwise F_{ST} values ranged from the single not statistically significant 0.0017 between the two Maryland sites (MLEB & MLMR), to 0.113 and 0.101 between Welsh (W) and Canadian (C), and Welsh (W) and Flaxfort (FF) *M. arenaria* respectively (Table 5). All samples of *M. arenaria* were significantly different from each other, excepting those from the two Maryland sites (MLEB & MLMR).

The population structuring suggested by the pairwise F_{ST} analysis was further supported in the un-rooted neighbour-joining tree based on inter sample D_a values, and showed the European and North American samples to be clearly separated, with the Netherlands (N) sample being in an intermediate position (Figure 3). Samples of *M. arenaria* from the two Irish sites, Bannow Bay (BB) and Flaxfort (FF), clustered together on the neighbour-joining tree, as did the two samples from Maryland, North America (MLEB &

MLMR). The hierarchical STRUCTURE plots (Figure 4) provide further support to these inter sample relationships. Samples from the two continents separated at the first level (i.e. Level 0 in Figure 4). Significant within continent variation was also identified at the second hierarchical level (Level 1, groups 1 and 2) with no further obvious differentiation between the two Maryland or Irish samples (i.e. third hierarchical level - Level 2). The significant level of population genetic structuring observed with pairwise F_{ST} estimates and confirmed from both neighbour-joining and STRUCTURE analyses was further supported by the independent results of Factorial Component Analysis (SI Figure 1), which shows clear separation among samples from Europe and North America in addition "within major regions" sub structuring (SI Figure 1).

Hierarchical AMOVA analyses corroborate previous analysis by indicating significant differences in genetic partitioning of *M. arenaria* at all levels of geographic organisation (with regional partitioning as used above for *COX*I data), between the two regions, Europe and North America, among populations within regions, and within populations (SI Table 5). As with mtDNA *COX*1 AMOVA results, most variation was due to differentiation within populations (88.29%), while 5.48% and 6.23% of total variation was represented by variation among Europe and North American samples, and among populations within regions respectively.

Discussion

As noted in the Introduction, previous studies of *M. arenaria* failed to demonstrate significant inter- and intra-continental differences between samples. Conversely, in the present study virtually all samples were significantly different for pairwise multi-locus microsatellite F_{ST} values, except those in close proximity in Chesapeake Bay (Table 5). The greatest difference was between samples from either side of the Atlantic (see AMOVA results in SI Table 4). This suggests that all but two of these samples should be regarded as largely distinct populations, with little contemporary gene flow between them.

In the present investigation, the utilisation of a combination of two independent molecular methods in this species was novel. Another innovation with these markers, which have previously proven to be discriminatory in eastern North America (Strasser & Barber

2009, St-Onge et al. 2011), was the inclusion of more than one European sample. This allows consideration of aspects such as possible colonisation modes and routes, and population variability within Europe. Both molecular methods indicate major or exclusive colonisation from eastern North America, in that European samples shared the same common mitochondrial haplotypes or common microsatellite alleles with eastern North American samples. In addition, North American samples displayed statistically significant higher levels of genetic diversity in comparison to European samples in a pattern that is consistent with source populations. Thus, we suggest that, based on current data, eastern North American populations were the major source of post Pleistocene colonisers to northern Europe.

Routes of European colonisation

Previous authors (Petersen et al. 1992, Strasser 1999) have suggested that *M. arenaria* was introduced into Europe by Vikings returning to Europe in pre-Columbian times. Alternatively, natural colonisation could have occurred by larval drift from west to east via northern coastal locations. Future sampling of *M. arenaria* in the more northerly parts of the contemporary range in eastern North America and western Europe, and in transitional regions such as Iceland and Greenland, might clarify the possible routes of colonisation, as has been the case in the molecular investigation of *Arctica islandica* (Dahlgren et al. 2000), *Cerastoderma edule* (Krakau et al. 2012), *Macoma balthica* (Luttikhuizen et al. 2003), and *Littorina saxatilis* (Panova et al. 2011).

Regardless of whether European colonisation was human mediated or natural, data obtained in the current study would suggest that colonisation occurred from the north of the eastern North American range. Analysis of the relationship at microsatellite loci between the *M. arenaria* samples suggests a closer relationship between the Prince Edward Island sample (north of the eastern North American range) and the Netherlands sample (North Sea, Europe), than for samples from further south in the eastern North American range (Maryland, USA) (Figure 3 and SI Fig 1).

Assuming that variability decreases from the point of first colonisation (Hewitt 2000), levels of microsatellite variability (i.e. allele diversity) and genetic similarity observed in the four European samples may give some indication of the sequence of north western European colonisation. The average level of genetic diversity was significantly higher in the North Sea (Netherlands) sample, than in other European samples (from the Celtic Sea, southern Ireland

and Wales (Table 1)). Following the theory put forward by Hewitt (2000), this suggests that the North Sea could have been the area of first European colonisation from which other areas would be subsequently colonised. While no statistically significant differences were observed among the other European samples, it is interesting to note the sample from Wales displays the lowest level of genetic diversity numerically, thus suggesting that this area may have been the last to be colonised. The resulting neighbour joining (NJ) tree (Figure 3) and the results from the STRUCTURE analysis (Figure 4) support this interpretation, as do pairwise F_{ST} values for microsatellites (Table 5). Indeed, the F_{ST} estimates might seems to suggest a 'stepping stone' sequence of progressive colonisation, similar to that proposed by Allendorf & Luikart (2008).

The situation is far less clear cut for the mitochondrial data, possibly because of the lower variability observed in *COX*1 sequences and the use of substantially lower sample sizes. Mitochondrial DNA pairwise Φ_{st} estimates for Canada compared with three of European samples (the Netherlands, Bannow Bay and Flaxfort) provided the same trend of increasing Φ_{st} estimates as observed for the microsatellites (Table 5). In contrast, the pairwise mitochondrial DNA Φ_{st} between Canada and Wales was zero, while an F_{ST} value of 0.113 for microsatellite data from the same samples was higher than those between the Canadian and other European samples (Table 5). The examination of genetic variability (*h* values) for *COX*1 also showed a complex pattern (Table 2), where *h* values of the Irish samples were particularly high. These results, however, have to be interpreted with caution because of the small sample involved. One possibly hypothesis is that at *COX*1, the Irish samples displays evidence of a separate colonisation event, either from eastern North America or from a possible southern European refugium (Maggs et al, 2008)

Possible European refugia?

The concept that European refugia, if they existed, should have occurred at the south of the range assumes that coastal marine species spread northwards post-glacially (Hewitt, 2000). Many marine species, including the northern quahog *Mercenaria mercenaria* (Baker et al. 2008), American lobster *Homarus americanus* (Kenchington et al. 2009), the marine gastropod *Acanthinucella spirata* (Hellberg et al. 2001), and the dog whelk *Nucella ostrina* (Marko 2004), appear to conform to the model proposed by Hewitt (2000). However, it has been proposed that several marine species instead had northern peri glacial refugia, on the

basis of high genetic variability in contemporary northern populations. These include the ocean quahog *Arctica islandica* (Dahlgren et al. 2000), *Macoma balthica* (Luttikhuizen et al. 2003), *Cerastoderma edule* (Krakau et al. 2012), *Pagurus longicarpus* (Young et al. 2002) and *Littorina saxatilis* (Panova et al. 2011). In addition, peri-glacial refugia have been suggested for many other marine species including *Carcinus maenas* (Roman & Palumbi 2004), seaweed species (Hoarau et al. 2007, Olsen et al. 2010, Coyer et al. 2011), the common shrimp, *Crangon crangon* (L) (Luttikhuizen et al. 2008), the genus *Patella* (Sa-Pinto et al. 2005) and tubiferous polychaetes (Jolly et al. 2006).

It is interesting to note that the molecular data produced by St-Onge et al. (2013) suggest a southern refugium for *M. arenaria* in eastern North America. In the present study, it was not feasible to determine whether there has also been a limited contribution from one or more refugia on European shores, if the species did persist in Europe during the Pleistocene. While some haplotypes were almost or totally exclusive to Europe (see Irish samples in Table 3), the majority were only one mutational step removed from haplotypes common in eastern North America (see Figure 2). For microsatellites, only a minority of rare alleles was exclusive to European samples. More extensive European sampling, particularly further south in the range, will be necessary to address this question.

Genetic variability may indicate size of introduction

The level of genetic variability at the microsatellite loci, observed in the Pacific North American sample is informative in the context of size of introductions. In this case there is documented evidence of recent human introduction from the southern part of the eastern North American range, when a large number of *M. arenaria* were introduced to areas through the commercial oyster industry, and dedicated plantings were made to initiate a commercial fishery (Hanna, 1966, Carlton, 1989, Strasser 1999). The sample analysed in the present study from Oregon on the Pacific coast had a level of microsatellite variability similar to that from Maryland on the Atlantic coast, giving genetic confirmation of a large introduction from an eastern North American area. In contrast, the European clam samples showed substantially lower microsatellite variability, suggesting a lower number of individuals introduced to this area (see Table 1).

The large inter- and intra-continental difference demonstrated in the current study may be indicative of intermittent movements of small numbers of individuals from North

America to Europe. The results from the analysis of *COX*1 sequences are less clear than those of microsatellites, which provided better support for discrimination among samples within continents in comparison to the single mitochondrial region. Interestingly, however, both mitochondrial DNA and microsatellites provide almost equal support to the inter region (continent) differences.

What is clearly demonstrated by the present results, is that the use of different marker types provides more detailed insights about mode of colonisation and contemporary population structure. We therefore suggest that a genomic approach be applied to the investigation of *M. arenaria* in the future, perhaps by identifying and screening a large number of SNPs (Beaumont & Balding 2004). This, in addition to modern statistical approaches based on Bayesian hypothesis inference (e.g. Cornuet et al 2014) may further clarify the ecology and previous movements of what is now an ecologically important and abundant soft sediment infaunal species in north western Europe. It would also provide genomic information that would be of use in *M. arenaria* aquaculture, which is widespread and commercially important in North America, and is to be developed in Europe.

Acknowledgements:

This project was part-funded by the Ireland Wales Programme 2007-2013 INTERREG 4A European Regional Development Fund.

Many thanks to Chris Dungan, John Chapman, Dante Mateo, Pauline Kamermans, Ruth Stockdale, Allen Whitaker, Eileen Dillane, and Jens and Nettan Carlsson for providing *M. arenaria* samples and practical help.

Caroline Bradley and Philip McGinnity were supported by the Beaufort Marine Research Award in Fish Population Genetics funded by the Irish Government under the Sea Change Programme.

References:

Allendorf FW, Luikart G (2008) Conservation and the Genetics of Populations. Malden, MA, Blackwell.

506 Arias-Pérez A., Fernández-Tajes J, Gaspar M.B, Méndez J (2012) Isolation of Microsatellite
507 Markers and Analysis of Genetic Diversity among East Atlantic Populations of the
508 Sword Razor Shell *Ensis siliqua*: A Tool for Population Management. Bio Genetics
509 50 (5-6): 397-415.

510 Baker P, Austin JD, Bowen BW, Baker SM (2008) Range-wide population structure and
511 history of the northern quahog (*Merceneria merceneria*) inferred from mitochondrial
512 DNA sequence data. ICES J Mar Sci 65(2): 155-163.

513 Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific
514 phylogenies. Mol Biol Evol 16 (1): 37-48.

515 Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among
516 populations from genome scans. Mol Ecol 13(4): 969-980.

517 Beebee TJC, Rowe G (2008) An introduction to molecular ecology. Oxford University Press.

518 Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05,
519 logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome,
520 Populations, Interactions, CNRS UMR 5000, Université de Montpellier II,
521 Montpellier (France).

522 Blakeslee A, Byers J, Lesser M (2008) Solving cryptogenic histories using host and parasite
523 molecular genetics: the resolution of *Littorina littorea's* North American origin. Mol
524 Ecol 17: 3684-2696.

525 Blakeslee A, Byres J (2008) Using parasites to inform ecological history: comparisons among
526 three congeneric marine snails. Ecology 89: 1068-1078.

527 Bologa AS, Bodeanu N, Petran A, Tiganus V, Zaitsev YP (1995) Major modifications of the
528 Black Sea benthic and planktonic biota in the last three decades. Bulletin de l'Institut
529 Océanographique: 85-110.

530 Brawley S, Coyer J, Blakeslee A, Hoarau G, Johnson L, Byers J, Stam W, Olsen J (2009)
531 Historical invasions of the intertidal zone of Atlantic North America associated with
532 distinctive patterns of trade and emigration. Proc Nat Acad Sci, Early edition 1-6.

533 Briski E, Ghabooli S, Bailey SA, MacIsaac HJ (2012) Invasion risk posed by
534 macroinvertebrates transported in ships' ballast tanks. Biol Invasions: 1-8.

535 Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide
536 addition by Taq DNA polymerase: primer modifications that facilitate genotyping.
537 BioTechniques 20: 1004-1010.

538 Caporale DA, Beal BF, Roxby R, Van Beneden R.J (1997) Population structure of *Mya*
539 *arenaria* along the New England coastline. Mol Mar Biol Biotech 6(1): 33-39.

540 Carlton JT (1989) Man's role in changing the face of the ocean: biological invasions and
541 implications for conservation of near-shore environments. Conserv Biol 3: 265-273.

542 Chapman J (2007) The mysterious periwinkle (*Littorina littorea*): Norse or later introduction
543 from Europe? Arc Stud Centre News 14:22.

544 Chapman JW, Blakeslee AMH, Carlton JT, Bellinger MR (2008) Parsimony dictates a human
545 introduction: on the use of genetic and other data to distinguish between the natural
546 and human-mediated invasion of the European snail *Littorina littorea* in North
547 America. Biol Invasions 10(2): 131-133.

548 Chapman JW, Carlton JT, Bellinger MR, Blakeslee AMH (2007) Premature refutation of a
549 human-mediated marine species introduction: the case history of the marine snail
550 *Littorina littorea* in the northwestern Atlantic. Biol Invasions 9: 737-750.

551 Conde A, Novais J, Domínguez J (2010) Southern limit of distribution of the soft-shell clam
552 *Mya arenaria* on the Atlantic East Coast. Biol Invasions 12(3): 429-432.

553 Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup
554 A (2014) DIYABC v2.0: a software to make Approximate Bayesian Computation
555 inferences about population history using Single Nucleotide Polymorphism, DNA
556 sequence and microsatellite data. Bioinformatics 30(8): 1187–1189.

557 Coyer JA, Hoarau G, Van Schaik J, Luijckx P, Olsen JL (2011) Trans- Pacific and trans-
558 Arctic pathways of the intertidal macroalga *Fucus distichus* L. reveal multiple glacial

559 refugia and colonizations from the North Pacific to the North Atlantic. J Biogeog
560 38(4):756–771.

561 Crocetta F, Turolla, E (2011) *Mya arenaria* Linné, 1758 (Mollusca: Bivalvia) in the
562 Mediterranean Sea: its distribution revisited. J of Biol Res-Thessaloniki: 16 188-193.

563 Dahlgren TG, Weinberg JR, Halanych KM (2000) Phylogeography of the ocean quahog
564 (*Arctica islandica*): influences of paleoclimate on genetic diversity and species range.
565 Mar Biol 137(3): 487-495.

566 Diniz F M, Iyengar A, Sarmanho da Costa Lima P, Maclean N, Bentzen P (2007) Application
567 of a double-enrichment procedure for microsatellite isolation and the use of tailed
568 primers for high throughput genotyping. Gen Mol Biol 30: 380–384.

569 Dupont L, Ellien C, Viard F (2007) Limits to gene flow in the slipper limpet *Crepidula*
570 *fornicata* as revealed by microsatellite data and a larval dispersal model. Mar Ecol
571 Prog Ser 349: 125-138.

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

574 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of
575 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.
576 Mol Mar Biol Biotech 3(5): 294.

577 Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis
578 program for Windows 95/98/NT. Nucleic Acids symposium series.

579 Hanna GD (1966) Introduced mollusks of western North America. Occ Papers Calif Acad Sci
580 48: t-108.

581 Harley CDG, Anderson KM, Lebreton CA-M, MacKay A, Ayala-Díaz M, Chong SL, Pond
582 LM, Maddison JHA, Hung BHC, Iversen SL (2013) The introduction of *Littorina*
583 *littorea* to British Columbia, Canada: potential impacts and the importance of biotic
584 resistance by native predators. Mar Biol: 1-13.

585 Hellberg D, Balch P, Roy K (2001) Climate-driven range expansion and morphological
586 evolution in a marine gastropod. Science 292: 1707–1710.

587 Hewitt GM (1999) Post-glacial re-colonisation of European biota. Biol J Linn Soc 68:87–
588 112.

589 Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405(6789): 907-913.

590 Hoarau G, Coyer JA, Veldsink JH, Stam WT, Olsen JL (2007) Glacial refugia and
591 recolonization pathways in the brown seaweed *Fucus serratus*. Mol Ecol 16:3606–
592 3616.

593 Holleley CE, Geerts PG (2009) Multiplex Manager 1.0: a crossplatform computer program
594 that plans and optimizes multiplex PCR. BioTechniques 46 (7): 511-517.

595 Jolly MT, Viard F, Gentil F, Thiébaut E, Jollivet D (2006) Comparative phylogeography of
596 two coastal polychaete tubeworms in the Northeast Atlantic supports shared history
597 and vicariant events. Mol Ecol 15(7): 1841-1855.

598 Kalinowski ST, (2004) Counting alleles with rarefaction: Private alleles and hierarchical
599 sampling designs. Conserv Genet 5:539-543.

600 Kalinowski ST, (2005) HP-RARE 1.0: a computer program for performing rarefaction on
601 measures of allelic richness. Mol Ecol Notes 5:187-189.

602 Keenan K, McGinnity P, Cross TF, Crozier WW, Prodohl PA (2013) diveRsity: An R
603 package for the estimation and exploration of population genetics parameters and
604 their associated errors. Methods Ecol Evol 4 (8): 782-788.

605 Kenchington EL, Harding EL, Jones MW, Prodohl PA (2009) Pleistocene glaciation events
606 shape genetic structure across the range of the American lobster, *Homarus*
607 *americanus*. Mol Ecol 18(8): 1654-1667.

608 Krakau M, Jacobsen S, Jensen KT, Reise K (2012) The cockle *Cerastoderma edule* at
609 Northeast Atlantic shores: genetic signatures of glacial refugia. Mar Biol 159: 221-
610 230.

611 Krapal AM, Popa OP, Iorgu EI, Costache M, Popa LO (2012) Isolation and Characterization
612 of New Microsatellite Markers for the Invasive Softshell Clam, *Mya arenaria* (L.)
613 (Bivalvia: Myidae). Internatl J Mol Sci 13(2): 2515-2520.

614 Lasota R, Hummel H, Wolowicz M (2004) Genetic diversity of European populations of the
615 invasive soft-shell clam *Mya arenaria* (Bivalvia). J Mar Biol Assoc United Kingdom
616 84(5): 1051-1056.

617 Levinton J (1973) Genetic variation in a gradient of environmental variability: marine
618 bivalvia (Mollusca). Science, 180(81):75-6.

619 Linnaeus C (1758) Systema Naturae per regna tria naturae, secundum classes, ordines,
620 genera, species, cum characteribus, differentiis, synonymis, locis. Editio decima,
621 reformata. Laurentius Salvius: Holmiae. ii, 824 pp.

622 Luttikhuizen, P. C., Campos, J., van Bleijswijk, J., Peijnenburg, K. T., & van der Veer, H. W.
623 (2008). Phylogeography of the common shrimp, *Crangon crangon* (L.) across its
624 distribution range. Molecular phylogenetics and evolution, 46(3), 1015-1030.

625 Luttikhuizen PC, Drent J, Van Delden W, Piersma T (2003) Spatially structured genetic
626 variation in a broadcast spawning bivalve: quantitative vs molecular traits. J Evol Biol
627 16(2): 260-272.

628 MacNeil FS (1965) Evolution and distribution of the genus *Mya*, and tertiary migrants of
629 mollusca. Geol Survey Prof Paper: 483-G, 1-51.

630 Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W,
631 Väinölä R (2008) Evaluating signatures of glacial refugia for North Atlantic benthic
632 marine taxa. Ecology 89(sp11): 108-122.

633 Marko PB (2004) 'What's larvae got to do with it?' Disparate patterns of post-glacial
634 population structure in two benthic marine gastropods with identical dispersal
635 potential. Mol Ecol 12:597–611.

636 Morgan RP, Block SB, Ulanowicz NI, Buys C (1978) Genetic variation in the soft-shelled
637 Clam, *Mya arenaria*. Estuaries and Coasts: 1.

638 Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular
639 data. J Mol Evol 19(2): 153-170.

640 Olsen JL, Zechman FW, Hoarau G, Coyer JA, Stam WT, Valero M, Aberg P (2010) The
641 phylogeographic architecture of the fucoid seaweed *Ascophyllum nodosum*: an
642 intertidal 'marine tree' and survivor of more than one glacial-interglacial cycle. J
643 Biogeog 37:842–856.

644 Panova M, Blakeslee AMH, Miller AW, Mäkinen T, Ruiz GM, Johannesson K, André C
645 (2011) Glacial history of the North Atlantic marine snail, *Littorina saxatilis*, inferred
646 from distribution of mitochondrial DNA lineages. PLoS ONE 6(3): e17511.

647 Petersen KS, Rasmussen KL, Heinemeier J, Rud N (1992) Clams before Columbus? Nature
648 (London) 359: 679.

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

651 Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for
652 exact tests and ecumenicism. J Hered 86: 248-249.

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

655 Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism
656 analyses by the coalescent and other methods. Bioinformatics 19(18): 2496-2497.

657 Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist
658 programmers. Methods Mol Biol 132: 365–386.

659 Rice WR (1989) Analyzing tables of statistical tests. Evolution 43 (1): 223-225.

660 Ruiz GM, Carlton JM, Grosholz ED, Hines AH (1997) Global Invasions of Marine and
661 Estuarine Habitats by Non-Indigenous Species: Mechanisms, Extent, and
662 Consequences. Amer Zoo 37:621-632.

663 Sa-Pinto A, Branco M, Harris DJ, Alexandrino P (2005) Phylogeny and phylogeography of
664 the genus *Patella* based on mitochondrial DNA sequence data. J Exp Mar Biol Ecol
665 325: 95-110.

Sparagano OAE, Purdom I, Priest FG, Kingston PF (2002) Molecular characterization of the bivalves *Mya arenaria, Mya truncata* and *Hiatella arctica*. J Mollusc Studies 68(2): 190.

St-Onge P, Parent E, Sévigny J-M, Tremblay R, Rioux-Paré R (2011) Isolation and characterization of eight novel microsatellite markers for the softshell clam (*Mya arenaria*). Mol Ecol Res.

St-Onge, P, Sevigny, JM, Strasser C, Tremblay R (2013) Strong population differentiation of softshell clams (*Mya arenaria*) sampled across seven biogeographic marine ecoregions: possible selection and isolation by distance. Mar Biol DOI 10.1007/s00227=012-2157-5: 1-17.

Strasser M (1999) *Mya arenaria*—an ancient invader of the North Sea coast. Helgo Mar Res 52(3): 309-324.

Strasser C, Barber P (2009) Limited genetic variation and structure in softshell clams (*Mya arenaria*) across their native and introduced range. Cons Gen 10(4): 803-814.

Strauch F (1972) Phylogenese, adaption und migration eineger nordischer mariner Mollusken genera (*Neptunea, Panomya, Cyrtodaria* und *Mya*) (in German). Abh Senck Natur Gesell: 531, 1-211.

Tozaki T, Mashima S, Hirota K (2001) CharMiura, N., Choi-Miura, N. H., Tomita, M. 2001. Chacterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. DNA Res 8: 33–45.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nuc Acid Res 35: W71–74.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Microchecker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4(3): 535-538.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of population structure. Evolution 38 (6): 1358-1370.

693 Weston S, Buttner JK (2010) Softshell Clam Culture: Basic Biology and General Culture
694 Considerations. NRAC Publication 201.

695 Young A, Torres C, Mack J, Cunningham C (2002) Morphological and genetic evidence for
696 vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit
697 crab *Pagurus longicarpus*. Mar Biol 140(5): 1059-1066.

698 Zaitsev Y, Mamaev V (1997) Marine biological diversity in the Black Sea: A Study of
699 Change and Decline. United Nations Publications, New York.

700

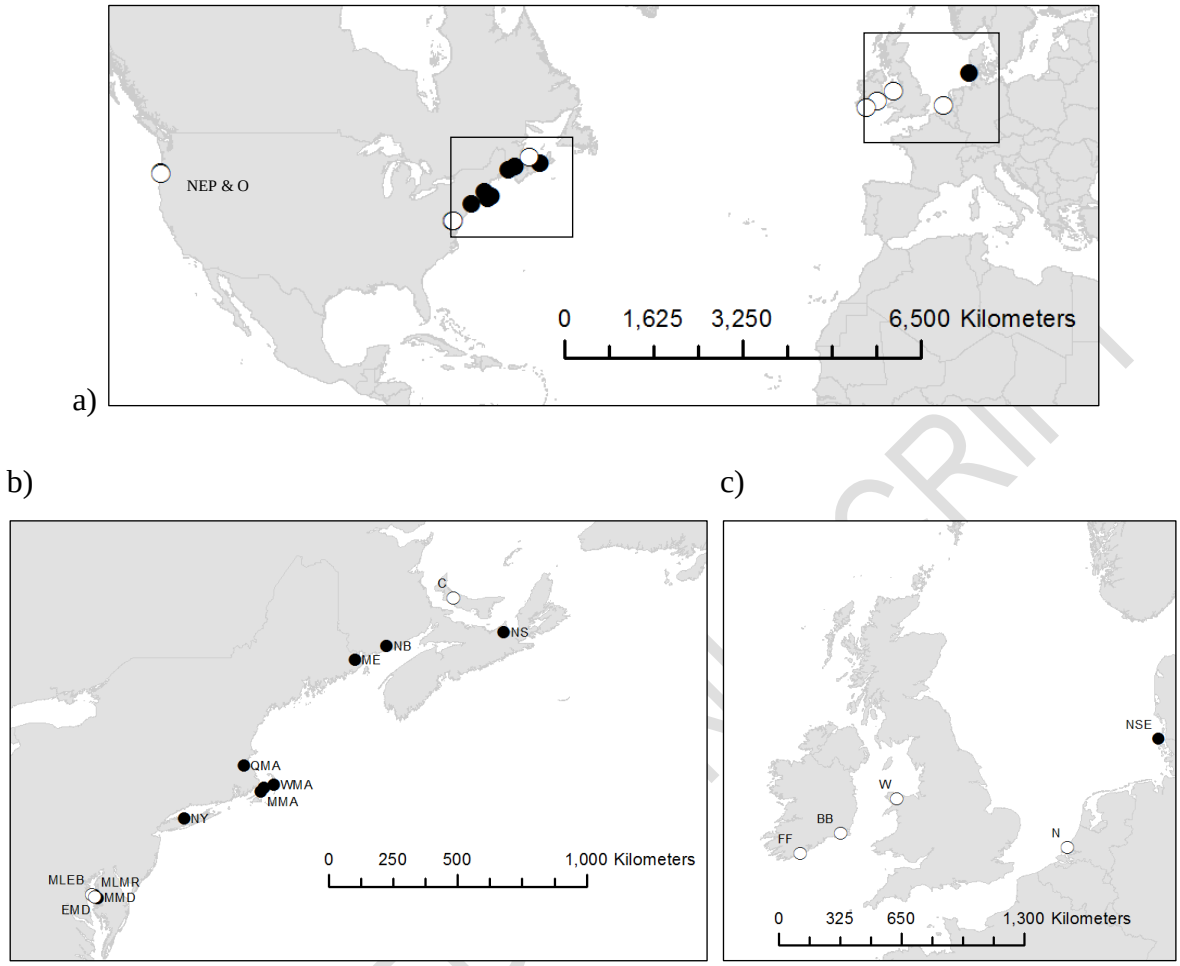
701

702

703

704

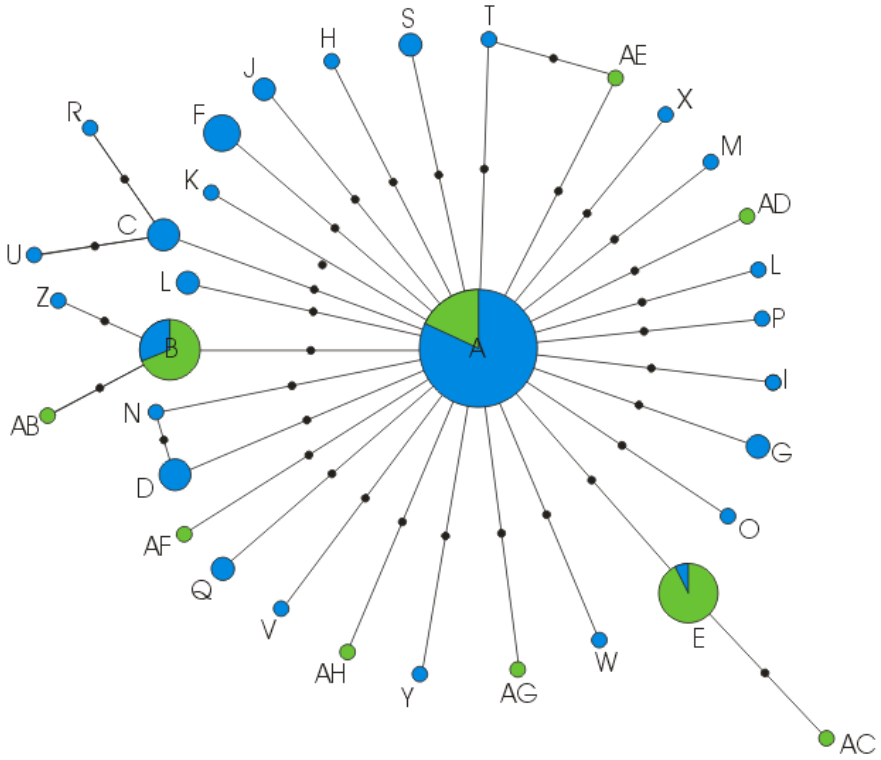
705



706

707 Figure 1. Sampling locations of *Mya arenaria* (a) throughout the range in the (b) western
708 North Atlantic region and (c) on western European coast. Black circles represent samples
709 analysed in the previous study of Strasser and Barber (2009) and white circles samples
710 collected for the present study. Sampling codes for current samples are as provided in Table
711 2. The NEP site in Oregon is obscured in Figure 1 (a) and some other overlap of
712 neighbouring sites occurs in Table 1 (a) and (b)

713



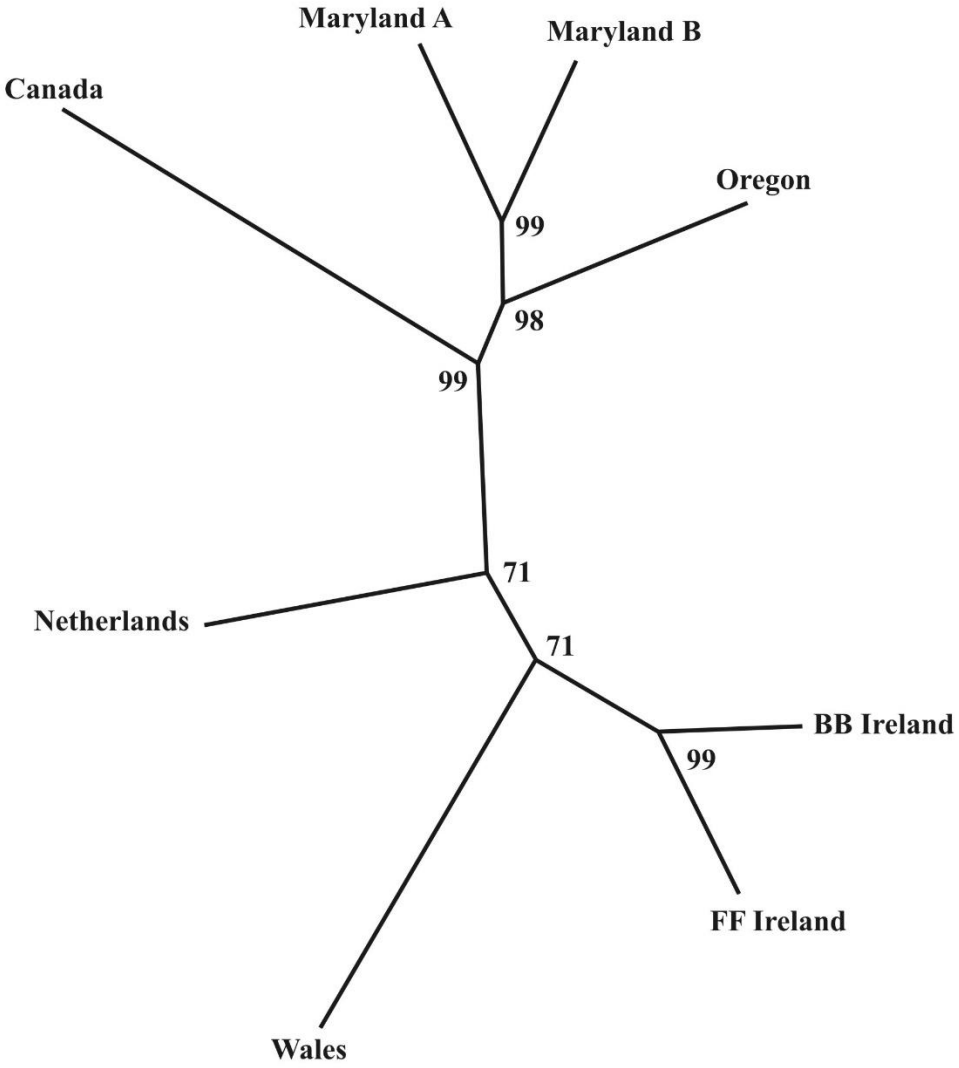
714

715 Figure 2. Median-joining network of *Mya arenaria* mtDNA haplotypes derived from *COX*I
716 sequencing data. The circle area is proportional to the number of individuals
717 characterised by a particular haplotype. The small black filled dots in the connecting
718 lines represent mutational steps. Colours represent geographic location where
719 haplotypes were found (blue – North America, green – Europe). Reticulations for
720 unresolved links (e.g. T-A-AE and N-A-D) among haplotypes are also shown.
721 Haplotype designations (A to AH) are as used in Table 3.

722

723

724



725

726

727 Figure 3. Unrooted neighbour-joining tree based on Nei *et al*. (1983) genetic distances
728 (D_A).Values on nodes represent percentage bootstrap support for groupings (out of 10,000
729 bootstraps).

730

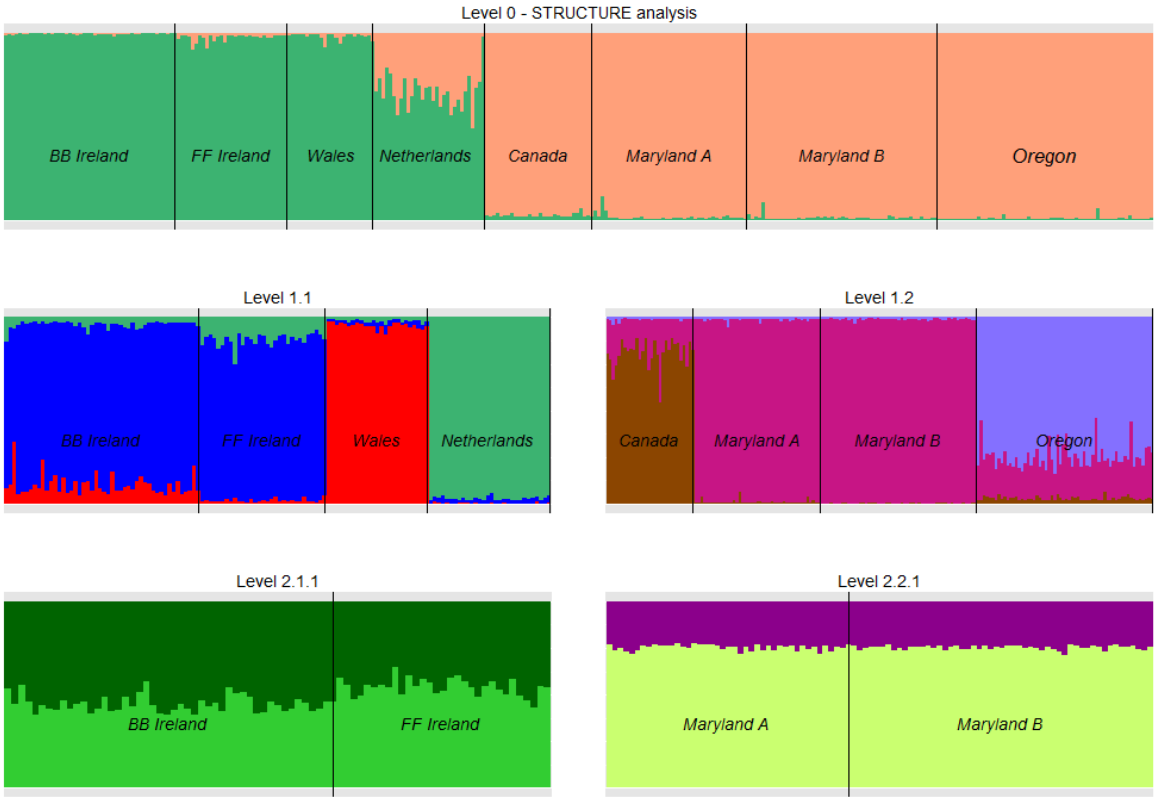


Figure 4. Estimated population structure from hierarchical STRUCTURE analyses for the full data set (Level 0) and independent STRUCTURE runs subsequently carried out on samples representing the major identifed groups (i.e. Hierarchical Levels 1 and 2 and respective subgroups). In each instance (i.e. hierarchial level and group), each particular individual is represented by a thin vertical line, which is divided into k coloured segments that represent the individual's estimated membership fractions in k clusters. Black thicker lines separate individuals from different samples.

748 Table 1. Sampling locations and numbers of *Mya arenaria* specimens analysed in this study,
749 including summary statistics for ten microsatellite loci (*a* = average number of alleles, *Rs* =
750 average allelic richness, *He* = expected heterozygosity, *Ho* = observed heterozygosity)

Sample	**(n)**	**Sample Area**	**Lat/Long**	***a***	***Rs***	***He***	***Ho***
MLEB	43	Eastern Bay, Maryland, eastern USA, North America	38.886 N 76.340 W	22.2	11.14	0.901	0.798
MLMR	53	Miles River, Maryland, eastern USA, North America	38.816 N 76.237 W	24.8	11.15	0.901	0.831
C	30	Prince Edward Island, Gulf of St. Lawrence, eastern Canada	46.446 N 63.767 W	14.6	9.01	0.844	0.747
N	30	Hinderplaat, The Netherlands, Europe	51.843 N -04.009 W	12	8.04	0.837	0.737
BB	48	Wexford, Ireland, Europe	52.251 N 06.763 W	8.9	5.99	0.752	0.678
FF	31	Cork, Ireland, Europe	51.648 N 08.698W	8.9	6.21	0.733	0.686
W	25	Menai Bridge, Wales, Europe	53.238 N 04.091 W	7.2	5.65	0.706	0.759
O	60	Alsea Bay, Oregon, western USA, North America	44.584 N 123.964 W	21.9	10.10	0.894	0.914

751
752
753
754
755
756
757
758
759
760
761
762
763
764
765

766 Table 2. Summary sample statistics derived from the mitochondrial *COX*1 sequencing data
767 including: number of individuals (N); number of haplotypes (nh); haplotype diversity ($h \pm$s.d.);
768 nucleotide diversity (π 10^{-2}); and average number of nucleotide differences (k). Bold font
769 identifies samples collected in the current investigation.

Sample Area	Sample Code	N	nh	*h*	π	*k*
Maryland A	MLEB	12	4	0.455 ± 0.17	0.094	0.500
Maryland B	MLMR	12	8	0.894 ± 0.08	0.244	1.303
Eastern Bay, MD	EMD	9	4	0.583 ± 0.18	0.167	0.889
Miles River, MD	MMD	15	4	0.371 ± 0.15	0. 096	0.514
Stony Brook, NY	NY	15	5	0.476 ± 0.16	0.100	0.533
Wareham, MA	WMA	25	7	0.430 ± 0.12	0.119	0.633
Mashpee, MA	MMA	20	6	0.447 ± 0.14	0.110	0.600
Barnstable, MA	BMA	19	3	0.001 ± 0.12	0.039	0.211
Quincy, MA	QMA	22	3	0.177 ± 0.11	0.034	0.182
Pembroke, ME	ME	21	4	0.271 ± 0.12	0.071	0.381
St. John, NB	NB	11	4	0.491 ± 0.03	0.102	0.545
Antigonish, NS	NS	20	1	-	-	-
Canada	C	16	2	0.125 ± 0.11	0.023	0.125
Sylt, Germany	NSE	15	3	0.530 ± 0.14	0.108	0.576
Netherlands	N	12	3	0.530 ± 0.14	0.108	0.576
Bannow Bay, Irl	BB	10	4	0.800 ± 0.08	0.226	1.200
Flaxfort, Irl	FF	10	3	0.711 ± 0.09	0.167	0.889
Wales	W	12	1	-	-	-
Newport, OR	NEP	20	5	0.574 ± 0.74	0.0013	0.668
Overall		296	34	0.422 ± 0.04	0.1071	0.521

770
771
772
773
774
775

776 Table 3. MtDNA haplotype distribution of the analysed *COX*I region of *M. arenaria.*
777 Samples are aligned geographically to represent the hypothesised (Strasser & Barber, 2009)
778 post Pleistocene spread of the species. Sampling sites abbreviations are given in Table 1 and
779 2. The shaded columns represent European sampling sites while the shaded rows represent
780 haplotypes unique to European sites. Underlined sampling sites represent novel samples
781 collected in the current study. 'Nhap' = No. haplotypes per sample. Other samples are from
782 Genbank (see text for details).

783

784

mtDNA Haplotype	Sample location																			
	MLEB	MLMR	EMD	MMD	NY	WMA	MMA	BMA	QMA	ME	NB	NS	C	NSE	N	BB	FF	W	NEP	All
A	9	4	6	12	11	19	15	17	20	18	8	20	15	8	8	3	4	12	13	222
B				1	1	1					1			3	1	3	2		1	14
C			1							1									2	4
D							1				1								2	4
E													1	4	2	3	4			14
F		2			1	1	1													5
G																			2	2
H											1									1
I										1										1
J		1								1										2
K									1											1
L									1											1
M								1												1
N								1												1
O							1													1
P							1													1
Q		1					1													2
R						1														1
S	1					1														2
T						1														1
U						1														1
V					1															1
W					1															1
X				1																1
Y				1																1
Z			1																	1
AA		1	1																	2
AB																1				1
AC															1					1
AD	1																			1
AE	1																			1
AF		1																		1
AG		1																		1
AH		1																		1
Total	12	12	9	15	15	25	20	19	22	21	11	20	16	15	12	10	10	12	20	296
Nhap	4	8	4	4	5	7	6	3	3	4	4	1	2	3	4	4	3	1	5	

Table 4. Pairwise population Φ_{st} values for mtDNA sequence data, showing population pairwise Φ_{st} values below the diagonal and associated P values above. The values with a '+' sign above the diagonal are significant after standard Bonferroni correction. Site abbreviations are as given in Table 1 and 2.

	MLEB	MLMR	EMD	MMD	NY	WMA	MMA	BMA	QMA	ME	NB	NS	C	NSE	N	BB	FF	W	NEP
MLEB		-	-	+	-	-	-	+	+	-	-	+	+	-	-	+	+	-	+
MLMR	0.017		-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	-	-
EMD	-0.008	0.008		-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
MMD	0.035	0.018	-0.042		-	-	-	-	-	-	-	-	-	-	-	+	+	-	-
NY	0.000	-0.001	-0.010	-0.016		-	-	-	-	-	-	+	-	-	-	+	+	-	-
WMA	0.023	-0.017	-0.020	0.001	-0.015		-	-	-	-	-	-	-	-	+	+	+	-	-
MMA	0.003	-0.003	0.013	0.015	-0.013	-0.001		-	-	-	-	-	-	+	+	+	+	-	-
BMA	0.049	0.014	0.047	0.033	0.007	0.001	-0.008		-	-	-	-	-	+	+	+	+	-	-
QMA	0.061	0.021	0.060	0.040	0.012	0.005	0.003	0.001		-	-	-	-	+	+	+	+	-	+
ME	0.033	0.005	0.012	0.024	0.003	-0.010	0.001	-0.001	0.001		-	-	-	+	+	+	+	-	-
NB	0.013	0.000	-0.024	-0.028	-0.023	-0.010	-0.018	0.006	0.028	0.008		+	-	-	-	+	+	-	-
NS	0.073	0.045	0.097	0.060	0.020	0.000	0.000	0.003	-0.004	-0.002	0.058		-	+	+	+	+	-	-
C	0.043	0.014	0.047	0.032	0.003	-0.006	-0.008	-0.003	-0.003	-0.007	0.021	0.014		-	-	+	+	-	-
NSE	0.061	0.078	-0.008	-0.021	0.022	0.043	0.071	0.128	0.144	0.098	0.000	0.208	-		-	-	-	-	-
N	0.061	0.078	0.043	0.058	0.059	0.062	0.071	0.128	0.144	0.098	0.051	0.208	0.057	0.013		-	-	-	+
BB	0.146	0.202	0.094	0.128	0.166	0.193	0.214	0.293	0.319	0.259	0.132	0.375	0.258?	0.001	0.038		-	+	+
FF	0.130	0.191	0.111	0.152	0.167	0.177	0.190	0.283	0.308	0.239	0.146	0.385	0.226	0.049	-0.043	-0.044		+	+
W	0.023	0.000	0.034	0.017	-0.016	-0.025	-0.028	-0.026	-0.031	-0.029	0.008	0.000	-0.019	0.136	0.136	0.280	0.290		-
NEP	0.044	0.021	-0.015	0.017	0.013	-0.001	0.009	0.022	0.039	0.013	-0.023	0.045	0.028	0.052	0.075	0.184	0.175	0.010	

Table 5: F_{ST} estimates for ten microsatellite loci, after correction for the presence of null alleles.

	MLEB	MLMR	C	N	BB	FF	W
MLMR	0.0017*						
C	0.0381	0.0379					
N	0.0316	0.0377	0.0648				
BB	0.0701	0.0742	0.0923	0.0513			
FF	0.0763	0.0771	0.0996	0.0549	0.0165		
W	0.0910	0.0832	0.1131	0.0975	0.0909	0.1009	
O	0.0174	0.0182	0.0486	0.0355	0.0773	0.0816	0.0980

Pairwise F_{ST} values are shown. *The value is not statistically significantly different after standard Bonferroni correction.