Population genetic structure of brown crab (*Cancer pagurus*) in Irish waters

Bruce Moran

MSc. in Molecular Ecology

Galway-Mayo Institute of Technology

Supervised by:
Dr. Elizabeth M. Gosling

Submitted to the Higher Education and Training Awards Council, April, 2009

Population genetic structure of brown crab (*Cancer pagurus*) in Irish waters

Bruce Moran

Abstract

The brown crab (*Cancer pagurus*) fishery in Ireland is one of the most important financially and socio-economically, with the species worth approximately €15m per year in the first half of the decade. Only mackerel (*Scomber scombrus*) and Dublin Bay prawn (*Nephrops norvegicus*) are of greater value. Despite this, very little research has been conducted to describe the stock structure of brown crab on a national scale. In this study a country-wide assessment of genetic population structure was carried out. Sampling was conducted from commercial fishing boats from 11/06 to 04/08 at seven sample sites representing the central Irish brown crab fisheries, with one sample site from the UK also included in the study. Six microsatellite markers, specifically developed for brown crab, were used to assess genetic diversity and estimate population differentiation parameters. Significant genetic structuring was found using *F*-statistics (F_{ST} = 0.007) and exact tests, but not with Bayesian methods. Samples from the UK and Wexford were found to be genetically distinct from all other populations. Three northern populations from Malin Head and Stanton Bank were genetically similar with F_{ST} estimates suggesting connectivity between them. Also, Stanton Bank, again on the basis of F_{ST} estimates, appeared to be connected to populations down the west coast of Ireland, as far south as Kerry. Two Galway samples, one inside and one outside of Galway Bay, were genetically differentiated despite their close geographic proximity. It is hypothesised that a persistent northerly summer current could transport pelagic larvae from populations along the southwest and west coasts of Ireland towards Stanton Bank in the North, resulting in the apparent connectivity observed in this study.

Table of Contents

Introduction

Background

The attempt to manage the Irish brown crab (*Cancer pagurus*) population is a necessary and welcome one because the industry generates some of the highest revenue of all commercially fished species in Irish waters: €15.4m in 2004, €12.7 m in 2003, €15.4m in 2002 and €19.2 in 2001 (CSO 2007), as well as providing a source of employment in local rural communities. Currently there is regulation in the form of technical conservation measures (TCMs) on the minimum legal size (MLS) of crab carapace width that can be landed (130mm), restriction on fishing effort for boats greater than 15m, a requirement to hold a polyvalent fishing licence (Tully *et al.* 2006) and a stipulation that no more than 5% of the catch can be comprised of unattached claws (Anon 2007). Various catch data from buyers and fishermen's logbooks have been analysed and used to assess the historical and current states of the various brown crab fisheries around Ireland. The Stock Book (Anon 2007) states clearly that the brown crab stock in Ireland "requires a management plan".

Brown crab are migratory and females can hatch between 1–4 million eggs at one spawning period with mating occurring over the winter months (Tully *et al.* 2006). Eaton *et al.* (2003) report that during a larval distribution study of brown crab in the North Sea, hatching occurred at the beginning of July and by mid August the main hatching zones were still discernable (i.e. hatching was still occurring then). This would give over two months when larvae are present within the water column, with post-larvae settling inshore (Robinson 1999).

This introduction will focus on three main topics: stock assessment methods typically used in fisheries management; a discussion of how molecular genetic technology can be used to determine important information in terms of fisheries management issues which are otherwise difficult to elucidate, including a presentation of underlying theory; presentation of the pertinent data used to describe the situations of the various fisheries around Ireland.

Describing Fisheries: Abundance Indices and Methods

Probably the most basic index describing the state of a fishery is catch per unit effort (CPUE). This index allows one to ascertain the recent history (past decade(s)) of a fishery using two parameters: reported catch (from fishermens' logbooks or from buyers (called landings per unit effort, LPUE); and reported effort, typically measured as the number of days fished, kilowatt hours employed in making the catch, or more simply, as is usually the case in Irish brown crab fisheries, number of pots used to fish. The potential of CPUE/LPUE as an assessment tool is that these indices can highlight a reduction in catch that is indicative of a decline in the stock being fished. They can also relate the reduction of a catch/landing to the increase in effort, if apparent, or can highlight that too great a catch is being taken if no increase in effort is seen. The main disadvantage of these methods as management tools is that they are not particularly powerful because they can only relate historic data to the situation, so if catch begins to fall with a constant effort it may be too late to reverse the trend. Their strength is that they do not require much data, and do not use statistical models to assess the situation in a fishery. This makes them quite transparent processes although reliability is based on the available data (i.e. data from fishermens' logbooks are anecdotally less prone to error than data from buyers). It should also be noted that buyer data does not account for where fishing occurred and so no information on fleet behaviour can be summarised from such information.

Traditionally defining maximum sustainable yield (MSY) has been the 'goal' of fisheries management (Hilborn and Walters 1992). MSY relates to the greatest catch (yield) that can be taken by a fishery that is sustainable *ad infinitum*, i.e. the MSY can be taken year after year without affecting the growth and reproduction of the stock being fished. In theory this is because there is an equilibrium between parameters that positively affect the stock such as growth and reproduction and those that negatively affect it such as natural and fishing mortality (King 1995). The 'surplus' of the particular fishery is taken, leaving a theoretically stable and healthy stock to produce the same surplus for the next fishing season. This is a difficult concept to put into practice as the models, most often referred to as production or surplus models, are based on several spurious assumptions: stocks are closed (i.e. no immigration/emigration) and recruitment/mortality dynamics are constant (King 1995). These are surprising assumptions given the basic nature of fisheries. Fish stocks are complex and dynamic systems that are very difficult to accurately measure because they are influenced by

external forces such as hydrology, food abundance and predation (themselves influenced by climate changes), as well as intrinsic factors such as recruitment and migration.

Production models require a continual small increase in fishing to allow determination of MSY. This can lead to a situation where a reduction in effort and catch is required, demands not easily implemented in a fishery (Hilborn and Walters 1992). Other more conservative forms of MSY can also be used in fisheries management, for example $F_{0.1}$ which models the yield based on 10% of the slope of the yield-per-recruit function at the origin (Hilborn and Walters 1992, p.459). This may however be as arbitrary as MSY, or may indeed reduce catches below what is economically viable for fishermen. Even conservative production models fall foul of a basic problem: they do not account for the dynamic nature and complexity of the system they are modelling.

Two primary research methods are used to attempt to define several key parameters of a stock; tag/recapture and depletion studies allow estimation of abundance and movement of stock in a fishery. Tag/recapture studies involve taking individuals from a catch, attaching a unique tag stating where and when they were caught, then releasing the individuals. Once recaptured, the direction and duration of movement can be quantified. This may elucidate a general trend in migration/movement of a stock if enough individuals are recaptured (Fahy *et al.* 2004). Stock depletion studies can be used to estimate stock abundance; individuals from a distinct mapped area are captured and tagged as before and then released. Fishing is then carried out in the area over days, weeks and/or months until the relative abundance of tagged individuals is 0 (i.e. until CPUE reaches 0), all tagged individuals are either caught or leave the area. Models can then be used to estimate abundance (Hilborn and Walters 1992, p.396 for further information). Similar models can be used on tag/recapture data, but are harder to implement since relative abundance is over the entire fishery and return of tagged individuals is in commercial fishermens' hands, and they may not be too enthusiastic to return part of their catch for what are sometimes paltry rewards (Tully *et al.* 2006).

Stocks and Population Structure

Knowledge of stock boundaries is necessary for good management practice, but there are several ways in which such boundaries are defined. They are often anthropogenic and

cartographic in nature. Anthropogenic boundaries refer to 'patches' fished historically by one family or group, whereas cartographic boundaries refer to specific areas on maps considered to represent one stock (M. Hayes pers. comm. 2007). Such boundaries may be difficult to regulate as they are arbitrarily defined and do not take into account the underlying population structure of the species being fished. Therefore, a biologically based approach to defining stock structure is now regarded as the most appropriate for proper management of fisheries (Begg *et al.* 1999).

The concept of 'unit stock' is continually evolving and now appears to be based on the fishery that it supplies and the controls put in place to manage it (Cadrin *et al.* 2005). Booke (1999) distils the issue down to a single point: the need for a marker, either phenotypic or genetic, which will not change greatly over generations thus allowing stocks to be distinguished from one another. This thesis deals with a biological definition: a stock can be defined as a group of individuals of a single species that is reproductively independent of similar groups and therefore has little physical connection with them (King, 1995). It is implied here that loss of physical contact results in reproductive independence, an idea that is mirrored in Mayr (1973) on speciation, as well as in population genetic theory such as island models of migration (Hudson 1998).

While Coyle (1998) argues the necessity for more than one method of stock identification due to the very low levels of migration needed to effectively genetically homogenise such groups (Wang 2004), defining stock structure genetically has a fundamental role to play in stock assessment and fishery management. Genotyping individuals using highly polymorphic genetic markers, together with the recent development of powerful and computationally intensive statistical software, will allow managers to make more informed choices. A primary necessity for Irish brown crab fisheries should thus be the understanding of the genetic structure of stock(s) upon which the entire fishery is based.

Applications of Population Genetics

The above definition of stock places importance on restriction of movement (and hence breeding outside local populations), which results in a distinct 'gene pool' in which the

various evolutionary forces may act to diversify, genetically, one stock from another. The term 'population' rather than 'stock' will be used henceforth in this section since it is better defined and more commonly used in population genetics literature. Here population is used to mean a part of a total population, which is, as in the definition of stock, reproductively and geographically isolated (to an extent) from other such populations. The evolutionary forces impacting on populations are: random genetic drift, mutation, migration, inbreeding and natural selection. In classic population genetics theory a population in which none of these forces are acting, and so which has no change of allele frequencies between generations, is said to be in Hardy-Weinberg equilibrium (HWE). This is usually an assumption made under a null hypothesis when testing for the action of any of the evolutionary forces listed on a sample population.

Selectively neutral, non-coding regions of DNA are used as molecular markers and all alleles at a selectively neutral 'locus' can be used to genotype individuals without interference from selection. In terms of population genetic theory selection therefore does not have to be modelled and therefore can be excluded from the above list. However, selection can be important in terms of population genetic studies, even when neutral markers are used, if a particular locus is 'linked' (i.e. is in close proximity on a chromosome) to another locus which is not selectively neutral, for example a gene. This phenomenon, termed 'linkage disequilibrium', should be tested for prior to analysis of data to determine if any such linkage, and subsequent selection acting upon the marker locus, is causing an over-representation of alleles.

Kimura's neutral theory (Kimura 1968; 1983) dismisses the effect of classic 'Darwinian' selection on genes (first proposed at the individual and not genetic level by Darwin (1856)) and puts random genetic drift ('drift' henceforth) at the forefront of the modern synthesis of evolution. Drift is defined as a stochastic fluctuation of alleles, which operates to ultimately 'fix' one allele (i.e. remove all but one allele from the population). Fixation occurs at the rate of $4N_e\mu$, N_e being the effective population size, μ the mutation rate, resulting in a mutation/fixation equilibrium in which mutation creates new variety and drift acts to bring one of the alleles in the population towards fixation. Therefore, within populations, new mutations occur and are brought towards fixation resulting in differentiation at the genetic level between such groups.

Migration is a force that acts to reduce differentiation between adjacent populations by introducing new variants into their respective gene pools. If differentiation between populations is a result of new mutations specific to one or other population (not yet shared by way of migration, or by random mutations occurring to generate two synonymous mutations in different populations), and mutation and drift are in equilibrium, then we can use measures of inbreeding within populations to determine how connected populations are.

Inbreeding occurs in a population when related individuals mate producing offspring that are 'homozygous', or have the same allele at a particular locus on each member of a pair of homologous chromosomes. The level of homozygosity increases at a rate determined by the number of mating individuals within the population and the relatedness of these individuals (Falconer and Mackay 1996). If no new individuals are introduced into the population then it is likely to become genetically distinct from other populations over many generations, 10-100,000s dependent on population size, not a particularly long time in evolutionary terms. Populations tend to become genetically dissimilar based on their isolation. This is due to: inbreeding and the related increase in homozygosity, which can be quantified; mutation introducing new alleles which are then brought towards fixation (or at least different frequencies than in other populations) by drift. We can therefore infer levels of migration between populations from the degree of inbreeding.

Genetic Methodologies

The above theory can be implemented in natural populations because of the development of several laboratory procedures that allow the genotype of an individual to be identified. Of these techniques 'molecular markers', 'polymerase chain reaction' (PCR) and 'electrophoresis' are the foundation. Applying molecular markers at multiple loci means that an overall allele frequency can be quantified for samples of individuals collected at a number of locations. Using these data it is possible to statistically analyse the degree of differentiation between these samples and thus obtain information on the population genetic structure of the study species.

Over the past 20 years selectively neutral 'microsatellite' marker loci or 'simple sequence repeats' (SSRs) have been widely used in population studies due to their high

polymorphism (numbers of alleles per locus, 10-50) and mutation rates (from 10^{-2} to 10^{-6} per generation (Schlötterer 2000)). Microsatellites are randomly repeated sequences of 1-6 base-pairs (e.g. ATAT or GAG). They are found in all prokaryotic and eukaryotic cells and are present in high numbers in non-coding regions of the genome. Because of their high mutation rates and high polymorphism it is possible to detect structuring between populations when the analysis of other less polymorphic markers (e.g. allozymes) cannot.

Microsatellite regions are bounded by a specific sequence of nucleotides from which a 'primer pair' is constructed. These primers (one at either end of the sequence, called the 'three prime' and 'five prime' (denoted 3', 5')) are used to make multiple copies of, or 'amplify' the repeated sequence of base-pairs. DNA is extracted from a tissue sample and combined with several PCR reagents in a 'master mix'. During PCR the complex of DNA and reagents is heated and cooled multiple times. Each heating event denatures the DNA causing it to 'unzip' and become two single strands instead of the usual double helix structure. The two primers 'hybridize' (attach) to the related areas on the DNA (to the complementary nucleotides). An enzyme called *Taq polymerase* then functions in 'extension', the addition of DNA nucleotides to the single strands beginning at the annealed primer sites. This creates new DNA strands (and twice the amount before denaturing), one strand of which is from the original 'template' DNA and one of which is the newly extended strand (Arnheim and Erlich 1992).

PCR product containing an exponentially increased amount of the target (microsatellite) region can be easily visualised using the technique of electrophoresis. This technique consists of running a sample through a thin gel matrix. As DNA is slightly negatively charged larger pieces move more slowly towards the positive pole. After running a current through the gel for an amount of time one can visualize (using various methods) the resultant 'bands' which correspond to a specific length in base pairs. Given a size standard base-pair sizes can be determined for a particular set of primers. Using multiple primers pairs, each specifically developed for an individual microsatellite locus, the genotype (either single-banded homozygote or double-banded heterozygote) of an individual can be scored. Ultimately, the genotypic data are compiled for all individuals in a sample (or population) and analysed using various statistical methods.

Statistical Methodologies

There are two categories of methodology used to determine population structure, fixation statistics (commonly called *F*-statistics) and Bayesian methods. *F*-statistics (Wright 1951; Weir and Cockerham 1984) are described as an indirect method of determining gene flow between populations because they use genotypes from a sample of individuals and do not actually physically track groups of individuals. These methods relate the degree of genetic variation in the total population to that of the sub-populations and individuals making up the total population. F_{ST} is the most commonly used *F*-statistic and refers to the ratio of genetic variation in a subpopulation (*S*) to that of the total population (*T*). The other terms, F_{IS} and F_{IT}, relate *F* (the fixation index or genetic variation based on the measured homozygosity) of *S* (subpopulation), *T* (total population) and *I* (individual) to one another. F_{ST} is a measure of the genetic differentiation within subpopulations (or, the reduction in heterozygosity) caused by inbreeding due to population subdivision. F_{IS} and F_{IT} are measures of deviation of individuals (in sub- and total populations) from Hardy-Weinberg equilibrium (Hedrick, 1999). F_{ST} has a relationship with migration rate (*m*, per generation):

$$F_{ST} \approx \frac{1}{1+4N_e m}$$

under the assumptions of a Wright-Fisher population (random-mating, constant size, $m>>\mu$, with constant *m* per generation), although not all agree, see Whitlock and McCauley (1999).

Statistical techniques such as *F*-statistics, along with other measures of diversity such as allelic richness, A_e (the mean number of alleles per locus based on the minimum number of individuals genotyped in a subpopulation), heterozygosity, H_o (heterozygosity of subpopulation) or H_s (heterozygosity of total population), are used to infer the state of inbreeding, and therefore migration, in a structured population. Gene flow is acting to counter inbreeding, which is what is essentially measured. F_{ST} has a value of 0-1; 0 denotes no inbreeding and so the population is constantly mixing with random mating; a value of 1 denotes completely inbreeding sub-populations. A good discussion of *F*-statistics in fish populations can be found in Waples (1998), who reports a low mean F_{ST} estimate of 0.062 from 57 marine species, in contrast to a value of 0.222 from 49 species of freshwater fish, thus highlighting the greater potential for migration in marine species.

Likelihood is an important concept because it gives the 'conditional probability' of a particular situation, given that one or more random variables are set at a value. Thus, we can set the parameters of the model dictating the number of subpopulations to a value between 1 and 10 and determine the likelihood of that situation given the data; which means that one is changing the hypothesis to ask a number of specific questions of the data, instead of asking if the null hypothesis is true or not, as frequentist methods do.

Bayesian methods for analysing population structure mainly use methods called Monte Carlo Markov chain methods. The Markov chain generates random variables, the generation being dependent only on the previously generated variable. This results in a step-wise changing of the particular variable, moving up or down whilst the other variables are held at a specific value (Beaumont and Rannala 2004). The method is used to model changes in allele frequencies and how these can change the potential sub-population that an individual may belong to. Changes to the MCMC result in changes to the likelihood. Generally, if the MCMC method is implemented well enough i.e. if the number of steps the MCMC takes is large enough, a most likely situation is found. This results in a most likely population of origin for an individual, given their genotype, based on the allele frequency distribution from the entire data set.

The program STRUCTURE (Pritchard *et al.* 2000) will be discussed in greater detail in the Results section, detailing how the program models the various parameters and how this is used to assign individuals to populations.

Irish Brown Crab Fisheries: Historic and Current States

The Irish brown crab fisheries may be split into several constituent parts: North West (Donegal and Mayo; NW), South West (Cork and Kerry; SW), South East (Waterford and Wexford; SE) and Mid West (Clare and Galway; MW) (Fig. 1). Of these, the NW component has accounted for between 50-80% of landings since 1990, and so the majority of interest is in this fishery, and most research has been conducted on it. This fishery, along with the South East fishery, is comprised of an inshore and an offshore component (inshore being <12 miles of coast). Offshore fishing is undertaken from 'vivier' vessels, usually <18m in length, during 5-6 day trips. Catches are held in tanks on board for the duration. The vessel captains freely

contribute their logbooks containing catch and effort data as well as information on areas fished to fisheries researchers for analysis. Logbooks have not historically been available for inshore fisheries but are becoming increasingly so. Much of the early analysis for inshore fisheries is derived from buyers' data, and so all catch effort analysis is in the form of LPUE.

From these data it is possible to review the history of the brown crab fisheries in Ireland. Tonnages are given in Figure 2, with LPUE shown in Figure 3 for NW Ireland. For fisheries outside of the NW LPUE data is only available for 2002, 2003 and/or 2004. These are: inshore SE ~1.08kg/pot falling to ~1.01kg/pot in 2002, 2003 respectively. Offshore SE was stable over the period 1990-2002 at 1.4-1.6kg/pot, these data coming from a French vivier fleet. SW showed 1.88-2.14kg/pot from 2002-2004 in Kerry from fishing records, but Cork data were poor (from surveys only) and although there was a relatively high LPUE of ~1.32 and ~2.39kg/pot in 2002 and 2003, respectively, however this data is unreliable (Tully *et al.* 2006). MW from 2002 was ~0.76kg/pot in Galway, no other data are available (all data Tully *et al.* 2006 except SE offshore, Fahy *et al.* 2002). Although Figure 2 gives the impression of a steady rise in landings up to 2004, the LPUE for NW fisheries gives a different picture. On the whole there has been a general decrease in LPUE from 1990 on. A similar trend is observed in other fisheries (Tully *et al.* 2006). Speculation has been that early 'exploratory' fishing caused a 'mining' of older and larger individuals from the breeding stock. This could have a significant impact on abundance as brown crab are quite long lived (Edwards 1979). During the 90's fishing settled into a stable phase. However, this stability may be due to an increased effort with greater returns, causing the decline evident from 2000 to 2004.

Figure 1: Location and extent of crab fisheries in Irish waters, divided into constituent regions (URL 1).

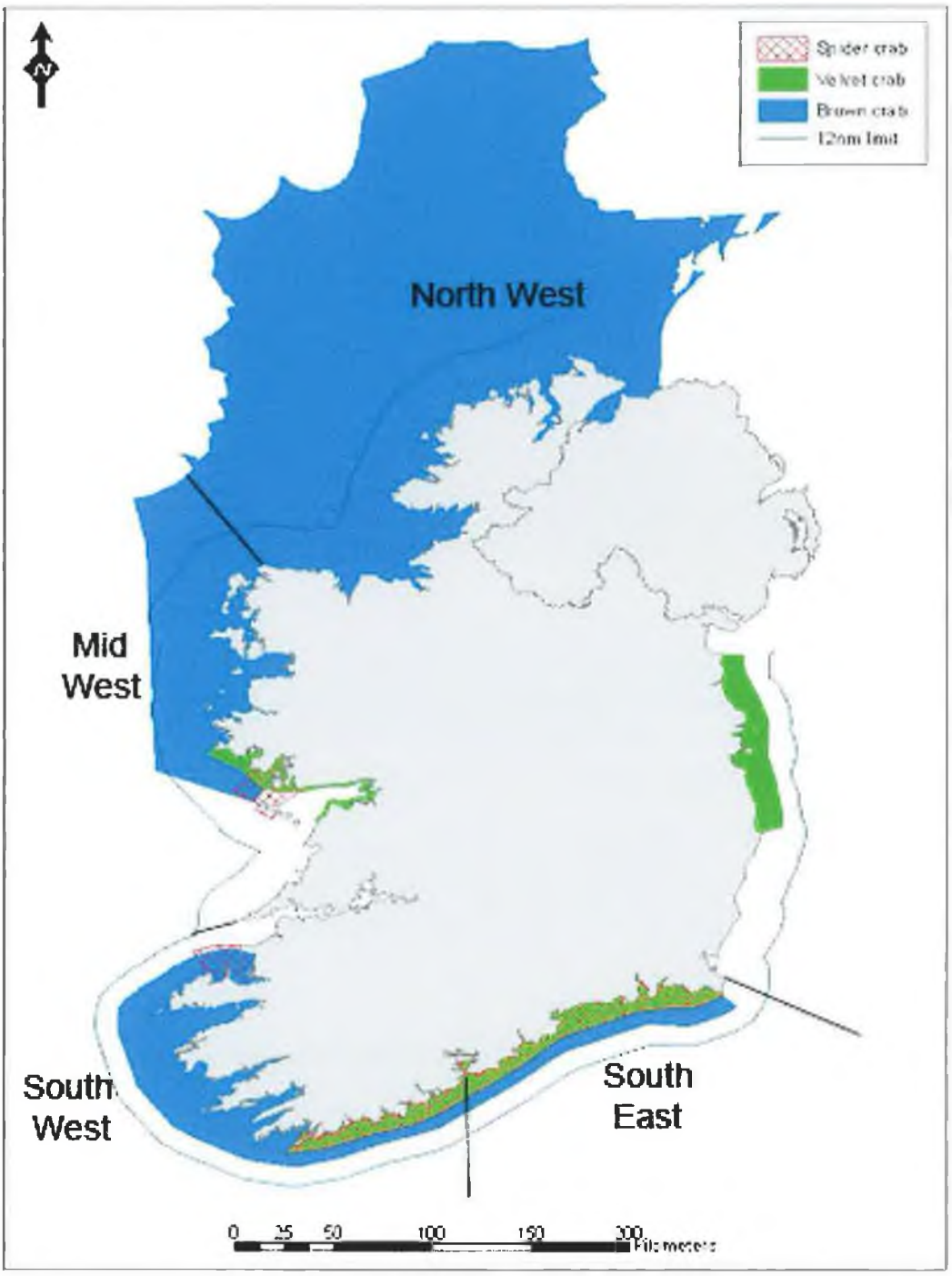
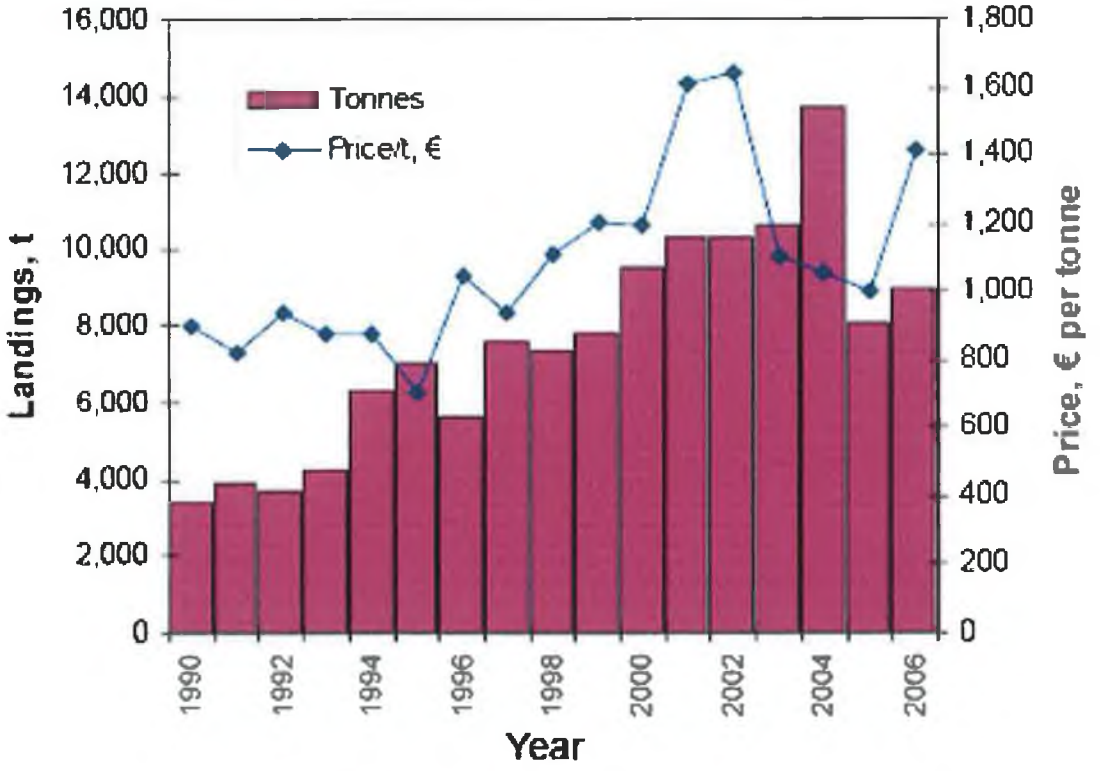


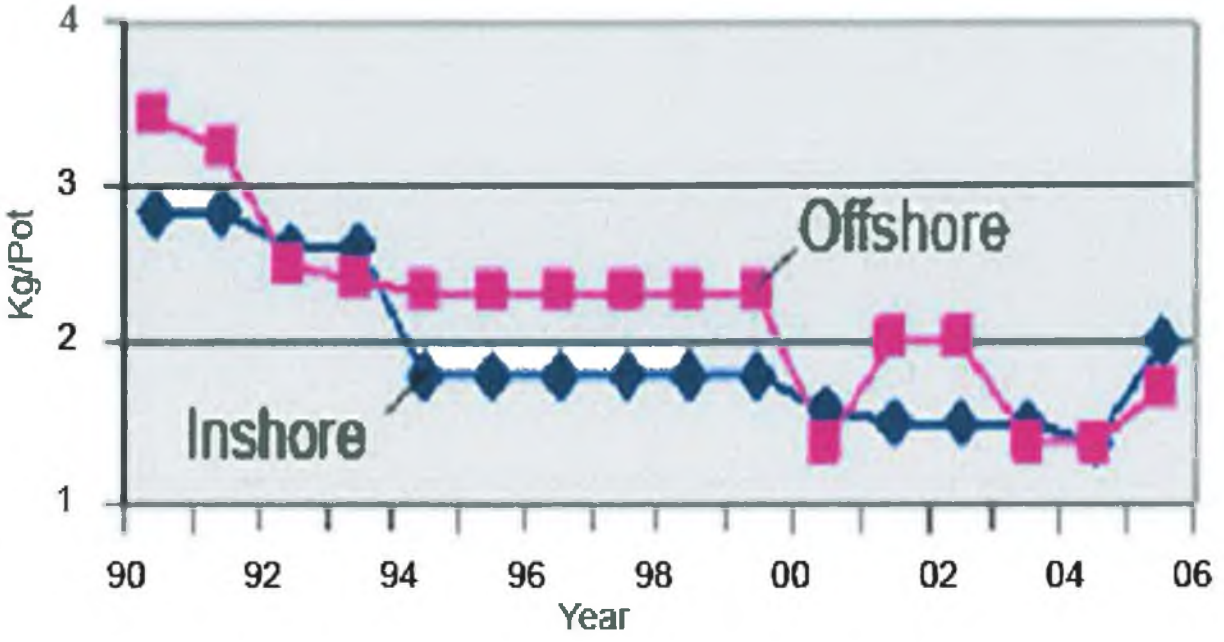
Figure 2: Total landings in tonnes and first sale value of brown crab (1990-2006 inclusive) (Anon. 2007) for Ireland.



There has been a reported recovery in 2005 for inshore and offshore NW fisheries with LPUE of ~2kg/pot, ~1.7kg/pot, respectively (ICES 2006), although this needs to be verified when the next two years LPUE data are made available. The introduction of effort controls in Irish law that allow immediate closure of fisheries in the NW when arbitrary landing limits, determined for each season and based on previous seasons' CPUE, are exceeded (Crab Fisheries Management and Conservation Regulations 2005, SI 676) may signal some degree of turn around although the extent of their use and how strictly they are enforced are important.

One interesting final note should be made regarding the findings of Tully *et al.* (2006) on the variance in catch and LPUE. A general linear model (GLM) was used to assess the variance components of catches made on different vessels, using different soak times (the time that pots are left in the water) at different times of year or when two sets of 'gear' i.e. strings of pots, were in competition.

Figure 3: Landings per unit effort (LPUE) in kilograms per pot (Kg/pot) in the Northwest inshore and offshore brown crab fisheries (from data in Tully *et al.* 2006; ICES 2006).



The model found that the most important factors were time of year and time of month, as well as soak time, but that these factors accounted for only 28% of variance among the data, the rest being random. This indicates that while regulations on effort and gear used may be important, this is not the major factor influencing the numbers of crab caught in Irish fisheries, and it may be necessary to begin use of other regulatory devices such as quotas or total allowable catch (TAC) to help the Irish brown crab fisheries to consistently reach their economic potential.

Aims of the Study

This study aims to:

1. Describe genetic diversity of samples of brown crab from around Ireland using genotype data from six polymorphic microsatellite markers.
2. Describe population genetic structure using 'traditional' methods of *F*-statistics and exact tests, novel Bayesian methods, and isolation-by-distance regression analysis.
3. Relate structuring found to factors potentially affecting the process, for example current or other barriers.

Materials and methods

Sampling Procedure

Sampling took place from November 2006 to April 2008. Samples of *Cancer pagurus* adults were collected at each of seven sample sites around Ireland (Figure 1; Table 1); sample sites represent: Stanton Bank, Malin Head (two sets of samples), Galway (two sets of samples), Kerry and Wexford. A further sample from Bridlington, Northumberland, UK was sent by N. McKeown (Royal Holloway, University of London, Surrey). The data set included 106 individuals; DNA samples for 82 of these individuals were genotyped in this study to allow cross-calibration of data. The final number used from this data set following calibration was 106 individuals. Sampling from all Irish populations was conducted on commercial catches during single days fishing (inshore vessels) and during week-long trips (vivier vessel). All fishing was carried out in distinct socially and historically defined 'patches' within each sample site. Samples of one of the eight walking legs were collected during fishing on inshore vessels and while transferring crab from live holding tanks to boxes on vivier vessels. Legs were stored in 96% ethanol to preserve the samples prior to DNA extraction. Sex of individuals was recorded upon sampling. The areas fished by vessels were recorded following contact with vessel captains after sampling had taken place. Two location points indicate the start and end of the string of pots from which samples were taken.

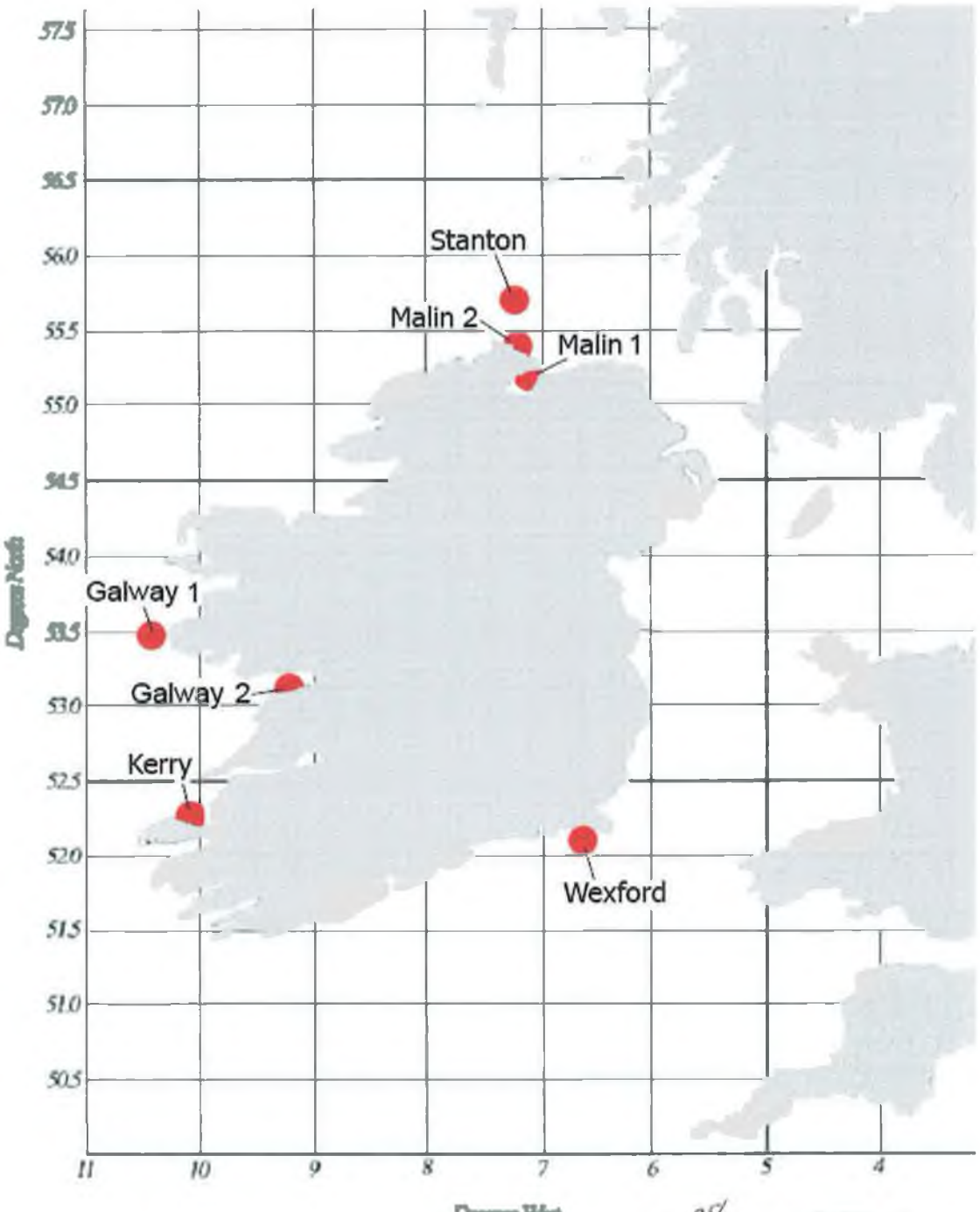
DNA Extraction

The method of DNA extraction was very similar to that of Sambrook *et al.* (1989) except that no sodium acetate was added before first phase of ethanol cleanup. DNA pellets, stored at –4°C prior to PCR, were resuspended in 100 µl autoclaved SDW, a 1:10 dilution of which was used in the initial PCR protocol.

Table 1: Sample site names, numbers of individuals (male:female), dates of sampling and locations (in decimal degrees).

Sample site	Number of indivs. (m:f)	Date	Location	
Stanton Bank	100 (21:79)	Jun-07	55.70N	7.26W
			55.68N	7.24W
Malin 1	100 (20:80)	Jun-07	55.20N	7.08W
			55.20N	7.07W
Malin 2	62 (12:50)	Nov-06	55.31N	7.20W
			55.30N	7.21W
Galway 1	78 (24:54)	Mar-08	**53.48N**	**10.32W**
Galway 2	86 (86:0)	Apr-08	53.14N	9.15W
Kerry	100 (27:73)	Jul-07	52.19N	10.05W
			52.18N	**10.06W**
Wexford	100 (18:82)	Jul-07	52.07N	06.41W
			52.05N	**06.42W**
UK Bridlington	106 (0:106)	Jun-06	53.52N	00.31E

Figure 4: Map of Ireland showing sampling locations (red dots) of sites from Table 1 (all approximate). UK Bridlington not shown.



PCR Protocol

PCR was undertaken using six microsatellites: Cpag-5D8, Cpag-3A2, Cpag-1B9, Cpag-3D7, Cpag-2A52 and Cpag-6C4B (McKeown and Shaw 2007). Reverse primer sequences were labelled at the 5' end with either 700 or 800 IRDye™ (Licor) to allow visualisation. These dyes allow two primer products, which amplify at the same base pair range, to be run on a single gel without conflict of signals, thus reducing cost and number of gels run by 50%. Primers Cpag-38, Cpag-5D8, Cpag-6D4B and Cpag-3A2 were 700 IRDye labelled, the remainder with 800 IRDye. Cpag-6D4B, Cpag-3A2, Cpag-1B9 and Cpag-2A52 were run together as a four- locus set, the rest as a two- locus set.

PCR amplifications were carried out in 10 µl reaction volumes containing 1mM dNTPs (Bioline), between 1 and 10pM/µl of each primer (MWG Biotech), 5U/µl BIOTAQ DNA polymerase (Bioline), 1.5 mM or 2 mM $MgCl_2$ solution (Bioline) and NH_4 buffer (Bioline) and 1 µl DNA at ~3pmol. PCR products were stored at –4° C prior to use. Annealing temperatures, primer and $MgCl_2$ concentrations for PCR reactions are given in Table 2.

Analysis of PCR Products

PCR products were run on a LiCor 4300 DNA Analyzer using 0.2mm polyacrylamide gels. Gels were prepared using 20 ml 6.5% polyacrylamide gel (KB-plus™), adding 150 µl 10% ammonium persulphate (APS) and 15 µl TEMED (Sigma) to induce polymerisation. The gel was left for 2 h to polymerise and was then pre-run in 1L 1xTBE buffer (10.8 g Tris-base, 5.5 g boric acid, 0.95g EDTA disodium salt, 750 ml SDW). PCR products were diluted 1:10 in loading buffer (bromophenol blue dye in formamide) and denatured at 95°C for 4 min prior to loading following the pre-run to allow the gel to come to running temperature of ~45°C. One µl of product:buffer was loaded into each of the 48 lanes. Allele size was calibrated using a 350 bp molecular weight ladder (LiCor). To ensure reliability of allele scoring two heterozygous reference individuals were used for each primer pair representing both the most common alleles and the size limit in the initial test runs. SAGA™ software was used to visualise gels and to database results.

Data Compilation

Genotype data were compiled from text file printouts from SAGA into three and then six numbered codes relating to allele size (bp) using Microsoft Excel. Text files were then constructed in GENEPOP 4.0.6 format (Raymond and Rousset 1995; Rousset 2008). Data from 82 individuals of known genotype previously determined by N. McKeown from the UK Bridlington population were used to cross-calibrate with data generated in this study. These 82 individuals were genotyped using the LiCor system as for the Irish populations. When differences were found between McKeowns' genotypes and our own (determined using both allele frequency analyses and checking by eye using MS Excel) the number of basepairs in the difference was noted (all were between 3 – 9 bp). The remaining 24 individuals from the UK sample which were not genotyped in this study were calibrated using this information. Cross-calibration information is given in Appendix 1. The calibration exercise was undertaken to potentially allow inclusion of further data from N. McKeown's work.

A further two GENEPOP input files were constructed containing male only and female only samples. These files were used to test whether any further structuring was evident based on either of the sexes alone.

MICROCHECKER (van Oosterhout *et al.* 2004) was used to check for null alleles and errors e.g. wrong sized or misnumbered alleles. The entire database of individuals was then checked using the filter function in MS Excel to find any redundant genotypes, characterised as being either: (i) exactly synonymous genotypes (three individuals) or (ii) genotypes at only one locus i.e. missing data for all but one locus (two individuals). One single GENEPOP file was constructed containing all 728 valid individuals sorted into the eight sample sites previously defined. This file was used to create input files for FSTAT 2.9.3 (Goudet 1996, Goudet 2001), STRUCTURE (Pritchard *et al.* 2000) and BAPS 5.0 (Corander *et* al 2003) which were constructed using CONVERT (Glaubitz 2004).

Table 2: Annealing temperatures (Ta), primer concentration, $MgCl_2$ concentration, repeat sequence, bp size range and dye type.

Locus ID	Repeat Motif	Ta (°C)	$MgCl_2$ (mM)	Primer conc. (pM/µl)	Size range (bp)	IRDye© Wavelength
Cpag-3A2	ATG	59	2	10	248-260	700
Cpag-2A52	ATG	59	2	10	146-185	800
Cpag-1B9	ATCT	56	2	5	214-282	800
Cpag-6C4B	AGTT	55	1.5	1	160-192	700
Cpag-5D8	TAC	59	2	10	156-234	700
Cpag-3D7	TCTG	56	2	10	158-214	800

Genetic Diversity Analysis

GENEPOP was used to determine total and mean number of alleles per locus and departure from Hardy-Weinberg equilibrium (HWE) using the MCMC method of Guo and Thompson (1992) with 10,000 dememorization steps, 500 iterations and 10,000 iterations per batch. The null hypothesis of this method is random union of gametes, the alternate hypothesis (H1) being non-random union of gametes inferring an underlying evolutionary force disrupting HWE (versus the other possible H1 of heterozygote deficiency/excess causing this disruption as in the other available tests).

FSTAT was used to determine: allelic richness; allele frequencies; linkage disequilibrium (LD) using the randomisation method, P-values being at 5% nominal level (P-values with Rice (1989) Bonferroni correction) and F_{ST} estimates (over all loci and at individual loci), P-values being again at 5% nominal level with Bonferroni correction. F_{ST} tests were carried out for each locus individually to determine if any loci showed any particular deviance from the general trend seen in the overall analysis.

GENETIX (Belkhir *et al.* 1996) was used to determine observed (H_O) and expected (H_E) heterozygosities, based on actual counts and allelic frequencies, respectively. Expected frequencies were calculated using Nei's (1978) unbiased method, which reduces error due to sample size.

BOTTLENECK (Cornuet and Luikart 1996) was used to determine if any of the loci in any of the populations showed signs of having undergone a population bottleneck. The stepwise mutation model (SMM) and infinite alleles/stepwise mutation model (IAM+SMM) were used with 1000 iterations. The SMM is thought to be more representative of the method of mutation in microsatellites. Wilcoxon tests for heterozygote excess and the mode-shift reported by the program were used to determine if the allele frequencies showed signs of recent bottlenecks. The mode-shift, from an L-shaped allele frequency distribution assumed under a mutation–drift equilibrium model (which should be evident in neutral molecular markers) to a non-L-shaped distribution, is caused by a sudden increase in alleles into a population. If a bottleneck has occurred alleles will be eliminated with greater potential for less frequent alleles to be lost. Heterozygosity will be reduced subsequently but there will be a stage, lasting for several dozen generations, where an excess of heterozygotes is seen in terms of the allele frequencies in the population. New alleles will quickly be introduced by

migrants, leading to a rise in heterozygosity and a shift in mode, which is tested for by the program.

Where stated analyses underwent sequential Bonferroni correction to avoid Type 1 errors (false positives) which otherwise occur randomly with x frequency where x is 1-confidence limit (Rice 1989). This correction factor is used despite criticism of the method (Moran, 2003) and potential reduction of power (Nakagawa, 2004).

Independent two-tailed *t*-tests were carried out for genetic diversity analyses at the marker and population levels to allow construction of confidence intervals (CI). Results between these CI are considered statistically significant. This test was used to determine if a particular locus or populations' results could be considered reliable in terms of the population structure evident.

Population Differentiation Analysis

Population differentiation was analysed using global and pairwise F_{ST} (Weir and Cockerham, 1984), and Fisher's exact tests of genotypic distributions (Raymond and Rousset, 1995) between pairs of populations. A multi-dimensional scaling (MDS) plot, conducted using the isoMDS command from the MASS package (Venables and Ripley 2002) in *R* 2.8.1 statistical software (R Development Core Team, 2008) was used to visualize connections apparent from the F_{ST} analysis. This method uses a set of dissimilarities (F_{ST} results) and returns a set of points. The distances between these points are approximately equal to the dissimilarities. The LOSITAN package (Beaumont and Nichols, 1996) was used to determine if any of the loci were under selection, based on F_{ST} and heterozygosity estimates, which may account for outlying results.

An isolation-by-distance model was assessed by linear regression analysis in MS Excel. An ANOVA was constructed using $F_{ST}/(1-F_{ST})$ and ln (distance in km) to determine whether there was a significant correlation between the two parameters. Distances between sample sites were calculated using the sample site coordinates and a 'shortest-straight-line' distance through water involving the least number of points between the two sample sites as possible. Points between sample sites are given in Appendix 2, along with a map showing

their relative positions to sample populations. Distances between points were determined using www.googlemaps.com.

Both STRUCTURE and BAPS were used to determine the most likely number of 'clusters' (K) of populations for the entire data set without giving the sampling site of origin of individuals as a prior.

STRUCTURE was run under the admixture model with parameters (random variables): Z, a vector consisting of $z^{(i)}$, the (unknown) population of origin of individual i; P, a vector consisting of p_{klj}, the (unknown) frequency of allele j at locus l in population k; Q, a vector consisting of $q_k^{(i)}$, the proportion of individual i's genome coming from population k. The data, X, is a vector consisting of $(x_l^{(i,1)}, x_l^{(i,2)})$, the genotype of i at locus l.

The allele frequencies are modelled with the Dirichlet distribution using a uniform prior equal to 1, resulting in an assumption of Hardy-Weinberg equilibrium. This also results in an assumption of unlinked marker loci, although Falush *et al.* (2003) have introduced a method to remove this assumption in STRUCTURE.

The analysis in STRUCTURE was carried out to estimate the number of populations, K, for potential K values of 1 – 7. Runs for each K value were independently replicated three times. A burn-in of 100,000 and run length of 1,000,000 iterations were used. Sampling population information is not used by the program. The output of $LnP(D)$ equivalent to the log likelihood of $P(K/X)$ determined above was used to determine the most likely K value.

Parameters under BAPS (for the same i, j, l and k as above, and N_p sampling populations) are: θ, a vector of all (unknown) allele frequencies (θ_k) consisting of p_{klj}; N, a vector consisting of all observed marker allele counts in a population, n_{klj}; Q, the admixture proportion represented by the vector $Q \in [0, 0.01,\ldots,0.99]$.

The data is represented by a prior for the allele frequencies, X (as in STRUCTURE), by v_p, the number of populations with different allele frequencies (where $v_p = 1,\ldots, N_p$) and by S, which is how Corander *et al.* (2003) model the 'partition' (or more generally the structure) of the sampling populations. For N_p sampling populations, S is an N_p x N_p matrix, the elements of which are 0 or 1 based on whether allele frequencies are equal in the $(1,\ldots,N_p)$ populations (i.e. if $\theta_r = \theta_w$, the matrix entry $S_{rw} = 0$). Therefore, when two sampling populations have the same allele frequencies they are joined to form one

population. Corander *et al.* (2003) state that under the multinomial model they construct, the assumptions they make, of Hardy-Weinberg equilibrium and unlinked loci, arise "naturally from the basic modelling principles of the Bayesian framework" (p.368).

The output of the program for determining K is derived from enumerative calculations when N_p<10, which it is in all data sets analysed here. The method computes the marginal likelihood $P(v_p, K)$ divided by the sum of marginal likelihoods of all possible partitions and outputs the ten K values with highest probabilities. This "equals the sum $\sum_{vp-1}^{Np} \sigma_{vp}^{Np}$, where σ_{vp}^{Np} is the Stirling number of the second kind" (Corander *et al.* 2003, p.373).

BAPS' analysis involves a mixture analysis using the 'Clustering of Groups of Individuals' option, which calculates the probabilities of K as described above. The maximum number of clusters (inferred populations, K) is specified by the user and allows multiple entries to be made, so values of K of 1 – 7 were entered five times each. The posterior probability is thus exactly calculated, as opposed to an estimate made by MCMC. The results give only the five highest marginal likelihoods, which means that graphical representation is difficult. The results are also not directly comparable to those of STRUCTURE.

The main difference between STRUCTURE and BAPS are the algorithms used and the underlying methods of calculation. STRUCTURE uses a Dirichlet function to model priors whereas BAPS uses a Stirling number method. The direct result for the user is that BAPS is far faster in calculating marginal likelihoods, in some cases by 12 hours per run depending on the K values involved.

Results

Genetic Diversity

From 728 individuals genotyped a total of 89 alleles were detected across all loci with over half of this polymorphism owing to two loci (Cpag-5D8 and 1B9). The number of alleles per locus ranged from eight (Cpag-6C4B and Cpag-3A2) to 33 (Cpag-5D8) which was at least twice the number reported by McKeown and Shaw (2007) for the same loci. Across all loci 19 private alleles were found, occurring at all loci except Cpag-6C4B and in all populations except Stanton and Malin 2. Frequencies of private alleles were low (Table 3) at between 0.0014 (Cpag-5D8 and 1B9) and 0.0041 (Cpag-3A2) with the highest number occurring in the Galway 1 sample set (0.0041; Table 4). The mean numbers of alleles per locus (Table 3) ranged from 4.5 (Cpag-3A2) to 22.88 (Cpag-5D8). Mean observed (H_O) and expected heterozygosity (H_E), using Nei's unbiased method (1978), across all populations was 0.659 and 0.662, respectively, with no significant deviations apparent between the two for loci or populations. Allelic richness (Table 3), scaled to 59 individuals per population, ranged from 4.31 (Cpag-2A5B) to 20.63 (Cpag-5D8). Null alleles and large allele dropout, which may cause an increase in homozygosity and occur due to mutation in the primer region, were not detected in any of the sample populations using MICROCHECKER, which also resolved all genotyping errors. Several outliers were found (alleles that were not at the 'correct' size being +/- 1 bp)) and were corrected to the nearest 'most likely' allele. This error occurred due to the Saga software used to score gels. Genetic diversity indices are summarised in Tables 3 and 4 for marker loci and sampling populations, respectively.

Scoring of gels was 96.97% successful with 268 alleles not scored out of a total of 8826 (728 individuals at 12 alleles). Complete genotypes with all 12 alleles scored were established for 86.64% of individuals. Non-scored alleles were due to lack of PCR product. No significant linkage disequilibrium was evident for any locus pair in any of the eight samples after sequential Bonferroni correction (see Appendix 3). Significant departures from Hardy-Weinberg equilibrium were observed in 3/48 exact tests (Table 5).

Table 3: Genetic diversity indices for loci: k = number of alleles, H_O = observed heterozygosity, H_E = expected heterozygosity, Mean k = mean number of alleles per locus, P_A = proportion of private alleles (over all populations). T-test CI = confidence interval for t-test.

Locus	k	H_O	H_E	Allelic Richness	Mean k	P_A
3D7	11	0.70	0.66	5.47	5.88	0.0034
5D8	33	0.90	0.89	20.63	22.88	0.0014
3A2	8	0.59	0.60	4.31	4.50	0.0041
1B9	18	0.57	0.56	10.17	11.50	0.0014
2A52	11	0.54	0.53	4.19	4.88	0.0027
6C4B	8	0.68	0.71	5.83	6.38	0
Average	15	0.66	0.66	9.34	8.87	0.0022
T-test CI	11-19	0.60-0.71	0.61-0.72	6.44-12.24	6.19-11.56	0.0016-0.0028

Table 4: Genetic diversity indices for sample populations: H_O = observed heterozygosity, H_E = expected heterozygosity, Mean k = mean number of alleles per locus, P_A = proportion of private alleles. T-test CI = confidence intervals for t-test.

Population	Sample size (n)	H_O	H_E	Allelic Richness	Mean k	P_A
Stanton	100	0.67	0.67	9.00	10.00	0
Malin 1	99	0.64	0.66	8.40	9.50	0.0007
Malin 2	62	0.68	0.64	7.30	7.33	0
Galway 1	78	0.66	0.68	8.91	9.67	0.0041
Galway 2	86	0.65	0.66	8.21	8.83	0.0027
Kerry	99	0.71	0.67	8.61	9.83	0.0014
Wexford	99	0.64	0.68	8.81	9.83	0.0021
UK Brid.	106	0.62	0.63	8.22	9.67	0.0021
Average	91	0.66	0.66	8.43	9.33	0.0016
T-test CI	78-104	0.63-0.68	0.65-0.68	7.94-8.92	8.54-10.12	0.0004-0.0029

Table 5: P-values for Hardy-Weinberg exact tests. Bold indicates departure from equilibrium i.e. significant values following Bonferroni correction, P< 0.0012.

Population	3D7	5D8	3A2	1B9	2A52	6C4B
Stanton	**0.000**	0.173	0.436	0.043	0.526	0.706
Malin 1	0.274	0.127	0.001	0.396	0.618	0.101
Malin 2	0.223	0.052	0.619	0.570	0.069	0.039
Galway 1	0.301	0.935	0.123	0.913	0.108	0.413
Galway 2	0.142	0.722	0.003	0.064	0.104	0.631
Kerry	0.032	0.251	0.369	0.157	0.733	**0.000**
Wexford	0.093	0.031	0.007	0.559	**0.000**	0.095
UK Brid.	0.237	0.612	0.349	0.315	0.239	0.342

There was no evidence for population bottlenecks in any of the populations at any locus for either the SMM or the SMM/IAM models. Wilcoxon tests were not significant for heterozygote excess in any population, and the mode-shift reported in all populations was the standard L-shaped distribution, indicating that no recent reductions in alleles at low frequencies had occurred.

From the LOSITAN analysis it was found that none of the loci used were under selection.

Population Differentiation

Pairwise F_{ST} values are given in Table 6. The overall F_{ST} value was low at 0.007, but a 95% confidence interval of 0.005-0.009 showed that the value was significantly different from zero.

Based on F_{ST} permutation test p-values there is evidence of significant differentiation between: UK Bridlington and all other populations; Wexford and all populations except Malin 2; Malin 1 and all other populations except Stanton and Malin 2; and Malin 2 was significantly differentiated from Galway 1 and Galway 2.

F_{ST} permutation test p-values were also calculated for each locus individually (Appendix 4) to determine if any locus showed any departure from global F_{ST} results. Pairwise F_{ST} estimates differed for individual loci; for example, for comparisons between Malin 2 and all other populations significant F_{ST} estimates (0.016 – 0.042; Appendix 4) were observed for Cpag-3A2 . Because allelic richness and mean allele number at Cpag-3A2 were lower than the lower confidence limit of the *t*-test this locus was removed to test the effect of the locus in determining overall F_{ST} estimates. Of the other five loci, there were four significant estimates for Cpag-2A52 (Wexford versus Malin 2, Galway 1, Galway 2 and Kerry). Even removing Cpag-3A2 and Cpag-2A52, which also showed low diversity, resulted in no significant F_{ST} estimates between Malin 2 and any other population, except UK Bridlington (Appendix 4).

Figure 5 shows MDS plots for visualisation of spatial patterns. These plots are based on F_{ST} values and show proximity of sample sites based on the results in Table 6 (upper diagonal). The closer the points in the plot are, the more similar the populations.

Genotypic exact tests have a null hypothesis of genotypes being drawn from the same distribution in the two populations being tested, i.e. genotypic frequencies in the two populations are the same. A significant result here indicates that the pairs of populations being tested are likely to be distinct from each. In all, nineteen population pairs were significantly different after sequential Bonferroni correction: UK Bridlington was significantly different to all other populations; Wexford to all but Malin 2; Malin 1 to all but Malin 2; and Galway 1 was significantly different to Malin 1, Malin 2 and Galway 2 indicating that these populations' genotypes are not drawn from the same distribution and so may potentially be different populations.

Given that the overall F_{ST} value is significant, but low, it is not surprising to find 1/3 of population pairs to be potentially connected. The exact test results show a very similar trend. Indeed, almost identical results were obtained from pairwise F_{ST} analysis and exact tests (Table 6).

Results from the two sexes analysed separately do not give any further insight into structuring. Males and females both show non-significant F_{ST} (0.003, 0.004 respectively) and so are not considered as informative as the complete data set.

Using a regression analysis of pairwise $F_{ST}/(1-F_{ST})$ values, the dependent variable, against $\log_{10}$ geographical distance in km (see Appendix 4), the independent variable (Figures 6a and 6b) between populations highlights the effect of physical separation of populations, and so determines whether an isolation-by-distance model of gene flow is plausible (Sokal and Wartenberg, 1983). This model implements a distance-based barrier past which gene flow is not possible (a realistic model given the various geographic and hydrological barriers potentially affecting this study). Regression of $F_{ST}/(1-F_{ST})$ against $\log_{10}$ geographical distance did showed significant correlation ($R^2 = 0.205$, $P < 0.0357$) when all sample sets were included. Removing the UK Bridlington sample set resulted in a non-significant result ($R^2 = 0.105$, $P > 0.0476$), indicating that an isolation-by-distance model of gene flow between Irish brown crab populations is not likely.

Figure 5: Plots derived from multi-dimensional scaling analysis. The more proximal the points, the more connected the populations are inferred to be.

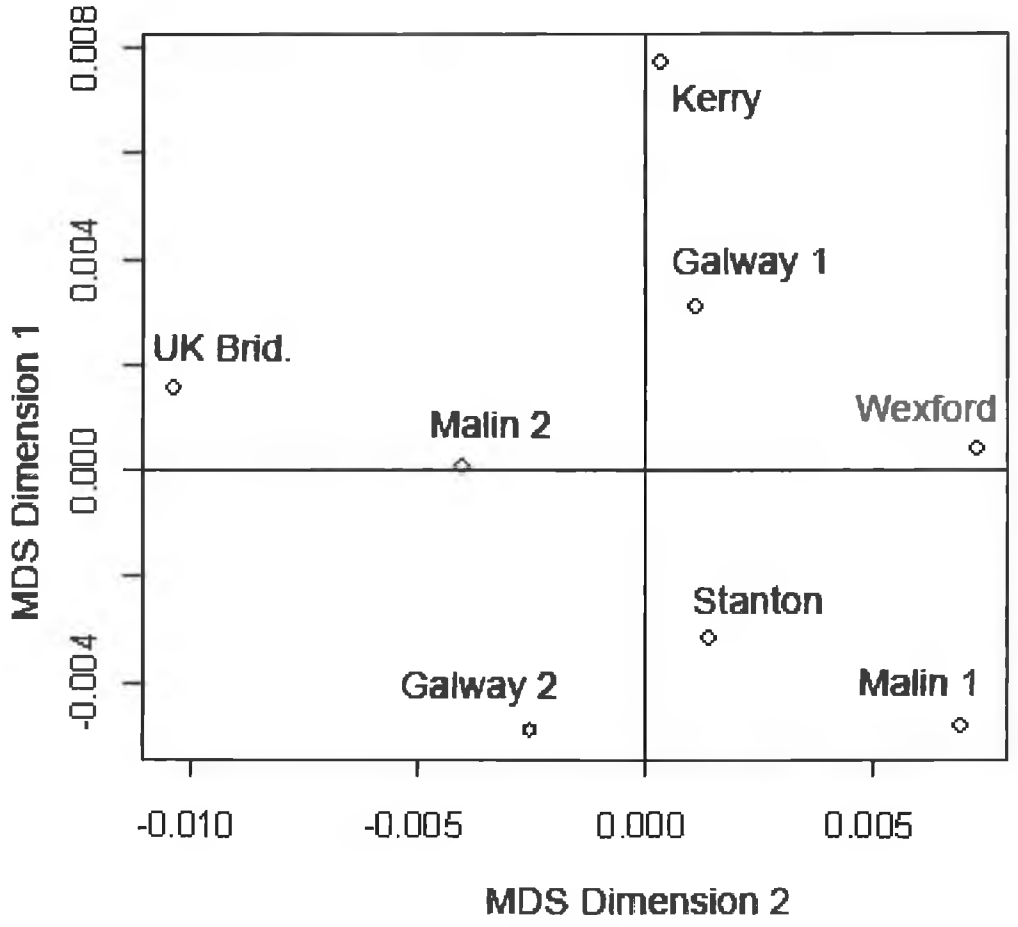


Figure 6a: Regression analysis of pairwise $F_{ST}/(1-F_{ST})$ values versus ln geographic distance in km. P-value for significance of $R^2 = 0.205$ at 5% level = 0.0155.

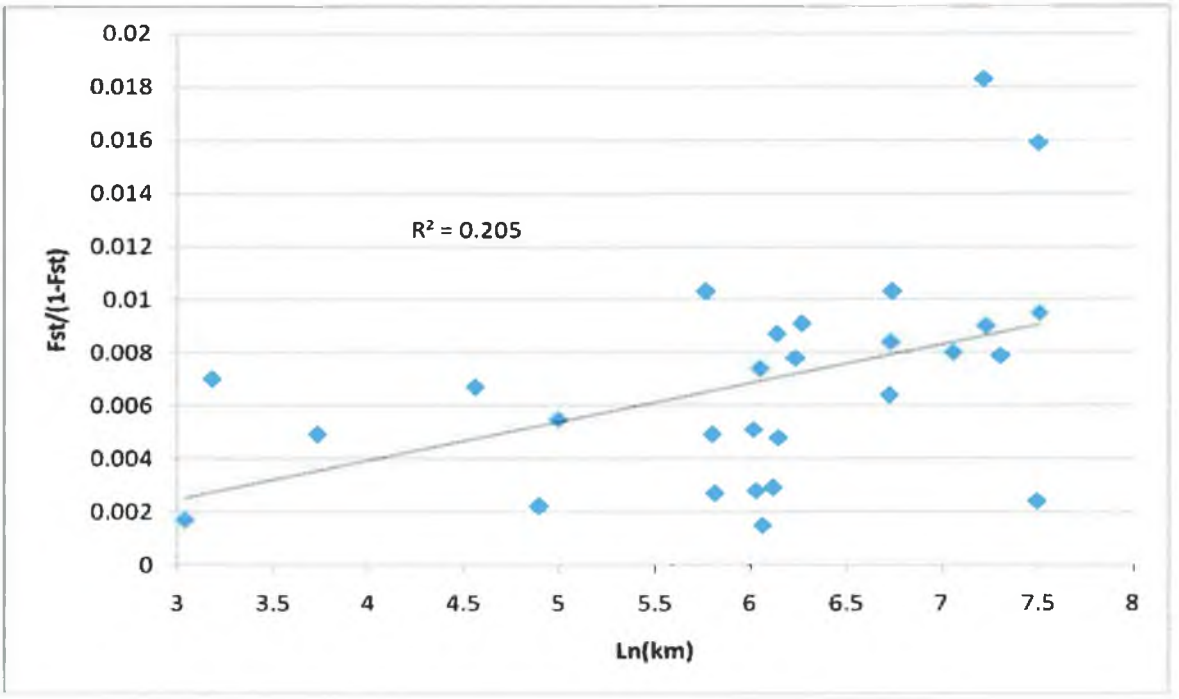


Figure 6b: Regression analysis of pairwise $F_{ST}/(1-F_{ST})$ values versus *ln* geographic distance in km for Irish samples. P-value for significance of $R^2 = 0.105$ at 5% level = 0.1515.

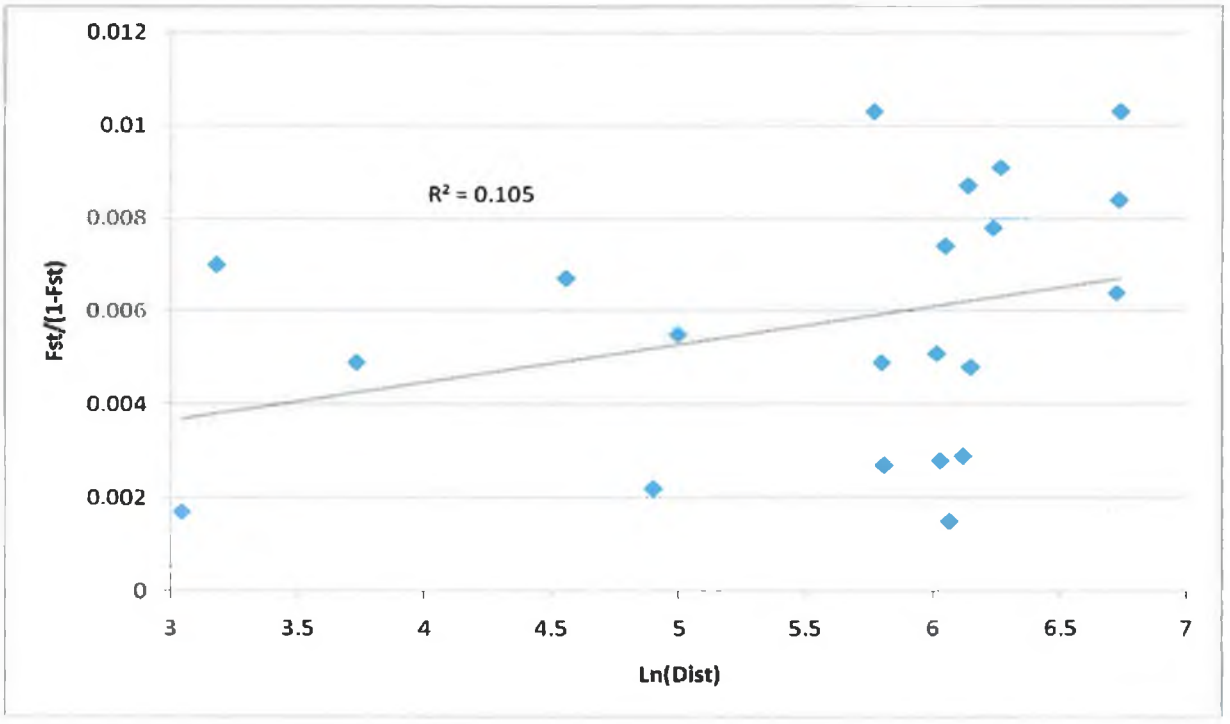
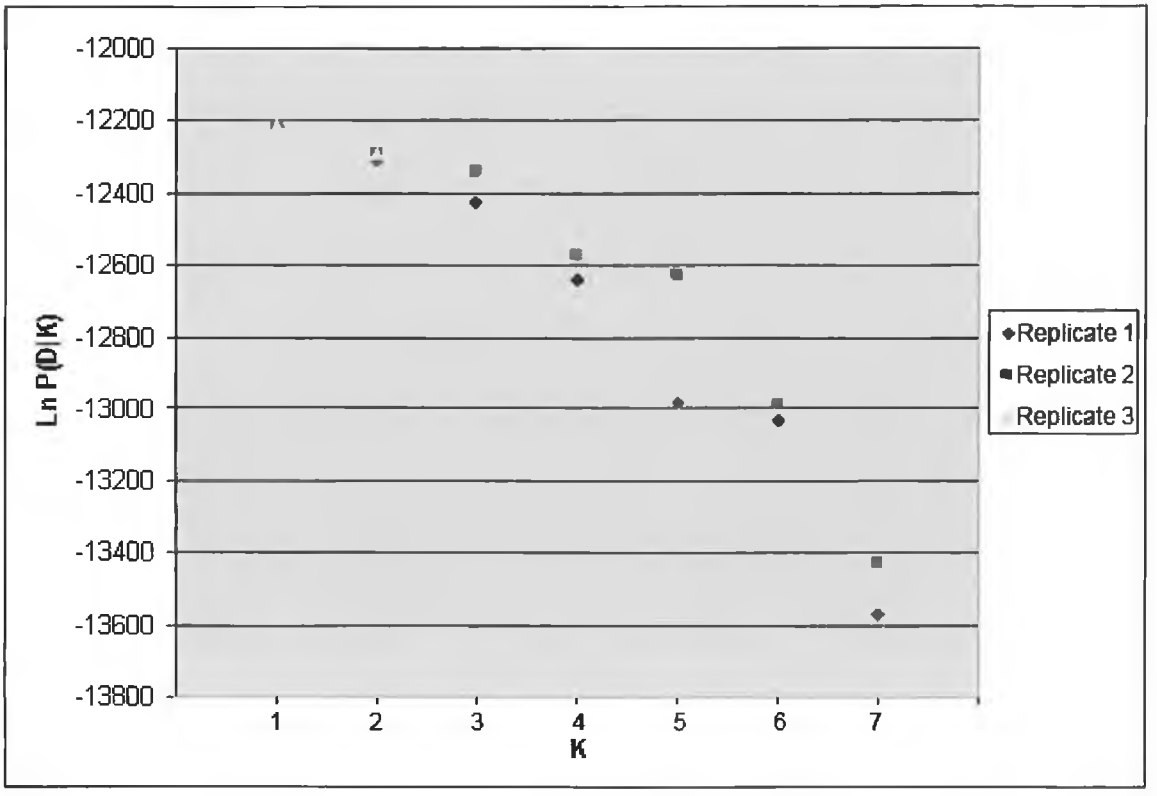


Table 6: Probabilities for Fisher's exact tests for pairwise genotypic differentiation (below the diagonal) and pairwise F_{ST} values (above the diagonal). Values for F_{ST} in bold are significant following the permutation test implemented in FSTAT which undergoes sequential Bonferroni correction ($P < 0.0018$); values for exact tests in bold are significant following sequential Bonferroni correction ($P < 0.0018$), highly significant values ($P < 1 \times 10^{-5}$) denoted as *.

	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford	UK Brid.
Stanton		0.0049	0.0017	0.0027	0.0015	0.0048	**0.0103**	**0.0095**
Malin 1	*		0.0070	**0.0049**	**0.0074**	**0.0087**	**0.0084**	**0.0159**
Malin 2	0.1560	0.0041		**0.0103**	**0.0028**	0.0029	0.0064	**0.0024**
Galway 1	0.3279	*	**0.0004**		**0.0067**	0.0022	**0.0078**	**0.0183**
Galway 2	0.1327	*	0.0066	**0.0009**		**0.0055**	**0.0091**	**0.0079**
Kerry	0.0038	*	0.0102	0.0956	*		**0.0051**	**0.0090**
Wexford	*	*	0.0063	**0.0003**	*	**0.0001**		**0.0080**
UK Brid	*	*	**0.0002**	*	*	*	*	

Analyses in both STRUCTURE and BAPS5 gave the most likely number of populations (from log likelihood values) to be K=1 which means that there is no evidence for substructure within the total sample set. Figure 7 shows log likelihoods of the probability of the data given each value of K from 1 to 7. This range of K values was used as higher values would take in the order of weeks to calculate. This is the definitive result from the analysis using STRUCTURE, tests being run at K = 1-7 giving a log likelihood value for each K value. These likelihoods depend on how well the data fit the K value in the model and so, when compared with one another, give the overall most likely value for K and so the most likely number of populations present.

Figure 7: Log likelihood values vs. K value (1-7). Most likely number of populations, K, indicated by highest Ln P(D|K), -12200 is for K = 1. Three replicates per K, runs of 100,000 burn-in repetitions followed by 500,000 MCMC repetitions. Ordinate shows log likelihood values, abscissa shows K values. Difference between replicates indicates level of variance.



The analysis from BAPS agreed with this *K* value. The result file from BAPS5 analysis only outputs the highest Ln P(D|K) value per run, which is not comparable to such values from STRUCTURE. The results do however confirm that no population structuring was evident using both programs,

In summary:

1. Bayesian analyses show no differentiation between populations, while results from F_{ST} analysis and exact tests did show evidence of significant genetic structure in Irish and UK populations.

2. From the F_{ST} analysis the UK sample differed significantly from all Irish samples.

3. In SE Ireland the Wexford sample differed significantly from all other samples, with the exception of Malin 2.

4. Along the west and north coasts of Ireland, the Stanton Bank sample showed genetic homogeneity with all other samples.

Discussion

Genetic Diversity in Populations

Diversity indices, represented by allelic richness and mean number of alleles per locus, show that over half of the polymorphism observed at the six microsatellite markers comes from two loci, Cpag-5D8 and Cpag-1B9 (Table 3). One population, Malin 2, showed a significantly lower diversity with an allelic richness and mean allele number of 7.3, 7.33, respectively; 95% confidence intervals from an independent two tailed *t*-test = 7.94–8.99, 8.44–10.33, respectively (Table 4). This is the only population which falls outside the CI.

Low diversity indices are typically an indication of a reduction in population size. This could be accounted for by a bottleneck which eliminates a component of the genetic diversity but which is not apparent in this study, or by inbreeding which reduces heterozygosity. Few heterozygote deficits were observed in tests for HWE and average heterozygosities for populations were all within the confidence intervals for *t*-tests (Table 4) from which it can be concluded that reduction in diversity is not likely to result from high levels of inbreeding (borne out in F_{ST} estimates). The low diversity may therefore be due to random sampling error. While the diversity at the loci used in this study may be low, the data set is not compromised by linkage disequilibrium or HW disequilibrium. F_{ST} and Bayesian analyses assume unlinked molecular markers that are in HWE and both of these assumptions are realised in this study and so results can be trusted.

The life history of brown crab does not suggest that inbreeding should occur over a small scale in the species. Brown crab are migratory and females hatch millions of eggs. A larval distribution study (Eaton *et al.* 2003) determined that larvae are in the water column for over two months, with post-larvae settling inshore (Robinson 1999). These factors should be important in terms of the degree of connectivity between populations as adult and larval distribution by directed and externally mediated movements would potentially increase levels of out-breeding between sample populations. The non-occurrence of an isolation-by-distance pattern highlights that this idea is probably accurate.

Population Genetic Structure from F_{ST} Analysis and Exact Tests

Overall findings from F_{ST} *analysis and exact tests*

Evidence for structuring between populations is supported from the analysis using F_{ST} analysis. Pairwise F_{ST} estimates were low (0.0015–0.0183; Hartl and Clark 1997) but 19/28 were significant following sequential Bonferroni correction. Results from exact tests were almost identical to those obtained from F_{ST} analysis. The exact test is considered as a more powerful test of genetic differentiation than F_{ST} analysis (Goudet *et al.* 1996). A significant result from a exact test means that the two populations being compared do not share the same genotypic frequency distribution (Goudet 2001). The F_{ST} analysis estimates the effect of population subdivision based on reduced heterozygosity caused by inbreeding due to non-connectivity of populations (Weir and Cockerham 1984). Thus the exact test and F_{ST} estimator use different methods to determine whether populations are subdivided, and their congruence here supports the overall validity of significant population differentiation.

Pairwise F_{ST} estimates for individual loci were wide ranging (0–0.04). Of the six loci two showed very distinct trends not seen in other loci: Cpag-3A2 showed significant estimates for Malin 2 with every other population; Cpag-2A52 showed significant estimates for Wexford with Malin 2, Galway 1, Galway 2 and Kerry (Appendix 4). Because selection is ruled out as a cause of these discrepancies Cpag-3A2 was removed from a subsequent analysis to determine its effect on global F_{ST} estimates which turned out to be negligible (not shown) as similar pairwise estimates were observed with or without this locus. Removing both Cpag-3A2 and 2A52 resulted in no differentiation between any populations except UK Bridlington versus all others (not shown). Due to the spurious nature of the Cpag-3A2 result in the Malin 2 population, along with the lower genetic diversity and sample size compared to other populations, the results from Malin 2 are considered distorted by this locus and therefore will not be discussed further.

Stanton Bank, the west coast and Malin Head

The majority of samples on the west and north coasts were genetically differentiated from each other. However, Stanton Bank was not significantly different from any other populations on the west coast down to and including Kerry, only being significantly different from Wexford and UK Bridlington. Because the rest of the samples on the west coast were significantly differentiated this result is difficult to explain. There is support for the idea that

migration north to, rather than south from, Stanton Bank caused this structuring. It is possible that coastal currents, discussed in detail in the next section, result in an influx of larvae to Stanton Bank. No tagging studies have investigated movement from south to north and so it is not possible to determine if such a direction of migration occurs.

In a study of brown crab in Swedish waters, based on two data sets, one tagging over 3500 individuals between 1968 and 1973, and another tagging over 8000 individuals in 2003, Ungfors *et al.* (2007) reported consistent, directed movement of females which was on average 2-8 times greater than males, with the majority of movements by both sexes <36 km over the entire range of times until recapture (>12 months). The mean rate of movement, measured in meters per day for individuals recaptured within the first month, was 325–345 in females and 202–299 in males. This figure declined by roughly half for females and by three quarters for males recaptured during the subsequent 2–6 months period. Females were considered to undertake directed migrations. Males were shown to move in both a directed and random manner, with an average distance of ~1 km until recapture. It is thus plausible that directed migrations by females over 10s of kilometres per month may result in gene flow between Stanton Bank and other populations on the west coast as tagging studies have found individuals, albeit very few, on the west coast that were tagged and released off Malin Head (M. Robinson pers. comm. November 2007). The splitting of the data set used in this study into male and female components revealed no further structuring to that seen when the whole data set was analysed, possibly due to the small male components of samples.

The three northern populations are the closest geographically in this study with distances between the two Malin samples and Stanton Bank being about 20–40 km. Tagging studies of over 8000 crab released from Malin Head reported by Tully *et al.* (2006) indicated that migrations occur in westerly, north westerly and southerly direction in Autumn, with counter-migrations easterly and north easterly also occurring. Recaptures over a three-year period have found migrants as far south as north Mayo, where the distribution of the 'northwest stock' is thought to end. Directed migration by adult (anecdotally most likely female) crabs almost certainly results in connectivity between local populations that are less than 50 km apart, and most likely explains the genetic similarity observed between the three northern sampling locations. The seclusion of Malin 1 in Lough Foyle may mean that it is isolated from migrants from the west coast and solely connected with Stanton Bank and Malin 2. However, if a counter migration back to Malin 1 was occurring from Stanton Bank then it might be expected that more of the west coast larvae/adults would be transported to Malin 1.

Galway and Kerry

Significant differentiation was observed between the two Galway samples, one collected within Galway Bay and the other in offshore waters. Given the potential for exchange of individuals through migrating adults and dispersing larvae and juveniles it seems highly unlikely that inshore brown crab are in some way isolated from offshore crab. However, in a study of brown crab focussing on the English Channel and North Sea, from pairwise F_{ST} estimates of 30 sample sites, Shaw and McKeown (unpublished data) found that two inshore populations at Newlyn, UK and Brest, France, were the only significantly differentiated samples. Beacham *et al.* (2008) also found one isolated and significantly differentiated population at Alison Sound, BC. It has also been shown that vertical migration of larvae in other decapods (dos Santos *et al.* 2008) may result in retention of larvae inshore. These examples give credibility to the idea that whilst it seems very unlikely that two populations so close to one another as Galway 1 and Galway 2 may not be in migratory contact, this may indeed be the case. That Galway 1 and Kerry are not significantly different but Galway 2 and Kerry are gives even more credence to the idea.

Grainger (1980) presented a circulatory pattern of Galway Bay which suggests an inflow at the South Sound (south of Inisheer, the smallest of the Aran Islands) and an outflow at the North Sound. The depth given for the bay is 36m, with a step down to 91m west of the Aran Islands. It is stated in the paper that there is a very low probability of (herring) larvae remaining in the Bay due to short 'flushing-times' (20 – 67 days in August and March respectively). Whether this is applicable to crab larvae is unknown as the depths at which they reside is unclassified.

One further issue with the Galway samples that must be dealt with is that the Galway 2 sample is composed entirely of males, which may mean that they may be less likely to move outside of the Bay. That said, they will still represent the local population because of the female contribution to ~50% of their genome. From the F_{ST} analysis conducted on the male only data set the small sample sizes (12 – 27 individuals) in all but the Galway 2 sample may have resulted in a large component of variation being missed which may elucidate the issue.

UK Bridlington and Wexford

There is significant differentiation between the UK and Irish samples. Previous work at this geographic scale has been conducted on brown crab using mitochondrial DNA (P. Shaw and N. McKeown, unpublished data) where it was found that three geographical groups exist, one in the north and east of the North Sea, one in the south of the North Sea, and one in the English Channel and Celtic Sea. A significant difference between North Sea and English Channel samples was found, which supports the differentiation found in this study between UK Bridlington and all Irish populations. McKeown (pers. comm. 2008) believes a post-glacial expansion from northern (north/east North Sea) and southern (Channel/Celtic Sea) refugia, with subsequent slow mixing of the two, has resulted in the patterns seen with mitochondrial DNA. However, their research only included three Irish populations (Wexford, Kerry and Stanton Bank, from the same DNA as those used in this study) of which Wexford and Kerry were both found to be in the Channel/Celtic Sea group, with Stanton Bank not located in any of their three groupings.

In Ireland, this Wexford sample was significantly differentiated from all of the samples along the west and north coasts with the exception of (the now discounted) Malin 2. This conflicts with the results of the mitochondrial DNA study referred to above, which grouped Wexford and Kerry together. Because Shaw and McKeown's study used 30 individuals per sample it is possible that less diversity was apparent then was found in the study presented here. The sequence used in Shaw and McKeown's work, a 765 bp sequence in cytochrome oxidase (COI 1), was far less polymorphic than the markers used here with nine related haplotypes found, and thus the Wexford sample may not be in current contact with Kerry but may have come from the same southern refuge. It is likely that the F_{ST} estimates presented here give a more accurate view of population structuring because of the larger sample sizes and concordant F_{ST} and exact test results.

Failure of Bayesian Methods

There are two plausible reasons for the failure of the Bayesian analyses seen in this study (Pritchard *et al.* 2007, help file of STRUCTURE 2.2). Firstly, data supporting an isolation-by-distance model lead to STRUCTURE having difficulty in determining clusters of individuals. This is because genotypes of individuals under such a model tend to represent more than one

population. The data here do not support an isolation by distance model but sample sites are relatively geographically close together, and subsequently individuals sampled may have "mixed membership in multiple groups". The second and more likely reason is that there may not be enough data to allow STRUCTURE to differentiate the various populations. As discussed, a low level of polymorphism was evident at several loci used in the study. The Bayesian methods use the allele frequency distributions in the data, from which models are constructed. Likelihoods are then calculated for these models, which include the *K* value (number of clusters/populations) being assessed (Pritchard *et al.* 2000). Low numbers of alleles (range 4.5-6.4) at four of the six loci are likely to confound the analysis as their distributions will appear very similar. This may also be true of highly polymorphic loci that may have very similar allelic frequencies, but this scenario is theoretically less likely if random mating and little migration are evident. The failure of the Bayesian methods can be construed as the data not showing sufficient difference between the sampled populations. This is supported by the analysis failing to recognise the UK Bridlington population, included for the very reason that it is geographically very distant and thus unlikely to receive migrants from Irish populations. Puebla *et al.* (2008) found a replicated clustering pattern in their data. Three populations from Greenland were clustered together, with the remaining 10 Canadian populations being considered by the analysis as one cluster. The authors believed that genetic population breaks can occur over a 'sea-scale' (i.e. between Canada and Greenland). This is akin to the geographical distance between the UK Bridlington population and the west of Ireland populations (both ~1000km) and so it could be concluded that a larger data set (in terms of loci) may well allow the program to perform efficiently.

Impact of Coastal Currents on Genetic Structure

Puebla *et al.* (2008) hypothesise that in snow crab larval dispersal is the primary method by which gene flow occurs due to the long pelagic larval phase of 3-5 months. The exact duration of the larval phase in brown crab is not known, but it is believed to be about two months long, between July and September (Eaton *et al.* 2003). As gene flow through migration is one of the primary factors influencing population genetic structuring, the most obvious impact upon this study, aside from directed migration of adults, are the currents operating during the time of year when pelagic larvae are in the water column and larval dynamics within the water column (i.e. vertical distribution). Fernand *et al.* (2006) discussed

the potential for transport via currents along the west coast of Ireland and found that during the summer months "fronts and jet-like flows extend continuously along the western seaboard of Ireland" (Figure 8). The paper highlighted the impact that this may have on plankton and larvae, calling it a "rapid transport mechanism". These largely baroclinic flows are narrow, about 10–20 km in width, with relatively little influence from wind (13–25%), and are continuous north along the western seaboard. Similarly, Brown *et al* (2003) reported a jet-like circulation in the southeast of Ireland, where the Celtic Sea and St. George's Channel meet, during the summer months into late autumn, that results in a current running parallel to the coast in a south westerly direction (see Figure 8). Fernand *et al.* (2006) reported that coupling with this flow could allow transport from the English Channel to the Irish north coast.

In terms of connectivity of populations, as indicated by F_{ST} estimates and exact test results, it seems possible that transport by the current reported by Fernand *et al.* (2006) could result in a movement of larvae from south/southwest Ireland as far north as the Stanton Bank site. Thus larvae in the water column, originating from any population, could be transported northward. What is not certain is the numbers of larvae which would remain at points along this migrationary route, or whether other (localised) currents moving in different directions would result in other directions of transport. It is difficult to test the effect of the larval transport hypothesis without sampling larvae throughout the summer months and determining their origin. The result of this northerly transportation of larvae could be a genetic representation of populations from the southwest/west in the northerly samples, as seen at Stanton Bank. The exchange of few migrants per generation is sufficient to maintain genetic similarity between populations by disrupting the effects of genetic drift and inbreeding (Wang 2004).

The effect of vertical migrations within the water column of larvae may result in a greater or lesser extent of migration based on where within the water column the larvae are found, and what environmental factors influence this (dos Santos *et al.* 2008). A study of larval decapod vertical distribution in a bay in central Chile (Yannicelli *et al.* 2006), comparable to Galway Bay in size, found that species vertical depth could be equated to inflow/outflow to/from the bay, and that the age and species of larvae were determining factors in vertical distribution. Larval distribution studies have not been conducted in *C. pagurus* and so it is not possible to determine the effect of vertical distribution on the potential for transport. The genetic structuring found in the Galway 1 and Galway 2 samples would make for an interesting case study if coupled with vertical migration data as the bay

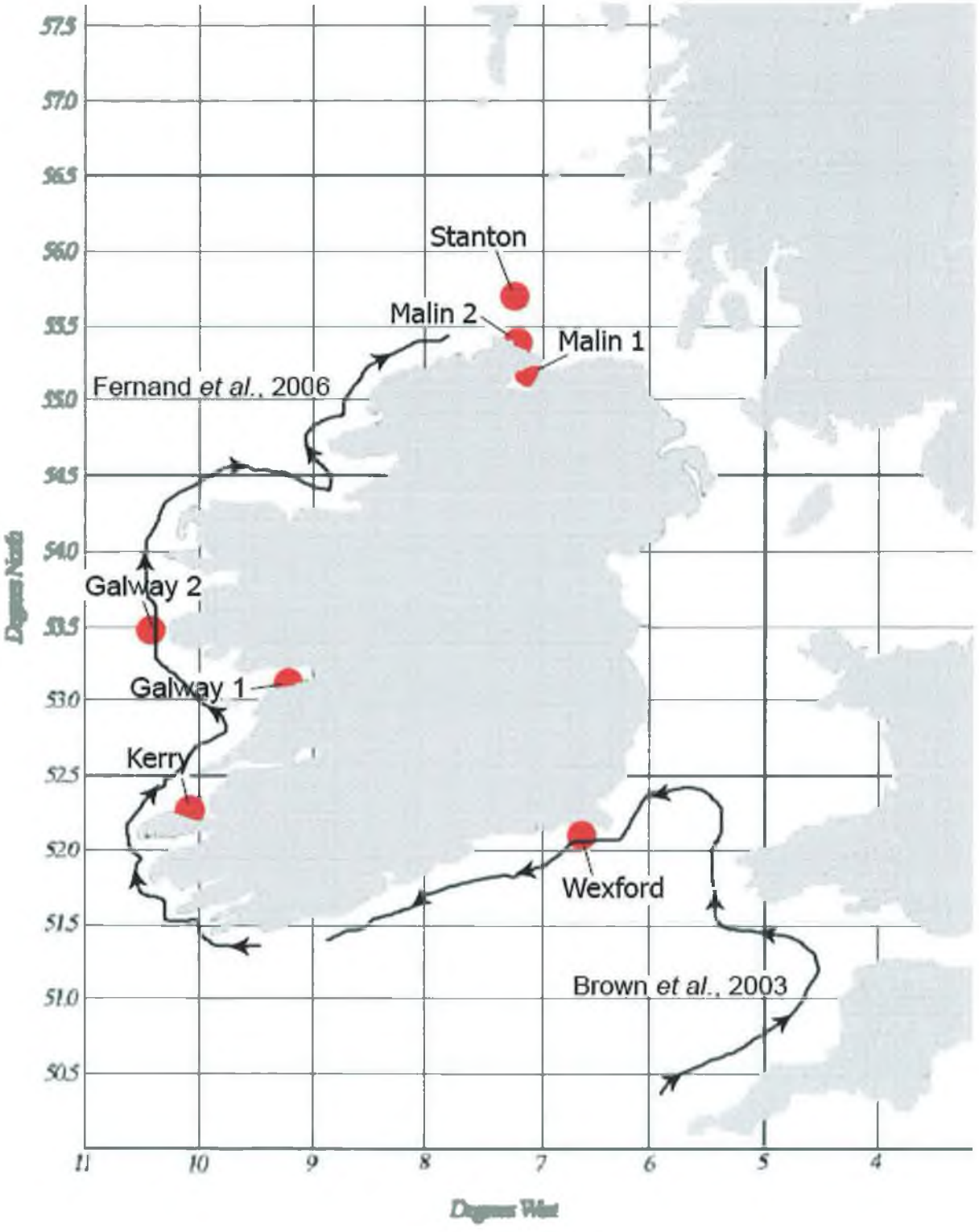
has a defined circulatory pattern and the results from this study suggest that populations in the bay may have restricted gene flow with the other (well connected) Irish populations.

Further to the issue of vertical migration, it is difficult to assess where the front proposed by Fernand *et al.* (2006) begins and ends relative to the coastline, as the continental shelf and the boundary between the saline Atlantic and fresher coastal water, in which the fronts occur, run at different distances to the coastline. Assuming that pelagic larvae are transported by currents 'offshore', considered to be between 5-18km by Fernand *et al.* (2006), it is possible that larvae being moved closer to the coast than 5km would not be caught in such a passive migration and so inshore areas could receive migrants. However, it is also possible that in more secluded areas less larvae could be moved inshore resulting in more genetically distinct and isolated populations. Again, this is difficult to determine without actual larval sampling and enough data from molecular markers to assign individuals to a population of origin (Waples 2004) but it could be one explanation for the genetic differentiation seen between Galway 1 and Galway 2 populations, which are geographically close but may not be in significant migratory contact.

Whilst the northerly migration route by summer currents is speculative it does explain the connectivity between Stanton Bank and Galway 1, Galway 2 and Kerry, which is the most difficult result to resolve provided the Malin 2 result is considered as an anomaly in the data. The connection of Stanton Bank with all other populations down the west coast may be more likely in terms of a northern current during the summer, to which southwest and westerly populations (starting with the Kerry population) would all contribute larvae which could move northwards ending at Stanton Bank. Under this hypothesis larvae from Galway 2 may be released into this current and be carried northwards to Stanton Bank, but not to other northern populations. Again larval sampling and assignment of larvae to a population of origin would be required to test this hypothesis.

Migration of adults is known to occur from Malin Head to Stanton Bank (Tully *et al.* 2006) but movement in the opposite direction is far less frequent (M. Robinson pers. comm. August 2008). If a general directed migration of adults occurs northwards along the west coast it is not unreasonable to expect them to end up at Stanton Bank if individuals are migrating there from Malin Head at this time at any rate.

Figure 8: Currents and directions reported in Brown *et al.* 2003 and Fernand *et al.* 2006. Both studies were conducted during summer months when brown crab larvae are in the water column.



In conclusion there is definite connection of crab populations around the Irish coast. Populations as far apart as Stanton Bank and Kerry appear to be in migratory contact. However, two close populations, Galway 1 and Galway 2, appear not to be in such contact. If an offshore coastal current running the length of the west coast occurs during the summer months, when pelagic larvae are in the water column then this mechanism of transport may allow the connectivity between sample sites that is apparent. This is dependent on multiple factors such as larvae dynamics within the water column and other environmental factors, but owing to the life history of the brown crab it is likely that enough larvae could be distributed along the coast to affect the genetic composition of local populations such that inbreeding, and hence local differentiation, does not occur, as is evident in this study.

References

Anon (2007) The Stock Book. Report to the Minister for Agriculture, Fisheries and Food Annual Review of Fish Stocks in 2007 with Management Advice for 2008. The Marine Institute, November 2007.

Arnheim, N. and Erlich, H. (1992). Polymerase chain reaction strategy. *Annual Review of Biochemistry* **61**, 131-165.

Beacham, T.D., Supernault J. and Miller K.M. (2008). Population structure of Dungeness crab (*Cancer magister*) in British Columbia. *Journal of Shellfish Research* **27(4)**, 901–906.

Beaumont, M.A. and Nichols, R.A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceeding of the Royal Society of London B, Biological Sciences* **263**, 1619–1626.

Beaumont M.A. and Rannala B. (2004). The Bayesian revolution in genetics. *Nature Reviews Genetics* **5**, 251–261.

Begg, G.A., Friedland, K.D. and Pearce, J.B. (1999). Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries Research* **43**, 1–8

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (1996-2004).GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France). Available at: http://www.univ-montp2.fr/~genetix/genetix/constr.htm#download

Booke, H.E. (1999). The stock concept revisited: perspectives on its history in fisheries. *Fisheries Research* **43**, 9-11.

Brown, J., Carillo, L., Fernand, L., Horsburgh, K.J., Hill, A.E., Young, E.F., 2003. Observations of the physical structure and seasonal jet-like circulation of the Celtic Sea and St. George's Channel of the Irish Sea. *Continental Shelf Research* **23**, 533–561.

Cadrin, S.X., Friedland K.D. and Waldman, J. (eds.) (2005). *Stock Identification Methods: applications in fishery science*. Elsevier Academic Press, Amsterdam.

Corander, J., Waldmann, P. and Sillanpää, M.J. (2003) Bayesian analysis of genetic differentiation between populations. *Genetics* **163**, 367-374.

Cornuet, J.M. and Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001–2014.

Coyle, T. (1998). Stock identification and fisheries management: the importance of using several methods in a stock identification study. In: *Taking Stock: defining and managing shared resources*. (Ed. Hancock, D.A.). Australian Society for Fishery Biology, Sydney pp. 173-182.

Crab Fisheries Management and Conservation Regulations (2005), SI 676. Available at: www.dcmnr.ie

CSO (2007). *Fishery Statistics 2003 and 2004*. Ref. 60/2007, ISSN 1393-5674.

Darwin, C. (1859). *On the origin of species*. John Murray, London.

dos Santos, A., Santos A.M.P, Conway, D.V.P., Bartilloti, C., Lourenço, P. and Quieroga, H (2008). Diel vertical migration of decapod larvae in the Portuguese coastal upwelling ecosystem: implications for offshore transport. *Marine Ecology Progress Series* **359**, 171–183.

Eaton, D.R., Brown, J., Addison, J.T., Milligan, S.P. and Fernand, L.J. (2003). Edible crab (*Cancer pagurus*) larvae surveys off the east coast of England: implications for stock structure. *Fisheries Research* **65**, 191–199.

Edwards, E (1979). *The edible crab and its fishery in British waters*. Fishing News Books, Farnham, Surrey.

Fahy, E., Hickey, J., Perella, N., Hervas, A., Carroll, J. and Andray, C. (2004). Bionomics of brown crab *Cancer pagurus* in the southeast Ireland inshore fishery. *Irish Fisheries Investigations* **12**, 30pp.

Falconer, D.S. and Mackay, T.H. (1996). *Introduction to quantitative genetics (4th ed.)*. Longmans Green, Harlow, Essex.

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

Fernand, L.J., Nolan, G.D., Raine, R., Chambers, C.E., Dye, S.R., White, M. and Brown, J. (2006). The Irish coastal current: A seasonal jet-like circulation. *Continental Shelf Research* **26**, 1775–1793.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004) *Bayesian data analysis* (2nd ed.). Chapman & Hall/CRC, Boca Raton.

Glaubitz, J.C. (2004). Convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes* **4**, 309–310.

Goudet, J. (1995). FSTAT (vers. 1.2): a computer program to calculate F-statistics. *The Journal of Heredity* **86**, 485-486.

Goudet, J. University of Lausanne; Lausanne: 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available at: http://www.unil.ch/dee/page6767_en.html.

Goudet, J., Raymond, M., de Meeus, T. and Rousset, R. (1996). Testing differentiation in diploid populations. *Genetics* **144**, 1933–1940.

Goudet, J., Perrin, N. and Waser, P. (2002). Tests for sex-biased dispersal using bi-parentally inherited genetic markers. *Molecular Ecology* **11**, 1103–1114.

Guo, S. W. and Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361–372.

Hartl, D. L. and Clark, A.G. (1997). *Principles of Population Genetics*. (3rd ed.) Sinauer Associates, Sunderland, Massachusetts.

Hedrick, P.W. (1999). *Genetics of Populations* (2nd ed.). Jones and Bartlett, Sudbury, Massachusetts.

Hilborn, R. and Walters, C.J. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*. Springer.

Hudson, R.R. (1998). Island models and the coalescent process. *Molecular Ecology* **7**, 413–418.

Huang, W.G., Cracknell, A.P., Vaughan, R.A. and Davies, P.A. (1991). A satellite and field view of the Irish shelf fronts. *Continental Shelf Research* **11**, 543-562.

ICES (2006). Report of the Study Group on the Biology and Life History of Crabs (SGCRAB) by Correspondence. ICES CM 2006/LRC:13.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**, 624–626.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.

King, M (1995). *Fisheries biology, assessment and management*. Fishing News Books, Oxford.

McKeown, N.J. and Shaw, P.W. (2007). Polymorphic nuclear microsatellite loci for studies of brown crab, *Cancer pagurus* L. *Molecular Ecology Notes* **8**, 653–655.

Mayr, E (1963) *Animal Species and Evolution*. Belknap Press (Harvard University), Cambridge, MA.

Moran, M.D. (2003). Arguments for rejecting the sequential Bonferroni in ecologicalStudies. *OIKOS* **100(2)**, 403–405.

Nakagawa, S (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology* **15**, 1044–1045.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590.

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

Pritchard, J. K., Wen, X. and Falush, D (2007). Documentation for structure software: Version 2.2. Available at: http://pritch.bsd.uchicago.edu/software

Puebla, O., Sévigny, J-M., Sainte-Marie, B., Brêthes, J-C., Burmeister, A., Dawe, E.G. and Moriyasu, M. (2008). Population genetic structure of the snow crab (*Chionoecetes opilio*) at the Northwest Atlantic scale. *Canadian Journal of Fisheries and Aquatic Science* **65**, 425–436.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Raymond M., and Rousset F. 1995. Genepop (version 1.2), population genetics software for exact tests and ecumenicism. *The Journal of Heredity* **86**, 248–249.

Rice, W.R. (1989). Analysing table of statistical tests. *Evolution* **43(1)**, 223–225.

Robinson, M. 1999. *Community structure and recruitment of decapods in shallow sub-littoral habitats*. Ph.D Thesis, University of Dublin, 149pp.

Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Notes* **8(1)**, 103–106.

Sambrook, J.; Fritsch, E.F. and Maniatis, T. *Molecular Cloning: a laboratory manual.* (2nd ed.). Cold Spring Harbor Laboratory Press, N.Y.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365–371.

Sokal, R.R. and Wartenberg, D. (1983). A test of spatial autocorrelation using an isolation-by-distance model. *Genetics* **105**, 219–237.

Tully, O., Robinson, M., O'Keefe, E., Cosgrove, R., Doyle, O. and Lehane, B. (2006). The Brown Crab (Cancer pagurus L.) Fishery: Analysis of the resource in 2004 - 2005. *Fisheries Resource Series*, Bord Iascaigh Mhara (Irish Sea Fisheries Board), Dun Laoghaire, Ireland Vol. 4, 48pp.

Ungfors, A., Halback, H. and Nilsson, P.G. (2007). Movement of adult edible crab (Cancer pagurus L.) at the Swedish West Coast by mark-recapture and acoustic tracking. *Fisheries Research* **84**, 345–357.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. and Shipley, P.F. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535–538.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S (Fourth Ed.)*. Springer, New York. ISBN 0-387-95457-0

Wang, J. (2004). Application of the One-Migrant-per-Generation Rule to Conservation and Management. *Conservation Biology* **18(2)**, 332–343.

Waples, R.S. (1998). Separating the wheat from the chaff: Patterns of genetic Differentiation in high gene flow species. *The Journal of Heredity* **89(5)**, 438–450.

Weir, B.S. and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* **15**, 323-354.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: FST doesn't equal 1/(4Nm+1). *Heredity* **82**, 117–125.

Yannicelli, B., Castro, L.R., Valle-Levinson, A., Atkinson, L. and Figueroa, D. (2005). *Journal of Plankton Research* **28**, 19–37.

URL1:http://www.bim.ie/uploads/text_content/docs/Map%20of%20Crab%20distribution%20around%20the%20Irish%20coast.html

Appendix 1

Table A1: Cross calibration information for UK Bridlington sample (McKeown, pers. comm. 2007) and Irish samples.

Locus	Change in basepairs (bp)
Cpag-3D7	No change
Cpag-5D8	+7 bp on UK samples (e.g. 176 bp UK = 183 bp IRL)
Cpag-3A2	+9 bp on UK samples (e.g. 248 bp UK = 257 bp IRL)
Cpag-1B9	+8 bp on UK samples (e.g. 246 bp UK = 254 bp IRL, EXCEPT 238UK which = 238IRL)
Cpag-2A52	No change
Cpag-6C4B	No change

Appendix 2

Geographic points between sample sites to allow for determination of in-water shortest-straight-line distance between sample sites used in the isolation-by-distance analysis. Points were used to connect sample sites e.g. point (iv) was used to calculate distance from Galway 1 to Galway 2 and Galway 2 to Kerry; points (v), (vi), (vii) were used to calculate distance from Kerry to Wexford. All distances reported are approximate.

Table A2: Points between sample sites used for determination of distances between sample sites. Points (i) – (vii) can be seen in Figure A1 below. Points (viii), (ix) and (x) are found in the English Channel and south North Sea and are used to calculate distance from UK Bridlington site to the Irish sites.

(i)	55.39N	6.85W
(ii)	55.28N	8.27W
(iii)	54.25N	10.2W
(iv)	53.16N	9.9W
(v)	52.3N	10.7W
(vi)	51.62N	10.45W
(vii)	51.3N	9.5W
(viii)	49.5N	5.8W
(ix)	50.95N	1.43E
(x)	52.7N	1.8E

Figure A1: Map of Ireland showing sampling locations (red dots) of sites from Table 1. UK Bridlington not shown. Points between sampling sites (black dots (i) – (vii)) in Table A2 were used to allow determination of shortest in-water distance between sample sites. All points are approximate.

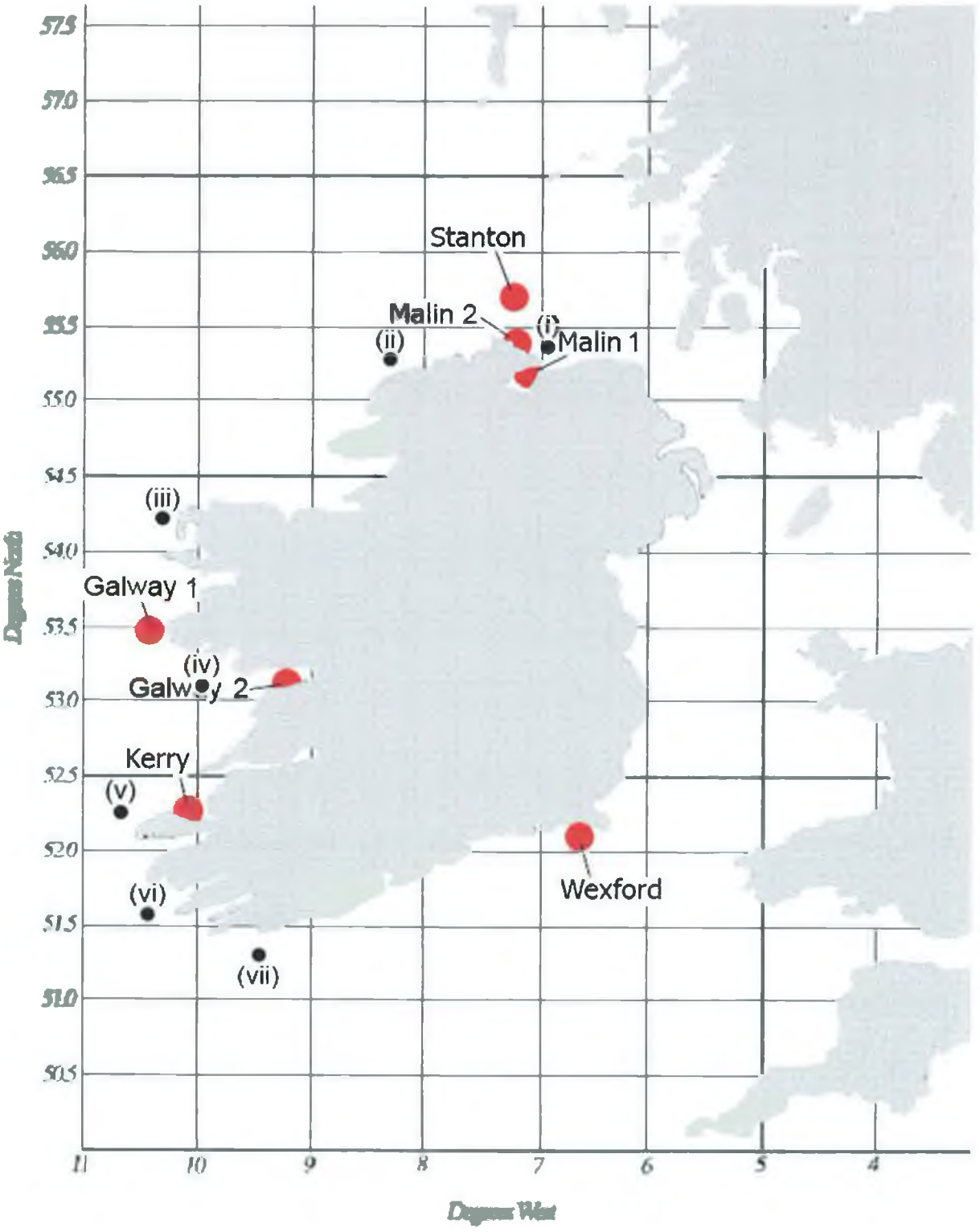


Table A3: Distances between sample sites based on data in Table 1 and points (i) – (x) shown above. Below the diagonal shows kilometres (km), above the diagonal shows $\log_{10}$ km.

	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford	UK Brid.
Stanton		3.729	3.040	5.810	6.062	6.147	6.738	7.504
Malin 1	41.7		3.178	5.766	6.028	6.117	6.721	7.496
Malin 2	20.9	24.0		5.796	6.051	6.138	6.733	7.501
Galway 1	333.6	319.4	329.1		4.559	4.897	6.235	7.215
Galway 2	429.1	414.9	424.6	95.5		4.996	6.266	7.303
Kerry	467.5	453.3	463.0	133.9	147.9		6.014	7.230
Wexford	844.1	829.9	839.6	510.5	526.6	409.3		7.055
UK Brid.	1814.8	1800.6	1810.3	1359.6	1484.8	1379.9	1158.1	

Appendix 3

Table A4: P-values for genotypic (linkage) disequilibrium, based on 2400 permutations. Adjusted p-value for 5% nominal level following Bonferroni correction = 0.000417. No significant pairwise tests.

Locus	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford	UK Brid.
3D7x5D8	0.6892	0.0192	0.1296	0.4029	0.9600	0.5163	0.0946	0.6092
3D7x3A2	0.3233	0.1042	0.8638	0.5942	0.3879	0.2117	0.0671	0.0050
3D7x1B9	0.9104	0.3633	0.3229	0.1233	0.4392	0.1146	0.3513	0.1938
3D7x2A52	0.9229	0.3246	0.4446	0.5479	0.6813	0.6738	0.4342	0.9129
3D7x6C4B	0.8017	0.0054	0.9679	0.0708	0.9233	0.6921	0.8554	0.7338
5D8x3A2	0.7721	0.4721	0.1238	0.4633	0.7471	0.8233	0.0642	0.7029
5D8x1B9	0.5017	0.4425	0.8192	0.9546	0.6383	0.1979	0.2258	0.5179
5D8x2A52	0.9629	0.4596	0.5800	0.5729	0.3942	0.7708	0.5938	0.2808
5D8x6C4B	0.7092	0.5788	0.6767	0.3750	0.6942	0.3004	0.0608	0.1113
3A2x1B9	0.2629	0.0400	0.5483	0.5283	0.3663	0.5479	0.0646	0.6679
3A2x2A52	0.3333	0.2483	0.5021	0.0433	0.2992	0.8354	0.6754	0.8129
3A2x6C4B	0.0004	0.1804	0.1208	0.6054	0.4129	0.0563	0.3404	0.0188
1B9x2A52	0.9263	0.5242	0.8954	0.8967	0.2200	0.1017	0.8033	0.2254
1B9x6C4B	0.2683	0.2392	0.2542	0.1100	0.1392	0.3438	0.0771	0.6921
2A52x6C4B	0.6042	0.8138	0.5817	0.3296	0.3150	0.7475	0.2121	0.0313

Appendix 4

Tables A5–10: Pairwise F_{ST} estimates for individual loci from GENEPOP 4.0.7 analysis following the standard ANOVA method of Weir and Cockerham (1984). Values for F_{ST} in bold are significant following the permutation test implemented in FSTAT 2.9.3, which uses sequential Bonferroni correction of $P < 0.0018$.

Locus:	Cpag-3D7						
	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford
Malin 1	-0.0064						
Malin 2	0.0028	-0.0010					
Galway 1	-0.0005	-0.0036	-0.0043				
Galway 2	0.0068	0.0029	-0.0046	0			
Kerry	0.0047	0.0006	-0.0035	-0.0052	0.0012		
Wexford	0.0097	0.0053	-0.0040	0.0011	-0.0056	0.0014	
UK Brid.	**0.0082**	0.0076	**0.0119**	**0.0176**	0.0073	**0.0237**	0.0106

Locus:	Cpag-5D8						
	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford
Malin 1	0.0010						
Malin 2	-0.0011	0.0033					
Galway 1	-0.0007	0.0008	**0.0081**				
Galway 2	0.0036	0.0030	0.0068	0.0117			
Kerry	0.0019	0	0.0062	-0.0019	0.0122		
Wexford	0.0020	0.0001	0.0010	0.0086	0.0031	0.0089	
UK Brid.	0.0042	0.0054	**0.0083**	**0.0146**	**0.0081**	**0.0135**	**0.0071**

Locus: Cpag-3A2

	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford
Malin 1	-0.0034						
Malin 2	**0.0380**	**0.0217**					
Galway 1	0.0118	0.0028	**0.0156**				
Galway 2	-0.0008	-0.0008	**0.0338**	0.0023			
Kerry	-0.0013	-0.0015	**0.0336**	0.0028	-0.0049		
Wexford	0.0027	-0.0029	**0.0216**	-0.0019	-0.0009	-0.0008	
UK Brid.	-0.0043	-0.0014	**0.0423**	0.0115	-0.0030	-0.0029	0.0034

Locus: Cpag-1B9

	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford
Malin 1	0.0057						
Malin 2	0.0008	-0.0022					
Galway 1	-0.0048	0.0138	0.0055				
Galway 2	-0.0048	0.0023	0.0005	-0.0017			
Kerry	0.0013	-0.0016	-0.0042	0.0057	0.0019		
Wexford	0.0173	0.0013	0.0015	0.0251	0.0169	0.0013	
UK Brid.	0.0283	0.0003	0.0094	**0.0398**	0.0231	0.0101	0.0031

Locus: Cpag-2A52

	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford
Malin 1	0.0030						
Malin 2	-0.0050	0.0052					
Galway 1	0.0172	0.0138	0.0493				
Galway 2	-0.0039	-0.0042	-0.0029	0.0304			
Kerry	-0.0029	0.0113	-0.0036	0.0064	0.0016		
Wexford	0.0257	0.0253	**0.0400**	**0.0144**	**0.0327**	**0.0161**	
UK Brid.	-0.0043	-0.0002	-0.0037	0.0201	-0.0045	-0.0020	0.0212

Locus: Cpag-6C4B

	Stanton	Malin 1	Malin 2	Galway 1	Galway 2	Kerry	Wexford
Malin 1	0.0110						
Malin 2	-0.0032	0.0161					
Galway 1	-0.0017	0.0001	0.0029				
Galway 2	0.0047	0.0059	0.0145	-0.0002			
Kerry	**0.0218**	0.0019	**0.0299**	0.0080	**0.0162**		
Wexford	0.0091	-0.0033	0.0112	0.0008	0.0135	0.0022	
UK Brid.	**0.0245**	-0.0044	**0.0304**	0.0105	0.0155	**0.0055**	0.0022

Acknowledgements

I would like to express my gratitude first and foremost to Dr. Elizabeth Gosling who has been an integral part of each stage of this project and without whom this research would never have been undertaken in the first place. I would like to thank Dr. Mary Hayes and Dr. Martin Robinson for their help in coordinating sampling and discussions in the early stages of the project. I would like to thank Dr. Niall McKeown of Royal Holloway, London, for the primer information and general help which he gave, advice on PCR conditions and specifically his help with resolving PCR issues. I would like to thank Dr. Anna Was for extending her knowledge of molecular DNA techniques to me. I would like to also thank Brendan Allen, Dave Tully, Ed Helps, Roisin Ní Bhriain, Ciara Ní Chualain and Rebecca O'Connor for providing me with samples, and to the fishermen for allowing access to their boats and fishing grounds. I would finally like to thank everyone working in the MERG laboratory, particularly Sandra Doherty and Carmel O'Connor for their unfaltering good humour and outlook.