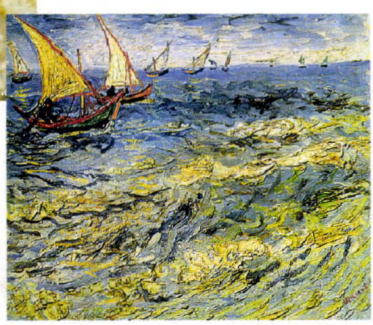
KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculteit Wetenschappen

Departement Biologie
Laboratorium voor Diversiteit en Systematiek van Dieren
Ch. Deberiotstraat 32
B-3000 Leuven

Sea (in)sight

From phylogeographical insights to visual local adaptation in marine gobies



Promotor:
Prof. Dr. Filip A.M. Volckaert

Co-Promotor:
Dr. Jeroen K.J. Van Houdt

Proefschrift voorgedragen tot het behalen van de graad van Doctor in de wetenschappen, 2009

Door:
Maarten H.D. Larmuseau

Katholieke Universiteit Leuven
Groep Wetenschap & Technologie
Faculteit Wetenschappen
Departement Biologie
Afdeling Dierenecologie en -systematiek
Laboratorium voor Diversiteit en Systematiek van Dieren

Sea (in)sight

From phylogeographical insights to visual local adaptation in marine gobies

(In)zicht op zee

Van fylogeografische inzichten naar visuele lokale adaptatie bij mariene grondels

Promotor: Prof. Dr. Filip A.M. Volckaert

Co-promotor: Dr. Jeroen K.J. Van Houdt

Proefschrift voorgedragen tot het behalen van de graad van Doctor in de Wetenschappen door:

Maarten H.D. Larmuseau

2009

ISBN 978-90-8649-272-5
Wettelijk depotnummer D/2009/10.705/50

Groep Wetenschap & Technologie
Faculteit Wetenschappen
Departement Biologie
Afdeling Dierenecologie en -systematiek

Laboratorium voor Diversiteit en Systematiek van Dieren
Ch. Deberiotstraat 32
B-3000 Leuven

KATHOLIEKE
UNIVERSITEIT
LEUVEN

Promotor: Prof. Dr. Filip A.M. Volckaert
Laboratorium voor Diversiteit en Systematiek van Dieren, K.U.Leuven, Leuven, België

Co-promotor: Dr. Jeroen K.J. Van Houdt
Joint Experimental Molecular Unit, Koninklijk Museum voor Midden-Afrika, Tervuren, België

Overige leden van de examencommissie:

Prof. Dr. Ronny Decorte
Departement Menselijke Erfelijkheid, K.U.Leuven, Leuven, België

Prof. Dr. Frans Ollevier
Laboratorium voor Aquatische Ecologie en Evolutiebiologie, K.U.Leuven, Leuven, België

Dr. Christophe Pampoulie,
Marine Research Institute, Reykjavík, Iceland

Dr. Joost A.M. Raeymaekers
Laboratorium voor Diversiteit en Systematiek van Dieren, K.U.Leuven, Leuven, België

Dr. Tom Wenseleers
Laboratorium voor Entomologie, K.U.Leuven, Leuven, België

Het onderzoek voorgesteld in dit proefschrift werd gesteund door een specialisatiebeurs verleend door het Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen).

'Ik snak naar adem als ik alleen wetenschappelijke kennis mag hebben. Biologie beschrijft het leven mechanisch, maar het leven is juist afwijken van het patroon. Leven is speelsheid, poëzie, kunst.'

Gerard Bodifée (Flemish physicist, 1946)

Dit proefschrift dat het resultaat is van 4 jaar multidisciplinair onderzoek, is niet alleen mijn verdienste, integendeel. Daarom wens ik hierbij verscheidene mensen te bedanken voor hun grote hulp bij het tot stand komen van dit werk.

Allereerst bedank ik mijn promotor Filip Volckaert voor zijn jarenlange steun, zijn vertrouwen en zijn streven om van zijn onderzoeksteam ook een hechte groep te maken wat ik altijd sterk geapprecieerd heb.

Daarnaast gaat mijn grootste dank ook uit naar mijn co-promotor Jeroen Van Houdt die het hele project mee vorm heeft gegeven, dit werk naar een veel hoger niveau heeft gebracht en die mij op gepaste tijd kon uitdagen maar vooral ook zelfvertrouwen kon doen geven.

Ik bedank ook mijn overige juryleden voor hun sterke inhoudelijke inbreng: Prof. Frans Ollevier voor zijn interesse en de inhoudelijke besprekingen. Prof. Ronny Decorte voor het vertrouwen en steun bij het schrijven van de postdoctoraat-aanvraag, wat mij tevens een ruimer beeld bood betreffende fylogeografie en populatiegenetica. Thank you Christophe for coming all the way from Iceland to give critical comments on this thesis. I hope this work can motivate you and others to continue with research on sand gobies in the future. Tom voor zijn interesse gedurende mijn volledige doctoraatsperiode en Joost voor zijn inbreng in het rodopsine verhaal, de statistiek en de verbeteringen van de manuscripten.

Mijn twee naaste collega's van eiland drie, ongelooflijk bedankt Els & Griet voor jullie grote hulp en opvang tijdens en na het werk.

Jef Guelinckx om het Flemish Goby Consortium uit te bouwen tot een goed werkende groep, inclusief het organiseren van vele staalname-campagnes, het verbeteren van talloze grondelmanuscripten en vooral het uitwisselen van kennis over het dikkopje.

Daarnaast dank ik ook al mijn overige collega's van het lab: Tine voor het delen van fylogenetische kennis, haar 'zandgrondel' stalen en haar sterke inbreng in hoofdstuk 5. Greg voor zijn aanstekelijke interesse in moleculaire ecologie. Bart en Sarah voor de zeer goede werking in het lab en het oplossen van alle moleculaire problemen. Erica for the technical assistance in the lab and support in Valencia. Anton, Dirk, Maarten, Nellie & Sara voor alle hulp in PC-problemen, bij genetische analyses en voor de inhoudelijke discussies. Special thanks also to Luisa to read and correct two chapters of my thesis.

Mijn thesisstudente Kim voor al haar DNA-werk en voor de uiterst goede samenwerking, en verder ook mijn twee bachelorstudentes Dorien & Eveline.

Many thanks to Dr. Kevin Ruddick (MUMM) for the excellent cooperation and the remote sensing analysis. Thanks to your contribution we could realize a very important progress in our research.

Veel dank aan Conny, Frans, Sinne, Eddy & Kim voor hun jarenlange hulp in administratie en IT.

Special thanks to Anne Christine Utne-Palm, Carin Magnhagen, Maria Järvi-Laturi, Kai Lindström, Lotta Kvarnemo, Katarzhyna Waligóra-Borek, David Rodgers, Ian McCarthy,

Mario Lepage, Maria Freitas, Alain Crivelli, Pilar Drake, Sofia Giakoumi, Alcibiades Economou, Henk van der Veer, Maarten Vanhove, Emilie Egea, Jan Breine, Els Cuveliers, Tine Huyse, Jef Guelinckx & Bram Vanschoenwinkel for providing many samples of *Pomatoschistus* sp.

Ik bedank al mijn collega's van de LADS en LAEE en al mijn vrienden voor de fantastische opvang bij alle moeilijke momenten tijdens het doctoraat (in totaal willekeurige volgorde): Gilles, Stefanie (Fie), Geert, Els (Cuviken), Jef (Jeffrie), Bert, Dino, Sara, Ine (de godin van 't tweede), Pieter (Bollie), Anton (Antonarctica), Kim, Jeroen (Casteels!), Dirk, Tom, Els, Anneleen, Mieke (Mie), Maarten, Mieke, Bert, Sarah, Steven (Soulmate), Joost, Emilie (good luck with your own PhD), iedereen aan de koffietafel, al de goeie vrienden en vriendinnen van UP en CDS, ...

Benedikt (Bik), Filip (Boeckie Boy), Ilse (Isch), Tom, David (Dé), Rebekka, Michiel, Pieterjan, Jonas, An-Sofie, Liselot, Joren, Evelien, Niels, Siemen & Lieselore voor hun oneindig aantal aanmoedigingen.

Uiteindelijk dien ik ook nog mijn pa en ma te bedanken die gewoonweg alles deden wat mogelijk was opdat dit proefschrift tot stand zou kunnen komen met zowel de sterkste inhoudelijke inbreng als de belangrijkste praktische hulp. Ik draag dit werk dan ook volledig aan hen beide op.

Maarten Larmuseau (Zandbergen, 14 juni 2009)

List of abbreviations

A1	A1 chromophore, the aldehyde of vitamin A1
A2	A2 chromophore, the aldehyde of vitamin A2
AA	Amino acid
Acc. nr.	Accession number in GenBank
AFLP	Amplified fragment length polymorphism
AIC	Akaike information criterion
AMOVA	Analysis of molecular variance
AO	Atlantic Ocean
B	Belgium
bp	Base pair
BSP	Bayesian skyline plot
c.	Circa
C	Cytoplasmic side of the cell membrane
CAGE	Cellulose acetate gel electrophoresis
CDOM	Colored dissolved organic matter
CI	Confidence intervals
C.I.	Consistency index
CMDS	Classical multi-dimensional scaling analysis
cyt *b*	Cytochrome *b*
D_{CE}	F_{ST}-linked pairwise genetic distances according Cavalli-Sforza & Edwards (1967)
d_N	Number of nonsynonymous substitutions per nonsynonymous site
D-test	Tajima's (1989) neutrality test
DNA	Deoxyribonucleic acid
d_S	Number of synonymous substitutions per synonymous site
$d\mu^2$	R_{ST}-linked pairwise genetic distances according Goldstein *et al.* (1995)
E	Extracellular side of the cell membrane
EG	Expansion growth
ESU	Evolutionary significant unit
FCA	Factorial correspondence analysis
FIASCO	Fast isolation by AFLP of sequences containing repeats protocol
F_S-test	Fu's (1997) neutrality test
F_{IS}	Inbreeding coefficient
F_{ST}	Fixation index
GPCR	G-protein-coupled receptor
GPI	Glucose-6-phosphate isomerase
h	Haplotype diversity
H_E	Expected heterozygosity
$H_{E\ n.b.}$	Unbiased expected heterozygosity
HKA	The test of Hudson, Kreitman & Aguadé (1987)
H_O	Observed heterozygosity
*Hsc*70	Heat-shock cognate protein
HWE	Hardy-Weinberg equilibrium
IAM	Infinite allele model
IBD	Isolation by distance
IB	Iberian Peninsula

IPTG	Isopropyl ß-D-1-thiogalactopyranoside
ITS	Internal transcribed spacer
k	Mean number of pairwise differences
K	Number of clusters calculated in Structure
Lap	Leucine aminopeptidase
LD	Linkage disequilibrium
Ldh	Lactate dehydrogenase
LGM	Last glacial maximum
Ln L	Log-likelihood value
LRT	Likelihood-ratio test
LWS/MWS	Red cone opsin
Ma	Million years ago
mAI_C	Mean assignment index
MCMC	Markoc chain Monte Carlo
miRNA	MicroRNA
ML	Maximum likelihood
MP	Maximum parsimony
MS	Mediterranean Sea
MSC	Messinian salinity crisis
MT	Methallothionein
mtDNA	Mitochondrial DNA
N	Number of individuals
n	Number of samples
NA	North Atlantic
Na	Not available
NAP	Non-algal particles
N_e	Effective population size
NJ	Neighbour-joining
NL	The Netherlands
NMDS	Non-metric multidimensional scaling analysis
Nr	Number of individuals wherefore the *RH1* gene is sequenced
Ns	Number of individuals wherefore microsatellites are genotyped
NTP	Nucleoside triphosphate
p	P-value
PA	Procrustes analysis
Pan I	Vesicular membrane protein
PCR	Polymerase chain reaction
QTL	Quantitative trait loci
r	Relatedness
r	Pearson or Spearman rank correlation coefficient
rag	Raggedness index
R.C.I.	Rescaled consistency index
rDNA	Ribosomal DNA
RFLP	Restriction fragment length polymorphism
R.I.	Retention index
RH1	Rod opsin
RH2	Green cone opsin
RNA	Ribonucleic acid

R_{ST}	Analogue of F_{ST} with taking account of the stepwise mutation model
S	Number of segregating sites
SDD	Sum of the squared deviations under the hypothesis of sudden expansion
SGE	Starch gel electrophoresis
SMM	Stepwise mutation model
Sn	Number of non-synonymous segregating sites
SNP	Single nucleotide polymorphism
SNPnon	All nonsynonymous SNPs
SNPsyn	All synonymous SNPs
SNPsyn+non	All SNPs of a gene fragment
Sqrt	Square root
Ss	Number of synonymous segregating sites
SWS1	UV-cone opsin
SWS2	Blue cone opsin
TBR	Tree bisection-reconnection branch swapping
Te	Absolute time since expansion
TM	Transmembrane helices
tMRCA	Time to the most recent common ancestor
tRNA	Transfer RNA
UK	United Kingdom
UV	Ultraviolet light
v.	Version
vAI_C	Variance of the assignment index
VP	Visual pigment
WMTL	Wavelength of maximally transmitted light
θ	Estimator of F_{ST}
Θ	Genetic variance parameter calculated per site
λ_{max}	Peak of maximal absorbance
π	Nucleotide diversity
ϱ	Estimator of R_{ST}
τ	Time since expansion measured in mutational time units
ω	Selection parameter

Table of content

Chapter 1

General introduction and outline of the thesis

'The Mediterranean Sea is a mackerel colour: in other words, changeable – you do not always know whether it is green or purple, you do not always know if it is blue, as the next moment the ever-changing sheen has assumed a pink or a gray tint'

Vincent Van Gogh (Dutch painter, 1853-1890) about his visit in Les Saintes-Marie-de-la-Mer, a French fishing village where he painted the effects of light on the sea.

Understanding the genetic basis of **local adaptation** is a prime interest of biology, as it involves the role of natural selection in promoting evolutionary change (Mayr 1963). The present knowledge about adaptive evolution for marine organisms is limited since population genetic surveys in these species have typically not applied genetic markers subject to selection (Hemmer-Hansen *et al.* 2007). Nevertheless, this knowledge is crucial to understand how evolution operates in the ocean (Conover *et al.* 2006).

Visual local adaptation should play an important evolutionary role in marine organisms. The water column of coastal habitats shows a range of optical characteristics which put special demands for visual predators or animals with a visually based mating system. Therefore, one of the few promising models to elucidate the mechanism and importance of selection as evolutionary force is the spectral tuning mechanism of **visual pigments** (VP) in vertebrates (Yokoyama 2000).

The aim of this thesis was to contribute to the actual research about the importance of natural selection as evolutionary force in marine organisms by studying the possibility and characteristics of visual local adaptation in marine vertebrates. A candidate gene approach was performed to demonstrate local adaptation at the rhodopsin gene, the VP which determines the spectral sensitivity of dim-light vision. The sand goby *Pomatoschistus minutus*, an abundant marine demersal fish along the European coasts, was selected as study species to realize the objectives of the thesis.

In this first chapter we introduce local adaptation in the marine environment and address the value of a candidate gene approach to quantify adaptation. The choice of the visual pigments of vertebrates as candidate genes is motivated, followed by an introduction about the study species the sand goby *P. minutus*. Finally, the objectives, the strategy and the outline of the thesis are addressed.

1. DETECTION OF LOCAL ADAPTATION IN MARINE ORGANISMS

Natural selection and evolutionary change

Natural selection is the process by which favorable heritable traits become more common in successive generations of reproducing organisms; unfavorable heritable traits become less common, due to differential reproduction of genotypes. Ancestors varied in their design and function, and the ones with the best designs passed on their genes in greater numbers and will have a higher Darwinian fitness (Freeman & Herron 2001). The **fitness** of an individual is defined as the relative contribution of its genotype to the next generation relative to the contributions of other genotypes (Begon *et al.* 2006). The phenotypic trait that has been favored by natural selection is known as an **adaptative trait** and can be identified by being variable, heritable and responsible for the variation in fitness (Howe & Brunner 2005). **Local adaptation** refers to the process when there is a greater fitness for individuals in their local habitats compared with the performance of immigrants due to natural selection (Carvalho 1993). Local adaptation may occur in species exposed to diverse abiotic or biotic environments – and may contribute to speciation (Howe & Brunner 2005).

Although mutation is responsible for the initial generation of the vast amounts of genetic variation that characterizes biological diversity, natural selection organizes this variation into functional individuals that are able to meet environmental challenges (Grant & Waples 2000). Yet powerful though the principle may be, natural selection is not the only cause of evolutionary change, and may, in many cases, be overtaken by other forces. Next to natural selection, the major evolutionary forces that can lead to differentiation in allele frequency are genetic drift and gene flow (Lenormand 2002). **Genetic drift** is a stochastic effect that arises from the role of random sampling in the production of offspring (Neigel 1997). Drift is fundamentally the result of finite population size and therefore most important in small populations. Migration leading to genetic homogenizing between partially isolated populations is called **gene flow** (Grant & Waples 2000). Gene flow will tend to homogenize

genetic variation among populations, whereas genetic drift and natural selection will increase the genetic differentiation between populations (Freeman & Herron 2001).

Local adaptation is essentially the result of the relative magnitude of the three main evolutionary forces. Pressure from natural selection promotes local adaptation. On the other hand, genetic drift and gene flow counteract natural selection, imposing a limit on local adaptation (Lenormand 2002). Genetic drift can wipe out adaptive polymorphisms just by random processes, especially in small populations (Freeman & Herron 2001). Gene flow among populations experiencing different selection regimes decreases local adaptations, depending on the strength of local selection relative to the rate at which non-adaptive alleles are being brought into the population through gene flow (Kingsolver *et al.* 2002). Thus, evolutionary scenarios with temporally stable neutral genetic structuring among populations could indicate that strong local selection would be able to override the effects of genetic drift and gene flow, resulting in adaptive population divergence, that is local adaptation (Grant & Waples 2000; Hemmer-Hansen *et al.* 2007). Therefore, understanding the genetic basis of local adaptation is of prime interest in biology as it involves the role of natural selection in promoting evolutionary change in contrast to neutral processes (Mayr 1963).

Is it possible to be locally adapted to the marine environment?

Traditionally, the genetic structure of marine organisms has been thought to be **homogeneous** due to the lack of obvious barriers to gene flow in the environment. Since gene flow is expected to hamper adaptive population divergence, the conventional idea was that local adaptation may be rare or absent in marine species although they are potentially affected by numerous selective pressures such as temperature and salinity (Guinand *et al.* 2004; Zane 2007).

In recent years, an increasing number of population genetic studies has described significant genetic structuring in several marine species, primarily with the aid of highly variable genetic markers such as microsatellites (Knutsen *et al.* 2003; Pampoulie *et al.* 2008). One major factor responsible for the present genetic structure of marine species is the geological and climatological history during the Pleistocene glaciations (Hoarau *et al.* 2007; Debes *et al.* 2008;

Luttikhuizen *et al.* 2008). Also contemporary factors maintain and promote genetic differentiation among marine populations on various geographical scales. The marine environment shows heterogeneity in response to climate, hydrodynamics and topography (Cowen *et al.* 2000), and biological traits, such as sex-dependent migration, site philopatry and assortative mating enhance genetic structuring (Ruzzante *et al.* 1998). Therefore the observed temporally stable **neutral genetic structuring** in several marine species may preclude adaptive divergence among populations and local adaptation to particular habitats (Conover *et al.* 2006). Moreover, the large effective population sizes of many marine species would tend to favor the effects of natural selection over the random effect of drift. However, any inference about the evolutionary significance of the results of neutral population studies in terms of adaptive population divergence have been based mostly on speculations regarding the potential for local adaptations to be present (Hemmer-Hansen *et al.* 2007).

Little is known about local adaptation in marine organisms since population genetic surveys in marine species have typically not applied genetic markers subject to selection (Hemmer-Hansen *et al.* 2007). Nevertheless, local adaptation in marine organisms has become increasingly documented over the past few years, indicating that selection may also be a **potent evolutionary force** in the marine environment (Canino *et al.* 2005; Hemmer-Hansen *et al.* 2007; Zane 2007; Sherman & Ayre 2008). Evidence for local adaptation in marine organisms is still scarce and knowledge of the spatial and temporal scale of adaptive genetic variation in marine systems remains limited, yet crucial to improve our understanding of how evolution operates in the ocean (Conover *et al.* 2006). Moreover, knowledge on local adaptation in marine organisms is crucial in order to predict if depleted or extinct populations can be effectively replaced by recolonization from other populations. Are local populations so locally adapted that it is unlikely that they can be replaced on a historic time-scale by individuals from other populations of the species (ICES 2005; Hauser & Carvalho 2008). Therefore, more research on the characteristics of local adaptation in marine organisms is highly recommended.

2. THE CANDIDATE GENE APPROACH IN MARINE ORGANISMS

Detecting local adaptation in marine organisms

To detect local adaptation in the wild, the demonstration of **three classical steps** is necessary. It has been theoretically described in Endler (1986) as follows:

Step 1: demonstrating that the populations differ for a **heritable trait**.

Step 2: demonstrating that the population differences are due to selection as **opposed to neutral processes**.

Step 3: establishing a link between functional variation and **selection regimes**.

Based on the requirement to fulfill all three steps, the documentation of variation in phenotypically plastic traits specific to distinct environments can not be taken as the ultimate proof that natural selection has occurred. **Molecular tools** may help to overcome this problem and, in certain circumstances, record the footprints of selection (Guinand *et al.* 2004). Molecular approaches of 'selection detection' should be able to link patterns of selection at particular loci to environmental features. This is the basic principle of a method that is called the '**candidate gene approach**', which has the potential to fulfill all three steps to demonstrate local adaptation even in non-model organisms (organisms with low genomic knowledge). If genetic variation is studied at candidate genes that are supposed to be coding genes with an ecologically important function, then this can provide evidence for local selection and thereby local adaptation (Ford 2002). Therefore it makes the candidate gene approach an obvious means to study natural selection in the field.

It is important to distinguish 'selection detection' methods from **'gene hunting' molecular tools**. These latter include tools to detect specific loci or genomic regions for which the neutral hypothesis is rejected. However, they will not provide evidence for natural selection and local adaptation at these sites (Volis 2008). In other words, step 2 to demonstrate local adaptation is fulfilled with 'gene hunting' tools, but not step 1 and 3. Nevertheless, these methods can provide important information on questions such as how much of the genome

is actually subjected to local selection (Storz 2005). Moreover, they can also lead to identifying functionally important variation. Rapid progress in genomics and related fields has resulted in a wave of new high technological methods and strategies to identify candidate genes, e.g. genome scans, quantitative trait loci (QTL) and micro-array approach (Vasemägi & Primmer 2005). Today, those methods are widely used in marine organisms to enhance the efficiency in the identification of candidate genes:

* **Genomic scans** (or multiple-marker based neutrality tests) denote analyses of a large number of genetic markers randomly distributed throughout the genome. If genomic scans are conducted for different populations, then the degree of genetic differentiation can be estimated for each individual marker, and it can be assessed if some of the markers are 'outliers' and thereby potentially subject to local selection (Storz 2005). The combination of multiple marker information enables distinguishing locus-specific effects, such as selection, from the genome-wide effects, as random genetic drift and gene flow (Vasemägi & Primmer 2005). Different genomic scans are already realized on marine species, which provided candidate genes that need to be validated (Oetjen & Reusch 2007; Wood *et al.* 2008; Galindo *et al.* 2009).

* **QTL analyses** (also known as linkage mapping) identify particular regions of the genome that are associated with a trait being assessed or measured within a pedigree (family). Accordingly, QTLs can be used for finding indications for local adaptation on a trait but not to prove it (Vasemägi & Primmer 2005). QTL analyses are increasingly used in marine species e.g. seabream (*Sparus auratus*) and sea bass (*Dicentrarchus labrax*) (Franch *et al.* 2006; Sarropoulou *et al.* 2008).

* If studies of gene expression variation in natural populations using **DNA microarrays** represent an interesting tool, they do not prove that observed individual variation is based on changes at the DNA level (sequence variation leading to distinct alleles distributed within and among populations) (Guinand *et al.* 2004; Roberge *et al.* 2007). Nevertheless, Larsen *et al.* (2007) showed the usefulness of microarrays in marine non-model organisms to identify regulatory genes which control the production of proteins at the centre of various key physiological pathways and which are potentially subjected under natural selection.

Methods in candidate gene approach

An extensive and multidisciplinary analysis is required to fulfill all three steps demonstrating local adaptation at a particular gene in the wild by using the candidate approach.

First, the direct link between genetic and phenotypic variation for a specific trait under study has to be revealed. This is a crucial step since natural selection only operates when there are phenotypic differences between individuals with a heritable basis (Howe & Brunner 2005). To realize the link, physiological and biochemical assays that identified amino acid (AA) changes responsible for phenotypic variation, need to be constructed (Yokoyama 2000). After that, population differentiation on the functional variation for the particular trait has to be observed to fulfill **step 1** (Endler 1986).

Second, to demonstrate that the population differences are due to selection (**step 2**), two classical categories of methods to detect signatures of positive selection are used in the candidate gene approach: sequenced-based neutrality tests and comparative tests with a neutral baseline.

* DNA sequence data of the candidate gene have been extensively used to test the neutral null hypothesis to infer the evidence of selection using a wide range of **sequenced-based neutrality tests** (Vasemägi & Primmer 2005). One category of statistical methods is based on allele frequency distributions and level of variability, and include Ewens-Watterson test, HKA test, Tajima's D-test and their later developments (Tajima 1989; Fu 1997). However, those tests had very limited success in providing unambiguous evidence for selection due to strong assumptions regarding the demography of the populations (Nielsen 2001; Ryynänen & Primmer 2004). In contrast, a second group of methods has been successful in providing evidence for selection and is based on comparisons between different classes of mutations within a locus and includes tests based on the nonsynonymous and synonymous substitutions ratio (d_N/d_S) (Yang & Bielawski 2000; Nielsen 2001). Estimates of the average nonsynonymous and synonymous substitution ratio between two sequences has been used to infer whether the particular gene has been under stabilizing or purifying selection ($d_N < d_S$), neutral evolution ($d_N = d_S$) or positive selection ($d_N > d_S$) (Nei & Gojobori 1986). As the

majority of amino acid sites are highly conserved and only few substitutions are expected to enhance the function of the protein, the $d_N > d_S$ calculated over the whole gene provides an extremely stringent criterion for inferring the presence of positive selection. Today, more elaborate approaches have been developed to infer lineage-specific episodes of positive selection from multiple sequences in a phylogenetic framework and to identify specific regions or even single sites under positive selection (Yang *et al.* 2000a).

* On the other hand, it is possible to separate the effects of natural selection from neutral (demographic) effects by comparing population differentiation in markers presumably under selection with a **'neutral baseline'** generated from neutral markers in these species (Hemmer-Hansen *et al.* 2007). Given the inherent difficulty of comparing levels of structuring across different marker types, the evidence of selection is considerably strengthened by the finding of diverging patterns as well as diverging levels of structuring (Lemaire *et al.* 2000).

Finally, to establish a direct link between the functional variation and a selection regime (**step 3**), it is necessary to discover the function of the gene in physiological assays and to link this variation to a specific environmental condition. This is a crucial step in demonstrating local adaptation on a particular gene due to the possibility of linkage disequilibrium between the marker loci and those directly under selection, by a process that is known as **'hitch-hiking'** (Carvalho 1993). The strategy to provide evidence for the fitness relevance of the functional variation of a candidate gene is very locus-dependent and therefore often the most difficult task to fulfill in the candidate gene approach (Endler 1986; Nager *et al.* 2000).

The candidate gene approach in marine ecology

Marine organisms have not been often the subject of studies on population differences at candidate genes believed to be directly affected by selection (Guinand *et al.* 2004). Moreover, although the high potential of this approach, almost all performed candidate gene studies on marine species did not succeed to fulfill all three steps to demonstrate local adaptation (Table 1.1).

Table 1.1 List of published candidate gene studies in obligate marine organisms (excluding anadromous and catadromous fishes). For each analyzed candidate gene the steps to demonstrate local adaptation that are fulfilled and the used type of the test that fulfilled step 2 are given.

Genes	Taxa	Steps	Tests for step 2	References
Bindin	Sea urchins *Echinometra mathaei*	1 + 2	d_N/d_S tests	Metz & Palumbi 1996; Palumbi 1999
Egg-laying hormone	Sea snake *Aplysia californica*	2	d_N/d_S tests	Endo *et al.* 1996
Glucose-6-phosphate isomerase (*GPI*)	Lozano's goby *Pomatoschistus lozanoi*	1 (p) + 2 (p)	Neutral baseline*	Gysels *et al.* 2004b
	Common goby *Pomatoschistus microps*	1 (p) + 3 (p)	"	Al-Hassan *et al.* 1987
Heat-shock cognate protein (*Hsc70*)	European flounder *Platichthys flesus*	1 (p) + 2	Neutral baseline	Hemmer-Hansen *et al.* 2007
Hemoglobin	Atlantic cod *Gadus morhua*	1 (p) + 3 (p)	"	Brix *et al.* 1998
	Turbot *Scophthalmus maximus*	1 (p) + 3 (p)	"	Imsland *et al.* 2000
Ldh B gene	Killifish *Fundulus heteroclitus*	1 + 2 (p) + 3	HKA tests	Schulte *et al.* 1997; Schulte *et al.* 2000
Ldh C gene	Sand goby *Pomatoschistus minutus*	1 (p) + 2 (p)	Neutral baseline*	Gysels *et al.* 2004b
Leucine aminopeptidase (*Lap*)	Marine mussel *Mytilus edulis*	1 + 2 (p) + 3	Neutral baseline*	Hilbish & Koehn 1985
Metallothionein (*MT*)	Common sole *Solea solea*	1 (p) + 2 (p)	Allele frequency distribution	Guinand *et al.* 2008
Vesicular membrane protein (*Pan* I)	Atlantic cod *Gadus morhua*	1 (p) + 2	Neutral baseline	Pogson & Fevolden 2003; Pampoulie *et al.* 2006
	Walley pollock *Theragra chalcogramma*	1 (p) + 2	Neutral baseline	Canino *et al.* 2005

(p) = step partially fulfilled; 'Neutral baseline' = comparison with neutral baseline; 'HKA test' = neutrality test based on the comparisons between nonsynonymous and synonymous substitutions within and between related species (Hudson *et al.* 1987); *, comparison with a baseline of <7 allozyme markers wherefore the neutrality could not be fully guaranteed.

The difficulty in **step 1** is to demonstrate a direct link between genotypic with phenotypic variation. Most studies listed in Table 1.1 demonstrated population differentiation on a marker linked to a candidate gene without proof of functional differences on the gene with phenotypic consequences (e.g. Guinand *et al.* 2004; Gysels *et al.* 2004c; Hemmer-Hansen *et al.* 2007). On the other hand, based on extensive functional assays a clear link could be established between genotypic and phenotypic variation for the bindin protein in sea urchins (Palumbi 1999), *Ldh*B protein in killifish (Schulte *et al.* 2000) and the *Lap* protein in the blue mussels (Hilbish & Koehn 1985).

To fulfill **step 2** in the approach, the most reliable evidence was given by d_N/d_S ratios on the egg-laying hormone and bindin gene (Endo *et al.* 1996; Metz & Palumbi 1996), but positive results are rare due to the conservative character of the test. There is a great potential for comparing candidate genes *versus* presumably neutral variation as done for the *Pan*I locus in

Gadus morhua and *Theragra chalcogramma* (Canino *et al.* 2005; Pampoulie *et al.* 2006), and the *Hsc70* locus in *Platichthys flesus* (Hemmer-Hansen *et al.* 2007). However, the use of different marker types in these studies introduced statistical and interpretational problems, especially with the comparison of the levels of structuring. Therefore, it is suggested that combining different methods will increase the evidence that the population differences on functional variation are due to selection as opposed to neutral processes.

In the few studies published on marine species, the function of the genes in question and/or the links between the traits and selection regimes were mostly obscure (**step 3**) (Guinand *et al.* 2004). For example, it has not been possible to link selection at the *Pan*I gene or *Hsc70* with any environmental factor, because little is known about the function of these loci in fishes (Pampoulie *et al.* 2006; Hemmer-Hansen *et al.* 2007). There are only two well known examples where a clear link was found between the genetic differentiation on a candidate gene and a specific selection regime in the marine environment, namely on the leucine aminopeptidase (*Lap*) gene in the marine mussel *Mytilus edulis* and *Ldh*B gene in killifish, *Fundulus heteroclitus*. *Lap* catalyses protein degradation and thus the production of amino acids. Different allozyme variants of *Lap* in *M. edulis* correlate clearly with salinity, and their frequencies in mussel populations respond within months to salinity changes in coastal regions (Hilbish & Koehn 1985). Next, an observed latitudinal cline of *Ldh*B alleles and expression of the *Ldh*B gene in *F. heteroclitus* accounts for metabolic differences between individuals that significantly affect their fitness under the various temperature regimes in the ocean along the eastern side of North America (Schulte *et al.* 2000; Beebee & Rowe 2004).

Although almost all these candidate gene studies on marine systems could not provide evidence for local adaptation, these studies highlight the usefulness of the candidate gene approach for demonstrating local adaptation in non-model organisms such as most marine organisms (Zane 2007).

Visual pigments as potential candidate genes for marine organisms

The modest list of studies in Table 1.1 accentuates that remarkably little evidence is found at this moment for local adaptation at the genetic level in marine organisms. The reason for the lack of knowledge about adaptive evolution in the ocean is the scarcity of genetic systems to prove natural selection (Yokoyama 2002). Therefore, there is need to select good candidate genes which has the potential to demonstrate local adaptation in the wild (Volis 2008).

One of the few promising models to elucidate the mechanisms of adaptive evolution and the importance of selection as evolutionary force is the spectral tuning mechanism of **visual pigments** (VP) in vertebrates (Yokoyama 2002). In VPs there is a direct link between genotypic and phenotypic variation due to the possibility of measuring the phenotypic effects of variation on the VPs genes by mutagenic experiments (step 1). Moreover, VPs have a well-defined role in nature as organisms detect differences in the spectral composition of the environment (Bowmaker 2008). The adaptive significance of visual traits is therefore obvious: in many animal species, individuals with good eyesight will be advantaged to find food and mates and to avoid predators than individuals with poor eyesight (step 3) (Freeman & Herron 2001). They have a strong effect on the evolution of organisms, providing an excellent system to study adaptive evolution at the molecular level. For all these reasons, opsin genes serve as a prime model to elucidate the mechanisms of adaptive evolution and the importance of selection as evolutionary force in vertebrates.

3. THE VISUAL TUNING SYSTEM OF VERTEBRATES

Visual pigments

In the animal world, there is a remarkable diversity of light **sensitive organs and tissues**, from simple photoreceptor organs in invertebrates to the compound eye of insects and the vertebrate camera eye (Shubin 2008). In vertebrates, a ray of light entering the eye passes through several relatively transparent elements to reach the retina: these elements include a thin film of tear-water, the cornea, the aqueous humour, the lens, and finally the vitreous

humour. The retina is a thin sheet of tissue that lines the back of the eye and contains the light-sensitive cells, namely the rod and cone cells (Connors 2003). The photoreceptor cells capture photons and this initiates a cascade of molecular events, resulting in the transduction of a signal through the optic nerve to the brain for higher-order processing (Connors 2003).

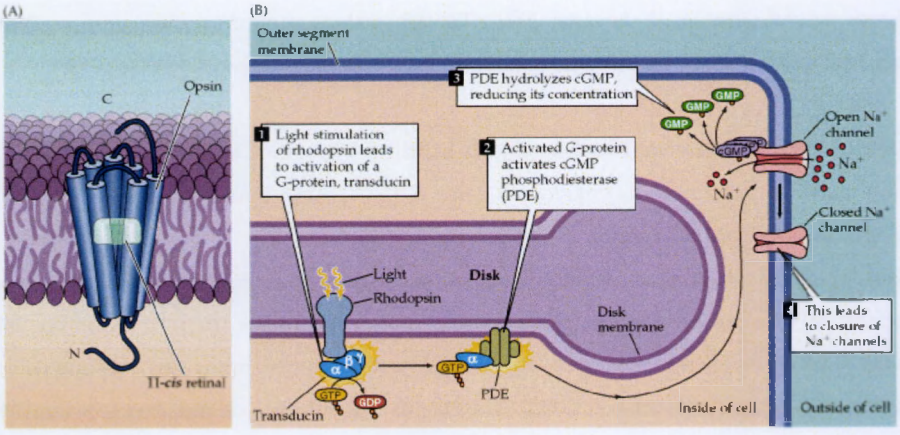


Figure 1.1 (a) Visual pigment with the opsin protein bound to the chromophore 11-*cis* retinal (A1 chromophore); (b) Overview of the cascade of molecular events when light activates the opsin receptor (Connors 2003).

Vision starts when photons are absorbed by **visual pigment (VP) molecules**, bound in the dense membrane of the rod and cone cells (Figure 1.1a). VPs are members of the protein superfamily known as G-protein-coupled receptors (GPCRs). GPCRs are defined by their heptahelical transmembrane structure and their ability to activate a GTP-binding protein (G-protein). In vertebrates, the family of GPCRs is one of the largest and most diverse protein families, with thousands of members predicted in several mammalian genomes (Connors 2003). GPCR function ranges from hormone and neurotransmitter detection, to sensory system receptors such as visual, gustatory and olfactory receptors (Park *et al.* 2008). The VP, as a subclass of GPCRs, consists of a protein moiety, the opsin, that has the ability to bind a light-sensitive molecule, the vitamin A-derived chromophore (Nickle & Robinson 2007). In vertebrate visual pigments, two different types of chromophores have been found. The most common is 11-*cis*-retinal (A1 chromophore), which is the aldehyde of vitamin A1. Many

fishes, amphibians and reptiles have also 11-cis-3,4-dehydroretinal (A2 chromophore, the aldehyde of vitamin A2). The chromophore is a light-sensitive molecule that changes his conformation by the absorption of a photon (Fig. 1.2). Due to this conformation change, the chromophore confers light sensitivity to the photopigment; its absorption of light triggers a transition in the photopigment from the inactive dark state to an active state (Fig. 1.1b) (Yokoyama 2000).

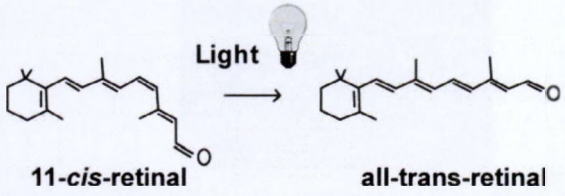


Figure 1.2 Absorption of a photon by 11-*cis*-retinal (A1 chromophore) causes the molecule to isomerize to all-*trans*-retinal (Connors 2003).

Each pigment shows a characteristic peak of **maximal absorbance (λ_{max})**. The precise location of this peak depends on the interactions between the chromophore and the opsin protein (Connors 2003). Based on their specific λ_{max} values and their amino acid compositions, the VPs are classified into five evolutionary distinct clusters (Yokoyama 2000; Nickle & Robinson 2007):

(i) **RH1 group** (**rod opsin**, consisting of mostly rhodopsins with λ_{max} values at about 500 nm);

(ii) **RH2 group** (**green cone opsin**, a mixture of rhodopsin-like pigments with λ_{max} values at 470-510 nm);

(iii) **SWS1 group** (**UV-cone opsins**, short wavelength- or blue-sensitive pigments with λ_{max} values at 360-430 nm);

(iv) **SWS2 group** (**blue cone opsin**, SWS1-like pigments with λ_{max} values at 440-460 nm);

(v) **LWS/MWS group** (**red cone opsin**, long wavelength- or middle wavelength-sensitive, or red- and green-sensitive pigments at 510-560 nm).

The rod opsin gene lineage *RH1* is expressed in the rod photoreceptors and produces monochromatic dim light vision. The cone opsin lineages are expressed in cone cells and are responsible for bright light colour vision (Trezise & Collin 2005).

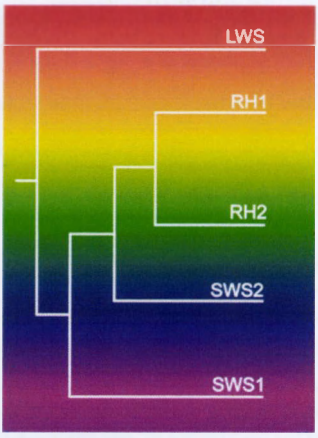


Figure 1.3 Phylogenetic relationships between the five major opsin gene lineages. A series of four gene duplication events progressively produced the vertebrate *LWS*, *SWS1*, *SWS2*, *RH2* and *RH1* opsin genes. The position of each branch on the background spectrum approximates the spectral sensitivity of each opsin: *LWS* (red cone); *SWS1* (UV-cone); *SWS2* (blue cone); *RH2* (green cone) and *RH1* (rod) (Trezise & Collin 2005).

Following the divergence of the ancestral vertebrate and invertebrate species, about 700 million years ago, the ancestral opsin gene underwent a series of independent duplication and diversification events. The gene duplications followed by amino acid substitutions are the basis for the functional differentiation of the five lineages of VP (Figure 1.3) (Yokoyama 2000). The vertebrate *RH1* and *RH2* genes were produced by duplication of the ancestral RH opsin gene approximately 500 million years ago, after the separation of the jawed (sharks and rays) and the jawless (lampreys and hagfishes) vertebrates. Therefore, colour vision clearly evolved first in the ancestral jawless vertebrates, with dim light vision appearing only after the evolution of the jawed vertebrates (Trezise & Collin 2005). Within each pigment group, further functional differentiations and sometimes additional opsin gene duplications or even gene loss are also detected, as it occur in a high frequency in cichlids (Parry *et al.* 2005).

Spectral tuning mechanism of marine vertebrates

The marine environment provides a **wide range of photic habitats**. Shallow water receives a wide range of wavelengths of light from the sun, but as we go deeper, the light is filtered by the water. Water acts as a monochromator, absorbing both long- and short-wave light, with the maximum transmission of pure water located in the blue region of the spectrum at approximately 460 nm (Figure 1.4) (Jerlov 1976). However, the marine environment is never pure and contains many contaminants such as suspended particles that will scatter short wavelengths, and dissolved substances that may colour the water, such as dissolved organic material staining the water yellow or brown, and phytoplankton containing chlorophyll. Consequently, seas are varying in many aspects such as turbidity, colour and brightness (Kirk 1996).

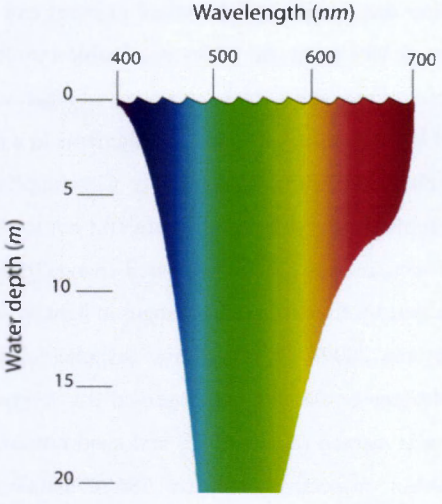


Figure 1.4 The visual spectrum ranges from 400-700 nm at the coastal water surface, but downwelling sunlight loses both long- and short-wave components rapidly within 10-15 m (Michiels *et al.* 2008).

Animals have evolved their visual sensitivity to match aspects of the photic environment of their habitat by a process that is called the **spectral tuning**. A clear correlation between the λ_{max} of the VPs and the photic characteristics of the habitat depth or habitat behaviour of their possessors is observed in several marine species (Crescitelli *et al.* 1985). Some of the

most dramatic examples have been found in fish rod photoreceptors (Spady *et al.* 2005). Extensive survey data show that the λ_{max} values of rhodopsins of mesopelagic marine fishes are more blue-shifted than those of shallow-water species (Yokoyama & Takenaka 2004). Similarly, the bottlenose dolphin (*Tursiops truncatus*) (Fasick & Robinson 1998) and the coelacanth (*Latimeria chalumnae*) (Yokoyama *et al.* 1999) have rod cells in which the peak sensitivity of the rhodopsins is blue-shifted in comparison with shallow-living relative species.

In vertebrates, spectral tuning of the pigment can be achieved on a physiological time scale by exchanging the nature of the chromophore (A1 and A2) or on an evolutionary time scale by amino acid substitutions in the protein part, the opsin (Bowmaker & Hunt 2006):

* **Physiological adaptation** to alter the function of the visual pigment can be realized by replacing A1 by A2 chromophores, or vice versa. An additional double carbon-carbon bond in the β-ionone ring of the A2 chromophore shifts the λ_{max} value to a longer wavelength. The exact λ_{max} of the VP is determined by the ratio of A1- and A2-pigments in a mixture, giving the organism the luxury of being able to shift its visual sensitivity. Chromophore changes are apparently due to regulation of chromophore synthesis enzymes and not to dietary changes. The increase of the A2-pigments is consistent with an anticipated change for a more reddish environment and can be brought about by environmental changes in light, season, migration, temperature and hormone (Yokoyama 2000). Chromophores switches also occur during migrations from freshwater to saltwater or *vice versa*, as observed for several catadromous fishes such as American eels (*Anguilla rostrata*) (Beatty 1975) and anadromous fishes such as several salmon species (Salmonidae) respectively (Beatty 1966). Finally, physiological adaptations can also be associated to different visual tasks as males and 'sneaker' males of the peacock blenny (*Salaria pavo*) differ in their A1-A2 chromophore ratio (White *et al.* 2004). As a general rule, visual pigments of marine species are dominated by A1-pigments, whereas those of freshwater fish possess only A2 or both A1 and A2 as their chromophores (Toyama *et al.* 2008).

* **Evolutionary adaptation** can be realized by amino acid substitutions in the protein part of the VP, the opsin. To date, amino acid changes at a total of 25 sites are known to be

involved in the spectral tuning of the VPs. These sites are generally known as '**spectral tuning sites**'. Most of these sites are located within or near the retinal-binding pocket (Nickle & Robinson 2007; Bowmaker 2008). However, also amino acids that are distributed elsewhere in the pigment, as amino acids in the C-terminus, are likely to interact with the chromophore (Yokoyama *et al.* 2007). The effects of substitutions at many of the potential tuning sites have been characterized through site-directed mutagenesis studies (Yokoyama *et al.* 1995). The magnitude of the λ_{max}-shifts caused by identical amino acid changes can differ significantly between opsin lineages and between species (Fasick & Robinson 1998; Yokoyama *et al.* 1999).

Intraspecific evolutionary adaptation in marine species?

Several phylogenetic studies were meant already to detect evolutionary adaptation in a range of marine vertebrates (Hunt *et al.* 2001; Yokoyama & Takenaka 2004). On the other hand, little work has been done on incipient evolutionary adaptation between populations of a single vertebrate species (Jokela *et al.* 2003). Not any single observation of **intraspecific evolutionary adaptation** has been found for marine vertebrates, although, it is thought to occur commonly. Many marine fishes inhabiting spectrally different waters do not have the possibility to adapt physiologically by chromophore change because they only possess A1 chromophores (Bowmaker 1995).

One clear indication for intraspecific evolutionary adaptation was found for a common marine fish, the sand goby ***Pomatoschistus minutus*** (Pallas 1770; Gobiidae, Teleostei) (Fig. 1.5). In a microspectrophotometric study of Jokela *et al.* (2003), the absorbance spectra in retinal rods were measured for various sand goby populations living in a wide range of photic environments. Considerable variation in λ_{max} values was found within and between populations, especially between Baltic and Atlantic populations (λ_{max} in English Channel = 506.2±0.3 nm and λ_{max} in northern Baltic Sea = 508.3±0.5 nm). The shapes of the absorbance spectra indicated polymorphism at the rhodopsin (*RH1*) gene rather than admixture of A1 and A2 chromophores or an extra rhodopsin gene (Jokela *et al.* 2003). Consequently, evolutionary adaptation, rather than physiological change, is presumed to be

responsible for spectral tuning in the sand goby. *P. minutus* seems to be a good model species to study the possibility of visual local adaptation in the marine environment.

Fig. 1.5 A male of the sand goby *Pomatoschistus minutus (Pallas, 1770) (©Vilda)*

4. THE SAND GOBY *POMATOSCHISTUS MINUTUS*

Pomatoschistus minutus was selected as study target due to the good knowledge of its biology, ecology and genetics justifies a study on intraspecific evolutionary adaptation of the rhodopsin gene. For many years the sand goby has a **model species** status in studies of sexual behavior (Pampoulie *et al.* 2004b; Lindström *et al.* 2006; Singer *et al.* 2006), ecology (Pasquaud *et al.* 2004; Ehrenberg *et al.* 2005; Ehrenberg & Ejdung 2008), estuarine migration (Maes *et al.* 2005; Guelinckx *et al.* 2008), parasite ecology (Huyse 2002; Zander 2005) and phylogeography and population genetics (Stefanni & Thorley 2003; Gysels *et al.* 2004b; Pampoulie *et al.* 2004a). Its phylogenetic status in the 'sand goby' group is well known (Webb 1980; McKay & Miller 1997; Huyse *et al.* 2004). Knowledge gained from all these studies gives an opportunity to demonstrate successfully (visual) local adaptation.

The 'sand goby' group

The **Gobiidae** (Actinopterygii, Teleostei, Perciformes) represent one of the most diverse families of fish, occupying marine, brackish and freshwater habitats in the tropical and temperate seas of the world (Hoese 1984; Miller 1986). Gobies are relatively small, demersal fish characterized by a typical elongated cylindrical body form and dorsolateral protruding

eyes. The pelvic fins are thoracic and fused, forming an adhesive disk. This allows them to remain stationary on the substrate in relatively strong currents (Miller 1986; Hamerlynck 1993). That the morphological concept of the Gobiidae is successful is justified by its status as the largest family of marine fishes (Nelson 2006). It comprises about 200 genera and 1875 species (www.fishbase.org). Due to the huge morphological and ecological diversity in rapidly diversifying lineages, the family Gobiidae is recognized as an example of marine adaptive radiation (Ruber *et al.* 2003; Taylor & Hellberg 2005).

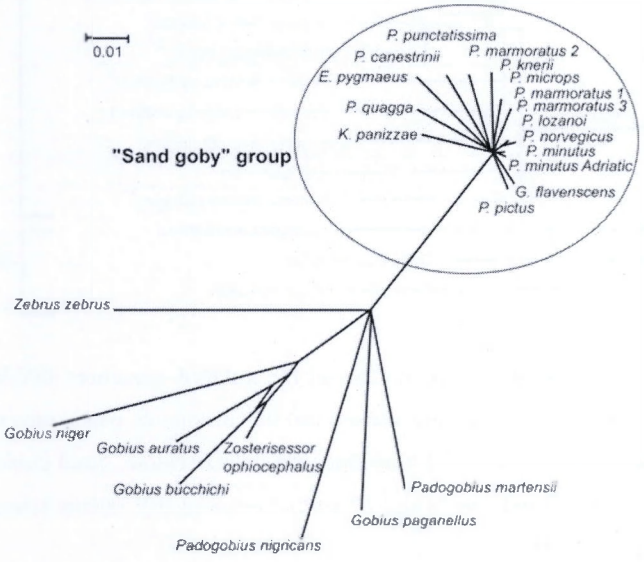


Fig. 1.6 Splits graph obtained from the 12S and 16S mtDNA sequences (750 bp) of the 'sand goby' group and related gobiids (after Huyse *et al.* 2004).

At least 52 species of Gobiidae are recorded from the north-eastern Atlantic and Mediterranean coasts. Most of these gobies belong to a so-called **'sand goby' group**, consisting of four genera with 12 described *Pomatoschistus* (Gill) species, one *Gobiusculus* (Duncker) species, 16 *Knipowitschia* (Ljin) species and two *Economidichthys* (Bianco, Bullock, Miller and Roubal) species (www.fishbase.org)(Miller 2009). The evolutionary gap between this 'sand goby' group and the other Atlantic-Mediterranean gobies is substantial (Fig. 1.6) (Huyse *et al.* 2004). Based on the morphologic and genetic analyses, the most likely sister group of the 'sand goby' group is the Indo-Pacific genus *Nesogobius* or *Tridentiger*. Closure of the Atlantic-Mediterranean part of the early Tethys (*c.* 15-14 Ma; at the end of the early

Miocene) is most likely the major vicariant event that separated the 'sand goby' from the *Nesogobius-Tridentiger* stocks (McKay & Miller 1997; Simonovic 1999; Huyse *et al.* 2004).

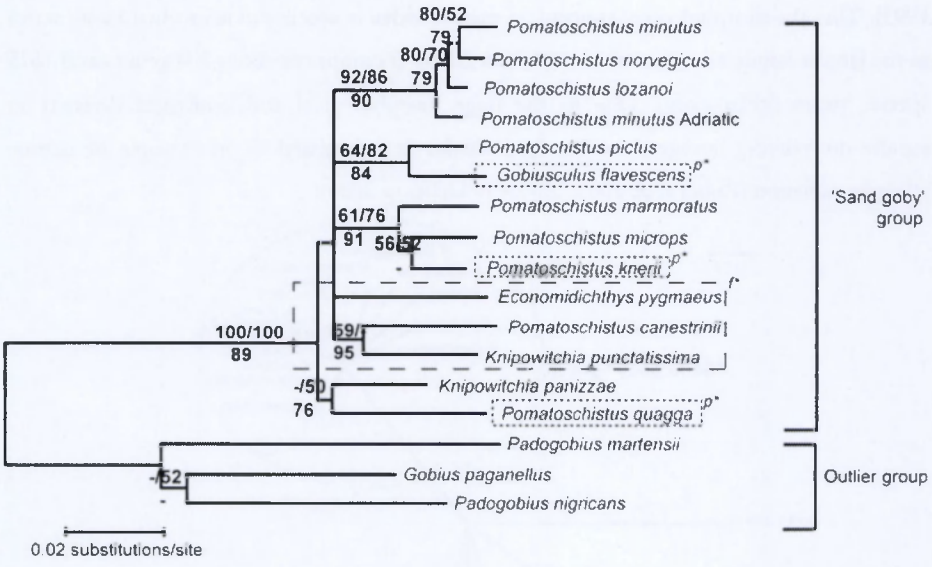


Fig. 1.7 Maximum-likelihood phylogram of 12S and 16S mtDNA sequences (800 bp) of the 'sand goby' group. Bootstrap values are shown for the maximum parsimony/neighbor-joining analyses above the branches and posterior probabilities below. 'Sand gobies' with a freshwater and a pelagic life style are boxed (*f** and *p** respectively); salinity tolerances are indicated (after Huyse *et al.* 2004).

Based on the phylogeny constructed with nuclear DNA (ITS1 locus) and mitochondrial DNA (12S and 16S fragments), the 'sand gobies' clustered **monophyletically**, as proposed on morphological grounds (Miller 1986; Huyse *et al.* 2004). *Gobiusculus flavescens*, *Economidichthys pygmaeus* and *Knipowitschia punctatissima* clustered within the *Pomatoschistus* species, pointing to a paraphyletic origin of these genera within the 'sand gobies' (Fig. 1.7). The 'star' phylogeny of the whole 'sand goby' group (Fig. 1.6) might suggest that these gobies evolved in a very short time period, most likely linked to the drastic alterations in the Mediterranean Sea during and immediately after the Messinian Salinity Crisis (MSC) at the end of the Miocene (Fig. 1.8). The MSC refers to the desiccation of the Mediterranean Sea – between 5.96 and 5.33 Ma - leading to the origin of several hyper- and hyposaline lakes with a transition to the 'Lago Mare' (sea-sized lake) (Banarescu 1992; Duggen *et al.* 2003). It is

hypothesized that the freshwater life-style that appeared to have a monophyletic origin in the group (Fig. 1.7) originated during this salinity crisis (Huyse *et al.* 2004). Moreover, the many endemic 'sand goby' species in the Mediterranean Sea are most likely linked to sympatric speciation events after the Pliocene reflooding when newly and freely isolated Mediterranean niches became available (Por & Dimentman 1985).

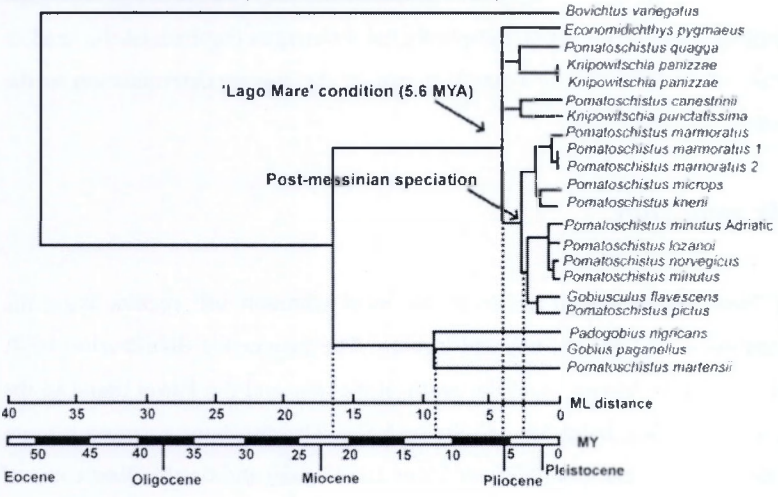


Fig. 1.8 Clock-constrained maximum likelihood tree constructed of 12S and 16S mtDNA sequences (800 bp) of the 'sand goby' group (after Huyse *et al.* 2004). The scale bars show the patristic distances and the time scale based on the assumption of the molecular clock of Huyse *et al.* (2004).

The shortest branch length in the phylogeny was found within the ***'Pomatoschistus minutus* complex'**, including *P. minutus*, the Adriatic *P. minutus*, *P. lozanoi* and *P. norvegicus* (Fig. 1.7) (Huyse *et al.* 2004). The genetic divergence between *P. minutus* from the Adriatic Sea versus the Atlantic-Mediterranean region was as high as the divergence between Atlantic *P. minutus* and *P. lozanoi*, suggesting that *P. minutus* from the Adriatic Sea should be considered as a distinct cryptic species in the *Pomatoschistus* genus (Wallis & Beardmore 1983; Stefanni & Thorley 2003; Gysels *et al.* 2004b; Huyse *et al.* 2004). The origin of the '*P. minutus* complex' dated back to the glacial cycling during the Early Pleistocene around 1.94-1.18 Ma (Fig. 1.8) (Huyse *et al.* 2004). The complex is always been thought to have speciated recently because of the observed hybridization between *P. minutus* and *P. lozanoi* in captivity (Fonds 1973) as in nature (Wallis & Beardmore 1980), and between *P. norvegicus* and *P. lozanoi* (Webb

1980). Current identification keys for these cryptic species are based on the pattern of the sensory papillae (Miller 1986) and pigmentation patterns (Hamerlynck 1990). Despite the morphological similarity, the genetic and karyological differentiations between the species are substantial, with differences in the number of chromosomes (*P. minutus* = 46 chromosomes; *P. lozanoi* = 37 chromosomes; *P. norvegicus* = 32 chromosomes) (Webb 1980). The high genetic divergences in contrast to cryptic morphological differences emphasized the need to consider molecular characterization as a standard part of the species determination in the 'sand goby' group.

Biology of the sand goby

The sand goby *Pomatoschistus minutus* is one of the most common fish species along the European Atlantic coast and its estuaries and lagoons. The geographic **distribution** of *P. minutus* includes the eastern Atlantic from the north of Norway and the Faroe Island to the south of Spain, the North Sea, Baltic Sea and the Irish Sea. The distribution pattern is more fragmented in the Mediterranean Sea (Gulfs of Lions and Genoa) and on the West coast of the Black Sea (Miller 1986). The precise geographical distribution of this fish is difficult to define because it has been often mistaken for other goby species (Hamerlynck 1993).

The sand goby prefers sandy and muddy sediments, up to depths of about 70 m (Miller 1986). The small fish (adult length 5-10 cm) is perfectly disguised with its semi-transparent body with orange and rusty chromatophores. During spawning time, the males have 5-10 dark-grey transverse stripes on the side of the body and blue coloration in the anal and dorsal fin (Fig. 1.5). The sexes also differ in the shape of the genital papilla, which is elongated and pointed in the male, short and blunt in the female (Rodrigues *et al.* 2006).

Due to its short lifetime, most of the sand goby have mainly one single **reproductive season** during which they undertakes several brooding cycles (Fonds 1973). This species usually spawns at sea, although spawning in lagoons and estuaries has been reported as well (Elliott & Hemmingway 2002). The male builds a nest underneath an empty bivalve shell or a flat stone and covers it with sand or mud. He then tries to attract several females sequentially into the nest where they attach the pear-shaped eggs (2000-4000 eggs per batch)

in a single layer to the ceiling of the shell (reviewed in Magnhagen 1999). The male attracts females by courtship display including erect fins, tail beats and swimming rapidly in short bouts close to the female (Forsgren 1997). Sand goby females are known to use multiple cues in their choice of a mate (Lindström & Lugli 2000; Pampoulie *et al.* 2004b; Svensson 2004), including the breeding colouration (Forsgren 1992). Sneak mating by small males with rudimentary breeding coloration ('sneakers') is a common phenomenon in sand gobies; half of all nests contained eggs fertilized by another male than the nest-holding male (Jones *et al.* 2001). After spawning, the females leave and the male provides all parental care for the eggs until hatching. Larvae are about 3 mm at hatching and they are pelagic for at least one month before they become demersal after metamorphosis at about 12 mm (Webb 1980).

Sand gobies have an important **ecological significance** in the marine food web and ecosystem because of their high density (Pasquaud *et al.* 2004; Ehrenberg *et al.* 2005). Due to the abundance of food in estuarine and coastal areas it was concluded that they have relatively little impact on the prey community and that goby populations are probably top-down regulated (Doornbos & Twisk 1987; Jaquet & Raffaelli 1989; Ehrenberg *et al.* 2005). The sand goby is a generalist and opportunistic consumer feeding on endo- and epibenthic prey, yet zooplankton and hyperbenthic prey items also occur in its diet. As they grow the diet become more diverse and consists of amphipods, mysids, polychaetes, shrimps, copepods and siphons of bivalves, but during their (post)larval stage they almost exclusively depend on copepods (Fonds 1973; Doornbos & Twisk 1987; Jaquet & Raffaelli 1989; Hamerlynck & Cattrijsse 1994; Beyst *et al.* 1999). Sand gobies are visual feeders (Healey 1971; Aarnio & Bonsdorff 1993) and nocturnal (Gibson & Hesthagen 1981). Nocturnal foraging seems advantageous in approaching prey and especially in avoiding predators (Thetmeyer 1997). The sand goby is quite important as prey for piscivorous fish species (e.g. *Gadus morhua*, *Trisopterus luscus*, *Merlangius merlangus*), crabs, piscivorous birds (e.g. *Phalacrocorax carbo*, *Podiceps cristatus*, *Mergus merganser*) and young harbour seals (Hamerlynck & Hostens 1993; Hamerlynck & Cattrijsse 1994; Beyst *et al.* 1999). Moreover, *P. minutus* is also an intermediate host to important parasites of (commercial) fish species (Huyse 2002; Zander 2005). Therefore they are considered to be an important link between benthic invertebrates and higher trophic levels. The sand goby does not have any (direct) commercial value but is a common by-catch species of the beam trawl fishery (Cabral *et al.* 2002).

Considerable **within-species morphological, ecological and behavioural differences** have been recorded for geographically distinct sand gobies. Miller (1986) considered a division between sand gobies from the Atlantic Ocean, *P. m. minutus*, and from the Mediterranean and Black Sea, *P. m. elongatus* (Canestrini 1861) based on morphological and ecological differences. In contrast to the Atlantic sand gobies, Mediterranean *P. minutus* carry a dark chin spot in the females and breast pigmentation in both sexes. The life cycle of the sand goby in the Mediterranean region is rather 'contracted' with fast growth and rapid maturity, increased spawning effort and shorter life spans (8-14 months). In contrast, north of the English Channel they have a more 'protracted' life cycle with shorter spawning seasons but longer life spans (20-24 months) (Fonds 1973; Bouchereau & Guelorget 1998; Pampoulie *et al.* 1999; Ehrenberg *et al.* 2005). The Portuguese populations showed life history characteristics situated in between those of the Mediterranean and the North Atlantic (Arruda *et al.* 1993; Leitão *et al.* 2006; Dolbeth *et al.* 2007). Furthermore, as temperature is important for spawning, the breeding period changes gradually with latitude, between April to August in the Baltic Sea (Parmanne & Lindström 2003; Waligóra-Borek & Sapota 2005); between February to June in the North Sea (Fonds 1973), and between December to April in southern Europe (Mediterranean, Iberian Peninsula) (Pampoulie *et al.* 1999). Finally, sand gobies in northern Europe are considered to make thermal migrations from estuaries and shallow waters into deeper warmer waters during winter and subsequent spawning migrations to shallow coastal waters after winter (Healey 1971; Fonds 1973; Maes *et al.* 1998). Emigration from estuaries and saline lagoons as observed in southern Europe is most likely only for the purpose of reproduction (Pampoulie *et al.* 1999; Dolbeth *et al.* 2007). These differences illustrate the possibility of intraspecific genetic differentiation in the sand goby.

Phylogeography and population genetics of the sand goby

Mitogenic phylogeographic patterns of the sand goby were studied throughout its full range by means of a 283 bp fragment of the mitochondrial cytochrome *b* (cyt *b*) locus (Gysels *et al.* 2004b) and an 840 bp fragment of the D-loop locus (Stefanni & Thorley 2003). Limited genetic differentiation with a weak pattern of isolation-by-distance was recorded throughout the distributional range of *P. minutus*. The highest degree of divergence was found between Atlantic and Mediterranean samples. However, the two studies showed

conflicting results for the present relationship between Mediterranean and Atlantic populations. A high number of endemic cyt *b* haplotypes as well as the most common Atlantic haplotype were recorded in appreciable frequencies for the sand gobies of the western Mediterranean Sea, suggesting secondary contacts between the two groups (Gysels *et al.* 2004b). However, Stefanni & Thorley (2003) found only endemic D-loop haplotypes in the western Mediterranean Sea, suggesting no subsequent gene flow between the two basins after the divergence of the two phylogeographical groups. In the Atlantic basin, both loci pointed to a range expansion into northern areas with a loss of variation at higher latitudes.

Sand gobies from nine localities throughout its total distribution have also been investigated with starch (SGE) and cellulose acetate (CAGE) gel electrophoresis at 13 enzyme systems. A low degree of **nuclear genetic differentiation** was observed between the different sample locations, which was interpreted as a result of a high level of gene flow within and between the different marine systems (Gysels 2003; Stefanni *et al.* 2003). Significant differentiation was noticed between Atlantic and western Mediterranean samples, suggesting a barrier for dispersal across the Strait of Gibraltar and/or an effect of isolation-by-distance. Within the Atlantic basin, Gysels (2003) recorded indications for a group structure with a slight (although not significant) effect of isolation-by-distance and with a clustering of the southern North Sea samples in comparison with samples from the Atlantic coast and the northern North Sea.

Small-scale genetic patterns in *P. minutus* were only studied along the Belgian coast using allozyme and microsatellite markers (Gysels *et al.* 2004c; Pampoulie *et al.* 2004a). Allozyme markers did not show a significant population structuring along the Belgian coast for the sand goby (Gysels *et al.* 2004c). On the other hand, microsatellite analysis yielded some evidence for small-scale population structuring with two spatially separated breeding units in the southern bight of the North Sea, namely one in the Oosterschelde and another one in the Belgian coastal area (Pampoulie *et al.* 2004a). Although the microsatellites suffered from problems that make them not fully reliable for studying subtle population structures (Larmuseau 2005), catchment data of the sand goby supported the presence of two different breeding units in the Southern Bight of the North Sea. One unit is breeding along the

western Belgian coast between Ostend and De Panne; the other breeding unit was found in the Voordelta between Renesse and Domburg (Vanden Eede 2006; Guelinckx 2008).

5. RESEARCH QUESTION, STRATEGY AND OUTLINE OF THE THESIS

The aim of the thesis was to contribute to the general understanding of the role of natural selection in the evolution of marine organisms. To study adaptive within-species change in a gene with direct phenotypic penetration, the rhodopsin gene, which determines the spectral sensitivity of dim-light vision, was selected. Because of indications for intraspecific polymorphism at the rhodopsin gene (Jokela *et al.* 2003) and its model status, the sand goby *Pomatoschistus minutus*, a common marine demersal fish, was chosen as study organism. Therefore the concrete research question of the thesis is: **'Is local adaptation detectable at the rhodopsin gene in *P. minutus*?'** We provide an answer through the following 'top-down' approach (starting with phenotypic observations) with five defined tasks:

TASK 1. Establishing a high quality neutral marker-derived phylogeographic and population genetic analysis for the sand goby.

Background:

The phylogeographic pattern and 'neutral' genetic structure of a particular species provide information of how neutral processes such as gene flow and genetic drift structured the intraspecific neutral genetic variation. Estimating the magnitude of contemporary gene flow and effective population size using neutral molecular markers is an indirect way to assess the potential of local adaptation (Adkinson 1995; Hansen *et al.* 2002). Moreover, neutral markers are influenced by genetic drift and gene flow in contrast to adaptive markers which are influenced by genetic drift, gene flow and natural selection. By comparing the genetic pattern of presumably neutral markers with the pattern on potentially adaptive markers, the signature of selection can be identified (Storz 2005). Additionally, there is also potential for investigating the origin and spread of adaptive traits based on the historical and present population structure (Beebee & Rowe 2004). To realize this first task the population genetic

study must be of high quality because the genetic signal from population differentiation is weak in marine species (Ward *et al.* 1994). Consequently, various errors associated with estimating population genetic parameters that might normally be safely ignored assume a relatively greater importance (Waples 1998).

Strategy:

First, two technical improvements were realized to assure partially a high quality for the phylogeographic and population genetic structure (*chapter 2*). The first part of *chapter 2* (*subchapter 2a*) describes a cost effective and fast molecular tool to determine four cryptic *Pomatoschistus* species unambiguously. In the second part of *chapter 2* (*subchapter 2b*), we described the isolation of nine new high quality microsatellites for *P. minutus*, which may be useful as genetic markers for the analysis of the population structure.

Second, a large-scale phylogeographic and population genetic survey covering the entire European range of the sand goby was conducted in *chapter 3* using both mitochondrial DNA (*subchapter 3a*) and microsatellite markers (*subchapter 3b*). We asked how recent climatic history and contemporary factors shaped the distribution of *P. minutus*. We also asked to what extent historical imprints of refuges, recolonization patterns and demographic expansion were detectable. Based on this information, the neutral genetic structure of the sand goby was reconstructed.

TASK 2. Demonstrating that the sand goby populations differ in functional variation at the rhodopsin gene (*RH1*).

Background:

Natural selection cannot operate unless there are genetically-based phenotypic differences between individuals. As long as heritability for a trait cannot be firmly established in the field, it will remain difficult to prove anything about local adaptation in natural populations (Endler 1986). In other words, populations have to differ for a heritable trait and the possibility of plasticity has to be ruled out. Nevertheless, although physiological and biochemical studies can be used to test the functional relationships between specific genes and phenotypes, the tests do not indicate as such whether these genes have any role in explaining adaptive genetic variation in nature (Howe & Brunner 2005). The task is

grounded on step 1 to demonstrate local adaptation in the wild (see part 2 of the introduction).

Strategy:

The genetic basis for the population differentiation in the λ_{max} of the rod opsin in the sand goby had to be confirmed, as suggested in Jokela *et al.* (2003). In *chapter 4*, the differentiation based on the functional variation of the *RH1* gene is described for sand goby populations living in various marine systems (the macro-scale analysis of *subchapter 4a*) and for populations within the marine systems of the Baltic-North Sea region (micro-scale analysis of *subchapter 4b*).

TASK 3. Demonstrating that the differentiation between sand goby populations for functional variation on *RH1* is due to natural selection as opposed to neutral processes.

Background:

One difficult step to demonstrate local adaptation is to differentiate unquestionable the signature of selection from neutral processes as genetic drift (Guinand *et al.* 2004). An observed (functional) differentiation between populations might be the result of either random processes or natural selection, two evolutionary forces that differentiate populations. Therefore, the possibility of population differentiation by chance had to be ruled out for the candidate gene variation (Endler 1986). To complete this task, it is important to combine different methods that will increase the evidence that the population differences on functional variation are due to selection (Vasemägi & Primmer 2005). This task is grounded in step 2 to demonstrate local adaptation in the wild (see part 2 of the introduction).

Strategy:

As described in *chapter 4* (*subchapter 4a*), two approaches were used: (1) by neutrality tests based on comparisons between the different classes of mutations within the rhodopsin gene sequences; and (2) comparisons of the geographical distributions of the rhodopsin variation with the neutral marker-derived phylogeographic and population genetic structure of sand gobies.

TASK 4. Establishing a link between the functional variation at the *RH1* gene and selection regimes that the sand goby populations experience.

Background:

The functional differentiation between populations must lead to fitness improvements for the locally adapted individuals. The possibility of linkage disequilibrium between the particular candidate gene and the loci or genomic regions directly under selection (genetic 'hitch-hiking') had to be ruled out (Carvalho 1993). This task is grounded in step 3 to demonstrate local adaptation in the wild (see part 2 of the introduction).

Strategy:

This task is realized in *chapter 4* by comparing the differentiation on the rhodopsin variation between sand goby samples with the local photic environment, measured by satellite remote sensing analysis.

TASK 5. Demonstrating directional selection in a phylogenetic framework to determine whether selection has played a significant role in the evolution of the *RH1* gene within the 'sand goby' group.

Background:

Through the study of the evolution of the candidate gene in related species, the arguments that selection on the gene influences contemporary population structure should be enforced (Endler 1986). Moreover, further insights in the selective character of the candidate gene can be gained to understand also better intraspecific distribution of adaptive variation.

Strategy:

In *chapter 5* the phylogenetic framework was used to determine whether there is evidence for differential adaptive molecular evolution in the rhodopsin gene among related 'sand goby' species inhabiting different environments in terms of salinity, depth, turbidity, substrate, and so on. By studying the evolution of the gene in the 'sand goby' group, the adaptive character

and the geographical distribution of the rhodopsin variation in *P. minutus* is better understood.

Finally, *chapter 6* summarizes the results for the five different tasks to find evidence for visual local adaptation on the rhodopsin gene in the sand goby. The implications of these results for our understanding of (visual) local adaptation in marine organisms are addressed. To conclude, perspectives for future studies are presented.

Chapter 2

Keeping an eye on the quality of genetic studies in sand goby

'Science is a way of trying not to fool yourself. The first principle is that you must not fool yourself, and you are the easiest person to fool.'

Richard Feynman (American physicist, 1918-1988)

SUBCHAPTER 2A

FAST PCR-RFLP METHOD FACILITATES IDENTIFICATION OF *POMATOSCHISTUS* SPECIES FROM THE NORTH ATLANTIC

Maarten H. D. Larmuseau[1], Jef Guelinckx[2], Bart Hellemans[1], Jeroen K. J. Van Houdt[1,3] and Filip A. M. Volckaert[1]

[1]Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Leuven, Belgium

[2]Laboratory of Aquatic Ecology and Evolutionary Biology, Katholieke Universiteit Leuven, Leuven, Belgium

[3]Royal Museum for Central Africa, Tervuren, Belgium

Published manuscript: *Journal of Applied Ichthyology* (2008), **24**, 342-344

Introduction

One of the main challenges in ichthyology is the identification of individuals to species level (Burton 1996). Many research ideas in ecology cannot be realised because identification of taxa poses a difficult problem. Moreover, drawing the wrong conclusions due to misinformation in basic identification of species could have serious consequences. Fernandez *et al.* (2006) emphasize the need to make molecular characterization of species a standard part of ecological analyses of population and communities. Molecular tools allow for the unambiguous identification of individuals (Lucentini *et al.* 2007) and create new opportunities in a broad spectrum of ecological studies (Bickford *et al.* 2007). A wide range of species-identification problems, from cryptic life stages (Chow *et al.* 2006) to prey species in gut contents (Zaidi *et al.* 1999), can be eliminated by molecular markers.

The *Pomatoschistus* genus (Perciformes, Gobiidae) includes some of the most abundant demersal fishes along the European coasts and estuaries (Ehrenberg *et al.* 2005; Leitão *et al.* 2006). Four *Pomatoschistus* species are difficult to distinguish due to similarities in

morphological appearances and the use of similar habitats in the North Atlantic (Hamerlynck 1990). These include the sand goby *Pomatoschistus minutus* (Pallas, 1770), Lozano's goby *Pomatoschistus lozanoi* (deBuen, 1923), common goby *Pomatoschistus microps* (Krøyer, 1838) and painted goby *Pomatoschistus pictus* (Malm, 1965). The taxonomic status of species belonging to the 'sand goby' group is well resolved (Miller 1986; Huyse *et al.* 2004). Currently, identification keys are based on different morphological characteristics such as pigmentation patterns (Hamerlynck 1990) and the pattern of the sensory papillae (Miller 1986). Once familiar with these morphological features the keys can be applied to identify (sub)adult specimens. However, they cannot be used to identify (post)larvae, damaged individuals, individuals found in gut contents and biopsies. Thus it was impossible to discriminate among the *Pomatoschistus* species in previous studies on hyperbenthos (Beyst *et al.* 1999; Beyst *et al.* 2001). Because of their abundance and ecological role in coastal waters of the North Atlantic, *Pomatoschistus* spp. are often used as model species in behavioural (Jones *et al.* 2001; Lindström *et al.* 2006), ecological (Leitão *et al.* 2006; Rodrigues *et al.* 2006) and population genetic studies (Gysels *et al.* 2004a; Gysels *et al.* 2004b; Pampoulie *et al.* 2004a). Recently *P. minutus* is increasingly being investigated as a model for estuarine migration dynamics and evolutionary adaptation (Guelinckx *et al.* 2007; *subchapter 2b*). Therefore it is useful to create a simple, cost effective test to distinguish these four *Pomatoschistus* species unambiguously.

Allozyme studies have shown that some diagnostic loci are useful for the identification of *Pomatoschistus* species (Wallis & Beardmore 1984). However, allozyme techniques require deep-frozen samples (-80°C) collected from freshly killed specimens, to preserve the maintenance of enzyme activities until analysis. Therefore, tests based on the PCR-RFLP (Restriction Fragment Length Polymorphism) method have recently been developed for species identification in many taxa, including fish (Céspedes *et al.* 2000; Cocolin *et al.* 2000; Aranishi 2005). As an alternative to laborious and costly DNA sequencing, RFLP is a highly reliable and easy method for identifying polymorphism in DNA sequences using restriction enzymes and gel electrophoresis.

Material & Methods

To identify potential diagnostic restriction sites, a limited amount of sequence data is available for *Pomatoschistus* species, except for those loci (cytochrome *b*, ITS, 16S, 12S and the mitochondrial control region) used in recent phylogeographical and phylogenetic studies (Stefanni & Thorley 2003; Gysels *et al.* 2004a; Gysels *et al.* 2004b; Huyse *et al.* 2004). However, no species-specific enzyme restriction sites were found in these loci, except in the mitochondrial control region. The high mutation rate in this locus makes a quick accumulatation of point mutations possible, allowing the discrimination of closely related species with PCR-RFLP.

Sequences of the control region are available on GenBank for *P. minutus* and *P. lozanoi* (accession numbers *P. minutus* AY033004–AY033036 and AY033043–AY033048, *P. lozanoi* AY033049–AY033053). For *P. pictus* and *P. microps*, genomic DNA was extracted from a fin-clip using the NucleoSpin Extraction kit (Machery-Nagel GmbH). The D-loop fragment of 840 bp was amplified using the HN20 (Bernatchez & Danzmann 1993) and the Pro19-primers (Bernatchez *et al.* 1995), located in the phenylalanine tRNA-gene and the proline tRNA-gene, respectively. The PCR reactions were carried out on a GeneAmp PCR System 2700 thermocycler (Applied Biosystems) in a total volume of 25 µl, containing 1 µl of genomic DNA, 1x PCR buffer, 0.2 mM dNTPs, 0.8 µM of each primer, 2.0 mM $MgCl_2$, 0.5 U of Taq DNA polymerase (Silverstar, Eurogentec) and mQ-H_2O. The PCR cycle was as follows: 94°C for 4 min followed by 35 cycles of 30 s at 96°C, 30 s at 52°C and 1 min at 72°C; with a final 10 min extension period at 72°C. After purification with the 'GFX PCR DNA and Gel Band Purification kit' (Amersham Biosciences), seven PCR products from *P. pictus* and nine from *P. microps* were sequenced in both directions using the BigDye Terminator v. 3.1 Cycle Sequencing Kit (Applied Biosystems). Four haplotypes were detected (GenBank accession numbers: EF122407–EF122410). Three restriction enzymes were selected to identify the species (Table 2.1). Because one haplotype of *P. minutus* was not restricted by *Eco*1471, an additional restriction with *Mph*1103I was necessary to ensure the correct identification of *P. minutus*.

For the identification of field samples, the PCR reactions were set to the same conditions as above and were subsequently followed by digestion with the restriction enzymes. An individual reaction contained 0.3 µl restriction enzyme, 7 µl PCR-product, 1 µl Y+/Tango™ buffer, Buffer B+ or Buffer R+ (Fermentas) for *Eam*1104I, *Eco*147I and *Mph*1103I respectively. DNA was digested at 37°C for 4 h. Following digestion, products were separated by electrophoresis on a 2% agarose gel for 20 min at 200 V. Fragment sizes were compared to a 100 bp ladder (Fermentas). Essential individual runs with known controls of the species were made to avoid erroneous identification owing to inhibition of the enzyme or incomplete digestion.

Table 2.1 Fragment sizes (bp) in the PCR reaction analysis (control) and the PCR-RFLP analysis using restriction enzymes: *Eam*1104 I, *Eco*147 I and *Mph*1103 I

	Pomatoschistus minutus	*Pomatoschistus lozanoi*	*Pomatoschistus microps*	*Pomatoschistus pictus*
No restriction enzyme	840	840	840	840
Eam 1104I	840	770 + 70	840	840
Eco 147I	210 + 630 or 840	840	840	740 + 100
Mph 1103I	600 + 240	840	840	840

Results & Discussion

The RFLP-PCR identification tool was blind-tested using 30 samples each of adult *P. minutus*, *P. lozanoi*, *P. microps* and *P. pictus* collected in the North Sea (Ostend, Belgium) and 20 samples each of adult *P. microps* and *P. minutus* from the Mediterranean Sea (Vaccarès lagoon, France). These 160 specimens were all unambiguously identified based on morphological characteristics following Miller (1986) and Hamerlynck (1990). The results of the three tests fully corroborated the morphological identification, demonstrating that the developed identification tool for the four *Pomatoschistus* species is correct and reproducible. Our results did not show intraspecific polymorphism for the three restriction enzymes. Except for the Adriatic sand goby, which is considered as a distinct species of the *Pomatoschistus* genus (Huyse *et al.* 2004), no population structure was revealed for *P. minutus* using a direct sequence analysis on the mitochondrial control region and on cytochrome *b* (Stefanni & Thorley 2003; Gysels *et al.* 2004b). An analysis of cytochrome *b* and allozymes also did not reveal a relevant population structure for *P. microps* (Gysels *et al.* 2004a). Comparison

between the intra- and interspecific variation of the control region sequences thus supports the assumption that the high mutation rate within this locus resulted in different RFLP-patterns among the four *Pomatoschistus* species and no substantial differences within species.

Application of the PCR-RFLP method in an ongoing population dynamic study for *P. minutus* in the Scheldt estuary already yielded new insights in the population structure of this species. Preliminary results based on 200 individuals of unknown identity revealed that *P. minutus* postlarvae are present in the estuary during autumn. This implies that sand goby spawns during summer in the southern North Sea, which is outside the assumed spawning season for this species (March–June) (Fonds 1973). Consequently, there is probably no temporal segregation in the spawning season with *P. lozanoi* (May–August) (Fonds 1973). Hence, previous assignments of *Pomatoschistus* postlarvae to species level based on the presumed difference in spawning time might be erroneous (Beyst *et al.* 1999).

The applicability of the PCR-RFLP method was not tested on digested or fixed material. However, we believe that the tool is promising for these types of samples as mtDNA is present at high copy numbers in cells and is more likely to preserve for a longer period than nuclear DNA. On the other hand, the length of the PCR-fragment might be too large to amplify when the DNA is highly degraded. It is then necessary to develop additional internal primers to amplify shorter fragments that include the species-specific restriction sites (Jehaes *et al.* 2001; Deagle *et al.* 2006).

The restriction enzyme-based identification tool described here will be of great use to those working on ecology and population genetics of the *Pomatoschistus* species because laborious direct sequencing procedures can now be avoided. This method can further be developed for other problematic taxa, both within and without the Gobiidae.

Acknowledgements

The first two authors acknowledge a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). We thank Alain Crivelli for providing samples and Gregory Maes and Annelies Ghesquiere for helpful

discussions. This study was supported by BELSPO no. SD/BN/01A (WESTBANKS). This study is part of the WESTBANKS project financed by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (contract number SD/BN/01A).

CHAPTER 2B

DEVELOPMENT AND CHARACTERIZATION OF NINE POLYMORPHIC MICROSATELLITE MARKERS IN THE SAND GOBY *POMATOSCHISTUS MINUTUS* (GOBIIDAE)

Maarten H. D. Larmuseau[1], Bart Hellemans[1], Jeroen K. J. Van Houdt[1,2] and Filip A. M. Volckaert[1]

[1]Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Leuven, Belgium

[2]Royal Museum for Central Africa, Tervuren, Belgium

Published manuscript: *Molecular Ecology Notes* (2007), **7**, 147-149

Abstract

A microsatellite-enriched genomic library was constructed for the sand goby, *Pomatoschistus minutus* (Pallas 1770), and nine polymorphic DNA microsatellite markers of high quality were successfully optimized. Characterization of 96 individuals from the Vaccarès lagoon (France) showed moderate to high levels of polymorphism (two to 54 alleles). All the markers conformed to Hardy–Weinberg equilibrium and showed no evidence of null alleles, large allele dropout, stuttering and linkage disequilibrium between pairs of loci. These markers successfully amplify in three closely related species and can be employed to investigate population genetic structure and to clarify paternity in *Pomatoschistus* species.

Keywords: enriched library, FIASCO, Gobiids, microsatellites, parentage, population structure

The sand goby, *Pomatoschistus minutus* (Pallas 1770; Gobiidae, Teleostei), is a small marine demersal fish, common in shallow waters along European coasts. Sand gobies are known to play an important role in the marine ecosystem and are a model for reproductive tactics

(Gysels *et al.* 2004b). Recently, *P. minutus* is increasingly being investigated as a model for evolutionary adaptation. This requires knowledge on neutral evolution, and hence an expansion of reliable, analysable polymorphic markers.

Eleven hypervariable microsatellites of sand goby have been developed previously (Pampoulie *et al.* 2004a; Berrebi *et al.* 2006). However, all these markers but one, to some extent suffer from problems that make them unreliable for studying the subtle population structure in sand gobies. The main problems are the presence of null alleles, difficulties with fragment calling and large allele dropout. Here, we describe the isolation of nine new high quality microsatellites for *P. minutus*, which may be useful as genetic markers for the analysis of the population structure and behaviour studies in sand gobies.

Genomic DNA was extracted from a fin-clip of *P. minutus* captured in Ostend, Belgium (51°17'N, 02°51'E) using the NucleoSpin Extraction kit (Machery-Nagel GmBH). A microsatellite-enriched library was constructed using the fast isolation by AFLP of sequences containing repeats (FIASCO) protocol (Zane *et al.* 2002) with some modifications. Briefly, we performed a selective hybridization with di- [$(CA)_{15}$], tetra- [$(CTTT)_5$; $(ACGC)_5$; $(GTCT)_5$] and penta-[$(TAACC)_4$] repeat probes. Fifty picomoles of each 5' biotinylated oligonucleotide probe was hybridized with 1.0 μg of adapter-ligated DNA. Hybridized probe–DNA was captured using magnetic beads (Streptavidin MagneSpher Paramagnetic Particles, Promega). Nonspecific DNA was removed by three nonstringency washes in 400 μL TEN 1000 (10 mM Tris-HCL, 1 mM EDTA, 1M NaCl) and three stringency washes in 400 μL 0.2×SSC and 0.1% SDS. Eluted DNA was thereafter amplified through polymerase chain reaction (PCR), and cloned into bacterial vectors using the TOPO-TA cloning kit (Invitrogen). Recombinant plasmids were screened on ampicillin-supplemented agar plates with X-gal and IPTG for blue-white selection.

To extract plasmid DNA, each bacterial recombinant clone was put into a 96-well plate in 100 μL of water and incubated at 96°C for 5 min. Ten microlitres of the plasmid extract was added to a 50-μL reaction volume containing 20 mM Tris-HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 U of *Taq* DNA polymerase (Invitrogen Life Technologies) and 20 pmol each of standard M13 forward and reverse primers. The PCR cycling was carried out

on a GenAmp PCR system 2700 thermocycler (Applied Biosystems) with 30 cycles of 30 s at 96°C, 30 s at 60°C and 60 s at 72°C. A total of 64 positive colonies were scanned for microsatellite repeats.

The PCR products were sequenced using the SequiTherm Excel II Long-Read DNA Sequencing Kit (Epicentre Technologies) with fluorescently labelled standard T7 promotor and M13 reverse primers. The samples were loaded on an automated DNA sequencer model 4200 (LI-COR) and the sequences from both strands were edited in the E-SEQ v. 2.0 software. Of the 64 positive colonies, 31 contained repetitive elements. For 24 clones, PCR primers were designed using the PRIMER 3 program (Rozen & Skaletsky 1998).

Table 2.2 Genetic characterization of nine polymorphic microsatellite loci in *Pomatoschistus minutus*, tested on 96 individuals from the Vaccarès lagoon

Locus	Primer sequence (5'-3')	Core motif	Size-range (bp)	Ta (°C)	nA	H_O	H_E	F_{IS}	P
Pmin03	F: GAGCTCATCAGCCAGGAGAG R: GCAGAGCTGTGAGCCATCTT	$(ac)_7atgctc(ac)_{23}$	154-261	57	39	0.943	0.950	0.008	0.267
Pmin04	F: TCTAACAGGCAGCTGAACGA R: GGCTCAAACACACACGAAAA	$(gt)_7aca(tg)_8$	195-243	57	15	867	0.819	-0.059	0.139
Pmin09	F: GCCGTGGATGCATTATCAGT R: GGGATGTGTGAGTGTGCAAG	$(tg)_2cg(tg)_{20}$	196-264	56	26	0.945	0.942	-0.003	0.405
Pmin16-2	F: CTTTTAGAGTGAGCAGAAAAGAGTG R: TGAGACATGAGAGGGGGAAG	$(at)_3gtat(gt)_3(ga)_4(gt)_8(ct)_7(gt)_8$	169-206	56	22	0.902	0.912	0.026	0.143
Pmin20	F: CAGATCTGTGGAAATCCAACC R: GGCCACAGATACGACCTAGC	$(gt)_3gc(gt)_4gc(gt)_8$	216-382	56	54	0.933	0.970	0.040	0.055
Pmin29	F: GGGCTCCACTTTGTTAGCAG R: CGTGGGAATTCCTTGATTGT	$(ca)_8$	210-219	54	9	0.738	0.792	0.069	0.129
Pmin31	F: GCTCTGTTGGCTCTGAATGA R: CTGGTTCTGCAGGAAAGTCC	$(gt)_3gcgtgg(gt)_{10}tagcagtccat(gt)_3gc(gt)_4$	155-213	54	18	0.843	0.907	0.071	0.030
Pmin35	F: GTGACTGGGAGCGTTTGAGT R: GCCCTATCTGCCTGACAAAG	$(ca)_3ta(ca)_{20}$	158-213	54	20	0.837	0.867	0.033	0.220
Pmin38	F: TGAATCCGAAGCCTGGTAAC R: TCCCTTCTGCTTCCTTTTGA	$(tg)_4ta(tg)_5$	180-182	54	2	0.032	0.031	-0.011	0.980

T_a, annealing temperature; N_a, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, fixation index; P, probability (based on 1000 permutations) of observing a more extreme F_{IS} value under the assumption of Hardy–Weinberg equilibrium (Bonferroni corrected 5% significance level = 0.00833)

PCRs were carried out in a total volume of 10 µL, containing 100 ng of *P. minutus* genomic DNA, 2 pmol of each locus-specific primer, 2.0 mm $MgCl_2$, 0.5 U of *Taq* DNA polymerase (Invitrogen Life Technologies) and other reagents as mentioned above. Amplifications were performed on a GeneAmp PCR System 2700 thermocycler (Applied Biosystems) and were carried out under the following conditions: 3 min at 95°C; 30 cycles of 30 s at 95°C, 30 s at

the annealing temperature (Table 2.2), 45 s at 72°C; and 5 min at 72°C. Individuals were genotyped on an ABI 3130 Genetic Analyser (Applied Biosystems). Allele sizes were determined by means of an internal GENESCAN 500-LIZ size standard and genotypes were obtained using GENEMAPPER v. 3.7 (Applied Biosystems).

The developed microsatellites were validated on 96 mature individuals from the Vaccarès lagoon, France (43°32'N, 04°35'E). All loci were amplified according to the optimized conditions (Table 2.2) and generated high quality products. The diversity measures and deviations from Hardy–Weinberg expectations were computed using the software package GENETIX v. 4.05.2 (Belkhir *et al.* 2004), and tests for linkage disequilibrium were conducted using GENEPOP v. 3.3 (Raymond & Rousset 1995). Table 2.2 summarizes the features of the nine polymorphic loci. None of the loci showed deviations from HWE after Bonferroni correction, or linkage between loci. The genotype data were checked for scoring errors with MICRO-CHECKER v. 2.2.3 (Van Oosterhout *et al.* 2004). There were no indications for genotyping errors due to the presences of null alleles, large allele dropout or stuttering at all nine loci. Thus, the optimized microsatellites appear to be useful and reliable for the further investigation of the population structure and genetic variability of *P. minutus* as well as for parentage and kinship analyses.

Table 2.3 Cross-species amplification using nine microsatellite primer pairs designed for *Pomatoschistus minutus*. The polymorphic loci are shown with the allele size. Monomorphic loci are listed in bold.

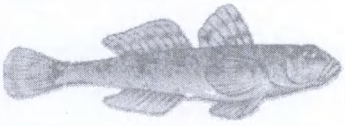
Species	N	Pmin03	Pmin04	Pmin09	Pmin16-2	Pmin20	Pmin29	Pmin31	Pmin35	Pmin38
Common goby *(Pomatoschistus microps)*	4	179-192	206-224	205-239	173-190	226-290	211-215	150-215	162-192	**176-176**
Painted goby *(Pomatoschistus pictus)*	4	177-189	208-224	203-276	173-185	219-303	210-215	155-176	181-198	180-186
Lozano's goby *(Pomatoschistus lozanoi)*	4	175-189	213-215	231-237	148-156	292-382	210-215	155-168	152-182	**182-182**

In addition, cross-species amplifications were tested on three related *Pomatoschistus* species. Four individuals of the Lozano's goby *P. lozanoi*, the common goby *P. microps* and the painted goby *P. pictus* were screened at nine loci without any modification of the PCR conditions as described above. The individuals were caught in the Southern Bight of the North Sea, Belgium. All loci could be successfully amplified in all species (Table 2.3) but more

pronounced stuttering was observed for the loci Pmin09, Pmin16-2 and Pmin20 in *P. microps* and *P. pictus*. This proves that the markers are suitable for other *Pomatoschistus* species.

Acknowledgements

This research was funded by a PhD grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). We thank Alain Crivelli and Jef Guelinckx for providing the samples and Gregory Maes for helpful discussions.



Chapter 3

Phylogeography and population genetics of the sand goby

'But where shall wisdom be found?
And where is the place of understanding?
The deep says, 'It is not in me'.
And the sea says, 'It is not with me'.

Job, 28: 12-14

SUBCHAPTER 3A

DISTRIBUTIONAL AND DEMOGRAPHIC CONSEQUENCES OF PLEISTOCENE CLIMATE FLUCTUATIONS FOR A MARINE DEMERSAL FISH IN THE NORTH-EASTERN ATLANTIC

Maarten H. D. Larmuseau[1], Jeroen K. J. Van Houdt[1,2], Jef Guelinckx[3], Bart Hellemans[1] and Filip A. M. Volckaert[1]

[1]Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Leuven, Belgium

[2]Royal Museum of Central Africa, Tervuren, Belgium

[3]Laboratory of Aquatic Ecology and Evolutionary Biology, Katholieke Universiteit Leuven, Leuven, Belgium

Accepted manuscript: *Journal of Biogeography* (2009), **36**, 1138-1151.

Abstract

Aim The Pleistocene glaciations were the most significant historical event during the evolutionary life span of most extant species. However, little is known about the consequences of these climate changes for the distribution and demography of marine animals of the north-eastern Atlantic. The present study focuses on the phylogeographic and demographic patterns of the sand goby, *Pomatoschistus minutus* (Teleostei: Gobiidae), a small marine demersal fish.

Location North-eastern Atlantic, Mediterranean, Irish, North and Baltic seas.

Methods Analysis was carried out by sequencing the mtDNA cytochrome *b* gene of sand gobies from 12 localities throughout the species' range, and using this information in combination with published data of allozyme markers and mtDNA control region sequences. Several phylogenetic methods and a network analysis were used to explore the

phylogeographic pattern. The historical demography of *P. minutus* was studied through a mismatch analysis and a Bayesian skyline plot.

Results Reciprocal monophyly was found between a Mediterranean Sea (MS) clade and an Atlantic Ocean (AO) clade, both with a Middle Pleistocene origin. The AO Clade contains two evolutionary significant units (ESUs): the Iberian Peninsula (IB) Group and the North Atlantic (NA) Group. These two groups diverged during Middle Pleistocene glacial cycles. For the NA Group there is evidence for geographic sorting of the ancestral haplotypes with recent radiations in the Baltic Sea, Irish Sea, North Sea and Bay of Biscay. The demographic histories of the Mediterranean Clade and the two Atlantic ESUs were influenced mainly by expansions dated as occurring during the Middle Pleistocene glaciations and post-Eem, respectively.

Main conclusions The pre-LGM (Last Glacial Maximum) subdivision signals were not erased for *P. minutus* during the LGM. Middle Pleistocene glaciations yielded isolated and differently evolving sets of populations. In contrast to the case for most other taxa, only the northern Atlantic group contributed to the post-glacial recolonization. The historical demography of Mediterranean sand gobies was influenced mainly by Middle Pleistocene glaciations, in contrast to that of the Atlantic populations, which was shaped by Late Pleistocene expansions.

Keywords: Bayesian skyline plot, glaciations, Gobiidae, mismatch analysis, mtDNA, north-eastern Atlantic, phylogeography, *Pomatoschistus minutus*, radiation, sand goby

Introduction

Organisms live in a constantly changing environment, and the genetic patterns we observe today are the result of both contemporary and historical factors (Avise 2000). The glaciation cycles during the Pleistocene (1800–11.5 ka) were arguably the most important climatological events during the evolutionary life span of most extant species (Hewitt 2000). Although the impact of the Pleistocene climate fluctuations on the distribution and demography of continental species is firmly established in the literature (Bernatchez & Wilson 1998; Taberlet

et al. 1998; Hewitt 2004), little is known about the impact of the glaciation cycles on the distribution and demography of north-eastern Atlantic marine species (Wilson 2006).

During glaciation events, ice-covered areas to the north were unsuitable habitats for near-shore marine species, forcing their extinction or a southward migration into one or more refugial areas. Although it is difficult to identify potential refugia from genetic data, several southern refugia have been suggested for north-eastern Atlantic marine species: around the Iberian Peninsula (Gysels *et al.* 2004a; Hoarau *et al.* 2007), Macaronesia (Domingues *et al.* 2005; Domingues *et al.* 2007) and the Mediterranean Sea (Olsen *et al.* 2004; Sa-Pinto *et al.* 2005). In addition to the large southern glacial refugia, small northern peri-glacial refugia have been proposed, such as the Bay of Biscay (Nesbø *et al.* 2000), the south-western coast of Ireland (Jolly *et al.* 2006; Hoarau *et al.* 2007) and Hurd Deep in the English Channel, a 150-km-long depression about 100 m deeper than the adjoining sea floor (Coyer *et al.* 2003; Provan *et al.* 2005b). When temperatures increased during the interglacials, populations from these refugia recolonised previously ice-covered areas in the north. These recolonization patterns remain poorly understood and are likely to be species-specific (Chevolot *et al.* 2006). Moreover, the role of the Mediterranean Sea for Atlantic species during the Pleistocene glaciation cycles is still controversial. During the glaciations, the decrease in sea level isolated the Mediterranean and Atlantic basins. Within the Mediterranean basin, extinction and recolonization played a major role during the glacial cycles, as organisms were isolated and could not follow favourable isothermals. No clear relationship is apparent between biological traits and observed patterns of partial or complete genetic isolation of the Atlantic and Mediterranean populations (Patarnello *et al.* 2007).

Owing to the open character of the sea, it is difficult to study the historical distribution and demography of species. The absence of obvious barriers to gene flow in the marine realm seems to facilitate extensive gene flow among marine populations, dimming the influence of historical events (Palumbi 1994). Most studies on marine species of the north-eastern Atlantic have focused on the genetic patterns of fishes and invertebrates with high dispersal potential. It is reasonable to assume that species showing reduced levels of contemporary gene flow and a fast evolutionary rate are better suited for elucidating phylogeographic patterns. Therefore, a marine demersal fish, the sand goby, *Pomatoschistus minutus* (Pallas,

1770; Gobiidae, Teleostei), was selected for the present study on the influence of the Pleistocene climate changes on the distribution and demography of marine demersal fishes. Sand gobies reach high densities along their geographical range: the Atlantic coasts from Norway to Spain, the North Sea, the Baltic Sea and the Irish Sea. The distribution pattern is more fragmented in the Mediterranean Sea and on the west coast of the Black Sea (Miller 1986) (Fig.3.1). In connection with its ecological significance (Pasquaud *et al.* 2004; Ehrenberg *et al.* 2005) and use as a model organism in ecology (Jones *et al.* 2001; Singer *et al.* 2006; Guelinckx *et al.* 2008), various phylogeographic and population genetic studies have been carried out on this species (Stefanni *et al.* 2003; Stefanni & Thorley 2003; Gysels *et al.* 2004b).

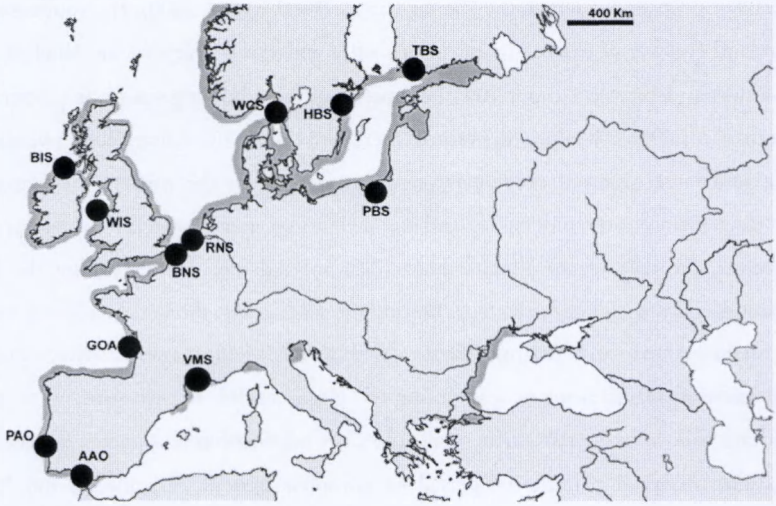


Figure 3.1 Geographical distribution of the 12 sampling locations in eight European marine systems for the sand goby, *Pomatoschistus minutus*. The shaded area represents the distribution range of *P. minutus* conforming to Miller (1986). See Table 3.1 for definitions of the sampling locations.

The main objective of the present study was to analyse how the Pleistocene climatic fluctuations shaped the regional distribution and demography of *P. minutus* along north-eastern Atlantic shores and in the Mediterranean basin. Three hypotheses based on former studies on the sand goby (Stefanni & Thorley 2003; Gysels *et al.* 2004b) were critically evaluated: (1) the Mediterranean Sea was recolonised by Atlantic sand gobies after the Last

Glacial Maximum (LGM; 20 ka) – earlier studies show conflicting results for the present relationship between Mediterranean and Atlantic populations; (2) there is no population differentiation within the Atlantic sand gobies – earlier studies revealed no significant genetic differentiation between sand goby populations in the Atlantic basin, although the species is expected to show reduced levels of contemporary gene flow owing to its life style; and (3) a population expansion of *P. minutus* in the North Atlantic occurred after the LGM – this was hypothesized despite a lack of demographic analyses for the sand goby.

Compared with a previous study on the phylogeography of *P. minutus* by Gysels *et al.* (2004b), we sequenced a much larger fragment of the cytochrome *b* (cyt *b*) gene from a higher number of sand gobies and collected at more sampling sites. These new data were pooled with D-loop (Stefanni & Thorley 2003) and allozyme (Stefanni *et al.* 2003; Gysels *et al.* 2004c) data for statistical analysis.

Materials & Methods

Sampling and mtDNA sequencing

A total of 263 *Pomatoschistus minutus* individuals were caught at 12 locations along European coasts between October 2005 and February 2007 (Table 3.1, Fig. 3.1). Samples were taken by fyke, hand net or beam trawling. Three individuals of *Pomatoschistus lozanoi* were caught as an outlier group for the phylogenetic analyses, one in the Gironde estuary, France (45°36'N, 01°01'W) (LO1), and two in Ostend, Belgium (51°17'N, 02°51'E) (LO2 and LO3). *Pomatoschistus* species were identified morphologically, based on the dermal head papillae (Miller 1986) and on pigmentation patterns (Hamerlynck 1990), and genetically according to the molecular tool described in *subchapter 2a*. For each sample, an 850 bp fragment of the mitochondrial DNA (mtDNA) cyt *b* was sequenced as detailed in Appendix S3.1. All sequences were deposited in the GenBank database (EU736948–EU737080). To verify the advantage of the longer sequenced gene fragment, the 850 bp sequences of the cyt *b* gene were compared with the 283 bp sequences of the same gene (AJ555096–AJ555123) determined by Gysels *et al.* (2004b). A sliding-window analysis of 100 bp and 10 bp steps was performed to illustrate nucleotide diversity variation along the whole cyt *b* gene with DnaSP

v. 4.10.9 (Rozas *et al.* 2003). For the phylogenetic and phylogeographical analyses only the long fragments of cyt *b* were used. The D-loop sequences of Stefanni & Thorley (2003) were recovered from GenBank (AY033004–AY033053).

Table 3.1 Overview of the *Pomatoschistus minutus* samples collected from 12 sampling sites in the north-eastern Atlantic.

Code	Area	Country	Location	Date	Longitude	Latitude	N	R
TBS	Northern Baltic Sea	Finland	Tvärminne	jul/06	59°50'N	23°12'E	29	11
HBS	Northern Baltic Sea	Sweden	Värtan	nov/05	58°59'N	17°27'E	24	12
PBS	Southern Baltic Sea	Poland	Sopot, Gulf of Gdańsk	feb/07	54°27'N	18°35'E	12	9
WCS	Skagerrak	Sweden	Bökevik Bay, Skaftö island	jun/06	58°14'N	11°26'E	17	12
RNS	North Sea	Netherlands	Renesse	nov/05	51°44'N	03°47'E	22	14
BNS	North Sea	Belgium	Oostduinkerke	nov/06	51°08'N	02°40'E	12	5
BIS	Irish Sea	North Ireland (UK)	River Bann estuary	sep/06	55°09'N	06°45'W	9	5
WIS	Irish Sea	Wales (UK)	Llanfairfechan	nov/06	53°59'N	03°59'W	30	13
GOA	Bay of Biscay	France	Gironde estuary	aug/06	45°36'N	01°01'W	30	18
PAO	Iberian Peninsula	Portugal	Alcochete, Tagus estuary	okt/05	38°45'N	08°57'W	10	5
AAO	Iberian Peninsula	Spain	Guadalquivir river estuary	nov/06	36°58'N	06°10'W	35	23
VMS1	Mediterranean Sea	France	Vaccarès lagoon	jan/06	43°32'N	04°35'E	25	24
VMS2	Mediterranean Sea	France	Vaccarès lagoon	jan/07	43°32'N	04°35'E	8	8

N, sample size; R, number of unique cyt *b* haplotypes

Phylogenetic and network analyses

Phylogenetic reconstruction was conducted using neighbour-joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) analyses. The trees were rooted with the sister species *P. lozanoi* (Huyse *et al.* 2004). NJ trees were constructed in MEGA v. 4.0 (Tamura *et al.* 2007). Bootstrap analyses were performed with 10,000 replicates. MP was performed using PAUP* v. 4.0b10 (Swofford 2002) with the following heuristic search settings: 10^3 random taxon addition replicates followed by tree bisection–reconnection (TBR) branch swapping and 1000 bootstrap replicates. According to the Akaike information criterion (AIC), MODELTEST v. 3.7 (Posada & Crandall 1998) selected the TrN + I + Γ

model as the most suitable model for further analysis. The ML analysis with 100 bootstrap replicates was performed using PhyML v. 2.4.4 (Guindon & Gascuel 2003), via a web server (Guindon *et al.* 2005). TCS v. 1.3 (Clement *et al.* 2000) was used to construct a statistical parsimony network, and a median-joining network analysis was performed using the software NETWORK v. 4.5.0.1 (http://www.fluxus-engineering.com).

The times to the most recent common ancestor (tMRCA) of *P. minutus* and the major sand goby mtDNA lineages were estimated using Bayesian inference as implemented by the software program BEAST v. 1.4.6 (Drummond & Rambaut 2006). Because of limitations in the BEAST program, all analyses were performed using the GTR model as best alternative for the TrN model. The rate variation among sites was modelled using a gamma distribution with 10 rate categories. The divergence times and their credibility intervals were estimated under the coalescent model with constant population size and with expansion growth, using the strict molecular clock. The posterior distributions of the dates being estimated were approximated by sampling parameter values at every 500th generation over 10^7 Markov chain Monte Carlo (MCMC) steps, after discarding 10^6 burn-in steps. Convergence of the sampled parameters was verified using the program TRACER v. 1.4 (Rambaut & Drummond 2007). The effective sample size for each parameter was found to exceed 100 individuals, which is the minimum recommended effective sample size (Weinstock *et al.* 2005). Dates of divergences between the clades were calculated using a conventional clock for the mitochondrial cyt *b* gene in bony fishes, namely 2% sequence divergence per million years (Bowen *et al.* 2001; Domingues *et al.* 2005). As the mtDNA of *Pomatoschistus* gobies has a much higher rate of evolution than that of other teleosts (Gysels *et al.* 2004a; Huyse *et al.* 2004), divergence times were also estimated using higher rates of 4% and 6% for the cyt *b* locus. Then, the specific molecular clock for the cyt *b* locus of *Pomatoschistus* gobies was calibrated based on the divergence of the Indo-Pacific and the Atlantic–Mediterranean gobies through the closure of the Atlantic–Mediterranean part of the Tethys Sea *c.* 15–14 Ma (McKay & Miller 1997). Therefore, the tMRCA was estimated with the various selected clocks between the two *Pomatoschistus* species and two Indo-Pacific gobies, *Tridentiger bifasciatus* (Acc. nr. AB021254) and *Tridentiger obscurus* (Acc. nr. AB021255).

After analyses on the cyt *b* sequences, phylogenetic and network analyses were conducted on the reconstructed data set of all available D-loop sequences and compared with the mtDNA-based results of Stefanni & Thorley (2003). The divergence times of the sand goby lineages were estimated using various rates (2%, 4%, 6% and 8% sequence divergence per million years) to calibrate the specific molecular clock for the D-loop of the *Pomatoschistus* species.

The genetic diversity and population differentiation in *P. minutus* were analysed based on the cyt *b* and D-loop data using several common indices (π, *h* and F_{ST}) and non-metric multi-dimensional scaling (NMDS) analyses (for a detailed methodology see Appendix S3.1). In addition, these statistical analyses were also performed on two original allozyme data sets of Stefanni *et al.* (2003) and Gysels *et al.* (2004c), which were obtained under the same laboratory conditions and for which high sample quality was guaranteed (Gysels 2003) (detailed in Appendix S3.1).

Demographic analyses

Demographic expansions were first investigated on the cyt *b* and the published control region sequences by means of Tajima's (1989) *D*-test and Fu's (1997) F_s-test of neutrality using ARLEQUIN v. 3.11 (Excoffier *et al.* 2005). For neutral markers, significant negative values can be expected in cases of population expansion (Tajima 1989; Fu 1997). Furthermore, the demographic history was examined using the frequency distribution of pairwise differences among sequences (Mismatch distribution) (Harpending 1994). Because the sand goby probably experienced demographic as well as range expansions and contractions during successive cycles of glaciations, both demographic and spatial mismatch analyses (Excoffier 2004; Excoffier *et al.* 2005) were conducted using ARLEQUIN. Past population demographics of *P. minutus* were also inferred for the cyt *b* and the D-loop data using the coalescent Bayesian skyline plot (BSP) model (Drummond *et al.* 2005) as implemented in BEAST and visualized in TRACER. Final analyses were run for 3×10^7 generations, sampling every 1000th generation, and a burn-in of 3×10^6 generations (after a pre-burn-in of 3×10^5 generations). The analyses were repeated using different values for the number of grouped intervals (m = 5, 10, 20) and different clock models (strict clock and

relaxed clock) with uncorrelated rates drawn from a lognormal distribution, conducting two independent MCMC runs for each parameter combination.

Results

MtDNA haplotypes

The alignment of sequences was straightforward as there were no gaps and translation into amino acids did not indicate nonsense or stop codons. The sequence characteristics matched the general properties of the *P. minutus* cyt *b* gene (Gysels *et al.* 2004b; Keith *et al.* 2005), suggesting a functional mtDNA cyt *b* gene and not a nuclear pseudogene (Zhang & Hewitt 1996). The alignment of all cyt *b* gene sequences revealed 130 unique haplotypes within the 263 sand goby individuals. A total of 130 variable sites (15.3% of the total) were detected, of which 70 positions (8.2% of the total) were parsimony-informative. Of the 130 variable sites, 18 (13.8%) were in first codon position, four (3.1%) were in second codon position, and 108 (83.1%) were in third codon position. Translation using the vertebrate mitochondrial genetic code indicated that 17 out of 282 amino acid residues were polymorphic.

The first 158 bp of our sequences match with the short fragment of cyt *b* sequenced by Gysels *et al.* (2004b) (see Table S3.1 in Appendix S3.2). Sliding-window analysis showed a low variability in this first part of the gene. Moreover, the original sequence data of Gysels *et al.* (2004b) often revealed low quality at the end of the sequence, in particular for the Mediterranean Sea samples. Because of conservative scoring, some mutations characteristic for the Mediterranean population were not detected by Gysels *et al.* (2004b). The high precision of the capillary DNA sequencer, the SeqScape software and the negative tests for detecting contamination guarantee the high quality of the sequences in the present study.

Phylogenetic and network analyses

Phylogenetic analyses of the data set, including all sequences of *P. minutus* and the outgroup taxon, revealed consistent results for the tree methods (Fig. 3.2). MP analysis generated a consensus tree of 339 steps (consistency index = 0.6012, retention index = 0.8746, rescaled

consistency index = 0.5258). A similar tree was obtained with the NJ and ML algorithms ($-\ln L = 3219.065$). In all phylogenetic trees, two main distinct mitochondrial lineages were identified within the ingroup: the Mediterranean Clade (MS Clade), with 30 haplotypes from the sampling location in the Mediterranean Sea (VMS1 and VMS2), and the Atlantic Clade (AO Clade), with 100 haplotypes from the Iberian Peninsula, the Bay of Biscay, Irish Sea, North Sea, Skagerrak and Baltic Sea (Fig. 3.2; Table S3.2 in Appendix S3.2).

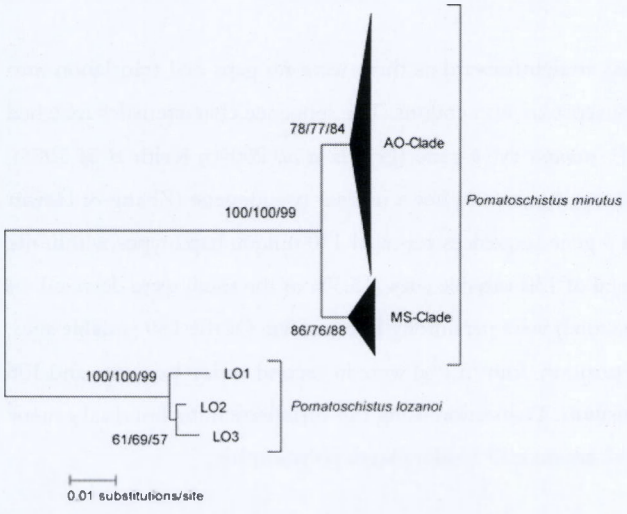


Figure 3.2 Maximum likelihood tree of all *Pomatoschistus minutus* cyt *b* haplotypes and three haplotypes of *Pomatoschistus lozanoi* as the genetic outgroup. Bootstrap values are indicated for statistically supported groupings (≥ 50%) for maximum likelihood (ML), maximum parsimony (MP) and neighbour joining (NJ) (ML/MP/NJ). Because of the high number of haplotypes, the clades for which no supported groupings were detected are compressed. MS Clade, Mediterranean Sea Clade; AO Clade, Atlantic Ocean Clade.

The network of the MS Clade based on cyt *b* displays a star-like pattern, with a high number of haplotypes compared with the number of sequenced individuals (Fig. 3.3). The network analysis of the AO Clade reveals two separate clusters of haplotypes, the Iberian Peninsular Group (IB Group), including all individuals caught in the Guadalquivir estuary (sample AAO) and in the Tagus estuary (sample PAO), and the North Atlantic Group (NA Group), comprising all sampling locations from the Bay of Biscay to the Baltic Sea (Fig. 3.4). The IB Group reveals a star-like pattern with one central haplotype (IB02), which was found with

high frequency in both sampled locations (AAO and PAO). Network analysis for all haplotypes of the NA Group displays a pattern with several highly frequent haplotypes instead of a star-like pattern with one central haplotype. The network shows one main ancestral haplotype (haplotype NA01), which occurred in each North Atlantic location, and a few highly frequent haplotypes, for example NA28 and NA09, which are very common in the northern Baltic Sea and the southern North Sea, respectively. Many haplotypes are connected with single mutation steps to those ancestral haplotypes. Such radiations were found in each marine system (Fig. 3.4). In the network of the NA Group, the clear geographic association of the haplotypes suggests limited contemporary gene flow between marine systems. The network analyses gave congruent results between the different methods. The phylogenetic and network analyses conducted on the reconstructed data set of all control region (D-loop) sequences were comparable with the initial analysis of Stefanni & Thorley (2003).

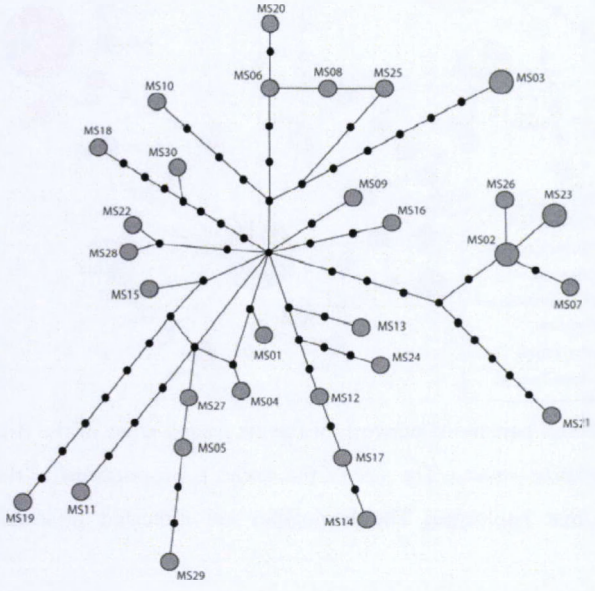


Figure 3.3 Statistical parsimony network of the cyt *b* haplotypes of the Mediterranean Sea (MS) Clade of *Pomatoschistus minutus*. The size of the circles is proportional to the number of sand gobies sharing that haplotype. The haplotypes are indicated by numbers as given in Table S3.1.

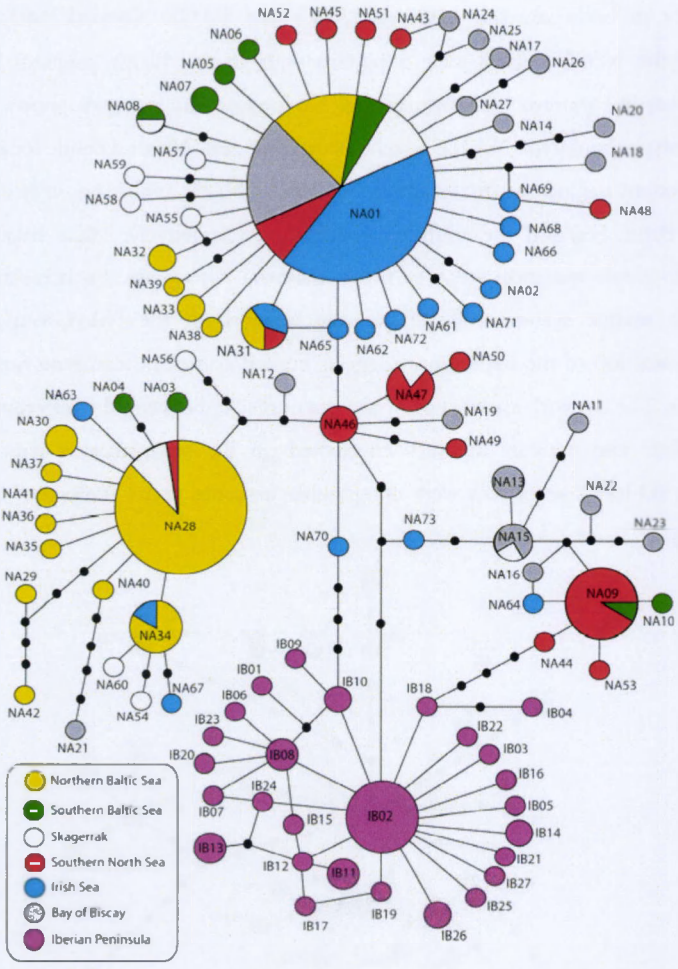


Figure 3.4 Statistical parsimony network of the cyt *b* haplotypes of the Atlantic Ocean (AO) Clade of *Pomatoschistus minutus*. The size of the circles is proportional to the number of sand gobies sharing that haplotype. The haplotypes are indicated by numbers as given in Table S3.1.

For dating the various historical events, the molecular clock was calibrated by calculating the tMRCA of the two *Pomatoschistus* species and two Indo-Pacific gobies (*T. bifasciatus* and *T. obscurus*) with the cyt *b* locus. The tMRCA was estimated at 49.93, 24.95 and 16.63 Ma with the 2%, 4% and 6% molecular clocks, respectively. The tMRCA of the data set including all

cyt *b* haplotypes of *P. minutus* and *P. lozanoi* was dated at *c.* 6.37, 3.23 and 2.14 Ma. Based on the D-loop data, the tMRCA of *P. minutus* and *P. lozanoi* was dated at *c.* 13.6, 6.8, 4.5 and 3.4 Ma with the 2%, 4%, 6% and 8% molecular clocks respectively. Because the Atlantic–Mediterranean and Indo-Pacific gobies diverged at least 14–15 Ma (McKay & Miller 1997) and *P. minutus* and *P. lozanoi* diverged between 2.5 and 1.1 Ma (Wallis & Beardmore 1984; Huyse *et al.* 2004), a realistic and conservative molecular clock might be between 4% and maximally 6% for the cyt *b* locus and between 6% and 8% for the D-loop. It might be postulated that the fast generation time (1 or 2 years) of gobies and their small body size would compensate for the slower clock generally assumed for fish (Martin & Palumbi 1993).

Table 3.2 Diversity indices and times to most recent common ancestor (tMRCAs) for the observed phylogeographic units and the various north-eastern Atlantic marine systems of *Pomatoschistus minutus* based on cyt *b* data.

	N	R	h	π	tMRCA (4%)	tMRCA (6%)
Total data set	263	130	0.947 (0.009)	0.01054 (0.00062)	0.827 (1.032-0.602)	0.551 (0.688-0.401)
MS-Clade	33	30	0.994 (0.009)	0.01069 (0.00067)	0.465 (0.609-0.306)	0.310 (0.406-0.204)
AO-Clade	230	100	0.931 (0.012)	0.00641 (0.00019)	0.398 (0.516-0.277)	0.265 (0.344-0.185)
IB-Group	45	27	0.925 (0.032)	0.00244 (0.00026)	0.144 (0.212-0.091)	0.096 (0.141-0.061)
NA-Group	185	73	0.897 (0.017)	0.00518 (0.00021)	0.345 (0.503-0.268)	0.230 (0.335-0.178)
Northern Baltic	53	16	0.800 (0.048)	0.00391 (0.00043)	-	-
Southern Baltic	12	9	0.939 (0.058)	0.00563 (0.00107)	-	-
Skagerrak	17	12	0.919 (0.057)	0.00555 (0.00080)	-	-
Southern North Sea	34	15	0.870 (0.044)	0.00504 (0.00033)	-	-
Irish Sea	39	17	0.687 (0.086)	0.00269 (0.00065)	-	-
Bay of Biscay	30	18	0.887 (0.052)	0.00587 (0.00068)	-	-

N, number of individuals per region; R, number of cyt *b* haplotypes observed; h, haplotype diversity; π, nucleotide diversity; tMRCA in Ma was estimated using molecular clocks of 4% and 6% sequence divergence per million years. The credibility intervals of tMRCA and the standard deviations of haplotype and nucleotide diversity values are given between brackets. MS-Clade, Mediterranean Sea Clade; AO-Clade, Atlantic Ocean Clade; IB-Group, Iberian Peninsula Group; NA-Group, North Atlantic Group.

The tMRCA of all *P. minutus* cyt *b* sequences was estimated as between 0.827 and 0.551 Ma, and those of each clade separately between 0.464 and 0.265 Ma (both respectively with the 4% and 6% clock). This suggests that the divergence between the MS Clade and the AO Clade occurred in the first half of the Middle Pleistocene. The tMRCA of the NA Group and IB Group was estimated at 0.345–0.230 and 0.144–0.096 Ma, respectively (Table 3.2),

also indicating a Middle Pleistocene divergence between the two lineages. Different coalescent models resulted in highly consistent estimates for the tMRCAs. The tMRCAs of the D-loop sequences of all genuine *P. minutus* samples, the MS Clade and the AO Clade were comparable with those estimated from the cyt *b* data (see Table S3.3 in Appendix S3.2).

Genetic diversity and population differentiation

For all cyt *b* haplotypes, the nucleotide diversity (π) and the haplotype diversity (h) were 0.01054 and 0.947, respectively. The genetic diversity was higher for the MS Clade (π = 0.01069 and h = 0.994) than for the AO Clade (π = 0.00641 and h = 0.931). The Irish Sea has low π- and h-values, whereas the Bay of Biscay, the southern Baltic, Skagerrak and the southern North seas have high π- and h-values. The Iberian Peninsula and the Northern Baltic have low π- but high h-values (Table 3.2). There was no significant decrease in nucleotide diversity northwards within the range.

The pairwise F_{ST}-values for the cyt *b* data are given in Table S3.4 in Appendix S3.2. The Mantel tests (Mantel 1967) revealed significant correlations for all data and those of the North Atlantic (respectively, $r^2 = 0.6617$, $p < 0.05$; $r^2 = 0.4220$, $p < 0.05$). The NMDS analysis based on the cyt *b* data (see Fig. S3.1a, b in Appendix S3.3) showed different clusters corresponding to the *a priori* groups of *P. minutus* based on its geography (Table 3.1). The samples WCS, PBS and GOA clustered centrally in the NMDS. These three sampling locations of the Bay of Biscay, Skagerrak and southern Baltic Sea contain haplotypes occupying the entire network of the North Atlantic, in contrast to the other sample locations (BNS, RNS, WIS, BIS, TBS and HBS) (Fig. 3.4).

The genetic variability and differentiation based on the reconstructed D-loop haplotypes and allozymes were highly consistent with the results of Stefanni & Thorley (2003), Stefanni *et al.* (2003) and Gysels *et al.* (2004c). Additional NMDS analyses for the allozyme data revealed a clear genetic differentiation between Mediterranean Sea and North Atlantic populations, despite the high variability within the latter group (see Fig. S3.1c in Appendix S3.3).

Demographic analyses

The mismatch distribution of all sand goby haplotype sequences was clearly bimodal (see Fig. S3.2a in Appendix S3.3): one mode corresponded with the number of differences between the MS and AO clades, and the other mode with the number of differences among individuals within both lineages. As expected from the star-like statistical parsimony network (Fig. 3.3), the mismatch distribution for the MS lineage (see Fig. S3.2b in Appendix S3.3) fitted the predicted distribution under a model of sudden expansion. Because of this, and because of the negative neutrality tests ($p < 0.05$) (Table 3.3), the sudden expansion model could not be rejected for the MS Clade. Its time of expansion was estimated at 0.283–0.187 Ma (Middle Pleistocene) (Table 3.3).

Table 3.3 Results of the neutrality tests and mismatch analysis for cyt *b* sequences.

			Demographic mismatch distribution					Spatial mismatch distribution				
	D	F_S	τ	t_e (4%)	t_e (6%)	P(rag)	P(SDD)	τ	t_e (4%)	t_e (6%)	P(rag)	P(SDD)
MS-Clade	**-1.732**	**-20.404**	9.623	0.283	0.189	0.923	0.884	9.508	0.280	0.187	0.857	0.765
AO-Clade	**-1.914**	**-24.974**	7.246	0.213	0.142	0.591	0.322	5.846	0.172	0.115	0.943	0.304
IB-Group	**-1.781**	**-26.156**	2.668	0.079	0.052	0.067	0.103	2.669	0.079	0.052	0.059	0.103
NA-Group	**-2.026**	**-25.473**	6.324	0.186	0.124	0.805	0.673	4.074	0.120	0.080	0.933	0.565

D, Tajima's D-test and F_S, Fu's F_S-test (bold: p-value < 0.05); τ, time since expansion measured in mutational time units; t_e, absolute time (Ma) since expansion estimated with molecular clocks of 4% and 6% sequence divergence per million years; P(rag), p-value for the raggedness index; P(SDD), p-value for the sum of the squared deviations under the hypothesis of sudden expansion. MS-Clade, Mediterranean Sea Clade; AO-Clade, Atlantic Ocean Clade; IB-Group, Iberian Peninsula Group; NA-Group, North Atlantic Group.

The mismatch analysis of the AO Clade showed a clear bimodal distribution (see Fig. S3.2c in Appendix S3.3), and therefore all haplotypes of the IB Group were separated from those of the NA Group. Despite the non-unimodal distribution of the mismatch analysis of the NA Group (see Fig. S3.2e in Appendix S3.3), the mismatch distributions of both groups were not significantly different from the predicted distribution under a model of sudden expansion. Moreover, the neutrality tests were significantly negative in both groups (Table 3.3). The expansions occurred 0.0785–0.0523 Ma (Late Pleistocene) and 0.186–0.0800 Ma (Middle and Late Pleistocene) for the IB Group and the NA Group, respectively. Only for the North Atlantic was a significant difference found between the demographic and

the range expansion analyses (Table 3.3). The various mismatch analyses of the D-loop showed the same global pattern, despite the low number of Mediterranean haplotypes (Table S3.3 in Appendix S3.2; Fig. S3.3 in Appendix S3.3).

The Bayesian skyline plot (BSP) analysis was used to date shifts in population size of the *P. minutus* clades. The BSP of the MS Clade with the cyt *b* data shows two periods of population growth. The oldest expansion is dated before 110 ka, and a more recent, much smaller, expansion phase occurred between 75 and 30 ka (Fig. 3.5a). The BSP of the IB Group reveals a growth of population size before 25–17 ka (Fig. 3.5b). The plot of the NA Group based on the cyt *b* data shows a sharp increase in the effective number of individuals between 70 and 30 ka and between 105 and 40 ka for the 6% and 4% clocks, respectively (Fig. 3.5c). The results were robust, as different coalescent models in the BEAST analysis resulted in similar estimates. The results of the BSP analyses of the cyt *b* data were comparable with those based on the D-loop data (see Fig. S3.4 in Appendix S3.3).

Discussion

Consequences of Pleistocene glaciations for the distribution of Pomatoschistus minutus

The phylogenetic analysis clearly shows two monophyletic clades within the species *P. minutus* (Fig. 3.2). The Mediterranean Sea Clade (MS Clade) is represented only by individuals from the Vaccarès lagoon (Gulf of Lion). The Atlantic Ocean Clade (AO Clade) comprises all populations spanning from the Iberian Peninsula to the Baltic Sea. Miller (1986) previously suggested this division between sand gobies from the Atlantic Ocean, *P. m. minutus*, and those from the Mediterranean and Black seas, *P. m. elongatus* (Canestrini, 1861), based on morphological and ecological differences. Earlier genetic studies based on mtDNA sequences and allozymes failed to prove a strong genetic divergence between the West Mediterranean and the Atlantic sand gobies owing to a limited resolution of the applied genetic markers (Stefanni *et al.* 2003; Stefanni & Thorley 2003; Gysels *et al.* 2004b; Huyse *et al.* 2004).

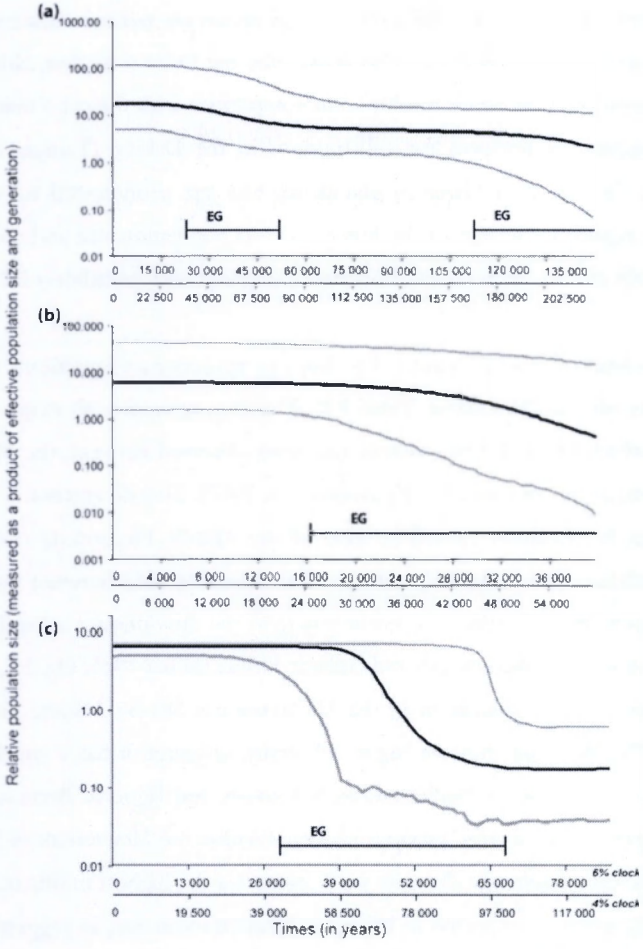


Figure 3.5 Bayesian skyline plots (BSPs) showing changes in population size through time: (a) Mediterranean Sea (MS) Clade; (b) Iberian Peninsula (IB) Group; (c) North Atlantic (NA) Group. These historical demographic trends of the cyt *b* lineages in *Pomatoschistus minutus* are represented by the 6% and 4% molecular clocks and by indication of the period of expansion growth (EG); the relative population size is measured as a product of effective population size and generation time. x-axis: time (in years); the upper axis is the time measured with the 6% molecular clock and the lower axis shows the time with the 4% molecular clock; y-axis: relative population size (measured as a product of effective population size and generation). The black line represents the median estimate; upper and lower limits (95% HPD) are drawn in grey.

The reanalyses of the D-loop and the allozyme data herein are not in conflict with the new cyt *b* data. Reciprocal monophyly was also shown for the D-loop marker, although with a much lower resolution. This lower resolution in comparison with the cyt *b* marker could be caused by homoplasy at hotspots for substitutions in the D-loop (Tamura & Nei 1993; Wakeley 1993). The signal of historical divergence was less pronounced for the allozyme data, which we suggest is because of the higher nuclear population size and the consequent lower genetic drift effects among nuclear markers compared with matrilinear DNA.

The divergence time of the two sand goby clades or evolutionary significant units (ESUs) was dated to the Middle Pleistocene (Table 3.2). This is consistent with many other marine organisms for which reciprocal monophyly has been observed between the Mediterranean and Atlantic populations (reviewed in Patarnello *et al.* 2007). This divergence is probably the result of geographic isolation caused by one of the Middle Pleistocene sea-level drops. Following this differentiation, subsequent gene flow was disrupted between the two clades. Contemporary gene flow is unlikely to occur owing to the discontinuous distribution of the sand goby between the Mediterranean and Atlantic basins (Miller 1986; Fig. 3.1). Hypothesis 1, which concerns the recolonization of the Mediterranean Sea by Atlantic sand gobies, is thus rejected. The MS Clade has the highest diversity, although it has a small distribution range along the north-western Mediterranean Sea coasts and lagoons. Because the highest species diversity of the 'sand goby' group is observed within the Mediterranean Sea (Huyse *et al.* 2004) and the cyt *b* haplotype diversity of *P. minutus* is the highest in this marine system, the origin of this species is expected to be in the Mediterranean Sea, as suggested by Gysels *et al.* (2004b). Similar to other estuarine systems, the Mediterranean lagoons play an important role in the life cycle of gobies (Bouchereau & Guelorget 1998). These lagoons, however, suffer from increasing human pressure through overexploitation (gobies are a common by-catch) and reduced water quality (Bouchereau & Guelorget 1998; Bouchereau 2001). Moreover, Mediterranean sand gobies are probably also directly threatened by the contemporary climate change (Philippart *et al.* 2007, and references therein), as the water temperature approaches critically high values for sand gobies in the Mediterranean system (Pampoulie *et al.* 1999). A further increase in water temperature could be pernicious as northward migration is excluded.

Two groups were identified within AO Clade: the Iberian (IB) Group and the North Atlantic (NA) Group. The IB Group contained the individuals collected along the coasts of the Iberian Peninsula. The NA Group comprised all populations spanning from the Bay of Biscay to the Baltic Sea (Fig. 3.4). This divergence is probably the result of population restriction within different refugia during glaciations in the Middle Pleistocene (Table 3.2). Hypothesis 2, which concerns the absence of population differentiation within the North Atlantic sand gobies, is therefore rejected. All haplotypes of the IB Group are unique for the Iberian Peninsula coast, suggesting a glacial refugium in this region for the sand goby. A growing body of studies on north-eastern Atlantic marine taxa [e.g. seagrasses (Coyer *et al.* 2003; Olsen *et al.* 2004), invertebrates (Zane *et al.* 2000), rays (Chevolot *et al.* 2006) and teleosts (Consuegra *et al.* 2002; Cortey & Garcia-Marin 2002; Gysels *et al.* 2004a)] suggests an Iberian refugium during the Pleistocene. In contrast to the case for most of the organisms with a presumed Iberian refugium, including the related common goby (*P. microps*) (Gysels *et al.* 2004a), the Iberian sand gobies appear not to have contributed to the post-glacial distribution expansion in northern Europe.

The northward spatial expansion to the North Sea, Irish Sea and Baltic Sea appears to have involved only the NA Group of the sand goby. The cyt *b* haplotype network of this NA Group showed no regional clustering, but there were population-specific haplotypes scattered throughout the whole network and radiations in the Baltic Sea, Irish Sea, North Sea and Bay of Biscay around a variety of common haplotypes (Fig. 3.4). This complex pattern suggests different genetic bottlenecks and subsequent recolonization phases, driven by the periodic climate fluctuations during the Middle and Late Pleistocene. It is very difficult and rather speculative to reconstruct these various historical patterns by locating the northern peri-glacial refugia and recolonization routes, based on present-day genetic information alone. However, based on the high present-day genetic diversity and the presence of an assortment of unique haplotypes (Table 3.2; Fig. 3.4), the Bay of Biscay and the Hurd Deep in the English Channel appear more likely to have been glacial refugia than south-western Ireland (Table 3.2, Fig. 3.4). There is, however, no evidence that the sand goby was able to withstand the much lower temperatures that have been reconstructed for this region of the English Channel for the glacial maxima of the Late Pleistocene. Modelling reconstructions suggest that sea temperatures near the shelf in this region of the English Channel were 1–

2°C in winter, rising to 5–6°C in summer during the LGM (Sarnthein 2001). The difference in the haplotype network between southern and northern Baltic Sea samples (Fig. 3.4) suggests that the Baltic Sea has been colonized in two phases over a period of 8000 years, with a stronger founder effect in the north. The presence of only two common haplotypes in the northern samples HBS and TBS (haplotypes NA01 and NA28) with unique derived haplotypes suggests that only a few individuals that were able to adapt to the severe abiotic conditions of the northern Baltic (Johannesson & André 2006) founded the population. A better sampling strategy with a higher spatial resolution in the North Atlantic region and the use of a genetic marker with an adequate variability (e.g. microsatellites) to detect genetic differences are needed to locate the refugial areas more accurately and to understand the complex recolonization routes in the north-eastern Atlantic.

Consequences of the Pleistocene glaciations for the demography of Pomatoschistus minutus

The various demographic analyses show that the intra-assemblage genetic structure of *P. minutus* contains signatures of demographic expansion events (Fig. 3.5; Fig. S3.2 in Appendix S3.3; Table 3.3). For the MS Clade, the main demographic expansion period was consistently dated as having occurred during the Middle Pleistocene using the mismatch analysis. This strongly corroborates the scarce data on the demographic expansion events of other Mediterranean Sea organisms: for the sea urchin *Paracentrotus lividus* (Calderón *et al.* 2008) and the tuna *Thunnus thynnus thynnus* (Carlsson *et al.* 2004) the demographic expansion was estimated as occurring between 0.135 and 0.360 Ma. However, the BSP of the cyt *b* sequences also suggests an additional smaller expansion in the Mediterranean region in the past 50,000 years (Fig. 3.5a).

It is generally assumed that the current demography of Atlantic populations results mainly from the LGM, because this period probably blurred the signals of earlier demographic events. Until now, the demographic expansions of marine species along the north-eastern Atlantic coast have always been dated as pre-LGM events, between 1.7 and 0.11 Ma; for example algae (Provan *et al.* 2005a; Hoarau *et al.* 2007; Calderón *et al.* 2008), polychaetes (Jolly *et al.* 2006), bivalves (Luttikhuizen *et al.* 2003), urchins (Calderón *et al.* 2008), crustaceans (Stamatis *et al.* 2004), rays (Chevolot *et al.* 2006) and teleosts (Gysels *et al.* 2004a;

Aboim *et al.* 2005; Bremer *et al.* 2005; Charrier *et al.* 2006). For the Iberian sand gobies (IB Group), an expansion period was dated herein as occurring in the Eem interglacial (128–67 ka) or later (Table 3.3, Fig. 3.5b), depending on the molecular clock used. For the NA Group, the bimodal mismatch distribution of all North Atlantic haplotypes corresponds to the network analysis, which showed star-like patterns around several ancestral haplotypes (Fig. 3.4). This suggests lineage sorting followed by a recent expansion period. In the case of a non-unimodal pattern of the mismatch distribution, the BSP is a better method with which to date this expansion period than the mismatch analysis. The BSP ignores the pattern of the mismatch distribution and is based on coalescent events (Drummond *et al.* 2005). Using BSP, the expansion event for the NA Group was estimated to have occurred during the Eem interglacial (Fig. 3.5c). The Eem interglacial followed the Warthe/Saale glaciation (180–128 ka), which is thought to have had more widespread effects on the distribution and demography of organisms than the LGM (Kellaway *et al.* 1975). This indicates that hypothesis 3 of the present study is not confirmed, and that the LGM did not substantially blur the signals of earlier demographic events. A similar conclusion was also reached in analyses of the demography of marine taxa in the north-eastern Atlantic (Hoarau *et al.* 2007).

Acknowledgements

The authors wish to thank the following persons who kindly provided the samples used in this study: Pilar Drake (CSIC, Spain), Maria Järvi-Laturi (University of Helsinki, Finland), Kai Lindström (University of Helsinki, Finland), Katarzyna Waligóra-Borek (University of Gdańsk, Poland), Lotta Kvarnemo (Göteborgs Universitet, Sweden), David Rodgers (Environment & Heritage Service, UK), Ian McCarthy (University of Wales, Bangor, UK), Mario Lepage (Ifremer, France), Martha Freitas (University of Algarve, Portugal) and Alain Crivelli (Station Biologique de la Tour du Valat, France). We are also grateful to Tine Huyse and three anonymous referees, whose suggestions greatly improved this manuscript. This work was supported by a PhD fellowship from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This study is part of the WESTBANKS project financed by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (contract number SD/BN/01A).

Appendix of Subchapter 3a

Appendix S3.1 Additional information on the mtDNA amplification and sequencing; and on the analysis of the genetic diversity and population differentiation

MtDNA amplification and sequencing

Genomic DNA was extracted from fin clips stored in 100% ethanol using the NucleoSplin Extraction Kit (Machery-Nagel GmBH, Düren, Germany). A fragment of 1188 bp of the mtDNA cytochrome *b* (cyt *b*) gene was polymerase chain reaction (PCR) amplified with primers GobycytbF (Gysels *et al.* 2004b) and H16526 (Briolay *et al.* 1998). The PCR was carried out on a GeneAmp PCR System 2700 thermocycler (Applied Biosystems, Foster City, CA, USA) in a total volume of 25 µl, containing 1 µl of genomic DNA, 1 X PCR buffer, 0.2 mM dNTPs, 0.8 µM of each primer, 2.0 mM $MgCl_2$, 0.5 U of Taq DNA polymerase (Silverstar, Eurogentec, Seraing, Belgium) and mQ-H_2O. The PCR cycle was: 4 min at 94°C followed by 35 cycles of 30 s at 96°C, 30 s at 54°C and 1 min at 72°C; with a final 10 min extension period at 72°C. To avoid contamination, different pipettes, aerosol barrier tips and different sections of the laboratory were used for pre- and post-PCR work. Every 15th individual (corresponding with one every two rows of a PCR-plate) a negative control was used to detect contamination. No contamination occurred during the screening procedures. All PCR products were visualized on agarose gels with ethidium bromide. After purification with the 'GFX PCR DNA and Gel Band Purification kit' (GE Healthcare, Piscataway, NJ, USA), the PCR products were sequenced using the BigDye Terminator v. 3.1 Cycle Sequencing Kit according to the manufacturer's protocol on an ABI 3130 automated capillary DNA sequencer (Applied Biosystems). Sequences of 850 bp were checked and aligned to each other with SeqScape v. 2.1 (Applied Biosystems). For the samples wherefore an ambiguity occurred, the complementary strand was sequenced.

Genetic diversity and population differentiation

Using the program Arlequin v. 3.11 (Excoffier *et al.* 2005), cyt *b* polymorphism was estimated as nucleotide (π) (Nei 1987) and haplotype diversity (h) (Nei & Tajima 1981) for

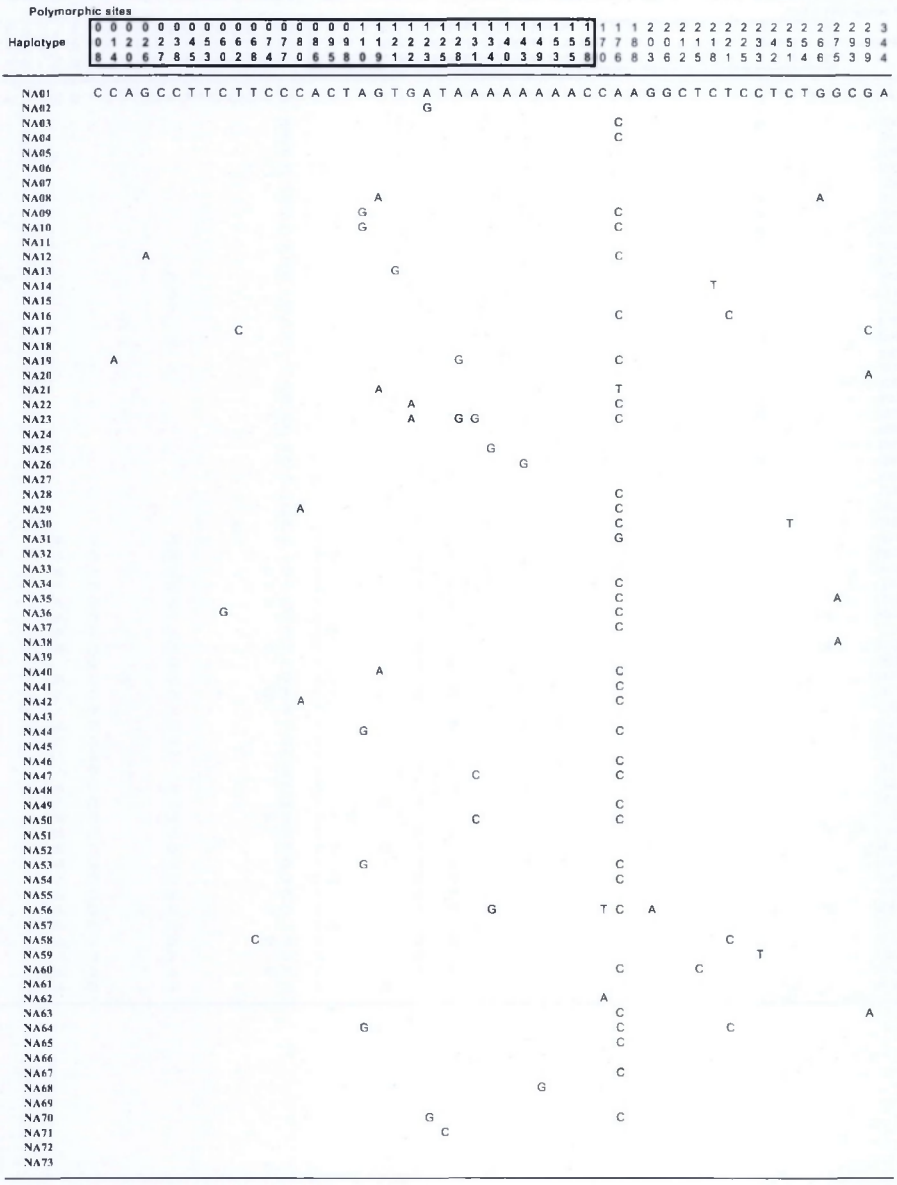
the total dataset, the different marine systems and each location separately. These indices were tested for correlation with the mean longitude of sampling locations in the North Atlantic. Genetic differentiation between populations was evaluated by estimating pairwise values of F_{ST} (Weir & Cockerham 1984) and the pairwise genetic distances according to Tamura & Nei (1993) with Arlequin. The statistical significance of the estimates was assessed by 10^4 permutations. A sequential Bonferroni test was applied to correct significance levels for multiple testing (Rice 1989). To analyse the effect of geographical distance on genetic distance (Tamura & Nei 1993), the Mantel test in GENETIX v. 4.05 (Belkhir *et al.* 2004) was used, which computes the correlation between distance matrices by means of a permutation procedure (Mantel 1967; Smouse *et al.* 1986). Geographical distances were obtained as the shortest coastal distances between sites using the electronic atlas ENCARTA (Microsoft, Redmond, WA, USA). The pairwise genetic distances were used in a non-metric multidimensional scaling analysis (NMDS) in order to reveal a group structure in STATISTICA v. 6.0 (StatSoft, Inc., Tulsa, OK, USA). Ordination plots with a stress value below 0.20 provide interpretable information concerning intersite relationships (Clarke 1993).

Population genetic analyses were also conducted on the original allozyme data of a selected group of populations for which high quality for a high number of samples was guaranteed (Gysels 2003). The following locations of Stefanni *et al.* (2003) were selected: Bergen (Ber: Northern North Sea), Oban (Oba: Irish Sea), Texel (Tex: North Sea) and Pérols (Per: Mediterranean Sea). Additionally, three populations of Gysels *et al.* (2004c) from the Southern North Sea were selected: Stroombank (Str), Kwintebank (Kwi) and Raan (Raa). In both studies nine enzyme systems, corresponding to 15 loci screened by means of Cellulose Acetate Gel Electrophoresis (CAGE), were used as a basis for genetic comparisons. Allele frequencies and the observed and unbiased expected heterozygosity were calculated in GENETIX. Pairwise F_{ST}-values and genetic distances (Slatkin 1995) were calculated and significance was assessed with permutation tests (1000 replicates) in GENETIX. A NMDS was applied on a matrix of genetic distances in STATISTICA.

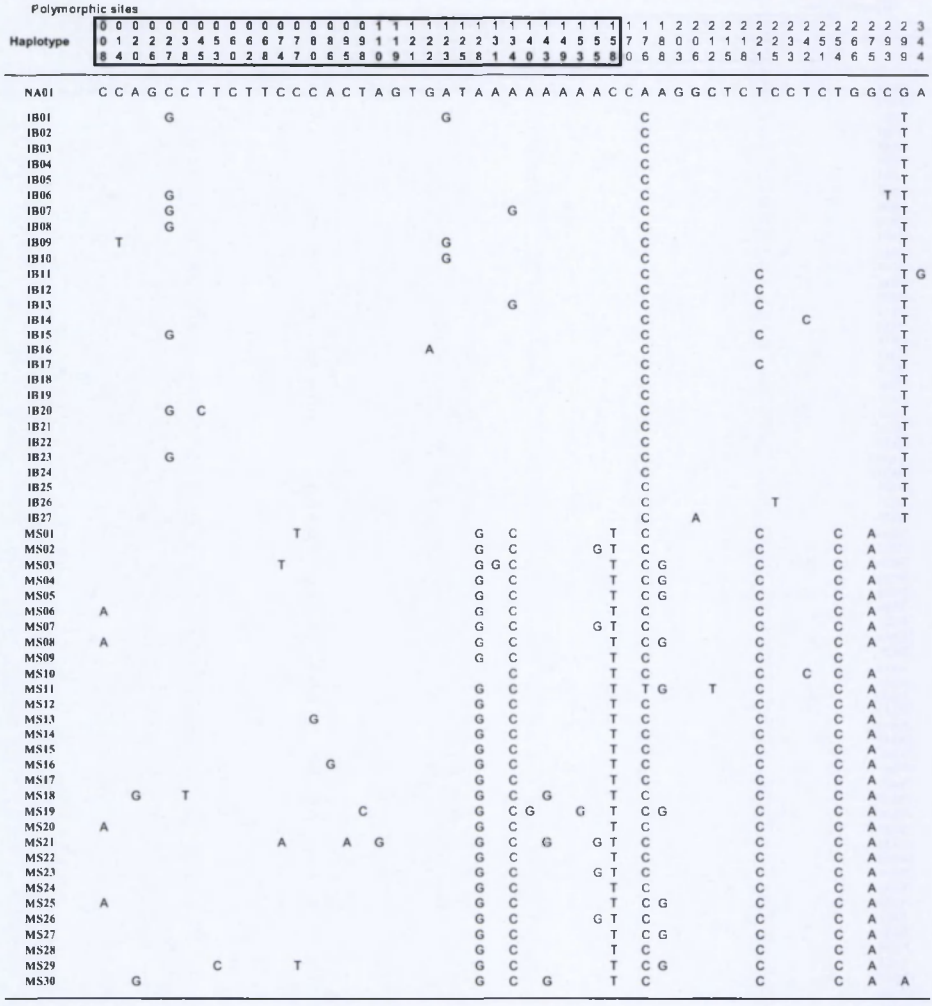
Appendix S3.2 Supplementary tables

Table S3.1 Variable nucleotide positions of the cyt *b* haplotypes observed in *Pomatoschistus minutus* with indication of the GenBank accession numbers of the haplotypes. The polymorphic sites in the overlap region with the fragment studied by Gysels *et al.* (2004b) are given in bold and in a square.

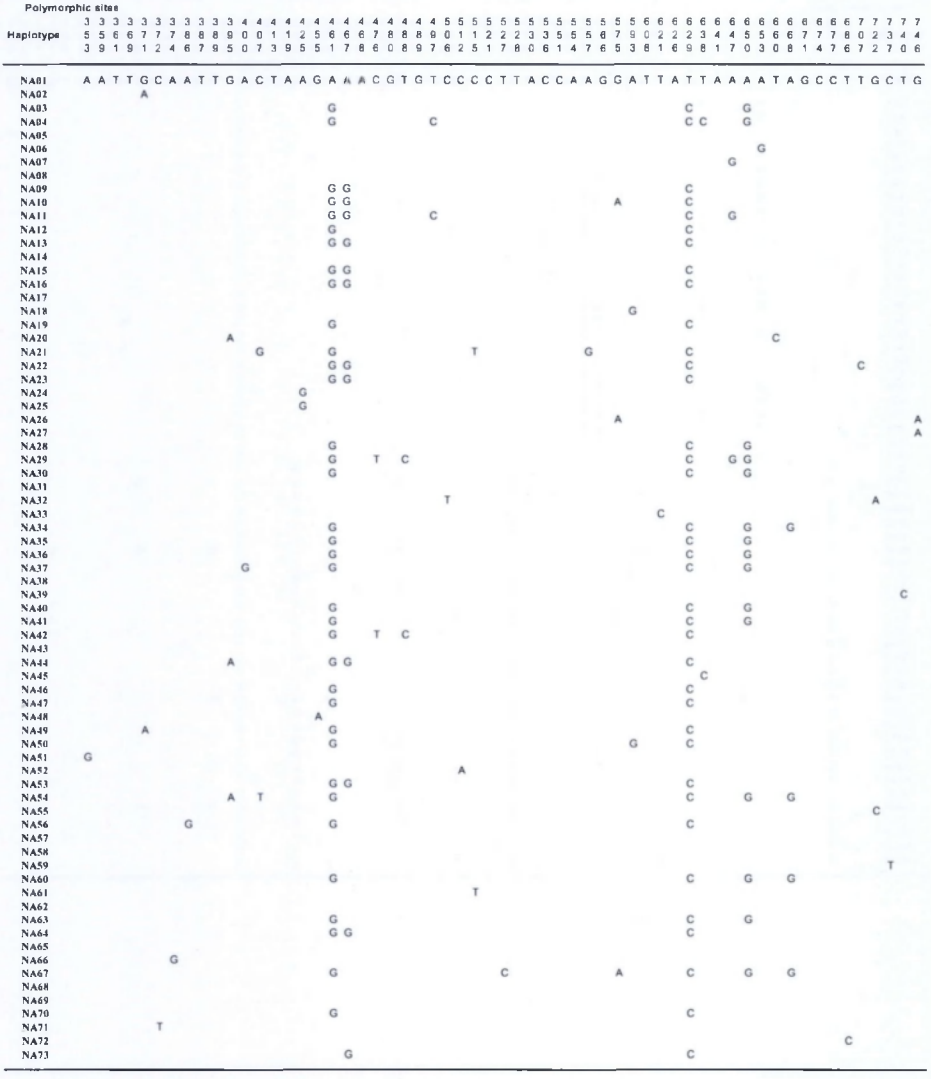
Part 1


Polymorphic sites
Haplotype
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 3
0 1 2 2 2 3 4 5 6 6 6 7 7 8 8 9 9 1 1 2 2 2 2 2 3 3 4 4 4 5 5 5 7 8 0 0 1 1 1 2 2 3 4 5 5 6 7 9 9 4
8 4 0 6 7 8 5 3 0 2 8 4 7 0 6 5 8 0 9 1 2 3 5 8 1 4 0 3 9 3 5 8 0 8 3 6 2 5 8 1 5 3 2 1 4 6 5 3 9 4
NA01 C C A G C C T T C T T C C C A C T A G T G A T A A A A A A A A C C A A G G C T C T C C T C T G G C G A
NA02 G
NA03 C
NA04 C
NA05
NA06
NA07
NA08 A A
NA09 G C
NA10 G C
NA11
NA12 A C
NA13 G
NA14 T
NA15
NA16 C C
NA17 C C
NA18
NA19 A G C
NA20 A
NA21 A T
NA22 A C
NA23 A G G C
NA24
NA25 G
NA26 G
NA27
NA28 C
NA29 A C
NA30 C T
NA31 G
NA32
NA33
NA34 C
NA35 C A
NA36 G C
NA37 C
NA38 A
NA39
NA40 A C
NA41 C
NA42 A C
NA43
NA44 G C
NA45
NA46 C
NA47 C C
NA48
NA49 C
NA50 C C
NA51
NA52
NA53 G C
NA54 C
NA55
NA56 G T C A
NA57
NA58 C C
NA59 T
NA60 C C
NA61
NA62 A
NA63 C A
NA64 G C C
NA65 C
NA66
NA67 C
NA68 G
NA69
NA70 G C
NA71 C
NA72
NA73

Part 2


Polymorphic sites
Haplotype
NA01 C C A G C C T T C T T C C C A C T A G T G A T A A A A A A A A C C A A G G C T C T C C T C T G G C G A
IB01 G G T
IB02 C T
IB03 C T
IB04 C T
IB05 C T
IB06 G C T T
IB07 G G C T
IB08 G C T
IB09 T G C T
IB10 G C T
IB11 C C T G
IB12 C C T
IB13 G C C T
IB14 C C T
IB15 G C C T
IB16 A C T
IB17 C C T
IB18 C T
IB19 C T
IB20 G C C T
IB21 C T
IB22 C T
IB23 G C T
IB24 C T
IB25 C T
IB26 C T T
IB27 C A T
MS01 T G C T C C C A
MS02 G C G T C C C A
MS03 T G G C T C G C C A
MS04 G C T C G C C A
MS05 G C T C G C C A
MS06 A G C T C C C A
MS07 G C G T C C C A
MS08 A G C T C G C C A
MS09 G C T C C
MS10 C T C C C C A
MS11 G C T T G T C C A
MS12 G C T C C C A
MS13 G G C T C C C A
MS14 G C T C C C A
MS15 G C T C C C A
MS16 G G C T C C C A
MS17 G C T C C C A
MS18 G T G C G T C C C A
MS19 C G C G G T C G C C A
MS20 A G C T C C C A
MS21 A A G G C G G T C C C A
MS22 G C T C C C A
MS23 G C G T C C C A
MS24 G C T C C C A
MS25 A G C T C G C C A
MS26 G C G T C C C A
MS27 G C T C G C C A
MS28 G C T C C C A
MS29 C T G C T C G C C A
MS30 G G C G T C C C A A

Part 3


Polymorphic sites
Haplotype
NA01 A A T T G C A A T T G A C T A A G A A A C G T G T C C C C T T A C C A A G G A T T A T T A A A A T A G C C T T G C T G
NA02 A
NA03 G C G
NA04 G C C C G
NA05
NA06 G
NA07 G
NA08
NA09 G G C
NA10 G G A C
NA11 G G C C G
NA12 G C
NA13 G G C
NA14
NA15 G G C
NA16 G G C
NA17
NA18 G
NA19 G C
NA20 A C
NA21 G G T G C
NA22 G G C C
NA23 G G C
NA24 G
NA25 G
NA26 A A
NA27 A
NA28 G C G
NA29 G T C C G G
NA30 G C G
NA31
NA32 T A
NA33 C
NA34 G C G G
NA35 G C G
NA36 G C G
NA37 G G C G
NA38
NA39 C
NA40 G C G
NA41 G C G
NA42 G T C C
NA43
NA44 A G G C
NA45 C
NA46 G C
NA47 G C
NA48 A
NA49 A G C
NA50 G G C
NA51 G
NA52 A
NA53 G G C
NA54 A T G C G G
NA55 C
NA56 G G C
NA57
NA58
NA59 T
NA60 G C G G
NA61 T
NA62
NA63 G C G
NA64 G G C
NA65
NA66 G
NA67 G C A C G G
NA68
NA69
NA70 G C
NA71 T
NA72 C
NA73 G C

Part 4

Polymorphic sites	3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7
Haplotype	3 5 6 6 7 7 8 8 8 9 0 0 1 1 2 5 6 6 7 8 6 0 8 8 9 7 6 2 5 1 7 8 0 6 1 4 7 6 5 3 8 1 6 9 8 1 7 0 3 0 8 1 4 7 6 7 2 7 0 6
	3 9 1 9 1 2 4 6 7 9 5 0 7 3 9 5 5 1 7 8 6 0 8 9 7 6 2 5 1 7 8 0 6 1 4 7 6 5 3 8 1 6 9 8 1 7 0 3 0 8 1 4 7 6 7 2 7 0 6
NA01	A A T T G C A A T T G A C T A A G A A A C G T G T C C C C T T A C C A A G G A T T A T T A A A A A T A G C C T T G C T G
IB01	G A G G C
IB02	G A G G C
IB03	G A G G C A
IB04	G A G G A C
IB05	G A G G A C
IB06	G A G G C
IB07	G A G G C
IB08	G A G G C
IB09	G A G G C
IB10	G A G G C
IB11	G A G G C
IB12	G A G G C
IB13	G A G C
IB14	G A G G C
IB15	G A G G C
IB16	G A G G C
IB17	G A G G C A
IB18	G A G G C
IB19	G A G G C A
IB20	G A G G C
IB21 C	G A G G C
IB22	G A G G G C C
IB23	G A G G C
IB24	G A G G C
IB25	G A G G C
IB26	G A G G C
IB27	G A G G C
MS01 G	A G G A T G C A
MS02 G	T G A T G G C G G A
MS03	G G A T T G C A
MS04 G	G G A T G C
MS05 G	G G A A C T G C
MS06	G G T G C T
MS07 G	T G A T C G G C G G A
MS08	G G T G C T
MS09 G	G G G A T G C C
MS10	G G A T G G G C
MS11 G	G G A T G C A
MS12 G	G A G G A T G C A
MS13 G	G G A T G G C A
MS14 G G	G G G A T G C G
MS15 G	C G G G A T G C
MS16 G	G G A T A G C
MS17 G G	G A G G A T G C
MS18 G A	G G A T A G C C G
MS19 G C	C G G G A T G C A
MS20	G G T A G C T T
MS21 G	G A T G T T G C G
MS22 G	G G A T G C C
MS23 G	G G A T G G C G G
MS24 G	G G G A T G G C C A
MS25	G G T G C
MS26 G	G C G A T G G C G G
MS27 G	G G A C T G C
MS28 G	G G A T G C C
MS29 G G	G G A C T G C
MS30 G	G G A T A G C

Part 5

Haplotype	Polymorphic sites 7 4 7	7 5 8	7 5 9	7 6 4	7 6 7	7 7 3	7 7 6	7 7 9	7 8 1	7 9 7	8 0 3	8 1 2	8 1 5	8 2 1	8 2 4	8 3 0	8 3 1	8 3 9	8 4 1	8 4 2	Accession Number
NA01	T	A	A	G	A	T	T	G	T	A	G	G	C	C	C	T	G	G	C	C	EU736948
NA02																					EU736949
NA03	C											A									EU736950
NA04												A									EU736951
NA05					G																EU736952
NA06																					EU736953
NA07																					EU736954
NA08																					EU736955
NA09												A						A			EU736956
NA10												A						A			EU736957
NA11				A								A						A			EU736958
NA12												A									EU736959
NA13												A						A			EU736960
NA14																					EU736961
NA15												A						A			EU736962
NA16												A						A			EU736963
NA17																					EU736964
NA18										C											EU736965
NA19																					EU736966
NA20			G																		EU736967
NA21												A									EU736968
NA22												A						A			EU736969
NA23												A						A			EU736970
NA24													T								EU736971
NA25																					EU736972
NA26																					EU736973
NA27																					EU736974
NA28												A									EU736975
NA29												A									EU736976
NA30												A									EU736977
NA31																					EU736978
NA32																					EU736979
NA33																					EU736980
NA34												A									EU736981
NA35												A									EU736982
NA36												A									EU736983
NA37												A									EU736984
NA38																					EU736985
NA39	C																				EU736986
NA40												A									EU736987
NA41												A	T								EU736988
NA42	C		G									A									EU736989
NA43													T								EU736990
NA44												A						A			EU736991
NA45																					EU736992
NA46																					EU736993
NA47																					EU736994
NA48										G											EU736995
NA49									C												EU736996
NA50																					EU736997
NA51																					EU736998
NA52																					EU736999
NA53								A				A						A			EU737000
NA54												A									EU737001
NA55																					EU737002
NA56												A									EU737003
NA57																			G		EU737004
NA58																					EU737005
NA59																					EU737006
NA60												A									EU737007
NA61																					EU737008
NA62																					EU737009
NA63												A									EU737010
NA64												A						A			EU737011
NA65																					EU737012
NA66																					EU737013
NA67												A									EU737014
NA68																					EU737015
NA69										G											EU737016
NA70																					EU737017
NA71																					EU737018
NA72																					EU737019
NA73																					EU737020

Part 6

Haplotype	Polymorphic sites 7457	7589	7656	7747	7736	7791	7817	7890	8013	8125	8141	8223	8301	8331	8419	8442	Accession Number
NA01	T	A	A	G	A	T	T	G	T	A	G	G	C	C	C	T	G	G	C	C	EU736948
1B01													T								EU737021
1B02																					EU737022
1B03																					EU737023
1B04												A									EU737024
1B05																					EU737025
1B06																					EU737026
1B07																					EU737027
1B08																					EU737028
1B09																		A			EU737029
1B10																					EU737030
1B11																					EU737031
1B12																					EU737032
1B13																					EU737033
1B14																					EU737034
1B15																					EU737035
1B16																					EU737036
1B17																					EU737037
1B18												A									EU737038
1B19																					EU737039
1B20																					EU737040
1B21																					EU737041
1B22																					EU737042
1B23																				T	EU737043
1B24																					EU737044
1B25	C																				EU737045
1B26																					EU737046
1B27																					EU737047
MS01						C	A					A			T			A			EU737048
MS02						C	A					A						A			EU737049
MS03						C	A					A						A			EU737050
MS04		G				C	A					A			T			A			EU737051
MS05						C	A					A						A			EU737052
MS06						C	A					A						A			EU737053
MS07						C	A					A						A			EU737054
MS08						C	A					A						A			EU737055
MS09						C	A					A						A			EU737056
MS10						C	A					A						A			EU737057
MS11					G	C	A					A						A			EU737058
MS12						C	A				A	A						A			EU737059
MS13						C	A					A						A			EU737060
MS14						C	A				A	A						A			EU737061
MS15						C	A					A						A			EU737062
MS16						C	A											A			EU737063
MS17						C	A				A	A						A			EU737064
MS18						C	A					A						A			EU737065
MS19						C	A					A						A			EU737066
MS20						C	A					A						A			EU737067
MS21						C	A					A						A			EU737068
MS22						C	A	A				A				C		A			EU737069
MS23				A		C	A					A						A			EU737070
MS24						C	A									C		A			EU737071
MS25						C	A					A						A			EU737072
MS26						C	A					A						A			EU737073
MS27						C	A					A						A			EU737074
MS28						C	A	A				A					C	A			EU737075
MS29	C					C	A					A						A			EU737076
MS30						C	A					A						A			EU737077

Table S3.2 Haplotype distribution of the cyt *b* haplotypes of *Pomatoschistus minutus*, the columns refer to a single population, while the rows refer to the distribution of the haplotypes. The haplotypes are indicated by numbers as given in Table S3.1.

Part 1

Haplotype	Locations	TBS	HBS	PBS	WCS	RNS	BNS	BIS	WIS	GOA	PAO	AAO	VMS1	VMS2
NA01		3	5	3	5	1	3	5	17	10				
NA02								1						
NA03				1										
NA04				1										
NA05				1										
NA06				1										
NA07				2										
NA08				1	1									
NA09				1		7	4							
NA10				1										
NA11										1				
NA12										1				
NA13										2				
NA14										1				
NA15					1					3				
NA16										1				
NA17										1				
NA18										1				
NA19										1				
NA20										1				
NA21										1				
NA22										1				
NA23										1				
NA24										1				
NA25										1				
NA26										1				
NA27										1				
NA28		14	8		2	1								
NA29			1											
NA30		2	1											
NA31		1	1			1			2					
NA32		1	1											
NA33		1	1											
NA34		3	2						1					
NA35			1											
NA36			1											
NA37			1											
NA38			1											
NA39		1												
NA40		1												
NA41		1												
NA42		1												
NA43						1								
NA44						1								
NA45						1								
NA46						3	1							
NA47					1	1	3							
NA48						1								
NA49						1								
NA50						1								

Part 2

Haplotype	Locations												
	TBS	HBS	PBS	WCS	RNS	BNS	BIS	WIS	GOA	PAO	AAO	VMS1	VMS2
NA51					1								
NA52					1								
NA53						1							
NA54				1									
NA55				1									
NA56				1									
NA57				1									
NA58				1									
NA59				1									
NA60				1									
NA61								1					
NA62								1					
NA63								1					
NA64								1					
NA65								1					
NA66								1					
NA67								1					
NA68								1					
NA69								1					
NA70								1					
NA71							1						
NA72							1						
NA73							1						
IB01											1		
IB02										5	7		
IB03											1		
IB04											1		
IB05											1		
IB06											1		
IB07											1		
IB08											3		
IB09											1		
IB10											2		
IB11											2		
IB12											1		
IB13											2		
IB14											2		
IB15											1		
IB16											1		
IB17											1		
IB18											1		
IB19											1		
IB20											1		
IB21											1		
IB22											1		
IB23											1		
IB24										1			
IB25										1			
IB26										2			
IB27										1			

Part 3

Haplotype	Locations												
	TBS	HBS	PBS	WCS	RNS	BNS	BIS	WIS	GOA	PAO	AAO	VMS1	VMS2
MS01												1	
MS02												1	1
MS03												2	
MS04												1	
MS05												1	
MS06												1	
MS07												1	
MS08												1	
MS09												1	
MS10												1	
MS11												1	
MS12												1	
MS13												1	
MS14												1	
MS15												1	
MS16												1	
MS17												1	
MS18												1	
MS19												1	
MS20												1	
MS21												1	
MS22												1	
MS23												1	1
MS24												1	
MS25													1
MS26													1
MS27													1
MS28													1
MS29													1
MS30													1

Table S3.3 Results of the times to Most Recent Common Ancestor (tMRCAs) and results of the neutrality tests and mismatch analysis for the total data set and two phylogeographic groups of *Pomatoschistus minutus* based on D-loop data. The tMRCAs (Ma) were estimated using the 6% and 8% molecular clocks, with the credibility intervals of tMRCA between brackets. D, Tajima's D-test and F_s, Fu's F_s-test (both in bold when p-value < 0.05); τ, time since expansion measured in mutational time units; t_e, absolute time (Ma) since expansion estimated with the 6% and 8% molecular clocks; P(rag), p-value for the raggedness index; P(SDD), p-value for the sum of the squared deviations under the hypothesis of sudden expansion.

					Demographic mismatch distribution					Spatial mismatch distribution				
	tMRCA (6%)	tMRCA (8%)	D	F_S	τ	t_e (6%)	t_e (8%)	P(rag)	P(SDD)	τ	t_e (6%)	t_e (8%)	P(rag)	P(SDD)
Total data set	0.475 (0.612-0.450)	0.356 (0.459-0.394)												
Mediterranean Sea	0.382 (0.507-0.286)	0.287 (0.380-0.214)	-0.098	**-3.599**	10.969	0.217	0.163	0.733	0.434	10.971	0.217	0.163	0.825	0.843
North Atlantic	0.373 (0.480-0.284)	0.280 (0.360-0.213)	-0.171	**-12.319**	6.537	0.130	0.098	0.686	0.507	4.252	0.085	0.064	0.623	0.312

Table S3.4 Pairwise F_{ST}-values among the sand goby populations for the cyt *b* locus (significant p-values ($p<0.05$) are in bold; * significant after Bonferroni correction). VMS is the combination of samples VMS1 and VMS2.

	HBS	PBS	WCS	RNS	BNS	BIS	WIS	GOA	PAO	AAO	VMS
TBS	*-0.00709*	**0.22236**	**0.16530**	**0.19646***	**0.24319***	**0.46590***	**0.38441***	**0.25724***	**0.66470***	**0.65877***	**0.68771***
HBS	-	**0.11068**	*0.06524*	**0.14228**	**0.18876**	**0.33582***	**0.25943***	**0.16527***	**0.63442***	**0.64053***	**0.67006***
PBS		-	*-0.02943*	*0.06609*	*0.10899*	*0.05634*	*0.02551*	*-0.01165*	**0.60164***	**0.63699***	**0.64266***
WCS			-	**0.09255**	**0.13266**	*0.07533*	*0.03582*	*0.01545*	**0.58653***	**0.61581***	**0.65041***
RNS				-	*-0.04899*	**0.27558***	**0.24451***	**0.08644**	**0.53784***	**0.55565***	**0.64462***
BNS					-	**0.37736**	**0.33046***	**0.11463**	**0.58116***	**0.59389***	**0.62738***
BIS						-	*0.00205*	*0.05785*	**0.75634***	**0.73889***	**0.68100***
WIS							-	**0.06409**	**0.72378***	**0.70895***	**0.72407***
GOA								-	**0.55549***	**0.57908***	**0.66075***
PAO									-	**0.14374***	**0.70567***
AAO										-	**0.75616***
VMS											-

Appendix S3.3 Supplementary figures

Figure S3.1 Non-metric multidimensional scaling (NMDS) plots of the sand goby populations based on genetic distances (Slatkin 1995); (a) entire cyt *b* dataset of *Pomatoschistus minutus* (stress value: 0.003) (b) cyt *b* dataset of North Atlantic (NA-) Group (stress value: 0.004), (c) entire allozyme dataset *P. minutus* (stress value: 0.013). VMS is the combination of samples VMS1 and VMS2. Ber: Bergen (Northern North Sea); Oba: Oban (Irish Sea); Tex: Texel (North Sea); Per: Pérols (Mediterranean Sea); Str: Stroombank, Kwi: Kwintebank and Raa: Raan (Southern North Sea).

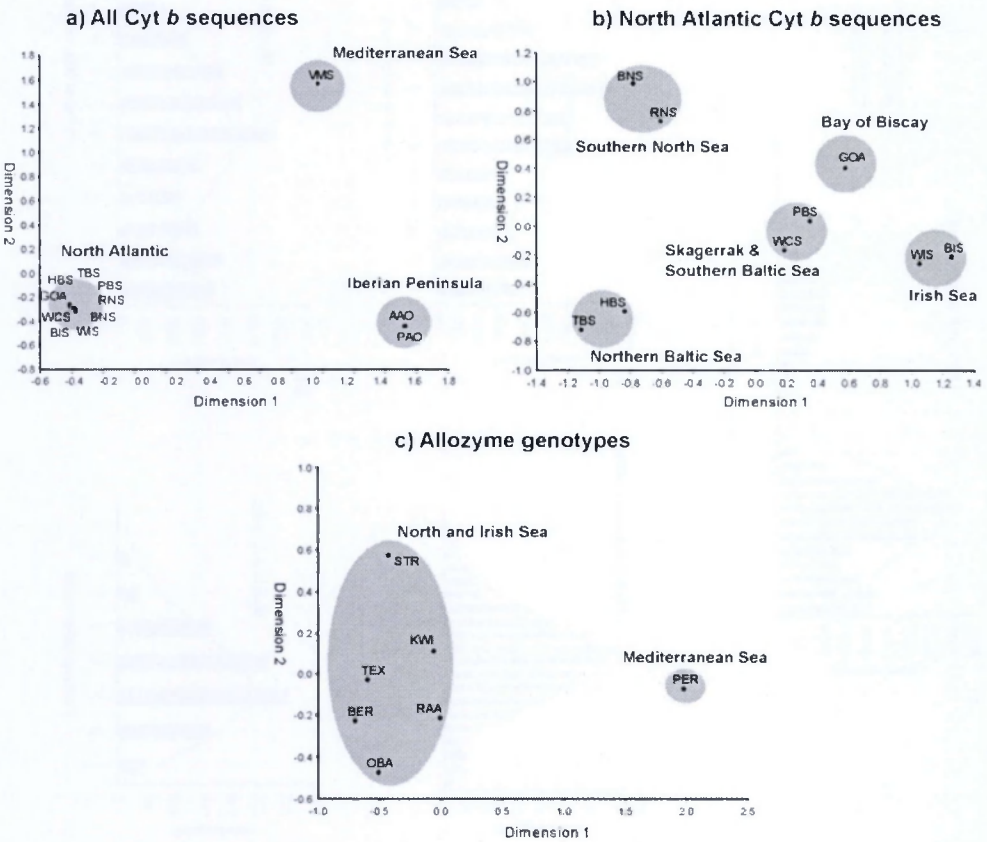


Figure S3.2 Mismatch distribution of cyt *b* for the significant geographical partitions of *Pomatoschistus minutus*; (a) complete data set, (b) Mediterranean Sea (MS-) Clade, (c) Atlantic Ocean (AO-) Clade, (d) Iberian Peninsula (IB-) Group and (e) North Atlantic (NA-) Group. Bars represent the observed frequencies of nucleotide differences between pairs of individuals, curves correspond to the fitted distribution following a model of population expansion.

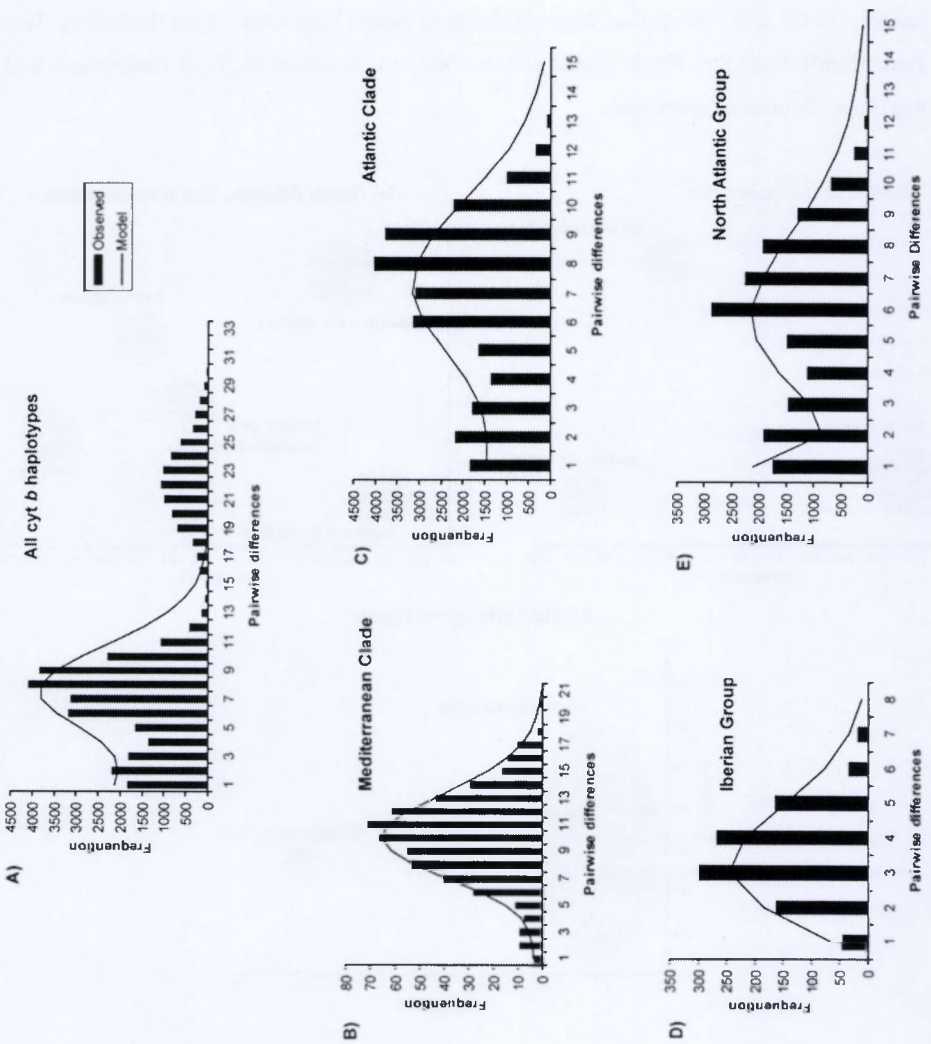


Figure S3.3 Mismatch distribution of the D-loop sequences for the significant geographical partitions of *Pomatoschistus minutus*; a) entire dataset, b) North Atlantic haplotypes and (c) Mediterranean haplotypes. Bars represent the observed frequencies of nucleotide differences between pairs of individuals, curves correspond to the fitted distribution under a model of population expansion.

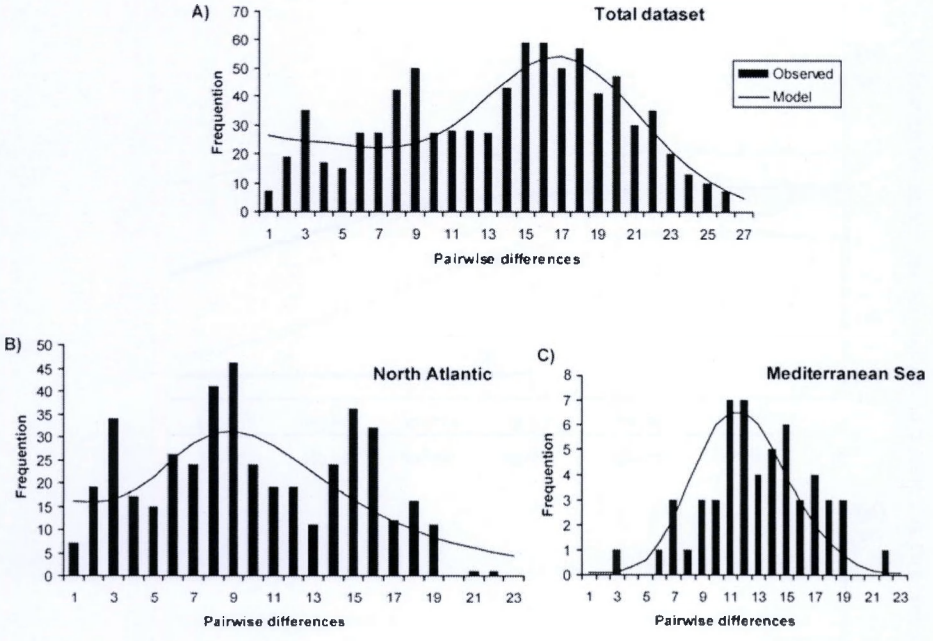
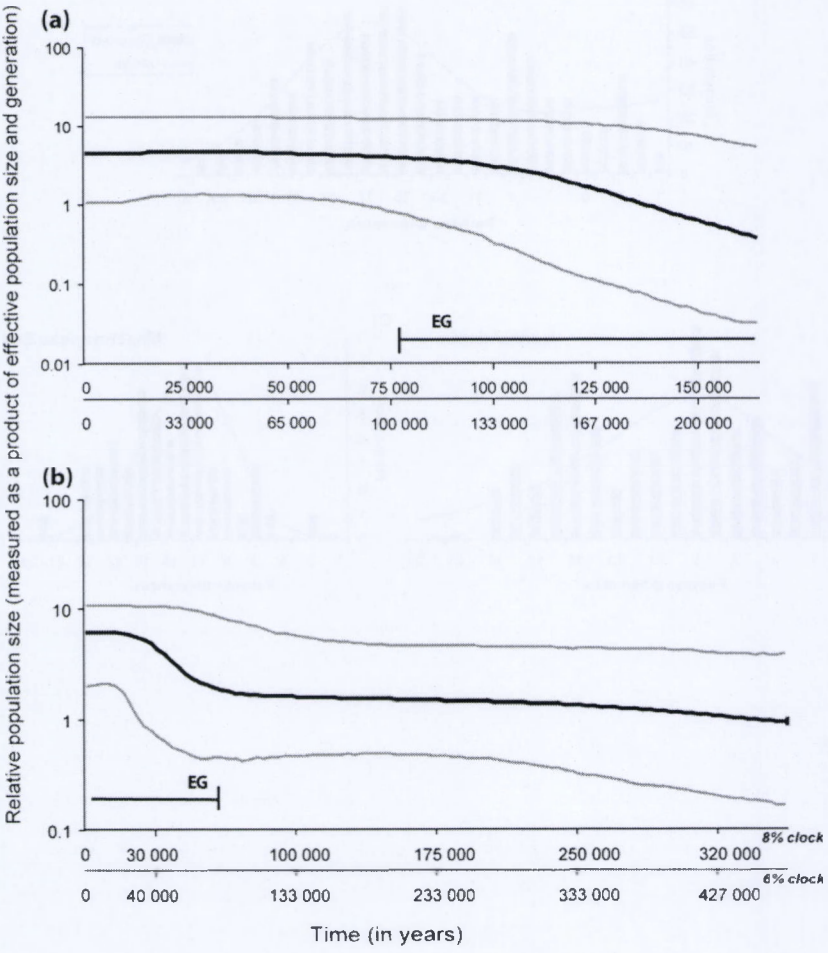


Figure S3.4 Bayesian skyline plot of the D-loop sequences in *Pomatoschistus minutus* with the 8% and 6% molecular clock and with indication of the period of expansion growth (EG). X axis: time (in years), the upper axis is the time measured with the 8% molecular clock, the lower axis shows the time with the 6% molecular clock; Y axis: relative population size (measured as a product of effective population size and generation); (a) Mediterranean haplotypes and (b) North Atlantic haplotypes.



SUBCHAPTER 3B

MITO-NUCLEAR DISCORDANCE IN THE DEGREE OF POPULATION DIFFERENTIATION IN A MARINE GOBY *POMATOSCHISTUS MINUTUS*

Maarten H.D. Larmuseau[1], Joost A.M. Raeymaekers[1], Bart Hellemans[1], Jeroen K.J. Van Houdt[1,2], & Filip A.M. Volckaert[1]

[1]Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Leuven, Belgium.

[2]Royal Museum for Central Africa, Tervuren, Belgium.

Submitted manuscript

Abstract

An increasing number of phylogeographic studies on marine species shows discordant patterns in the degree of population differentiation between nuclear and mitochondrial markers. To understand better which factors have the potential to cause these patterns of discordance in marine organisms, a population genetic study was realised on the sand goby *Pomatoschistus minutus* (Pallas 1770; Gobiidae, Teleostei). Sand gobies from eight European locations were genotyped at eight microsatellite markers. Microsatellites confirmed the global phylogeographical pattern of *P. minutus* observed with mitochondrial DNA (mtDNA) markers and nuclear allozyme markers. Three groups consistent with the mitochondrial lineages were defined (Mediterranean, Iberian and North Atlantic group) and indications for a recent founder-event in the northern Baltic Sea were found. Nevertheless, differences in the degree of population differentiation between the nuclear and mitochondrial markers were large (global F_{ST}-values for microsatellites = 0.0121; for allozymes = 0.00831; for mtDNA = 0.4293). Selection, sex-biased dispersal, homoplasy and a high effective population size are generally accepted as explanations for this mito-nuclear discrepancy in the degree of population differentiation. Selection on mtDNA and microsatellites, and

homoplasy on microsatellite markers, however, are unlikely to be a main cause for *P. minutus*. Also there was no indication for male-biased dispersal in the sand goby. The most likely reason for the discordant pattern is a recent demographical expansion resulting in high effective population sizes and preventing microsatellites to reach mutation-drift equilibrium.

Keywords: Bayesian assignment test, effective population size, genetic distance, Gobiidae, marine fish, north-eastern Atlantic, microsatellites, Procrustes analyses

Introduction

Traditionally, the genetic structure of marine organisms has been thought to be homogeneous due to the lack of obvious barriers to gene flow in the 'open' marine environment. In the last decade, however, an increasing number of population genetic studies described distinct genetic structuring for several marine species on large and small geographical scales (Hauser & Carvalho 2008). Those observed population genetic structures reflect both historical and contemporary processes (Balloux & Lugon-Moulin 2002). Geographical and climatic factors acting during the Pleistocene glaciations (1800 - 11.5 ka) are the major factors responsible for the present genetic structure of most extant marine species (Hewitt 2000). Heterogeneity in the marine environment owing to the influence of climate, hydrodynamics and topography, together with biological traits, such as sex-dependent migration, philopatry and assortative mating, which may counteract gene flow, enhance genetic structuring (Ruzzante *et al.* 1998).

These recent insights are mainly due to the increased popularity of polymorphic microsatellite markers in marine population and landscape genetics (Jørgensen *et al.* 2005). Microsatellites have proven for many species to be more powerful for resolving population structure than mitochondrial DNA (mtDNA) and allozyme markers (e.g. Nesbø *et al.* 2000; De Innocentiis *et al.* 2004). However, this is not applicable to all marine organisms (Lukoschek *et al.* 2008). There is almost certainly a publication bias for marine studies detecting micro-scale population structures with microsatellites (Hauser & Carvalho 2008). An increasing number of studies shows a large difference in the order of magnitude for the population divergence between nuclear and mitochondrial markers. Among those mtDNA

shows a higher sensitivity to resolve the phylogeographical and population genetic structure (Peijnenburg *et al.* 2006; Lukoschek *et al.* 2008). This discordance in resolution among markers may result from the differential effects of genetic drift, mutation and migration on a marker class, or may result from selection or sex-biased dispersal (Buonaccorsi *et al.* 2001). More research is required to study how common and important those factors are in the marine environment.

The sand goby *Pomatoschistus minutus* (Pallas 1770; Gobiidae, Teleostei) is a small marine demersal fish common in shallow waters along European coasts (Miller 1986, Fig. 3.6). Although earlier analyses with allozyme markers showed very low population differentiation values for *P. minutus* (Stefanni *et al.* 2003), a recent study revealed that the sand goby is highly structured at the mtDNA cyt *b* locus (*subchapter 3a*). Based on mtDNA, middle Pleistocene glaciations yielded three isolated and different evolving sets of sand goby populations. Reciprocal monophyly was observed between a Mediterranean Sea (MS) and an Atlantic Ocean (AO) Clade. The AO-Clade contains two ESUs: the Iberian Peninsula (IB) Group and the North Atlantic (NA) Group. For the NA-Group, there is evidence for geographic sorting of the ancestral mtDNA haplotypes with recent independent radiations in the Baltic Sea, Irish Sea, North Sea and Bay of Biscay (*subchapter 3a*). Analyses with allozyme markers only showed divergence between Mediterranean and Atlantic populations for *P. minutus* but with very low detected differentiation, attributed to high contemporal gene flow throughout its whole distributional range (Stefanni *et al.* 2003). However, the biology and morphology of sand gobies indicate reduced levels of contemporary gene flow (Miller 1986). Moreover, the mitochondrial phylogeography invalidates the interpretation of the allozyme data because of the high F_{ST} values between sand goby populations and the lack of common haplotypes between the three ESUs (*subchapter 3a*). The reasons for this discordant pattern can be surveyed by using another set of nuclear markers, such as DNA microsatellites. In contrast to allozymes, DNA microsatellite markers promise a higher resolution for population differentiation in the sand goby because of a higher mutation rate (Chistiakov *et al.* 2006). Moreover, Pampoulie *et al.* (2004a) detected genetic structure on a micro-scale for *P. minutus* in the southern North Sea using microsatellites, while Gysels *et al.* (2004c) did not in the same region using allozyme markers.

In this study, two hypotheses are tested: a) is the phylogeographical structure of the sand goby observed by using two markers, the mtDNA cyt *b* locus (*subchapter 3a*) and allozymes (Stefanni *et al.* 2003) confirmed with another type of nuclear marker, namely microsatellites and b) do microsatellite markers in comparison with mitochondrial markers and allozymes show a superior resolution to reveal the phylogeographical and population genetic structure.

Materials & Methods

Sampling and species/sex identification

A total of 696 *Pomatoschistus minutus* individuals were caught at eight locations along the European coast between September 2002 and January 2007 (Table 3.4, Fig. 3.6). Many of the samples have already been included in a previous mtDNA study (*subchapter 3a*). New samples were identified as *P. minutus* morphologically, based on the dermal head papillae (Miller 1986) and pigmentation pattern (Hamerlynck 1990), and genetically according to a molecular tool described in *subchapter 2a*. The sex of each sand goby was determined by the shape of the urogenital papilla as drawn in Rodrigues *et al.* (2006).

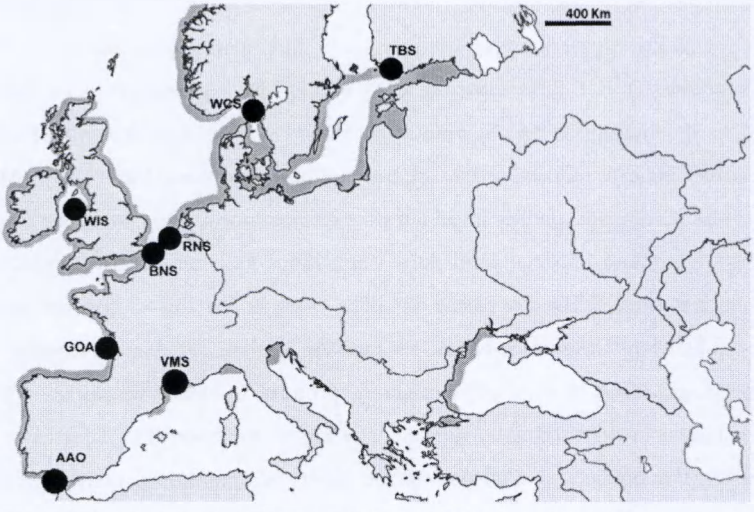


Figure 3.6 Geographical distribution of the eight sampling locations of sand goby *Pomatoschistus minutus*. The shaded area represents the distribution range of *P. minutus* according to Miller (1986). See Table 3.4 for sample codes.

Table 3.4 Overview of the ten samples of *Pomatoschistus minutus* collected at eight different locations. N indicates sample size.

Code	Area	Country	Location	Date	Coordinates	N
TBS	Northern Baltic Sea	Finland	Tvärminne	Jul/06	59°50'N - 23°12'E	96
WCS	Skagerrak	Sweden	Bökevik Bay, Skaftö island	Jun/06	58°14'N - 11°26'E	88
RNS	North Sea	The Netherlands	Renesse	Nov/05	51°44'N - 03°47'E	48
BNS1	North Sea	Belgium	Oostduinkerke	Nov/06	51°08'N - 02°40'E	47
BNS2	North Sea	Belgium	Oostduinkerke	Sep/02	51°08'N - 02°40'E	95
WIS	Irish Sea	Wales (UK)	Llanfairfechan	Nov/06	53°59'N - 03°59'W	45
GOA	Bay of Biscay	France	Gironde estuary	aug/06	45°36'N - 01°01'W	40
AAO	Iberian Peninsula	Spain	Guadalquivir river estuary	Nov/06	36°58'N - 06°10'W	95
VMS1	Mediterranean Sea	France	Vaccarès lagoon	Jan/06	43°32'N - 04°35'E	94
VMS2	Mediterranean Sea	France	Vaccarès lagoon	Jan/07	43°32'N - 04°35'E	48

Microsatellite genotyping

Genomic DNA was extracted from fin clips stored in 100% ethanol using the NucleoSplin Extraction Kit (Machery-Nagel GmBH, Düren, Germany). Each individual was genotyped at nine multiplexed microsatellite loci (Pmin03, Pmin04, Pmin09, Pmin16-2, Pmin20, Pmin29, Pmin31, Pmin35 and Pmin38) (*subchapter 2b*) on an ABI 3130 automated capillary DNA sequencer (Applied Biosystems).

Several methods were used to mitigate genotyping errors in the dataset. First, to avoid contamination, different pipettes, aerosol barrier tips and different sections of the laboratory were used for pre- and post-PCR work. Every 15th individual (corresponding with one every two rows of a PCR-plate) a negative control was used to detect contamination. Accordingly, only high DNA quantity and quality was used because contaminant molecules have a higher probability of being amplified when the number of template DNA molecules is low (Pompanon *et al.* 2005). No contamination occurred during the screening procedures. Second, care was taken to ensure that no allele-calling error was present in the dataset. The allele class boundaries were manually identified because of the high error rate with

automated binning approaches (Amos *et al.* 2007). Third, the potential occurrence of null alleles and scoring errors due to stuttering or large allele dropout in the data set was assessed using the software MICRO-CHECKER v. 2.2.3 (Van Oosterhout *et al.* 2004) and DROPOUT v. 2.0 (McKelvey & Schwartz 2005). No scoring errors were detected in the dataset. The used microsatellites were already selected on high quality in *subchapter 2b.* Fourth, rarefaction curves were realized to determine the minimal sample size required for the degree of accuracy in allele numbers in a particular population. The curves were made using the DOH program (Brzustowski 2002) for the populations with more than 90 individuals. For all loci, 40-45 individuals per population were the minimum to genotype, except for locus Pmin20 that still was in the exponential phase even after genotyping 90 individuals. Therefore, this locus was excluded from further statistical analysis. Finally, to verify the reproducibility of results, 10-30% samples of the analyses were run twice for all microsatellite markers, as was recommended by Pompanon *et al.* (2005). These replicates were carried out with a systematic duplication of the samples during sample collection. No difference was noticed in the comparison between the blind samples and the original experiments, suggesting a negligible error rate.

Genetic diversity

We used genotype and allele frequencies of the microsatellite loci to obtain standard estimates of genetic diversity within and between sample sites. Genetic variation in each population was measured by calculating the mean number of alleles per locus, the observed (H_O) and unbiased expected (H_E) heterozygosities and the F_{IS}. The deviation from Hardy-Weinberg equilibrium was assessed with GENETIX v. 4.05 (Belkhir *et al.* 2004). Allelic richness, which corrects the number of alleles for sample size, was assessed using FSTAT v. 2.9.3.2 (Goudet 2001). Exact tests of linkage disequilibrium (LD) between pairs of loci were calculated at each location, each region and across all samples using GENEPOP v. 3.4. (Raymond & Rousset 1995).

Patterns of population subdivision

Different methods were used to reveal the population substructure of *P. minutus*. First, a factorial correspondence analysis (FCA) of individual multilocus genotypes was performed using GENETIX to reveal the portion of the hyperspace of all genotypes occupied by each group of individuals. Second, population differentiation was quantified in GENETIX using the standardized allelic variation F_{ST}, estimated as θ (Weir & Cockerham 1984) and in SPAGeDi 1.2g (Hardy & Vekemans 2002) using an analogue of F_{ST} for microsatellites R_{ST}, estimated as ϱ (Slatkin 1995). F_{ST}-linked pairwise genetic distances were calculated according to Cavalli-Sforza & Edwards (1967) (D_{CE}) with GENETIX, and R_{ST}-linked pairwise genetic distances calculated according to Goldstein *et al.* (1995) ($d\mu^2$) were obtained with SPAGeDi. F_{ST}s and D_{CE}s were tested for significance against 10^4 random permutations of the data in GENETIX. The significance for R_{ST}s and $d\mu^2$s were tested in SPAGeDi also against 10^4 random permutations. A sequential Bonferroni test was applied to correct significance levels for multiple testing (Rice 1989). Standardized genetic differentiation measures were obtained by dividing F_{ST} measures by the maximum values for F_{ST} (Hedrick 2005; Meirmans 2006), calculated using the pragmatic recoding approach suggested by Meirmans (2006). In order to assess the influence of stepwise-like mutations versus drift on genetic differentiation, we performed a permutation test available in the software SPAGeDi. Allele size at each locus was randomly permutated among allelic states (2000 mutations) to simulate a distribution of R_{ST} values (pR_{ST}) and 95% confidence intervals (CI) under the null hypothesis that differences in allele sizes do not contribute to population differentiation (Hardy *et al.* 2003). Third, the classical multidimensional scaling analyses (CMDS) based on the two types of genetic distances were obtained using the Vegan package in R (Oksanen *et al.* 2007). Ordination plots with a stress value below 0.20 provide interpretable information concerning intersite relationships (Clarke 1993). Fourth, to analyse the effect of geographical distance on genetic distance, the Mantel test in GENETIX (Belkhir *et al.* 2004) was used, which computes the correlation between distance matrices by means of a permutation procedure (Mantel 1967; Smouse *et al.* 1986). Geographical distances were obtained as the shortest coastal distances between sites using the electronic atlas ENCARTA (Microsoft 2001). Both types of pairwise genetic distances, D_{CE} and $d\mu^2$, were used for the Mantel test. Finally, a Bayesian clustering analysis was realised for the microsatellite data using the program

STRUCTURE v. 2.2 (Pritchard *et al.* 2000). This approach that estimates the number of independent genetic clusters in the data set does not require *a priori* information about population structure, and thus, provides an estimate of genetic structure independent of the origin of samples. We used the no-admixture algorithm without prior population information and used 10,000 runs as burn-in and 100,000 runs for each of three Markov chains. For each simulation of K = 1-10 (no. of clusters), we used ten replicates. We selected the most likely number of clusters given the data by choosing the number of clusters where we observed the largest difference in log likelihoods (ΔK) (Evanno *et al.* 2005).

Comparison between microsatellite and mitochondrial data

MtDNA cyt *b* data were available for eight out of ten samples (TBS, WCS, RNS, BNS1, WIS, GOA, AAO and VMS1; *subchapter 3a*). On the other hand, the allozyme analyses in Gysels (2003) and in Stefanni *et al.* (2003) were restricted to samples from other locations. Therefore, it was only possible to compare the results of microsatellites and mtDNA cyt *b* data. Two statistical methods were used to compare the degree of population differentiation between the types of genetic distances calculated for the microsatellite data (D_{CE} and $d\mu^2$) and the genetic distances of Tamura & Nei (1993) calculated for the mtDNA data. First, the pairwise F_{ST} matrices were correlated by using simple Mantel procedures (Mantel 1967) in the Vegan library in R (Oksanen *et al.* 2007). Permutations (n = 10,000) were used to evaluate statistical significance. Then, two-dimensional MDS ordinations of the two types of genetic marker were compared by a Procrustes Analysis (PA) (Gower 1975) using R software. PA is searching for the best match between two configurations of points in a multivariate Euclidean space using rotation, translation, reflection, and dilation of one configuration. The criterion used to assess the best fit is the minimization of the sum of squares between the differences for each observation (m^2). The significance of the result, an optimal superposition of one configuration on the other (reference) configuration is obtained through a permutation test (PROTEST, Jackson 1995). PROTEST is using $R = \text{sqrt}(1-m^2)$ as a test statistic, which can be interpreted as a correlation.

Impact of selection, sex-biased dispersal and effective population size on the degree of population divergence estimates

To verify if the variation on the microsatellites can be influenced by selection, two different approaches were obtained. First, all microsatellite flanking regions were compared to sequences in GenBank by means of the BLAST program (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul *et al.* 1990) to verify if the microsatellites potentially are located within described functional regions of the genome. Second, potential outlier microsatellites were identified by using the selection detection workbenches LOSITAN (Antao *et al.* 2008) and BayeScan (Foll & Gaggiotti 2008). Analyses are performed for LOSITAN with 10,000 simulations for IAM and SMM, both with the options 'Neutral' mean F_{ST} and force mean F_{ST}. For BayeScan, 10,000 iterations were done with a thinning interval of 20 and with 10 pilot runs.

To detect differences in migration rates between females and males, deviation from Hardy-Weinberg equilibrium (F_{IS}), differentiation among populations (F_{ST}), relatedness (r) (Queller & Goodnight 1989), mean assignment index (mAI_C) and variance of the assignment index (vAI_C) were quantified separately for both sexes over all populations (Goudet *et al.* 2002). Statistical significance of differences in these within-population indices was determined with 10,000 permutations using the randomization method implemented in FSTAT (Goudet 2001).

Effective population sizes were estimated using different point methods, which do not require samples spaced over at least one temporal interval, and temporal methods, where samples are taken from the same location at two or more points in time separated by a specified number of generations (one or more). Two different point estimation methods were estimated for the ten samples. The program NeEstimator v. 1.3 (Peel *et al.* 2004) was used for the Heterozygote Excess method that examines the excess of heterozygotes in the sample compared to the proportion predicted under Hardy-Weinberg equilibrium (Luikart & Cornuet 1999). For the estimation of effective population size based on linkage disequilibrium data, the program LDNe v. 1.31 (Waples & Do 2008) was used because it implements a recently developed bias correction (Waples 2006). Finally, N_e was estimated

using two temporal methods in NeEstimator v. 1.3: (i) based on the Moment Approach (Waples 1989) and (ii) based on the pseudo-likelihood method (MLNE) developed by Wang & Whitlock (2003). These two methods require at least two temporally spaced samples and therefore this analysis was only possible for VMS and BNS (Table 3.4).

Results

Genetic diversity

Mean allelic richness per location, corrected for sample size, varied between 12.088 (northern Baltic Sea) and 15.454 (Mediterranean Sea) (mean = 13.622) (Table 3.5). Mean expected heterozygosity was relatively uniform among the different sampling sites with the lowest value in the northern Baltic Sea sample (0.740) and the highest in the Mediterranean Sea samples (0.781) (mean = 0.756) (Table 3.5). Locations WCS and BNS2 showed a significant departure from Hardy-Weinberg equilibrium (Table 3.5) (both multilocus F_{IS} = 0.050). Pairwise comparisons between loci revealed no significant linkage disequilibrium after sequential Bonferroni corrections.

Table 3.5 Estimates of genetic diversity of the ten samples of *Pomatoschistus minutus* based on eight microsatellite markers. $H_{E\,n.b.}$ is the unbiased expected heterozygosity, H_O the observed heterozygosity, and F_{IS} measures deviation from Hardy-Weinberg equilibrium. Statistically significant F_{IS} values are listed in bold. See Table 3.4 for sample codes.

Population	Sample Size	$H_{E\,n.b.}$	H_O	F_{IS}	Allelic richness
TBS	96	0.740	0.718	0.030	12.088
WCS	88	0.760	0.722	**0.050**	13.211
RNS	48	0.752	0.755	-0.003	12.735
BNS1	47	0.749	0.689	0.081	12.319
BNS2	95	0.762	0.724	**0.050**	13.404
WIS	45	0.754	0.724	0.040	13.534
GOA	40	0.752	0.732	0.027	14.087
AAO	95	0.752	0.714	0.051	14.015
VMS1	94	0.781	0.766	0.019	15.454
VMS2	48	0.766	0.723	0.057	15.375

Patterns of population subdivision

First, the graphical distribution of populations from the factorial correspondence analysis showed that the two Mediterranean samples (VMS1 and VMS2) clustered together, as well as all Atlantic samples, except the one from the Iberian Peninsula (AAO) (Fig. 3.7). Disregarding the AAO sample, the sample of the Bay of Biscay (GOA) was the most aberrant among the Atlantic samples. The distribution of all individuals in a FCA plot (graph not shown), shows a small overlap between the Mediterranean and Atlantic samples as well as between AAO and the other Atlantic samples.

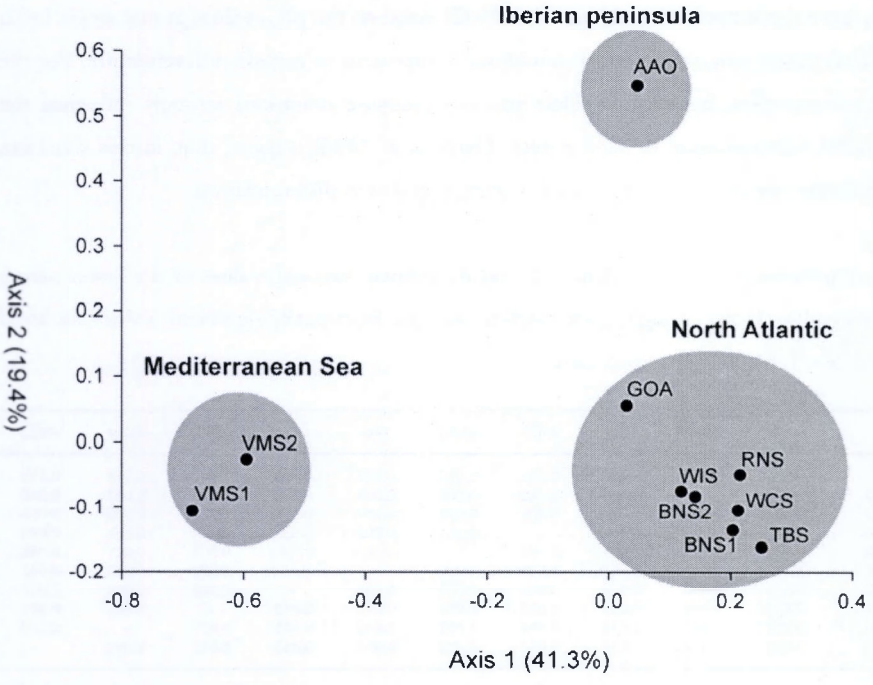


Figure 3.7 Factorial correspondence analysis plot based on microsatellite data of the 10 samples of *Pomatoschistus minutus*. See Table 3.4 for sample codes.

Second, the global θ and ϱ values across all samples, excluding the temporal samples (BNS2 and VMS2), were 0.0121 and 0.0426 respectively (both highly significant, $p<0.001$). The global standardised multilocus F_{ST} was 0.0484. The pairwise F_{ST} value was significant after

Bonferroni correction for all population pairs including TBS, VMS and AAO (the latter site has no significant difference with WIS) (Table 3.6). Pairwise R_{ST} values were significant after Bonferroni correction for a limited number of population pairs between VMS and Atlantic populations (Table 3.6). The pairwise F_{ST} and R_{ST} values between the two temporal samples were not statistically different ($p>0.05$) for BNS (0.0038 and 0.0005 respectively) and VMS (-0.0012 and -0.0025 respectively) (Table 3.6). The pairwise standardised F_{ST} values between the different samples are listed in Table 3.7. Jackknife analysis revealed that locus Pmin16-2 was responsible for the largest divergence as calculated with R_{ST}, but not with F_{ST} (average pairwise R_{ST} = 0.02974 with locus Pmin16-2 and R_{ST} = 0.00528 without this locus). Random permutation of different allele sizes among allelic states at each locus revealed that estimates of R_{ST} were significantly larger than the 95% CI range of the pR_{ST} values at one single locus Pmin16-2 (Table 3.8), suggesting a mutational component to genetic differentiation. For the other microsatellite, neither loci allele size nor stepwise mutations strongly influence the population differentiation of sand gobies. Hardy *et al.* (2003) suggest that, in this situation, F_{ST} should be preferred over R_{ST} for estimating population differentiation.

Table 3.6 Pairwise F_{ST} (below diagonal) and R_{ST} (above diagonal) values of the *Pomatoschistus minutus* samples based on eight microsatellite markers. Statistically significant values are listed in bold. See Table 3.4 for sample codes.

	TBS	WCS	RNS	BNS1	BNS2	WIS	GOA	AAO	VMS1	VMS2
TBS	-	0.017	0.012	0.003	0.023	0.017	0.044	0.018	**0.118**	**0.118**
WCS	**0.007**	-	-0.002	-0.002	0.006	0.004	0.013	0.001	**0.093**	**0.085**
RNS	**0.009**	0.002	-	0.004	0.014	-0.001	0.008	0.005	0.069	0.066
BNS1	**0.010**	0.003	0.001	-	0.001	0.008	0.022	0.004	**0.095**	**0.090**
BNS2	**0.007**	0.004	0.001	0.004	-	-0.001	0.0061	0.002	**0.081**	**0.068**
WIS	**0.013**	**0.012**	0.007	0.004	0.006	-	-0.004	0.001	0.055	0.047
GOA	**0.013**	0.008	**0.010**	0.004	0.005	0.009	-	0.002	0.042	0.027
AAO	**0.012**	**0.010**	**0.009**	**0.012**	**0.008**	0.010	**0.013**	-	**0.086**	**0.071**
VMS1	**0.025**	**0.017**	**0.019**	**0.018**	**0.017**	**0.015**	**0.015**	**0.017**	-	-0.003
VMS2	**0.018**	**0.011**	**0.013**	**0.013**	**0.012**	**0.014**	**0.013**	**0.012**	-0.001	-

Table 3.7 Standardised pairwise F_{ST} estimates (Hedrick 2005) based on the microsatellite data (below diagonal) and the F_{ST} estimated from mitochondrial cyt *b* sequence data (above diagonal) between the eight locations for *P. minutus*. Mitochondrial divergence estimates were calculated as described in *subchapter 3a.*

	TBS	WCS	RNS	BNS1	BNS2	WIS	GOA	AAO	VMS1	VMS2
TBS	-	0.165	0.196	0.243	-	0.384	0.257	0.659	0.688	-
WCS	0.027	-	0.093	0.133	-	0.036	0.015	0.616	0.650	-
RNS	0.024	0.001	-	-0.049	-	0.245	0.086	0.556	0.645	-
BNS1	0.030	0.010	0.002	-	-	0.330	0.115	0.594	0.627	-
BNS2	0.024	0.020	0.002	0.019	-	-	-	-	-	-
WIS	0.032	0.047	0.018	0.019	0.021	-	0.064	0.709	0.724	-
GOA	0.050	0.042	0.031	0.014	0.030	0.034	-	0.579	0.661	-
AAO	0.045	0.043	0.042	0.054	0.037	0.038	0.055	-	0.756	-
VMS1	0.053	0.041	0.040	0.053	0.049	0.053	0.042	0.044	-	-
VMS2	0.036	0.025	0.029	0.043	0.037	0.055	0.036	0.029	-0.006	-

Table 3.8 Mean single locus and multilocus pairwise estimates of F_{ST}, R_{ST}, pR_{ST} (95% distribution of central values in parentheses) between the ten samples of sand gobies following 2000 permutations of the microsatellite alleles (Hardy *et al.* 2003).

	F_{ST}	R_{ST}	pR_{ST} (95% range)
Multilocus	0.012	0.042*	0.011 (0.002 - 0.024)
Pmin09	0.008	0.014	0.008 (-0.004 - 0.029)
Pmin16-2	0.028	0.225*	0.027 (-0.002 - 0.103)
Pmin35	0.016	0.001	0.016 (-0.001 - 0.045)
Pmin38	-0.002	0.003	-0.002 (-0.004 - 0.003)
Pmin29	0.009	0.021	0.010 (-0.003 - 0.031)
Pmin31	0.005	-0.004	0.005 (-0.004 - 0.021)
Pmin03	0.006	0.006	0.006 (-0.004 - 0.026)
Pmin04	0.007	0.008	0.007 (-0.003 - 0.022)

* Statistically significant (p = 0.01)

Third, the CMDS plots based on both types of genetic distances separated clearly the group of the two Mediterranean samples from the populations of the Atlantic (Fig. 3.8). Both plots had a stress value below 0.20 suggesting interpretable information concerning intersite relationships.

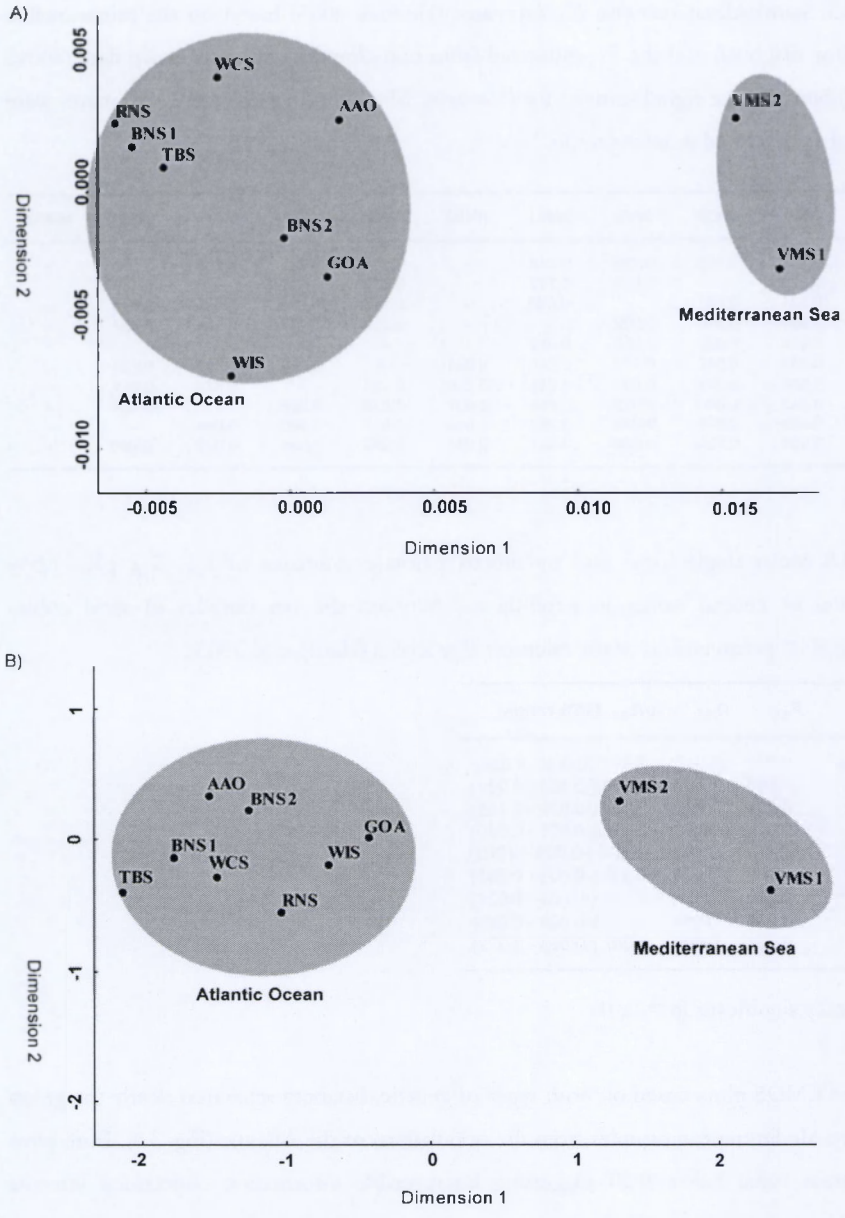


Figure 3.8 Classical multidimensional scaling plots of pairwise genetic distance for the microsatellite data calculated according to (a) Cavalli-Sforza & Edwards (1967) (D_{CE}) and (b) Goldstein *et al.* (1995) ($d\mu^2$) among the 10 samples of *Pomatoschistus minutus*. See Table 3.4 for sample codes.

Fourth, the global Mantel test revealed a significant isolation by distance (IBD) pattern with D_{CE} ($r = 0.713$, $p<0.05$) but not with $d\mu^2$ ($r = 0.166$, $p>0.05$). However, no significant IBD was found for both types of genetic distances when the Mediterranean Sea or/and Iberian Peninsula samples were excluded from the dataset.

Finally, testing the significance of the stepwise clustering procedure performed in STRUCTURE resulted in a separation of the samples into three hypothetical clusters (highest ΔK for $K = 3$). The lowest proportion of assignment to a particular cluster was 0.411 to cluster 1 for sample WCS. The highest proportion was 0.897 to cluster 2 for sample TBS (Table 3.9). All Atlantic samples (excluding TBS) had the highest assignment value for cluster 1; the two Mediterranean samples for cluster 2 and the northern Baltic sample (TBS) for cluster 3 (Table 3.9; Fig. 3.9).

Table 3.9 Summary of the assignment analysis following STRUCTURE v. 2.2 based on microsatellite data of ten *Pomatoschistus minutus* samples. The proportion of individuals assigned to the three hypothetical clusters is given. The highest assignment value for each population is listed in bold. See Table 3.4 for sample codes.

	Inferred Clusters		
Population	1	2	3
TBS	0.059	**0.897**	0.044
WCS	**0.411**	0.388	0.201
RNS	**0.514**	0.295	0.191
BNS1	**0.510**	0.336	0.154
BNS2	**0.442**	0.324	0.234
WIS	**0.516**	0.236	0.248
GOA	**0.448**	0.226	0.326
AAO	**0.492**	0.154	0.353
VMS1	0.125	0.057	**0.819**
VMS2	0.160	0.079	**0.762**

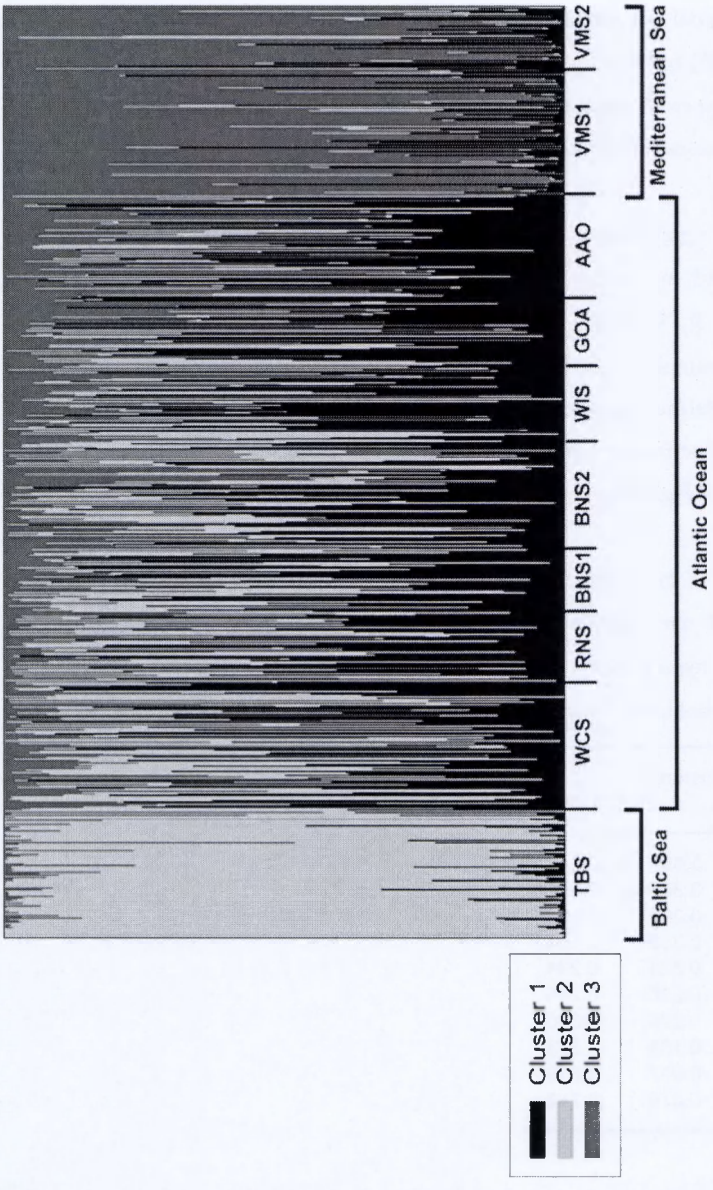


Figure 3.9 Results of the clustering analysis conducted in STRUCTURE 2.2 (Pritchard *et al.* 2000) based on the microsatellite data. In the bar plot, each of the 696 individuals is represented by a vertical bar indicating its estimated proportion of membership to the three clusters ($K = 3$). See Table 3.4 for sample codes.

Comparison between types of genetic distances and mtDNA

F_{ST}-values are much lower for microsatellites (global F_{ST} = 0.0121 [confidence interval CI = 0.0066 - 0.0174] and global standardised F_{ST} = 0.0484 [CI = 0.0264 - 0.0696], both without the temporal samples VMS2 and BNS2) than for mtDNA (global F_{ST} = 0.4293 [CI = 0.3231 – 0.5302]). Pairwise F_{ST} values between all samples based on the mtDNA data are presented together with the standardised pairwise F_{ST} estimates for microsatellites in Table 3.7. Each pairwise F_{ST} value for mtDNA is on average ten times higher than the value for microsatellites. The pairwise genetic distances of the microsatellite data ($d\mu^2$ and D_{CE}) correlated well with each other (Procrustes R = 0.8339, p-value = 0.004; Mantel R = 0.8574, p-value<0.001). The D_{CE} pairwise genetic distances of the microsatellite data correlated well with Tamura & Nei (1993) genetic distances based on the mtDNA data (Procrustes R = 0.746, p-value<0.001; Mantel R = 0.6804, p-value = 0.032). In contrast, the $d\mu^2$ pairwise genetic distances of the microsatellite data did not correlate significantly with the distances based on the mtDNA data (Procrustes R = 0.660, p-value = 0.086; Mantel R = 0.645, p-value = 0.081).

Impact of selection, sex-biased dispersal and effective population size on the degree of population divergence estimates

No indication for selection for some of the microsatellite markers was observed in the analysis. The results of Lositan and BayeScan were congruent with this pattern by revealing no outlier loci. Moreover, a blast search of the flanking regions of the microsatellite loci in de GenBank database revealed that they are not linked to a functional region already sequenced for any organism available in GenBank.

A total of 352 males and 344 females were genotyped and the population genetic parameters were estimated separately for each sex. No parameter was significantly different between the sexes (Table 3.10) suggesting no sex-biased dispersal.

Table 3.10 F-statistics, relatedness (r), mean assignment (mAI_C) and variance assignment (vAI_C) for each sex based on the microsatellite data. Significance (p) was assessed using the randomization method of Goudet *et al.* (2002).

Sex	F_{IS}	F_{ST}	r	mAI_C	vAI_C
Males	0.040	0.011	0.022	-0.078	8.643
Females	0.042	0.009	0.017	0.080	9.819
p	0.890	0.380	0.378	0.478	0.255

Estimates from the various point and temporal methods to estimate the N_e (Table 3.11) are not congruent with each other. Both point methods cannot exclude the possibility that the population sizes are infinite. The two temporal methods are similar for the BNS population (ca. 400 individuals) and for VMS (>1000 individuals).

Table 3.11 Estimates and the 95% confidence intervals of the effective population sizes (N_e) of the populations of *Pomatoschistus minutus* using point methods (heterozygote excess and linkage disequilibrium) and temporal methods (moment based approach and MLNE) on the microsatellite data. The temporal methods were only estimated for the BNS and VMS locations. See Table 3.4 for sample codes. Na, not available.

	Heterozygote excess		Linkage disequilibrium		Moments based approach		MLNE	
Populations	**Estimate**	**95% CI**	**Estimate**	**95% CI**	**Estimate**	**95% CI**	**Estimate**	**95% CI**
TBS	infinity	Na	724.3	324.5 - infinity				
WCS	9927	Na	infinity	78350.4 - infinity				
RNS	145.7	Na	infinity	421.2 - infinity				
BNS1	15.9	Na	infinity	-729.2 - infinity				
BNS2	infinity	Na	1159.9	245.1 - infinity	470	192.5 - infinity	361.9	187.4 - 1504
WIS	infinity	Na	196.8	115.9 - 557.7				
GOA	infinity	Na	2146.3	247.3 - infinity				
AAO	infinity	Na	547.1	297.8 - 2505.8				
VMS1	13.5	Na	infinity	1014.3 - infinity				
VMS2	infinity	Na	1413.3	291.5 - infinity	infinity	113.3 - infinity	1651.3	161.3 - infinity

Discussion

Phylogeography of the sand goby

The current study with microsatellite markers revealed a subtle but significant genetic structure within *Pomatoschistus minutus*. The highest divergence in the analysis was found between populations of the Atlantic, comprising all sand goby populations spanning from the Iberian Peninsula to the Baltic Sea, and the Mediterranean Sea, represented by individuals from the Vaccarès lagoon (Gulf of Lion) (Table 3.6; Fig. 3.7, 3.8, 3.9). Microsatellite marker Pmin16-2 shows a clear differentiation with a small overlap in allele size between the MS-Clade and AO-Clade, clearly indicating the historical divergence of both regions. In addition, the highest divergence noticed in the genetic analysis of allozyme markers were also found between Mediterranean and Atlantic sand gobies (Stefanni *et al.* 2003; reanalysed in *subchapter* 3a). The results are also congruent with the phylogenetic analysis based on mtDNA cyt *b* that clearly demonstrated two monophyletic clades within *P. minutus*, the Mediterranean Sea Clade (MS-Clade) and the Atlantic Ocean Clade (AO-Clade) (*subchapter 3a*). Divergence between the two sand goby clades is most likely the result of geographic isolation caused by one of the Middle Pleistocene sea-level drops (*subchapter 3a*). The Mediterranean populations showed the highest variability in the microsatellite markers (Table 3.5) despite the small distribution range along the north-western Mediterranean Sea coasts and lagoons (Miller 1986; Fig. 3.6). *Subchapter 3a* noticed the highest diversity in mtDNA in the Mediterranean samples as well. Furthermore, the highest species diversity of the 'sand goby' group is observed within the Mediterranean Sea (Huyse *et al.* 2004) supporting a Mediterranean origin of *P. minutus*, as suggested by Gysels *et al.* (2004b). Huyse *et al.* (2004) estimated the origin of the species between 1.94-1.18 mya (Early Pleistocene). The Pleistocene glaciations were the most significant historical events during the evolutionary lifespan of most Holarctic species and are believed to have accelerated the speciation process in the present day sister taxa (Avise 2000).

Limited genetic differentiation was observed between the Atlantic samples with microsatellites, but the FCA and the pairwise F_{ST} values suggest that the Iberian Peninsula' sample is different from all other Atlantic populations (Figure 3.7; Table 3.6). These results

are congruent with the two mitochondrial groups within the Atlantic samples: the Iberian (IB-Group) and the North Atlantic Group (NA-Group). This divergence is likely the result of population decline within different refugia during glaciations in the Middle Pleistocene (*subchapter 3a*). Based on Bayesian assignment test and pairwise F_{ST} values, the position for the sample of the northern Baltic Sea (TBS) was remarkable (Table 3.6, 3.9; Figure 3.9). This is most likely the result of a founder-event, based on the low variation on the microsatellite loci in this population (Table 3.5) and as suggested in *subchapter 3a* based on the mtDNA data. The difference in the cyt *b* haplotype network between southern and northern Baltic Sea samples suggested that the Baltic Sea has been colonized in two phases over a period of 8000 year, with a stronger founder effect in the north. The presence of only two common cyt *b* haplotypes in the northern samples HBS and TBS (haplotypes NA01 and NA28) with their many uniquely derived haplotypes suggested that only a few individuals were able to adapt to the severe abiotic conditions of northern Baltic (Johannesson & André 2006). They founded the population still inhabiting the area today.

The Mantel test and Procrustes analysis illustrated that the microsatellite results are congruent with the mtDNA phylogeographic pattern. However, those tests were not significant when the $d\mu^2$ genetic distance was used for microsatellites. This suggests that, although both measures are correlated, F_{ST} correlates better with mtDNA than R_{ST}. In theory F_{ST} is more sensitive than R_{ST} for recent intraspecific divergence (Gaggiotti *et al.* 1999; Balloux & Lugon-Moulin 2002). Moreover, R_{ST} can be less accurate in reflecting population differentiation due to its higher associated variances (Balloux & Lugon-Moulin 2002). Therefore, the number of loci screened has to increase before a consistent pattern is reached (Gaggiotti *et al.* 1999; Balloux & Goudet 2002). This has already been empirically observed with European grayling *Thymallus thymallus* (Koskinen *et al.* 2004).

Overall, the results confirm the first hypothesis of this study that microsatellite markers reveal a phylogeographic pattern congruent with the patterns based on mtDNA and allozyme data.

Differences in the degree of divergence estimates among nuclear and mtDNA markers

The difference in the degree of population differentiation between the nuclear and mitochondrial markers is remarkable. The mitochondrial differentiation values are more than an order of magnitude higher than the nuclear differentiation, even after standardization of the microsatellite results. The standardised measure allows comparison between loci with different levels of genetic variation (Hedrick 2005). Therefore, several other factors can be the cause of the observed discrepancy.

First, selection may have differential effects on genetic markers. It has been suggested that balancing selection may significantly influence the distribution of allozyme diversity (De Innocentiis *et al.* 2001). Loci experiencing balancing selection will have allele frequencies more similar than expected under neutrality, reducing the F_{ST} estimates. Allendorf & Seeb (2000) concluded that estimates of population structure produced by allozymes were generally comparable to those obtained with other nuclear markers, including microsatellites. They noted that when differences between marker classes did occur, they were usually due to one or a few exceptional loci and not all of them. Each used microsatellite marker was tested for positive or balanced selection in comparison with the other markers in the Lositan and BayeScan analysis, but no marker seemed to be an outlier. Similar to microsatellites, it is not safe to assume *a priori* that mtDNA evolves as a strictly neutral marker (Ballard & Whitlock 2004). Selection on mtDNA may accelerate the coalescence of lineages, and thus increase levels of differentiation observed between populations (Peijnenburg *et al.* 2006). The observation of very low nuclear differentiation in sand gobies while mtDNA data revealed no gene flow between the three isolated Middle Pleistocene lineages, cannot be explained by selective evolution for mtDNA. The rejection of the null hypothesis in different neutrality tests for the mtDNA data of *P. minutus* was assigned to demographical expansions instead of selection (*subchapter 3a*). Various demographic analyses on the mtDNA data showed that the intra-assemblage genetic structure of *P. minutus* contains signatures of demographic expansion events.

Second, studies documenting weaker population subdivision for nuclear than maternally inherited genetic markers often attribute these discrepancies to male-biased dispersal. Sex-

biased dispersal is common in nature (reviewed in Cano *et al.* 2008), however, it has only been described for a limited number of marine fishes (listed in Consuegra & de Leaniz 2007). No indications for sex-biased dispersal were found for *P. minutus* with microsatellite markers (Table 3.10). Contemporary gene flow between the Mediterranean and Atlantic basins is also unlikely due to the discontinuous distribution of the sand goby (Miller 1986; Fig. 3.6). Moreover, only males have to migrate successfully to explain the pattern. However, females are expected to be the most mobile sex, especially during the spawning period when males are guarding their nest (Lindström *et al.* 2006).

Third, various technical problems, including homoplasy, may have reduced the signal of differentiation detected by the microsatellite markers. Homoplasy occurs when different copies of a locus are identical in state, although not identical by descent. The situations where size homoplasy is most prevalent involve high mutation rates and large population sizes together with strong allele size constraints (Estoup *et al.* 2002). Therefore, effects of homoplasy are expected to be common for microsatellites in marine fishes (O'Reilly *et al.* 2004), which has implications for the identification of genetic structuring (Carreras-Carbonell *et al.* 2006). Microsatellites probably suffered from higher levels of homoplasy than mtDNA because of higher mutation rates and larger effective population sizes (Balloux *et al.* 2000). The various microsatellite markers of *P. minutus* indicate homoplasy because of the high allele numbers and the limited size range of all markers (in average 1 allele per 2,053 bp). However, simulation studies suggest that size homoplasy will have much less effect on estimates of population differentiation than gene migration or genetic drift (Estoup *et al.* 2002). Therefore, it is unlikely that homoplasy is the main cause of the observed differences.

Finally, mitochondrial markers can be more sensitive in detecting differentiation because of a lower effective population size than nuclear markers (Shaw *et al.* 2004). Genetic drift effects are linked to effective population size (N_e) and therefore it is possible that ecologically relevant population structure remains undetectable by using neutral markers when the N_e is high (Bentzen 1998). Marine fish has the potential to have high N_e and therefore recently separated large populations may appear genetically homogeneous even in the complete absence of contemporary gene flow (Hauser & Carvalho 2008). Simulations in Buonaccorsi *et al.* (2001) showed that differences in the magnitude of estimated population subdivision

from nuclear and mitochondrial markers could be accounted for entirely by differences in effective population sizes and polymorphism on F_{ST} estimates. Due to the haploid and maternal-only inheritance of mtDNA it has an effective population size of one-quarter that of nuclear DNA, making mtDNA more susceptible to effects of genetic drift (Shaw *et al.* 2004). This explanation has been invoked for the discordant patterns in population differentiation between nuclear and mitochondrial markers for marine organisms, such as the blue marlin *Makaira nigricans* (Buonaccorsi *et al.* 2001), Patagonian toothfish *Dissostichus eleginoides* (Appleyard *et al.* 2002; Shaw *et al.* 2004) and the olive sea snake *Aipysurus laevis* (Lukoschek *et al.* 2008). Simulations in EASYPOP v. 1.7 (Balloux 2001) with specific biological information about the sand goby showed no difference with the simulations of Buonaccorsi *et al.* (2001) (results not shown). Moreover, *P. minutus* is known to be one of the most abundant fish species along almost its full range (Pasquaud *et al.* 2004; Ehrenberg *et al.* 2005; Maes *et al.* 2005). Therefore, high N_e most likely explains the discordant patterns between nuclear and mitochondrial data for the sand goby. Point methods to estimate the N_e of the present sampled populations can not invalidate the null hypothesis of an infinite population size for *P. minutus* (Table 3.11). Nevertheless, point methods are not always reliable and biased (Wang & Whitlock 2003), especially when sample size is small (<100 individuals) and below the true N_e (England *et al.* 2006). Both temporal N_e estimates show a limited population size, especially for the BNS location (100-1000 individuals) (Table 3.11). Still, there are crucial differences between the results of the two temporal methods and the assumption of a closed system without migration could not be fulfilled. Migration inside the marine system can therefore cause a strong underestimation of the N_e (Wang & Whitlock 2003). On the other hand, the high number of alleles for microsatellites confirms the hypothesis of a high effective population size of more than thousand individuals to maintain the high genetic variation (Ewens 1972; Poulsen *et al.* 2006). The Pmin20 locus was excluded from the analysis because after genotyping more than 90 individuals, the number of alleles approaching the number of fish genotyped. The studies of Jones *et al.* (2001) and Pampoulie *et al.* (2004a) also observed microsatellite markers with a very high number of alleles.

Our present results do not confirm the second hypothesis, which put that microsatellite markers are more sensitive for population differentiation on a macro-scale than mitochondrial and allozyme markers. The most likely reason for the discordant pattern

between nuclear and mitochondrial loci is that recent demographical expansion reached such high effective population sizes in *P. minutus*.

Conclusions

Marine organisms have a high potential for gene flow and population size. However, it is not known whether the observed low genetic differentiation for so many marine species reflects high effective population sizes and low gene flow, high effective population sizes and high rates of gene flow, or low effective population sizes and high rates of gene flow (Hauser & Carvalho 2008). In this study on *P. minutus* the scenario of high effective population sizes and low gene flow best explains the observed genetic pattern. However, this conclusion could only be reached with more than one genetic marker. Therefore, one has to be cautious when interpreting present-day genetic structure in terms of gene flow while using one single type of marker and/or statistical method.

By organizing a better sampling strategy and using straightforward N_e-estimation methods incorporating migration (Wang & Whitlock 2003), better estimates of the N_e might confirm the expected high N_e for *P. minutus*. Also SNPs and microsatellite markers with less alleles covering a limited size range are useful to understand the power of homoplasy in the analysis. It will allow verifying the significance of each explanation for the discordant pattern between nuclear and mitochondrial markers.

Acknowledgements

The authors thanks the following persons who kindly provided the samples used in this study: Pilar Drake (CSIC, Spain), Maria Järvi-Laturi (University of Helsinki, Finland), Kai Lindström (Åbo Akademi University, Finland), Lotta Kvarnemo (Göteborgs Universitet, Sweden), Ian McCarthy (University of Wales, Bangor, UK), Mario Lepage (Ifremer, France), Jef Guelinckx and Alain Crivelli (Station Biologique de la Tour du Valat, France). We also thank Luisa Orsini, Gregory Maes and Dirk Schaerlaekens for interesting discussions and Dorien Daneels for the simulation analyses. The first author received a PhD-fellowship of the Institute for the Promotion of Innovation through Science and Technology in Flanders

(IWT-Vlaanderen). This study is part of the WESTBANKS project financed by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (contract nr SD/BN/01A).

Chapter 4

Local adaptation on the rhodopsin gene of the sand goby

'To suppose that the eye, with all its inimitable contrivances for adjusting the focus to different distances, for admitting different amounts of light, and for the correction of spherical and chromatic aberration, could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree…'

Charles Darwin (English naturalist, 1809-1882), 'On the Origin of Species' (1859)

SUBCHAPTER 4A

TO SEE IN DIFFERENT SEAS: SPATIAL VARIATION IN THE RHODOPSIN GENE OF THE SAND GOBY (*POMATOSCHISTUS MINUTUS*)

Maarten H.D. Larmuseau[1], Joost A.M. Raeymaekers[1], Kevin G. Ruddick[2], Jeroen K.J. Van Houdt[1,3] & Filip A.M. Volckaert[1]

[1]Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Leuven, Belgium.

[2]Management Unit of the North Sea Mathematical Models (MUMM), Royal Belgian Institute for Natural Sciences (RBINS), Brussels, Belgium

[3]Royal Museum for Central Africa, Tervuren, Belgium.

Accepted manuscript: *Molecular Ecology*, in press.

Abstract

Aquatic organisms living in a range of photic environments require specific mechanisms to tune their visual pigments. Maximum absorbance (λ_{max}) of retinal rods in populations of the marine demersal sand goby (*Pomatoschistus minutus;* Gobiidae, Teleostei) correlates with the local optic environment. It has been shown that this is not regulated through a physiological response by exchanging the rhodopsin chromophore. To test for evolutionary adaptation, the sequence of the rhodopsin (*RH1*) gene was analysed in 165 *Pomatoschistus minutus* individuals from seven populations across its distribution range. Analysis showed a high level of intraspecific polymorphism at the *RH1* gene, including non-synonymous mutations on amino acids, known as spectral tuning sites. Population differentiation at these sites was in agreement with the observed differentiation in λ_{max} values. Analyses of d_N/d_S substitution rate ratios and likelihood ratio tests under site-specific models detected a significant signal of positive Darwinian selection on the *RH1* gene. A strong discrepancy in differentiation was noticed between *RH1* gene variation and the presumably neutral microsatellites and

mitochondrial data. Samples did not cluster according to geographical or historical proximity with regards to *RH1* but according to the general photic conditions of the habitat environment of the sand goby. The present study highlights the usefulness of sensory genes, like rhodopsin, for studying the characteristics of local adaptation in marine non-model organisms.

Keywords: Adaptive evolution, candidate gene, Gobiidae, marine fish, photoreceptor, remote sensing, vision

Introduction

For too long, the genetic structure of marine organisms has been thought to be homogeneous due to the lack of obvious barriers to gene flow in the environment. Since gene flow is expected to hamper adaptive population divergence, the traditional idea was that local adaptation may be rare or absent in marine fishes (Hemmer-Hansen *et al.* 2007). Lately, however, an increasing number of population genetic studies have described complex genetic structures in several marine species (Knutsen *et al.* 2003; Pampoulie *et al.* 2008). One major factor responsible for the present genetic structure of marine species is the geological and climatological history during the Pleistocene glaciations (Debes *et al.* 2008; Luttikhuizen *et al.* 2008). Also contemporary factors maintain and promote genetic differentiation among marine populations on various geographical scales. The marine environment shows heterogeneity in response to climate, hydrodynamics and topography (Cowen *et al.* 2000), and biological traits, such as sex-dependent migration, site philopatry and assortative mating enhance genetic structuring (Ruzzante *et al.* 1998). Stable neutral genetic structuring among populations may indicate that local selection is overriding the effects of drift and gene flow, resulting in adaptive divergence. Local adaptation in marine organisms has become increasingly documented, indicating that selection is also a potent evolutionary force in the marine environment (Canino *et al.* 2005; Hemmer-Hansen *et al.* 2007; Zane 2007; Sherman & Ayre 2008). Nevertheless, knowledge of the spatial and temporal scale of adaptive genetic variation in marine systems remains scant, yet crucial to improve our understanding of how evolution operates in the ocean (Conover *et al.* 2006).

The sand goby *Pomatoschistus minutus* (Pallas 1770; Gobiidae, Teleostei) is a common small marine demersal fish inhabiting the shallow waters along European coasts (Miller 1986) (Fig. 4.1). A recent study has shown relatively low levels of gene flow and high genetic structuring in this species compared to other marine fish species (*subchapter 3a*). Therefore it is a suitable model for studying the characteristics of local adaptation in the marine environment. Middle Pleistocene glaciations yielded three isolated and differently evolving sets of sand goby populations. Reciprocal mitogenic monophyly was observed between a Mediterranean Sea (MS) and an Atlantic Ocean (AO) clade (*subchapter 3a*). The AO-Clade contains two major phylogeographic groups: the Iberian Peninsula (IB) group and the North Atlantic (NA) group. For the NA-Group there is evidence for geographic sorting of the ancestral mitochondrial DNA (mtDNA) haplotypes with recent radiations in the Baltic Sea, Irish Sea, North Sea and Bay of Biscay. Northern Baltic Sea sand gobies are considered to belong to an isolated population with clear evidence for founder effects (*subchapter 3a*). Allozyme and microsatellite analyses largely corroborated this phylogeographic pattern (Stefanni *et al.* 2003) (*subchapter 3b*).

Sand gobies are visual feeders (Healey 1971; Aarnio & Bonsdorff 1993) and mostly nocturnal. Patterns of activity are largely influenced by tides and light intensity (Ehrenberg & Ejdung 2008). For *P. minutus,* nocturnal foraging is advantageous in approaching prey and in avoiding predators (Thetmeyer 1997). Relative to body size, the protruding eyes are large and might be capable of detecting prey organisms in very dim light (Thetmeyer 1997). The geographical distribution of the sand goby includes a wide range of photic environments, varying in turbidity, colour and brightness. Therefore, adaptation to the local spectral environment may be crucial. Vertebrates have visual pigment (VP) molecules bound in dense membrane stacks in retinal photoreceptors to mediate vision. The VP protein moiety is opsin, which is a G protein linked receptor, bound to a light-sensitive chromophore, 11-*cis* retinal (A1) or 11-*cis* 3, 4-dehydroretinal (A2) (Park *et al.* 2008). Each pigment shows a characteristic peak of maximal absorbance (λ_{max}), its precise location depending on the interactions between the chromophore and the opsin protein. The pigment that mediates vision in dim light and absorbs light with λ_{max} of about 500 nm, is rhodopsin; It is located in rod cells.

Vertebrates have various possibilities to modify their visual system to cope with the photic environment. The spectral tuning of the VP proteins can be assessed on a physiologically time scale through exchange of the chromophore (A1 or A2), consistent with an anticipated change in photic environment (Bowmaker 1995). Tuning can also be achieved at the DNA level on an evolutionary time scale through amino acid (AA) substitutions in the protein part (the opsin) (Yokoyama 2000). The first possibility seems unlikely in sand gobies. Jokela *et al.* (2003) measured the absorbance spectra microspectrophotometrically in retinal rods of various sand goby populations. They found considerable variation in λ_{max} values within and between populations. The shapes of the absorbance spectra indicated polymorphism at the rhodopsin gene rather than admixture of A1 and A2 chromophores, suggesting that the variation in λ_{max} values is genetic. Therefore, evolutionary adaptation, rather than physiological change, is presumed to be responsible for spectral tuning.

The tuning mechanism of visual pigments should be a suitable candidate to understand the opportunities and characteristics of local adaptation in the marine environment. The aim of this study was to assess if sand gobies are evolutionary adapted to local photic environments on the rhodopsin gene (*RH1*). Our strategy to demonstrate local adaptation on *RH1* consists of three steps: first we demonstrate differentiation in the functional variation of the *RH1* gene between sand goby populations. Next we demonstrate that the population differentiation of *RH1* is due to selection. Finally we establish a link between the functional variation of *RH1* and selection regimes.

Material & Methods

Sampling and species identification

A total of 165 *Pomatoschistus minutus* individuals were caught at seven locations along the European coast between January 2006 and February 2007 (Table 4.1, Fig. 4.1). Samples were taken either by fyke, hand net or beam trawling. The sand gobies were distinguished from other cryptic *Pomatoschistus* species morphologically, based on the dermal head papillae (Miller 1986) and pigmentation pattern (Hamerlynck 1990), and genetically, based on a PCR-RFLP species identification protocol described in *subchapter 2a.*

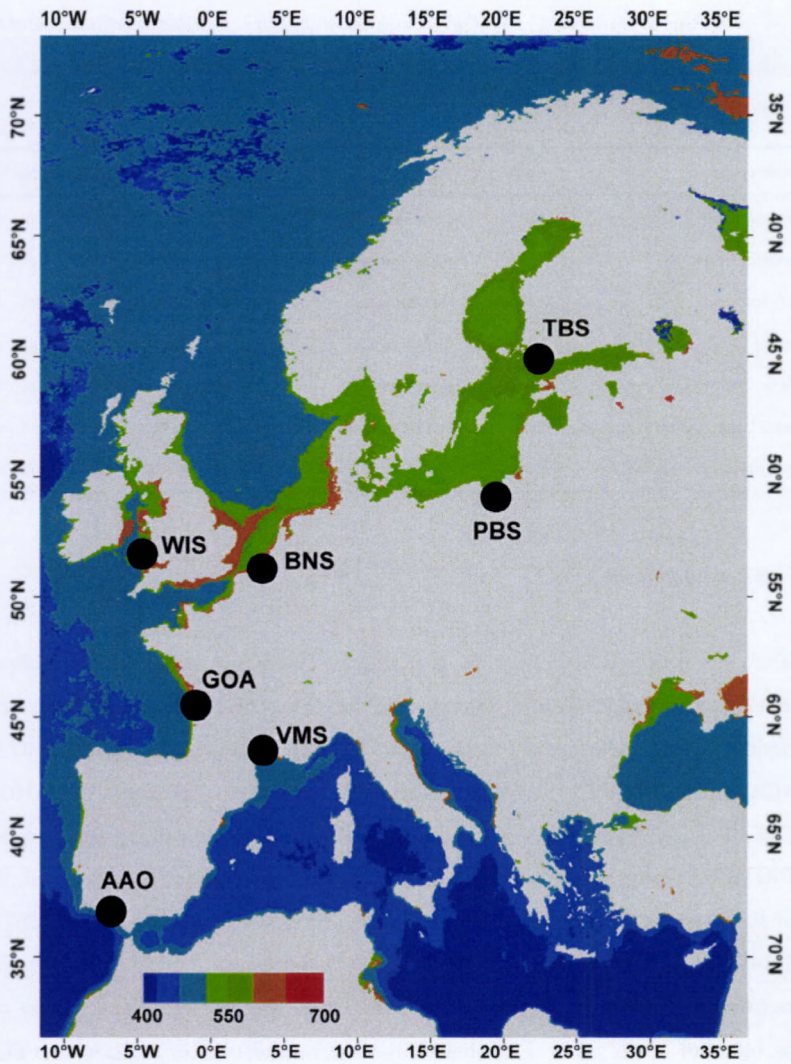


Figure 4.1. Geographical distribution of the seven sampling locations in seven European marine systems for sand goby *Pomatoschistus minutus*. The wavelength (nm) of maximally transmitted light estimated from the MODIS annual composite radiance data of 2007 is shown for all European seas. Discrete colours correspond to the MODIS wavelengths 412 nm (deep blue), 443 nm (blue), 488 nm (pale blue), 531 nm (bright green), 551 nm (dark green), 667 nm (red) and the default value 615 nm (orange) substituted in case of saturated data at 531 nm, 555 nm or 667 nm.

Table 4.1. Collection information for the seven populations of *Pomatoschistus minutus*, including code, marine system, country, site, date, longitude and latitude, and sample sizes for the *RH1* and microsatellite analyses (Nr and Ns respectively).

Code	Marine system	Country	Site	Date	Latitude	Longitude	Nr	Ns
TBS	(Northern) Baltic Sea	Finland	Tvärminne	Jul/06	59°50'N	23°12'E	20	96
PBS	(Southern) Baltic Sea	Polen	Sopot, Bay of Gdańsk	Feb/07	54°27'N	18°35'E	10	-
BNS	North Sea	Belgium	Oostduinkerke	Nov/06	51°08'N	02°40'E	27	47
WIS	Irish Sea	UK (Wales)	Llanfairfechan	Nov/06	53°20'N	03°59'W	21	45
GOA	Atlantic Ocean (Bay of Biscay)	France	Gironde estuary	Aug/06	45°36'N	01°01'W	22	40
AAO	Atlantic Ocean (Iberian Peninsula)	Spain	Guadalquivir river estuary	Nov/06	36°58'N	06°10'W	37	95
VMS	Western Mediterranean Sea	France	Vaccarès lagoon	Jan/06	43°32'N	04°35'E	28	94

Gene amplification and sequencing

Genomic DNA was extracted from fin clips, stored in 100% ethanol, using the NucleoSplin Extraction Kit (Machery-Nagel GmBH, Düren, Germany). An 868 bp fragment of the *RH1* gene was amplified in polymerase chain reactions (PCR) with the forward primer PminRh1F GCGCCTACATGTTCTTCCTT and the reverse primer Rh1039r TGCTTGTTCATGCAGATGTAGA (Chen *et al.* 2003). The forward primer was designed using the PRIMER 3 program (Rozen & Skaletsky 1998) on conserved regions of the alignment of *RH1* gene sequences from *P. minutus* (Acc. nr. X62405), *Gobius niger* (Y18675), *Zeus faber* (Y14484), *Sargocentron diadema* (U57537) and *S. microstoma* (U57542). Developing new primers to amplify a larger fragment of the *RH1* gene was not successful due to co-amplification of other opsin genes. PCR reactions were carried out on a GeneAmp PCR System 2700 thermocycler (Applied Biosystems, Foster City, CA, USA) in a total volume of 25 µl, containing 1 µl of genomic DNA, 1 X PCR buffer, 0.2 mM dNTPs, 0.8 µM of each primer, 2.0 mM $MgCl_2$, 0.5 U of Taq DNA polymerase (Silverstar, Eurogentec, Seraing, Belgium) and mQ-H_2O. The PCR profile was: 4 min at 94°C followed by 35 cycles of 30 s at 96°C, 30 s at 54°C and 1 min at 72°C; with a final 10 min extension at 72°C. To avoid contamination, different pipettes, aerosol barrier tips and different sections of the laboratory were used for pre- and post-PCR work. Every other 15th individual (corresponding with one

every two rows of a PCR-plate) a negative control was inserted to detect contamination. No contamination occurred during the screening procedures. All PCR products were visualized on agarose gels with ethidium bromide. After purification with the 'GFX PCR DNA and Gel Band Purification kit' (GE Healthcare, Piscataway, NJ, USA), the PCR products were sequenced in both directions using the BigDye Terminator v. 3.1 Cycle Sequencing Kit on an ABI 3130 automated capillary DNA sequencer (Applied Biosystems). Sequences of 756 bp (252 AA) were checked and aligned to each other with SEQSCAPE v. 2.1 (Applied Biosystems). The full rhodopsin sequence of *P. minutus* counts 1056 bp and thus 352 amino acids (AAs) (Archer *et al.* 1992). The 252 AA' fragment under study thus represent 72% of the protein. However, all known 25 AAs involved in the spectral tuning of the visual pigments are included in this gene fragment (Yokoyama *et al.* 2007 and references herein). Automated detection of point mutations was realized with the GAP4 subprogram embedded in the STADEN package (http://sourceforge.net/projects/staden) and checked manually by eye.

For several reasons we are convinced that no other member of the opsin gene family than the *RH1* gene was co-amplified and analyzed. First, when designing primers for rhodopsin, sites were selected that differ among paralogous genes. Second, other opsin genes have introns, unlike the rhodopsin genes of bony fishes (Bowmaker 1995). Third, the duplication event separating rhodopsin from other opsin genes occurred before the diversification of vertebrates (Yokoyama 2000). If we had sequenced by mistake a paralogous opsin gene, the sequence alignment would have shown this extreme divergence. Finally, different PCR-products were cloned to control for a recent duplication event of the *RH1* gene in *P. minutus*. A total of 21 individuals with more than one heterozygote single nucleotide polymorphism (SNP) locus was cloned into bacterial vectors using the TOPO-TA cloning kit (Invitrogen). Five to ten clones originating from two independent PCR reactions (for protocol see above) per specimen were sequenced. No more than two haplotypes were observed in each reaction, suggesting that only one gene was sequenced and analyzed. All 38 rhodopsin haplotypes determined in this study were deposited in the GenBank database (accession numbers: FJ410451-FJ410488; Table S4.1).

Microsatellite genotyping and analysis

Variation at eight high-quality microsatellite markers (Pmin03, Pmin04, Pmin09, Pmin16-2, Pmin29, Pmin31, Pmin35 and Pmin38) (*subchapter 2b*) was assessed for 417 sand gobies of six populations (Table 4.1). Deviation from Hardy-Weinberg equilibrium and population differentiation quantified as F_{ST} were quantified with GENETIX v. 4.05 (Belkhir *et al.* 2004).

Haplotype reconstruction and network analysis

Rhodopsin haplotypes of the 21 cloned individuals and 105 sequenced individuals with less than two heterozygous sites were available. The haplotypes of the 39 remaining individuals were inferred from the genotypes using the Bayesian statistical methods in the program PHASE v. 2.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003). Using this program, haplotypes have been resolved based on the assumption that unsolved haplotypes tend to be more similar to previously sampled known haplotypes. Runs were conducted separately for each population, with known haplotype information (i.e. homozygous haplotypes and cloned haplotypes) being included as prior information. Ten independent runs per population were conducted, each with a burn-in-period of 1000 followed by 10,000 iterations with a thinning interval of 100 steps. The results and the goodness-of-fit values were very similar among runs, indicating that the run lengths were sufficient. Haplotypes of individuals with more than one heterozygous site for which the phase could not be determined with a probability of >95% (averaged over the ten runs) were excluded from the haplotype network (19 out of 165 analyzed individuals). A haplotype network of the rhodopsin haplotypes was constructed using the statistical parsimony method implemented in the program TCS v. 1.21 (Clement *et al.* 2000). Interpopulation relationships were assessed by estimating pairwise F_{ST} values based on the haplotype distributions with ARLEQUIN v. 3.11 (Excoffier *et al.* 2005). These values were then used for a classical multidimensional scaling (CMDS) analysis in the Vegan package in R (Oksanen *et al.* 2007) for detecting group structure. CMDS plots having a stress value less than 0.20 provide interpretable information concerning intersite relationships (Clarke 1993).

Genetic diversity and neutrality tests

The number of segregating sites (S), the mean number of pairwise differences (k) and estimates of nucleotide polymorphism (π, θ) were calculated using DNAsp v. 4.10.9 (Rozas *et al.* 2003).

Several analyses were performed to determine if positive selection was involved in the evolution of *RH1* in *P. minutus*. The number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) were estimated using the Z-test implemented in MEGA v. 4.0 (Tamura *et al.* 2007) according to Nei & Gojobori (1986) with the correction of Jukes & Cantor (1969) for multiple substitutions. The variances of d_S and d_N were computed by bootstrap (10,000 replicates). With this information, the null hypothesis of neutral evolution (H_0: $d_N = d_S$) *versus* the hypothesis of positive selection (H_1: $d_N > d_S$) was tested using a Z-test: $Z = (d_N - d_S)/\mathrm{SQRT}(\mathrm{Var}(d_S) + \mathrm{Var}(d_N))$. The maximum-likelihood method (Yang *et al.* 2000a) implemented in the program CODEML of the PAML 4.1 software package (Yang 2007) was used to test whether codon sites on the *RH1* gene were affected by positive selection (Yang *et al.* 2005). The models were M7 (beta) and M8 (beta and ω) (Yang *et al.* 2000a). While recombination can potentially generate false-positives in the detection of positive selection, these models are more robust against the occurrence of recombination than the other models implemented in CODEML (Anisimova *et al.* 2003). The models M7 and M8 are compared pairwise using the likelihood-ratio test (LRT) (Nielsen & Yang 1998). To provide phylogenetic information for the analysis, the best tree for *RH1* sequences was identified with the maximum likelihood method under the one-ratio model (M0) in CODEML. Positively selected codons ($\omega>1$ with $p>95\%$) were identified through an empirical Bayesian approach implemented in CODEML (Yang *et al.* 2005).

Another method used to test for the effects of differential selection among populations is to compare the distribution of the variation on *RH1* and neutral nuclear markers. If SNPs of the *RH1* gene code for adaptive variation, the *RH1* gene is expected to reveal aberrant population structures in comparison to nuclear markers (such as microsatellite markers), which may be behaving neutral (Bamshad & Wooding 2003). To compare the degree of

population differentiation between the *RH1* and the microsatellite markers, several methods were applied. First, correlations between pairwise F_{ST}-values of the two markers were calculated and tested using simple Mantel procedures (Mantel 1967) in the Vegan package in R (Oksanen *et al.* 2007). Because the number of Mantel test permutations is limited for small sample sizes (n = 6) (Legendre 2000), complete enumeration of all possible 6! = 720 permutations was carried out for all tests. Second, two-dimensional CMDS ordinations of pairwise F_{ST}-values of different marker types were compared by a Procrustes Analysis (PA) (Gower 1975) with the Vegan package. PA searches for the best match between two configurations of points in a multivariate Euclidean space using rotation, translation, reflection, and dilation of one configuration. The criterion used to assess the best fit is the minimization of the sum of squares between the differences for each observation (m^2). The significance of the result, an optimal superposition of one configuration on the other (reference) configuration, is obtained through a permutation test (PROTEST) (Jackson 1995). PROTEST uses $R = sqrt(1-m^2)$ as a test statistic, which can be interpreted as a correlation. Finally, a selection detection workbench LOSITAN (Antao *et al.* 2008) based on the FDIST F_{ST} outlier methods of Beaumont and Nichols (1996) was used to evaluate the neutrality of the microsatellites and the presumed outlier status of *RH1*. Different runs were assessed: one run with only microsatellite data, 14 different runs with all microsatellites and a polymorphic SNP of the *RH1* gene, and a final run with all microsatellites and all polymorphic SNPs of the *RH1* gene. For all runs 30,000 simulations were generated with 'neutral mean F_{ST}' and 'force mean F_{ST}', to increase the reliability of the mean F_{ST}.

Environmental light measurements

To correlate the differences in the rhodopsin gene variation between *P. minutus* populations with the light transmittance of the respective waters, the spectral distribution of environmental ambient light was measured in the north-eastern Atlantic Ocean, Mediterranean Sea and adjacent seas. Lindström (2000) introduced the concept of 'wavelength of maximally transmitted light' (WMTL) in order to characterize the spectral content and depth variation of the underwater light climate by a single parameter for comparison with the spectral sensitivity of the eyes of marine animals. Whereas a combination of underwater light measurements and optical modelling is used by Audzijonyte

et al. (2005) to estimate the WMTL at various locations, a new method is described in Appendix 4.1 to estimate this parameter from satellite remote sensing data. The method has the advantage of providing information at almost any location on earth without the need for *in situ* measurements or *a priori* knowledge, and of relying on a more uniform methodology. Water-leaving radiance data as measured by the MODIS-AQUA satellite sensor was downloaded on 19th November 2008 from the NASA 'Ocean Color' web site (http://oceancolor.gsfc.nasa.gov/) as the annual composite for 2007 (4 km standard map image file) for each of the available spectral bands (412 nm , 443 nm, 488 nm, 531 nm, 551 nm and 667 nm).

Results

Nucleotide diversity of the RH1 *gene*

Sequences matched the general properties of the *Pomatoschistus minutus RH1* gene (X62405) (Archer *et al.* 1992). In total, 19 segregating sites or SNPs were noticed across all genotypes (Table S4.2). Five SNPs (SNP 2, 3, 5, 8 and 18) were not polymorphic according to the 99% criterion; three polymorphic SNPs were part of the same codon and were merged (written further as SNP_9_10_11). The alignment in amino acids (AA) shows five non-synonymous AA substitutions; four are located in the transmembrane helices and one in the C-II loop (Fig. 4.2). The sequencing and cloning reactions revealed 38 confidently resolved haplotypes (Table S4.1). Nucleotide diversity (π) of the *RH1* gene fragment was estimated to be 0.0074 in total. The within-population *RH1* nucleotide diversity values were highest in populations BNS, WIS and GOA (southern North Sea, Irish Sea and Bay of Biscay, respectively); the lowest value was found in PBS (southern Baltic Sea; Table 4.2).

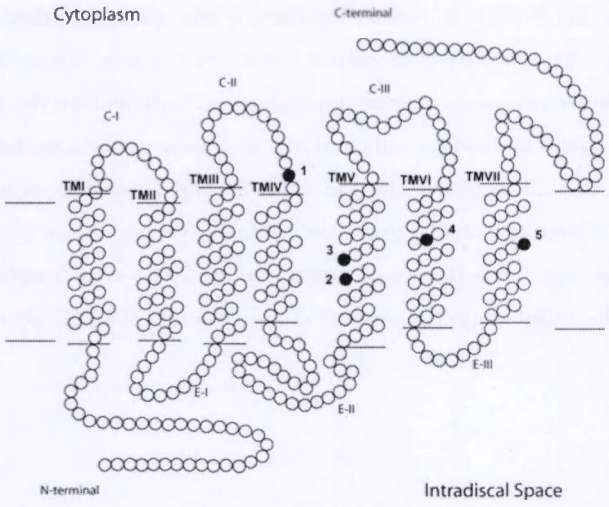


Figure 4.2. Two-dimensional model of the seven transmembrane α–helices of the bovine rhodopsin (*RH1*) as in Hargrave & McDowell's (1992). The seven transmembrane helices (TM) are numbered, as well as the three loops at the cytoplasmic side (C) and the extracellular side (E) of the cell membrane. The different non-synonymous mutations found in *Pomatoschistus minutus* are shown in filled circles. (1) AA151 (SNP4); (2) AA214 (SNP9_10_11); (3) AA217 (SNP12); (4) AA261 (SNP14); (5) AA299 (SNP19).

Table 4.2. Summary of diversity indices for the 19 polymorphic sites analyzed for variation at the population level of the sand goby *RH1* gene. N, number of individuals surveyed; Sn, number of non-synonymous segregating sites; Ss, number of synonymous segregating sites; k, mean number of pairwise differences; л, average number of nucleotide differences per site; θ, theta per site. For site codes see Table 4.1.

		No. of haplotypes						
Population	N	Total	Private	Sn	Ss	k	л	θ
TBS	20	8	2	5	3	1.763	0.0023 ± 0.0015	0.0028
PBS	10	4	1	4	2	0.895	0.0012 ± 0.0010	0.0012
BNS	27	14	8	5	8	4.197	0.0056 ± 0.0006	0.0038
WIS	21	7	1	5	7	3.987	0.0053 ± 0.0008	0.0037
GOA	22	8	4	4	6	3.755	0.0048 ± 0.0007	0.0030
AAO	37	9	5	6	8	1.287	0.0017 ± 0.0006	0.0038
VMS	28	12	6	4	4	1.449	0.0019 ± 0.0003	0.0023
Total	165	38	27	6	13	5.608	0.0074 ± 0.0001	0.0039

Population differentiation on the RH1 *gene*

A maximum parsimony network of confidently resolved haplotypes revealed that three haplotypes are common in at least four populations (Fig. 4.3). Haplotype RhP7 is shared among all populations. Haplotypes clustered roughly into two groups, one containing the majority of haplotypes observed in the Iberian Peninsula, North Sea and Irish Sea, and another group containing haplotypes occurring mainly in the Mediterranean and Baltic Sea (Fig. 4.3). Accordingly, pairwise F_{ST} values (Table 4.3) showed a clustering of Mediterranean and northern Baltic Sea (VMS and TBS) samples; the Atlantic samples (GOA, WIS and AAO) clustered together. Finally, the frequency of non-synonymous SNPs and polymorphic AA differed between the populations (Table 4.4).

Table 4.3. Pairwise F_{ST} estimates based on the *RH1* gene (below diagonal) and based on the microsatellite loci (above diagonal) between sand goby populations (significant p-values after Bonferroni correction are in bold and in cursive). For site codes see Table 4.1.

	TBS	PBS	BNS	WIS	GOA	AAO	VMS
TBS	-	-	***0.0095***	***0.0128***	***0.0132***	***0.0125***	***0.0251***
PBS	***0.1322***	-	-	-	-	-	-
BNS	***0.2976***	***0.2013***	-	0.0038	0.0038	***0.0122***	***0.0184***
WIS	***0.6551***	***0.6205***	***0.2893***	-	0.0086	0.0097	***0.0153***
GOA	***0.6647***	***0.6317***	***0.2961***	-0.0213	-	***0.0130***	***0.0145***
AAO	***0.8598***	***0.8732***	***0.6119***	***0.1653***	***0.1668***	-	***0.0166***
VMS	***0.1509***	***0.3108***	***0.3809***	***0.6942***	***0.7030***	***0.8688***	-

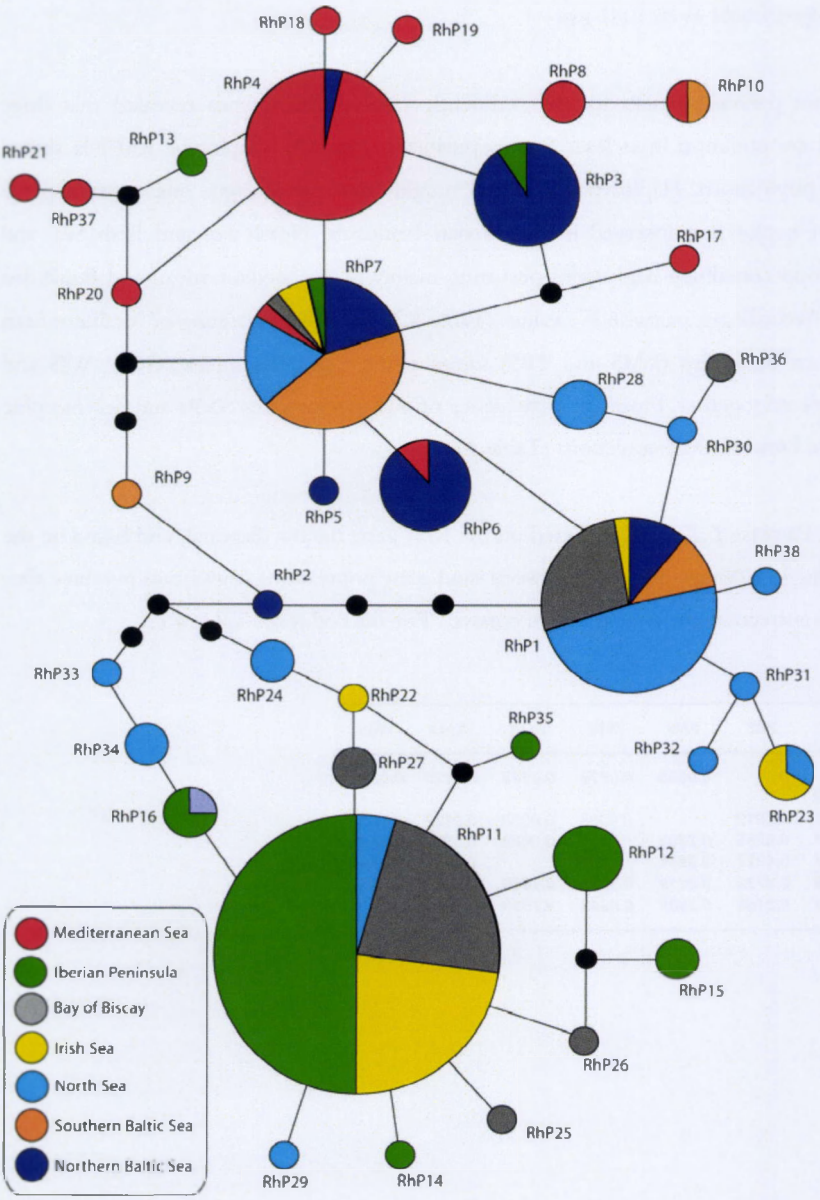


Figure 4.3. Statistical parsimony network of the rhodopsin haplotypes of *Pomatoschistus minutus*. The size of the circles is proportional to the number of gobies sharing that haplotype. Haplotypes are indicated by numbers as given in Table S4.1. Black dots are undetected haplotypes.

Neutrality tests

The null hypothesis of evolution according to the neutral model could not be rejected with a general Z-test for all samples combined as well as for the samples from the Baltic Sea, North Sea and Mediterranean Sea separately. The d_S values were significantly larger than the d_N values only for the Z-tests with the Atlantic samples WIS, GOA and AAO. However, the LRT of the maximum-likelihood analysis demonstrated that M8, the model that accounts for sites under positive selection, showed a significantly better fit than model M7, which does not allow for positive selection. The ω ratio is more than 1 (Table 4.5), indicating positive selection in the *RH1* sequences of *P. minutus*. Bayes identification showed that sites AA151, 214 and 299 of the *RH1* gene were significantly under positive selection.

Table 4.4. Frequency of the AA substitutions detected at the *RH1* gene in seven sand goby populations. The highest frequency in a population is given in bold for each AA. For site codes see Table 4.1.

	Population TBS	PBS	BNS	WIS	GOA	AAO	VMS
AA151 (or SNP4)							
Asn	0.125	0.250	**0.741**	**0.929**	**0.977**	**0.960**	0.107
Thr	**0.875**	**0.750**	0.259	0.071	0.023	0.040	**0.893**
AA214 (or SNP9_10_11)							
Ala	**0.975**	**0.950**	**0.704**	0.286	0.273	0.040	**0.893**
Ile	0.025	0.050	0.296	**0.714**	**0.727**	**0.960**	0
Thr	0	0	0	0	0	0	0.107
AA217 (or SNP 12)							
Ile	**0.625**	**0.950**	**0.981**	**0.976**	**1**	**0.973**	0.179
Thr	0.375	0.050	0.019	0.024	0	0.027	**0.821**
AA261 (or SNP 14)							
Phe	0.450	**1**	**1**	**1**	**1**	**0.973**	**0.857**
Tyr	**0.550**	0	0	0	0	0.027	0.143
AA299 (or SNP 19)							
Ala	**1**	**1**	**0.667**	0.238	0.250	0.054	**1**
Ser	0	0	0.333	**0.762**	**0.750**	**0.946**	0

Table 4.5. Log-likelihood values and parameter estimates for the *RH1* gene sequences of *P. minutus*. Ln L is the log-likelihood value, ω is the selection parameter and p_n is the proportion of sites that fall into ω_n site class. Sites inferred to be under positive selection are given at the 99% (**) confidence interval level.

Model	Ln L	Estimates of paramaters	Positively selected sites
M7 (beta)	-1361.766	p = 0.02286	
		q = 0.14384	
M8 (beta & ω)	-1337.268	q_0 = 0.97818	151**
		(p_1 = 0.02182)	214**
		p = 0.00500	261
		q = 0.27884	299**
		ω = 6.02381	

Table 4.6. Summary of the Mantel tests and Procrustes Analyses correlating pairwise F_{ST}-values based on microsatellites markers versus the *RH1* gene in *Pomatoschistus minutus*. Significant p-values (<0.05) are given in bold. μsats, microsatellite markers; SNPsyn+non, all rhodopsin SNPs of the *RH1* gene; SNPsyn, all synonymous SNPs of the *RH1* gene; SNPnon, all nonsynonymous SNPs of the *RH1* gene; (s), a polymorphic synonymous SNP of the *RH1* gene; (n), a polymorphic nonsynonymous SNP of the *RH1* gene.

Baseline	Marker	Mantel	*P-value Mantel*	Procrustes R	*P-value Procrustes R*
μsats	SNPsyn+non	0.307	0.100	0.218	0.910
	SNPsyn	0.202	0.167	0.143	0.972
	SNPnon	0.428	0.053	0.245	0.919
	SNP1 (s)	0.087	0.315	0.158	0.916
	SNP4 (n)	0.340	0.111	0.153	0.939
	SNP6 (s)	0.188	0.197	0.174	0.946
	SNP7 (s)	0.199	0.190	0.169	0.932
	SNP9 (n)	0.240	0.152	0.601	0.092
	SNP10 (n)	0.208	0.171	0.220	0.916
	SNP11 (s)	0.208	0.172	0.221	0.894
	SNP12 (n)	0.633	**0.048**	0.453	0.433
	SNP13 (s)	0.175	0.242	0.112	0.977
	SNP14 (n)	0.300	0.273	0.351	0.524
	SNP15 (s)	0.012	0.395	0.400	0.410
	SNP16 (s)	0.629	0.108	0.581	0.175
	SNP17 (s)	0.184	0.237	0.288	0.753
	SNP19 (n)	0.223	0.162	0.172	0.931

Overall population differentiation was considerably higher for the full fragment of the *RH1* gene (F_{ST} = 0.4549) than at the nuclear microsatellite markers (F_{ST} = 0.0126). Exact tests showed that no locus or sample exhibited consistent deviations from Hardy-Weinberg equilibrium with respect to microsatellites. The pairwise F_{ST}-values of the *RH1* gene and the microsatellite data (Table 4.3) did not correlate with each other (Procrustes R = 0.218, p-value = 0.910; Mantel R = 0.307, p-value = 0.100) (Table 4.6). Additionally, the simulation-based LOSITAN tests confirmed the neutrality of the microsatellites and the highly significant (>0.99) outlier position of each SNP of the *RH1* gene in comparison with the microsatellites (data not shown).

Light measurements

Figure 4.1 plots the broad spatial variation of the wavelength of maximally transmitted light (WMTL) in the north-eastern Atlantic Ocean and Mediterranean Sea. The default value of 615 nm (coloured in orange) was substituted in case of saturated data at 531 nm, 555 nm or 667 nm. This means that it is difficult to interpret the results along the coasts of Belgium, The Netherlands and the United Kingdom. The underwater light climate is mainly blue in the deep offshore oligotrophic waters of the Mediterranean and the Atlantic Ocean, west of the continental shelf break. Greener waters are found in the Southern North Sea, the southern Baltic Sea, near river mouths and in various shallow coastal waters in the North Sea, Mediterranean Sea and Black Sea. A few isolated areas with an underwater light climate shifted towards red light are found in nearshore regions of the northern Baltic Sea and for a few inland waters. The map therefore shows congruent but more detailed results than previous maps of optical water types (Jerlov 1976).

Discussion

Functional polymorphism of the RH1 *gene in* P. minutus

Jokela *et al.* (2003) found individual differences in the λ_{max} values of the retinal rods of *Pomatoschistus minutus*. Since the differences could not be explained by a chromophore change, they suggested polymorphism of the opsin gene instead of physiological changes.

Sequence analysis revealed substantial variation at the *RH1* gene with 19 SNPs, of which 14 were polymorphic, in seven sand goby populations (Table S4.2). This is the first observation of such a high level of intraspecific variation at a spectral opsin gene in vertebrates. There are five amino acid (AA) replacements, of which some are known to have a significant effect on the λ_{max} values of retinal rods in aquatic vertebrates.

One of the AA substitutions present in sand goby is a phenylalanine to tyrosine substitution of AA261 (SNP14), known for causing a strong red-shift of the λ_{max} values in retinal rods of many teleost families (Hunt *et al.* 1996; Hunt *et al.* 2001; Yokoyama & Takenaka 2004). A comparative study on Salmonidae showed that the Phe261Tyr substitution causes a red-shift of ca. 10 nm in *Salmo salar* in comparison with *Oncorhynchus sp.* (Dann *et al.* 2004). A mutagenesis experiment of this mutant in *Astyanax fasciatus* confirmed a red-shift of 8 nm in λ_{max} values (Yokoyama *et al.* 1995). The second well known mutation is on AA299 (SNP19). This site is localized towards the interior of the retinal binding pocket in helix VII (Fig. 4.2) and close to the Schiff base linkage between the opsin and the chromophore (Hunt *et al.* 2007). It suggests that this AA directly interacts with the chromophore (Fasick & Robinson 1998). A weak blue-shift of the λ_{max} values of retinal rods caused by the Ala299Ser/Thr substitution has already been documented in many teleost families (Yokoyama *et al.* 1995; Hunt *et al.* 2001) and in the bottlenose dolphin (*Tursiops truncatus*) (Fasick & Robinson 1998).

Limited information is available for the three other AA-substitutions in the dataset (AA151, AA214 and AA217). A comparative analysis of all available *RH1* genes of Teleostei on GenBank showed that those three AA sites are not conserved in Teleostei (results not shown). The effect on λ_{max} values of substituted AA214 and AA217 (SNP9_10_11 and SNP12 respectively) has been tested by mutagenesis experiments on red/green opsins (Yokoyama 2000). Only substitution Ile214Thr caused a substantial difference of <5 nm from red to green in red/green opsins (Asenjo *et al.* 1994). However, the effect of a substitution on the AA214 in the rhodopsin gene remains unknown. AA151 (SNP4) is the only AA-substitution that is not located in the helix structure (Fig. 4.2), but may still affect λ_{max} values (Yokoyama *et al.* 2007). Mutagenic experiments on the *RH1* gene are required to study the effect of these five non-synonymous mutations on the λ_{max} values of retinal rods in *Pomatoschistus minutus.*

Network analysis of the significant *RH1* haplotypes (Fig. 4.3) and the highly significant F_{ST} values (Table 4.3) revealed that variation at the *RH1* gene is not randomly distributed. Several populations are differentiated for several AA sites of *RH1*, including the two well known AA (AA261 and 299) that most likely influence the λ_{max} values of retinal rods (Table 4.4). The differentiation is consistent with the λ_{max} values on the retinal rods as measured by Jokela *et al.* (2003). North Sea and Atlantic samples have blue-shifted AA-substitutions instead of the red-shifted substitution in the northern Baltic Sea sample (TBS) (Table 4.4), consistent with the larger λ_{max} values in the northern Baltic Sea gobies compared to their Atlantic relatives.

Based on the wide rod λ_{max} distributions within sand goby populations, Jokela *et al.* (2003) suggested the presence of within-population polymorphism on the *RH1* gene. AA variation on *RH1*, including on AA261 and AA299, is indeed polymorphic in various populations, demonstrating a genetic basis for within-population variation in spectral sensitivity (Table 4.4). The equal distribution of tyrosine and phenylalanine on AA261 (SNP14) in the northern Baltic individuals, can explain the particularly broad λ_{max} distribution spanning 5.7 nm of the spectrum in the population (Jokela *et al.* 2003).

Population differentiation on RH1 *due to selection*

Our results suggest that interpopulation allelic variation of the *RH1* gene is linked to selection as opposed to neutral processes like genetic drift. The d_N/d_S substitution rate ratios of the complete *RH1* fragment in *P. minutus* did not reveal selection. However, tests of neutrality are generally conservative because substitution rates are averaged across all amino-acid sites tested (Bamshad & Wooding 2003). Consequently, analyses of d_N/d_S ratios and likelihood ratio tests under site-specific models detected a significant signal of positive Darwinian selection on the *RH1* gene. Bayesian analysis identified three individual positively selected sites in *RH1*, including AA299, which was verified as a true tuning site for rhodopsin (Table 4.5).

Moreover, selective forces most likely influence the rhodopsin gene of *P. minutus*, since the samples did not group according to geographical or historical proximity with regards to *RH1*

variation (Hemmer-Hansen *et al.* 2007). Strong discrepancies were found between the distribution of the variation at *RH1* and the phylogeographic pattern of the sand goby based on the distribution of the variation at the mtDNA Cyt *b* gene (*subchapter 3a*) and nuclear allozyme (Stefanni *et al.* 2003) and microsatellite markers (*subchapter 3b*). Samples of the northern Baltic Sea and Mediterranean Sea carry a similar allelic profile of the *RH1* gene (Fig. 4.3; Table 4.2, 4.3), although historically the Mediterranean *P. minutus* individuals belong to a different phylogeographic mtDNA clade (MS-Clade) than the Atlantic and Baltic sand gobies (AO-Clade) (*subchapter 3a*). The *RH1* gene was also congruent between sand gobies from the Iberian Peninsula and the Irish Sea - Bay of Biscay. However, the Iberian sand gobies belong to a different historical unit (IB-Group) compared to the North Atlantic gobies (NA-Group), which includes all the populations from the Bay of Biscay to the northern Baltic. Additionally, distributions of the *RH1* and microsatellite variation were statistically significant different from each other (Table 4.6). No convincing evidence was found for non-neutrality of any of the microsatellites used to represent the neutral baseline. In contrast, each SNP of the *RH1* gene was clearly identified as an outlier locus in comparison with the microsatellite markers. Therefore, random processes may be ruled out to explain the functional differentiation on the *RH1* gene between the different sand goby populations.

The link between functional variation on RH1 *and environmental light climate*

The significant discrepancies at various levels between the distribution of neutral markers and *RH1* gene data suggest that its variation is influenced by the optical environment instead of genetic drift. Differences in optical characteristics clustered the sampling locations into three groups: the Mediterranean Sea, Iberian Peninsula, the Bay of Biscay and Irish Sea (VMS, AAO, GOA and WIS) with a mainly blue light climate; the southern Baltic and North Sea (PBS and NBS) with greener water; and the northern Baltic Sea (TBS) with water with the highest WMTL values (Figure 4.1, Jerlov 1976). In general, these robust differences in environmental light transmittance correspond well with the differences in the absorbance spectra of the retinal rods (Jokela *et al.* 2003) and with functional variation at the *RH1* gene. The blue-shift of the λ_{max} values of the dim-light receptors and the highest frequency of the Ala299Ser substitution in the Bay of Biscay, Irish Sea and the Iberian Peninsula is characteristic for offshore blue water. The red-shift of the λ_{max} values of the rods and the

highest frequency of Phe261Tyr substitution in the northern Baltic might be an adaptation to the red-shifted light condition in this region. One remarkable observation is that the individuals of the Mediterranean Sea clustered with the northern Baltic samples based on the *RH1* variation. In the Mediterranean Sea, coastal lagoons play an important role in the life cycle of sand gobies as nurseries and feeding sites (Bouchereau & Guelorget 1998), while this is not the case for Atlantic sand gobies (Guelinckx *et al.* 2008). These lagoons, which are not included in Figure 4.1, are characterised by a much higher turbidity than offshore (Poizat *et al.* 2004). Such conditions are thought to require spectral adaptations similar to those in the Northern Baltic.

Conclusion

The three conditions to demonstrate local adaptation at the rhodopsin gene of the sand goby are fulfilled. First, functional polymorphism was observed in the *RH1* gene. Then, it was demonstrated that population differentiation at the *RH1* gene was due to selection. Finally, a correlation was found between *RH1* variation and the specific spectral characteristics of the habitat environment of the sand goby. Therefore, there are good indicators for local adaptation of the rhodopsin gene in *P. minutus*. Further molecular research with a higher sampling resolution in space and time is required to disentangle the temporal variability of the *RH1* polymorphism and the small-scale differentiation on the *RH1* gene for *P. minutus* inside the various marine systems.

The hypothesis that sand gobies are evolutionary adapted to their optical environment implies that rapid changes in optical habitat characteristics may have negative consequences. For example, increased water turbidity by algal blooms in the highly polluted Baltic Sea negatively influenced sexual behaviour of fishes with a visual mating system. Cases have been documented for sand goby (Järvenpää & Lindström 2004) and three-spined stickleback (*Gasterosteus aculeatus*) (Engström-Öst & Candolin 2007; Candolin *et al.* 2008). Water quality of lagoonal and coastal waters can be influenced by anthropogenic changes in the nutrient load and by climatic factors. Therefore if temperature continues to rise, spectral transmission of the water may shift with temperature (Archer *et al.* 2001). A marine monitoring program for water clarity and optical properties is therefore recommended, not only to consider their

effect on primary productivity, but also because of their direct influence on the visual capacity of the fish community (Aksnes 2007) and other organisms.

Acknowledgements

We thank the following persons who generous provided samples: Pilar Drake (CSIC, Spain), Maria Järvi-Laturi (University of Helsinki, Finland), Kai Lindström (Åbo Akademi University, Finland), Katarzyna Waligóra-Borek (University of Gdańsk, Poland), Ian McCarthy (University of Wales, Bangor, UK), Mario Lepage (Ifremer, France), Jef Guelinckx (Flemish Goby Consortium, Belgium) & Alain Crivelli (Station Biologique de la Tour du Valat, France). We also thank the 'Marine Genomics Europe' Network of excellence, subject editor Michael Hansen, five anonymous referees and Anne Christine Utne-Palm, Luisa Orsini, Gregory Maes, Hendrik Larmuseau, Bart Hellemans & Erika Souche for useful discussions and technical assistance. The first author received a PhD-fellowship of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The second author received a postdoctoral fellowship of Fund for Scientific Research – Flanders (FWO-Vlaanderen). This study was part of the WESTBANKS project financed by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (contract nr SD/BN/01A).

Appendix of Subchapter 4a

Appendix S4.1 Estimates of wavelengths of maximally transmitted light for European seas: theory and description of method.

Underwater light field

At any point and time the underwater light field can be fully defined by the radiance distribution, which gives light intensity for all viewing directions and wavelengths. The vertical decrease of this light intensity is approximately exponential with the decay constant for each wavelength defined via a diffuse attenuation coefficient, the definition of which depends on how the radiance distribution is directionally integrated (downwelling/scalar, upward/downward/total). These attenuation coefficients depend in turn on the inherent optical (scattering and absorption) properties of the water and its constituents and, to a lesser extent, on the angular distribution of radiance. Detrital and mineral suspended particles (NAP or Non-Algal Particles) and Coloured Dissolved Organic Matter (CDOM) absorb blue light (~420-480 nm) strongly and green light (~480-570 nm) weakly. Phytoplankton absorbs mainly blue light. Pure water molecules absorb red light (~570-670 nm) strongly and green light weakly but absorbs very little blue light. As a result, in the clearest natural waters with low suspended particulate concentrations and low CDOM, it is blue light that is least attenuated with depth and the underwater light climate is dominated by blue light. This is typical of deep oceanic waters, which are predominantly blue except during phytoplankton blooms when there is a shift towards green. In more coastal waters there may be sources of NAP (tidal resuspension, river discharges) and/or CDOM (associated with low salinity water from rivers). As NAP or CDOM increases, the water colour and underwater light climate shift from blue to green. For very turbid waters (high NAP) the colour may even shift to brown/red as discussed by Kirk (1996). Lindström (2000) introduced the concept of 'wavelength of maximally transmitted light' (WMTL) in order to characterize the spectral content and depth variation of the underwater light climate by a single parameter for comparison with the spectral sensitivity of the eyes of marine animals.

Wavelength of maximally transmitted light - theory

Whereas a combination of underwater light measurements and optical modelling is used by Audzijonyte *et al.* (2005) to estimate the WMTL at various locations, a method is described here to estimate this parameter from satellite remote sensing data. This method has the advantages of providing information at almost any location on earth without the need for *in situ* measurements or *a priori* knowledge and of giving a more uniform methodology.

The WMTL, λ_{lmax}, can be formally defined as the spectral minimum of a diffuse attenuation coefficient. While the latter depends on both the definition of angular integration and on depth, it is convenient to note (see Mobley 1994 and references therein) that all diffuse attenuation coefficients tend with increasing depth towards a single asymptotic diffuse attenuation coefficient, denoted here as $k_\infty(\lambda)$, where λ represents wavelength. Theoretical and experimental work summarized by Mobley (1994) shows that $k_\infty(\lambda)$ is a function of the total absorption $a(\lambda)$ and scattering $b(\lambda)$ coefficients of the medium (pure water, CDOM and particles) and of the scattering phase function. For the purposes of the current study a simplification of this dependence is proposed here which gives a reasonable fit to the numerical solution of the asymptotic radiance distribution problem for a commonly used particle phase function shown in Figure 9.2 of Mobley (1994):

$$k_\infty(\lambda) = a(\lambda) + 0.1 * b(\lambda) \tag{1}$$

Ocean colour sensors such as SeaWiFS, MODIS and MERIS provide routinely the normalized water-leaving radiance at the sea surface, $nLw(\lambda)$. This can be related to the absorption and backscattering, $b_b(\lambda)$ coefficients via theoretical models such as the model of Gordon *et al.* (1988) which gives

$$\frac{nLw}{E_0^{TOA}}(\lambda) = \left(\frac{\pi f' \Re}{Q}\right)\frac{b_b}{a + b_b} \tag{2}$$

where $\pi f' \Re / Q$ groups terms representing air-sea interface refraction and reflection (Morel & Gentili 1996) as well as the empirical relationship between directional water-leaving reflectance and inherent optical properties; $E_0^{TOA}(\lambda)$ is the extra terrestrial solar irradiance.

The main assumptions of the method proposed here are that, for estimation of λ_{lmax}:

a) Wavelength variation of the factor $\pi f' \Re / Q$ can be neglected

b) Wavelength variations of the scattering and backscattering coefficients, $b(\lambda)$ and $b_b(\lambda)$, can be neglected.

Under these assumptions the wavelength variation of both k_∞ and nLw/E_0^{TOA} is determined entirely by the wavelength variation of the absorption coefficient, $a(\lambda)$. The wavelength λ_{lmax} giving the spectral minimum of $k_\infty(\lambda)$ will, therefore, give the spectral maximum for nLw/E_0^{TOA}. Put more simply, the wavelength of light with strongest reflectance as seen from above the surface is also the wavelength that predominates at depth because it is the least absorbed.

Wavelength of maximally transmitted light – estimation from MODIS data

Data for $nLw(\lambda)$ as measured by the MODIS-AQUA satellite sensor was downloaded on 19th November 2008 from the NASA 'Ocean Color' web site (http://oceancolor.gsfc.nasa.gov/) as the annual composite for 2007 (4 km standard map image file) for each of the available spectral bands, 412 nm , 443 nm, 488 nm, 531 nm, 551 nm and 667 nm. This data was then georeferenced, cropped to the European Seas region, radiometrically scaled using the embedded slope coefficients and divided by tabulated values for $E_0^{TOA}(\lambda)$ to obtain $nLw/E_0^{TOA}(\lambda)$ spectra. For each pixel in this image dataset the wavelength of maximum nLw/E_0^{TOA} was identified and output as λ_{lmax}. The resulting map of λ_{lmax} is shown in Figure 4.1. For some turbid water regions the MODIS dataset is saturated at wavelengths 531 nm, 551 nm or 667 nm and is truncated at a maximum value. In this case λ_{lmax} is set equal to the arbitrary value of 615 nm (orange colour in Figure 4.1) because it cannot be estimated correctly using this data set. In reality λ_{lmax} may lie somewhere between the MODIS wavelengths and spectral modelling could be added to this method to interpolate, particularly for the spectral range 551-667 nm where there are no MODIS wavelengths.

Appendix S4.2 Supplementary tables

Table S4.1. Rhodopsin haplotypes and their geographical distribution in *Pomatoschistus minutus*. The numbers of haplotypes reconstructed with a probability ≥ 95% (estimated using PHASE 2.0.2) are listed in bold. The numbers of haplotypes that were non-significantly reconstructed are given in italic and between brackets. Dots indicate homology with haplotype RhP1. For site abbreviations see Table 4.1.

		SNP nr. and mutation																	Haplotype tally within populations						
		1	2	3	4	5	6	7	8	9_10_11	12	13	14	15	16	17	18	19							
Haplotype nr.	Acc. Nr.																		TBS	PBS	BNS	WIS	GOA	AAO	VMS	All
RhP1	FJ410451	A	G	C	A	C	T	T	C	GCT	T	T	T	C	C	A	C	G	**4**	**4**	**18** *(+2)*	**3** *(+1)*	**10**	-	*(1)*	**39** *(+4)*
RhP2	FJ410452	.	.	.	.	.	.	.	.	ATC	.	.	.	.	.	.	.	.	1	-	-	-	-	-	-	**1**
RhP3	FJ410453	.	.	.	C	.	.	.	.	.	C	.	A	.	.	.	.	.	**10** *(+3)*	-	-	-	-	**1**	-	**11** *(+3)*
RhP4	FJ410454	.	.	.	C	.	.	.	.	.	C	.	.	.	.	.	.	.	**1**	-	-	-	-	-	**29** *(+6)*	**30** *(+6)*
RhP5	FJ410455	.	.	.	C	A	.	.	.	.	.	.	.	.	.	.	.	.	**1**	-	-	-	-	-	-	**1**
RhP6	FJ410456	.	.	.	C	.	.	.	.	.	.	.	A	.	.	.	.	.	**9**	-	-	-	-	-	**1** *(+1)*	**10** *(+1)*
RhP7	FJ410457	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	**6** *(+4)*	**14**	**8**	**2** *(+1)*	**1**	**1**	**1**	**33** *(+5)*
RhP8	FJ410458	.	.	.	C	.	.	.	.	.	C	.	.	.	T	.	.	.	*(1)*	-	*(1)*	-	-	-	**2**	**2** *(+2)*
RhP9	FJ410459	.	.	.	C	.	.	.	.	ATC	.	.	.	.	.	.	.	.	-	1	-	-	-	-	-	**1**
RhP10	FJ410460	.	.	.	.	.	.	.	.	.	C	.	.	.	T	.	.	.	-	**1**	-	*(1)*	-	-	**1** *(+1)*	**2** *(+2)*
RhP11	FJ410461	G	.	.	.	.	A	A	.	ATC	.	G	.	.	.	C	.	T	-	-	**8**	**26** *(+1)*	**27**	**58**	-	**119** *(+1)*
RhP12	FJ410462	G	.	.	.	.	A	A	.	ATC	.	G	.	T	.	C	.	T	-	-	-	-	-	**6**	-	**6**
RhP13	FJ410463	.	.	C	.	.	.	.	.	.	C	G	.	.	.	.	.	.	-	-	-	-	-	**1**	-	**1**
RhP14	FJ410464	G	.	G	.	.	A	A	.	ATC	.	G	.	.	.	C	.	T	-	-	-	-	-	**1**	-	**1**
RhP15	FJ410465	G	.	.	.	.	.	.	.	ATC	.	G	.	T	.	C	.	T	-	-	-	-	-	**2**	-	**2**
RhP16	FJ410466	.	.	.	.	.	A	A	.	ATC	.	G	.	.	.	C	.	T	-	-	-	-	**1**	**3**	-	**4**
RhP17	FJ410467	.	.	.	C	.	.	.	.	.	.	.	A	.	T	.	.	.	-	-	-	-	-	-	*1 (+5)*	*1 (+5)*
RhP18	FJ410468	.	.	.	C	.	.	.	.	.	C	.	.	.	.	.	T	.	-	-	-	-	-	-	**1**	**1**
RhP19	FJ410469	.	.	.	C	.	.	A	.	.	C	.	.	.	.	.	.	.	-	-	-	-	-	-	**1**	**1**
RhP20	FJ410470	.	.	.	C	.	.	.	.	ACT	C	.	.	.	.	.	.	.	-	-	-	-	-	-	*1 (+1)*	*1 (+1)*
RhP21	FJ410471	.	.	.	.	.	.	.	.	ACT	C	G	.	.	.	.	T	.	-	-	-	-	-	-	*1 (+1)*	*1 (+1)*
RhP22	FJ410472	G	.	.	.	.	A	A	.	ATC	.	G	.	.	.	.	.	.	-	-	-	**1**	-	-	-	**1**
RhP23	FJ410473	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	-	-	**1**	**2** *(+2)*	-	-	-	**3** *(+2)*
RhP24	FJ410474	G	.	.	.	.	A	A	.	ATC	.	.	.	.	.	.	.	.	-	-	**2**	*(2)*	-	-	-	**2** *(+2)*
RhP25	FJ410475	G	.	.	.	.	A	A	.	ATC	.	.	.	.	.	C	.	T	-	-	-	-	**1**	-	-	**1**
RhP26	FJ410476	G	.	.	.	.	.	A	.	ATC	.	G	.	.	.	C	.	T	-	-	-	-	**1**	-	-	**1**
RhP27	FJ410477	G	.	.	.	.	A	A	.	ATC	.	G	.	.	.	.	.	T	-	-	-	-	**2**	-	-	**2**
RhP28	FJ410478	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	-	-	**3** *(+1)*	-	-	-	-	**3** *(+1)*
RhP29	FJ410479	G	.	.	C	.	A	A	.	ATC	.	G	.	.	.	C	.	T	-	-	**1**	-	-	-	-	**1**
RhP30	FJ410480	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	-	-	**1**	-	-	-	-	**1**
RhP31	FJ410481	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	-	-	**1**	-	-	-	-	**1**
RhP32	FJ410482	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	-	-	**1**	-	-	-	-	**1**
RhP33	FJ410483	.	.	.	.	.	A	.	.	ATC	.	G	.	.	.	.	.	T	-	-	**1**	-	-	-	-	**1**
RhP34	FJ410484	.	.	.	.	.	A	.	.	ATC	.	G	.	.	.	C	.	T	-	-	**2** *(+2)*	-	-	-	-	**2** *(+2)*
RhP35	FJ410485	G	.	.	.	.	A	A	.	ATC	.	G	A	.	.	C	.	.	-	-	-	-	-	**1**	-	**1**
RhP36	FJ410486	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	.	-	-	-	-	**1**	-	-	**1**
RhP37	FJ410487	.	.	.	.	.	.	.	.	ACT	C	G	.	.	.	.	.	.	-	-	-	-	-	-	**1**	**1**
RhP38	FJ410488	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	**1**	-	-	-	-	**1**
Total																			**32** *(+8)*	**20**	**48** *(+6)*	**34** *(+8)*	**44**	**74**	**40** *(+16)*	**292** *(+38)*

Table S4.2 Nucleotide polymorphisms at the *RH1* gene in seven sand goby populations. Dots indicate homology with the reference sequence (Acc. number X62405 or haplotype RhP1). AA numbers are listed for the non-synonymous mutations, which are listed in bold. For site codes see Table 4.1.

SNP nr.	1	2	3	**4**	5	6	7	8	**9**	**10**	11	**9_10_11**	**12**	13	**14**	15	16	17	18	**19**
Nucleotide position	60	126	171	**296**	297	339	357	373	**484**	**485**	486	**484-486**	**494**	525	**626**	648	669	684	738	**739**
Amino acid nr.	-	-	-	**151**	-	-	-	-	-	-	-	**214**	**217**	-	**261**	-	-	-	-	**299**
Reference	A	G	C	**A**	C	T	T	C	**G**	**C**	T	**GCT**	T	T	T	C	C	A	C	**G**
TRS	.	.	.	**./C**	./A	.	.	.	**./A**	**./T**	./C	**./ATC**	**./C**	.	**./A**	.	./T	.	.	.
PBS	.	.	.	**./C**	.	.	.	.	**./A**	**./T**	./C	**./ATC**	**./C**	.	.	.	./T	.	.	.
BNS	./G	./A	.	**./C**	.	./A	./A	.	**./A**	**./T**	./C	**./ATC**	**./C**	./G	.	.	./T	./C	.	**./T**
WIS	./G	.	.	**./C**	.	./A	./A	.	**./A**	**./T**	./C	**./ATC**	**./C**	./G	.	.	./T	./C	.	**./T**
GOA	./G	.	.	**./C**	.	./A	./A	.	**./A**	**./T**	./C	**./ATC**	.	./G	.	.	.	./C	.	**./T**
AAO	./G	.	./G	**./C**	.	./A	./A	.	**./A**	**./T**	./C	**./ATC**	**./C**	./G	**./A**	./T	.	./C	.	**./T**
VMS	.	.	.	**./C**	.	.	.	./A	**./A**	.	.	**./ACT**	**./C**	./G	**./A**	.	./T	.	./T	.

SUBCHAPTER 4B

DIFFERENTIAL MODE OF ADAPTATION TO THE RHODOPSIN GENE IN COASTAL BALTIC AND NORTH SEA POPULATIONS OF THE SAND GOBY, *POMATOSCHISTUS MINUTUS*

Maarten H.D. Larmuseau[1], Kim Vancampenhout[1], Jeroen K.J. Van Houdt[1,2] & Filip A.M. Volckaert[1]

[1]Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Leuven, Belgium.

[2]Royal Museum for Central Africa, Tervuren, Belgium.

Submitted manuscript

Abstract

Knowledge of the spatial and temporal scale of adaptive genetic variation in marine systems remains scarce, yet crucial to improve our understanding of how evolution operates in the marine environment. An excellent model to elucidate the mechanisms and importance of selection as evolutionary force is the spectral tuning mechanism of the visual pigment (VP) in marine vertebrates. In the sand goby *Pomatoschistus minutus* (Teleostei; Gobiidae) evidence was recently found for adaptation of the rhodopsin (*RH1*) gene to the specific light environment of marine systems. The aim of this study is to characterize the mode of selection at the *RH1* gene on a micro-scale in time and space for sand gobies of the Baltic - North Sea region. *RH1* sequences of 491 *Pomatoschistus minutus* individuals from 15 locations were analysed. The broad sampling supported local adaptation at the rhodopsin gene in the Baltic Sea and North Sea despite spatial and temporal variation on a micro-scale in the light regime. Sand gobies may have adapted to the specific Baltic environment by selection on a combination of standing genetic variation and possibly *de novo* mutation in the rhodopsin gene. In contrast to the Baltic Sea population, balancing selection at the *RH1* gene seemed to characterise the sand gobies of the North Sea since no functional differentiation was found

between samples in spite of the occurrence of a neutral population structure. The high level of polymorphism on the rhodopsin gene might be maintained either by the heterogeneity of the light regime along the coastline of the North Sea or by the migration patterns of juvenile sand gobies between the open sea and turbid coastal environments. The tendency that the genomes of estuarine migrants carry rhodopsin variants associated with more turbid brackish water, illustrates the importance of turbidity and visual capacity of sand gobies in their habitat preferences.

Keywords: Adaptive evolution, candidate genes, *de novo* mutation, genetic structure, Gobiidae, marine fish, North Sea, photoreceptor genes, postglacial expansion, vision

Introduction

Understanding the genetic basis of local adaptation is of prime interest in biology as it involves the role of natural selection in promoting evolutionary change (Mayr 1963). Since gene flow is expected to slow down adaptive population divergence, the traditional idea was that local adaptation may be rare or absent in marine organisms (Hemmer-Hansen *et al.* 2007). Today, local adaptation in marine organisms has become increasingly documented, indicating that natural selection is as well a potent evolutionary force in the 'open' ocean (Canino *et al.* 2005; Pampoulie *et al.* 2006; Hemmer-Hansen *et al.* 2007; Sherman & Ayre 2008). Nevertheless, knowledge of the spatial and temporal scale of adaptive genetic variation in marine systems remains scant, yet crucial to improve our understanding of how evolution operates in the ocean (Conover *et al.* 2006; Zane 2007).

Adaptive evolution in the ocean is poorly known because of the scarcity of genetic systems to evaluate natural selection (Yokoyama 2002). One of the few well known models to elucidate the mechanism and importance of selection as evolutionary force is the spectral tuning mechanism of the visual pigment (VP) in marine vertebrates (Yokoyama 2000). This model can identify specific amino-acid changes that are responsible for the adaptation of organisms to specific environments. Vertebrates have VP molecules bound in dense membrane stacks in retinal photoreceptors to mediate vision. The VP protein moiety is opsin, which is a G protein linked receptor, bound to a light-sensitive chromophore (Park *et*

al. 2008). Each pigment shows a characteristic peak of maximal absorbance (λ_{max}), its precise location depending on the interactions between the chromophore and the opsin protein. To modify their visual system to cope with the photic environment, spectral tuning of the VP proteins can be assessed at the physiological and at the DNA level. Spectral tuning can be achieved on a physiological time scale by exchanging the chromophore (A1 or A2 pigments), or on an evolutionary time scale by amino acid (AA) substitutions in the opsin part that change the λ_{max} of VP (Bowmaker 2008). Evolutionary adaptation of VP genes has been established in several phylogenetic studies of marine organisms (Hunt *et al.* 2001; Yokoyama & Takenaka 2004).

Intraspecific local adaptation at the rhodopsin (*RH1*) gene seems to be influencing fitness of the sand goby *Pomatoschistus minutus* (Pallas 1770; Gobiidae, Teleostei) (*subchapter 4a*). The sand goby is a relatively small demersal species that is very abundant along the European coasts and its estuaries (Bouchereau & Guelorget 1998). It forms an important ecological link between benthic invertebrates and larger predatory fish such as cod and whiting (Maes *et al.* 2003). Its nocturnal foraging has advantages for approaching prey and avoiding predators (Ehrenberg & Ejdung 2008). As visual predator (Healey 1971; Aarnio & Bonsdorff 1993), this requires a good sight in very dim light (Thetmeyer 1997). The geographical distribution of the sand goby encompasses a wide range of photic environments, varying in turbidity, colour and brightness. Therefore, adaptation to the local spectral environment was presumed to be crucial for *P. minutus*. As expected, positive selection was recently demonstrated for the sand goby at the *RH1* gene, which is crucial for dim light vision in vertebrates (*subchapter 4a*). A clear link was found between the functional differentiation at the *RH1* gene and the photic climate of the environment, characterized by the value of the 'wavelength of the maximum transmitted light' (WMTL).

A strong difference in the *RH1* gene was found in sand goby samples from the Baltic and North Sea. Differentiation of the polymorphic spectral tuning with a known effect on the λ_{max} of the rod opsin was congruent with the on average higher WMTL values in the Baltic than in the North Sea (Jerlov 1976). However, only one sample per marine system was included in the macro-scale spatial analysis of the *RH1* variation, despite the presence of differences in the WMTL-pattern within both marine systems (*subchapter 4a*). In the Baltic

Sea, the photic regime of the whole Baltic and Bothnian Sea has a stable temporal pattern based on WMTL values (Lindström 2000) in contrast to the habitat for the sand goby in the southern North Sea. The latter habitat area, with its highly turbid estuaries, is characterized by heterogeneity in light conditions in space and time (*subchapter 4a*). Therefore, it is presumed that the mode of adaptation at the *RH1* gene will be differentiated between and within both marine systems.

Further molecular research with a higher sampling resolution is required to disentangle the spatial and temporal variability of the *RH1* polymorphism in *P. minutus* inside the Baltic-North Sea region. The aim of the present study was to find answers to the following research questions: (1) What is the *RH1* variation of the sand goby within marine systems in comparison with the variation in the combined Baltic-North Sea region? (2) What is the mode of adaptation at the *RH1* gene of the sand goby in a temporally stable and unstable environment?

Material & Methods

Sampling and species identification

A total of 460 *Pomatoschistus minutus* individuals were caught at 13 locations along the coasts and estuaries of the Baltic-North Sea region (Table 4.7, Fig. 4.4). Samples were taken either by fyke, hand net, beam trawl or from the cooling-water intake screens of nuclear power plants (Borssele, NL and Doel, B). Sand gobies were distinguished from other cryptic *Pomatoschistus* species morphologically based on the dermal head papillae (Miller 1986) and pigmentation pattern (Hamerlynck 1990), and genetically, based on a PCR-RFLP species-specific identification protocol (*subchapter 2a*). The sex was determined by the shape of the urogenital papilla (Rodrigues *et al.* 2006).

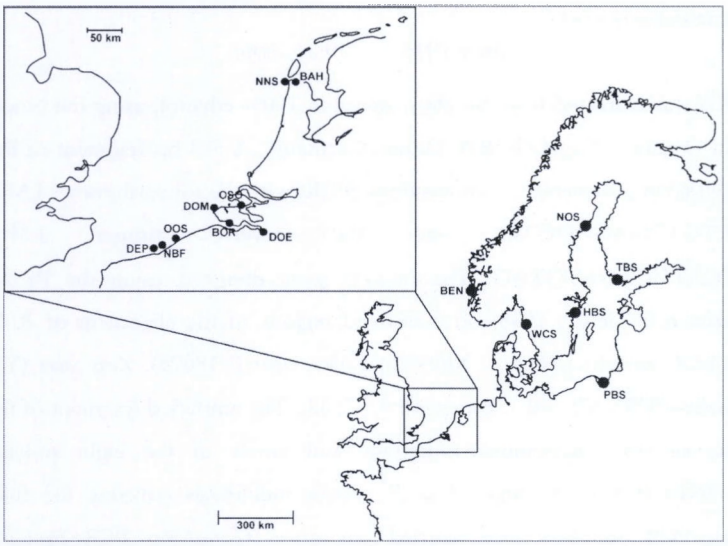


Figure 4.4 Geographical distribution of the 15 sampling locations in the Baltic-North Sea region for the sand goby, *Pomatoschistus minutus*. See Table 4.7 for definitions of the sampling locations.

Table 4.7 Overview of the *Pomatoschistus minutus* samples collected at 15 sampling sites along the coasts of the North Sea and Baltic Sea. N, sample size.

Code	Area	Country	Location	Date	Coordinates	N
NOS	Bothnian Sea	Sweden	Nordmaling	July/2008	63°24' N-19°45' E	28
TBS	Northern Baltic Sea	Finland	Tvärminne, Vargskar island	July/2006	59°50' N-23°12' E	20
HBS	Northern Baltic Sea	Sweden	Nasaäng	October/2005	58°59' N-17°27' E	26
PBS	Southern Baltic Sea	Poland	Sopot, Bay of Gdańsk	February/2007	54°27' N-18°36' E	10
WCS	Kattegat	Sweden	Bökevik Bay, Skaftö island	July/2006	58°14' N-11°26' E	30
BEN	Northern North Sea	Norway	Bergen	July/2008	60°16' N-04°59' E	20
NNS	Southern North Sea	Netherlands	Stuifdijk	March/2003	52°58' N-04°42' E	36
BAH	Southern North Sea	Netherlands	Balgzand	August/2007	52°56' N-04°53' E	40
DOM	Southern North Sea	Netherlands	Domburg	March/2004	51°34' N-03°29' E	42
OSO	Oosterschelde	Netherlands	Kattedijke	April/2004	51°33' N-03°58' E	42
BOR	Scheldt Estuary	Netherlands	Borssele	March/2004	51°24' N-03°43' E	43
DOE	Scheldt Estuary	Belgium	Doel	March/2004	51°18' N-04°16' E	39
OOS	Southern North Sea	Belgium	Oostende	March/2005	51°15' N-02°55' E	42
NBF	Southern North Sea	Belgium	Oostduinkerke	August/2008	51°08' N-02°39' E	39
DEP	Southern North Sea	Belgium	De Panne	March/2004	51°06' N-02°34' E	34

Gene amplification and sequencing

Genomic DNA was extracted from fin clips, stored in 100% ethanol, using the NucleoSpin Extraction Kit (Machery-Nagel GmBH, Düren, Germany). A 548 bp fragment of the *RH1* gene was amplified in polymerase chain reactions (PCR) with the forward primer LMCercleF GTCCTGGCTGTTGAGAGGTG and the reverse primer LMCercleR TGCTTGTTCATGCAGATGTAG. The primers were designed using the PRIMER 3 program (Rozen & Skaletsky 1998) on conserved regions of the alignment of *RH1* gene sequences from *P. minutus* (Acc. nr. X62405), *Gobius niger* (Y18675), *Zeus faber* (Y14484), *Sargocentron diadema* (U57537) and *S. microstoma* (U57542). The amplified fragment of the *RH1* gene contains all non-synonymous mutations and seven of the eight polymorphic synonymous SNPs that were observed in *P. minutus* individuals covering the full range (*subchapter 4a*). PCR reactions were carried out on a GeneAmp PCR System 2700 thermocycler (Applied Biosystems, Foster City, CA, USA) in a total volume of 25 µl, containing 1 µl of genomic DNA, 1 X PCR buffer, 0.2 mM dNTPs, 0.8 µM of each primer, 2.0 mM $MgCl_2$, 0.5 U of Taq DNA polymerase (Silverstar, Eurogentec, Seraing, Belgium) and mQ-H_2O. The PCR profile was: 4 min at 94°C followed by 35 cycles of 30 s at 96°C, 30 s at 56°C and 1 min at 72°C; with a final 10 min extension at 72°C. To avoid contamination, separate pipettes, aerosol barrier tips and sections of the laboratory were used for pre- and post-PCR work. Every other 15th individual (corresponding with one every two rows of a PCR-plate) a negative control was inserted to detect contamination. No contamination occurred during the screening procedures. All PCR products were visualized on agarose gels with ethidium bromide. After purification with the 'GFX PCR DNA and Gel Band Purification kit' (GE Healthcare, Piscataway, NJ, USA), the PCR products were sequenced in both directions using the BigDye Terminator v. 3.1 Cycle Sequencing Kit on an ABI 3130 automated capillary DNA sequencer (Applied Biosystems). Sequences were checked and aligned to each other with SEQSCAPE v. 2.1 (Applied Biosystems). Automated detection of point mutations was realized with the GAP4 subprogram embedded in the STADEN package v. 1.7 (http://sourceforge.net/projects/staden) and rechecked manually by eye. For the statistical analyses also the samples TBS (northern Baltic Sea) and PBS (southern Baltic Sea) from *subchapter 4a* were included in the further statistical analysis (Table 4.7).

The haplotypes of each individual were inferred from the genotypes using the Bayesian statistical methods in the program PHASE v. 2.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003). Runs were conducted separately for each population, with known haplotype information being included as prior information. Ten independent runs per population were conducted, each with a burn-in-period of 1000 followed by 10,000 iterations with a thinning interval of 100 steps. The results and the goodness-of-fit values were very similar among the runs, indicating that the run lengths were sufficient. The haplotypes of the samples TBS and PBS were already reconstructed in *subchapter 4a* and could be included as prior information for the PHASE analysis.

Population diversity and deviation from Hardy-Weinberg

The number of segregating sites (S), the mean number of pairwise differences (k) and estimates of nucleotide polymorphism (π, θ) were calculated for the haplotype data of the *RH1* gene using DNAsp v. 4.10.9 (Rozas *et al.* 2003). Tests for Hardy-Weinberg equilibrium (HWE) were calculated in FSTAT v. 2.9.3 (Goudet 2001).

Genetic differentiation

Four different methods were used to reveal the population substructure in *P. minutus* based on the SNP data at the *RH1* gene. First, population differentiation was quantified in GENETIX v. 4.05 (Belkhir *et al.* 2004) using the standardized allelic variation F_{ST}, estimated as Θ (Weir & Cockerham 1984). F_{ST}-linked pairwise genetic distances were calculated according to Cavalli-Sforza & Edwards (1967) (D_{CE}) with GENETIX. F_{ST}s and D_{CE}s were tested for significance against 10^4 random permutations of the data in GENETIX. Pairwise F_{ST} values between the samples were as well estimated separately for synonymous SNPs as for nonsynonymous SNPs of the *RH1* fragment. Second, the pairwise genetic distances were used in a non-metric multidimensional scaling analysis (NMDS) in order to reveal a group structure in STATISTICA v. 6.0 (StatSoft, Inc., Tulsa, OK, USA). Third, the impact of the observed structuring in the NMDS on the genetic structure was tested using a hierarchical analysis of molecular variance (AMOVA; ARLEQUIN, v. 3.0) (Excoffier *et al.* 2005). Other AMOVA-tests were performed based on *a priori* knowledge about the neutral genetic

population structure and population dynamics of populations from the southern North Sea (Pampoulie *et al.* 2004; Guelinckx 2008). Fourth, the correlation between the genetic and geographic distances among all 15 populations of the dataset and among the populations of the southern North Sea was tested using a Mantel test with 1000 permutations in GENETIX. Geographical distances were measured as the shortest coastal distances between sites using the electronic atlas Google Earth (http://earth.google.com).

Finally, to detect sex-specific differences in inheritance and migration rates, deviation from Hardy-Weinberg equilibrium (F_{IS}), differentiation among populations (F_{ST}), relatedness (r) (Queller & Goodnight 1989), mean assignment (mAI_C) and variance of the assignment (vAI_C) were quantified separately for both sexes over all populations (Goudet *et al.* 2002). Statistical significance of differences in these indices was determined with 10,000 permutations using the randomization method implemented in FSTAT.

Results

Nucleotide diversity of the RH1 *gene*

The sequences matched the general properties of the *Pomatoschistus minutus RH1* gene (X62405) (Archer *et al.* 1992). In total, 19 segregating sites or SNPs were noticed across all genotypes (Table S4.3); 14 of them had already been observed for the sand goby and are named as in *subchapter 4a*. Five synonymous SNPs (SNP 13b, 13t, 13q, 17b, 17t) were observed for the first time but they were not polymorphic according to the 99% criterion, similar to the SNPs 5 and 18. The alignment in amino acids (AA) showed the same five non-synonymous AA substitutions as in *subchapter 4a*. PHASE analysis revealed 35 confidently resolved haplotypes (Table S4.4). Due to the lack of information on three variable nucleotide sites (SNPs 1, 2 and 3), the haplotypes are called 'haplogroups' to differentiate with the haplotypes of *subchapter 4a*.

The lowest *RH1* nucleotide diversity (π) values were found in the populations of the Bothnian Sea and Baltic Sea (0.00131-0.00267), the highest values were found in the populations of the North Sea and Kattegat (0.00487-0.00719). The value for population

DOE (Doel, Scheldt estuary) was much lower than the values for the other samples in the North Sea (Table 4.8). A significant deviation from expected HWE was only indicated for populations BEN (Bergen, northern North Sea) and BAH (Balgzand, southern North Sea), with an excess of haplogroup heterozygotes.

Table 4.8 Diversity indices for the 15 samples of *Pomatoschistus minutus* in the Baltic-North Sea region based on haplotype data of the rhodopsin gene. N, number of individuals surveyed; Sn, number of non-synonymous segregating sites; Ss, number of synonymous segregating sites; k, mean number of pairwise differences; л, average number of nucleotide differences per site; θ, theta value per site. For site abbreviations see Table 4.7.

		No. of haplogroups							
Population	N	Total	Private	Sn	Ss	k	π	Θ	F_{IS}
NOS	28	6	1	3	1	0.719	0.001 ± 0.0003	0.002	-0.087
TBS	20	8	2	5	3	1.463	0.003 ± 0.0004	0.003	0.050
HBS	26	5	1	3	2	0.864	0.002 ± 0.0002	0.002	-0.004
PBS	10	4	1	4	2	0.895	0.002 ± 0.0006	0.003	0.265
WCS	30	10	1	5	5	2.842	0.005 ± 0.0004	0.004	-0.009
BEN	20	10	3	4	5	2.745	0.005 ± 0.0006	0.003	-0.277*
NNS	36	6	1	5	6	3.642	0.007 ± 0.0006	0.004	0.009
BAH	40	6	0	4	5	3.155	0.006 ± 0.0007	0.003	-0.265*
DOM	42	9	1	5	7	3.939	0.007 ± 0.0005	0.004	0.109
OSO	42	7	0	4	5	3.826	0.007 ± 0.0005	0.003	-0.086
BOR	43	8	1	4	6	3.818	0.007 ± 0.0005	0.004	0.011
DOE	39	8	3	4	8	2.670	0.005 ± 0.0008	0.004	0.077
OOS	42	7	0	5	5	3.781	0.007 ± 0.0005	0.004	0.038
NBF	39	5	0	4	5	3.716	0.007 ± 0.0006	0.003	0.092
DEP	34	3	0	4	5	3.604	0.007 ± 0.0007	0.003	0.022

Genetic differentiation

The substitution frequency on the five polymorphic AAs each sample showed the highest differentiation between Baltic and North Sea samples (Table 4.9). Within the North Sea and Kattegat, no differentiation was found based on the frequencies of the AA-substitutions between samples. However, the most diverged sample in the North Sea was DOE (Doel, Scheldt estuary). This sample had a much higher frequency for alanine on AA214 and on AA299 in comparison with the other samples from the North Sea (Table 4.9). Pearson's chi-square tests in STATISTICA confirmed this trend on both AA sites between the sample DOE and the other North Sea samples ($p<0.05$).

Table 4.9 Frequency of the amino acid substitutions detected at the rhodopsin gene of sand gobies from 15 sampling sites. The highest frequency in a sampling site is given in bold for each amino acid. For site abbreviations see Table 4.7.

	Population NOS	TBS	HBS	PBS	WCS	BEN	NNS	BAH	DOM	OSO	BOR	DOE	OOS	NBF	DEP
AA151 (or SNP4)															
Asn	0.077	0.125	0.019	0.250	**0.875**	**0.675**	**0.917**	**0.936**	**0.842**	**0.905**	**0.833**	**0.885**	**0.881**	**0.842**	**0.845**
Thr	**0.923**	**0.875**	**0.981**	**0.750**	0.125	0.325	0.083	0.064	0.158	0.095	0.017	0.115	0.119	0.158	0.155
AA214 (or SNP9_10_11)															
Ala	**1**	**0.975**	**1**	**0.950**	**0.714**	**0.700**	**0.694**	**0.750**	**0.655**	**0.655**	**0.674**	**0.821**	**0.655**	**0.692**	**0.697**
Ile	0	0.025	0	0.050	0.286	0.300	0.306	0.250	0.345	0.345	0.326	0.179	0.345	0.308	0.303
AA217 (or SNP 12)															
Ile	**0.893**	**0.625**	**0.923**	**0.950**	**0.965**	**1**	**0.986**	**1**	**0.988**	**1**	**1**	**1**	**0.988**	**1**	**1**
Thr	0.107	0.375	0.077	0.050	0.035	0	0.014	0	0.012	0	0	0	0.012	0	0
AA261 (or SNP 14)															
Phe	**0.893**	0.450	**0.596**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Tyr	0.107	**0.550**	0.404	0	0	0	0	0	0	0	0	0	0	0	0
AA299 (or SNP 19)															
Ala	**1**	**1**	**1**	**1**	**0.672**	**0.675**	**0.700**	**0.756**	**0.707**	**0.647**	**0.683**	**0.820**	**0.655**	**0.684**	**0.697**
Ser	0	0	0	0	0.328	0.325	0.300	0.245	0.293	0.354	0.317	0.180	0.345	0.316	0.300

Four statistical methods were used to describe the population differentiation at *RH1*. First, the pairwise F_{ST}-values revealed clustering in the three groups; the Baltic group with all samples of the Baltic and Bothnian Sea, the northern North Sea group with the samples of Bergen (BER) and Kattegat (WCS), and the southern North Sea group with all samples from Belgium and The Netherlands (Table 4.10). No pairwise F_{ST}-values were significant for samples within the southern North Sea, except for several pairs including sample DOE (Doel, Scheldt estuary) (Table 4.10). Remarkably, the pairwise F_{ST}-values based on the non-synonymous SNPs were differentiated from those based on the synonymous SNPs of the *RH1* gene fragment (Table 4.11). The F_{ST}-values between Baltic and North Sea samples were twice as high based on non-synonymous SNPs as for synonymous SNPs. Also the values between Baltic samples were only significant based on non-synonymous SNPs. This contrasted with the F_{ST}-values between samples from the northern North Sea and southern North Sea, which were only significant based on synonymous SNPs (Table 4.11). Second, the NMDS plots based on D_{CE} distances revealed the same clustering of samples in three groups as did the pairwise F_{ST}-values (Fig. 4.5a). The NMDS analysis confirmed also that the group with the Baltic and Bothnian samples is more variable than the group with all samples from the southern North Sea (Fig. 4.5a). NMDS analysis of the samples of the southern North Sea showed no clustering, but appointed the sample DOE (Doel, river Scheldt) as the most diversified population (Fig. 4.5b). Both NMDS plots had a stress value below 0.20 suggesting interpretable information concerning intersite relationships. Third, the impact of the observed structuring in the NMDS on the genetic structure was tested using a hierarchical analysis of molecular variance. The clustering of samples in the three groups of the NMDS plot for all samples revealed that 25.24% of the variation in the dataset was distributed among groups. No significant results were found for other AMOVA-tests based on *a priori* knowledge about the neutral genetic population structure and population dynamics of populations from the southern North Sea. Fourth, the global Mantel test revealed a significant isolation by distance (IBD) pattern with D_{CE} ($r = 0.916$, $p<0.05$). No significant IBD was found when only southern North Sea populations were analysed.

Finally, population genetic parameters were estimated separately for males and females. No parameter was significantly different between the sexes suggesting no difference in inheritance of the *RH1* variation or sex-biased dispersal.

Table 4.10 Pairwise F_{ST} estimates based on the rhodopsin gene (above diagonal) and pairwise geographical distances in km (below diagonal) between sand goby populations. For site abbreviations see Table 4.7. *Significant (p <0.05); **remains significant after Bonferroni correction (Rice 1989).

	NOS	TBS	HBS	PBS	WCS	BEN	NNS	BAH	DOM	OSO	BOR	DOE	OOS	NBF	DEP
NOS	-	0.206**	0.094**	0.033	0.431**	0.344**	0.402**	0.399**	0.370**	0.408**	0.362**	0.363**	0.391**	0.369**	0.384**
TBS	606	-	0.084**	0.244**	0.423**	0.360**	0.400**	0.407**	0.377**	0.408**	0.373**	0.385**	0.393**	0.376**	0.386**
HBS	672	326	-	0.191**	0.471**	0.392**	0.431**	0.435**	0.401**	0.435**	0.395**	0.408**	0.420**	0.401**	0.416**
PBS	1220	689	1073	-	0.282**	0.197**	0.275**	0.273**	0.254**	0.292**	0.246**	0.226**	0.274**	0.245**	0.257**
WCS	1770	1534	1144	1058	-	0.007	0.075**	0.060**	0.089**	0.100**	0.075**	0.046**	0.084**	0.072*	0.078**
BEN	2421	2099	1828	1656	592	-	0.089*	0.079**	0.092**	0.110**	0.077**	0.056**	0.093**	0.074*	0.081**
NNS	2731	2460	2118	1900	912	1091	-	-0.003	-0.009	-0.009	-0.010	0.025	-0.010	-0.012	-0.012
BAH	2668	2394	2043	1843	877	1068	10	-	0.009	0.013	0.005	0.002	0.009	-0.001	0.001
DOM	2845	2518	2197	2017	989	1266	194	194	-	-0.009	-0.011	0.041*	-0.011	-0.012	-0.011
OSO	2791	2497	2162	1988	1059	1293	198	203	38	-	-0.008	0.051**	-0.010	-0.009	-0.008
BOR	2876	2562	2305	2104	1118	1318	213	225	34	70	-	0.032	-0.010	-0.013	-0.013
DOE	2897	2594	2224	2083	1132	1313	269	277	85	122	53	-	0.043*	0.025	0.026
OOS	2900	2664	2335	2200	1124	1349	239	255	63	100	62	113	-	-0.011	-0.010
NBF	3037	2701	2377	2246	1147	1391	266	281	87	122	82	132	20	-	-0.015
DEP	2969	2190	2346	2704	1187	1370	268	280	88	125	89	138	26	6	-

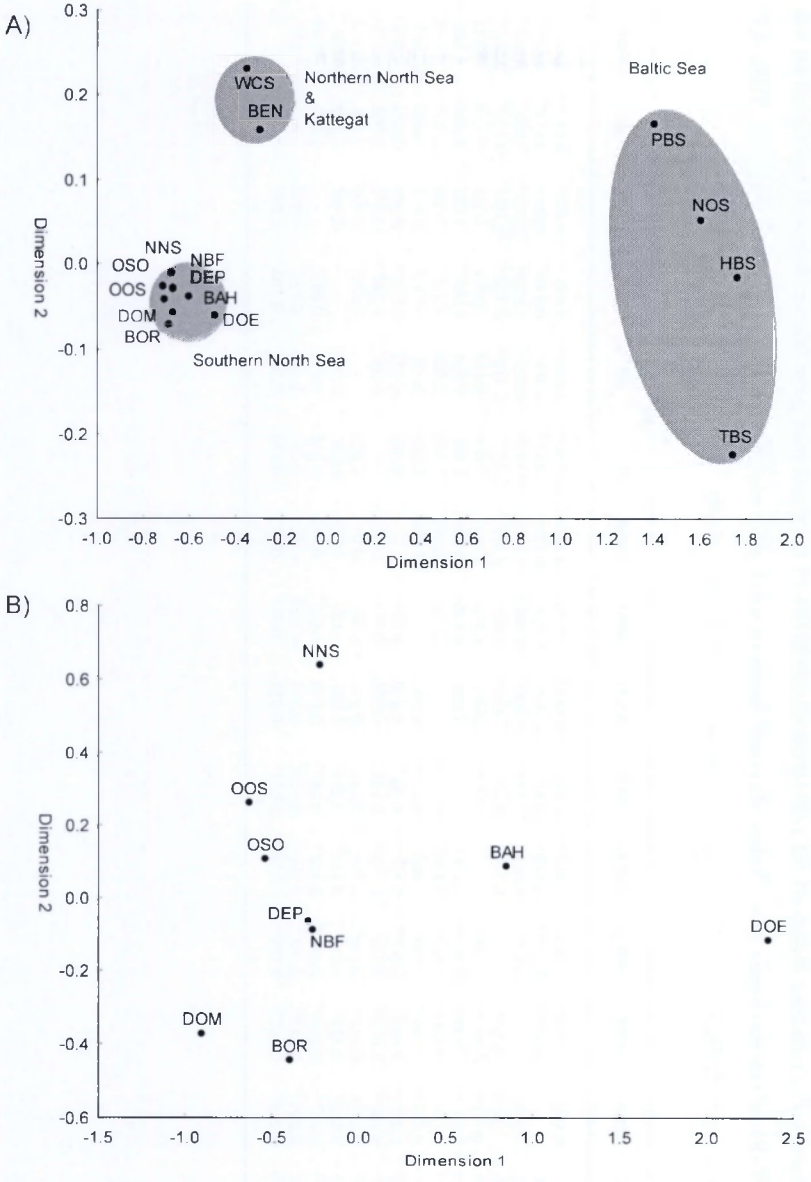


Figure 4.5 Non-metric multidimensional scaling (NMDS) plots of sand goby populations based on genetic distances (Cavalli-Sforza & Edwards 1967); (a) entire rhodopsin dataset of *Pomatoschistus minutus* from the Baltic and North sea, (b) rhodopsin dataset of the southern North Sea group. For site abbreviations see Table 4.7.

Table 4.11 Pairwise F_{ST} estimates based on the non-synonymous SNPs of the rhodopsin gene (above diagonal) and based on the synonymous SNPs of the rhodopsin gene (below diagonal) between sand goby populations. For site abbreviations see Table 4.7. *Significant (p <0.05); **remains significant after Bonferroni correction (Rice 1989).

	NOS	TBS	HBS	PBS	WCS	BEN	NNS	BAH	DOM	OSO	BOR	DOE	OOS	NBF	DEP
NOS	-	0.230**	0.120**	0.049*	0.497**	0.394**	0.524**	0.545**	0.465**	0.520**	0.463**	0.513**	0.496**	0.470**	0.470**
TBS	-0.004	-	0.096**	0.263**	0.475**	0.399**	0.508**	0.534**	0.471**	0.511**	0.472**	0.515**	0.492**	0.473**	0.475**
HBS	-0.020	0.004	-	0.231**	0.543**	0.449**	0.564**	0.586**	0.513**	0.559**	0.512**	0.563**	0.539**	0.518**	0.527**
PBS	-0.045	-0.026	-0.027	-	0.334**	0.227**	0.375**	0.401**	0.313**	0.377**	0.312**	0.358**	0.351**	0.315**	0.322**
WCS	0.183**	0.165**	0.173**	0.132**	-	0.016	-0.012	0.001	-0.010	0.009	-0.012	0.021	-0.011	-0.015	-0.013
BEN	0.149**	0.133**	0.139**	0.101**	-0.011	-	0.035	0.059	0.008	0.031	0.004	0.062**	0.020	0.005	0.009
NNS	0.253**	0.229**	0.246**	0.188**	0.139**	0.131**	-	0.004	-0.007	-0.009	-0.006	0.021	-0.009	0.009	-0.009
BAH	0.199**	0.181**	0.194**	0.149**	0.105**	0.095**	-0.002	-	0.008	0.012	0.010	0.003	0.011	0.005	0.004
DOM	0.270**	0.248**	0.264**	0.207**	0.159**	0.152**	-0.010	0.001	-	-0.006	-0.012	0.029**	-0.009	-0.012	0.012
OSO	0.283**	0.260**	0.276**	0.220**	0.173**	0.166**	-0.009	0.013	-0.012	-	0.006	0.046*	-0.012	0.008	0.006
BOR	0.249**	0.229**	0.243**	0.190**	0.138**	0.134**	-0.012	0.002	-0.011	-0.009	-	0.030*	-0.009	0.014	-0.013
DOE	0.132**	0.116**	0.128**	0.087**	0.068*	0.051*	0.028	0.002	0.048	0.054*	0.033*	-	0.046*	0.023	0.020*
OOS	0.269**	0.247**	0.263**	0.207**	0.153**	0.147**	-0.010	0.007	-0.012	-0.010	-0.011	0.044*	-	0.010	-0.009
NBF	0.249**	0.226**	0.242**	0.182**	0.137**	0.129**	-0.014	-0.004	-0.011	-0.010	-0.013	0.027	-0.012	-	0.015
DEP	0.269**	0.244**	0.261**	0.199**	0.148**	0.141**	-0.014	-0.001	-0.011	-0.010	-0.013	0.030	-0.011	-0.015	-

Discussion

Differentiation at the RH1 gene between the Baltic Sea and the North Sea

The highest degree of differentiation at the *RH1* gene was found between the sand gobies of the North Sea and Baltic Sea. This differentiation was mainly linked to the functional variation of *RH1* (Table 4.11), which points to directional selection towards the light regime of the local marine system. Here, random processes do not play a key role. The Baltic environment is characterized by an on average higher value of 'wavelength of the maximally transmitted light' (WMTL) in comparison with the North Sea (Jerlov 1976). The distribution of the variation on the spectral tuning sites of the *RH1* gene of the sand gobies matches with these environmental differences. All individuals from the North Sea were fixed for the blue-shifted mutation at AA261 (Y261F) and all individuals of the Baltic Sea were fixed for the red-shifted allele at AA299 (S299A) (Table 4.9). Differentiation between the samples of the two systems was also observed at other polymorphic AAs, but the effect of AA-substitutions awaits testing on rod opsins, for example by mutagenic experiments. For example AA214 is known to function as a spectral tuning site on cone opsins (Asenjo *et al.* 1994); it is almost fixed for alanine in the samples of the Baltic Sea in contrast to the North Sea samples (Table 4.9).

Our broad sampling coverage in time and space in the Baltic-North Sea region showed a convincing association between *RH1* variation and the global light environment. The macro-scale differentiation between the Baltic and North Sea populations is a permanent feature, regardless of the micro-scale and temporal variation of the local light regime. It remains stable for generations given the short life span of the sand goby (Fonds 1973; Hamerlynck 1990). Therefore, the 'snapshot' sampling design of *subchapter 4a* did not constrain the spatial analysis of the rhodopsin variation.

Most marine organisms, including gobies, immigrated into the Baltic Sea during the last marine (Littorina) phase, starting some 8000 years ago and gradually evolving into the present-day brackish water condition (Björck 1995). Despite this short geological history, populations inhabiting the estuarine Baltic Sea have evolved substantially different from populations of the fully marine North Sea e.g. in herring *Clupea harengus* (Jørgensen *et al.*

2005), turbot *Scophthalmus maximus* (Nielsen *et al.* 2004) and green alga *Cladophora rupestris* (Johansson *et al.* 2003). The genetic differences are most likely as a consequence of isolation and bottlenecks, as well as selection on adaptive traits (Johannesson & André 2006). The short period of transition suggests that all these organisms adapted to the changing environmental conditions by selection on pre-existing genetic variation instead of new mutations. Standing variation leads to more rapid evolution in novel environments because it is available immediately, whereas more time is required for a new beneficial mutation to arise (Barrett & Schluter 2008). All AAs of the *RH1* gene of the sand goby, except AA261, were polymorphic in the North Sea, suggesting indeed a shift in adaptation from North Sea to Baltic Sea light conditions by standing variation. Remarkably mutation F261Y, which is known to have a strong phenotypic effect on the λ_{max} of the rod opsins (Yokoyama *et al.* 1995), is only observed in the northern Baltic and Bothnian Sea. Therefore the scenario that might explain the origin of this important functional mutation, either *de novo* or standing variation, remains unknown (Barrett & Schluter 2008). The total absence of Phe261Tyr in all populations of the NA-Group outside the northern Baltic may argue for a *de novo* mutation. Among all 834 haplotypes from the southern Baltic Sea, Kattegat and the North Sea, the allele with a red-shifted effect on the λ_{max} of rod opsins was not observed, although sampling included regions with a high local WMTL value. Nevertheless, the mutation might be present in the source population at a frequency below the detection threshold (< 0.12%). Such occurrence of allele sorting is not at all impossible. For example surveys of large numbers of three-spined stickleback *Gasterosteus aculeatus* belonging to the ancestral and fully marine ecotype showed that low-plated alleles of the *Eda* gene are present at comparably modest frequencies (circa 0.2%). The low value is sufficient to introduce repeatedly the allele into freshwater stickleback populations by marine founders (Colosimo *et al.* 2005). Finally, the F261Y mutation can also be introduced through hybridization and backcrossing with other species (Gibson & Dworkin 2004). *P. minutus* hybridizes with *P. lozanoi* (Wallis & Beardmore 1980) but there are no observations for such specific substitution in *P. lozanoi* (*chapter 5*).

Differentiation at the RH1 *gene within the Baltic Sea and North Sea*

The sand goby samples of the Baltic Sea were significantly differentiated from each other based at the *RH1* variation (Table 4.10). However, the F_{ST}-values were only significant for the six non-synonymous SNPs (Table 4.11). This indicates that the genetic differentiation between the Baltic samples was most likely the result of adaptation to the local spectral environment and not of random processes. The photic regime of the complete Baltic and Bothnian Sea has a stable temporal pattern based on WMTL values (Lindström 2000). This robust pattern should provide an opportunity for local populations to adapt to their environment; it has been suggested to explain the observed differentiation in absorbance of VP opsins between several Baltic Sea samples of *Mysis relicta* (Jokela-Määttä *et al.* 2005). The robust pattern observed in the Baltic strongly contrasts with the unstable pattern in WMTL along the North Sea coast (*subchapter 4a*). No differentiation was observed between the North Sea samples based on functional variation at the *RH1* gene (Table 4.11).

The frequency of the various alleles at the polymorphic AAs was relatively stable across the North Sea. In contrast to the lack of functional differentiation, two groups were significantly differentiated from each other based on the synonymous SNPs of *RH1* (F_{ST}-value = ±0.10; Table 4.11); a northern North Sea group with the samples from Bergen and Kattegat, and a southern North Sea group with all Belgian and Dutch samples. Also neutral markers have been shown to differentiate between Kattegat and the Southern North Sea (*subchapter 3a*). Moreover, within the Southern Bight of the North Sea, two breeding units were observed based on microsatellite and catchment data (Pampoulie *et al.* 2004a; Vanden Eede 2006; Guelinckx 2008). Therefore, the lack of differentiation on functional variation of the *RH1* gene between several populations is an indication for stabilizing selection. It has already been observed for other vertebrates in several regions of the genome, e.g. reproductive proteins and the Major Histocompatibility complex (Alcaide *et al.* 2008; Turner & Hoekstra 2008; Oliver *et al.* 2009). The results show similar selective pressures shaping the non-synonymous *RH1* variation in the whole North Sea, which fits with the similar light regimes at the North Sea locations sampled (*subchapter 4a*).

North Sea populations maintain stable frequencies of rod opsin phenotypes. This can be attributed to environmental heterogeneity in the light regime that characterizes the North Sea coastline (*subchapter 4a*). On the other hand, polymorphism at *RH1* may also be the product of the heterogeneity in habitats that sand gobies encounter during their life history. Like many marine fish, juvenile sand gobies grow up in the turbid waters of the North Atlantic estuaries and coastal surf zones (Healey 1971; Fonds 1973; Maes *et al.* 1998). Although females seem to dominate in brackish water (Vanden Eede 2006), no significant difference was observed between the sexes in the genetic composition and population differentiation at the *RH1* gene. However, the estuary (Scheldt estuary in Doel, sample DOE) was clearly differentiated from the open sea in the southern North Sea (Table 4.10, 4.11), suggesting an important role of rhodopsin in the estuary. In the estuarine mesohaline zone (Doel), there was a higher frequency of the red-shifted allele at AA299 and of the allel at AA214 fixed in sand goby populations in the brackish environments of the Baltic Sea and Mediterranean Sea. The trend for more red-shifted alleles was not observed for the estuarine polyhaline zone (BOR). However, the mesohaline is characterized by above average turbidities (Herman & Heip 1999). The trend was also not found in the Oosterschelde estuary (OSO), where turbidity is much lower than in the Scheldt estuary (Smaal & Nienhuis 1992).

The results based on *RH1* prove that the estuarine sand gobies are not just a random group of marine residents. Gobies with red-shift AA-substitutions at the *RH1* gene may prefer to migrate deeper into the estuary, which enjoys a more reddish light environment in comparison with the open sea. Based on the isotope composition and dynamics of muscle and liver tissue (Guelinckx *et al.* 2008) and on the chemical composition of otoliths (Guelinckx 2008), the sand goby displays a facultative use of the estuary. There seems to be no obligate estuarine stage but a highly individual movement pattern in estuarine habitat use. It is supported by the simultaneous presence of sand goby in the estuary and adjacent marine habitats during the various life stages (Hostens 2000; Hostens 2003). It was thought that the sand goby exploits the estuary rather opportunistically when estuarine conditions (e.g. temperature and dissolved oxygen concentration) are suitable. Hence, individuals may respond quickly to changes in climate conditions, food availability or predation risk and shift rapidly between coastal and estuarine nursery areas in order to increase their individual state

and fitness (Guelinckx *et al.* 2008). Our results provide strong support to the hypothesis that turbidity shapes estuarine habitat use by marine fish (Benaka 1999; Maes *et al.* 2005).

Conclusion

Local visual adaptation of *P. minutus* gobies differs between the Baltic and North Sea, despite the micro-scale and temporal variation in the light regime within these marine systems. In the North Sea region, selection at the *RH1* gene was balanced, since no differentiation on functional variation was found between samples, although genetic and catchment data suggested different populations within the region. The high level of polymorphism at the *RH1* gene could be maintained either by the high heterogeneity of the light regime along the coastline of the North Sea, or by the specific migration patterns of juvenile sand gobies between the open sea and the estuary. The tendency that estuarine migrants have *RH1* variants associated with brackish water illustrates the importance of turbidity and visual capacity. Future research might focus on turbidity gradients along an estuary in combination with genetic and functional analysis of vision. It may clarify whether there is a true link between visual capacity and the so much discussed differential estuarine use by sand goby and so many other fishes.

Acknowledgements

We thank the following persons who generously provided samples: Anne Christine Utne-Palm (Universitetet i Bergen, Norway), Carin Magnhagen (Swedish University of Agricultural Science, Sweden), Maria Järvi-Laturi (University of Helsinki, Finland), Kai Lindström (Åbo Akademi University, Finland), Lotta Kvarnemo (Göteborgs Universitet, Sweden), Katarzyna Waligóra-Borek (University of Gdańsk, Poland), Henk van der Veer (NIOZ, The Netherlands), Tine Huyse and Els Cuveliers (K.U.Leuven, Belgium), and Jef Guelinckx (Flemish Goby Consortium, Belgium). The first author received a PhD-fellowship of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This study was financed by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (WESTBANKS project, contract nr SD/BN/01A).

Appendix of Subchapter 4b

Appendix S4.3. Supplementary tables

Table S4.3 Nucleotide polymorphisms at the rhodopsin gene in sand goby of 15 sampling sites. Dots indicate homology with the reference sequence (Acc. nr. X62405). Amino acid numbers are listed for the non-synonymous mutations, which are listed in bold. For site abbreviations see Table 4.7.

SNP nr.	**4**	5	6	7	**9**	**10**	11	**9_10_11**	**12**	13	13b	13t	13q	**14**	16	17	17b	17t	18	**19**
Nucleotide position	**296**	297	339	357	**484**	**485**	486	**484-486**	**494**	525	528	531	564	**626**	669	684	696	702	738	**739**
Amino acid nr.	**151**	-	-	-	-	-	-	**214**	**217**	-	-	-	-	**261**	-	-	-	-	-	**299**
Reference	**A**	C	T	T	**G**	**C**	T	**GCT**	**T**	T	C	C	C	**T**	C	A	G	C	C	**G**
NOS	**./C**	.	.	.	.	.	.	.	**./C**	.	.	.	.	**./A**	./T	.	.	.	.	.
TBS	**./C**	./A	.	.	**./A**	**./T**	./C	**./ATC**	**./C**	.	.	.	.	**./A**	./T	.	.	.	.	.
HBS	**./C**	.	.	.	.	.	.	.	**./C**	.	./T	.	.	**./A**	./T	.	.	.	.	.
PBS	**./C**	.	.	.	**./A**	**./T**	./C	**./ATC**	**./C**	.	.	.	.	.	./T	.	.	.	.	.
WCS	**./C**	.	./A	.	**./A**	**./T**	./C	**./ATC**	**./C**	./G	.	.	.	.	./T	./C	.	.	.	**./T**
BEN	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	.	./G	.	.	.	.	.	./C	.	.	.	**./T**
NNS	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	**./C**	./G	.	.	./G	.	.	./C	.	.	.	**./T**
BAH	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	.	./G	.	.	.	.	.	./C	.	.	.	**./T**
DOM	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	**./C**	./G	.	.	.	.	./T	./C	.	.	./T	**./T**
OSO	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	.	./G	.	.	.	.	.	./C	.	.	.	**./T**
BOR	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	.	./G	.	.	.	.	./T	./C	.	.	.	**./T**
DOE	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	.	./G	.	./T	.	.	.	./C	./A	./A	.	**./T**
OOS	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	**./C**	./G	.	.	.	.	.	./C	.	.	.	**./T**
NBF	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	.	./G	.	.	.	.	.	./C	.	.	.	**./T**
DEP	**./C**	.	./A	./A	**./A**	**./T**	./C	**./ATC**	.	./G	.	.	.	.	.	./C	.	.	.	**./T**

Table S4.4 Rhodopsin haplogroups and their geographical distribution in *Pomatoschistus minutus*. Dots indicate homology with the rhodopsin reference of sand goby (Acc. nr. X62405). Non-synonymous SNPs are listed in bold. For site abbreviations see Table 4.7.

	SNP nr. and mutation																	Haplogroup tally within populations													
Haplogroup nr.	4	5	6	7	9_10_11	12	13	13b	13t	13q	14	16	17	17b	17t	18	19	NOS	TBS	HBS	PBS	WCS	BEN	NNS	BAH	DOM	OSO	BOR	DOE	OOS	NBF	DEP
Reference	A	C	T	T	GCT	T	T	C	C	C	T	C	A	G	C	C	G	NOS	TBS	HBS	PBS	WCS	BEN	NNS	BAH	DOM	OSO	BOR	DOE	OOS	NBF	DEP
HG 1	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	41	10	24	13	13	9	5	5	12	8	14	9	9	12	9
HG 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	4	1	5	24	16	43	53	42	45	43	52	46	41	39
HG 3	C	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	3	1	4	1	1										
HG 4	C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	2	9	22												
HG 5	C	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	3	13													
HG 6	C	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	1														
HG 7	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.		1			1		1		1				1		
HG 8	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		1													
HG 9	.	.	.	.	ATC	.	.	.	.	.	.	.	.	.	.	.	.		1													
HG 10	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.			1												
HG 11	C	.	.	.	ATC	.	.	.	.	.	.	.	.	.	.	.	.				1											
HG 12	.	.	.	.	ATC	.	G	.	.	.	.	.	C	.	.	.	T					2	2			1		1		2		
HG 13	.	.	.	.	ATC	.	G	.	.	.	.	.	.	.	.	.	T					12	3									
HG 14	.	.	J	.	ATC	.	G	.	.	.	.	.	.	.	.	.	T					2	2									
HG 15	.	.	.	.	ATC	.	.	.	.	.	.	.	.	.	.	.	T					3	3									
HG 16	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.					1	1									
HG 17	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	T					1										
HG 18	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T						2									
HG 19	C	.	.	.	ATC	.	.	.	.	.	.	.	.	.	.	.	T						1									
HG 20	.	.	.	.	ATC	.	G	.	.	.	.	.	.	.	.	.	.						1									
HG 21	.	.	A	A	ATC	.	G	.	.	.	.	.	C	.	.	.	T							21	18	22	26	23	12	24	23	20
HG 22	.	.	A	A	ATC	.	G	.	.	.	.	.	.	.	.	.	T							1		1	2	2				
HG 23	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.							1								
HG 24	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.								2		1				1	
HG 25	.	.	A	A	ATC	.	.	.	.	.	.	.	C	.	.	.	T								1		1		1			
HG 26	.	.	.	A	ATC	.	G	.	.	.	.	.	C	.	.	.	T								1					1	1	
HG 27	.	.	A	A	ATC	.	G	.	.	.	.	.	C	.	.	.	.									3						
HG 28	.	.	A	A	ATC	.	G	.	.	.	.	T	C	.	.	.	T									1		1				
HG 29	.	.	A	A	ATC	.	G	.	.	.	.	.	C	.	.	T	.									1						
HG 30	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.										1		1			
HG 31	.	.	A	.	ATC	.	G	.	.	.	.	.	C	.	.	.	T											1		1		
HG 32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T											1				
HG 33	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.												1			
HG 34	.	.	A	A	ATC	.	G	.	.	.	.	.	C	A	.	.	T												1			
HG 35	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.												1			

Chapter 5

Insights in the adaptive evolution of the rhodopsin gene in the 'sand goby' group (Teleostei, Gobiidae)

'The eye, which is the window of the soul, is the chief organ whereby the understanding can have the most complete and magnificent view of the infinite works of nature; and the ear comes second, which acquires dignity by hearing things the eye has seen.'

Leonardo Da Vinci (Italian polymath, 1452-1519).

INSIGHTS IN THE ADAPTIVE EVOLUTION OF THE RHODOPSIN GENE IN THE 'SAND GOBY' GROUP (TELEOSTEI, GOBIIDAE)

Maarten H.D. Larmuseau[1], Tine Huyse[1], Kim Vancampenhout[1], Jeroen K.J. Van Houdt[1,2] & Filip A.M. Volckaert[1]

[1]Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Leuven, Belgium.

[2]Royal Museum for Central Africa, Tervuren, Belgium.

Submitted manuscript

Abstract

The spectral tuning mechanism of visual pigments is an excellent model to elucidate the mechanisms of adaptive evolution and the importance of selection as evolutionary force. This mechanism can identify amino-acid changes which are responsible for the adaptation of organisms to specific environments. The aim of the study was to perform a phylogenetic analysis to determine whether there is evidence for differential adaptive molecular evolution on the rhodopsin (*RH1*) gene among closely related 'sand goby' species (Teleostei, Gobiidae) that inhabit different photic environments. Fragments of the *RH1* gene (868 bp) were sequenced and analysed for nine 'sand goby' species. A high level of interspecific polymorphism at the *RH1* gene was observed, including non-synonymous mutations on amino acids known as spectral tuning sites. Moreover, a link was found between some of the amino acid (AA) substitutions and the spectral conditions of the habitat environment of the species in terms of salinity and depth. The phylogeny based on the *RH1* fragments was compared with a 'neutral phylogeny' using partial 12S and 16S mtDNA (784 bp). 'Sand goby' species clustered together based on *RH1* sequences due to a similar habitat choice although they are not related in the 'neutral' phylogeny. Finally, analyses of d_N/d_S substitution rate ratios and a likelihood ratio test under site-specific models detected significant signal of positive Darwinian selection on specific AA of the *RH1* gene. This study identified a

number of promising candidate tuning sites for future study by site-directed mutagenesis and illustrates the molecular evolutionary dynamics of gobiid visual sensitivity and its relationship to the photic environment.

Keywords: Adaptation, gobies, opsin genes, phylogeny, photic environment, photoreceptor

Introduction

Adaptive evolution in vertebrates has been extremely difficult to study, due to the scarcity of genetic systems in which the functional effects of mutations can be evaluated (Yokoyama 2002). One of the few excellent models to elucidate the mechanisms of adaptive evolution and the importance of selection as evolutionary force is the spectral tuning mechanism of visual pigments (VP). Visual pigments have a well-defined role in nature as they allow the organism to detect differences in the spectral composition of the environment (Bowmaker 2008). Therefore they have a strong and direct effect on the evolution of organisms, providing an excellent system to study adaptive evolution at the molecular level.

Vertebrate VP molecules are bound in dense membrane stacks in retinal photoreceptors to mediate vision. The VP consists of a protein moiety, the opsin, bound to a light-absorbing molecule, the chromophore (Park *et al.* 2008). Each pigment shows a characteristic peak of maximal absorbance (λ_{max}), its precise value depending on the interactions between the chromophore and the opsin protein (Yokoyama 2000). Vertebrates have different possibilities to modify their visual systems (spectral tuning) to cope with their specific photic environments. These changes in absorbance maxima can be achieved at the physiological or DNA level. The first mechanism will exchange the chromophore (A1 or A2 pigments) (Bowmaker 1995) while at the DNA level amino acid (AA) substitutions in opsin proteins will occur (Hunt *et al.* 2001; Yokoyama & Takenaka 2004). From several site-directed mutagenic experiments, ca. 25 AA are known to be involved in the spectral tuning of the visual pigments by causing significant shifts in absorbance maxima (Yokoyama *et al.* 2007 and references herein).

A clear correlation between the λ_{max} of the VPs and the photic characteristics of the habitat location of their possessors is observed in many aquatic species. Fish are one of the best-studied species because of the diversity in clearly defined photic environments (Yokoyama 2000). These environments differ in turbidity and colour of the water, and the brightness of the downwelling light (Bowmaker 1995). Moreover, the intensity of downwelling light diminishes rapidly with increasing depth (Jerlov 1976). There is mounting evidence for a relationship between genetic variation at the VP opsins and the differences in environmental light conditions in several aquatic vertebrates, fish and marine mammals (Fasick & Robinson 1998; Hunt *et al.* 2001; Sugawara *et al.* 2005). The importance of this genetic spectral tuning in aquatic organisms has been emphasized recently by studies demonstrating positive selection on opsin genes in cichlids and salmonids (Sugawara *et al.* 2002; Dann *et al.* 2004; Spady *et al.* 2005) and by a recent study that showed the role of divergent selection on sensory genes in promoting speciation through sensory drive in cichlid fish (Seehausen *et al.* 2008). Phylogenetic analyses provide the ideal tool to understand and illustrate the role of adaptation on opsin genes in the evolution of aquatic taxonomic groups.

A good taxonomic group to study the molecular evolution of opsin genes in a phylogenetic framework is the 'sand goby' group (Teleostei, Gobiidae). Among the Eastern Atlantic-Mediterranean Gobioid fish, the monophyletic 'sand goby' group consists of four paraphyletic genera and 30 species have been described: 12 *Pomatoschistus* (Gill) species, one *Gobiusculus* (Duncker) species, 15 *Knipowitschia* (Ljin) species and two *Economidichthys* (Bianco, Bullock, Miller and Roubal) species (Miller 1986). Those species are differentiated in habitat preference, salinity tolerance (marine, euryhaline, freshwater), substrate (sand, mud, rocks) and depth (demersal, pelagic) (Miller 1986). All 'sand goby' species are visual feeders (Aarnio & Bonsdorff 1993; Utne-Palm 2002; Jackson & Rundle 2008), with some of them known to be nocturnal (Gibson & Ezzi 1981; Ehrenberg & Ejdung 2008). Therefore, the best candidate to study the evolution of an opsin gene in the 'sand goby' group is the rhodopsin gene (*RH1*). The *RH1* gene encodes the rod opsin protein, which mediates vision in very dim light and is located in rod cells.

The few data available on the λ_{max} values of the rod opsin for several 'sand goby' species show differences which are not linked to more than one rhodopsin gene. Moreover, the

measurements showed the presence of one single chromophore in the rod opsin (Jokela *et al.* 2003; Utne-Palm & Bowmaker 2006). Therefore, the differences in λ_{max} between the species indicate polymorphism in the protein part of the pigment instead of physiological adaptation. Moreover, evolutionary adaptation in the *RH1* gene has already been observed in *P. minutus*, one of the members of the 'sand goby' group (*subchapter 4a*). An unusual high level of intraspecific polymorphism at the *RH1* gene was detected and some AA known as spectral tuning sites were variable in *P. minutus*. The population differentiation at these tuning sites between seven sand goby populations across its distribution range was in agreement with the observed population differentiation in λ_{max} values of the rod opsins (Jokela *et al.* 2003) and the specific photic conditions at the sampling locations (*subchapter 4a*).

The aim of the present study was to perform a phylogenetic analysis to determine whether there is evidence of differential adaptive molecular evolution in the *RH1* gene among closely related 'sand goby' species inhabiting different environments in terms of for example salinity, depth, turbidity and substrate.

Material & Methods

Sampling and species identification

In total, 62 individuals belonging to eight species and four genera were collected for this study (Table 5.1). Samples were taken either by fyke, hand net, beam trawling or using the cooling-water intake screens of the nuclear power plant of Doel for the samples of the river Scheldt. The gobies were identified morphologically according to Miller (1986). The determination of the two cryptic species *P. minutus* and *P. lozanoi* was guaranteed by means of extra morphological characteristics based on pigmentation patterns (Hamerlynck 1990) and genetical tools according to *subchapter 2a*. *P. minutus* from the Adriatic Sea was considered as a distinct species of the 'sand goby' group, following Huyse *et al.* (2004). Fish species, geographic distribution and type of habitat of the sampling locations are listed in Table 5.1.

Table 5.1 Goby taxa used, including collection size, number of specimens sequenced, GenBank Accession number, natural distribution range and type of habitat of the sampling location (Miller 1986; Huyse *et al.* 2004; Kottelat & Freyhof 2007). N, sample size.

Species	Collection site	Country	N	Acc. Nr.	Distribution	Habitat
Pomatoschistus minutus (Pallas 1770)	-	-	-	FJ410451-488	Northeastern Atlantic; western Mediterranean Sea; Black Sea	1,2
P. minutus Adriatic	Venice	IT	4		Adriatic coast	1
P. microps (Kroyer 1838)	Vårtan	S	18		Northeastern Atlantic and western Mediterranean	2
	Scheldt River	B	3			2
	Dale	UK	2			2
	Roscoff	F	3			2
P. lozanoi (de Buen 1932)	Ostend	B	5		Northeastern Atlantic	1
P. pictus (Malm 1865)	Ostend	B	4		Northeastern Atlantic and Mediterranean Sea	1
P. canestrinii (Ninni 1883)	Chioggia	IT	2		Adriatic coast	2
Gobiusculus flavescens (Fabricius 1779)	Bergen	N	7		Northeastern Atlantic and Mediterranean Sea	3
Economidichthys pygmaeus (Holly 1929)	Acheloos River	G	6		Rivers and streams of western Greece	4
Knipowitschia milleri (Ahnelt & Bianco 1990)	Acheron River	G	8		Greece: lower stretch of River Acheron	4
Zosterisessor ophiocephalus (Pallas 1814)	-	-	-	Y18678	Mediterranean Sea, Black Sea and Sea of Azov	-
Gobius niger (Linnaeus 1758)	-	-	-	Y18675	Eastern Atlantic, Mediterranean Sea and Black Sea	-

B, Belgium; F, France; G, Greece; IT, Italy; S, Sweden; UK, United Kingdom
1, Demersal, marine; 2, Demersal, brackish, turbid; 3, Pelagic, marine; 4, Demersal, freshwater

Gene amplification and sequencing

Genomic DNA was extracted from fin clips stored in 100% ethanol using the NucleoSpin Extraction Kit (Machery-Nagel GmBH, Düren, Germany). An 868 bp fragment of the *RH1* gene was amplified in polymerase chain reactions (PCR) with the forward primer PminRh1F 5'-GCGCCTACATGTTCTTCCTT-3' (*subchapter 4a*) and the reverse primer Rh1039r 5'-TGCTTGTTCATGCAGATGTAGA-3' (Chen *et al.* 2003). The PCR reactions were carried out on a GeneAmp PCR System 2700 thermocycler (Applied Biosystems) in a total volume of 25 µl, containing 1 µl of genomic DNA, 1 X PCR buffer, 2.0 mM dNTPs, 0.8 µM of each primer, 2.0 mM $MgCl_2$, 0.5 U of Taq DNA polymerase (Silverstar, Eurogentec, Seraing, Belgium) and mQ-H_2O. The PCR profile was: 4 min at 94°C followed by 35 cycles of 30 s at 96°C, 30 s at 54°C and 1 min at 72°C; with a final 10 min extension at 72°C. To avoid contamination, aerosol barrier tips and different pipettes were used, and the PCR reactions were prepared in the pre-PCR section of the laboratory. Negative controls proved the lack of contamination. All PCR products were visualized on agarose gels with ethidium bromide. After purification with the 'GFX PCR DNA and Gel Band Purification kit' (GE Healthcare, Piscataway, NJ, USA), the PCR products were sequenced in both directions using the

BigDye Terminator v. 3.1 Cycle Sequencing Kit on an ABI 3130 automated capillary DNA sequencer (Applied Biosystems). Sequences of 702 bp (234 AA) were visually inspected and aligned to each other with SEQSCAPE v. 2.1 (Applied Biosystems). The 234 AA fragment under study represents 67% of the protein. All known 25 AA involved in the spectral tuning of the visual pigments are included in this gene fragment (Yokoyama *et al.* 2007 and references herein). Point mutations were automatically detected with the GAP4 subprogram embedded in the STADEN package (http://sourceforge.net/projects/staden) and re-inspected manually by eye. All rhodopsin sequences determined in this study were deposited in the GenBank database (accession numbers pending). The already 38 observed haplotypes of *P. minutus* were included in the statistical analysis (FJ410451-FJ410488) (*subchapter 4a*). In addition, partial 12S and 16S rRNA genes were amplified and sequenced for six specimens of *Economidichthys pygmaeus* and eight specimens of *Knipowitschia milleri*. Therefore, the same primers and protocol were used as described in Huyse *et al.* (2004). The already sequenced 12S and 16S rRNA fragments of the other species of the 'sand goby' group were included in the phylogenetic analysis (AJ616811-AJ616820, AJ616828-AJ61683) (Huyse *et al.* 2004).

Phylogenetic analysis

Phylogenetic tree reconstruction based on the *RH1* sequences was conducted using neighbour-joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) analysis. For species with more than one observed haplotype, only the two most divergent haplotypes were selected for the phylogenetic analysis. For *P. minutus*, only the two most frequent and the two most divergent haplotypes (RhP7, RhP11, RhP15 and RhP21; accession numbers FJ410457, FJ410461, FJ410465 and FJ414071 respectively) were included. Sequences of *Zosterisessor ophiocephalus* and *Gobius niger* were used as outgroup species for the phylogenetic analyses (Table 5.1). Only substitutions in residues that are located in the seven transmembrane helical regions appear important for the spectral tuning of the resulting pigment (Bowmaker 2008). Therefore, phylogenetic analyses were also performed on the dataset with only the codons inside the transmembrane regions ('helices' dataset). NJ trees were constructed in MEGA v. 4.0 (Tamura *et al.* 2007). Bootstrap analyses were performed with 10,000 replicates. MP was performed using PAUP* v. 4.0b10 (Swofford 2002) with the following heuristic search settings: 10^5 random taxon addition replicates followed by tree

bisection-reconnection (TBR) branch swapping and 1000 bootstrap replicates. For both datasets, Modeltest v. 3.7 (Posada & Crandall 1998) selected the HKY + Γ model as the most suitable model based on the Akaike information criterion (AIC). The ML analysis with 100 bootstrap replicates was performed using PhyML v. 2.4.4 (Guindon & Gascuel 2003) accessed through a web server (Guindon *et al.* 2005). Based on the observed phylogeny, the AA sequence of the ancestral pigment was inferred using a likelihood-based Bayesian method (Yang 1997) implemented in CODEML in PAML v. 4.2 (Yang 2007).

In addition, a phylogeny of the 'sand goby' group was constructed using the combined mitochondrial 12S and 16S rRNA dataset. Sequences of *Padogobius martensii* and *P. nigricans* were used as outgroup species for the phylogenetic analyses based on 12S and 16S data (Accession numbers: AF616828, AF616837, AF067261 & AF067270).

Tests of positive selection

Two types of analyses were performed to determine if positive selection was involved in the evolution of *RH1* gene in the 'sand goby' group. First, MEGA v. 4 (Tamura *et al.* 2007) was used to compare the relative abundance of synonymous and nonsynonymous substitutions between pairs of sequences of the 'total' and 'helices' dataset. The number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) were estimated using the Z-test implemented in MEGA. The variances of d_S and d_N were computed by bootstrap (10,000 replicates). With this information, the null hypothesis of neutral evolution (H0: $d_N = d_S$) *versus* the hypothesis of positive selection (H1: $d_N > d_S$) was tested using a Z-test: $Z = (d_N - d_S)/\text{sqrt}\,(\text{Var}(d_S) + \text{Var}(d_N))$. Secondly, the CODEML program of PAML was used. Two tests (among lineages and among sites) using two types of models ('branch-specific' models and 'site-specific' models) were performed to determine whether positive selection influenced the *RH1* variation. The 'branch-specific' models allow the d_N/d_S ratio (hereafter referred as ω ratio) to vary among branches in the phylogeny, and therefore are useful in detecting positive selection operating on a particular lineage. However, if different protein-coding regions within a gene experience different selective pressures, it is not optimal to consider the average ω ratio of an entire gene sequence. The level of selection also varies in different

amino acid positions along protein sequences, and several 'site-specific' models have been developed that account for ω ratio variation between particular codon sites (Nielsen & Yang 1998). Likelihood-ratio tests (LRT) were used to evaluate two-codon-based models of sequence evolution, as described by Yang and co-workers (Yang 2000; Yang *et al.* 2000b). For the test among lineages, the model M0, which assumes a single ω ratio for all nucleotide sites and branches of the phylogeny, was compared with a model that estimates two different ω ratios, one for the lineage of interest ('foreground lineage') and another for all the other lineages ('background lineages') (Yang 1998). The test among lineages was performed based on the 'total' as well on the 'helices' dataset. For the test among sites, parameters were estimated under two different models M7 (beta) and M8 (beta and ω) (Yang *et al.* 2000a). While recombination on the nuclear rhodopsin gene can potentially generate false-positives in the detection of positive selection, these models are more robust against the occurrence of recombination than the other models implemented in CODEML (Anisimova *et al.* 2003). Positively selected codons (ω>1 with p>95%) were identified through an empirical Bayesian approach (Yang *et al.* 2005).

Results

Identity of RH1 *gene sequence*

We used several approaches to guarantee that no other member of the opsin gene family than the *RH1* gene was co-amplified and analyzed. First, when designing primers for rhodopsin, sites were selected that differ among paralogous genes. Second, other opsin genes have introns, unlike the rhodopsin genes of bony fish (Bowmaker 1995). Third, the duplication event separating rhodopsin from other opsin genes occurred before the diversification of vertebrates (Yokoyama 2000). Therefore, if a paralogous opsin gene was sequenced by mistake, sequence alignment would have shown this extreme divergence. The alignment of the sequences was straightforward as there were no gaps and translation into AA did not show nonsense or stop codons. The sequences matched the general properties of the *P. minutus RH1* gene (X62405) (Archer *et al.* 1992).

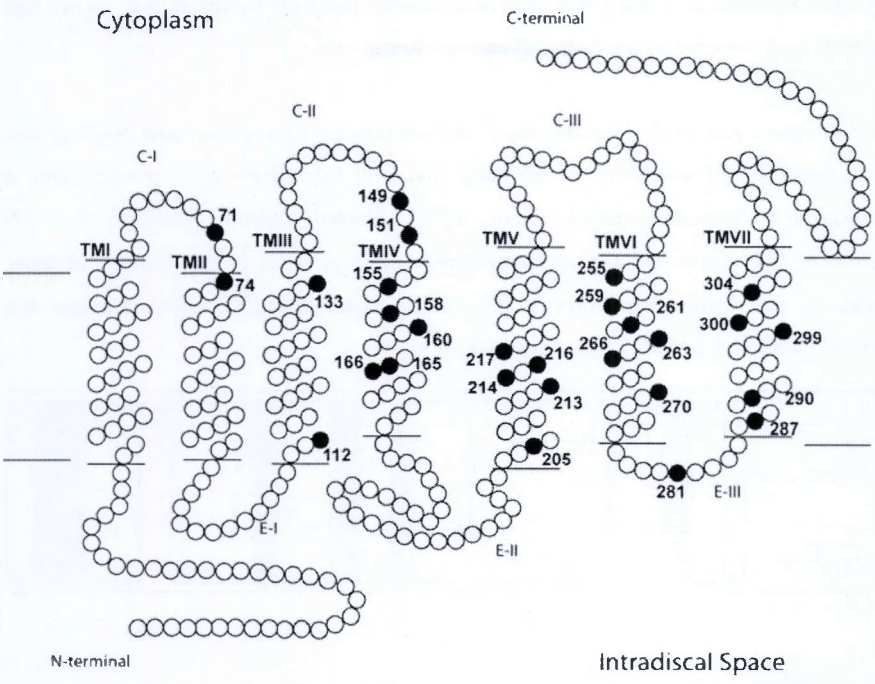


Figure 5.1 Two-dimensional model of the seven transmembrane α–helices of the rhodopsin protein (*RH1*) (Hargrave & McDowell 1992). The seven transmembrane helices (TM) are numbered, as well as the three loops at the cytoplasmic side (C) and the extracellular side (E) of the cell membrane. The 28 AA-substitutions found in the 'sand goby' group are shown with filled circles and numbered.

Polymorphism at the RH1 *gene*

In all available *RH1* sequences of the nine 'sand goby' species, 79 segregating sites or SNPs were noticed (11% of the total fragment). The alignment in AA showed 28 AA substitutions (12% of total AA), from which 24 are located in the transmembrane helices and four in the C-loops (Table 5.2; Fig. 5.1). Six polymorphic AA positions are close to the retinal-binding pocket (Table 5.2). Three examples of a correlation between AA substitution and the habitat type (Table 5.1, 5.2) were found: AA substitutions M155V and L216F were only present in *G. flavescens*, the only species in the 'sand goby' group that has a marine pelagic life style, the T/I281A substitution that was only present in the two obligatory freshwater species, *E.*

pygmaeus and *K. milleri*, and finally the I205L substitution that only occurred in *P. microps* and *P. canestrinii*, both sampled in highly turbid environments.

Table 5.2 Amino acid (AA) replacements at 28 polymorphic sites in the 'sand goby' group. The AA sequence of the ancestral pigment (ancestral haplotype) was inferred using a likelihood-based Bayesian method (Yang 1997). Shaded columns indicate the AA replacements that either line the chromophore-binding pockets or are located in close proximity to the chromophore (Spady *et al.* 2005; Bowmaker 2008). Dots indicate the identity of the AAs with those of the ancestral haplotype.

AA number	71	74	112*	118*	149	151	155	158	160	165*	166	205	213	214	216	217*	255	259	261	263	266	270	281*	287	290	292	300	304
Ancestral haplotype	P	Y	L	I	T	N	M	A	S	A	T	I	L	I	L	V	I	I	F	V	L	S	T	F	I	S	L	M
Pomatoschistus minutus	A	L	V	V	.	/T	.	.	.	.	.	.	.	/I/T	.	I/T	.	.	/Y	T	.	.	.	.	.	/A	.	.
P. minutus Adriatic	A	L	V	V	.	.	.	.	.	.	.	.	.	I	.	I/T	.	.	.	T	.	.	.	.	.	/A	.	.
P. microps	.	.	.	V	S	.	.	.	.	.	.	L	.	.	.	.	.	.	Y	.	.	.	.	.	/V	.	.	.
P. lozanoi	A	L	V	V	.	.	.	.	.	.	.	.	.	/A	.	I	.	.	.	T	.	.	.	.	.	/A	.	.
P. pictus	.	.	.	.	.	.	.	.	.	.	.	.	S	.	.	F	.	V	.	.	.	.	I	.	V	.	I	.
P. canestrinii	.	.	.	.	S	.	.	.	.	S	.	L	.	.	.	.	V	.	Y	.	.	.	.	.	.	.	.	.
Gobiusculus flavescens	.	.	.	V	S	.	/V	.	.	.	S	.	.	.	F	I	V	.	.	.	.	.	.	.	.	.	/I	.
Economidichthys pygmaeus	.	.	I	.	.	.	.	.	T	G/C	S	.	.	.	.	F	.	.	.	.	.	.	A	L	.	.	.	.
Knipowitschia milleri	.	.	V	.	.	.	.	G	.	.	.	.	V	.	.	.	.	.	.	.	S	G	A	.	.	.	.	V

Phylogenetic analysis

Phylogenetic analyses of the 'total' and 'helices' rhodopsin datasets revealed consistent results for the three methods (Fig. 5.2). For the 'total' dataset, maximum parsimony (MP) analysis generated a single consensus tree with tree length = 189, consistency index (C.I.) = 0.78, retention index (R.I.) = 0.87 and rescaled consistency index (R.C.I.) = 0.71. For the 'helices' dataset, MP generated a consensus tree with tree length = 111, C.I. = 0.81, R.I. = 0.86 and R.C.I. = 0.69. Similar trees were obtained with the neighbour joining (NJ) and maximum likelihood (ML) algorithms (ln L = -1997.62 and ln L = -994.19, respectively for the 'total' and 'helices' dataset).

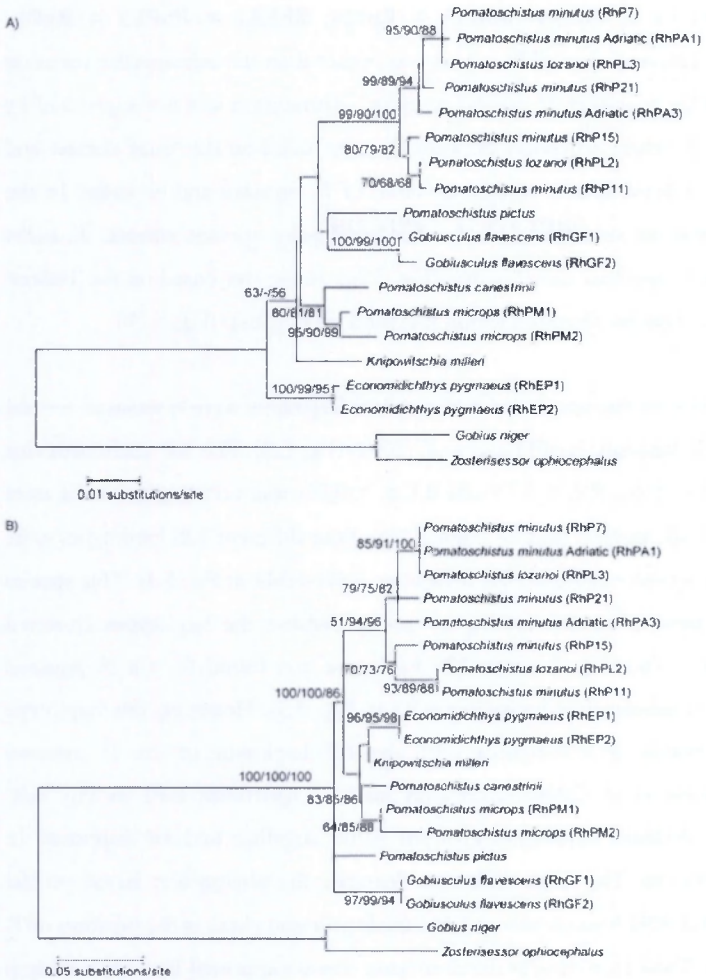


Figure 5.2 Maximum likelihood trees of the 'total' *RH1* dataset (A) and the 'helices' dataset (B) of the 'sand goby' group. *RH1* haplotypes of *Zosterisessor ophiocephalus* and *Gobius niger* were used as outgroup in both trees. Bootstrap values are indicated for statistically supported groupings (≥ 50%) for neighbour joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) analyses (NJ/MP/ML).

The genus *Pomatoschistus* was not monophyletic in both phylogenetic trees (Fig. 5.2a, b). The haplotypes of *P. minutus* were not monophyletic since they clustered together with haplotypes of *P. lozanoi* and the Adriatic *P. minutus*. Moreover, some haplotypes of both species were

identical to haplotypes of *P. minutus* (RhPA5 = RhP21; RhPA1 = RhPL3 = RhP1). Therefore, the intraspecific variation of *P. minutus* was higher than the interspecific variation of the three species of the so-called '*P. minutus* complex'. Although it was not supported by high bootstrap values, the main difference between the tree based on the 'total' dataset and the one based on the 'helices' dataset was the position of *E. pygmaeus* and *K. milleri*. In the first tree, *E. pygmaeus* was the sistergroup to all other 'sand goby' species, whereas *K. milleri* clustered together with *P. canestrinii* and *P. microps* (Fig. 5.2a). In the tree based on the 'helices' dataset, *K. milleri* and *E. pygmaeus* clustered within the 'sand goby' group (Fig. 5.2b).

The phylogenetic analyses of the combined 12S and 16S fragments were consistent for the three methods and with the analysis of Huyse *et al.* (2004) (Fig. 5.3). The MP consensus tree (tree length = 291, C.I. = 0.66, R.I. = 0.74 and R.C.I. = 0.55) was very similar to the trees obtained in the NJ and ML analysis (ln L = -2602.9776). Four different 12S haplotypes were found for *K. milleri* (Accession numbers after reviewing; KM1-KM4 in Fig. 5.3). This species can be considered a member of the 'sand goby' group because the haplotypes clustered firmly within this group (Fig. 5.3). A single 12S haplotype was found for six *E. pygmaeus* specimens (Accession numbers after reviewing; EP2 in Fig. 5.3). However, this haplotype differed in three nucleotides in comparison with the 12S haplotype of the *E. pygmaeus* specimens used in Huyse *et al.* (2004) (Accession number AJ616820; EP1 in Fig. 5.3). Nevertheless, both *E. pygmaeus* haplotypes grouped firmly together and are supposed to represent one single species. The main difference between the phylogenies based on the combined 12S and 16S mtDNA sequences and the rhodopsin sequences is the position of *P. canestrinii* and *P. microps*. Their clustering in the rhodopsin tree is supported by high bootstrap values, although they are not clustered in the 'neutral' phylogeny (Fig. 5.2 and 5.3) (Huyse *et al.* 2004).

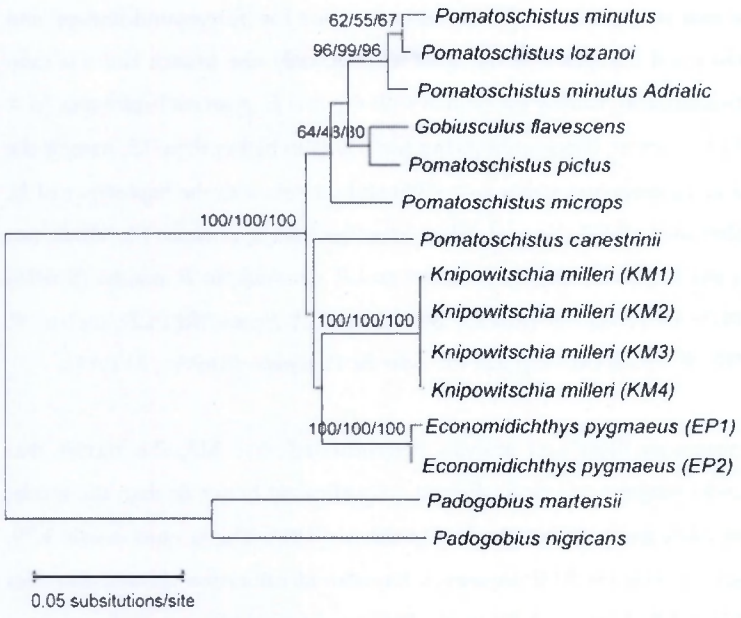


Figure 5.3 Maximum likelihood trees of the combined 12S and 16S mtDNA sequences of the 'sand goby' group. Bootstrap values are indicated for statistically supported groupings (≥ 50%) for neighbour joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) analyses (NJ/MP/ML).

Tests of positive selection

MEGA was used to test the null hypothesis of neutral evolution (H0: $d_N = d_S$) *versus* the hypothesis of positive selection (H1: $d_N > d_S$) for each pair of haplotypes. For the 'total' dataset, no haplotype pair could significantly reject the null hypothesis. For the 'helices' dataset, the null hypothesis was significantly rejected (p-value <0.05) between *K. milleri* and *E. pygmaeus*, and between *K. milleri* and *P. canestrinii*.

Also codon-based branch models of sequence evolution were used to detect positive selection. The single ω ratio of model M0 was higher for the tree based on the 'helices' dataset (ω = 0.46) than the one based on the 'total' dataset (ω = 0.28). For both datasets, no likelihood ratio test was significant in the comparison between the ω ratio of model M0 and

the two-ratio model that estimates two different ω ratios, one for 'foreground lineage' and another for the background lineages. For the 'total' dataset, only one branch had a ω ratio higher than 0.5 (not significant), namely the branch with the two *E. pygmaeus* haplotypes (ω = 0.6806). For the 'helices' dataset, different branches had a ω ratio higher than 0.5, namely the branch with the two *E. pygmaeus* haplotypes (ω = 0.96) and the one with the haplotypes of *E. pygmaeus* and *K. milleri* (ω = 0.97). Several other branches had a ω ratio >1, albeit not significantly; namely the branch leading to *P. microps* and *P. canestrinii*; to *P. minutus* (RhP15) and *P. lozanoi* (RhPL2); to *P. minutus* (RhP15, RhP11) and *P. lozanoi* (RhPL2); and to *P. minutus* (RhP7, RhP21), *P. lozanoi* (RhPL3) and the Adriatic *P. minutus* (RhPA1, RhPA3).

The LRT of the maximum-likelihood analysis demonstrated that M8, the model that accounts for sites under positive selection, showed a significantly better fit than the model M7, which does not allow for positive selection (p-value <0.01). The ω ratio equals 1.79, indicating positive selection on the *RH1* sequences. Bayesian identification showed that sites AA112, AA133, AA165, AA217 and AA281 of the *RH1* gene are significantly under positive selection.

Discussion

The results of three different approaches of phylogenetic analysis showed that the evolution of the rhodopsin gene in the 'sand goby' group is incompatible with a neutral model of sequence evolution. Clear indications for positive Darwinian selection were found first by linking functional variation on the *RH1* gene to the light environment; then by comparing the 'neutral' phylogeny with the phylogeny based on the *RH1* gene; and finally by using statistical tests to detect signatures of directional selection on the *RH1* gene.

Functional polymorphism at the RH1 *gene in the 'sand goby' group*

A microspectrophotometric study showed differences in the λ_{max} values of retinal rods between several species of the 'sand goby' group (Jokela *et al.* 2003; Utne-Palm & Bowmaker 2006). Since those differences could not be explained by an extra rod visual pigment or chromophore change, polymorphism on the rhodopsin gene instead of physiological

changes was expected in the spectral tuning mechanism of the 'sand goby' group. The results of the present study confirmed this hypothesis and showed that 11% of the nucleotide sites of the *RH1* fragment was polymorphic. A comparable percentage of variable sites in *RH1* was found in two related studies, namely for 31 seven-spined goby species of the *Elecatinus* genus (Taylor & Hellberg 2005) and nine squirrelfish species of the paraphyletic *Neoniphon* and *Sargocentron* genera, another group of nocturnal Teleostei (Yokoyama & Takenaka 2004). In the 'sand goby' group, 28 non-synonymous polymorphisms were found with 24 of them located in the transmembrane helices of the rhodopsin. Since 40% of the AA of the *RH1* fragment is located in the loops, the polymorphic AA outside the helices (11%) are strongly underrepresented (Table 5.2). This may illustrate that substitutions in residues located in the seven transmembrane helical regions instead of substitutions in the loops appear to be important for the spectral tuning of the resulting pigment (Bowmaker 2008).

Six AA that are polymorphic among the species of the 'sand goby' group are located closely to the retinal-binding pocket and are important as possible tuning sites on the *RH1* gene (Table 5.2) (Yokoyama 2000). The best-known tuning site is AA261, where a phenylalanine to tyrosine substitution causes a strong red-shift of the λ_{max} values in retinal rods of many families of Teleostei (Yokoyama *et al.* 1995; Hunt *et al.* 1996; Hunt *et al.* 2001; Dann *et al.* 2004; Yokoyama & Takenaka 2004). Based on the 'neutral' phylogeny (Huyse *et al.* 2004) and on the reconstructed ancestral haplotype (Table 5.2), the F261Y mutation has most likely arisen three times independently in the 'sand goby' group: in *P. microps*, *P. canestrinii* and in certain populations of *P. minutus*. The F261Y mutation was only found in Baltic and Mediterranean populations of *P. minutus* that are living in turbid environments and require a red-shifted spectral adaptation of the rhodopsin for an optimal sight in dim light. This substitution was predicted to be the main source of the 2.9 nm red-shift in the λ_{max} values of rod pigment of Baltic compared to Atlantic *P. minutus* (Jokela *et al.* 2003; *subchapter 4a*). The lack of the mutation in the sister taxa of *P. minutus*, suggests a *de novo* mutation scenario instead of the standing variation hypothesis to explain the origin of the functional mutation F261Y in *P. minutus* individuals of the Baltic Sea and Mediterranean Sea (*subchapter 4a*). Nevertheless, hybridization with other species is a potential route to preserve large amounts of standing variation despite negative selection (Barrett & Schluter 2008). In contrast to *P. minutus*, all haplotypes of *P. microps* and *P. canestrinii* include the F261Y mutation. This makes

sense because all samples of both species were caught in brackish environments, characterized by high turbid conditions. Therefore a clear link is found between a specific light environment and functional variation on rhodopsin in the 'sand goby' group.

The second polymorphic site in the 'sand goby' group is AA299. This is another well-identified tuning site in the retinal rods documented in many families of Teleostei (Yokoyama *et al.* 1995; Hunt *et al.* 2001) and in the bottlenose dolphin (*Tursiops truncatus*) (Fasick & Robinson 1998). The S299A substitution, that causes a weak red-shift of the λ_{max} value, is only found in the three (studied) species belonging to the '*P. minutus* complex'. This complex comprises four cryptic species, *P. minutus*, the Adriatic *P. minutus*, *P. lozanoi* and *P. norvegicus*, which all speciated during the early Pleistocene (Webb 1980; Gysels *et al.* 2004b; Huyse *et al.* 2004). Therefore it is thought that the S299A mutation originated only once in the 'sand goby' group in adaptation to marine waters with an above average wavelength of maximum transmission light (WMTL). The observed heterozygosity of this AA within the three species is most likely the result of the optical heterogeneity of their analogous habitat, a suggestion already made for the *P. minutus* populations in *subchapter 4a.*

The effects of substitutions on the four remaining polymorphic AA that are closely located to the retinal-binding pocket of the opsin (AA133, AA214, AA217 and AA300) have not yet been validated experimentally. Although the effect of variation at AA214 and AA217 on λ_{max} values has been tested by mutagenic experiments on red/green opsins of humans (Asenjo *et al.* 1994), the effect of substitutions on these AA are not yet known for rhodopsin. In addition to these six AA, other sites can also be involved in the spectral tuning mechanism of opsins. Candidate sites are those with a likely link between AA substitutions and spectral conditions of the habitat. In the dataset three examples were noticed (Table 5.2); first, AA substitutions M155V and L216F were only present in *G. flavescens*, the only species in the 'sand goby' group that lives pelagically; second, the T/I281A substitution was only present in the two obligatory freshwater species, *E. pygmaeus* and *K. milleri*; and finally, the I205L substitution occurred only in *P. microps* and *P. canestrinii*, both sampled in highly turbid environments (Table 5.1).

Only the phenotypical effect of substitutions on two ΛΛ is known for the 'sand goby' group (ΛΛ261 and ΛΛ299). However, other ΛΛ substitutions have to be involved to explain the differences in the measured λ_{max} values of retinal rods between species; e.g. the Baltic *P. minutus versus P. microps* (λ_{max} = 508.3±0.5 nm and 515.7±0.4 nm, respectively) (Jokela *et al.* 2003). Most likely one or more ΛΛ substitutions are responsible for the red-shift of 7.4 nm in *P. microps* in comparison with the Baltic *P. minutus*. Seven ΛΛ substitutions are different among the two species, with one at ΛΛ217, close to the retinal-binding pocket (Table 5.2). Future research with mutagenic experiments has to validate the adaptive significance of this specific and other ΛΛ substitutions at *RH1*. Moreover, absorbance spectra in retinal rods of all other species of the 'sand goby' group have to be measured microspectrophotometrically to link the phenotypical differences to the specific ΛΛ substitutions at the *RH1* gene.

A comparison between the 'neutral phylogeny' and RH1 *phylogeny*

The phylogeny of the 'sand goby' group based on the rhodopsin sequences was not congruent with the one based on the mtDNA and ITS rDNA sequences (Fig. 5.2 and 5.3; Huyse *et al.* 2004). One discordant pattern supported by high bootstrap values was the clustering of *P. canestrinii* and *P. microps*, collected in a highly turbid habitat, based on *RH1* genotypes, although this was not supported by the mtDNA tree (Huyse *et al.* 2004). Additionally, the clustering of those two species was supported by the *RH1* tree based on non-synonymous nucleotides in contrast to the *RH1* tree based on only synonymous SNPs (results not shown). Another discrepancy between the two phylogenetic analyses was the clustering of two obligatory freshwater fish, *K. milleri* and *E. pygmaeus*, in the phylogeny based on the 'helices' dataset in contrast to the 'neutral' tree, albeit with low bootstrap support. This indicates that 'sand goby' species clustered according to the photic conditions of their habitat with regards to the functional variation on *RH1* instead of phylogenetic proximity.

A final clear contrast with the 'neutral' phylogeny is that *RH1* haplotypes of *P. lozanoi* and the Adriatic *P. minutus*, considered as separate species, clustered within the *P. minutus* haplotypes (Fig. 5.2). The intraspecific *RH1* variation of *P. minutus* is higher than the interspecific variation of the *P. minutus* complex. Nevertheless, the species of this complex are morphologically and genetically diversified; they speciated during the early Pleistocene

(Webb 1980; Gysels *et al.* 2004b; Huyse *et al.* 2004). The low differentiation on *RH1* between those species can be the result of either the lack of lineage sorting because of the recent differentiation, the hybridization that still occurs between the species of the complex (Fonds 1973; Wallis & Beardmore 1980) or the local adaptation of *P. minutus* to different optic environments in contrast to the other species of the complex. More research on this *P. minutus* complex is required, including on *P. norvegicus* (Huyse *et al.* 2004), the fourth species of the complex that is missing in the present analysis. Living at a depth of 400 m, *P. norvegicus* encounters a light regime much different from the other species of the *P. minutus* complex.

The discordant patterns between the 'neutral' and the *RH1* phylogeny indicate that rhodopsin is not always a good genetic marker to infer the general phylogeny of aquatic taxa. Therefore it is recommended not to use rhodopsin as 'neutral' marker in phylogenetic studies, as has been done on a regular basis (Taylor & Hellberg 2005; Mayden *et al.* 2007; Schoenhuth *et al.* 2008).

Tests of selection on RH1

The d_N/d_S ratios for the whole *RH1* fragment analysed by Z-tests and 'branch-specific models' were not significant. However, these tests of neutrality are generally conservative because the substitution rates are averaged across all AA sites (Bamshad & Wooding 2003). Since substitutions in residues located in the seven transmembrane helical regions appear to be important for the spectral tuning of the pigment (Bowmaker 2008), tests revealed higher but insignificant d_N/d_S ratios for the 'helices' dataset in comparison with the 'total' dataset. Therefore, the low statistical power of those tests is most likely linked to the small number of AA replacements at critical sites involved in the spectral tuning of visual pigments (Yokoyama 2002). This hypothesis is fully confirmed by analyses of d_N/d_S ratios under 'site-specific models' that indeed provided evidence for positive Darwinian selection at specific codons of the rhodopsin gene. The Bayesian analysis in the site-specific model identified five individual positively selected sites at *RH1* of the 'sand goby' group, namely AA112, AA133, AA165, AA217 and AA281. Two of those sites, AA133 and AA217, are in close proximity of the chromophore. Remarkably, one codon that is identified as possibly under selection, AA281, is located in the loops, although it is believed that λ_{max} of VP are modulated by

interactions between the retinal and AA either within or near the retinal in the transmembrane helices (Bowmaker 2008). However, one study has shown already that AAs outside the helices may be involved in the spectral tuning of rhodopsins (Yokoyama *et al.* 2007). For AA281 there was even an unambiguous relationship between the specific habitat behaviour and the AA substitution T281A for the two obligatory freshwater fish. Therefore, the T281A may be a specific adaptation of the *RH1* gene to an exclusive freshwater lifestyle. Moreover, the MEGA analysis for the 'helices' dataset also showed a significant result for positive selection ($d_N/d_S > 1$) between the two freshwater fishes.

A recent intraspecific study of rhodopsin variation in *P. minutus*, also detected a significant signal of positive selection on *RH1*. Three AA were identified as positively selected codons (AA151, AA214 and AA299) (*subchapter 4a*). Because there are no common positively selected AA sites in the analyses for *P. minutus* and for the total 'sand goby' species, natural selection seems to influence other AA sites on *RH1* in *P. minutus* than at the phylogenetic level. This makes sense because the range of habitat types for all species is broader than for *P. minutus* only. It will be interesting to study the possibility for local adaptation in other species of the 'sand goby' group. Since no differences were found between the Baltic and Atlantic haplotypes of *P. microps*, it seems that unlike *P. minutus* not all species of the 'sand goby' group are locally adapted at the *RH1* gene. However, the habitat of *P. microps* is only restricted to estuaria and brackish water, in contrast to the range of habitat use known for *P. minutus*. Therefore, to study the possibility of local adaptation at *RH1* the focus has to lie on species with a high diversity in habitat use along their distribution range, as it is the case for *P. canestrini* and *G. flavescens*.

Conclusion

The present study highlights the usefulness of sensory genes, like rhodopsin, for studying adaptive molecular evolution in non-model taxonomic groups. Based on three different approaches to detect positive selection at the *RH1* gene, species of the 'sand goby' group are most likely adapted to their local light environment. Moreover, several new potential spectral tuning sites on the *RH1* were identified. They await validation with mutagenic studies. Most mutations considered in biochemical analyses to elucidate the spectral tuning mechanism are

not found in nature, and their relevance in the actual λ_{max} shifts of visual pigments is not necessarily clear (Yokoyama *et al.* 1995). Therefore, phylogenetic studies provide a valuable framework to suggest potential tuning sites of opsin genes based on actual polymorphism data.

Future research has to test if positive selection on opsin genes is driving the evolution of other aquatic species and to verify the role of the other types of opsin genes in speciation. This will help to answer whether the sensory drive hypothesis can explain fast speciation in taxonomic groups as the 'sand goby' group. Between the species of the 'sand goby' group there is a huge differentiation in visual mating systems with species-specific colour signals (Forsgren 1997; Amundsen & Forsgren 2001). Therefore, the present study suggests that natural selection might also influence the genetic variation on cone opsins between species and that sensory drive could even lead to the evolution of colour polymorphisms and hence speciation in the 'sand goby group' due to heterogeneous light environment.

Acknowledgements

The authors wish to thank the following persons who kindly provided the samples: Anne Christine Utne-Palm (University of Bergen, Norway), Sofia Giakoumi and Alcibiades Economou (Hellenic Centre for Marine Research, Greece), Emilie Egea (Centre Océanologique de Marseille, France), Els Cuveliers, Bram Vanschoenwinkel and Maarten Vanhove (K.U.Leuven), and Jef Guelinckx (VMM). This work was supported by a PhD fellowship of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This study is part of the WESTBANKS project financed by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (contract number SD/BN/01A).

Appendix of Chapter 5

Table S5.1 Inventory of the rhodopsin haplotypes and their distribution in the 'sand goby' group. N = number of haplotypes per sampling location.

Species	Sampling site	Haplotype	N
Pomatoschistus minutus Adriatic	Venice (IT)	RhPA1	2
		RhPA2	1
		RhPA3	1
		RhPA4	2
		RhPA5	1
		RhPA6	1
Pomatoschistus microps	Värtan (S)	RhPM4	4
		RhPM5	26
		RhPM6	6
	Scheldt River (B)	RhPM1	2
		RhPM6	4
	Dale (UK)	RhPM4	1
		RhPM5	3
	Roscoff (F)	RhPM2	1
		RhPM3	1
		RhPM5	2
		RhPM6	2
Pomatoschistus lozanoi	Ostend (B)	RhPL1	2
		RhPL2	6
		RhPL3	2
Pomatoschistus pictus	Ostend (B)	RhPP1	8
Pomatoschistus canestrinii	Chioggia (IT)	RhPC1	4
Gobiusculus flavescens	Bergen (N)	RhGF1	5
		RhGF2	2
		RhGF3	3
		RhGF4	2
		RhGF5	2
Economidichthys pygmaeus	Acheloos River (G)	RhEP1	1
		RhEP2	11
Knipowitschia milleri	Acheron River (G)	RhKM1	16

B, Belgium; F, France; G, Greece; IT, Italy; S, Sweden; UK, United Kingdom

Chapter 6

General discussion

'To say that we should drop the idea of truth as out there waiting to be discovered is not to say that we have discovered that, out there, there is no truth.'

Richard Rorty (American philosopher, 1931-2007).

Under natural selection, individuals tend to adapt to the local environmental conditions, resulting in a pattern of **local adaptation** (Lenormand 2002). Understanding the genetic basis of local adaptation is a prime interest in biology as it involves the role of natural selection in promoting evolutionary change (Mayr 1963). Since gene flow is expected to hamper adaptive population divergence, the traditional idea was that local adaptation may be rare or absent in organisms (Nielsen 2005), including marine (Hemmer-Hansen *et al.* 2007). Today, local adaptation in marine organisms has become increasingly documented, indicating that natural selection is a potent evolutionary force in the ocean (Canino *et al.* 2005; Pampoulie *et al.* 2006; Hemmer-Hansen *et al.* 2007; Sherman & Ayre 2008). Nevertheless, knowledge about adaptive evolution in marine organisms remains limited, yet is crucial to improve our understanding of how evolution operates in the ocean (Conover *et al.* 2006). The aim of this thesis was to contribute to the actual research about the importance of natural selection as evolutionary force in marine organisms and about the spatial and temporal scale of adaptive genetic variation in marine systems.

The reason for the lack of knowledge about adaptive evolution in the ocean is the scarcity of genetic systems to evaluate natural selection (Yokoyama 2002). One of the few excellent models to elucidate the mechanism and importance of selection as evolutionary force is the spectral tuning mechanism of **visual pigments** (VP) in vertebrates (Yokoyama 2000). This model can identify specific AA changes which are responsible for the adaptation of organisms to specific environments. Moreover, VPs have a well-defined role in nature as they allow the organism to detect differences in the spectral composition of the environment (Bowmaker 2008). Therefore they have a strong effect on the evolution of organisms, providing an excellent system to study adaptive evolution by means of molecular tools.

As model system in this thesis, we used the rhodopsin (*RH1*) gene, which determines the spectral sensitivity of dim-light vision, of the sand goby *Pomatoschistus minutus* was selected because microspectrophotometric measurements of the rods suggested local adaptation (Jokela *et al.* 2003). Therefore, the research question of the present thesis was: '**Is there local adaptation on the rhodopsin gene in the sand goby?**' An approach with five different tasks to demonstrate local adaptation in the wild was taken to resolve this research question.

TASK 1. A high quality neutral marker-derived phylogeographic and population genetic analysis for the sand goby

Neutral-marker derived phylogeographic patterns

The phylogeographic and population genetic analysis was carried out by sequencing a partial fragment of the mtDNA cytochrome *b* (cyt *b*) gene (*subchapter 3a*) and by genotyping eight microsatellite markers (*subchapter 3b*) of several sand goby samples throughout the full species range. Moreover, published data of allozyme markers and mtDNA control region sequences were implemented and reanalysed in *subchapter 3a*. The analysis of the results from different types of molecular markers and genetic statistic methods revealed one consistent genetic pattern for the sand goby.

The population genetic structure of *P. minutus* is the result of historical and contemporary processes. Middle Pleistocene glaciations yielded isolated and differently evolving sets of populations. The Mediterranean sand gobies (MS-Clade) diverged from the Atlantic sand gobies (AO-Clade), as suggested previously by Miller (1986) based on morphological and ecological differences. In the second half of the Middle Pleistocene, the AO Clade diverged further into two phylogeographical groups: the Iberian Peninsula (IB) Group and the North Atlantic (NA) Group. The historical demography of Mediterranean sand gobies was influenced mainly by Middle Pleistocene glaciations, in contrast to that of the Atlantic populations, which was shaped by Late Pleistocene expansions (*subchapter 3a*).

The pre-LGM (Last Glacial Maximum) subdivision signals were not erased for *P. minutus* during the LGM. In contrast to most organisms with a presumed Iberian refugium (e.g. Consuegra *et al.* 2002; Cortey & Garcia-Marin 2002; Gysels *et al.* 2004a; Olsen *et al.* 2004), the Iberian sand gobies appear not to have contributed to the post-LGM distribution expansion in northern Europe. Moreover, secondary contact after the divergence of the three phylogeographic groups did not take place and contemporary gene flow within these groups is presumed to be limited. The low level of genetic differentiation between sand goby populations observed with neutral nuclear markers resulted most likely from high effective population sizes (*subchapter 3b*).

The genetic pattern of the sand goby is more structured in contrast to other marine organisms of the northeastern Atlantic. Since most phylogeographic studies were focused on (commercial) fishes and invertebrates with high dispersal potential (Nesbø *et al.* 2000; Bremer *et al.* 2005; Chevolot *et al.* 2006), species showing reduced levels of contemporary gene flow, such as the sand goby, seem to be better suited for elucidating phylogeographic patterns. More surveys on sedentary and demersal marine organisms are required for a comprehensive view on the comparative phylogeography of northeastern Atlantic marine species (Patarnello *et al.* 2007).

Potential for local adaptation

The observation of low genetic differentiation with nuclear markers has always been assigned to high rates of gene flow. Therefore, the potential for local adaptation was expected to be limited for *P. minutus* because if gene flow is high it should preclude local adaptation by rapidly blurring allele frequency differences between populations (Gysels 2003; Stefanni *et al.* 2003). However, *chapter 3* showed that high effective population sizes instead of high gene flow explained the low values for nuclear population differentiation in the sand goby. Therefore, the evolutionary scenario with temporally stable neutral genetic structuring among sand gobies populations revealed that selection would be able to override the effects of drift and gene flow, resulting in adaptive population divergence.

Next, the robust phylogeographic and population genetic structure of the sand goby serves as a solid neutral background to compare with population differentiation based on presumably adaptive data (Beebee & Rowe 2004). The usefulness of the incorporation of demographic information (neutral processes) to identify evidence of selection in genes has been highlighted in human genomics (Muers 2009; Nielsen *et al.* in press). The authors identified novel targets of selection including some linked to psychiatric disorders. They remained largely undetected in previous studies that focussed specifically on positive selection without any approach that separated demography from selection (Muers 2009).

Quality in population genetic structure

The pattern in *chapter 3* provided an opposite answer about the potential for local adaptation in comparison with previous phylogeographic studies on the sand goby (Gysels 2003; Stefanni *et al.* 2003; Stefanni & Thorley 2003; Gysels *et al.* 2004b). Consequently, it is crucial to discuss the robustness of the new results, so that we can assure that our conclusions should still be valid when more genomic information of the sand goby will be available in the future. Moreover, the quality of the data and analysis is a crucial factor in the genetic analyses for marine species due to the weak detectable genetic signal for population differentiation (McKelvey & Schwartz 2004; DeWoody *et al.* 2006). As a consequence, various errors associated with estimating population genetic parameters that might normally be safely ignored assume a relatively greater importance (Waples 1998).

Based on the literature, there are three major quality-checks to ensure the quality of a phylogeographic and population genetic study: on the sampling strategy, the molecular marker and the statistical analysis. All previous studies on sand gobies suffered from shortcomings in relation to these quality-checks; in our new analysis we tried to address them. First, we optimized the **sampling strategy** with the certainty that we were dealing with a single taxon based on a molecular species identification tool (*subchapter 2a*). The necessity of this test is illustrated by the cross-amplification of the microsatellites in the different species whereby the conspecific *P. lozanoi* cannot be individually differentiated from *P. minutus* with microsatellite markers (*subchapter 2b*). In the sampling strategy, sand gobies from all marine systems and replicated in time were included in order to differentiate between spatio-temporal differentiation (Waples 1998). Second, we guarantee the data reproducibility of the **genetic markers**. For all genotyping and sequencing, pre- and post-PCR were separated to avoid contamination (*subchapter 3a, b*). Moreover, only high quality microsatellite markers were developed (*subchapter 2b*) and genotyped (*subchapter 3b*). Genotyping errors (null alleles, size homoplasy, allelic dropout, contamination, stuttering, etc.) were estimated with replicate genotyping (5-10% of the samples) together with software to detect and calculate the genotyping error rate (*subchapter 3b*). Third, we analysed various types of molecular markers (nuclear and mitochondrial markers) and used several statistical methods (F-statistics, Bayesian statistics and coalescent methods) for the **interpretation** of

the data. By combining the different types of markers and statistics, it was clear that not the dataset was incorrect in previous genetic analyses, but only the interpretation of the results.

The present study still suffered from some limitations: the sampling strategy was limited in space and time for some regions, and the number of nuclear markers (8) was low in comparison with the number of chromosomes (47) in *P. minutus* (Webb 1980). Nevertheless, the results of the different genetic markers and statistic methods showed a robust pattern for the sand goby. Therefore, this genetic pattern may serve as neutral background in our approach to demonstrate local adaptation. Task 1 of our approach is therefore fulfilled.

TASK 2. Sand goby populations differ in functional variation on the rhodopsin gene (RH1)

Population differentiation in the λ_{max} values of the retinal rods in *P. minutus* was observed by microspectrophotometry (Jokela *et al.* 2003). To realize the second task, the possibility of physiological adaptation had to be ruled out to explain this phenotypic population differentiation. Since these phenotypic differences could not be explained by a chromophore change (A1 *versus* A2 chromophores) or an additional rhodopsin gene by a recent gene duplication, polymorphism on the rhodopsin gene (*RH1*) was suggested. Sequence analysis of *RH1* revealed indeed polymorphism (*subchapter 4a, b*). The analysis showed even the highest level of intraspecific polymorphism at a spectral opsin gene observed in vertebrates. Five amino acid sites on *RH1* were polymorphic in the sand goby; two of them (AA261 and AA299) are known to be spectral tuning sites and to have a significant effect on the λ_{max} values of retinal rods in aquatic vertebrates (Hunt *et al.* 2001; Yokoyama & Takenaka 2004; Hunt *et al.* 2007); two others (AA214 and AA217) are known to be close to the retinal-binding pocket and are potential spectral tuning sites for rhodopsin in vertebrates (reviewed in Yokoyama 2000).

Sand goby samples of the European marine systems were differentiated from each other based on functional variation at *RH1*, including on the two well known spectral tuning sites (AA261 and AA299). The expected population differentiation on λ_{max} values of the retinal rods based on the effect of AA-substitutions on those two sites was consistent with the

phenotypic population differentiation of Jokela *et al.* (2003) (*subchapter 4a*). Next, the functional variation at the *RH1* gene was also polymorphic within populations (*subchapter 4a, b*), demonstrating a genetic basis for within-population λ_{max} variation (Jokela *et al.* 2003). The equal distribution of tyrosine and phenylalanine on AA261 in the northern Baltic individuals explains most likely the broad λ_{max} distribution spanning 5.7 nm of the spectrum in this particular population in contrast to the much smaller range of λ_{max} in North Sea samples (Jokela-Määttä *et al.* 2006; Jokela-Määttä *et al.* 2007). Additionally, the genetic basis for the observed phenotypic differences on the rod opsins between and within sand goby populations was confirmed. The direct link between phenotype and genotype was identified in mutagenesis and microspectrophotometric experiments that tested the effect of specific AA substitutions on the phenotypic variation (Yokoyama 2000). Therefore task 2 has been fulfilled.

TASK 3. The differences between sand goby populations for functional variation at RH1 *are due to natural selection as opposed to neutral processes*

Two independent approaches were used to detect the signature of natural selection at the *RH1* gene. The first approach was the use of **sequenced-based neutrality tests** based on comparisons between the different classes of mutations within the rhodopsin gene. The d_N/d_S substitution rates of the complete *RH1* fragment in *P. minutus* did not reveal selection (*subchapter 4a*). However, tests of neutrality are generally conservative because substitution rates are averaged across all amino-acid sites tested (Bamshad & Wooding 2003). It is believed that adaptation and major functional shifts occur by amino acid replacements at a small number of critical sites in proteins (Golding & Dean 1998), inclusive in visual pigments (Yokoyama 2002). Analyses of d_N/d_S ratios and likelihood ratio tests under 'site-specific' models indeed detected significant signals of positive Darwinian selection at the *RH1* gene. Bayesian analysis identified three individual positively selected sites at *RH1*, including AA299, which was verified as a true tuning site for rhodopsin (*subchapter 4a*). Applications of these site-specific methods represent very powerful tools to obtain evidence of positive selection and to study function of proteins (Clark *et al.* 2003; Sawyer *et al.* 2005).

The second approach was the comparison of the geographical distribution of the rhodopsin variation with the **neutral marker-derived phylogeographic and population genetic pattern**. There is a great potential for this candidate gene *versus* neutral variation approach to disclose the genetic basis of adaptive population divergence among neutral populations. Nevertheless, it has only rarely been applied in marine fish (but see Pampoulie *et al.* 2006; Hemmer-Hansen *et al.* 2007) due to the labors involved to make this comparison in non-model species. In the present study, strong discrepancies were found between the distribution of the variation at *RH1* and the phylogeographic pattern of the sand goby based on the distribution of the variation at the mtDNA cyt *b* gene (*subchapter 3a*), nuclear allozyme markers (reanalysed in *subchapter 3a*; Gysels 2003; Stefanni *et al.* 2003) and microsatellite markers (*subchapter 3b*). Samples of the northern Baltic and Mediterranean Sea carry a similar allelic profile of the *RH1* gene, although the Mediterranean *P. minutus* individuals belong to a different phylogeographic group (MS-Clade) than the Atlantic and Baltic sand gobies (AO-Clade). The *RH1* gene pattern was also congruent between sand gobies from the Iberian Peninsula, Irish Sea and Bay of Biscay. However, Iberian sand gobies belong to a different historical unit (IB-Group) compared to the North Atlantic gobies (NA-Group), which includes all the populations from the Bay of Biscay to the northern Baltic. Additionally, distributions of the *RH1* and microsatellite variation were different from each other based on Mantel tests, Procrustes analyses and outlier detection tests (*subchapter 4a*). On the other hand, there was no observation of functional differentiation between samples of the North Sea on the *RH1* gene (*subchapter 4b*). Nevertheless, catchment data (Vanden Eede 2006; Guelinckx 2008), neutral genetic data of mtDNA (*subchapter 3a*) and microsatellites (Pampoulie *et al.* 2004a) showed the existence of different sand goby populations in the North Sea. Therefore, it is shown that random processes are ruled out to explain *RH1* pattern between the sand goby populations within and between marine systems. Hence, task 3 is fulfilled.

Many studies using the candidate gene approach have the problem to rule out completely the possibility that neutral processes instead of natural selection differentiated populations on functional variation (Canino *et al.* 2005; Hemmer-Hansen *et al.* 2007; O'Malley *et al.* 2007; Jorgensen & Emerson 2008; Oliver *et al.* 2009). The use of different marker types (nuclear SNPs, microsatellites, mtDNA) with their specific mutation pattern and heterozygosity

introduced statistical and interpretational problems in outlier tests. Comparing patterns between populations statistically in Mantel and Procrustes analyses rather than comparing levels of population structuring (F_{ST}), appears to be an excellent and more reliable approach. Therefore, the strategy used in *subchapter 4a*, with the Procrustes analyses has a high potential to demonstrate that selection rather than only random processes influenced functional polymorphisms in other studies on candidate gene.

TASK 4. Establishing a link between the functional variation on the RH1 *gene and selection regimes*

The function of the *RH1* gene is unambiguous; it is the only visual pigment gene in the sand goby that is expressed in the rod photoreceptors and produces monochromatic **dim light vision**. No recent duplication event of the *RH1* gene lineage occurred in *P. minutus* and the 'sand goby' group (*subchapter 4a*; Jokela *et al.* 2003; Utne-Palm & Bowmaker 2006) in contrast to several cichlid genera (Spady *et al.* 2005), European eel *Anguilla anguilla* (Archer *et al.* 1995) and common carp *Cyprinus carpio (Lim et al. 1997)*. Therefore, functional variation at *RH1* may have a direct effect on the fitness of sand gobies. They are visual feeders (Healey 1971; Aarnio & Bonsdorff 1993), mostly nocturnal (Ehrenberg & Ejdung 2008) and hence should depend on the RH1 gene. For *P. minutus,* nocturnal foraging is advantageous in approaching prey and in avoiding predators (Thetmeyer 1997). The adaptive significance of dim light vision is thus obvious: individuals with good eyesight will be better able to locate food and mates and to avoid predators than individuals with poor eyesight in dim light (Freeman & Herron 2001).

Dim light vision should benefit from increased photon capture. One evident selection pressure on visual pigments for dim-light vision is that spectral absorbance should in some way 'match' the spectral distribution of available photons (Jokela-Määttä *et al.* 2005). The best concept to characterize the dim light conditions of different waters is the '**wavelength of maximally transmitted light**' (WMTL) (Lindström 2000; Audzijonyte *et al.* 2005). The WMTL differs between marine systems mainly because of variation in the concentration of detrital and mineral suspended particles (NAP or non-algal particles), and coloured dissolved organic matter (CDOM) (Kirk 1996). A combination of underwater light measurements and

optical modelling has been used by Lindström (2000), Jokela-Määttä *et al.* (2005) and Audzijonyte *et al.* (2005) to estimate the WMTL at various locations. A promising method using satellite remote sensing data in *subchapter 4a* resulted in a map with WMTL values anywhere on the European continental shelf. Based on this map and earlier published data (Jerlov 1976), the generality and robustness of the differences in light regimes between and within the marine systems are well known.

Differences in environmental light transmittance (WMTL) between the marine systems correspond well with the population differentiation in the absorbance spectra of the retinal rods (Jokela *et al.* 2003) and with the temporally stable differentiation on the functional variation at the *RH1* gene (*chapter 4*). The sequence results illustrate that based on the rhodopsin variation sand goby samples clustered according to the local photic climate instead of historical or geographic proximity (*chapter 3, 4*). The blue-shift of the λ_{max} values of the dim-light receptors and the high frequency for the blue-shifted AA substitutions on AA261 and AA299 in the samples of the Bay of Biscay, Irish Sea and Iberian Peninsula are conforming to low WMTL values in these regions. The red-shift of the λ_{max} values of the rods and the high frequency of the red-shifted AA substitutions on AA261 and AA299 in the samples of the northern Baltic and Mediterranean lagoons might be an adaptation to the higher local WMTL conditions (*subchapter 4a*). Within the North Sea region, the high level of polymorphism on *RH1* can most likely be maintained by the high heterogeneity of the light regime along the coastline of the North Sea (*subchapter 4b*). On the other hand, the heterogeneity can also be the outcome of the individual inshore-offshore migration patterns of juvenile sand gobies (Healey 1971; Fonds 1973; Maes *et al.* 1998). The trend that estuarine migrants have much more rhodopsin variants which are associated with brackish water than sand gobies sampled in open sea, showed that the observed genetic heterogeneity in the North Sea may be stabilized by different migration strategies within the populations. Recent ecological studies based on biochemical markers studying the migration pattern in the *P. minutus* between the estuary and open sea supported this hypothesis (Guelinckx 2008; Guelinckx *et al.* 2008). There seems to be no obligate estuarine stage but a highly individual movement pattern in estuarine habitat use. The results of *subchapter 4b* showed that their spectral sensitivity in turbid water may be a crucial factor in the individual exploitation of the estuarine habitat use. Moreover, these results revealed an extra indication for the strong

correspondence between genotypic variation and environmental conditions for the trait under study.

All the measured and presumed differences in λ_{max} of the retinal rods between sand goby populations are much smaller (<10 nm) than expected based on the observed differences in WMTL (circa 100 nm) between marine systems. Nevertheless, they are significant for the spectral sensitivity since a red-shift of 2.9 nm for λ_{max} measured for Baltic sand gobies in comparison with Atlantic relatives revealed a 13-19% better capture of quanta in two northern Baltic Sea locations (Jokela *et al.* 2003). Although the ecological value and fitness relevance are clear and relevant, it is worth noting that the differences in λ_{max} values of the ideally adapted visual pigments need to be similar as the differences in WMTL to be perfectly adapted to the photic environment. VPs are a good demonstration that natural selection cannot result in perfectly adapted organisms because most adaptive traits represent compromises among conflicting needs and have **physiological constraints** (Wolfe *et al.* 2007). Visual pigments vary with respect to two key functional properties: spectral sensitivity and thermal stability (Ala-Laurila *et al.* 2004). Spectral sensitivity expresses the relative probabilities for the pigment to be activated by the electromagnetic radiation of different wavelength. This defines the limits to the visible band of the electromagnetic spectrum. In addition to their activation by light, visual pigments can also be activated by their own thermal energy. At least in their physiological expression, thermal activations of the visual pigments are indistinguishable from activations by light (Donner 1992). They cause a random light-like activity ('dark' light) that constitutes an irreducible intrinsic noise of the visual system and sets an ultimate limit to visual sensitivity (Aho *et al.* 1988). It is clear that spectral sensitivity and thermal stability must both have been of critical importance in the evolution of visual pigments. Therefore, a pigment tries to be spectrally tuned for maximal quantum catch in a given environment and at the same time be as thermal stable as possible (Ala-Laurila *et al.* 2004).

The link between the functional variations on the *RH1* gene with documented selection regimes for the sand goby and the fitness relevance of the phenotypic difference have been clearly confirmed. Therefore, task 4 has been properly accomplished for our model system.

TASK 5. Directional selection in a phylogenetic framework to determine whether selection has played a significant role in RH1 *evolution within the 'sand goby' group*

Evolutionary adaptation has been proposed to explain the interspecific differences in the measured λ_{max} values of retinal rods between closely related species of the 'sand goby' group (Jokela *et al.* 2003; Utne-Palm & Bowmaker 2006). A high level of interspecific polymorphism at the *RH1* gene was observed, including non-synonymous mutations on amino acids known as spectral tuning sites (*chapter 5*). Three different approaches indicated an adaptive molecular evolution on the rhodopsin gene among the related 'sand goby' species that inhabit different photic environments in terms of salinity and depth. First, a link was found between some of the AA substitutions on (potential) spectral tuning sites and the spectral conditions of the habitat environment of the species. Second, phylogeny based on the *RH1* fragments was compared with a 'neutral phylogeny' using partial 12S and 16S mtDNA (Huyse *et al.* 2004). Different 'sand goby' species clustered together based on *RH1* sequences due to a similar habitat choice although they are not related in their 'neutral' phylogeny. Finally, analyses of d_N/d_S substitution ratios and likelihood ratio test under 'site-specific' models detected a significant signal of positive Darwinian selection on specific AAs of the *RH1* gene.

The rhodopsin study of the 'sand goby' group did not provide the first evidence of positive selection on opsin genes in vertebrates within a phylogenetic framework (Terai *et al.* 2002; Dann *et al.* 2004; Spady *et al.* 2005). Nevertheless, with the observation that species of the 'sand goby' group are evolutionary adapted to their photic habitat and that species with a similar selective regime will have a analogous pattern on the *RH1* gene, the arguments that selection on the candidate gene also influences contemporary population structure in *P. minutus* are reinforced. Moreover, sand goby populations living in a highly turbid environment, react similarly as related species living in similar environmental conditions. The red-shifted AA-substitution F261Y was only found in sand goby individuals living in the northern Baltic and Mediterranean lagoons, as well in the species *P. microps* and *P. canestrinii* (*chapter 4, 5*).

P. minutus is a typical example of an eurytopic species occurring in different habitats with specific photic conditions. If *P. minutus* is indeed locally adapted, then intraspecific variation on functional variation of *RH1* must be much higher than relative stenotopic species. The habitat of the common goby *Pomatoschistus microps* is restricted to estuaria and brackish water, in contrast to the wide habitat use of *P. minutus* (Miller 1986). No functional intraspecific variation and no genetic differentiation on *RH1* were observed for Baltic and Atlantic *P. microps* samples (*chapter 5*). Therefore, these results are a further indication for the specific character of adaptation on rhodopsin in *P. minutus*. This is also strengthened by the observation that there are no common AA sites assigned as positively selected in *P. minutus* and all 'sand goby' species. As expected, natural selection seems to influence other AA sites on *RH1* in *P. minutus* than at the phylogenetic level (*chapter 4, 5*).

The phylogenetic analysis provided more indications that rhodopsin variation is indeed influenced by local selective pressure in the sand goby. Therefore, task 5 is fulfilled.

Visual local adaptation in P. minutus

All five tasks of our approach to demonstrate local adaptation on the rhodopsin gene in *P. minutus* have been fulfilled. Therefore, there is a **clear indication for visual local adaptation in the sand goby**. No matter how strong the indications for visual local adaptation are found in *P. minutus*, new approaches still have to validate these results. More research is recommended, including common-garden experiments to measure empirically relative fitness of individuals from different populations in an identical environment. Nevertheless, the observed pattern for rhodopsin is one of the strongest genetic signatures of natural selection yet reported in marine organisms. There are many signals of natural selection in marine organisms but it is usually difficult to know what was being selected for (Bergstrom & Kautsky 2006; Pampoulie *et al.* 2006; Hemmer-Hansen *et al.* 2007; Sherman & Ayre 2008). Therefore this rhodopsin study suggests that local adaptation is possible in the sea, and that selective forces may explain function and shape patterns in the ocean. Moreover, the rhodopsin case provides valuable insights into the evolutionary processes in

marine fish about the molecular basis of adaptation, the speed of selection and the mode of selection in space and time.

The rhodopsin study shows that the ***molecular basis of adaptation*** of proteins can be ground on AA replacements at a small number of critical sites. In *P. minutus*, maximally five AAs of *RH1* were involved in the genetic mechanism to adapt to the local dim light environment. One observed AA substitution on AA261 is even assigned to a λ_{max}-shift of 5.7 nm on the retinal rods in the northern Baltic sand gobies (*subchapter 4a*). Rhodopsin can therefore be added to the list of major functional shifts that are caused by minor AA changes. Similar cases are the hemoglobin gene in the Andean goose (Jessen *et al.* 1991) and the ribonuclease gene in the douc langur monkey (Zhang *et al.* 2002). Those examples are based on information derived from protein phylogeny, structure and genetic engineering (mutagenic experiments), whose combination provides deeper insights into the structure-function relationship of evolving proteins.

The visual local adaptation of northern Baltic gobies is a marked occasion to study the ***speed*** with which a genetic mutation can be favored under conditions of strong natural selection in the marine environment. Gobies immigrated the Baltic Sea during its last truly marine (Littorina) phase, starting 8000 years ago and gradually developing into the present-day brackish water environment (Björck 1995; Johannesson & André 2006). Within this short period, northern Baltic sand gobies became locally adapted to the specific photic conditions of their environment (*chapter 4*). The high speed of local adaptation suggests that sand gobies adapted by selection on standing genetic variation instead of *de novo* mutations. Standing variation leads to more rapid evolution in novel environments because it is available immediately at the time that selective conditions change, whereas more time is required for a new beneficial mutation to arise (Barrett & Schluter 2008). Nevertheless, in addition to adaptation to standing variation, indications were also found for a single *de novo* mutation with a major phenotypic effect (Phe261Tyr) in the northern Baltic populations (*chapter 4, 5*). Therefore, rhodopsin demonstrated the possibility of a fast rate of evolutionary changes in marine organisms with selection on standing as well as on *de novo* variation. Adaptation at new mutations with a positive effect on the fitness of marine organism can occur at a rate comparable with terrestrial species. A good example with a similar timescale

of 7000 years, is the human lactase persistence adaptation in Europe, where natural selection would have favoured anyone in the European population with a mutation that kept the lactase gene switched on. The lactose mutant provides the ability to digest milk in adulthood (Tishkoff *et al.* 2007).

Finally, the rhodopsin gene provided a good opportunity to study the ***mode of selection*** in the marine environment. The results of *chapter 4* showed that sand goby populations living under the same dim light conditions will adapt similarly on the rhodopsin gene. First, the Mediterranean and northern Baltic sand gobies acquired the same strategy to adapt to their more reddish light environment (*subchapter 4a*). Based on the historical pattern (*chapter 3*), the populations realised this specific adaptation independently from each other. It is a clear example of **convergent evolution** due to strong selective pressure resulting from the specific photic environment in brackish water. Convergent adaptation on the rhodopsin gene has already been observed in the East African cichlids. Analysis of the rhodopsin gene of the Lake Malawi (clear water, short wavelength-rich) and Lake Victoria (turbid water, long wavelength-rich) cichlids has revealed that species living in the same environments resorted to the same genetical solution in spite of their phylogenetic correlation (Spady *et al.* 2005; Trezise & Collin 2005). Next, results of *subchapter 4b* showed that similar selection pressure in the North Sea and Kattegat shaped analogous functional variation on *RH1*. All populations of the North Sea and Kattegat displayed the same strategy to the heterogeneous light conditions in this region. Therefore, the lack of differentiation on functional variation of the *RH1* gene between different genetic populations in space and time clearly points to **stabilizing selection** on a micro-scale.

The importance of studying the spectral tuning mechanism for aquatic species has been emphasized recently when empirical evidence revealed that divergent selection on sensory genes can even promote speciation through sensory drive in cichlid fishes because they are likely to affect directly mate preferences (Seehausen *et al.* 2008). However, no matter how crucial the light sensitivity is for the fitness of an animal, it is just one particular aspect in addition to e.g. temperature and salinity tolerance in shaping the adaptive divergence between populations and the speciation of organisms.

Consequences of visual local adaptation

The main consequence of the evolutionary adaptation to the light environment is that rapid changes in optical habitat by human activities may be negative for the sand goby (*chapter 4*). Many aquatic environments have become increasingly eutrophic as a result of pollution and climate change (Archer *et al.* 2001; Philippart *et al.* 2007). Eutrophication leads to increased growth of planktonic algae. As a consequence, both turbidity and siltation of organic material increases, leading to a decrease in light intensity and a narrowing of the light spectrum in the water. If marine fishes do not have the possibility to adapt physiologically to the new conditions with their rod and cone opsins, it will affect their search for food and the visual breeding system of fishes (Järvenpää & Lindström 2004; Engström-Öst & Candolin 2007; Candolin *et al.* 2008). Consequently, changes in water clarity must be considered in monitoring programs to evaluate marine ecosystem change (Archer *et al.* 2001; Aksnes 2007). The main focus has to be on those areas that suffer the most from increasing human pressure, such as the Mediterranean lagoons (Bouchereau & Guelorget 1998; Bouchereau 2001) and the highly polluted Baltic Sea (Järvenpää & Lindström 2004). Because sand gobies of these regions have become adapted to their local photic environment in an evolutionary timescale, rapid changes will clearly have a (temporally) negative effect (*chapter 3*).

The description of the demography (genetic drift and gene flow) of wild populations has been the primary focus of population and conservation genetics over the last number of decades (Beaumont 2005). However, the rhodopsin case shows that adaptive variation can also be a significant component of intraspecific biodiversity in marine species. Patterns of differentiation at selected and neutral markers can be strikingly different. Frankham & Reed (2001) concluded that molecular measurements of genetic diversity on neutral markers only have a very limited ability to predict adaptive genetic variation. Thus, neutral genetic data cannot serve as a surrogate of adaptive genetic data due to the fact that neutral markers may fail to recognize locally adapted populations (Holderegger *et al.* 2006). This means that current marine management practices could misidentify management units, leading to erosion of genetic resources. Moreover, knowledge on local adaptation in marine populations is crucial in order to predict if depleted or extinct populations can be effectively replaced by recolonization from other populations on a historic time-scale (ICES 2005).

Therefore it is required that also genetic information of locally adapted traits is taken into account to manage marine stocks.

Perspectives

The present study has answered our research question but created at the same time new questions and opportunities. Therefore, to conclude this discussion we will list recommendations for further research.

The direct **link between genotype and phenotype** for the rhodopsin pigment has to be further characterized. *Chapter 4 and 5* suggested potential tuning sites of opsin genes based on actual polymorphism data. The phenotypic effects of amino acid substitutions on these sites could be validated using microspectrophotometric studies and mutagenic experiments. Most mutations considered in biochemical analyses to elucidate the spectral tuning mechanism are not found in nature, and their significance in the actual λ_{max} shifts of visual pigments is not necessarily clear (Yokoyama 2000; Park *et al.* 2008). Therefore, the observed polymorphisms on rhodopsin of *P. minutus* and 'sand goby' group can be applied to understand better the functioning and physiology of visual pigments and of G-protein-coupled receptors in general.

To understand better the link between genotypic and phenotypic variation, it is essential to study as well the **link between genotype and gene expression**. The results of Larsen *et al.* (2007) indicated that local adaptation in gene expression is common in marine organisms. At this moment, it is not clear what is happening in individuals with two different rhodopsin haplotypes, a status that has been found in many sand goby individuals (*subchapter 4a, b*). Consequently, the expression of the different rhodopsin alleles may be differentiated between sand goby populations and may as well be part of the visual local adaptation in *P. minutus*. Different transcription factors are known that regulate the expression of opsin genes, such as atonal and pax proteins (Hofmeyer & Treisman 2008). Therefore it is important to study as well the expression level of rhodopsin to study visual local adaptation.

The phenotypic effect of the AA-substitutions in *P. minutus* and the 'sand goby' group will provide additional information for a better **link between genotype and the environmental conditions**. Nevertheless, this will only be the case when more information will be gained about the photic characteristics of marine systems and the habitat preferences of gobies. The WMTL map for the European seas that was realised in *subchapter 4a* shows spatial data with a resolution of 4 km and integrated for one year. This may be quite different from the spatio-temporal scale that is relevant for the feeding and adaptation processes in the sand goby. More ecological research has to be done on the horizontal and vertical spatial ranges of feeding for the sand goby. An interesting model to survey the link between rhodopsin variation and the photic environment are the lagoons in the northern Mediterranean Sea. Mediterranean sand gobies spawn in open sea and mature in the various lagoons, each differentiated in turbidity and spectral conditions (Pampoulie *et al.* 1999; Banas *et al.* 2005).

Additional experimental studies will be crucial in the future to establish a better **link between the genotypic variation at the *RH1* gene and selection regimes** that sand gobies encounter in the wild. In theory, common-garden experiments are very useful to prove the fitness relevance of the genetic differentiation. Common-garden experiments denote set-ups where individuals from different populations are reared in a common environment. This should eliminate all environmental variation and leave only genetically based differences between populations (ICES 2005). In practice, however, maternal effects pose a significant problem which must be accounted for, especially due to the problems to rear sand gobies (Christophe Pampoulie, personal communication). Moreover, it is difficult to evaluate the fitness relevance for populations adapted to different light regimes, encountered during different life stages. This is true for the sand gobies from the North Sea (*subchapter 4b*).

An additional research topic is the demonstration of local adaptation on **other opsin genes** in *P. minutus*. Since sand gobies have a similar tristimulus colour vision to that of humans (Jokela-Määttä *et al.* 2007), there is potential for high selection pressure on cone opsins as well. In *P. minutus*, colour vision is important for nest choice by males (Wong *et al.* 2008) and mate choice by females (Kangas 2000). During the spawning period, sand goby males

develop a dark blue coloration on the pectoral fin and a bright blue spot on the first dorsal fin (Figure 1.5). Females favour colourful males, which change the intensity of the blue colour on its anal fin (Forsgren 1992; Forsgren 1997; Kangas 2000). Because it is primarily blue light that is filtered out with increasing turbidity (Järvenpää & Lindström 2004; Michiels *et al.* 2008), it is hypothesised that also sand goby populations will be locally adapted on the blue cone opsin in particular and on the other cone opsins in general. Moreover, it will be interesting to gain insights into the interaction between the variants on the opsin genes and the **(sexual) behaviour of the sand goby**. The sand goby has a model species status in studies of sexual behaviour (Pampoulie *et al.* 2004b; Lindström *et al.* 2006; Singer *et al.* 2006) and future behaviour studies should take the genetic variation on rhodopsin and cone opsins into account to understand the variation in mate choice and sneaking behaviour.

To understand the general visual sensitivity of marine vertebrates, evolutionary adaptation of visual pigments **in other marine species** merits attention. The generality of the patterns observed in this thesis awaits confirmation. Is the pattern that we observed for *P. minutus* a unique or a rare strategy to adapt to different photic environments or a general pattern in marine vertebrates? Based on the results of *chapter 5* for *P. microps*, the focus has to lie on species with a broad range in habitat use throughout its range, as is the case for e.g. *P. canestrini* and *Gobiusculus flavescens* of the 'sand goby' group, and for other teleosts such as cod *Gadus morhua*, European flounder (*Platichthys flesus*) and Atlantic herring *Clupea harengus* (Jokela-Määttä *et al.* 2007).

The positive outcome for the rhodopsin gene shows the usefulness of the candidate gene approach for studying the characteristics of **local adaptation in marine non-model organisms** (Yokoyama 2002). The approach provided a very good opportunity to elucidate the mechanisms of adaptive evolution and selection in the ocean. More research on candidate genes with a direct link between the molecular variation and environmental conditions, such as sensory genes, is therefore warranted. Next, with the recent advent of genetic approaches in this system as genome scans, QTL-analyses and DNA microarrays (*chapter 1*), it should be possible to identify several genes that underlie many ecologically important traits in marine organisms. This will bring a new molecular dimension to the study of the scale and magnitude of local adaptation in the ocean (Colosimo *et al.* 2005).

Furthermore, knowledge of the genetic basis of local adaptation will add to our ability to manage biodiversity efficiently in relation to human impact such as exploitation, pollution, fishery and global warming. It will improve our ability to predict how the distribution and abundance of species will change in response to human-mediated evolutionary forces.

Finally, not withstanding a rich literature describing empirical studies of quantitative genetic variation and expectations from theoretical treatments of adaptation, we know little about the genetic basis of adaptation in nature. Which classes of genes vary among ecotypes, ecoclines and species? Are the same classes of genes involved? Does adaptation and speciation involve either few genes with large effects or many genes with small effects (Howe & Brunner 2005) and what is the effect of epigenetics and regulatory factors such as miRNA? The - in the present study - successful approach has a high potential to solve research questions **in other domains of biology**, inclusively biomedical science. Evolutionary biology remains not fully recognized as a crucial basic science for medicine, however, within the last decade several important advances have made it possible to study 'modern' diseases from an evolutionary perspective based on candidate gene studies (Ding & Kullo 2009). The human body and its pathogens are evolving biological systems shaped by selection under the constraints of tradeoffs that produces specific compromises and vulnerabilities. Therefore, powerful insights from evolutionary biology generate new questions whose answers will help improve human health (Nesse *et al.* 2006). Therefore, the presented framework incorporating phylogeographical and phylogenic information into a candidate gene study will also generate important insights into the origin and expression of human diseases (Nielsen *et al.* in press).

SCIENTIFIC SUMMARY

Under natural selection, individuals tend to adapt to local environmental conditions, resulting in a pattern of **local adaptation**. Understanding the genetic basis of local adaptation is of prime interest in biology as it involves the role of natural selection in promoting evolutionary change. Since gene flow is thought to hamper adaptive population divergence, the established idea was that local adaptation might be rare or even absent in marine organisms. Today, local adaptation has become increasingly documented, indicating that natural selection is a potent evolutionary force in the 'open' ocean. Knowledge on adaptive evolution in marine organisms remains limited, yet crucial to improve our understanding of how evolution functions in the ocean. The aim of this thesis is to contribute to the knowledge on the importance of natural selection as evolutionary force in marine organisms and on the spatio-temporal scale of adaptive genetic variation in marine systems.

The reason for the lack of knowledge on adaptive evolution in the ocean is the shortage of suitable genetic systems to evaluate natural selection. One of the few promising models to elucidate the mechanism and importance of selection as evolutionary force is the spectral tuning mechanism of the **visual pigments** (VP) in marine vertebrates. This model identifies amino acid changes that are responsible for adaptation to specific environments. Moreover, visual pigments have a well-defined role in nature as they detect differences in the spectral composition of the environment. Therefore they have a strong effect on the evolution of organisms, providing an excellent system to study adaptive evolution at the molecular level.

A candidate gene approach was followed to demonstrate local adaptation on the rhodopsin (*RH1*) gene, the VP located in the rods which determines the spectral sensitivity of dim-light vision. The sand goby *Pomatoschistus minutus* (Gobiidae, Teleostei), a common marine demersal fish along the European coasts, was selected as study model to realize. The specific research question of the thesis was: '**Is visual local adaptation detectable on the rhodopsin gene of the sand goby**?' An approach was adopted with five steps to demonstrate local adaptation in the wild.

First, a **phylogeographic and population genetic analysis** derived from neutral markers was realized for the sand goby based on high quality mitochondrial and microsatellite genotypes. An evolutionary scenario with temporally stable genetic structuring among populations was observed. It indicated that strong local selection would be able to override the effects of drift and gene flow, resulting in adaptive population divergence. Moreover, the robust phylogeographic and population genetic structure of the sand goby served as an appropriate neutral background to detect the effect of natural selection on the candidate gene.

Second, the genetic basis for measured phenotypic differences in the maximum absorbance (λ_{max}) of the retinal rods between sand goby populations was confirmed based on sequence analysis of the *RH1* gene. Five amino acid sites were polymorphic at *RH1*, with some known to be **spectral tuning sites**, sites with a significant effect on the λ_{max} values of vertebrate rods. Moreover, sand goby sites were differentiated from each other based on the functional variation at *RH1*, including on the spectral tuning sites.

Third, based on two independent approaches it was demonstrated that the differences between sand goby populations on the functional variation at *RH1* are due to natural selection and not due to neutral processes. Sequence-based neutrality tests detected unambiguously significant **signals of positive Darwinian selection** at the *RH1* gene. A comparison of the geographical distributions of the rhodopsin variation with the neutral marker-derived phylogeographic and population genetic structure revealed a signature of local selection on the rhodopsin gene.

Fourth, a link was established between the functional variation at the *RH1* gene and a **selection regime** that the sand goby populations experience. The solid differences in maximum environmental light transmittance between the various marine systems correspond well with the differences in the absorbance spectra of the retinal rods and the temporally stable differentiation on the functional variation at the *RH1* gene. Based on *RH1* variation sand gobies clustered according to the photic conditions of their habitat in-stead of historical or geographic proximity.

Finally, adaptive molecular evolution on the rhodopsin gene was demonstrated in a **phylogenetic framework** among related species of the sand goby (the 'sand goby' group) that inhabit different photic environments in regards to salinity and depth. With the observation that different species are evolutionary adapted to their photic habitat at the rhodopsin gene and that species with a similar selective regime have also an analogous genetic pattern at the *RH1* gene, the arguments are reinforced that selection on the candidate gene influences the contemporary population level in *P. minutus*.

All five steps have been fulfilled to detect local adaptation at the rhodopsin gene in *P. minutus*. Consequently, **clear evidence was provided for visual local adaptation in the sand goby**. The rhodopsin gene showed that the molecular basis of adaptation of proteins may occur by amino-acid replacements at a small number of critical sites. Moreover, it demonstrated also a relatively fast rate of evolutionary change in marine organisms with selection on *de novo* and on standing variation. No matter how strong the indications are for visual local adaptation in *P. minutus*, new approaches are required to validate these results. They include common-garden experiments to measure relative fitness of individuals from different populations in an identical environment. Nevertheless, the observed pattern for rhodopsin is one of the strongest genetic signatures of natural selection yet reported in marine organisms. The signals of natural selection are common in marine organisms but it is usually difficult to know what was being selected for. Therefore this rhodopsin study suggests that local adaptation is possible in the sea, and that selective forces may explain function and shape patterns in the ocean.

The main **consequence** of the evolutionary adaptation to light environment is that rapid changes in spectral habitat by human interference may be negative for the sand goby and other marine organisms. If marine fishes do not have the possibility to adapt physiologically to the new conditions, it will affect their visual breeding system and search for food. Consequently, changes in water clarity must be considered in monitoring programs to evaluate marine ecosystem change. Subsequently, the rhodopsin case shows that adaptive variation can also be a significant component of intraspecific biodiversity. Since neutral genetic data often fail to recognize locally adapted populations, this means that marine management by using only neutral genotypes, might misidentify management units, leading

to erosion of genetic resources. Therefore it is essential that genetic information of locally adapted traits is taken into account to manage marine stocks.

Finally, the rhodopsin study stimulates further research to find additional evidence for local adaptation to other marine environmental conditions such as temperature and salinity tolerance. It will contribute to more insights on the importance of natural selection as a common evolutionary force that shapes life in the ocean.

WETENSCHAPPELIJKE SAMENVATTING

Lokale adaptatie is het proces waarbij individuen zich genetisch aanpassen aan lokale omgevingscondities als gevolg van natuurlijke selectie. Het begrijpen van de moleculaire basis van lokale adaptatie is biologisch van groot belang aangezien dit de rol van natuurlijke selectie in evolutie weergeeft. Daar genmigratie adaptieve divergentie tegengaat door zijn homogeniserend karakter, beweerden biologen voor lange tijd dat lokale adaptatie zeldzaam is of zelfs helemaal afwezig in mariene organismen. Lokale adaptatie wordt steeds vaker beschreven in mariene organismen, wat een aanwijzing is dat natuurlijke selectie wel degelijk een belangrijke evolutionaire kracht kan zijn voor het leven in zee. De kennis van adaptieve evolutie in mariene organismen blijft beperkt ondanks het belang om het proces van evolutie in de zee te begrijpen. Het doel van deze thesis bestond erin om bij te dragen aan het huidig onderzoek over de sterkte van natuurlijke selectie als evolutionaire kracht in mariene organismen en over de spatiële en temporele schaal van adaptieve genetische variatie in mariene systemen.

De reden voor het huidig gebrek aan kennis aangaande adaptieve evolutie in de zee is de schaarste aan genetische systemen om natuurlijke selectie te evalueren. Een van de weinige veel belovende modellen om de werking en het belang van selectie als evolutionaire kracht te begrijpen, is het afstemmingsmechanisme van de **visuele pigmenten** (VP) aan verschillende lichtcondities in mariene vertebraten. Dit model identificeert specifieke aminozuur-veranderingen die verantwoordelijk zijn voor de adaptatie aan specifieke omgevingsomstandigheden. Daarenboven hebben VPs een duidelijke rol in de natuur aangezien ze dieren toelaten verschillen in de fotische compositie van de omgeving te detecteren. Daardoor hebben VPs een sterk effect op de evolutie van organismen en zijn ze uitermate geschikt om adaptieve evolutie te bestuderen op moleculair niveau.

In deze thesis werd een kandidaatgenanalyse gerealiseerd om lokale adaptatie aan te tonen op het rodopsine (*RH1*) gen, het gen van het visuele pigment in de staafjes dat de spectrale sensitiviteit bepaald van het zicht in schemerlicht. Het dikkopje *Pomatoschistus minutus* (Gobiidae, Teleostei), een algemeen voorkomend bodemvis langs de Europese kusten, werd geselecteerd als studiesoort om de objectieven van de thesis te realiseren. De specifieke

onderzoeksvraag luidt daarom: '**Is visuele lokale adaptatie aanwijsbaar op het rodopsine gen van het dikkopje**?'. Een specifieke benadering met vijf opeenvolgende stappen om lokale adaptatie aan te tonen in de natuur werd ontwikkeld om een antwoord te bieden op deze onderzoeksvraag.

Eerst werd een **fylogeografische en populatie genetische analyse** van het dikkopje gerealiseerd met hoog kwalitatieve genotypes van neutrale merkers, namelijk mitochondriale en microsatelliet merkers. Een evolutionair scenario met een complexe populatiestructuur die temporeel stabiel is werd waargenomen voor *P. minutus*. Deze observatie toont de mogelijkheid aan dat sterke lokale selectie de effecten van drift en genmigratie kan overtreffen in deze soort. Daarenboven is de robuuste fylogeografische en populatiegenetische structuur een goede neutrale basis om de adaptieve populatie differentiatie op het kandidaatgen te identificeren.

In de tweede stap werd de genetische basis bevestigd van de fenotypische verschillen in de maximale absorptie (λ_{max}) van de staafjes tussen populaties van dikkopjes. Dit werd gerealiseerd door een uitgebreide sequentie analyse van het *RH1* gen. Vijf aminozuren op *RH1* waren polymorf; twee zijn gekend als cruciale aminozuren die de λ_{max} waarden van de staafjes in vertebraten kunnen wijzigen. Bovendien werd vastgesteld dat de populaties van dikkopjes in de verschillende Europese mariene systemen van elkaar verschillen op basis van de **functionele variatie** van *RH1*.

Ten derde werd aangetoond op basis van twee onafhankelijke benaderingen dat de verschillen tussen de populaties voor de functionele variatie op het *RH1* gen beïnvloed zijn door natuurlijke selectie en niet alleen door neutrale processen. Sequentie gebaseerde neutraliteits testen toonden een ondubbelzinnig en significant signaal aan van **positieve Darwiniaanse selectie** op het *RH1* gen. Daarnaast toonde de vergelijking van de geografische verspreiding van de rodopsine variatie met de fylogeografische en populatiegenetische structuur van de dikkopjes een signaal aan van natuurlijke selectie op *RH1*.

In de vierde stap werd een duidelijk verband aangetoond tussen de functionele variatie op het *RH1* gen en het **selectie regime** dat de populaties van dikkopjes ervaren in de verschillende mariene systemen. De robuuste verschillen in de lokale maximale licht transmissie tussen de mariene systemen zijn in overeenstemming met de verschillen in de absorptie-spectra van de staafjes en met de temporeel stabiele verschillen op de functionele variatie van het *RH1* gen. Het is duidelijk dat op basis van de *RH1* variatie de populaties van *P. minutus* zich groeperen volgens de fotische condities van hun habitat en niet volgens historische en geografische verwantschap.

In de laatste stap werd d.m.v. een **fylogenetische analyse** adaptieve moleculaire evolutie op het rodopsine gen aangetoond tussen verwante soorten van het dikkopje die in verschillende fotische omgevingen voorkomen op basis van saliniteit en diepte. Doordat deze verwante soorten evolutionair aangepast zijn aan hun fotische habitat op het *RH1* gen en dat soorten met een gelijkaardig selectie-regime analoge genetische patronen vertonen op het *RH1* gen, zijn de argumenten versterkt dat selectie op het kandidaatgen ook de huidige populatiestructuur van *P. minutus* beïnvloedt.

Uiteindelijk zijn alle vijf stappen vervuld om lokale adaptatie op het rodopsine gen aan te tonen. Daardoor is er een duidelijke aanwijzing voor **visuele lokale adaptatie in het dikkopje**. De rodopsine studie bij het dikkopje toont aan dat de moleculaire basis van adaptatie van proteïnen kan voorkomen door aminozuur vervangingen op een beperkt aantal belangrijke plaatsen in het gen. Daarenboven toont het ook de relatief hoge snelheid aan van evolutionaire veranderingen via selectie op zowel nieuwe als permanente variatie. Hoe sterk de indicaties ook zijn voor visuele lokale adaptatie in *P. minutus*, nieuwe analyses moeten deze resultaten valideren, waaronder transplantatie experimenten om de relatieve fitness van individuen van verschillende populaties in een zelfde omgeving te meten. Niettemin is het geobserveerde patroon voor rodopsine in de huidige stand van de wetenschap een van de sterkste genetische aanwijzingen voor natuurlijke selectie bij mariene organismen. Verscheidene signalen van natuurlijke selectie waren reeds eerder waargenomen in mariene organismen maar het was steeds moeilijk om te weten waarvoor geselecteerd werd. Deze rodopsine studie suggereert dan ook dat lokale adaptatie in zee mogelijk is en dat selectie de functie en de patronen in de zee kan verklaren.

Een belangrijk **gevolg** van de evolutionaire adaptatie aan de lichtomgeving is dat snelle veranderingen in de mariene fotische omstandigheden door menselijke activiteiten, negatieve gevolgen kunnen hebben voor het dikkopje en andere organismen. Als mariene vissen niet de mogelijkheid hebben om zich fysiologisch aan te passen aan nieuwe condities, dan kan dit de zoektocht naar voedsel en de voortplanting sterk beïnvloeden. Bijgevolg is het belangrijk om veranderingen in waterhelderheid te beschouwen in de huidige opvolgingsprogramma's die mariene ecosysteem-veranderingen in kaart brengen. Daarenboven toont de rodopsine studie aan dat adaptieve variatie een belangrijk onderdeel is van de intraspecifieke biodiversiteit in mariene soorten. Aangezien neutrale genetische data vaak falen in de aanduiding van lokaal aangepaste populaties, betekent dit dat het huidige mariene beheer, dat nog grotendeels gebruik maakt van neutraal genetische data, verkeerde beheerseenheden kan identificeren, met een erosie van de genetische variatie als gevolg. Daardoor is het noodzakelijk dat ook genetische informatie van adaptieve kenmerken wordt betrokken bij het beheer van mariene visbestanden.

Ten slotte zijn al deze vaststellingen een aanmoediging om analoge studies te verrichten naar de mogelijkheid van lokale adaptatie aan andere mariene omgevingsfactoren zoals het zoutgehalte en de temperatuur. Zo'n onderzoek kan de algemeenheid en intensiteit verduidelijken in welke mate natuurlijke selectie de evolutie van het leven in zee stuurt.

POPULAR SUMMARY

Exactly 150 years ago, Charles Darwin described natural selection as the motor of the evolution of life. Nevertheless, it is not yet clear how important natural selection is for the evolution of marine organisms. The genetic adaptation to local environmental conditions as a result of natural selection, a process known as local adaptation, will be reduced by the migration of organisms due to its homogeneous character. Because of the huge potential for migration in the 'open' sea, for a long time biologists declared that local adaptation is rare and even absent.

Nevertheless, current research shows that the sea is not as 'open' as it may seem. Many marine organisms are able to occupy a permanent place and hence occur in distinct populations. Since migration seems limited, the possibility of local adaptation in marine species presents an important research question. The most recent studies showed that natural selection might be an important evolutionary force in the ocean, however without any good scientific evidence.

The present thesis has the ambition to prove that marine species may indeed be genetically adapted to local conditions. A promising opportunity is the possibility for local adaptation to the light regime of the see. The light that organisms perceive varies between seas due to the differences in turbidity and the colour of the water. The importance of sight for marine animals is obvious, especially to find food and mates, and to avoid predators. Therefore, the aim of the thesis was to study local adaptation at the rhodopsin gene - the gene of the visual pigment that determines the visual capacity in dim-light - of a marine goby, the sand goby (*Pomatoschistus minutus*). The sand goby is a small and abundant fish species that lives along the European coasts.

The results showed strong evidence that sand goby populations are genetically adapted to their specific and local light environment. They are adapted to high turbidity in the Baltic Sea and the Mediterranean lagoons, and to the more blue light of the Bay of Biscay and along the coasts of Spain and Portugal. Moreover, the sand gobies of the North Sea reveal a strategy of adaptation to the unstable local light conditions. In the current state of science, the

rhodopsin gene provides one of the strongest indications that local adaptation occurs in the marine environment. Therefore, they encourage analogous studies to find further evidence for local adaptation to other marine environmental conditions such as salinity tolerance and temperature. Such studies will clarify the importance of natural selection as evolutionary force for marine life.

To conclude, this study reveals that the sand goby is evolutionary adapted to its light environment. There are strong indications that if the light environment changes due to either pollution or climate change, marine fishes won't likely be able to adapt rapidly to the new circumstances. Good management of the light conditions of the marine ecosystem will be essential to support a balanced ecosystem and healthy fish stocks.

POPULAIRE SAMENVATTING

Charles Darwin beschreef 150 jaar geleden natuurlijke selectie als motor van de evolutie van het leven. Anno 2009 blijft er onduidelijkheid in welke mate natuurlijke selectie invloed heeft op de evolutie van mariene organismen. De genetische aanpassing aan lokale omgevingsomstandigheden als gevolg van natuurlijke selectie, lokale aanpassing genoemd, wordt door de migratie van de organismen afgeremd. Aangezien de migratiemogelijkheden groot worden geacht voor het leven in de 'open' zee, verklaarden biologen voor lange tijd dat lokale aanpassing zeldzaam of zelfs uitgesloten was.

Huidig onderzoek toont echter aan dat de zee niet zo 'open' is zoals op het eerste zicht lijkt. Heel wat mariene organismen zijn aan een vaste plaats gebonden en komen voor onder de vorm van verschillende populaties. Aangezien de mate van migratie gering blijkt te zijn, vormt de mogelijkheid van lokale aanpassing bij zee-organismen een actuele onderzoeksvraag. Recente studies toonden immers aan dat er inderdaad een grote kans is dat lokale selectie de evolutie van het zeeleven kan beïnvloeden. Het bleek echter moeilijk om dit wetenschappelijk te onderbouwen.

Deze thesis heeft als doel duidelijke aanwijzingen te verschaffen dat mariene soorten zich genetisch aanpassen aan plaatselijke omgevingsfactoren. Een veel belovende aanpak is de genetische aanpassing aan de lokale lichtomstandigheden in zee te bestuderen. Het licht dat mariene dieren waarnemen varieert immers door verschillen in de troebelheid en kleur van het water. Uiteraard is het belang van een goed zicht uitermate groot om o.a. voedsel en geschikte partners te vinden, en om roofdieren te ontwijken. Daarom werd in deze thesis de mogelijkheid van lokale aanpassingen op het rodopsine gen - het gen van het visuele pigment dat instaat voor schemerzicht - onderzocht in een mariene grondel, het dikkopje (*Pomatoschistus minutus*). Het dikkopje is een klein bodemvisje dat leeft langs de kusten van de Europese zeeën en is de meest voorkomende vissoort in de Noordzee.

De resultaten van deze thesis tonen overtuigend aan dat de verschillende populaties van dikkopjes genetisch aangepast zijn aan de lokale lichtomstandigheden. Zo blijken dikkopjes goed aangepast te zijn aan de hoge troebelheid van de Baltische Zee en de lagunes van de

Middellandse Zee, en aan het meer blauwe licht in de Golf van Biskaje en langs de kusten van Spanje en Portugal. De dikkopjes van onze Noordzee houden er dan weer een aparte strategie op na om zich aan de plaatselijke onstabiele lichtomstandigheden aan te passen. In de huidige kennis van de wetenschap is daarom met het rodopsine gen in het dikkopje één van de sterkste aanwijzingen gevonden dat lokale aanpassing ook in zee kan voorkomen. Deze vaststellingen zijn daarom een aanmoediging om gelijkaardige studies te verrichten naar lokale aanpassing aan andere mariene omgevingsfactoren zoals het zoutgehalte en de temperatuur. Dit verder onderzoek kan dan de mate bepalen in hoeverre natuurlijke selectie de evolutie van het leven in zee stuurt.

Ten slotte wijst dit onderzoek erop dat dikkopjes evolutionair en dus op lange termijn zijn aangepast aan het licht van hun omgeving. Daardoor zijn er sterke aanwijzingen dat indien de omgeving op snelle wijze zou veranderen door vervuiling of door de opwarming van de aarde, mariene vissen zich moeilijk of zelfs niet zullen aanpassen aan de nieuwe omstandigheden waardoor hun verdere bestaan in gevaar komt. Om een evenwichtlig ecosysteem en gezonde zeevispopulaties mogelijk te maken moeten daarom snelle veranderingen in de lichtcondities van de Europese zeeën vermeden worden.

REFERENCES

Aarnio K. & Bonsdorff E. (1993) Seasonal-variation in abundance and diet of the sand goby *Pomatoschistus minutus* (Pallas) in a northern Baltic archipelago. *Ophelia*, **37**, 19-30.

Aboim M.A., Menezes G.M., Schlitt T. & Rogers A.D. (2005) Genetic structure and history of populations of the deep-sea fish *Helicolenus dactylopterus* (Delaroche, 1809) inferred from mtDNA sequence analysis. *Molecular Ecology*, **14**, 1343-1354.

Adkinson M. (1995) Population differentiation in Pacific salmon: local adaptation, genetic drift or the environment? *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 2762-2777.

Aho A.C., Donner K., Hyden C., Larsen L.O. & Reuter T. (1988) Low retinal noise in animals with low body-temperature allows high visual sensitivity. *Nature*, **334**, 348-350.

Aksnes D.L. (2007) Evidence for visual constraints in large marine fish stocks. *Limnology and Oceanography*, **52**, 198-203.

Ala-Laurila P., Donner K. & Koskelainen A. (2004) Thermal activation and photoactivation of visual pigments. *Biophysical Journal*, **86**, 3653-3662.

Alcaide M., Edwards S.V., Negro J.J., Serrano D. & Tella J.L. (2008) Extensive polymorphism and geographical variation at a positively selected MHC class IIB gene of the lesser kestrel (*Falco naumanni*). *Molecular Ecology*, **17**, 2652-2665.

Al-Hassan L.A.J., Webb C.J., Giama M. & Miller P.J. (1987) Phosphoglucose isomerase polymorphism in the common goby, *Pomatoschistus microps* (Krøyer) (Teleostei, Gobiidae), around the British-Isles. *Journal of Fish Biology*, **30**, 281-298.

Allendorf F.W. & Seeb L.W. (2000) Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution*, **54**, 640-651.

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.

Amos W., Hoffman J.I., Frodsham A., Zhang L., Best S. & Hill A.V.S. (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, **7**, 10-14.

Amundsen T. & Forsgren E. (2001) Male mate choice selects for female coloration in a fish. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 13155-13160.

Anisimova M., Nielsen R. & Yang Z.H. (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics*, **164**, 1229-1236.

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

Appleyard S.A., Ward R.D. & Williams R. (2002) Population structure of the Patagonian toothfish around Heard, McDonald and Macquarie Islands. *Antarctic Science*, **14**, 364-373.

Aranishi F. (2005) Rapid PCR-RFLP method for discrimination of imported and domestic mackerel. *Marine Biotechnology*, **7**, 571-575.

Archer S., Hirano J. & Vallerga S. (2001) The effects of variability of marine ecosystems on the visual sensitivity of fish. *Archo Oceanography Limnology*, **22**, 119-126.

Archer S., Hope A. & Partridge J.C. (1995) The molecular basis for the green-blue sensitivity shift in the rod visual pigments of the European eel. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **262**, 289-295.

Archer S.N., Lythgoe J.N. & Hall L. (1992) Rod opsin cDNA sequence from the sand goby (*Pomatoschistus minutus*) compared with those of other vertebrates. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **248**, 19-25.

Arruda L.M., Azevedo J.N. & Neto A.I. (1993) Abundance, age-structure and growth, and reproduction of gobies (Pisces Gobiidae) in the Ria de Aveiro lagoon (Portugal). *Estuarine Coastal and Shelf Science*, **37**, 509-523.

Asenjo A.B., Rim J. & Oprian D.D. (1994) Molecular determinants of human red/green color discrimination. *Neuron*, **12**, 1131-1138.

Audzijonyte A., Pahlberg J., Väinölä R. & Lindstrom M. (2005) Spectral sensitivity differences in two Mysis sibling species (Crustacea, Mysida): Adaptation or phylogenetic constraints? *Journal of Experimental Marine Biology and Ecology*, **325**, 228-239.

Avise J.C. (2000) *Phylogeography: the history and formation of species.* Harvard University Press, Cambridge, MA.

Ballard J.W.O. & Whitlock M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729-744.

Balloux F. (2001) EASYPOP (version 1.7): A computer program for population genetics simulations. *Journal of Heredity*, **92**, 301-302.

Balloux F., Brunner H., Lugon-Moulin N., Hausser J. & Goudet J. (2000) Microsatellites can be misleading: An empirical and simulation study. *Evolution*, **54**, 1414-1422.

Balloux F. & Goudet J. (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology*, **11**, 771-783.

Balloux F. & Lugon-Moulin N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155-165.

Bamshad M. & Wooding S.P. (2003) Signatures of natural selection in the human genome. *Nature Reviews Genetics*, **4**, 99-111A.

Banarescu P. (1992) *Zoogeography of fresh waters. Distribution and dispersal of fresh water animals in North America and Eurasia.* Aula-Verlag, Wiesbaden.

Banas D., Grillas P., Auby I., Lescuyer F., Coulet E., Moreteau J.C. & Millet B. (2005) Short time scale changes in underwater irradiance in a wind-exposed lagoon (Vaccarès lagoon, France): efficiency of infrequent field measurements of water turbidity or weather data to predict irradiance in the water column. *Hydrobiologia*, **551**, 3-16.

Barrett R.D.H. & Schluter D. (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38-44.

Beatty D.D. (1966) A study of succession of visual pigments in Pacific salmon (*Oncorhynchus*). *Canadian Journal of Zoology*, **44**, 429-455.

Beatty D.D. (1975) Visual pigments of American eel *Anguilla rostrata*. *Vision Research*, **15**, 771-776.

Beaumont M.A. (2005) Adaptation and speciation: what can F_{st} tell us? *Trends in Ecology & Evolution*, **20**, 435-440.

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

Beebee T. & Rowe G. (2004) *An introduction to molecular ecology.* Oxford University Press, Oxford.

Begon M., Townsend C.R. & Harper J.L. (2006) *Ecology: from individuals to ecosystems*, 4 edn. Wiley-Blackwell.

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (2004) *GENETIX 4.05, logiciel sous windows™ pour la génétique des populations.* Université de Montpellier II, Montpellier (France).

Benaka L.R. (1999) Fish habitat: Essential fish habitat and rehabilitation. In: *American Fisheries Society Symposium 22* (ed. Bethesda M.D.), p. 454 pp.

Bentzen P. (1998) Seeking evidence of local stock structure using molecular genetic methods. In: *The implications of localized fisheries stocks* (eds. Hunt von Herbing I., Kornfield I., Tupper M. & Wilson J.), pp. 20-30. Regional Agricultural Engineering Service, New York.

Bergstrom L. & Kautsky L. (2006) Local adaptation of *Ceramium tenuicorne* (Ceramiales, Rhodophyta) within the Baltic Sea. *Journal of Phycology*, **42**, 36-42.

Bernatchez L. & Danzmann R.G. (1993) Congruence in control region sequence and restriction site variation in mitochondrial DNA of brook charr (*Salvelinus fontinalis* Mitchill). *Molecular Biology and Evolution*, **10**, 1002-1014.

Bernatchez L., Glemet H., Wilson C.C. & Danzmann R.G. (1995) Introgression and fixation of Arctic charr (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook rout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 179-185.

Bernatchez L. & Wilson C.C. (1998) Comparative phylogeography of nearctic and palearctic fishes. *Molecular Ecology*, **7**, 431-452.

Berrebi P., Lasserre B., Barbisan F. & Zane L. (2006) Isolation of microsatellite loci and cross-species amplifications in three gobiid fish of the genus *Pomatoschistus*. *Molecular Ecology Notes*, **6**, 724-727.

Beyst B., Hostens K. & Mees J. (2001) Factors influencing fish and macrocrustacean communities in the surf zone of sandy beaches in Belgium: temporal variation. *Journal of Sea Research*, **46**, 281-294.

Beyst B., Mees J. & Cattrijsse A. (1999) Early postlarval fish in the hyperbenthos of the Dutch delta (south-west Netherlands). *Journal of the Marine Biological Association of the United Kingdom*, **79**, 709-724.

Bickford D., Lohman D.J., Sodhi N.S., Ng P.K.L., Meier R., Winker K., Ingram K.K. & Das I. (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148-155.

Björck S. (1995) A review of the history of the Baltic sea, 13.0-8.0 Ka Bp. *Quaternary International*, **27**, 19-40.

Bouchereau J.L. (2001) Le gobie buhotte *Pomatoschistus minutus* (Pallas, 1770). In: *Atlas des poissons d'eau douce de France* (eds. Keith P. & Allardi J.), pp. 336-337. Publications Scientifiques du Muséum National d'Histoire Naturelle, Paris.

Bouchereau J.L. & Guelorget O. (1998) Comparison of three Gobiidae (Teleostei) life history strategies over their geographical range. *Oceanologica Acta*, **21**, 503-517.

Bowen B.W., Bass A.L., Rocha L.A., Grant W.S. & Robertson D.R. (2001) Phylogeography of the trumpetfishes (*Aulostomus*): Ring species complex on a global scale. *Evolution*, **55**, 1029-1039.

Bowmaker J.K. (1995) The visual pigments of fish. *Progress in Retinal and Eye Research*, **15**, 1-31.

Bowmaker J.K. (2008) Evolution of vertebrate visual pigments. *Vision Research*, **48**, 2022-2041.

Bowmaker J.K. & Hunt D.M. (2006) Evolution of vertebrate visual pigments. *Current Biology*, **16**, R484-R489.

Bremer J.R.A., Vinas J., Mejuto J., Ely B. & Pla C. (2005) Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular Phylogenetics and Evolution*, **36**, 169-187.

Briolay J., Galtier N., Brito R.M. & Bouvet Y. (1998) Molecular phylogeny of Cyprinidae inferred from cytochrome *b* DNA sequences. *Molecular Phylogenetics and Evolution*, **9**, 100-108.

Brix O., Forås E. & Strand I. (1998) Genetic variation and functional properties of Atlantic cod hemoglobins: introducing a modified tonometric method for studying fragile hemoglobins. *Comparative Biochemistry and Physiology*, **119A**, 575-583.

Brzustowski J. (2002) *Doh assignment test calculator*, http://www2.biology.ualberta.ca/jbrzusto/Doh.php.

Buonaccorsi V.P., McDowell J.R. & Graves J.E. (2001) Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Molecular Ecology*, **10**, 1179-1196.

Burton R.S. (1996) Molecular tools in marine ecology. *Journal of Experimental Marine Biology and Ecology*, **200**, 85-101.

Cabral H.N., Teixeira C.M., Gamito R. & Costa M.J. (2002) Importance of discards of a beam trawl fishery as input of organic matter into nursery areas within the Tagus estuary. *Hydrobiologia*, **475**, 449-455.

Calderón I., Giribet G. & Turon X. (2008) Two markers and one history: phylogeography of the edible common sea urchin *Paracentrotus lividus* in the Lusitanian region. *Marine Biology*, **154**, 137-151.

Candolin U., Engström-Öst J. & Salesto T. (2008) Human-induced eutrophication enhances reproductive success through effects on parenting ability in sticklebacks. *Oikos*, **117**, 459-465.

Canino M.F., O'Reilly P.T., Hauser L. & Bentzen P. (2005) Genetic differentiation in walleye pollock (*Theragra chalcogramma*) in response to selection at the pantophysin (*Pan I*) locus. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 2519-2529.

Cano J.M., Mäkinen H.S. & Merilä J. (2008) Genetic evidence for male-biased dispersal in the three-spined stickleback (*Gasterosteus aculeatus*). *Molecular Ecology*, **17**, 3234-3242.

Carlsson J., McDowell J.R., Diaz-Jaimes P., Carlsson J.E.L., Boles S.B., Gold J.R. & Graves J.E. (2004) Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Molecular Ecology*, **13**, 3345-3356.

Carreras-Carbonell J., Macpherson E. & Pascual M. (2006) Population structure within and between subspecies of the Mediterranean triplefin fish *Tripterygion delaisi* revealed by highly polymorphic microsatellite loci. *Molecular Ecology*, **15**, 3527-3539.

Carvalho G.R. (1993) Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology*, **43**, 53-73.

Cavalli-Sforza L.L. & Edwards A.W.F. (1967) Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics*, **19**, 233-257.

Céspedes A., García T., Carrera E., González I., Fernández A., Asensio L., Hernández P.E. & Martín R. (2000) Genetic differentiation between sole (*Solea solea*) and Greenland

halibut (*Reinhardtius hippoglossoides*) by PCR-RFLP analysis of a 12S rRNA gene fragment. *Journal of the Science of Food and Agriculture*, **80**, 29-32.

Charrier G., Chenel T., Durand J.D., Girard M., Quiniou L. & Laroche J. (2006) Discrepancies in phylogeographical patterns of two European anglerfishes (*Lophius budegassa* and *Lophius piscatorius*). *Molecular Phylogenetics and Evolution*, **38**, 742-754.

Chen W.J., Bonillo C. & Lecointre G. (2003) Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution*, **26**, 262-288.

Chevolot M., Hoarau G., Rijnsdorp A.D., Stam W.T. & Olsen J.L. (2006) Phylogeography and population structure of thornback rays (*Raja clavata* L., Rajidae). *Molecular Ecology*, **15**, 3693-3705.

Chistiakov D.A., Hellemans B. & Volckaert F.A.M. (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, **255**, 1-29.

Chow S., Suzuki N., Imai H. & Yoshimura T. (2006) Molecular species identification of spiny lobster phyllosoma larvae of the genus *Panulirus* from the northwestern Pacific. *Marine Biotechnology*, **8**, 260-267.

Clark A.G., Glanowski S., Nielsen R., Thomas P.D., Kejariwal A., Todd M.A., Tanenbaum D.M., Civello D., Lu F., Murphy B., Ferriera S., Wang G., Zheng X.G., White T.J., Sninsky J.J., Adams M.D. & Cargill M. (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science*, **302**, 1960-1963.

Clarke K.R. (1993) Nonparametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, **18**, 117-143.

Clement M., Posada D. & Crandall K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1659.

Cocolin L., D'Agaro E., Manzano M., Lanari D. & Corni G. (2000) Rapid PCR-RFLP method for the identification of marine fish fillets (seabass, seabream, umbrine, and dentex). *Journal of Food Science*, **65**, 1315-1317.

Colosimo P.F., Hosemann K.E., Balabhadra S., Villarreal G., Dickson M., Grimwood J., Schmutz J., Myers R.M., Schluter D. & Kingsley D.M. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928-1933.

Connors B.W. (2003) Sensory Transduction. In: *Medical Physiology* (eds. Boron W.F. & Boulpaep E.L.), pp. 325-358. Saunders.

Conover D.O., Clarke L.M., Munch S.B. & Wagner G.N. (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*, **69**, 21-47.

Consuegra S. & de Leaniz C.G. (2007) Fluctuating sex ratios, but no sex-biased dispersal, in a promiscuous fish. *Evolutionary Ecology*, **21**, 229-245.

Consuegra S., De Leaniz C.G., Serdio A., Morales M.G., Straus L.G., Knox D. & Verspoor E. (2002) Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology*, **11**, 2037-2048.

Cortey M. & Garcia-Marin J.L. (2002) Evidence for phylogeographically informative sequence variation in the mitochondrial control region of Atlantic brown trout. *Journal of Fish Biology*, **60**, 1058-1063.

Cowen R.K., Lwiza K.M.M., Sponaugle S., Paris C.B. & Olson D.B. (2000) Connectivity of marine populations: Open or closed? *Science*, **287**, 857-859.

Coyer J.A., Peters A.F., Stam W.T. & Olsen J.L. (2003) Post-ice age recolonization and differentiation of *Fucus serratus* L. (Phaeophyceae; Fucaceae) populations in Northern Europe. *Molecular Ecology*, **12**, 1817-1829.

Crescitelli F., McFall-Ngai M. & Horwitz J. (1985) The visual pigment sensitivity hypothesis - Further evidence from fishes of varying habitats. *Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology*, **157**, 323-333.

Dann S.G., Allison W.T., Levin D.B., Taylor J.S. & Hawryshyn C.W. (2004) Salmonid opsin sequences undergo positive selection and indicate an alternate evolutionary relationship in *Oncorhynchus*. *Journal of Molecular Evolution*, **58**, 400-412.

De Innocentiis S., Lesti A., Livi S., Rossi A.R., Crosetti D. & Sola L. (2004) Microsatellite markers reveal population structure in gilthead sea bream *Sparus auratus* from the Atlantic Ocean and Mediterranean Sea. *Fisheries Science*, **70**, 852-859.

De Innocentiis S., Sola L., Cataudella S. & Bentzen P. (2001) Allozyme and microsatellite loci provide discordant estimates of population differentiation in the endangered dusky grouper (*Epinephelus marginatus*) within the Mediterranean Sea. *Molecular Ecology*, **10**, 2163-2175.

Deagle B.E., Eveson J.P. & Jarman S.N. (2006) Quantification of damage in DNA recovered from highly degraded samples - a case study on DNA in faeces. *Frontiers in Zoology*, **3**, 11.

Debes P.V., Zachos F.E. & Hanel R. (2008) Mitochondrial phylogeography of the European sprat (*Sprattus sprattus* L., Clupeidae) reveals isolated climatically vulnerable populations in the Mediterranean Sea and range expansion in the northeast Atlantic. *Molecular Ecology*, **17**, 3873-3888.

DeWoody J., Nason J.D. & Hipkins V.D. (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, **6**, 951-957.

Ding K. & Kullo I.J. (2009) Evolutionary genetics of coronary heart disease. *Circulation*, **119**, 459-467.

Dolbeth M., Martinho F., Leitão R., Cabral H. & Pardal M.A. (2007) Strategies of *Pomatoschistus minutus* and *Pomatoschistus microps* to cope with environmental instability. *Estuarine Coastal and Shelf Science*, **74**, 263-273.

Domingues V.S., Bucciarelli G., Almada V.C. & Bernardi G. (2005) Historical colonization and demography of the Mediterranean damselfish, *Chromis chromis*. *Molecular Ecology*, **14**, 4051-4063.

Domingues V.S., Faria C., Stefanni S., Santos R.S., Brito A. & Almada V.C. (2007) Genetic divergence in the Atlantic-Mediterranean Montagu's blenny, *Coryphoblennius galerita* (Linnaeus 1758) revealed by molecular and morphological characters. *Molecular Ecology*, **16**, 3592-3605.

Donner K. (1992) Noise and the absolute thresholds of cone and rod vision. *Vision Research*, **32**, 853-866.

Doornbos G. & Twisk F. (1987) Density, growth and annual food-consumption of Gobiid fish in the saline lake Grevelingen, the Netherlands. *Netherlands Journal of Sea Research*, **21**, 45-74.

Drummond A.J. & Rambaut A. (2006) *BEAST*, 1.4 edn., http://beast.bio.ed.ac.uk.

Drummond A.J., Rambaut A., Shapiro B. & Pybus O.G. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185-1192.

Duggen S., Hoernle K., van den Bogaard P., Rupke L. & Morgan J.P. (2003) Deep roots of the Messinian salinity crisis. *Nature*, **422**, 602-606.

Ehrenberg S.Z. & Ejdung G. (2008) Daily activity pattern of the sand goby, *Pomatoschistus minutus* (Pisces), at low light intensity. *Hydrobiologia*, **603**, 129-137.

Ehrenberg S.Z., Hansson S. & Elmgren R. (2005) Sublittoral abundance and food consumption of Baltic gobies. *Journal of Fish Biology*, **67**, 1083-1093.

Elliott M. & Hemmingway K.L. (2002) *Fishes in estuaries*. Blackwell Science, London.

Endler J.A. (1986) *Natural selection in the wild*. Princeton University Press, Princeton.

Endo T., Ikeo K. & Gojobori T. (1996) Large-scale search for genes on which positive selection may operate. *Molecular Biology and Evolution*, **13**, 685-690.

England P.R., Cornuet J.M., Berthier P., Tallmon D.A. & Luikart G. (2006) Estimating effective population size from linkage disequilibrium: severe bias in small samples. *Conservation Genetics*, **7**, 303-308.

Engström-Öst J. & Candolin U. (2007) Human-induced water turbidity alters selection on sexual displays in sticklebacks. *Behavioral Ecology*, **18**, 393-398.

Estoup A., Jarne P. & Cornuet J.M. (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591-1604.

Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Ewens W.J. (1972) Sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87.

Excoffier L. (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853-864.

Excoffier L., Laval G. & Schneider S. (2005) ARLEQUIN ver.3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47-50.

Fasick J.I. & Robinson P.R. (1998) Mechanism of spectral tuning in the dolphin visual pigments. *Biochemistry*, **37**, 433-438.

Fernandez C.C., Shevock J.R., Glazer A.N. & Thompson J.N. (2006) Cryptic species within the cosmopolitan desiccation-tolerant moss *Grimmia laevigata*. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 637-642.

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

Fonds M. (1973) Sand gobies in the Dutch Wadden Sea (*Pomatoschistus*, Gobiidae, Pisces). *Netherlands Journal of Sea Research*, 417-478.

Ford M.J. (2002) Applications of selective neutrality tests to molecular ecology. *Molecular Ecology*, **11**, 1245-1262.

Forsgren E. (1992) Predation risk affects mate choice in a Gobiid fish. *American Naturalist*, **140**, 1041-1049.

Forsgren E. (1997) Mate sampling in a population of sand gobies. *Animal Behaviour*, **53**, 267-276.

Franch R., Louro B., Tsalavouta M., Chatziplis D., Tsigenopoulos C.S., Sarropoulou E., Antonello J., Magoulas A., Mylonas C.C., Babbucci M., Patarnello T., Power D.M., Kotoulas G. & Bargelloni L. (2006) A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L. *Genetics*, **174**, 851-861.

Freeman S. & Herron J.C. (2001) *Evolutionary analysis*, Second edition edn. Prentice-Hall, New Jersey.

Fu Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915-925.

Gaggiotti O.E., Lange O., Rassmann K. & Gliddon C. (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, **8**, 1513-1520.

Galindo J., Moran P. & Rolan-Alvarez E. (2009) Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology*, **18**, 919-930.

Gibson G. & Dworkin I. (2004) Uncovering cryptic genetic variation. *Nature Reviews Genetics*, **5**, 681-691.

Gibson R.N. & Ezzi I.A. (1981) The biology of the Norway goby, *Pomatoschistus norvegicus* (Collett), on the west-coast of Scotland. *Journal of Fish Biology*, **19**, 697-714.

Gibson R.N. & Hesthagen I.H. (1981) A comparison of the activity patterns of the sand goby *Pomatoschistus minutus* (Pallas) from areas of different tidal range. *Journal of Fish Biology*, **18**, 669-684.

Golding G.B. & Dean A.M. (1998) The structural basis of molecular adaptation. *Molecular Biology and Evolution*, **15**, 355-369.

Goldstein D.B., Linares A.R., Cavalli-sforza L.L. & Feldman M.W. (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463-471.

Gordon H.R., Brown O.B., Evans R.H., Brown J.W., Smith R.C., Baker K.S. & Clark D.K. (1988) A semianalytic radiance model of ocean color. *Journal of Geophysical Research-Atmospheres*, **93**, 10909-10924.

Goudet J. (2001) *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.)*, http://www.unil.ch/izea/softwares/fstat.html.

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

Gower J.C. (1975) Generalized Procrustes Analysis. *Psychometrika*, **40**, 33-51.

Grant W.S. & Waples R.S. (2000) Spatial and temporal scales of genetic variability in marine and anadromous species: implications for fisheries oceanography. In: *Fisheries oceanography: An integrative approach to fisheries oceanography.* (eds. Harrison P.J. & Parson T.R.), pp. 61-93. Blackwell Science, Oxford.

Guelinckx J. (2008) *Estuarine habitat use by a goby species: a geochemical approach.* PhD-thesis, Katholieke Universiteit Leuven, Leuven, Belgium.

Guelinckx J., Maes J., Geysen B. & Ollevier F. (2008) Estuarine recruitment of a marine goby reconstructed with an isotopic clock. *Oecologia*, **157**, 41-52.

Guelinckx J., Maes J., Van Den Driessche P., Geysen B., Dehairs F. & Ollevier F. (2007) Changes in delta C^{13} and delta N^{15} in different tissues of juvenile sand goby *Pomatoschistus minutus*: a laboratory diet switch experiment. *Marine Ecology-Progress Series*, **341**, 205-215.

Guinand B., Lemaire C. & Bonhomme F. (2004) How to detect polymorphisms undergoing selection in marine fishes? A review of methods and case studies, including flatfishes. *Journal of Sea Research*, **51**, 167-182.

Guinand B., Rolland J.L. & Bonhomme F. (2008) Genetic structure of the common sole (*Solea solea*) in the Bay of Biscay: Nurseries as units of selection? *Estuarine Coastal and Shelf Science*, **78**, 316-326.

Guindon S. & Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696-704.

Guindon S., Lethiec F., Duroux P. & Gascuel O. (2005) PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research*, **33**, W557-W559.

Gysels E.S. (2003) *Genetic structure and evolutionary patterns of* Pomatoschistus *gobies of the European continental shelf and in the Mediterranean Sea.* PhD-thesis, Katholieke Universiteit Leuven, Leuven.

Gysels E.S., Hellemans B., Pampoulie C. & Volckaert F.A.M. (2004a) Phylogeography of the common goby, *Pomatoschistus microps*, with particular emphasis on the colonization of the Mediterranean and the North Sea. *Molecular Ecology*, **13**, 403-417.

Gysels E.S., Hellemans B., Patarnello T. & Volckaert F.A.M. (2004b) Current and historic gene flow of the sand goby *Pomatoschistus minutus* on the European Continental Shelf and in the Mediterranean Sea. *Biological Journal of the Linnean Society*, **83**, 561-576.

Gysels E.S., Leentjes V. & Volckaert F.A.M. (2004c) Small-scale clinal variation, genetic diversity and environmental heterogeneity in the marine gobies *Pomatoschistus minutus* and *P. lozanoi* (Gobiidae, Teleostei). *Heredity*, **93**, 208-214.

Hamerlynck O. (1990) The identification of *Pomatoschistus minutus* (Pallas) and *Pomatoschistus lozanoi* (de Buen) (Pisces, Gobiidae). *Journal of Fish Biology*, **37**, 723-728.

Hamerlynck O. (1993) *The ecology of* Pomatoschistus minutus *Pallas and* Pomatoschistus lozanoi *de Buen (Pisces, Gobiidae) in the shallow coastal waters and estuaries of the Southern Bight of the North Sea.* PhD-thesis, Rijksuniversiteit Gent, Gent.

Hamerlynck O. & Cattrijsse A. (1994) The food of *Pomatoschistus minutus* (Pisces, Gobiidae) in Belgian coastal waters, and a comparison with the food of its potential competitor *P. lozanoi*. *Journal of Fish Biology*, **44**, 753-771.

Hamerlynck O. & Hostens K. (1993) Growth, feeding, production, and consumption in 0-group bib (*Trisopterus luscus* L.) and whiting (*Merlangius merlangus* L.) in a shallow coastal area of the south-west Netherlands. *Ices Journal of Marine Science*, **50**, 81-91.

Hansen M.M., Ruzzante D.E., Nielsen E.E., Bekkevold D. & Mensberg K.L.D. (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. *Molecular Ecology*, **11**, 2523-2535.

Hardy O.J., Charbonnel N., Freville H. & Heuertz M. (2003) Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467-1482.

Hardy O.J. & Vekemans X. (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hargrave P.A. & McDowell J.H. (1992) Rhodopsin and phototransduction - a model system for G-protein-linked receptors. *FASEB Journal*, **6**, 2323-2331.

Harpending H.C. (1994) Signature of ancient population-growth in a low-resolution mitochondrial-DNA mismatch distribution. *Human Biology*, **66**, 591-600.

Hauser L. & Carvalho G.R. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333-362.

Healey M.C. (1971) Distribution and abundance of sand gobies, *Gobius minutus*, in Ythan estuary. *Journal of Zoology*, **163**, 177-229.

Hedrick P.W. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633-1638.

Hemmer-Hansen J., Nielsen E.E., Frydenberg J. & Loeschcke V. (2007) Adaptive divergence in a high gene flow environment: *Hsc70* variation in the European flounder (*Platichthys flesus* L.). *Heredity*, **99**, 592-600.

Herman P.M.J. & Heip C.H.R. (1999) Biochemistry of the maximum turbidity zone of Estuaries (MATURE): some conclusions. *Journal of Marine Systems*, **22**, 89-104.

Hewitt G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

Hewitt G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 183-195.

Hilbish T.J. & Koehn R.K. (1985) The physiological basis of natural selection at the *Lap* locus. *Evolution*, **39**, 1302-1317.

Hoarau G., Coyer J.A., Veldsink J.H., Stam W.T. & Olsen J.L. (2007) Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. *Molecular Ecology*, **16**, 3606-3616.

Hoese D.F. (1984) Gobioidei relationships. In: *Ontogeny and systematics of fishes.* (eds. Moser H.G., Richards W.J., Cohen D.M., Fahay M.P., Kendall Jr. A.W. & Richardson S.L.), pp. 588-591. American Society of Ichthyologists and Herpetologists, Lawrence.

Hofmeyer K. & Treisman J.E. (2008) Sensory systems: seeing the world in a new light. *Current Biology*, **18**, 919-921.

Holderegger R., Kamm U. & Gugerli F. (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, **21**, 797-807.

Hostens K. (2000) Spatial patterns and seasonality in the epibenthic communities of the Westerschelde (Southern Bight of the North Sea). *Journal of the Marine Biological Association of the United Kingdom*, **80**, 27-36.

Hostens K. (2003) *The demersal fish and macro-invertebrate assemblages of the Westerschelde and Oosterschelde estuaries (Southern Bight of the North Sea).* PhD-thesis, Ghent University (RUG), Gent, Belgium.

Howe G.T. & Brunner A.M. (2005) An evolving approach to understanding plant adaptation. *New Phytologist*, **167**, 1-5.

Hudson R.R., Kreitman M. & Aguadé M. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153-159.

Hunt D.M., Carvalho L.S., Cowing J.A., Parry J.W.L., Wilkie S.E., Davies W.L. & Bowmaker J.K. (2007) Spectral tuning of shortwave-sensitive visual pigments in vertebrates. *Photochemistry and Photobiology*, **83**, 303-310.

Hunt D.M., Dulai K.S., Partridge J.C., Cottrill P. & Bowmaker J.K. (2001) The molecular basis for spectral tuning of rod visual pigments in deep-sea fish. *Journal of Experimental Biology*, **204**, 3333-3344.

Hunt D.M., Fitzgibbon J., Slobodyanyuk S.J. & Bowmaker J.K. (1996) Spectral tuning and molecular evolution of rod visual pigments in the species flock of cottoid fish in Lake Baikal. *Vision Research*, **36**, 1217-1224.

Huyse T. (2002) *Evolutionary associations between Gyrodactylus and its goby host: Bound forever?* PhD-thesis, K.U.Leuven, Leuven.

Huyse T., Van Houdt J. & Volckaert F.A.M. (2004) Paleoclimatic history and vicariant speciation in the 'sand goby' group (Gobiidae, Teleostei). *Molecular Phylogenetics and Evolution*, **32**, 324-336.

ICES (2005) Report of the working group on the application of genetics in fisheries and mariculture (WGAGFM), p. 47 pp. ICES CM 2005/F:01, Silkeborg, Denmark.

Imsland A.K., Foss A., Stefansson S.O. & Nævdal G. (2000) Hemoglobin genotype of turbot (*Scophthalmus maximus*): consequences for growth and variations in optimal temperature for growth. *Fish Physiology and Biochemistry*, **23**, 75-81.

Jackson A.C. & Rundle S.D. (2008) Diet-shifts by an estuarine goby (*Pomatoschistus microps*) in the face of variable prey availability. *Journal of Experimental Marine Biology and Ecology*, **361**, 1-7.

Jackson D.A. (1995) Protest - a Procrustean randomization test of community environment concordance. *Ecoscience*, **2**, 297-303.

Jaquet N. & Raffaelli D. (1989) The ecological importance of the sand goby *Pomatoschistus minutus* (Pallas). *Journal of Experimental Marine Biology and Ecology*, **128**, 147-156.

Järvenpää M. & Lindström K. (2004) Water turbidity by algal blooms causes mating system breakdown in a shallow-water fish, the sand goby *Pomatoschistus minutus*. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **271**, 2361-2365.

Jehaes E., Toprak K., Vanderheyden N., Pfeiffer H., Cassiman J.J., Brinkmann B. & Decorte R. (2001) Pitfalls in the analysis of mitochondrial DNA from ancient specimens and the consequences for forensic DNA analysis: the historical case of the putative heart of Louis XVII. *International Journal of Legal Medicine*, **115**, 135-141.

Jerlov N. (1976) *Marine optics*. Elsevier, Amsterdam.

Jessen T.H., Weber R.E., Fermi G., Tame J. & Braunitzer G. (1991) Adaptation of bird hemoglobins to high-altitudes - Demonstration of molecular mechanism by protein engineering. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 6519-6522.

Johannesson K. & André C. (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, **15**, 2013-2029.

Johansson G., Sosa P.A. & Snoeijs P. (2003) Genetic variability and level of differentiation in North Sea and Baltic Sea populations of the green alga *Cladophora rupestris*. *Marine Biology*, **142**, 1019-1027.

Jokela-Määttä M., Pahlberg J., Lindstrom M., Zak P.P., Porter M., Ostrovsky M.A., Cronin T.W. & Donner K. (2005) Visual pigment absorbance and spectral sensitivity of the *Mysis relicta* species group (Crustacea, Mysida) in different light environments. *Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology*, **191**, 1087-1097.

Jokela-Määttä M., Smura T., Aaltonen A., Ala-Laurila P. & Donner K. (2007) Visual pigments of Baltic Sea fishes of marine and limnic origin. *Visual Neuroscience*, **24**, 389-398.

Jokela-Määttä M., Vartio A., Paulin L. & Donner K. (2006) Individual variation in rod absorbance spectra correlated with opsin gene polymorphism in sand goby (*Pomatoschistus minutus*). *Journal of Vision*, **6**, 53.

Jokela M., Vartio A., Paulin L., Fyhrquist-Vanni N. & Donner K. (2003) Polymorphism of the rod visual pigment between allopatric populations of the sand goby (*Pomatoschistus minutus*): a microspectrophotometric study. *Journal of Experimental Biology*, **206**, 2611-2617.

Jolly M.T., Viard F., Gentil F., Thiebaut E. & Jollivet D. (2006) Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology*, **15**, 1841-1855.

Jones A.G., Walker D., Kvarnemo C., Lindström K. & Avise J.C. (2001) How cuckoldry can decrease the opportunity for sexual selection: Data and theory from a genetic parentage analysis of the sand goby, *Pomatoschistus minutus*. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 9151-9156.

Jørgensen H.B.H., Hansen M.M., Bekkevold D., Ruzzante D.E. & Loeschcke V. (2005) Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. *Molecular Ecology*, **14**, 3219-3234.

Jorgensen T.H. & Emerson B.C. (2008) Functional variation in a disease resistance gene in populations of *Arabidopsis thaliana*. *Molecular Ecology*, **17**, 4912-4923.

Jukes T.H. & Cantor C.R. (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro H.N.), pp. 21-132. Academic Press, New York.

Kangas N. (2000) *Mechanisms of sexual selection in the sand goby* Pomatoschistus minutus. PhD-thesis, University of Helsinki, Helsinki.

Keith P., Galewski T., Cattaneo-Berrebi G., Hoareau T. & Berrebi P. (2005) Ubiquity of *Sicyopterus lagocephalus* (Teleostei: Gobioidei) and phylogeography of the genus *Sicyopterus* in the Indo-Pacific area inferred from mitochondrial cytochrome *b* gene. *Molecular Phylogenetics and Evolution*, **37**, 721-732.

Kellaway G.A., Redding J.H., Shephardthorn E.R. & Destombes J.P. (1975) Quaternary history of English-Channel. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, **279**, 189-218.

Kingsolver J.G., Pfennig D.W. & Servedio M.R. (2002) Migration, local adaptation and the evolution of plasticity. *Trends in Ecology & Evolution*, **17**, 540-541.

Kirk J.T.O. (1996) *Light and photosyntesis in aquatic ecosystems.* Cambridge University Press, Cambridge.

Knutsen H., Jorde P.E., Andre C. & Stenseth N.C. (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology*, **12**, 385-394.

Koskinen M.T., Hirvonen H., Landry P.A. & Primmer C.R. (2004) The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. *Hereditas*, **141**, 61-67.

Kottelat M. & Freyhof J. (2007) *Handbook of European freshwater fishes* Kottelat, Cornol and Freyhof, Berlin.

Larmuseau M.H.D. (2005) *Analyse van genmigratie bij de grondel* Pomatoschistus minutus *(Pallas, 1770) in de zuidelijke Noordzee en Westerschelde aan de hand van microsatelliet merkers.* M.Sc. thesis, Katholieke Universiteit Leuven, Leuven.

Larsen P.F., Nielsen E.E., Williams T.D., Hemmer-Hansen J., Chipman J.K., Kruhøffer M., Grønkjær P., George S.G., Dyrskjøt L. & Loeschcke V. (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Molecular Ecology*, **16**, 4674-4683.

Legendre P. (2000) Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation*, **67**, 37-73.

Leitão R., Martinho F., Neto J.M., Cabral H., Marques J.C. & Pardal M.A. (2006) Feeding ecology, population structure and distribution of *Pomatoschistus microps* (Krøyer, 1838) and *Pomatoschistus minutus* (Pallas, 1770) in a temperate estuary, Portugal. *Estuarine Coastal and Shelf Science*, **66**, 231-239.

Lemaire C., Allegrucci G., Naciri M., Bahri-Sfar L., Kara H. & Bonhomme F. (2000) Do discrepancies between microsatellite and allozyme variation reveal differential selection between sea and lagoon in the sea bass (*Dicentrarchus labrax*)? *Molecular Ecology*, **9**, 457-467.

Lenormand T. (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183-189.

Lim J., Chang J.L. & Tsai H.J. (1997) A second type of rod opsin cDNA from the common carp (*Cyprinus carpio*). *Biochimica Et Biophysica Acta-Gene Structure and Expression*, **1352**, 8-12.

Lindström K. & Lugli M. (2000) A quantitative analysis of the courtship acoustic behaviour and sound patterning in male sand goby, *Pomatoschistus minutus*. *Environmental Biology of Fishes*, **58**, 411-424.

Lindström K., St Mary C. & Pampoulie C. (2006) Sexual selection for male parental care in the sand goby, *Pomatoschistus minutus*. *Behavioral Ecology and Sociobiology*, **60**, 46-51.

Lindström M. (2000) Eye function of Mysidacea (Crustacea) in the northern Baltic Sea. *Journal of Experimental Marine Biology and Ecology*, **246**, 85-101.

Lucentini L., Vercillo F., Palombi A., Panara F. & Ragni B. (2007) A PCR-RFLP method on faecal samples to distinguish *Martes martes*, *Martes fiona*, *Mustela putorius* and *Vulpes vulpes*. *Conservation Genetics*, **8**, 757-759.

Luikart G. & Cornuet J.M. (1999) Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics*, **151**, 1211-1216.

Lukoschek V., Waycott M. & Keogh J.S. (2008) Relative information content of polymorphic microsatellites and mitochondrial DNA for inferring dispersal and population genetic structure in the olive sea snake, *Aipysurus laevis*. *Molecular Ecology*, **17**, 3062-3077.

Luttikhuizen P.C., Campos J., Van Bleijswijk J., Peijnenburg K.T.C.A. & van der Veer H.W. (2008) Phylogeography of the common shrimp, *Crangon crangon* (L.) across its distribution range. *Molecular Phylogenetics and Evolution*, **46**, 1015-1030.

Luttikhuizen P.C., Drent J. & Baker A.J. (2003) Disjunct distribution of highly diverged mitochondrial lineage clade and population subdivision in a marine bivalve with pelagic larval dispersal. *Molecular Ecology*, **12**, 2215-2229.

Maes J., de Brabandere L., Ollevier F. & Mees J. (2003) The diet and consumption of dominant fish species in the upper Scheldt estuary, Belgium. *Journal of the Marine Biological Association of the United Kingdom*, **83**, 603-612.

Maes J., Stevens M. & Ollevier F. (2005) The composition and community structure of the ichthyofauna of the upper Scheldt estuary: synthesis of a 10-year data collection (1991-2001). *Journal of Applied Ichthyology*, **21**, 86-93.

Maes J., Taillieu A., Van Damme P.A., Cottenie K. & Ollevier F. (1998) Seasonal patterns in the fish and crustacean community of a turbid temperate estuary (Zeeschelde Estuary, Belgium). *Estuarine Coastal and Shelf Science*, **47**, 143-151.

Magnhagen C. (1999) Alternative mating tactics, nest defence and courtship in gobies - just a matter of size? In: *Behaviour and conservation of littoral fishes* (eds. Almada V.C., Oliveira R.F. & Gonçalves E.J.), pp. 275-300. ISPA, Lisboa.

Mantel N. (1967) The detection of disease clustering and a generalised regression approach. *Cancer Research*, **27**, 209-220.

Martin A.P. & Palumbi S.R. (1993) Body size, metabolic-rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 4087-4091.

Mayden R.L., Tang K.L., Conway K.W., Freyhof J., Chamberlain S., Haskins M., Schneider L., Sudkamp M., Wood R.M., Agnew M., Bufalino A., Sulaiman Z., Miya M., Saitoh K. & He S. (2007) Phylogenetic relationships of *Danio* within the order cypriniformes: A framework for comparative and evolutionary studies of a model species. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, **308B**, 642-654.

Mayr E. (1963) *Animal species and evolution.* Harvard University Press, Cambridge, MA.

McKay S.I. & Miller P.J. (1997) The affinities of European sand gobies (Teleostei: Gobiidae). *Journal of Natural History*, **31**, 1457-1482.

McKelvey K.S. & Schwartz M.K. (2004) Genetic errors associated with population estimation using non-invasive molecular tagging: Problems and new solutions. *Journal of Wildlife Management*, **68**, 439-448.

McKelvey K.S. & Schwartz M.K. (2005) DROPOUT: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Molecular Ecology Notes*, **5**, 716-718.

Meirmans P.G. (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399-2402.

Metz E.C. & Palumbi S.R. (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Molecular Biology and Evolution*, **13**, 397-406.

Michiels N.K., Anthes N., Hart N.S., Herler J., Meixner A.J., Schleifenbaum F., Schulte G., Siebeck U.E., Sprenger D. & Wucherer M.F. (2008) Red fluorescence in reef fish: A novel signalling mechanism? *BMC Ecology*, **8**, 16.

Miller P.J. (1986) Gobiidae. In: *Fishes of the north-eastern Atlantic and Mediterranean* (eds. Whitehead P.J.P., Bauchot M.-L., Hureau J.C., Nielsen J. & Tortononese E.), pp. 1019-1085. UNESCO, Paris.

Miller P.J. (2009) A west Balkanian freshwater gobiid fish, *Knipowitschia mrakovcici* sp.nov. (Teleostei: Gobiidae). *Journal of Fish Biology*, **74**, 1499-1507.

Mobley C.D. (1994) *Light and water: radiative transfer in natural waters.* Academic Press, San Diego.

Morel A. & Gentili B. (1996) Diffuse reflectance of oceanic waters .III. Implication of bidirectionality for the remote-sensing problem. *Applied Optics*, **35**, 4850-4862.

Muers M. (2009) Population genetics: Separating demography from selection. *Nature Reviews Genetics*, **10**, 280-281.

Nager R.G., Keller L.F. & van Noordwijk A.J. (2000) Understanding natural selection on traits that are influenced by environmental conditions. In: *Adaptive genetic variation in the wild* (eds. Mousseau T.A., Sinervo B. & Endler J.A.), pp. 95-115. Oxford University Press US.

Nei M. (1987) *Molecular evolutionary genetics.* Columbia University Press, New York, USA.

Nei M. & Gojobori T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418-426.

Nei M. & Tajima F. (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, **97**, 145-163.

Neigel J.E. (1997) Population geentics and demography of marine species. In: *Marine biodiversity: patterns and processes* (eds. Ormond R.F.G., Gage J.D. & Angel M.V.), pp. 274-292. Cambridge University Press, Cambridge.

Nelson J.S. (2006) *Fishes of the world, 4th Edition.* John Wiley and Sons, New York.

Nesbø C.L., Rueness E.K., Iversen S.A., Skagen D.W. & Jakobsen K.S. (2000) Phylogeography and population history of Atlantic mackerel (*Scomber scombrus* L.): a genealogical approach reveals genetic structuring among the eastern Atlantic stocks. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267**, 281-292.

Nesse R.M., Stearns S.C. & Omenn G.S. (2006) Medicine needs evolution. *Science*, **311**, 1071.

Nickle B. & Robinson P.R. (2007) The opsins of the vertebrate retina: insights from structural, biochemical, and evolutionary studies. *Cellular and Molecular Life Sciences*, **64**, 2917-2932.

Nielsen E.E., Nielsen P.H., Meldrup D. & Hansen M.M. (2004) Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. *Molecular Ecology*, **13**, 585-595.

Nielsen R. (2001) Statistical tests of selective neutrality in the age of genomics. *Heredity*, **86**, 641-647.

Nielsen R. (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197-218.

Nielsen R., Hubisz M.J., Helllmann I., Torgerson D., Andrés A.M., Albrechtsen A., Gutenkunst R., Adams M.D., Cargill M., Boyko A., Indap A., Bustamante C.D. & Clark A.G. (in press) Darwinian and demographic forces affecting human protein coding genes. *Genome Research.*

Nielsen R. & Yang Z.H. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, **148**.

O'Malley K.G., Camara M.D. & Banks M.A. (2007) Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, **16**, 4930-4941.

O'Reilly P.T., Canino M.F., Bailey K.M. & Bentzen P. (2004) Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. *Molecular Ecology*, **13**, 1799-1814.

Oetjen K. & Reusch T.B.H. (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology*, **16**, 5156-5167.

Oksanen J., Kindt R., Legendre P. & O'Hara R.B. (2007) *vegan: Community Ecology Package version 1.8-5.*, http://cran.r-project.org/.

Oliver M.K., Lambin X., Cornulier T. & Piertney S.B. (2009) Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (*Arvicola terrestris*) metapopulations. *Molecular Ecology*, **18**, 80-92.

Olsen J.L., Stam W.T., Coyer J.A., Reusch T.B.H., Billingham M., Bostrom C., Calvert E., Christie H., Granger S., La Lumiere R., Milchakova N., Oudot-Le Secq M.P., Procaccini G., Sanjabi B., Serrao E., Veldsink J., Widdicombe S. & Wyllie-Echeverria S. (2004) North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Molecular Ecology*, **13**, 1923-1941.

Palumbi S.R. (1994) Genetic-divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547-572.

Palumbi S.R. (1999) All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 12632-12637.

Pampoulie C., Gysels E.S., Maes G.E., Hellemans B., Leentjes V., Jones A.G. & Volckaert F.A.M. (2004a) Evidence for fine-scale genetic structure and estuarine colonisation in a potential high gene flow marine goby (*Pomatoschistus minutus*). *Heredity*, **92**, 434-445.

Pampoulie C., Lindstrom K. & St Mary C.M. (2004b) Have your cake and eat it too: male sand gobies show more parental care in the presence of female partners. *Behavioral Ecology*, **15**, 199-204.

Pampoulie C., Rosecchi E., Bouchereau J.L. & Crivelli A.J. (1999) Life history traits of *Pomatoschistus minutus* in the Rhone Delta, France. *Journal of Fish Biology*, **55**, 892-896.

Pampoulie C., Ruzzante D.E., Chosson V., Jörundsdóttir T.D., Taylor L., Thorsteinsson V., Danielsdottir A.K. & Marteinsdottir G. (2006) The genetic structure of Atlantic cod (*Gadus morhua*) around Iceland: insight from microsatellites, the *Pan I* locus, and tagging experiments. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2660-2674.

Pampoulie C., Stefansson M.O., Jorundsdottir T.D., Danilowicz B.S. & Danielsdottir A.K. (2008) Recolonization history and large-scale dispersal in the open sea: the case study

of the North Atlantic cod, *Gadus morhua* L. *Biological Journal of the Linnean Society*, **94**, 315-329.

Park J.H., Scheerer P., Hofmann K.P., Choe H.-W. & Ernst O.P. (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*, **454**, 183-187.

Parmanne R. & Lindström K. (2003) Annual variation in goby larvae abundance in the Northern Baltic. *Journal of Fish Biology*, **62**, 413-426.

Parry J.W.L., Carleton K.L., Spady T.C., Carboo A., Hunt D.M. & Bowmaker J.K. (2005) Mix and match color vision: tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. *Current Biology*, **15**, 1734-1739.

Pasquaud S., Girardin M. & Elie P. (2004) Diet of gobies of the genus *Pomatoschistus* (*P. microps* and *P. minutus*), in the Gironde estuary (France). *Cybium*, **28**, 99-106.

Patarnello T., Volckaert F.A.M.J. & Castilho R. (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, **16**, 4426-4444.

Peel D., Ovenden J.R. & Peel S.L. (2004) *NeEstimator: software for estimating effective population size, Version 1.3.* Queensland Government; Department of Primary Industries and Fisheries.

Peijnenburg K.T.C.A., Fauvelot C., Breeuwer A.J. & Menken S.B.J. (2006) Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology*, **15**, 3319-3338.

Philippart C.J.M., Anadón R., Danovaro R., Dippner R., Drinkwater K.F., Hawkins S.J., Oguz T. & Reid P.C. (2007) Climate change impacts on the European marine and coastal environment. Marine Board-ESF Position Paper. p. 82, Strasbourg.

Pogson G.H. & Fevolden S.E. (2003) Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (*Pan*I) locus. *Molecular Ecology*, **12**, 63-74.

Poizat G., Rosecchi E., Chauvelon P., Contournet P. & Crivelli A.J. (2004) Long-term fish and macro-crustacean community variation in a Mediterranean lagoon. *Estuarine Coastal and Shelf Science*, **59**, 615-624.

Pompanon F., Bonin A., Bellemain E. & Taberlet P. (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847-859.

Por F.D. & Dimentman C. (1985) Continuity of Messinian biota in the Mediterranean basin. In: *Geological evolution of the Mediterranean basin.* (eds. Stanley D.J. & Wezel F.C.), pp. 545-557. Springer-Verlag, Berlin.

Posada D. & Crandall K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Poulsen N.A., Nielsen E.E., Schierup M.H., Loeschcke V. & Gronkjaer P. (2006) Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). *Molecular Ecology*, **15**, 321-331.

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

Provan J., Murphy S. & Maggs C.A. (2005a) Tracking the invasive history of the green alga *Codium fragile* ssp tomentosoides. *Molecular Ecology*, **14**, 189-194.

Provan J., Wattier R.A. & Maggs C.A. (2005b) Phylogeographic analysis of the red seaweed *Palmaria palmata* reveals a Pleistocene marine glacial refugium in the English Channel. *Molecular Ecology*, **14**, 793-803.

Queller D.C. & Goodnight K.F. (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258-275.

Rambaut A. & Drummond A.J. (2007) *Tracer*, 1.3 edn., http://beast.bio.ed.ac.uk/Tracer.

Raymond M. & Rousset F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Reed D.H. & Frankham R. (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, **55**, 1095-1103.

Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

Roberge C., Paez D.J., Rossignol O., Guderley H., Dodson J. & Bernatchez L. (2007) Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular Immunology*, **44**, 1374-1383.

Rodrigues P., Reis-Henriques M.A., Campos J. & Santos M.M. (2006) Urogenital papilla feminization in male *Pomatoschistus minutus* from two estuaries in northwestern Iberian Peninsula. *Marine Environmental Research*, **62**, S258-S262.

Rozas J., Sanchez-DelBarrio J.C., Messeguer X. & Rozas R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496-2497.

Rozen S. & Skaletsky H.J. (1998) *PRIMER 3*. http://www.genome.wi.met.edu/genome_software/other/primer3.html.

Ruber L., Van Tassell J.L. & Zardoya R. (2003) Rapid speciation and ecological divergence in the American seven-spined gobies (Gobiidae, Gobiosomatini) inferred from a molecular phylogeny. *Evolution*, **57**, 1584-1598.

Ruzzante D.E., Taggart C.T. & Cook D. (1998) A nuclear DNA basis for shelf- and bank-scale population structure in northwest Atlantic cod (*Gadus morhua*): Labrador to Georges Bank. *Molecular Ecology*, **7**, 1663-1680.

Ryynänen H.J. & Primmer C.R. (2004) Distribution of genetic variation in the growth hormone 1 gene in Atlantic salmon (*Salmo salar*) populations from Europe and North America. *Molecular Ecology*, **13**, 3857-3869.

Sa-Pinto A., Branco M., Harris D.J. & Alexandrino P. (2005) Phylogeny and phylogeography of the genus *Patella* based on mitochondrial DNA sequence data. *Journal of Experimental Marine Biology and Ecology*, **325**, 95-110.

Sarnthein M. (2001) Der Atlantic im Letzen Glazialen Maximum. *Nova Acta Leopoldina*, **88**, 35-43.

Sarropoulou E., Nousdili D., Magoulas A. & Kotoulas G. (2008) Linking the genomes of nonmodel teleosts through comparative genomics. *Marine Biotechnology*, **10**, 227-233.

Sawyer S.L., Wu L.I., Emerman M. & Malik H.S. (2005) Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 2832-2837.

Schoenhuth S., Doadrio I., Dominguez-Dominguez O., Hillis D.M. & Mayden R.L. (2008) Molecular evolution of southern North American Cyprinidae (Actinopterygii), with the description of the new genus *Tampichthys* from central Mexico. *Molecular Phylogenetics and Evolution*, **47**, 729-756.

Schulte P.M., Glemet H.C., Fiebig A.A. & Powers D.A. (2000) Adaptive variation in lactate dehydrogenase-B gene expression: Role of a stress-responsive regulatory element. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6597-6602.

Schulte P.M., Gomez-Chiarri M. & Powers D.A. (1997) Structural and functional differences in the promoter and 5' flanking region of *LdhB* within and between populations of the Teleost *Fundulus heteroclitus*. *Genetics*, **145**, 759-769.

Seehausen O., Terai Y., Magalhaes I.S., Carleton K.L., Mrosso H.D.J., Miyagi R., van der Sluijs I., Schneider M.V., Maan M.E., Tachida H., Imai H. & Okada N. (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620-U623.

Shaw P.W., Arkhipkin A.I. & Al-Khairulla H. (2004) Genetic structuring of Patagonian toothfish populations in the Southwest Atlantic Ocean: the effect of the Antarctic Polar Front and deep-water troughs as barriers to genetic exchange. *Molecular Ecology*, **13**, 3293-3303.

Sherman C.D.H. & Ayre D.J. (2008) Fine-scale adaptation in a clonal sea anemone. *Evolution*, **62**, 1373-1380.

Shubin N. (2008) *Your inner fish. A journey into the 3.5-billion-year history of the human body.* Pantheon Books, New York.

Simonovic P.D. (1999) Phylogenetic relationships of Ponto-Caspian gobies and their relationship to the Atlantic-Mediterranean Gobiinae. *Journal of Fish Biology*, **54**, 533-555.

Singer A., Kvarnemo C., Lindström K. & Svensson O. (2006) Genetic mating patterns studied in pools with manipulated nest site availability in two populations of *Pomatoschistus minutus*. *Journal of Evolutionary Biology*, **19**, 1641-1650.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457-462.

Smaal A.C. & Nienhuis P.H. (1992) The Eastern Scheldt (The Netherlands), from an estuary to a tidal bay - a review of responses at the ecosystem level. *Netherlands Journal of Sea Research*, **30**, 161-173.

Smouse P.E., Long J.C. & Sokal R.R. (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**, 627-632.

Spady T.C., Seehausen O., Loew E.R., Jordan R.C., Kocher T.D. & Carleton K.L. (2005) Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Molecular Biology and Evolution*, **22**, 1412-1422.

Stamatis C., Triantafyllidis A., Moutou K.A. & Mamuris Z. (2004) Mitochondrial DNA variation in northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus*. *Molecular Ecology*, **13**, 1377-1390.

Stefanni S., Gysels E.S., Volckaert F.A.M. & Miller P.J. (2003) Allozyme variation and genetic divergence in the sand goby, *Pomatoschistus minutus* (Teleostei: Gobiidae). *Journal of the Marine Biological Association of the United Kingdom*, **83**, 1143-1149.

Stefanni S. & Thorley J.L. (2003) Mitochondrial DNA phylogeography reveals the existence of an Evolutionarily Significant Unit of the sand goby *Pomatoschistus minutus* in the Adriatic (Eastern Mediterranean). *Molecular Phylogenetics and Evolution*, **28**, 601-609.

Stephens M. & Donnelly P. (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162-1169.

Stephens M., Smith N.J. & Donnelly P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978-989.

Storz J.F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688.

Sugawara T., Terai Y., Imai H., Turner G.F., Koblmuller S., Sturmbauer C., Shichida Y. & Okada N. (2005) Parallelism of amino acid changes at the *RH1* affecting spectral sensitivity among deep-water cichlids from Lakes Tanganyika and Malawi. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 5448-5453.

Sugawara T., Terai Y. & Okada N. (2002) Natural selection of the rhodopsin gene during the adaptive radiation of East African Great Lakes Cichlid fishes. *Molecular Biology and Evolution*, **19**, 1807-1811.

Svensson O. (2004) *Sexual selection in* Pomatoschistus - *nests, sperm competition, and paternal care.* PhD-degree, Stockholm University, Stockholm.

Swofford D.L. (2002) *PAUP*. Phylogenetic analysis using parsimony (* and other methods)* Sinauer Associates, Sunderland, Massachusetts.

Taberlet P., Fumagalli L., Wust-Saucy A.G. & Cosson J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453-464.

Tajima F. (1989) The effect of change in population-size on DNA polymorphism. *Genetics*, **123**, 597-601.

Tamura K., Dudley J., Nei M. & Kumar S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

Tamura K. & Nei M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512-526.

Taylor M.S. & Hellberg M.E. (2005) Marine radiations at small geographic scales: speciation in neotropical reef gobies (*Elacatinus*). *Evolution*, **59**, 374-385.

Terai Y., Mayer W.E., Klein J., Tichy H. & Okada N. (2002) The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 15501-15506.

Thetmeyer H. (1997) Diel rhythms of swimming activity and oxygen consumption in *Gobiusculus flavescens* (Fabricius) and *Pomatoschistus minutus* (Pallas) (Teleostei: Gobiidae). *Journal of Experimental Marine Biology and Ecology*, **218**, 187-198.

Tishkoff S.A., Reed F.A., Ranciaro A., Voight B.F., Babbitt C.C., Silverman J.S., Powell K., Mortensen H.M., Hirbo J.B., Osman M., Ibrahim M., Omar S.A., Lema G., Nyambo T.B., Ghori J., Bumpstead S., Pritchard J.K., Wray G.A. & Deloukas P. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, **39**, 31-40.

Toyama M., Hironaka M., Yamahama Y., Horiguchi H., Tsukada O., Uto N., Ueno Y., Tokunaga F., Seno K. & Hariyama T. (2008) Presence of rhodopsin and porphyropsin in the eyes of 164 fishes, representing marine, diadromous, coastal and freshwater species - A qualitative and comparative study. *Photochemistry and Photobiology*, **84**, 996-1002.

Trezise A.E.O. & Collin S.P. (2005) Opsins: Evolution in waiting. *Current Biology*, **15**, R794-R796.

Turner L.M. & Hoekstra H.E. (2008) Reproductive protein evolution within and between species: maintenance of divergent ZP3 alleles in *Peromyscus*. *Molecular Ecology*, **17**, 2616-2628.

Utne-Palm A.C. (2002) Visual feeding of fish in a turbid environment: Physical and behavioural aspects. *Marine and Freshwater Behaviour and Physiology*, **35**, 111-128.

Utne-Palm A.C. & Bowmaker J.K. (2006) Spectral sensitivity of the two-spotted goby *Gobiusculus flavescens* (Fabricius): a physiological and behavioural study. *Journal of Experimental Biology*, **209**, 2034-2041.

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

Vanden Eede S. (2006) *Populatiedynamica van* Pomatoschistus minutus *in de Noordzee en het Schelde-estuarium.* Masters thesis, Katholieke Universiteit Leuven, Leuven, Belgium.

Vasemägi A. & Primmer C.R. (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623-3642.

Volis S. (2008) Detection of signature of positive selection in naturally occurring genetic variation. In: *Population genetics research progress* (ed. Koven V.T.), pp. 279-310. Nova Science Publishers, Inc.

Wakeley J. (1993) Substitution rate variation among sites in hypervariable region-1 of human mitochondrial-DNA. *Journal of Molecular Evolution*, **37**, 613-623.

Waligóra-Borek K. & Sapota M. (2005) Breeding characteristics of *Pomatoschistus minutus* (sand goby) - one of the most abundant goby species in coastal waters of the Gulf of Gdańsk regio (south Baltic Sea). *Oceanological and Hydrobiological Studies*, **34**, 47-55.

Wallis G.P. & Beardmore J.A. (1980) Genetic evidence for naturally occuring fertile hybrids between two goby species, *Pomatoschistus minutus* and *P. lozanoi* (Pisces, Gobiidae). *Marine Ecology Progress Series*, **3**, 309-315.

Wallis G.P. & Beardmore J.A. (1983) Genetic differentiation between populations of *Pomatoschistus minutus* from the Bristol channel and the Adriatic. *Genetica*, **62**, 75-80.

Wallis G.P. & Beardmore J.A. (1984) An electrophoretic study of the systematic relationships of some closely related goby species (Pisces, Gobiidae). *Biological Journal of the Linnean Society*, **22**, 107-123.

Wang J.L. & Whitlock M.C. (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, **163**, 429-446.

Waples R.S. (1989) A generalized-approach for estimating effective population-size from temporal changes in allele frequency. *Genetics*, **121**, 379-391.

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

Waples R.S. (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, **7**, 167-184.

Waples R.S. & Do C. (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, **8**, 753-756.

Ward R.D., Woodwark M. & Skibinski D.O.F. (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology*, **44**, 213-232.

Webb C.J. (1980) Systematics of the *Pomatoschistus minutus* complex (Teleostei: Gobiidae). *Philosophical Transactions of the Royal Society of London Series B*, **291**, 201-241.

Weinstock J., Willerslev E., Sher A., Tong W.F., Ho S.Y.W., Rubenstein D., Storer J., Burns J., Martin L., Bravi C., Prieto A., Froese D., Scott E., Lai X.L. & Cooper A. (2005) Evolution, systematics, and phylogeography of Pleistocene horses in the New World: A molecular perspective. *Plos Biology*, **3**, 1373-1379.

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

White E.M., Goncalves D.M., Partridge J.C. & Oliveira R.F. (2004) Vision and visual variation in the peacock blenny. *Journal of Fish Biology*, **65**, 227-250.

Wilson A.B. (2006) Genetic signature of recent glaciation on populations of a near-shore marine fish species (*Syngnathus leptorhynchus*). *Molecular Ecology*, **15**, 1857-1871.

Wolfe S.L., Peter J.R., Hertz P.E. & Mcmilla S. (2007) *Biology: the dynamic science* Cengage Learning EMEA.

Wong B.B., Lehtonen T.K. & Lindström K. (2008) Male nest choice in sand gobies, *Pomatoschistus minutus*. *Ethology*, **114**, 575-581.

Wood H.M., Grahame J.W., Humphray S., Rogers J. & Butlin R.K. (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, **17**, 3123-3135.

Yang Z.H. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, **13**, 555-556.

Yang Z.H. (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, **15**, 568-573.

Yang Z.H. (2000) Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *Journal of Molecular Evolution*, **51**, 423-432.

Yang Z.H. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586-1591.

Yang Z.H. & Bielawski J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, **15**, 496-503.

Yang Z.H., Nielsen R., Goldman N. & Pedersen A.-M. (2000a) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431-449.

Yang Z.H., Swanson W.J. & Vacquier V.D. (2000b) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution*, **17**, 1446-1455.

Yang Z.H., Wong W.S.W. & Nielsen R. (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107-1118.

Yokoyama R., Knox B.E. & Yokoyama S. (1995) Rhodopsin from the fish, *Astyanax* - Role of tyrosine-261 in the red shift. *Investigative Ophthalmology & Visual Science*, **36**, 939-945.

Yokoyama S. (2000) Molecular evolution of vertebrate visual pigments. *Progress in Retinal and Eye Research*, **19**, 385-419.

Yokoyama S. (2002) Evaluating adaptive evolution. *Nature Genetics*, **30**, 350-351.

Yokoyama S., Tada T. & Yamato T. (2007) Modulation of the absorption maximum of rhodopsin by amino acids in the C-terminus. *Photochemistry and Photobiology*, **83**, 236-241.

Yokoyama S. & Takenaka N. (2004) The molecular basis of adaptive evolution of squirrelfish rhodopsins. *Molecular Biology and Evolution*, **21**, 2071-2078.

Yokoyama S., Zhang H., Radlwimmer F.B. & Blow N.S. (1999) Adaptive evolution of color vision of the Comoran coelacanth (*Latimeria chalumnae*). *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 6279-6284.

Zaidi R.H., Jaal Z., Hawkes N.J., Hemingway J. & Symondson W.O.C. (1999) Can multiple-copy sequences of prey DNA be detected amongst the gut contents of invertebrate predators? *Molecular Ecology*, **8**, 2081-2087.

Zander C.D. (2005) Comparative studies on goby (Teleostei) parasite communities from the North and Baltic Sea. *Parasitology Research*, **96**, 62-68.

Zane L. (2007) Adaptive divergence overcoming gene flow. *Heredity*, **99**, 565-566.

Zane L., Bargelloni L. & Patarnello T. (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**, 1-16.

Zane L., Ostellari L., Maccatrozzo L., Bargelloni L., Cuzin-Roudy J., Buchholz F. & Patarnello T. (2000) Genetic differentiation in a pelagic crustacean (*Meganyctiphanes norvegica*: Euphausiacea) from the North East Atlantic and the Mediterranean Sea. *Marine Biology*, **136**, 191-199.

Zhang D.X. & Hewitt G.M. (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, **11**, 247-251.

Zhang J.Z., Zhang Y.P. & Rosenberg H.F. (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genetics*, **30**, 411-415.

LIST OF PUBLICATIONS

Peer reviewed papers

Larmuseau M.H.D., Raeymaekers J.A.M., Ruddick K.G., Van Houdt J.K.J. & Volckaert F.A.M. (In Press) To see in different seas: spatial variation in the rhodopsin gene of the sand goby (*Pomatoschistus minutus*). *Molecular Ecology*, in press (Impact factor: 5.169).

Larmuseau M.H.D., Freyhof J., Volckaert F.A.M. & Van Houdt J.K.J. (In Press) Matrilinear phylogeography and demographic patterns of *Rutilus rutilus*: implications for taxonomy and conservation. *Journal of Fish Biology*, in press (Impact factor: 1.404).

Larmuseau M.H.D., Van Houdt J.K.J, Guelinckx J., Hellemans B. & Volckaert F.A.M. (2009) Distributional and demographic consequences of Pleistocene climate fluctuations for a marine demersal fish in the NE Atlantic. *Journal of Biogeography*, **36**, 1138-1151 (Impact factor: 3.539).

Larmuseau M.H.D., Guelinckx J., Hellemans B., Van Houdt J.K.J & Volckaert F.A.M. (2008) Fast PCR-RFLP method facilitates identification of *Pomatoschistus* species in the North Atlantic. *Journal of Applied Ichthyology*, **24**, 342-344. (Impact factor: 0.663).

Larmuseau M.H.D., Hellemans B., Van Houdt J.K.J. & Volckaert F.A.M. (2007) Development and characterization of nine polymorphic microsatellite markers in the sand goby *Pomatoschistus minutus* (Gobiidae). *Molecular Ecology Notes*, **7**, 147-149. (Impact factor: 1.257).

Raeymaekers J.A.M., Van Houdt J.K.J., **Larmuseau M.H.D.**, Geldof S. & Volckaert F.A.M. (2007) Divergent selection as revealed by P_{ST} and QTL-based F_{ST} in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. *Molecular Ecology*, **16**, 891-905. (Impact factor: 5.169).

Book sections

Vanaverbeke J., Courtens W., Deneudt K., Franco M.A., Gysels E.S., Hellemans B., Huyse T., **Larmuseau M.H.D.**, Mees J., Moodley L., Pampoulie C., Provoost P., Remerie T., Soetaert K., Stienen E.W.M., Vanden Berghe E., Van de Walle, M., Van Oevelen D., Volckaert F.A.M. & Vinckx M. (2007) TROPHOS: Fundamental research for policy and sustainable management. In: Science and sustainable management of the North Sea: Belgian case studies (Eds. Calewaert J.-B. & Maes F.). Academia Press, Gent, Belgium.

Reports

Vanaverbeke J., Braeckman U., Cuveliers E., Courtens W., Huyse T., Lacroix G., **Larmuseau M.H.D.**, Maes G., Provoost P., Rabaut M., Remerie T., Savina M., Soetaert K., Stienen E.W.M., Verstraete H., Volckaert F.A.M. & Vinckx M. (2009). WESTBANKS Final Report Phase I. Project of the Belgian Federal Office for Scientific, Technical and Cultural Affairs; Contract nr. SD/BN/01A.

Printed by PROCOPIA NV
Ambachtenlaan 29
3001 Heverlee
016/38.36.34
info@procopia.be
www.procopia.be

Laboratorium voor Diversiteit en Systematiek van Dieren
Ch. Deberiotstraat 32
B- 3000 Leuven, België
Aquabio@bio.kuleuven.be