TASK 6

Genetic differentiation and genetic barcoding of squid paralarvae

Différenciation génétique et code-barre génétique pour paralarves

Objective/*Objectifs*:

1. Development of DNA barcoding methods for identification of early life stages of squid to species.
2. Identification of field samples of squid eggs / larvae to species using genetic methods.
3. Analysis of genetic differentiation among spawning populations of cuttlefish (*Sepia officinalis*).

1. Développement de méthodes de codes-barres génétiques afin de relier à leur espèce des calmars aux premiers stades de leur vie
2. Identification d'échantillons d'œufs et de larves de calmar grâce à la méthode génétique
3. Analyse de la différence génétique au sein de populations de seiches (*Sepia officinalis*) frayant

2009-2012 Work progress / Etat d'avancé du Travail

I. Task Results Summary / *Résumé des résultats obtenus*

i. DNA based assays were developed for species identification of early life history stages of squid. The accuracy of methods was confirmed by analysis of reference specimens. The assays are inexpensive and accurate DNA methods for rapid and high throughout analyses of species identity that can be readily extended to further cepahlopod species.

ii. The DNA species identification assays were successfully applied to field caught samples of paralarvae and provided unambiguous typing in all cases.

iii. New genetic markers (microsatellite loci) were developed and represent powerful DNA tools for future studies of cuttlefish (*Sepia officinalis*).

iv. Population genetic analysis of *S. officinalis* indicated that :

- Cuttlefish in the Channel belong to a single genetic population
- There is high genetic connectivity between populations in the Channel and Bay of Biscay
- Sweepstakes recruitment processes, i.e. a small number of breeders contributing to recruitment, are occuring in the Golfe du Morbihan and warrant further study.

i. Des analyses basées sur l'ADN ont été développées afin de pouvoir identifier l'espèce de calmars en début de cycle de vie.
La justesse des méthodes employées a été confirmée grâce à l'analyse de spécimens de référence. Ces analyses sont peu onéreuses et correspondent à des méthodes ADN précises pour une analyse rapide et extensive de l'identité d'une espèce, et peuvent facilement être étendues à d'autres espèces de céphalopodes.

ii. Les analyses ADN d'identification d'espèce ont été appliquées avec succès sur des échantillons de paralarves, et ont donné des résultats sans ambigüité dans tous les cas.

iii. De nouveaux marqueurs génétiques (microsatellites loci) ont été développés and représentent des outils ADN puissants pour de futures études sur les seiches (*Sepia officinalis*).

iv. Les analises génétiques de populations de *S. officinalis* indiquent que :

- Les seiches de Manche appartiennent à une seule population génétique
- Il y a une forte connexité génétique entre les populations de la Manche et du Golfe de Gascogne
- Il y a une certaine « lotterie » dans le recrutement, c'est à dire qu'un petit nombre de géniteurs contribue au recrutement dans le Golfe du Morbihan, ce qui justifierait une étude plus poussée.

TASK 6

II. Actions carried out during the project/ *Bilan des opérations réalisées pendant le projet*

Task 6.1: Development of DNA barcoding methods for identification of early life stages of squid to species.

- Available sequence data for the four target species were downloaded from public databases.
- New sequence data collected for *A. media* samples from Moray Firth (Scotland) - sequence to be deposited in public database (GenBank) with publication 2 described below.
- DNA sequences aligned and putative universal primers (for PCR-RFLP method) and species-specific primers (for PCR-species-specific primer method) designed.
- Primers tested.
- Restriction enzymes predicted to produce species-specific cleavage patterns identified from available sequence data.
- Species specific PCR and PCR-RFLP assays evaluated on adult specimens previously identified to species.
- Performance of both methods for species identification of early life history stages assessed by comparative analysis of preserved paralarvae.

Tâche 6.1 : Développement de méthodes de codes-barres génétiques pour relier à leur espèce des calmars aux premiers stades de leur vie

- ★ les séquences de données disponibles pour les quatres espèces cibles ont été prises à partir de bases de données publiques
- ★ une nouvelle séquence de données pour des échantillons de *A. media* a été collectée à Moray Firth (Ecosse). Séquence qui sera déposée dans une base de données publique (GenBank) avec la publication 2 décrite ci-dessous.
- ★ Des séquences ADN ont été alignées, et des amorces supposées universelles (pour une méthode PCR-RFLP) ainsi que des amorces spécifiques aux espèces ont été conçues.
- ★ Amorces testées
- ★ Prévision des enzymes de restriction afin de produire des schémas de clivage spécifiques à l'espèce identifiés à partir des séquences de données disponibles.
- ★ Analyses PCR et PCR-RFLP spécifiques à l'espèce menées sur des spécimens adultes identifiés à l'espèce au préalable.
- ★ Evaluation par analyse comparative des paralarves conservées des résultats de ces deux méthodes pour l'identification des premiers stades de la vie des espèces.

Task 6.2: Identification of field samples of squid eggs / larvae to species using genetic methods.

- DNA extracted from 38 paralarvae collected during JUVECEPH (2008) sampling cruise.
- PCR-RFLP species assay performed.
- All 38 paralarvae identified as *Loligo vulgaris*.
- Genetic typing confirmed by DNA sequencing of Cytochrome oxidase I gene fragment and subsequent sequence comparisons using BLAST searches (sequences to be deposited in GenBank with publication 2 described below).
- DNA extracted from 96 paralarvae collected from Seine Bay (2011 cruise).
- PCR-RFLP species assay performed.
- 95 individuals identified as *A.subulata* with the remaining individual identified as *L.vulgaris.*

TASK 6

Tâche 6.2 : Identification d'échantillons d'œufs et de larves de calmar grâce à la méthode génétique

- ★ Extraction d'ADN des 38 paralarves collectées pendant la campagne JUVECEPH (2008)
- ★ Analyses PCR-RFLP effectuées
- ★ Chacune des 38 paralarves identifiées en tant que *Loligo vulgaris*.
- ★ Empreinte génétique confirmée par séquencage ADN du fragment de gene cytochrome oxydase I et comparaisons de séquence en découlant à l'aide de recherches BLAST (séquences devant être déposées dans GenBank avec la publication 2 décrite ci-dessous).
- ★ Extraction ADN de 96 paralarves collectées dans la Baie de Seine (campagne 2011)
- ★ Analyse PCR-RFLP effectuées
- ★ 95 individus identifiés en tant que *A.subulata,* et un individu identifié en tant que *L.vulgaris.*

Task 6.3: Analysis of genetic differentiation among spawning populations of cuttlefish (Sepia officinalis).

- Fourteen novel microsatellite DNA loci were developed for *S. officinalis* using a previously optimised protocol involving: (i) genomic library construction (cloning of target species DNA); (ii) selective hybridisation of DNA probes; (iii) DNA sequencing of clones; (iv) PCR primer design.

- DNA extracted from samples of spawning individuals collected in 2009 from three UK sites (Torbay, Selsey, Eastbourne).

- New microsatellite loci tested on subset of individuals from UK sites.

- Statistical simulations performed to assess Type I and Type II error rates in inferences of population structure under various data conditions (loci number and sample sizes).

- Using statistical simulations the optimal combination of loci were selected for genetic analysis of samples (i.e. the number of loci for conferring lowest Type I and II error rates for the predicted sample sizes to be employed).

- UK spawning samples genotyped at 9 loci.

- Genetic diversity statistically analysed (Hardy-Weinberg equilibrium, pairwise differentiation (exact tests, *Fst*), clustering and assignment tests).

- DNA extracted from spawning samples collected from French locations of Agon-Coutainville and Seine Bay.

- French spawning samples genotyped at 10 loci.

- Genetic diversity among UK and French spawning samples statistically analysed.

- DNA extracted from egg samples collected from 4 sites: Torbay (UK), Selsey (UK), Agon Coutanville (France) and Seine Bay (France).

- Egg samples analysed at 9 microsatellite loci.

- Egg sample genetic variation statistically assessed as for other samples.

TASK 6

- Based on initial results, kinship patterns within egg samples were investigated using maximum likelihood methods to identify related (full- and half-sib) individuals. Related individuals were removed in subsequent studies of population differences.

- DNA extracted from replicate samples of spawning individuals collected in Eastbourne and Torbay (2010), and a sample of overwintering adults from the central western channel (Casquets).

- DNA extracted from two samples collected in the Bay of Biscay: (i) Sene – North Biscay, (ii) Arcachon – South Biscay.

- Replicate samples and western Channel / Biscay samples genotyped at 9 loci.

- Genetic diversity statistically analysed.

- Genetic patterns compared among individual loci.

- Deficits of heterozygotes identified at three loci which may be due to technical artefacts known as null alleles that may bias estimates of population structure.

- Analyses repeated using various combinations of loci (excluding various combinations of loci with suspected null alleles) and additional statistical methods of null allele frequency inference.

Tâche 6.3 : Analyse de la différence génétique au sein de populations de seiches (Sepia officinalis) frayant

★ Quatorze nouveaux loci ADN microsatellites ont été développés pour les *S. officinalis* grâce à un protocole déjà optimisé : (i) construction d'une base de génomes (clonage ADN d'espèces cibles), (ii) hybridation sélective de sondes ADN, (iii) séquencage ADN de clones, (iv) conception d'amorce PCR

★ ADN extrait d'échantillons d'individus frayant collectés en 2009 dans trois sites britanniques (Torbay, Sesley, Eastbourne).

★ Nouveaux microsatellites loci testés sur un sous-ensemble d'individus des sites britanniques

★ Simulations statistiques effectuées afin d'évaluer les taux d'erreurs Type I et Type II dans les suggestions de structures de population sous diverses conditions de données (nombre loci et tailles des échantillons).

★ A l'aide de simulations statistiques la combinaison optimale de loci a été selectionnée pour une analyse génétique des échantillons (c'est-à-dire le nombre de loci nécessaire pour un taux d'erreur type I et II le plus bas pour que les tailles d'échantillons prévus soient utilisés)

★ Échantillons britanniques genotypés à 9 loci

★ Diversité génétique analysée au niveau statistique (équilibre Hardy-Weinberg, differentiation deux à deux (tests exacts, Fst), tests de regroupement et de répartition)

★ ADN extrait des échantillons collectés en France à Agon-Coutainville et en Baie de Seine

★ Echantillons français genotypés à 10 loci

- ★ Diversité génétique des échantillons français et anglais analysées au niveau statistique
- ★ ADN extrait d'échantillons d'oeufs collectés dans 4 sites : Torbay (RU), Sesley (RU), Agon-Countainville (France) et Baie de Seine (France)
- ★ Echantillons d'œufs analysés à 9 loci microsatellites
- ★ Variation génétique des oeufs échantillonnés évaluée statistiquement comme pour les autres échantillons
- ★ A partir des résultats initiaux, les liens de parenté des œufs échantillonnés ont été recherchés à l'aide de méthodes de vraissemblance maximale afin d'identifier des individus liés (frères et demi-frères). Les individus liés furent écartés des études de différences de populations suivantes.
- ★ ADN extrait d'échantillons réplicats/copiés d'individus collectés à Eastbourne et Torbay (en 2010), et d'un échantillon d'adultes hivernant au centre de la Manche-ouest (Les Casquets)
- ★ ADN extrait de deux échantillons collectés dans le Golfe de Gascogne, à Sene, au nord, et à Arcachon, au sud.
- ★ Échantillons réplicats ainsi que ceux de la Manche ouest et du Golfe de Gascogne génoytpés à 9 loci.
- ★ Diversité génétique analysée au niveau statistique
- ★ Schémas/modèles génétiques comparés au sein de chaque loci.
- ★ Déficit d'hétérozygotes identifié à 3 loci, peut être dû à un effet indésirable appelé allèles nuls qui peut biaiser les estimations d'une structure de population.
- ★ Analyses répétées à l'aide de différentes combinaisons de loci (excluant plusieurs combinaisons de loci avec des alleles nuls supspectés) et méthodes statistiques additionnelles afin de déterminer la cause des allèls nuls.

III. Scientific report / *Détail des travaux scientifiques*

Outputs

1 .Task 6 Progress reports for years 1 and 2, and Final (application) report (see below).

2. Articles in preparation for peer reviewed journals (*journal name*):

(i) Identification and characterisation of dinucleotide microsatellite markers for the common cuttlefish, *Sepia officinalis* L. (*Conservation Genetics Resources*)
(ii) Novel genetic methods for squid species identification: application to paralarvae reveals unexpected species ratios. *(ICES Journal of Marine Sciences*)
(iii) Population genetic structuring of the common cuttlefish (*Sepia officinalis* L.) in the English Channel and Bay of Biscay. (*ICES Journal of Marine Sciences*)

3. DNA sequences of COI gene of *Alloteuthis media* and microsatellites cloned from *Sepia officinalis* submitted to public database (GenBank): in line with standard procedures these sequences will be released to public access when the associated publications are in press.

CRESH Task 6 (Genetics) Final (application) report.

Executive Summary

Task 6.1: Development of DNA barcoding methods for identification of early life stages of squid to species.

Two DNA based methods (i. species-specific PCR; ii. species-specific PCR-RFLP) for species identification were developed for unambiguous genetic identification of individuals of *Loligo vulgaris*, *L. forbesi* and *Alloteuthis*. The species specific PCR involves the use of specially designed primer combinations that produce positive reactions only in the presence of DNA from a given species. The PCR-RFLP method involves the cutting of DNA with particular enzymes which produce species diagnostic patterns. The accuracy of both assays was initially demonstrated by analysis of known reference specimens. The assays were then tested on early life history stages. While both methods conferred 100% accuracy in positive species identification of paralarvae, the PCR-RFLP method was reported to be more robust (i.e. less reaction failures).
Both techniques represent inexpensive and accurate DNA methods for rapid and high throughout analyses of species identity. While DNA methods are applicable to a range of substrates (i.e. DNA extracted from blood, tissue, foodstuffs) it is suggested that the PCR-RFLP method be preferentially used where DNA quality may not be optimal. The PCR-RFLP method can also be readily extended to other cephalopod species through judicious selection of cutting enzymes that may be identified from analysis of documented species DNA sequences using publicly accessible databases (e.g. GenBank), or by sequencing of previously undocumented species.

Task 6.2: Identification of field samples of squid eggs / larvae to species using genetic methods.

The species identity of 38 whole paralarvae collected as part of the JUVECEPH 2008 cruise were investigated using both morphological and genetic (PCR-RFLP of sample DNA) species identification protocols. Genetic methods identified all individuals as *L. vulgaris*, which conflicted in a number of cases with inferences based on morphology. The validity of the genetic designations was subsequently confirmed by DNA sequencing. Such comparative analyses will be useful in future attempts to refine morphological keys for species identification.
The PCR-RFLP species assay was applied to test the species identify of 96 paralarvae collected from Seine Bay as part of the JUVECEPH 2011 cruise. Despite *a priori* prediction that *L. vulgaris* would predominate in the sample, only one individual was identified as *L. vulgaris* with the remaining individuals (n = 95) classified as *Alloteuthis subulata*. The results highlight the utility of the genetic species identification methods for apportioning species ratios, which may improve accuracy of spawning maps, stock assessments, and studies of ecological interactions.

Task 6.3: Analysis of genetic differentiation among spawning populations of cuttlefish (*Sepia officinalis*).

As a step beyond the original project plans, an additional 11 microsatellite loci were developed for *S. officinalis*, to complement previously existing loci. Simulation analyses indicated that the extra loci conferred a considerable increase in statistical power (>30%) to detect population structuring, and will benefit future studies of cuttlefish.

Based on results of simulations the most informative combination of loci (n = 9) were identified for subsequent genetic testing of population structure: (i) within the Channel; and (ii) between the Channel and Bay of Biscay.
Excluding possible null allele effects, tests of spatial and temporal genetic structuring indicated no significant differences among adult samples with the exception of the Sene (Golfe du Morbihan) Biscay sample which was differentiated from all other samples. Genetic indices indicate that the differentiation of Sene is most likely due to sweepstake recruitment processes, and thus unlikely to reflect cross-generational population isolation.
Genetic analysis of eggs represents a powerful approach for studies of population structure in *S. officinalis*. However, in this study the genotyping of eggs revealed large 'family effects' (i.e. samples consisted of genetically related individuals from a small number of families) precluding rigorous and unbiased population level analysis. Future studies should reduce the risk of family effects by increasing the number of spatial / temporal sampling points of eggs.

In conclusion the population genetic data indicate that:

(i) Channel cuttlefish belong to a single genetic population
(ii) Extensive dispersal and interbreeding between Biscay and the Channel
(iii) Sweepstake recruitment processes may be important on short timescales within sites such as the Golfe du Morbihan.

Detailed report

Task 6.1 Development of DNA barcoding methods for identification of early life stages of squid to species

Introduction

Aquatic species identifications have traditionally been performed using external morphological and anatomical characters. However, the accuracy of such phenotypic based approaches may be limited by considerable variation within species and/or small differences between species, as well as circumstances which may remove informative traits such as fishery processing, digestion, etc. Phenotypic based species identification of early life stages (eggs and larvae) is even more complicated than for adults as these stages often provide few consistent taxonomically important characters (e.g. Strauss & Bond 1990). DNA-based species identification methods offer an analytically powerful addition, or even an alternative, to phenotypic based methods and are increasingly being applied to taxonomic, ecological and forensic studies.

The development and application of genetic species identification methods to marine taxa has, to date, largely focused on a limited number of groups (e.g. Gadoid fish). At present such resources are limited for cephalopods, but are urgently needed given the fundamental importance of correct species identification to informing conservation and management goals. The objective of the present research

was to develop genetic species identification methods that could be routinely applied to the identification of individuals from three ecologically and economically important squid taxa with overlapping distributions: *Loligo vulgaris*, *L. forbesi* and *Alloteuthis spp.* A second objective was to empirically test the utility of these methods for species identification of early life history stages.

Materials & Methods

All mitochondrial DNA (mtDNA) sequences for the three taxa available on GenBank were downloaded and homologous sequences aligned using Bioedit (Hall 1999). The cytochrome oxidase I (COI) gene was selected as a marker for species identification. Based on COI sequence variation two different types of assay were developed: (1) Species specific PCR (Polymerase Chain Reaction amplification of specific gene regions); (2) PCR-RFLP (PCR combined with Restriction Fragment Length Polymorphism).

Species specific PCR: This assay utilises species specific primers which generate a product only in the presence of DNA from a given species. Using the aligned sequences putative species specific primer pairs for each of the three taxa were designed by eye. A secondary consideration in the design of such primers was the production of distinctive fragment sizes which could permit typing by means of multiplex PCR (i.e. single tube assay). Primer pairs were initially tested separately using a range of chemical reaction stringencies, specifically combinations of a range of $MgCl_2$ concentrations (1.0 to 3.0mM) and annealing temperatures (45 ^{o}C - 60 ^{o}C). After identifying primer pairs with robust species specificity the efficiency of a multiplex PCR assay was then evaluated.

PCR-RFLP: This assay utilises restriction enzymes which produce species specific cleavage patterns of the DNA. Based on the DNA sequence alignment PCR primers were designed from conserved regions of sequence to permit amplification of a homologous gene fragment in all species, which could then be cut using restriction enzymes. Sequences for each taxon were analysed using NEBcutter 2.0 (Vincze et al. 2003) to identify restriction enzyme cleavage sites present. Patterns of cleavage site presence were then compared across taxa to elucidate potentially species-diagnostic enzyme combinations for RFLP analysis. Restriction digests were performed following manufacturers (New England Biolabs) recommendations.

Reference specimens for optimisation and testing of the assays consisted of adult samples from each of the taxa collected throughout their range and which had been reliably morphologically identified. For the reference specimens DNA was extracted from ethanol preserved tissues using a CTAB-phenol/chloroform protocol. The performance of both assays for species assignment of early life history stages was empirically assessed using DNA extracted with a Chelex-proteinase K method from 38 paralarvae sampled during the JUVCEPH 2008 cruise. In all cases DNA fragments were resolved on 2% agarose gels and visualised using Gel red staining.

Results

TASK 6

Species specific PCR: Three primer pairs were identified which resulted in 100% correct species identification among reference individuals through the combination of: (i) positive reaction products in their target taxa (Fig. 1); (ii) no cross amplification in the two other taxa for any conditions (reaction stringencies) assayed (Fig. 1). Optimal positive amplification was achieved using 10µl total reaction volumes consisting of 1X PCR Buffer, 2mM MgCl2, 200 µM of each dNTP, 0.2 units of Taq (BioLine), 0.2 µM each primer, and 2 µl of sample DNA template. Amplifications involved an initial denaturation step (95 °C for 3 min) followed by 35 cycles of 30 s at 95 °C, 30 s at 55 °C and 30 s at 72 °C. As the species specific products were markedly different in size the three primers pairs were then combined in a multiplex PCR using the aforementioned reaction conditions. Multplex PCR analysis of reference individuals followed the individual PCR results in conferring 100% correct species identification.

PCR-RFLP: The primers CRESH-F (5'-gagcaggcttagttggtacttc-3') and CRESH-R (5'-atggctccagctaacacagg-3') permitted PCR amplification of a 544bp fragment of the COI gene in all individuals. PCR's were performed in 25µl total reaction volumes consisting of 1X PCR Buffer, 2mM MgCl2, 200 µM of each dNTP, 0.2 units of Taq (BioLine), 0.5 µM of each primer, and 2 µl of DNA template. Amplifications involved an initial denaturation step (95 °C for 3 min) followed by 35 cycles of 30 s at 95 °C, 30 s at 52 °C and 30 s at 72 °C. Based on sequence comparisons the combination of *Sfc*I and *Bcc*I restriction enzymes was predicted to reciprocally differentiate the three groups by a minimum of two site differences (Table 1). RFLP analysis of reference individuals produced expected fragment sizes and conferred 100% discrimination of the three groups. The enzyme *Bcc*I was also reported to distinguish between *A. subulata* and *A. media* (Table 1).

Upon application of both methods to DNA extracted from paralarve it was necessary to increase the number of PCR cycles to 55 in each case. While the species specific PCR method did not misidentify any individuals it did exhibit a large number (~30%) of false negatives (i.e. no PCR products produced for any primer pair) meaning that those individuals could not be identified to type. The PCR-RFLP method correctly identified all paralarvae to type.

Discussion

This work provides the first DNA based markers to distinguish *L. vulgaris*, *L. forbesi* and *Alloteuthis.* The methods are also, to our knowledge, the first for any cephalopod that does not rely on *de novo* DNA sequencing. This removal of a costly DNA sequencing step increases the potential for routine high throughput application. Although both methods are based on a small number of site differences, analysis of reference individuals from a wide geographical range did not indicate any ambiguity due to intraspecific polymorphisms and supported the specificity of both methods.

The markers are expected to be particularly useful for studies of early life history stages. Comparative analysis of both methods indicated that the PCR-RFLP method outperformed the species specific PCR method, which produced a large number of uninformative false negatives. This high rate of

false negatives is undoubtedly due to the respective primers being less efficient in PCR compared to the PCR-RFLP primers, a feature linked to the greater restrictions imposed on the sequences of the species specific primers. While additional species specific primers could be assayed in the future the PCR-RFLP method represents the more suitable approach to studies of early life history stages at present. Furthermore, PCR-RFLP method is expected to be readily applicable to other cephalopods though the judicious selection of restriction enzymes.

Task 6.2 Identification of field samples of squid eggs / larvae to species using genetic methods.

Introduction

This work reports the first application of the genetic species identification assays to field-caught samples of paralarvae. Two samples of ethanol-preserved paralarave were provided for analysis by project partners:

Sample 1: 38 paralarvae collected during the JUVECEPH 2008 survey.

Sample 2: 96 paralarvae collected during the JUVECEPH 2011 survey in Seine Bay.

Methods

Sample 1 was used in the testing and optimisation of the genetic species identification assays described in Task 6.1. This testing reported the PCR-RFLP method to outperform the species specific PCR method in analyses of earl life history stages. All paralarvae were therefore genotyped using the PCR-RFLP method following the protocol described in Task 6.1.

Results

Unambiguous genotypes were obtained for all samples (i.e. there were no reaction failures or unrecognised genotypes).

Sample 1: All individuals were identified as *L. vulgaris*.

Sample 2: 95 individuals were identified as *A. subulata*, one individual was identified as *L. vulgaris*.

Discussion

Morphological analysis of sample 1 prior to the genetic analysis identified a number of individuals as *L. forbesi*. The subsequernt PCR-RFLP classification of all individuals as *L. vulgaris* therefore represents an obvious discordance between genetic and phenotypic methods. However, the validity of the genetic designations was subsequently confirmed by DNA sequencing of the COI gene (using the PCR primers) for each of the samples with conflicting results, highlighting the accuracy of the PCR-RFLP assay and the difficulties of phenotypic based species identification of early life history stages in cephalopods. Such

comparative genetic and phenotypic analyses may be useful in future attempts to improve accuracy of phenotypic species identification methods.

The species ratio reported for sample 2 was very surprising as fishery information suggested, *a priori*, that most individuals would be *L. vulgaris*. Inaccurate species identification of early life history stages has been reported to severely compromise estimates of stock biomass in other commercially important taxa (e.g. Cod, Fox *et al.* 2008). Ideally, any species identification method should be capable of positively identifying close to 100% of all individuals sampled. The PCR-RFLP assay produced no null reactions demonstrating its robustness across varying DNA templates (DNA quality / quantity). Furthermore, as described in Task 6.1, the assay can readily be extended to identify other spatially / temporally overlapping cephalopod species. The results of this pilot study demonstrate how application of the PCR-RFLP assay is likely to produce new, and perhaps surprising, insights into the ecology of cephalopod species.

Task 6.3 Analysis of genetic differentiation among spawning populations of cuttlefish (*Sepia officinalis*).

Introduction

Populations constitute interbreeding units with more or less autonomous dynamics and recruitment and are frequently defined as harvest stocks in relation to fisheries management (Carvalho & Hauser 1994). In terrestrial and freshwater environments, populations are often well delimited by conspicuous physical barriers to mixing and interbreeding (e.g. Avise 2000). However, in the marine environment distinct populations are more difficult to detect and for many marine species it is unclear to what degree distinct populations exist at all, or whether they are organised into larger panmictic units (McQuinn 1997). This distinction is critical as different populations may possess different genetic, physiological, behavioural or other characteristics that may cause differences in life history traits such as fecundity and mortality rates, and ultimately production and abundance (Gold & Richardson 1998). As recruitment and sustainability may be properties specific to individual populations failure to identify, and independently manage, distinct populations can lead to local overfishing and ultimately to severe declines or population collapse (Hutchings 2000; Knutsen et al., 2003). This problem may be especially acute in systems with diverse and potentially locally adapted migratory components that overlap spatially and seasonally.

Genetic markers represent powerful tools for examining population structure that in the marine environment might otherwise be undetected due to difficulties in implementing standard ecological methods such as mark-recapture or behavioural observation (Shaklee & Bentzen 1998). Furthermore, genetic markers provide a unique ability to distinguish between non-reproductive dispersal and

effective dispersal (i.e. those individuals that survive and breed in the new population). Populations that are not linked by effective dispersal (i.e. interbreeding) may accrue different gene frequencies. Therefore, by characterising the geographical distribution of genetic variation population units can be identified and the processes defining such populations inferred.

Genetic studies of marine taxa have revealed that population structuring can be shaped by processes that are intrinsic (e.g. dispersal behaviour) or extrinsic (e.g. hydrographical fronts) to species (or an interplay of both) and may operate on ecological (10-100 generations) or evolutionary (>1000 generations) timescales. The common cuttlefish, *Sepia officinalis*, is a nektobenthic species inhabiting shallow coastal waters (to 200m depth), ranging from the Baltic Sea to the Mauritania-Senegal border in the NE Atlantic and throughout the Mediterranrean Sea. The species lacks a planktonic larval stage and although adults undertake seasonal migrations between deeper and shallower waters, they are considered weak swimmers. The limited dispersal ability of the species leads to predictions that gene flow will be restricted leading to local population structuring. In agreement with this, Perez-Losada et al. (2007) reported isolation by distance (increasing genetic differentiation with geographical distance) to be the predominant model explaining the partitioning of mtDNA variation among Mediterranean and NE Atlantic samples. Using microsatellite markers, Perez-Lozada et al. (2002) also reported population structuring around the Iberian Peninsula that could be attributed to an isolation by distance model but could not rule out introgression between historically isolated populations. Garoia et al. (2004) detected no significant population structuring of *S. officinalis* within the Adriatic but noted that the geographical scale of the study may not have been sufficient to include population boundaries (i.e. insufficient geographical scale for distance effects to be detected).

Despite such studies there remains considerable uncertainty regarding population structuring and recurrent patterns of connectivity throughout the range of *S. officinalis*. Data are particularly limited for the northern range of the species (Bay of Biscay; English Channel, North Sea). As the species supports an important commercial fishery and occupies a significant position in the trophic chain of Atlantic and Mediterranean ecosystems such information is urgently needed to inform management and conservation practises, particularly in light of indications of putative population declines linked to overexploitation (Perez-Losada et al. 2007). The objectives of this research were to investigate: (i) spatial and temporal patterns of population structure within the English Channel; (ii) regional population structuring between the Channel and Bay of Biscay.

Materials and methods

Sample collection and preparation: All samples were provided by project partners in the form of arm tips or mantle (for adults), or eggs preserved in absolute ethanol and are detailed in Table 2. Within the Channel *S. officinalis* adults undertaking return migrations were sampled in 2009 from 5 spawning sites with return migrant adults sampled from two of these sites again in 2010. Eggs were collected from 4 of

the spawning sites. Overwintering adults were also sampled from a deeper offshore location. For comparison of regional patterns adults were sampled from both a northern and southern location in the Bay of Biscay. Total DNA was extracted using a standard CTAB-phenol/chloroform method.

Microsatellite marker development: Prior to this project Shaw & Perez-Losada (2000) developed 7 microsatellite loci for *S. officinalis*, which while providing considerable insight into *S. officinalis* genetic structuring exhibit a number of non-desirable technical features (interlocus linkage, complex mutation processes) that could compromise inferences of fine scale patterns (Waples 1998). Therefore, as a step beyond the original project plans it was decided to develop additional microsatellite loci to increase investigative power. A partial genomic library was probed for TG and GA dinucleotide repeat sequences following Shaw & Perez-Losada (2000). Hybridising 'positive' clones were sequenced and PCR primers designed from microsatellite flanking regions following McKeown & Shaw (2008). PCR conditions were optimised for 11 new loci (Table 3).

Genetic Analysis of samples: Each locus was individually amplified in a 10µl reaction mixture containing ~100 ng of template DNA, 1X buffer, 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 pmol of each primer, and 0.2 U of *Taq* DNA polymerase (Bioline, UK). PCR amplifications involved an initial denaturation step (95 °C for 3 min) followed by 35 cycles of 30 s at 95 °C, 30 s at the locus optimal annealing temperature; and 30 s at 72 °C. Optimal annealing temperatures are given in Table 3 for the new loci (prefixed *Sof*-di) and in Shaw & Perez-Losada (2000) for the original loci (prefixed *Sof*). PCR products were size separated using an AB3500 DNA sequencer (Applied Biosystems) with allele inference performed using the GENEMAPPER software (version 4.1, Applied Biosystems).

Power Simulation: Prior to the analysis of all population samples a subset of individuals (n = 48) were genotyped at the new and original loci. The simulation method implemented in the program POWSIM (Ryman & Palm 2006) was then used to estimate sample size dependent probabilities of Type I and Type II errors under the different models of population structure (i.e. level of genetic differentiation among populations). This preliminary analysis was performed to choose the most informative combination of loci for the complete population analysis.

Statistical analysis of genetic diversity

1. Genetic variation within samples: Numbers of alleles (N_A), allelic richness (A_R; El Mousadik & Petit 1996 - the number of alleles estimated by rarefaction if all samples were equal to the smallest sample size in a set of comparisons), observed heterozygosity (H_O), and expected heterozygosity (H_E) (Nei 1978), were calculated using FSTAT 2.9.3.2 (Goudet 1995). Genotype frequency conformance at individual loci to Hardy-Weinberg equilibrium (HWE) expectations and genotypic linkage equilibrium between pairs of loci were tested using exact tests incorporating a Markov chain method (Guo & Thompson 1992) with default parameters in GENEPOP 3.3 (Raymond & Rousset 1995). Multilocus values of significance for HWE tests were obtained using Fisher's method (Sokal & Rohlf 1995) to

combine probabilities of exact tests. Deviations from HWE proportions were quantified by the unbiased inbreeding coefficient, F_{IS}, calculated according to Weir & Cockerham (1984) in FSTAT.

Genetic relatedness between all individuals, and mean relatedness values (and variances) for each sample were estimated using the triadic likelihood estimator (trioML, Wang 2007), calculated with the program COANCESTRY (Wang 2010). Differences among pairs of groups were then tested using the bootstrapping approach implemented in COANCESTRY.

Due to the increased chance of related individuals being sampled among eggs (non-random spatial distribution of related individuals) kinship patterns were investigated using the maximum likelihood kinship analysis method implemented in COLONY 2.0.0.1 (Wang 2004) to identify full-sib and half-sib groupings within the egg population samples.

2. Testing population structure: The null hypothesis of allele frequency homogeneity among samples was qualitatively tested for each locus by exact tests implemented in GENEPOP with values combined over loci according to Fischer's method. Genetic differentiation among samples was quantified by F_{ST} (Wright 1951), the coefficient of inter-sample genetic variation, which ranges from 0 (identical gene frequencies) to 1 (samples fixed for different genetic variants). Single locus and multilocus estimates of F_{ST} were calculated in GENEPOP 3.4. The significance of F_{ST} estimates were tested by permutation (following Goudet *et al*. 1996) whereby individuals/genotypes were shuffled among samples (10,000 iterations for each F_{ST} tested) and F_{ST} recalculated. Permutations were performed for each locus separately and for all loci simultaneously and the probability of the null hypothesis (F_{ST} = 0) was taken as the proportions of replicates that yielded a value of F_{ST} that was equal to, or higher than the observed value. The presence and frequency of null alleles (non-amplifying alleles) was investigated using the expectation maximisation (EM) algorithm implemented in FreeNA (Chapuis & Estoup 2007). FreeNA was also used to generate estimates of F_{ST} (and associated 95% confidence intervals) accounting for such null alleles.

An important consideration when sampling highly mobile adults is the possible effect of sampling migrants on estimates of population structure derived from comparisons between admixed samples. Population structure was also investigated using the Bayesian clustering analysis implemented in the program STRUCTURE (Pritchard *et al*. 2000) to estimate the most probable number of populations represented by the data, and membership of individuals to such clusters, without any *a priori* sample information. Such an approach would not be compromised by sampling of migrants. Both the 'no admixture model' (as recommended for low F_{ST}; Pritchard *et al*. 2000) and 'admixture model with correlated allele frequencies' were used independently to identify the number of clusters, *K* (from a range of 1-10), with the highest posterior probability. Each MCMC run consisted of a burn in of 10^{-6} steps followed by 5 X 10^{-6} steps: 3 replicates were conducted for each *K* to assess consistency. The *K* value best fitting the data set was estimated by the log probability of data [Pr(X/K].

TASK 6

Results

Microsatellite marker development: The new microsatellite loci all generated high quality products. Among the 48 test samples allelic variation was detected at all but one locus (*Sof*-di-1) with allele sizes differing by expected multiples of their repeated motifs. Standard diversity indices for each locus are presented in Table 3 along with primer sequences and allele size ranges. Tests for linkage disequilibrium (LD) and deviations of genotype proportions from expectations of Hardy–Weinberg equilibrium (HWE) were performed as for the full population analyses (described in the section 'statistical analysis of genetic diversity'). No significant LD was detected between any locus pair. Genotype proportions conformed to HWE expectations for 8 of the 10 polymorphic loci, with significant departure from HWE due to an excess of homozygotes found at loci *Sof*-di-4 and *Sof*-di-13 (Table 3).

Power analysis and locus selection: Simulations in POWSIM indicated that the new suite of microsatellite loci conferred a considerable increase in the power to detect subtle population structuring compared to the original loci (Fig. 2). Simulations also revealed that a combination of the new loci *Sof*-di-3, 4, 8, 13, 21, 26 and old loci *Sof*-1, 2, 6 was likely to be the most informative for the population level analysis. This combination conferred a 99% chance of detecting population structure at F_{ST} = 0.0025 with a Type I error rate of 0.03. Analysis of all loci was predicted to confer identical power to detect structure but would result in a greater Type I error rate (0.07). The subset of 9 loci were therefore genotyped in all samples.

Population genetic analysis: The total number of alleles per locus ranged from 10 - 27 (average = 17). Levels of genetic variation were generally similar among samples (Table 4 -7) though some individual locus indices were notably lower for the Seine Bay egg and Sene samples. 46 out of 126 locus/sample tests of Hardy-Weinberg equilibrium conformance indicated significant deviations (Table 8). Such deviations were particularly evident at two loci. Locus *Sof*-di-4 exhibited a significant deficit of heterozygotes in all 14 samples, while *Sof*-di-13 exhibited significant heterozygote deficiency in all samples apart from the Seine Bay eggs. The remaining deviations from HWE were randomly partitioned among loci and mostly due to heterozygote deficits. The heterozygote deficits at *Sof*-di4 and 13 were also evident in the large positive F_{IS} values at these loci. Upon exclusion of these loci significantly negative F_{IS} values were detected for the Seine Bay eggs and Sene samples. F_{IS} values were not significant for the three other egg samples, while significantly positive F_{IS} values were observed for a number of adult samples (Table 9). The Sene sample demonstrated a markedly higher number of cases of significant linkage disequilibrium between pairs of loci (13 out of 21 tests) compared to other samples (average for other samples was 3 out of 21 tests). All loci conformed to linkage equilibrium when tests were performed globally (excluding the egg samples).

Mean values of relatedness were noticeably higher for the Seine Bay egg and Sene samples. However, this difference was only significant for the Seine Bay egg sample (Fig. 4). Kinship analysis revealed extensive kin structuring within the egg samples (Figs.5-8). Among the Torbay eggs 11 pairs of

individuals (dyads) were called as full-sibs, 119 as half-sibs, and individuals could be grouped into 37 full-sib families. For Selsey eggs the numbers were 23 full-sib dyads, 131 half-sib dyads and individuals fitting into 32 full-sib families. For Agon Coutanville analysis revealed 55 full-sib dyads, 123 half-sib dyads, and individuals fitting into 28 family groups. Kin structuring among eggs was most extreme for Seine Bay which reported 442 full-sib dyads, 397 half-sib dyads and all individuals fitting into 13 family groups. For Seine Bay 40 out of the total of 48 eggs could actually be grouped into one full-sib family. As this kin structure could significantly bias pairwise tests of genetic differentiation the egg samples were omitted from such tests.

Analysis of genetic differentiation among samples of adults based on the raw data for the 9 loci revealed a number of significant differences between samples (Table 10). As *Sof*-di 4 and 13 had revealed widespread heterozygote deficiencies pairwise tests were repeated: (i) excluding both these loci (Table 11); and (ii) with the null allele correction method described (Table 12). Pairwise tests using the modified data sets revealed a markedly reduced number of significant outcomes. The salient feature of the test was the differentiation of Sene from all other samples. Excluding Sene sample no pairwise comparisons (spatial or temporal), yielded significant results that were replicated across tests, indicating a high degree of genetic similarity across the sample set.

The Bayesian clustering analysis implemented in STRUCTURE revealed that the most probable number of genetically distinct groups present in the dataset was 2 (P = 1, probabilities for all other models of K were ~ 0). However, this result was explained by the clear clustering of the Seine Bay egg genotypes into one group with all remaining samples clustering into the other group. This result was identical across the different models implemented.

Discussion

Two of the loci (*Sof*-di-4 and 13) used for the population analysis exhibited significant heterozygote deficits in all but one population sample analysed. Heterozygote deficits may reflect natural selection, non-random mating, Wahlund effect (inclusion in a sample of individuals derived from genetically differentiated populations), or technical artefacts known as null alleles. Given the locus specific nature of the deficits reported here null alleles are the most likely explanation. The most common source of null alleles is the failure of PCR to amplify some alleles. While the use of species specific primers reduces the probability of null alleles they are still reported to be common among cephalopods and other marine taxa. Shaw *et al*. (2010) demonstrated how the occurrence of null alleles can significantly compromise estimates of genetic structure and potentially result in Type I errors (i.e. the erroneous detection of significant genetic differentiation). Such an effect was evident here as the number of significant pairwise tests of differentiation was markedly reduced when performed: (i) excluding *Sof*-di-4 and 13; and (ii) when null allele correction methods were employed. Despite these null alleles the newly developed loci represent a valuable addition to the suite of microsatellite markers available for

studies of *S. officinalis*. The variable levels of polymorphism across loci will permit testing of hypotheses relating to phylogeography, population genetics and kinship in this species. Thus, these loci will potentially provide information vital to the design of sustainable fisheries management plans and conservation of the species long-term evolutionary potential.

Genetic structuring among adults was investigated using a range of pairwise tests. With the exception of comparisons involving Sene the majority of these pairwise tests revealed insignificant differentiation when performed with the modified data sets. Furthermore, in cases where a pairwise comparison indicated significant differentiation, this significance was not replicated across different tests or, where applicable, temporal comparisons. Collectively these features indicate that, with the exception of the unanimous differentiation of Sene, all other significant test results are likely to be biologically non-significant (Waples 1998), and therefore suggest substantial genetic connectivity among the respective samples. These data imply that *S.officinalis* within the Channel area represent a sinlge genetic population unit, and that cuttlefish within the Bay of Biscay may also be part of the same genetic population.

An important consideration in the genetic analysis of samples is how mechanical mixing of individuals may obscure detection of population structure by means of pairwise comparisons. The method implemented in STRUCTURE, by testing for population structure without *a priori* sample definition, provides a means of removing such potential bias. However, the resolution of such analysis may be constrained when the level of differentiation among populations is low (F_{ST} < 0.05; Latch et al. 2006).

Analysis of egg samples represents a potentially powerful approach to minimise noise due to intra-generation dispersal in studies of population structure for *S. officinalis*. The respective patterns of F_{IS} observed among adults (significantly positive for 5 Channel samples) and eggs (not significant for 3 egg samples) is certainly compatible with a model of admixture of genetically distinct populations among adults and reduced admixture among egg samples. However, the kinship analysis revealing that egg samples consisted of a large number of related individuals indicates that this would be an oversimplification. When a sample is composed of offspring from a few parents the discreteness of the number of possible genotypes may lead to excess of heterozygotes (relative to expected proportions in an infinite population with the same allele frequencies (Luikart & Cornuet 1999). However, when several family groups are sampled, as was reported here, allele frequencies are expected to vary among such groups, even if parents are all from the same genetic population. Therefore, by pooling the various groups the heterozygote excess may be countered by a family 'Wahlund effect' (Pudovkin et al., 1996). In the case of the eggs from Torbay, Selsey and Agon Coutanville, these opposing factors seem to have cancelled each other out, resulting in an artefactual Hardy-Weinberg equilibrium. For the Seine bay egg sample, the predominance of a single family group skewed the balance towards a heterozygore excess. The distinct clustering of the Seine Bay eggs into a discrete group by STRUCTURE was also driven by

this family sampling effect bias and not population differentiation. In light of these family effects, and the fact that positive, often unexplained, F_{IS} values are widely reported in marine invertebrates (Bohonak 1999; Addison *et al*. 2005), the respective patterns of F_{IS} for adults and juvenile can not be regarded as strong evidence for population structuring.

A salient feature of the data was the genetic differentiation of Sene from all other samples. In light of the lack of differentiation between the Arcachon and the Channel samples geographical distance could be ruled out as the primary factor underlying this. Sweepstakes recruitment has been invoked to explain chaotic genetic patchiness in a number of marine systems (Hedgecock 1994). Under the hypothesis of sweepstakes recruitment a low number of effective breeders results in skewed allele frequencies among recruits, which can lead to apparent genetic differentiation, even in systems where parents are essentially from a panmictic population (genetic differentiation without isolation). The Sene sample exhibited both a heterozygote excess and reduced allelic diversity, genetic signatures compatible with sweepstakes recruitment (Cornuet & Luikart 1996). Such features may also be associated with a genetic bottleneck within an isolated population. However, such a genetic bottleneck would be expected to result in increased within-sample relatedness over time and this was not evident here. Therefore, given the background of genetic similarity within the region, the model of sweepstakes recruitment without isolation is favoured to explain the apparent genetic differentiation of Sene. The Sene sample was collected from within the Golfe du Morbihan, an area with a distinctive environment (e.g. strong tidal currents) that could certainly introduce added stochasticity into recruitment patterns. Despite this, however, sweepstakes effects are typically resolved among juveniles and it is unusual to detect sweepstakes effects among dispersing adults. The patterns reported here call for further studies of dispersal of *S. officinalis* within the Golfe du Morbihan and genetic analysis of additional spatial/temporal samples.

In conclusion the genetic data indicate that: (i) although signs of genetic differences were detected between samples, a lack of a clear association with geographic area or cohort indicates that there is a lack of distinct genetic structuring within the Channel *S.officinalis* population (i.e. it represents a single genetic population); and (ii) substantial genetic connectivity is suggested between Channel and Bay of Biscay *S.officinalis* populations (at least within the studied area).

References

Addison JA, Hart MW (2005) Spawning, copulation and inbreeding coefficients in marine invertebrates. *Biology Letters*, **1,** 450–453.

Avise JC (2000) *Phylogeography*. Harvard University Press, Cambridge, MA.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. Quarterly Review in Biology,**74**, 21-45.

Carvalho GR, Hauser L (1994) Molecular genetics and the stock concept in fisheries.*Reviews in Fish Biology and Fisheries*, **4**, 326-350.

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621-631.

Fox CJ, Taylor MI, Pereyra R, *et al*. (2005) Taqman DNA technology confirms likely over-estimation of cod (*Gadus morhua* L.) egg abundance in the Irish Sea: implications for the assessment of the cod stock and mapping of spawning areas using egg-based methods. *Molecular Ecology*, **14**, 879-884.
Garoia F, Guarniero I, Ramsak A, Ungaro N, Landi M, Piccinetti C, Mannini P, Tinti F (2004) Microsatellite DNA variation reveals high gene flow and poanmictic populations in the Adriatic shared stocks of the European squid and cuttlefish (Cephalopoda). *Heredity*, **93**, 166-174.
Gold JR, Richardson LR (1998) Mitochondrial DNA diversification and population structure in fishes from the Gulf of Mexico and western Atlantic. *Journal of Heredity*, **89**, 404-414.
Guo SW, Thompon EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics*, **48**, 361-372.
Goudet J (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485-486.
Goudet J, Raymond M, de Meeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933-1940
Hall TA (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nuclear Acids Symp Ser, **41**,95-98.
Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: *Genetic and Evolution of Aquatic Organisms* (ed. Beaumont AR), pp. 122-134. Chapman & Hall, London, UK.
Hutchings JA (2000) Collapse and recovery of marine fishes. *Nature*, **406**, 882-885.
Knutsen H, Jorde PE, Andre C, Stenseth CHR (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology*, **12**, 385-394.
El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, **92**, 832–839.
Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE Jr (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**,295–302.
Luikart G, Cornuet JM (1999) Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics* **151**: 1211-1216.
McKeown NJ, Shaw PW (2008) Polymorphic nuclear microsatellite loci for studies of brown crab, *Cancer pagurus* L. *Molecular Ecology Resources*, **8**, 653-655.
McQuinn IH (1997) Metapopulations and the Atlantic Herring. *Reviews in Biology and Fisheries*, **7**, 297-329.
Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583.
Perez-Losada M, Guerra A, Carvalho GR, Sanjuan A, Shaw PW (2002) Extensive population subdivision of the cuttlefish *Sepia officnalis* (Mollusca:Cephalopoda) around the Iberian Peninsula indicated by microsatellite DNA variation. *Heredity*, **89**, 417-424.
Perez-Losada M, Nolte MJ, Crandall KA, Shaw PW (2007) Testing hypotheses of population structuring in the Northeast Atlatic Ocean and Mediterranean Sea using the common cuttlefish *Sepia officinalis*. *Molecualr Ecology*, **16**, 2667-2679.
Pritchard JK, Stephen M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959
Pudovkin AI, Zaykin DV, Hedgecock D (1996). On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* **144**: 383–387.
Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology*, **6,** 600-602.
Shaklee JB, Bentzen P (1998) Genetic identification of stocks of marine fish and shellfish. *Bulletin of Marine Science*, **62**, 589–621.

Shaw PW, Hendrickson L, McKeown NJ, Stonier T, Naud M-J, Sauer WHH (2010) Discrete spawning aggregation of loliginid squid do not represent genetically distinct populations. *Marine Ecology Progress Series*, **408**, 117-127.

Shaw P, Perez-Losada M (2000) Polymorphic microsatellite loci in the common cuttlefish *Sepia officinalis* (Cephalopoda). *Molecular Ecology*, **9,** 237-238
Sokal RR, Rohlf, FJ (1995) Biometry: the principles and practise of statistics in biological research. 3rd Edition
Strauss RE, Bond CE (1990) Taxonomic methods:morphology. In Schreck CB. Moyle PB (eds) *Methods for Fish Biology*. American Fisheries Society.
Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: a program to cleave DNA with restriction enzymes Nucleic acids Res 31:3688-3691.
Wang, J. (2004) Sibship reconstruction from genetic data with typing errors.*Genetics* 166: 1963-1979.
Wang, J. (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetical Research* 89: 135-153.
Wang, J. 2010. COANCESTRY: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, **11**, 141-145.
Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, **89**, 439–450.
Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **386**, 1358-1370.
Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323-354.

Tables

Table 1. Location of diagnostic restriction enzyme cleavage sites within the PCR amplified 544bp fragment

	L.vulgaris	*L. forbesi*	*A. media*	*A. subulata*
Sfc I	277;352	277	No cut site	No cut site
Bcc I	No cut site	385	254	254;436

Table 2. Details of samples included in population geentic analysis of *S. officinalis* . *N* denotes the number of individuals successfully genotyped at all loci for the respective sample

Sample site	Specimen type-year	*n*	Sample code
Torbay (UK)	Spawning adults-2009	65	Torbay-2009
	Spawning adults-2010	44	Torbay-2010
	Eggs-2010	48	Torbay-egg
Selsey (UK)	Spawning adults-2009	71	Selsey-2009
	Eggs-2010	48	Selsey-egg
Eastbourne (UK)	Spawning adults-2009	69	Eastbourne-2009
	Spawning adults-2010	48	Eastbourne-2010
Agon Coutanville (France)	Spawning adults-2009	48	Agon Countanville
	Eggs-2010	48	Agon C-egg
Seine Bay (France)	Spawning adults-2009	48	Seine Bay
	Eggs-2010	48	Seine Bay-egg
Centre of Channel	Overwintering juveniles-2010	35	Casquets
North Biscay - Golfe du Morbihan	Adults-2006	45	Sene
South Biscay - Bassin d'Arcachon	Adults-2006	48	Arcachon

TASK 6

Table 3. Primer sequences and characteristics of 11 microsatellite loci developed for *Sepia officinalis* , including repeat motif observed in the clone used to develop each locus and optimal annealing temperature (T_a) for each primer pair. Allele numbers (N_a) and size range, observed (H_O) and expected (H_E) heterozygosity and *P* - values for tests of Hardy-Weinberg equilibrium (P_{HW}) calculated from analysis of 48 individuals.

Locus	Primer Sequences (5''-3')	Repeat motif in cloned allele	T_a (^{0}C)	N_a	Size range (bp)	H_O	H_E	P_{HW}
Sof-di-1	F: CAAACAACGTAGACAAAGAAATGG R: TGACCAGGACGAGAGAATGC	$(GT)_{12}$	62	1	198	0	0	-
Sof-di-2	F: AGTCAGCAGCTCGACAAACTC R: TGGAAGTAGCAAGGAAAGCAG	$(GT)_9GA(GT)_3GA(GT)_9$	60	3	168-172	0.104	0.101	1
Sof-di-3	F: ATTGTGATGAAGGCCCAATC R: CTTTCAGACGTCTCACTACATGC	$(GT)_{12}ACAT(GT)_3...(GT)_{18}$	59	11	220-228	0.771	0.758	0.4482
Sof-di-4	F: AATCTTCCACACGGAAGTAAAG R: ACATTGGTCGCAGTCGAAAC	$(GT)_{28}$	58	14	138-188	0.438	0.806	<0.0001
Sof-di-6	F: GGTTTCACCGTCATGTTGTG R: CGAAAGGAAAAGAAGAGACTCG	$(GT)_{14}CT(GT)_5$	58	14	202-252	0.75	0.697	0.624
Sof-di- 8	F: TGTATGTCCGAATCACATCG R: AATCGATAGACCGCCTAAACC	$(GT)_{16}$	57	9	148-170	0.646	0.748	0.259
Sof-di-13	F: TGAAAAAGTTCCCCAAAACTTC R: TCCGTACAACCGACGTAGAAC	$(AT)_5AC(AT)_5(GT)_2AT(GT)_{11}$	59	6	194-210	0.319	0.639	<0.0001
Sof-di-14	F: CGATTAGGCGAACCTATCTACG R: ATTCTTTTTATTATTAAGGAAATGCTG	$GTAT(GT)_{12}$	61	5	188-204	0.125	0.121	1
Sof-di-21	F: CCAGATTTCTTAAAAACAGTTTCTCTC R: CGAATCGTTTAAAATTTAGTGCTG	$(TC)_4...(CA)_{24}TA(CA)_3(CT)_2TT(CT)_7$	59	12	214-252	0.792	0.792	0.279
Sof-di-22	F: GCCACTAAGACGTAAGAGGTTG R: TCGACTCTAGAGGATCCTTGC	$(GT)_{17}(GA)_7TA(GA)_9$	55	3	158-166	0.104	0.101	1
Sof-di-26	F: AATATGGGCATATAAGTGTCT R: GCACATAGCCTGACGATTGA	$(CA)_3C(CA)_{16}(TA)_8$	57	5	216-232	0.646	0.635	0.710

TASK 6

Table 4. Allele number (*Na)* per locus/sample

	Eggs				Adults/Juveniles									
	Torbay	Selsey	Agon Coutanville	Seine Bay	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Sof-2	8	8	8	5	8	9	10	9	9	9	10	9	6	9
Sof-1	8	9	7	4	10	9	8	7	7	6	9	7	7	7
Sof-6	7	8	7	4	7	7	10	7	7	7	7	6	4	8
Sof-di-3	10	12	8	7	14	13	13	11	7	11	12	8	7	10
Sof-di- 4	12	13	12	10	16	12	14	13	13	13	15	10	12	17
Sof-di-21	12	14	8	9	14	12	14	13	13	12	14	12	12	15
Sof-di-13	4	3	3	4	5	3	4	5	2	6	6	4	4	3
Sof-di-8	6	8	5	4	9	6	10	10	8	9	10	6	6	8
Sof-di-26	6	6	6	8	8	7	8	6	8	5	5	8	8	11

Table 5. Allelic Richness per locus and sample based on a minimum sample size of 29

	Eggs				Adults/Juveniles									
	Torbay	Selsey	Agon Coutanville	Seine Bay	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Sof-2	7.228	6.992	7.463	4.032	7.098	8.124	8.286	8.364	8.197	8.157	8.701	8.625	5.352	8.156
Sof-1	6.753	8.228	6.234	4	8.316	7.7	6.72	6.201	6.893	5.61	7.796	6.866	6.684	6.663
Sof-6	6.639	6.386	6.142	3.63	6.378	6.575	7.982	5.736	6.681	6.728	5.7	6	3.706	7.646
Sof-di-3	8.914	9.075	7.201	6.642	10.353	10.341	10.409	9.238	6.622	9.533	9.707	7.757	6.458	8.959
Sof-di- 4	10.985	11.466	11.414	9.313	12.982	11.245	11.384	9.92	11.634	11.361	12.825	9.804	10.572	14.72
Sof-di-21	10.509	12.626	6.851	7.791	10.299	9.805	10.766	10.752	10.602	10.1	11.798	11.8	10.405	12.79
Sof-di-13	3.845	2.846	2.997	3.583	3.905	3	3.536	4.615	2	5.42	5.205	3.987	3.674	2.784
Sof-di-8	5.2	6.877	4.843	3.622	6.889	5.127	7.633	7.432	7.247	7.503	7.888	5.871	5.331	7.204
Sof-di-26	5.47	5.505	5.391	7.054	6.433	6.514	6.13	5.321	7.479	4.601	4.823	7.867	7.679	9.421

TASK 6

Table 6. Expected heterozygosity (H_E) per locus and population

	Eggs				Adults/Juveniles									
	Torbay	Selsey	Agon Coutanville	Seine Bay	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Sof-2	0.512	0.606	0.673	0.159	0.611	0.572	0.667	0.687	0.557	0.656	0.624	0.61	0.461	0.638
Sof-1	0.72	0.761	0.728	0.682	0.789	0.765	0.767	0.746	0.748	0.749	0.756	0.782	0.759	0.774
Sof-6	0.661	0.616	0.605	0.654	0.634	0.701	0.652	0.554	0.641	0.671	0.604	0.672	0.532	0.713
Sof-di-3	0.783	0.733	0.671	0.497	0.764	0.723	0.76	0.709	0.676	0.758	0.751	0.752	0.675	0.726
Sof-di- 4	0.845	0.812	0.893	0.729	0.847	0.868	0.871	0.829	0.819	0.804	0.887	0.77	0.837	0.872
Sof-di-21	0.836	0.857	0.764	0.738	0.773	0.752	0.798	0.825	0.832	0.792	0.824	0.858	0.778	0.836
Sof-di-13	0.512	0.501	0.457	0.507	0.545	0.57	0.576	0.629	0.504	0.639	0.63	0.527	0.542	0.484
Sof-di-8	0.675	0.689	0.684	0.36	0.71	0.628	0.671	0.71	0.717	0.748	0.707	0.718	0.569	0.709
Sof-di-26	0.641	0.559	0.682	0.612	0.677	0.66	0.619	0.638	0.732	0.635	0.633	0.628	0.724	0.783

Table 7. Observed heterozygosity (H_O) per locus and population

	Eggs				Adults/Juveniles									
	Torbay	Selsey	Agon Coutanville	Seine Bay	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Sof-2	0.500	0.583	0.681	0.146	0.585	0.596	0.662	0.762	0.513	0.604	0.646	0.563	0.571	0.630
Sof-1	0.542	0.813	0.809	0.837	0.785	0.723	0.739	0.767	0.722	0.660	0.848	0.818	0.756	0.690
Sof-6	0.646	0.542	0.542	0.696	0.597	0.681	0.615	0.492	0.588	0.532	0.574	0.759	0.700	0.705
Sof-di-3	0.750	0.708	0.660	0.500	0.657	0.723	0.741	0.643	0.636	0.771	0.771	0.636	0.854	0.739
Sof-di- 4	0.391	0.362	0.468	0.486	0.522	0.261	0.746	0.500	0.268	0.438	0.617	0.441	0.682	0.340
Sof-di-21	0.915	0.894	0.574	0.872	0.790	0.688	0.702	0.825	0.750	0.792	0.667	0.742	0.841	0.867
Sof-di-13	0.313	0.333	0.354	0.457	0.194	0.271	0.241	0.322	0.257	0.319	0.386	0.364	0.372	0.162
Sof-di-8	0.667	0.689	0.604	0.250	0.688	0.609	0.589	0.635	0.610	0.646	0.596	0.613	0.610	0.628
Sof-di-26	0.766	0.467	0.875	0.727	0.567	0.711	0.586	0.444	0.730	0.646	0.667	0.667	0.907	0.833

TASK 6

Table 8. P values of tests of conformance of genotype proportion to Hardy-Weinberg equilibrium expectations

	Eggs				Adults/Juveniles									
	Torbay	Selsey	Agon Coutanville	Seine Bay	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Sof-2	0.1553	0.3978	0.3818	0.0746	0.0739	0.5921	0.2665	0.9715	0.0799	0.1375	0.9242	0.1224	0.7699	0.3397
Sof-1	0.0222	0.8568	0.1738	<0.00001	0.1344	0.6565	0.5206	0.7655	0.3516	0.5428	0.6023	0.8904	0.0037	0.6533
Sof-6	0.1952	0.0351	0.0212	0.0385	0.5018	0.1642	<0.00001	0.2121	0.7179	0.0441	0.0614	0.714	0.0004	0.1162
Sof-di-3	0.7244	0.4972	0.2865	0.2052	0.009	0.7713	0.5022	0.2057	0.0949	0.4732	0.5965	0.1895	0.2286	0.2385
Sof-di- 4	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001	<0.00001
Sof-di-21	0.5118	0.4932	<0.00001	0.0004	0.0422	0.5856	0.0619	0.1669	0.6194	0.2962	0.0021	0.0963	0.0713	0.9942
Sof-di-13	0.0003	0.0272	0.0066	0.4994	<0.00001	<0.00001	<0.00001	<0.00001	0.0055	<0.00001	0.0005	0.0073	0.0059	<0.00001
Sof-di-8	0.5579	0.1543	0.2702	0.0048	0.1715	0.7223	0.3067	0.1998	0.5119	0.2284	0.0222	0.6159	0.9456	0.1849
Sof-di-26	0.7939	0.0235	0.0344	0.6276	0.029	0.1869	0.583	0.0084	0.0916	0.7162	0.3153	0.7294	0.0006	0.6029

Table 9. F_{IS} per locus and sample

	Eggs				Adults/Juveniles									
	Torbay	Selsey	Agon Coutanville	Seine Bay	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Sof-2	0.024	0.038	-0.011	0.084	0.043	-0.041	0.008	-0.109	0.079	0.079	-0.034	0.078	-0.239	0.012
Sof-1	0.248	-0.068	-0.111	-0.227	0.005	0.055	0.037	-0.028	0.034	0.12	-0.122	-0.047	0.004	0.108
Sof-6	0.023	0.121	0.105	-0.063	0.059	0.028	0.056	0.113	0.082	0.208	0.048	-0.128	-0.316	0.012
Sof-di-3	0.043	0.033	0.018	-0.006	0.14	-0.001	0.024	0.094	0.059	-0.017	-0.026	0.154	-0.265	-0.018
Sof-di- 4	0.537	0.554	0.476	0.333	0.383	0.699	0.143	0.397	0.672	0.456	0.304	0.427	0.185	0.61
Sof-di-21	-0.095	-0.043	0.248	-0.181	-0.022	0.086	0.12	-0.001	0.099	0	0.191	0.135	-0.081	-0.037
Sof-di-13	0.39	0.335	0.226	0.1	0.645	0.525	0.582	0.488	0.49	0.501	0.387	0.309	0.313	0.665
Sof-di-8	0.012	0	0.117	0.306	0.031	0.03	0.122	0.106	0.15	0.136	0.157	0.146	-0.071	0.114
Sof-di-26	-0.194	0.165	-0.283	-0.189	0.163	-0.077	0.053	0.303	0.004	-0.017	-0.054	-0.061	-0.253	-0.064
Multilocus	0.113	0.121	0.096	-0.006	0.152	0.157	0.119	0.148	0.185	0.162	0.1	0.113	-0.071	0.144
Multilocus-7 loci	0.009	0.026	0.013	**-0.088**	**0.059**	0.015	**0.061**	**0.062**	**0.072**	**0.072**	0.027	0.044	**-0.165**	0.017

Multilocus - 7 loci refers to value obtained after exclusion of *Sof*-di-4 and *Sof*- di-13. Values in bold were found to be signficant after 1000 permutations.

TASK 6

Table 10. Pairwise tests for genetic differentiation based on 9 locus raw data. Below triangular matrix - F_{ST} with significant values after 1000 permutations in boldface. Top triangular matrix - P values from exact tests of genetic differentiation (significant values in bold).

	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	Torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Eastbourne-2009	-	**0.0001**	0.0975	**0.0002**	**<0.00001**	0.1100	**<0.00001**	0.0531	**<0.00001**	**<0.00001**
Eastbourne-2010	-0.0013	-	**<0.00001**	**0.0000**	**0.0029**	**<0.00001**	**<0.00001**	**0.0088**	**0.0043**	0.0255
Selsey-2009	-0.0021	-0.0027	-	**0.0001**	**0.0001**	**0.0012**	**<0.00001**	0.2671	**<0.00001**	**<0.00001**
Torbay-2009	**0.0042**	**0.0038**	**0.0015**	-	**0.0000**	**0.0054**	**<0.00001**	**0.0017**	**<0.00001**	**<0.00001**
Torbay-2010	-0.0009	0.0009	0.0031	**0.0063**	-	**0.0025**	**<0.00001**	**0.0055**	**0.0004**	0.5135
Agon Coutanville	-0.0036	**0.0011**	-0.0004	**0.0008**	-0.0017	-	0.2097	**0.0298**	**<0.00001**	**<0.00001**
Seine Bay	**0.0033**	**0.0025**	**0.0011**	**0.0054**	**0.0074**	-0.0005	-	**0.0001**	**<0.00001**	**<0.00001**
Casquets	0.0022	0.0016	0.0022	0.0052	0.0009	-0.0006	0.0066	-	**0.0004**	0.1124
Sene	**0.0113**	**0.0045**	**0.0094**	**0.0119**	**0.0059**	**0.0145**	**0.0139**	**0.0126**	-	**<0.00001**
Arcachon	**0.0058**	0.0005	**0.0032**	**0.0065**	-0.0033	**0.0042**	**0.0065**	0.0024	**0.0074**	-

Table 11. Pairwise tests for genetic differentiation based on 7 loci after exclusion of *Sof*-di-4 and *Sof*-di-13. Below triangular matrix - F_{ST} with significant values after 1000 permutations in boldface. Top triangular matrix - P values from exact tests of genetic differentiation (significant values in bold).

	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	Torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Eastbourne-2009	-	0.144	0.308	**0.0049**	0.118	0.349	0.0697	0.504	**<0.00001**	0.099
Eastbourne-2010	-0.0029	-	0.806	0.147	0.207	0.164	0.265	0.138	**<0.00001**	0.453
Selsey-2009	-0.0028	-0.0035	-	0.18	0.106	0.193	0.0537	0.66	**<0.00001**	0.205
Torbay-2009	**0.004**	0.0049	0.0006	-	0.126	**0.0134**	**0.001**	0.07	**<0.00001**	0.089
Torbay-2010	0.0001	0.0003	0.0006	0.0044	-	0.497	0.1588	0.502	**0.01257**	0.691
Agon Coutanville	-0.0039	-0.0007	-0.003	0.0009	-0.0029	-	0.524	0.566	**<0.00001**	0.868
Seine Bay	-0.0033	-0.0019	-0.0024	**0.0045**	-0.0007	-0.0049	-	0.139	**<0.00001**	0.251
Casquets	-0.0016	0.0005	-0.0003	0.0036	-0.0034	-0.0038	0.0007	-	**<0.00001**	0.764
Sene	**0.0085**	**0.0078**	**0.0107**	**0.0126**	**0.0036**	**0.0134**	**0.0135**	**0.0151**	-	**<0.00001**
Arcachon	-0.0006	-0.0007	-0.0032	0.0006	-0.0054	-0.0061	-0.0021	-0.003	**0.0094**	-

TASK 6

Table 12. Pairwise estimates of F_{ST} across 9 loci using the null allele correction methods in FreeNA. Significant values after 1000 bootstraps in boldface.

	Eastbourne-2009	Eastbourne-2010	Selsey-2009	Torbay-2009	Torbay-2010	Agon Coutanville	Seine Bay	Casquets	Sene	Arcachon
Eastbourne-2009	-									
Eastbourne-2010	-0.000147	-								
Selsey-2009	-0.0013	-0.001806	-							
Torbay-2009	0.003323	0.004265	0.000903	-						
Torbay-2010	0.000743	0.001744	0.003479	**0.005626**	-					
Agon Coutanville	-0.002912	0.001497	-0.001457	0.001209	-0.000168	-				
Seine Bay	0.003191	0.002463	0.00193	0.006285	0.005766	-0.000344	-			
Casquets	0.002641	0.003215	0.001776	0.005707	0.002373	-0.00003	0.00619	-		
Sene	**0.012435**	**0.007213**	**0.011409**	**0.012729**	**0.006908**	**0.016344**	**0.016121**	**0.013903**	-	
Arcachon	0.005725	0.001578	0.003811	0.004685	-0.003144	0.003651	0.006191	0.004073	**0.009929**	-

TASK 6

Figures

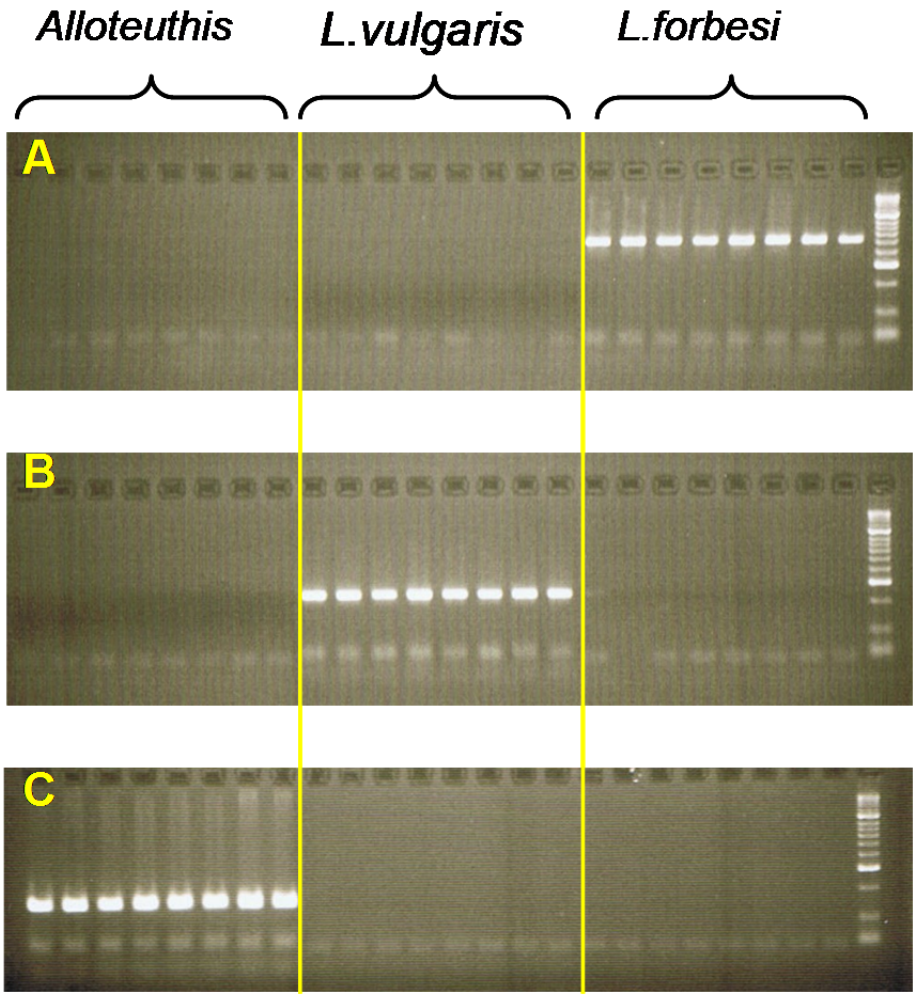

Figure 1. Comparative results of species specific PCR assays for 8 individuals from each taxon performed using the optimised reaction conditions. A – results for the *L. forbesi* primers, B – results for *L. vulgaris* primers, C – results for *Alloteuthis* primers. In all cases assays yielded correct positive identification and no non-specific products.

TASK 6

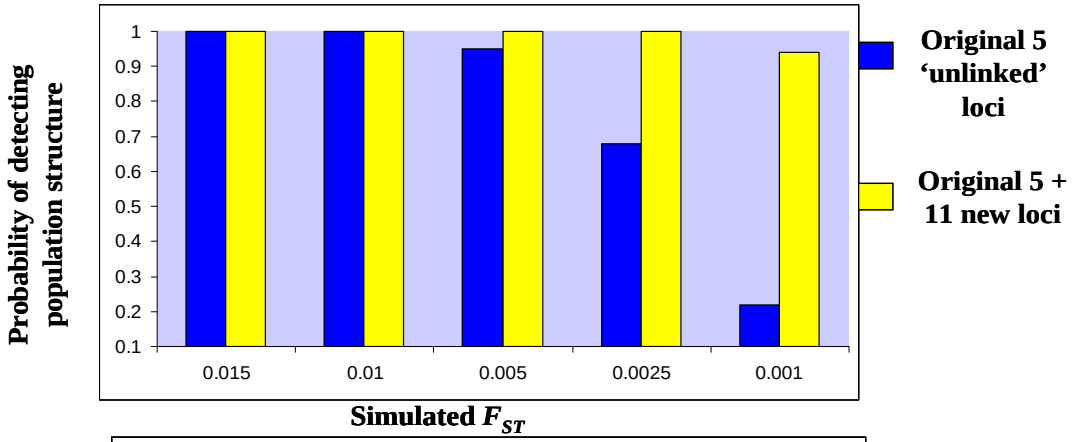

Figure 2. Results of POWSIM analysis estimating the probability of data from two suites of loci (1. original loci; 2. original loci + new loci combined) to detect various levels of population structure as represented by simulated F_{ST}

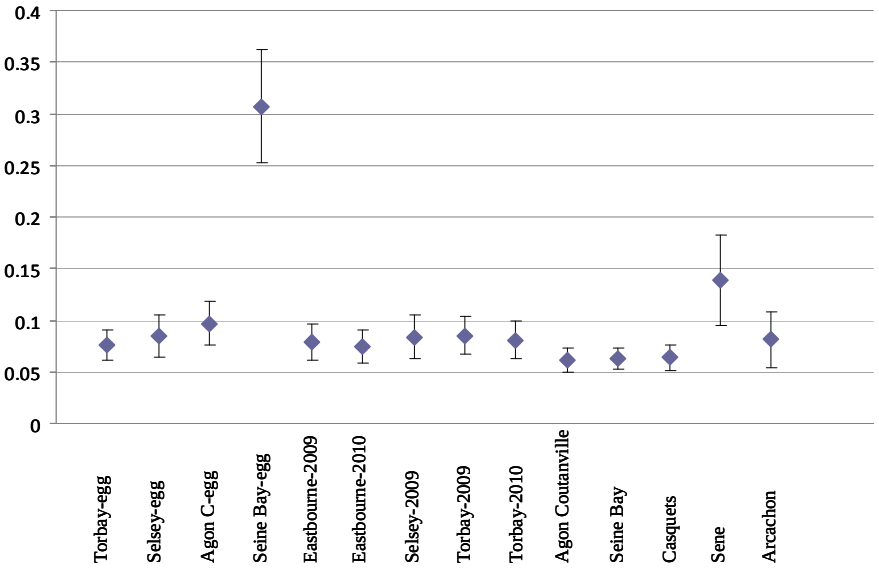

Figure 3. Sample mean relatedness values (and variances) estimated using trioML.

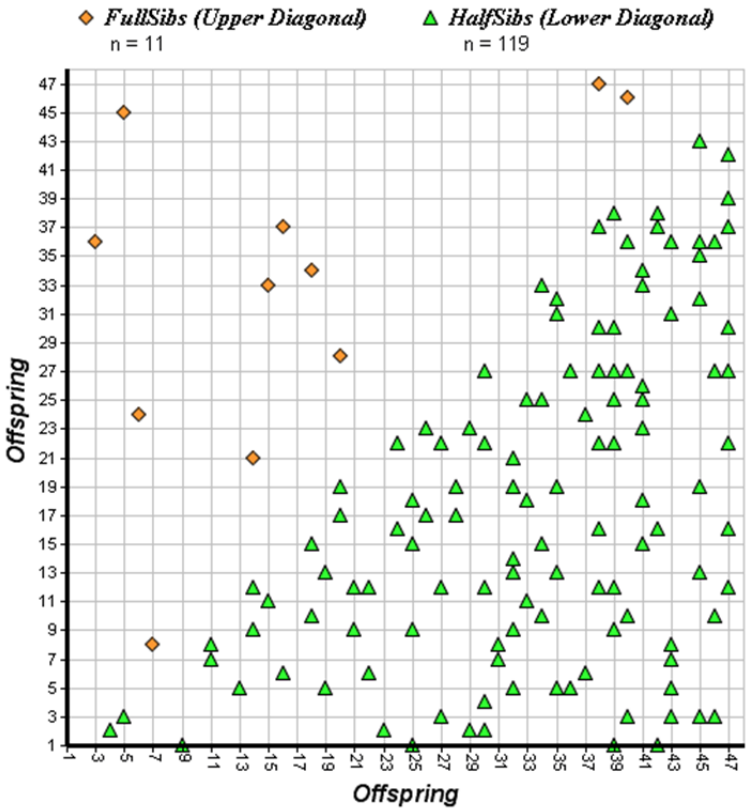

Figure 4. Fullsib and Halfsib dyads identified among Torbay eggs

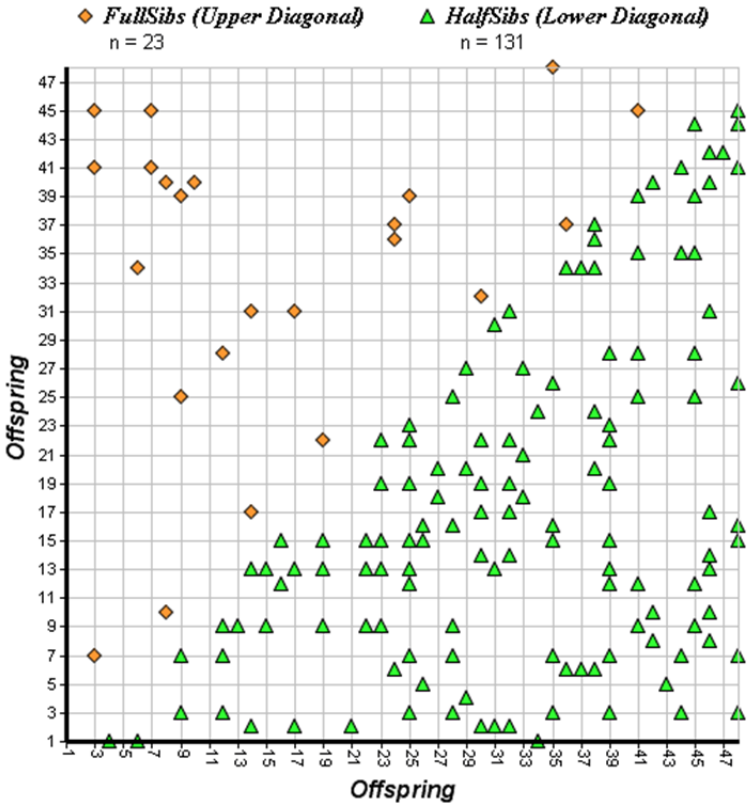

Figure 5. Fullsib and Halfsib dyads identified among Selsey eggs

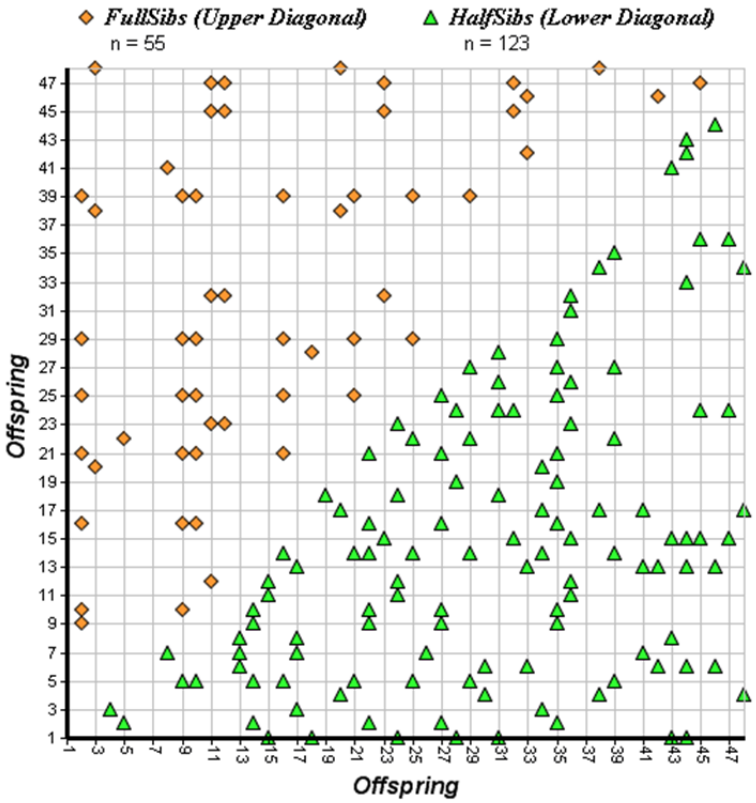

Figure 6. Fullsib and Halfsib dyads identified among Agon Coutanville eggs

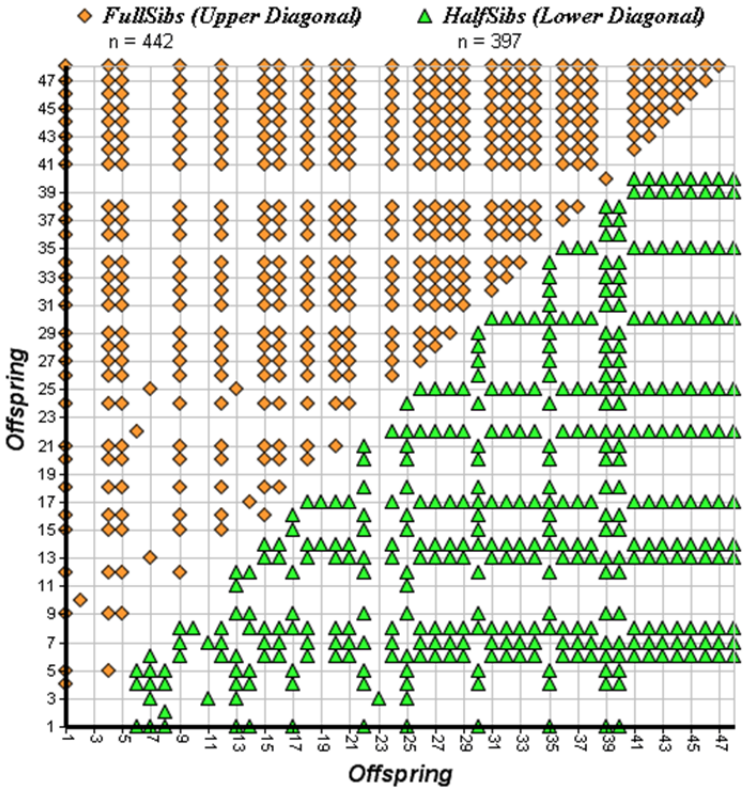

Figure 7. Fullsib and Halfsib dyads identified among Seine Bay eggs