

Phylogeography and Adaptive Divergence of Three-spined Stickleback Populations

Hannu Mäkinen
University of Helsinki
2007

Phylogeography and Adaptive Divergence of Three-spined Stickleback Populations

Hannu Mäkinen

Ecological Genetics Research Unit
Department of Biological and Environmental Sciences
Faculty of Biosciences
University of Helsinki
Finland

Academic Dissertation

To be presented, with permission of the Faculty of Biosciences of the University of Helsinki, for public criticism in the Auditorium 1041 of Biocenter 2, Viikinkaari 5, October 5th, at 12.00 o'clock

Helsinki 2007

© Hannu Mäkinen (chapters 0, III, V, VI)
© Blackwell Publishing Ltd (chapter I, IV)
© Elsevier (chapter II)

Author's address

Department of Biological and Environmental Sciences
P.O. Box 65 (Viikinkaari 1)
FI-00014 University of Helsinki
Finland

e-mail: hannu.makinen@helsinki.fi

ISBN 978-952-92-2635-1 (Paperback)

ISBN 978-952-10-4180-8 (PDF)

<http://ethesis.helsinki.fi>

Cover drawing © Jouni Heinänen

Cover layout © Anne Vääri

Layout Timo Päivärinta

Edita, Helsinki 2007

Phylogeography and Adaptive Divergence of Three-spined Stickleback Populations

Hannu Mäkinen

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:

- I Mäkinen HS, Cano JM, Merilä J (2006) Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. *Molecular Ecology* 15, 1519–1534.
- II Mäkinen HS, Merilä J (2007) Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe – evidence for multiple glacial refugia. *Molecular Phylogenetics and Evolution*, in press.
- III Cano JM, Mäkinen HS, Leinonen T, Freyhof J, Merilä J (2007) Extreme neutral genetic and morphological divergence supports classification of Adriatic three-spined stickleback (*Gasterosteus aculeatus*) populations as conservation units. *Manuscript*.
- IV Cano JM, Matsuba C, Mäkinen H, Merilä J (2006) The utility of QTL-linked markers to detect selective sweeps in natural populations — a case study of the *EDA* gene and a linked marker in threespine stickleback. *Molecular Ecology* 15, 4613-4621.
- V Mäkinen HS, Cano JM, Merilä J (2007) Identifying footprints of directional and balancing selection in marine and freshwater populations of the three-spined stickleback (*Gasterosteus aculeatus*). *Manuscript*.
- VI Mäkinen HS, Shikano T, Cano JM, Merilä J (2007) Hitchhiking mapping reveals a large genomic region affected by natural selection in three-spined stickleback chromosome VIII. *Manuscript*.

Contributions

| | I | II | III | IV | V | VI |
|------------------------|------------|-----------|------------|----------------|------------|----------------|
| Original Idea | JM, HM | JM, HM | HM, JC | JC | HM, JC | HM |
| Laboratory work | HM | HM | HM, TL | HM, CM | HM | HM, TS |
| Data analysis | HM, JC | HM | HM, JC | HM, JC, CM | HM, JC | HM, TS, JC |
| Manuscript preparation | HM, JC, JM | HM, JM | HM, JC, JM | HM, JC, CM, JM | HM, JC, JM | HM, TS, JC, JM |

Hannu Mäkinen (HM), José Manuel Cano (JC), Juha Merilä (JM), Chikako Matsuba (CM), Takahito Shikano (TS), Tuomas Leinonen (TL)

Supervised by:

Prof. Juha Merilä
University of Helsinki
Finland

Reviewed by:

Prof. Eric B. Taylor
The University of British Columbia
Canada

Prof. Michael M. Hansen
Danish Institute for Fisheries Research
Denmark

Examined by:

Prof. Filip Volckaert
Katholieke Universiteit Leuven
Belgium

Contents

| | |
|---|----|
| INTRODUCTION | 5 |
| <i>Determinants of genetic variation</i> | 5 |
| <i>Phylogeography – genetic variation in the historical context</i> | 6 |
| <i>Adaptive divergence in the Northern Hemisphere fish species</i> | 8 |
| <i>Molecular markers in genetic studies</i> | 9 |
| <i>The three-spined stickleback as a model species</i> | 11 |
| <i>Specific aims of the research</i> | 12 |
| SUMMARY OF THE MAIN RESULTS AND DISCUSSION | 13 |
| LARGE SCALE PHYLOGEOGRAPHY AND POPULATION STRUCTURE OF THREE-SPINED STICKLEBACKS IN EUROPE | 13 |
| <i>Genetic relationships among marine and freshwater populations</i> | 13 |
| <i>Mitochondrial DNA phylogeography in Europe</i> | 15 |
| <i>Identifying conservation units</i> | 18 |
| GENETIC BASIS OF ADAPTIVE DIVERGENCE | 20 |
| <i>The evolution of lateral plate number</i> | 20 |
| <i>Detecting targets of natural selection at the molecular level</i> | 22 |
| <i>An explorative genome scan in three-spined stickleback populations</i> | 23 |
| <i>A fine-scale mapping of a selective sweep</i> | 24 |
| CONCLUSIONS AND FUTURE DIRECTIONS | 28 |
| ACKNOWLEDGEMENTS | 29 |
| LITERATURE CITED | 31 |

INTRODUCTION

Determinants of genetic variation

The ability of a population to adapt to changing environmental conditions depends critically on the amount of heritable genetic variation in a given population (Lande and Shannon 1996; Lynch 1996). Thus, measuring the amount of genetic variation and understanding how it is divided within and among populations is one of the central tasks in evolutionary genetics. Assessing the relative roles random and deterministic processes (e.g. Merilä and Crnokrak 2001) as well historical factors (e.g. Hewitt 1996, 2000) shaping the patterns of genetic variation are important for understanding how population differentiation proceeds. In the long term, the process of population differentiation can lead to speciation, that is, evolution of new species, which is a central theme of all biological research (Coyne and Orr 2004). Furthermore, investigating patterns of genetic variation at the molecular level might help in identifying genes that are responsible for evolutionary change and reproductive isolation, and can thereby shed light on the genetic mechanisms underlying adaptation and speciation (Orr 2005a, b). The information found in the patterns of geographic distribution of genetic variation also has important applications in planning management and conservation strategies aiming to preserve biodiversity (Moritz 1994; Fraser and Bernatchez 2001).

The patterns in genetic variation and population differentiation are shaped by a complex interplay between molecular and population level processes, as well as by historical factors. Mutation and recombination acting on a DNA sequence are the ultimate sources of genetic

variation, which is then shaped by random genetic drift and natural selection. Random fluctuations in allele frequencies from generation to generation can result from the sorting of alleles during reproduction, commonly known as random genetic drift. As a result of random genetic drift, alleles can be lost, increase in frequency or become fixed in a population (Wright 1931). The magnitude of random fluctuations depends on the effective population size, i.e. the number of individuals reproducing in a population. In large populations only small changes in allele frequencies are expected in comparison to small populations, while the latter can undergo large and unpredictable changes in allele frequencies (Hedrick 2005). Thus, evolutionary change may proceed solely as a consequence of random processes. In contrast, gene flow between populations, i.e. individuals that are migrating between populations and incorporated to the breeding population can reduce the effect of genetic drift. Even a small number of individuals (ca. one gene copy per generation) can maintain allele frequencies similar between populations (Wright 1931). Later, this one-migrant-per-generation rule has been questioned especially in cases of small or fluctuating effective population sizes (Mills and Allendorf 1996; Vucetich and Waite 2000). A conservative estimate of 10 immigrants per generation would be needed to prevent loss of genetic variation due to random genetic drift in small populations (Vucetich and Waite 2000).

The traditional view of the evolutionary biology has emphasized natural selection as the primary mechanism underlying phenotypic evolution (Darwin 1859; Fisher 1930). The effect of natural selection on the level of genetic variation

depends strongly on the type of selection operating in a genetic locus (Nielsen 2005; Charlesworth 2006). Directional selection changes the allele frequencies by increasing the frequency of a beneficial allele relative to the others, and thus, leads to reduced levels of genetic variation at the selected locus but also at linked sites (Maynard – Smith and Haigh 1974; Schlötterer 2003; Nielsen 2005). Directional selection can operate on existing alleles (standing variation) or on new mutations in the population (Teshima et al. 2006). Balancing or stabilizing selection in turn tends to maintain similar allele frequencies between populations and increase the degree of within population genetic variation (Nielsen 2005; Charlesworth 2006). In addition, balancing selection is thought to be capable of maintaining genetic variation over long evolutionary timescales (Asthana et al. 2005). Like directional selection, balancing selection affects the patterns of genetic variation at the nearby genomic regions (Charlesworth 2006). Background selection, on the other hand, removes detrimental mutations from the population and thus reduces the genetic variation (Charlesworth 1994). Furthermore, background selection operating on slightly deleterious mutations may affect the patterns of neutral genetic variation via hitchhiking with a selected gene (Charlesworth 1993).

There has been intensive debate on the relative roles of genetic drift and natural selection underlying evolutionary change (e.g. Bernardi 2007). According to the neutral theory of evolution, most of the mutations are selectively neutral and thus the major force shaping the genetic variation is random genetic drift (Kimura 1968). The neutral theory was in contrast to the Darwinian model of evolution, which emphasized the role of directional selection

as a primary force underlying evolutionary change (Darwin 1859, Fisher 1930). The neutral theory was largely based on the observation that most of the mutations did not change the amino acid code, and thus, had little evolutionary significance (Kimura 1968). A later modification of the neutral theory, i.e. the nearly neutral theory of evolution developed by Ohta (1992, 2002) emphasized the role of effective population size. In populations with small effective size, random genetic drift would be the major evolutionary force. However, empirical studies have indicated the predominant role of selection over genetic drift in the evolution of phenotypic traits even in small populations (Merilä and Crnokrak 2001; Koskinen et al. 2002a). Yet, the development of the neutral theory has been a theoretical foundation - in the broad sense - for coalescent theory (Hudson 1983), molecular clock dating (Bromham and Penny 2003) and phylogenetic inference (Ohta and Gillespie 1996). Furthermore, neutral theory provides a null model for testing deviations from neutral expectations (Lewontin and Krakauer 1973; Kreitman 1996; Beaumont 2005).

Phylogeography – genetic variation in the historical context

Phylogeography aims to assess the processes behind the geographical distribution of genealogical lineages at the intraspecific level and between closely related species (Avice 2000; Hewitt 2000). The study of phylogeography is comprised of (i) estimating a gene genealogy of a species and (ii) interpreting the genealogy. The interpretation mainly focuses on historical factors such as Pleistocene ice ages, which have had an important role in shaping the patterns of genetic variation in terrestrial and aquatic

organisms both in Europe and Northern America (Avice 2000; Hewitt 2000). The period of the repeated glaciations during the Pleistocene (from ca. 1.9 million years to ca. 10 000 years ago) drastically changed the environment of the Northern Hemisphere (Dawson 1992). Advancing ice sheets devastated suitable terrestrial habitats and also reshaped the water systems such as the main river drainages in Europe (Dawson 1992; Gibbard 1998). During the warmer interglacial periods, vast proglacial lakes were formed in the glacial margins creating suitable colonization routes for aquatic organisms (Bernatchez and Wilson 1998; Mangerud et al. 2004). The changes in the sea level were at an order of magnitude of 100 m leaving large coastline areas dry in the North Sea and Mediterranean Sea (Dawson 1992). The genetic consequences or the legacy of the Pleistocene ice ages were diverse (Taberlet et al. 1998; Hewitt 2000). During the glacial maxima, species' distributions were restricted to southern refugial areas and species probably experienced reductions in effective population sizes. Reshaping the landscape by the ice sheets introduced new geographical barriers, which restricted gene flow between populations further enhancing the genetic divergence between populations (Hewitt 1996, 2000, 2001, 2004).

These effects of the Pleistocene ice ages can be detected in the present day patterns of genetic variation (Hewitt 2000). Phylogeographic studies have revealed several geographically distinct evolutionary lineages at the intraspecific level. For example in Scandinavia, freshwater fishes are often diverged into western, southern and eastern evolutionary lineages in their molecular characteristics, which reflect the colonizations from the different glacial refugia (e.g. Koskinen

et al. 2000; Kontula and Väinölä 2001). Lower genetic variability at the northern latitudes is likely a result of the gradual loss of genetic variation during the recolonization from the southern refugia (Hewitt 1996; Bernatchez and Wilson 1998; Avice 2000). Finally, the genetic divergence in southern latitudes appears to exceed the divergence observed in the northern latitudes indicating a more recent origin of the northern populations (Bernatchez and Wilson 1998).

The analysis methods in phylogeographic studies have been traditionally carried out by reconstructing phylogenetic trees or haplotype networks from molecular data (Posada and Crandall 2001). The interpretation has then typically relied in a narrow question of detecting monophyletic groups, which is considered to reflect strong isolation of genealogical lineages (Moritz 1994; Zink 2004). The development of coalescent theory coupled with Bayesian analysis methods have enabled to reconstruct more complex genealogical histories from molecular data (Hey and Nielsen 2004; Beaumont and Rannala 2004; Hey 2006). The picture emerging from applying coalescent methods in the phylogeographic studies indicates that intermediate stages of divergence (e.g. between polyphyly and monophyly) can reflect significant population isolation (Omland et al. 2006; Peters et al. 2005). In addition, the coalescent methods provide means to estimate e.g. gene flow, genetic diversity and divergence times between populations, which have remained relatively difficult to estimate in the traditional analysis framework (Beaumont and Rannala 2004).

The detection of distinct evolutionary lineages at the intraspecific level in the phylogeographic studies has also been the basis for the definition of evolutionary

significant units (ESUs; Moritz 1994; Avise 2000). The concept of ESU was developed to define valuable units for conservation at the intraspecific level, without warranting further taxonomical classifications (Ryder 1986). The difficulty in identifying such populations is how to set objective criteria for the definition of ESU. Moritz (1994) suggested that populations, which are monophyletic with respect to mtDNA haplotypes and show significant differentiation in nuclear genetic marker allele frequencies, could be considered as ESUs. This criterion is based on the idea that neutral genetic divergence indicates a long independent history of a population and it is likely that the population has accumulated unique characteristics that are relevant in conserving the genetic diversity of the species in interest. Later, this approach has been criticized because it overlooks phenotypic evolution, which can proceed faster than the divergence in neutral genetic markers (Crandall 2000; Merilä and Crnokrak 2001; Fraser and Bernatchez 2001). This is a highly relevant issue in the Northern Hemisphere fish species, where substantial adaptive divergence in young postglacial lakes is of commonplace occurrence (Bernatchez and Wilson 1998; Taylor 1999). It is unlikely that mtDNA genealogy would reach monophyly in such a short timescale.

Adaptive divergence in the Northern Hemisphere fish species

The climatic oscillations during the Pleistocene have also triggered phenotypic divergence among the Nearctic and Palearctic fishes (Bernatchez and Wilson 1998; Taylor 1999). Numerous aquatic environments were formed after the last glaciation (ca. 10 000 years ago) providing novel habitats for the colonizing

fish species (Bernatchez and Wilson 1998; Taylor 1999; Schluter 2000). The phenotypic divergence can be observed in e.g. body size and shape, feeding morphology in many fish genera including *Coregonus* (Østbye et al. 2005a,b; 2006; Landry et al. 2007), *Thymallus* (Koskinen et al. 2002a) and *Gasterosteus* (Taylor and McPhail 2000; Kristjánsson et al. 2002a,b; Leinonen et al. 2006). The main evolutionary force behind adaptive divergence is considered to be directional selection which can cause phenotypic evolution in post-glacial or even in very short timescales (Schluter 2000; Koskinen et al. 2002a; Kristjánsson et al. 2002a; Bell et al. 2004). Furthermore, the Northern Hemisphere fish species provide prime examples of sympatric, allopatric and parapatric specializations to different ecological niches (Bernatchez and Wilson 1998; Taylor 1999). Sympatric ecomorphs that have specialized to e.g. limnetic or benthic habitats within the same lake have been found in three-spined sticklebacks (*Gasterosteus aculeatus*, Taylor and McPhail 2000; Kristjánsson et al. 2002b), arctic charr (*Salvelinus alpinus*; Wilson et al. 2004) and in whitefish (*Coregonus lavaretus*; Østbye et al. 2005a; Rogers and Bernatchez 2007). For example, the limnetic and benthic three-spined stickleback ecomorphs in six lakes in the British Columbia are examples of parallel evolution in postglacial lakes (Taylor and McPhail 2000). The sympatric pairs have evolved independently in each of the lakes as a result of double invasions by their marine ancestors. First invaders adapted to the benthic lifestyle whereas the later colonizers adapted to limnetic habitat. Thus, the evolution of these ecomorphs proceeded with an allopatric phase and a subsequent sympatric phase (Taylor and McPhail 2000). The genetic mechanisms

underlying adaptive divergence have remained largely unresolved, but in recent years quantitative trait locus mapping (QTL) studies have revealed the number and identity of the genes coding for ecologically important traits such as morphology (Peichel et al. 2001; Colosimo et al. 2004, 2005; Shapiro et al. 2004; Cresko et al. 2004; Rogers and Bernatchez 2005, 2007, Rogers et al. 2007).

Molecular markers in genetic studies

Measuring genetic variation within and between populations has been typically carried out by using molecular genetic markers. They provide a simple way to estimate allele frequencies, genetic diversity and genetic distances between populations (Avisé 2004). In recent years, the number of genetic markers available for evolutionary studies even in non-model organisms has increased several fold due to the development of high-throughput technologies in marker screening and genotyping. The fast accumulation of sequence information in the public databases such as GeneBank at NCBI enables the use of existing sequence information for the search and development of genetic markers (Martins et al. 2006; Benson et al. 2007). Furthermore, whole genome sequencing projects are revolutionizing the genetic marker development, giving a chance to focus on specific genomic regions of interest (Hubbard et al. 2007). In the following, the properties of the genetic markers used in this thesis are reviewed

Microsatellites are probably the most versatile nuclear markers available for genetic studies. Microsatellites or simple sequence repeats (SSRs) are short (1-6 bp) tandem repeats, which are common in vertebrate genomes (Beckmann

and Weber 1992). Microsatellites are typically highly polymorphic, which makes them suitable for e.g. parentage analyses, individual identification, linkage mapping and in estimating population differentiation and genetic diversity (Chistiakov et al. 2006). The application of microsatellites in evolutionary studies may be compromised by the fast mutation rate resulting in the saturation of genetic distance estimates (Paetkau et al. 1997). However, some empirical findings suggest that microsatellites can reveal population history even in very long evolutionary time-scales (Hänfling et al. 2002; Koskinen et al. 2002b; Estoup et al. 2002). Moreover, technical problems such as non-amplifying alleles may hamper the inference based on these markers especially in individual identification (Hoffman and Amos 2005). Other genotyping errors resulting from individual mistakes, problems in allele scoring and DNA sample quality may further introduce bias to the inferences based on the microsatellite markers (Bonin et al. 2004; Pompanon et al. 2005).

Animal mitochondrial DNA (mtDNA) has been the marker of choice especially in phylogeographic studies (Avisé 2000). Approximately 70% of the phylogeographic analyses have been based on the mtDNA (Zhang and Hewitt 2003). The fast mutation rate, maternal inheritance, absence of recombination and the relatively easy laboratory methodology makes mtDNA suitable for phylogeographic studies (Avisé 2000; Hewitt 2000). Furthermore, the effective population size of mtDNA is roughly one fourth of the size of the nuclear markers enhancing the effect of genetic drift. On the negative side, mtDNA is effectively one locus and reflects only a small proportion of the total genome content implying that evolutionary inferences might be inflated

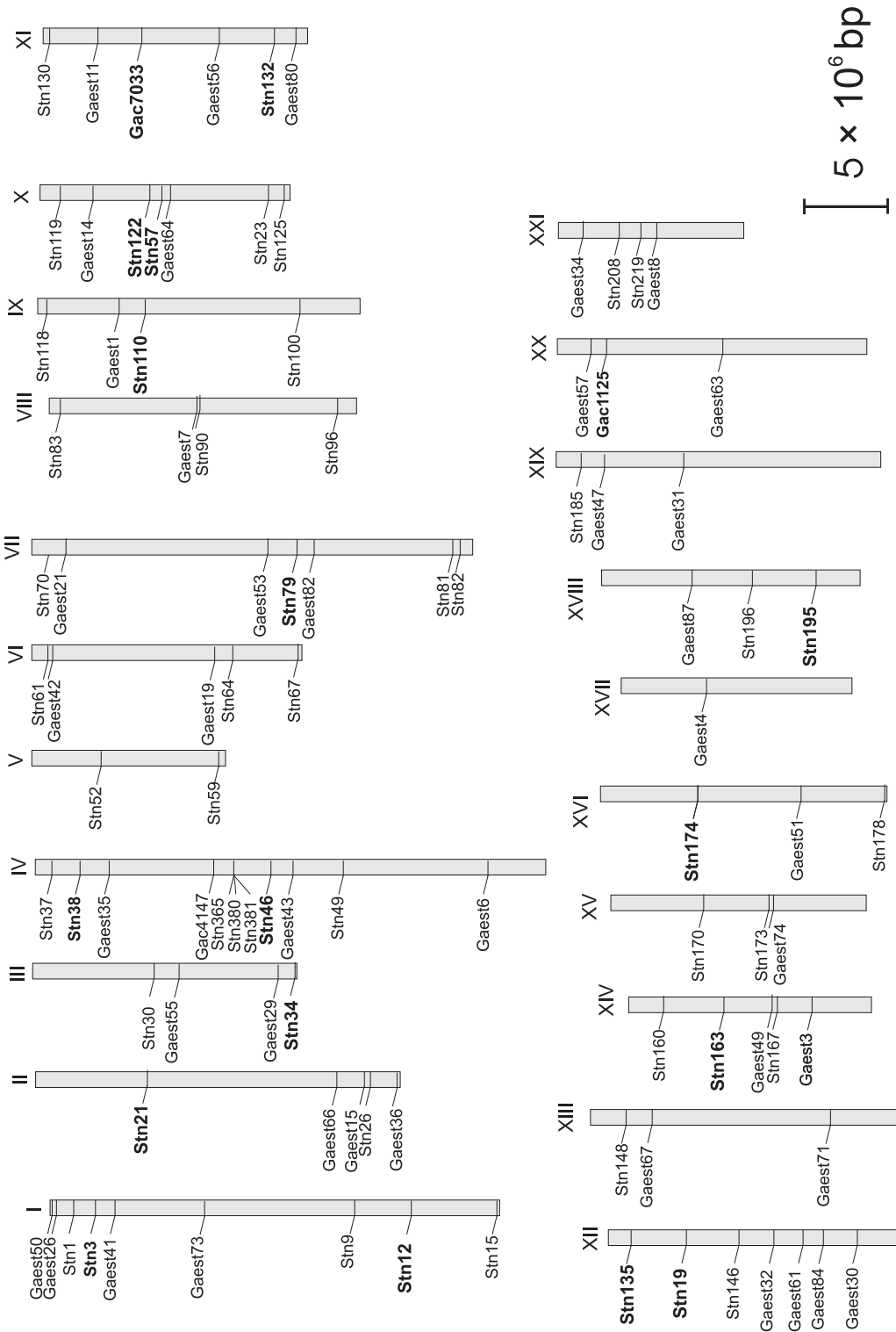


Fig. 1. A physical map of the microsatellite markers used in this thesis. The chromosome map was depicted from whole three-spined stickleback sequence available at http://www.ensembl.org/Gasterosteus_aculeatus/index.html. The complete set of markers were used in the Chapters V and VI and the markers indicated in bold were used in the Chapters I, III and IV.

due to random lineage sorting. In other words, the history of a mtDNA gene might not reflect species history (Pamilo and Nei 1988).

Single nucleotide polymorphisms (SNPs) are gaining popularity in population genetic research and they might have applications in phylogeographic studies as well (Morin et al. 2004; Brumfield et al. 2003). The SNPs are more densely spaced in eukaryotic genomes than microsatellites, providing a possibility for a more fine-scale analysis in linkage mapping studies and in specific genomic regions (Morin et al. 2004). On the other hand, the relatively low levels of polymorphisms in SNP markers require a larger number of SNP loci to be used to reach a similar level of resolution in e.g. individual identification in comparison to highly polymorphic microsatellites (Schlötterer 2004). However, SNP markers provide a method to identify the actual mutation (quantitative trait nucleotides, QTNs), which are responsible for the evolution of a phenotypic trait (Mackay 2001). SNP markers may suffer from ascertainment bias such that only the most polymorphic sites are discovered, which might lead to erroneous estimates of genetic variation as well as population differentiation (Morin et al. 2004; Clark et al. 2005).

It is typically assumed that molecular markers are evolving neutrally and thus are suitable for estimating the neutral processes such as gene flow and random genetic drift (Chistiakov et al. 2006). Recently, there has been growing concern that the assumption of neutrality might not hold true (Bazin et al. 2006; Meiklejohn et al. 2007). In mtDNA, no correlation was found between the expected effective population size and the level of polymorphisms in a meta-analysis including diverse set of taxa (Bazin et al. 2006). In short,

Bazin et al's (2006) study suggested that selection, rather than population history explained the observed pattern, and thus would indicate that application of mtDNA in phylogeographic studies would not be completely warranted. Furthermore, some studies have speculated about the possibility that microsatellites function as 'tuning knobs' of evolution. Due to their high mutation rate, microsatellite loci within genes and in regulatory elements can affect the gene expression levels and thus have impact on phenotypic evolution (Fondon and Garner 2004; Kashi and King 2006). In recent years, much attention has also been paid to the genetic hitchhiking, where the variation of neutral genetic markers is affected by natural selection due to linkage with a selected gene (Maynard-Smith and Haigh 1974). The principle of genetic hitchhiking can be applied in detecting genomic regions affected by natural selection, and thus, for mapping of genes underlying adaptive divergence (Schlötterer 2003)

The three-spined stickleback as a model species

Systematically, there are only two recognized species in the genus *Gasterosteus*: *G. aculeatus* and *G. wheatlandi* (Black-spotted stickleback), the latter of which lives in sympatry with North-eastern North American three-spined sticklebacks, (Mattern 2004). Many members of *Gasterosteidae* are important model species, but especially the three-spined stickleback is widely used in evolutionary biology research (Bell and Foster 1994; McKinnon and Rundle 2002; Östlund-Nilsson et al. 2007). The extreme phenotypic diversity of this species in post-glacial lakes in particular makes it suitable for studies of adaptive divergence,

speciation and vertebrate evolution (Bell and Foster 1994; McKinnon and Rundle 2002; Peichel 2005, Cresko et al. 2007; Östlund-Nilsson et al. 2007). In addition, the relative ease in which populations of the wild origin can be bred in laboratory enables experimental manipulations in quantitative trait mapping and also in physiological and behavioural studies (Peichel 2005; Cresko et al. 2007). The development of genomic resources such as a large number of microsatellite markers, linkage map, expressed sequence tag libraries and whole genome sequencing facilitates the use the three-spined stickleback as a model of vertebrate evolution (Kingsley et al. 2004).

Specific aims of the research

Despite the popularity of the three-spined stickleback as a model, many aspects of its population genetic structure and phylogeography have remained largely unresolved especially in the European distribution area. In addition, studies on the genetic basis of adaptation have just started to emerge, but are mainly restricted to morphological traits. The major scope of this thesis is to

understand the phylogeography of the three-spined sticklebacks in Europe using molecular marker based inferences on colonization history, genetic variation and population differentiation (Chapters **I** and **II**). The molecular data together with morphological measurements are used to identify evolutionary significant units or populations of special conservation value (Chapter **III**). In the Chapters **IV**, **V** and **VI**, the aim was to understand genetic mechanisms underlying phenotypic evolution and to identify genomic regions of adaptive significance. More specifically, in Chapter **IV** the focus is to understand the genetic basis and evolutionary mechanisms of the number of lateral plate evolution in Fennoscandian three-spined stickleback populations. In Chapter **V**, a generic genome scan is performed with microsatellite markers to identify genomic regions showing signals of directional and balancing selection. The final Chapter (**VI**) is a follow up of Chapter **V**, reporting results of a fine-scale hitchhiking mapping in nearby genomic regions of a putatively selected locus aiming to identify ecologically important genes involved in adaptive divergence.

SUMMARY OF THE MAIN RESULTS AND DISCUSSION

LARGE SCALE PHYLOGEOGRAPHY AND POPULATION STRUCTURE OF THREE-SPINED STICKLEBACKS IN EUROPE

Genetic relationships among marine and freshwater three-spined stickleback populations

The large scale phylogeography and the population genetic structure of the three-spined sticklebacks in Europe have been poorly understood. Previous studies have focused only on geographically proximate populations (Reusch et al. 2001; Raeymaekers et al. 2005; Malhi et al. 2006) or have employed a relatively sparse sampling on a global scale (Orti et al. 1994). Similarly, the early inferences based on the morphological traits might be biased due to homoplasy in the evolution of these characters (Münzing 1963; Bell and Foster 1994).

To explore the degree of genetic differentiation among marine and freshwater populations, samples from 73 locations were collected covering the whole species distribution area in Europe (Fig. 2 A). First, microsatellite markers were used to estimate the finescale genetic structure both within and between habitat types and among geographic areas (I). The patterns in allele frequency differentiation suggested that the genetic differentiation among marine (coastal and migratory) populations was low ($F_{ST} = 0.021 - 0.029$) in comparison to the substantial divergence in freshwater (lake and river) populations ($F_{ST} = 0.23 - 0.32$). In addition, genetic diversity among the marine populations ($H_E = 0.78 - 0.80$) was significantly higher than in the freshwater populations ($H_E = 0.56 - 0.64$). The explanation for

these differences might be the nature and geological features of the different habitat types. In the marine environment, gene flow between populations is not restricted due to geographical barriers in comparison to isolated freshwater habitats. It is therefore likely that genetic drift in isolated freshwater populations has resulted in the substantial divergence in microsatellite allele frequencies. Similar patterns in genetic differentiation and diversity have been detected also in local scale studies in Europe and in the Pacific distribution area (Taylor and McPhail 2000; Reusch et al. 2001; Raeymaekers et al. 2005). The low genetic divergence among marine populations is also in line with the results obtained from other marine fish species like cod (*Gadus morhua*; Knutsen et al. 2003) and European hake (*Merluccius merluccius*; Lundy et al. 1999).

The analyses of molecular variance (AMOVA) results suggest that the freshwater populations have been derived multiple times and independently from the relatively uniform ancestral marine population. Only 0.2% of the total variance in allele frequencies was explained by habitat type whereas the geographical origin explained slightly higher percentage of variation (2.7%). Also, the phylogenetic analysis indicated a relatively random clustering of populations from different habitat origins. This pattern is in agreement with the global pattern (McKinnon et al. 2004; Colosimo et al. 2005) and also with local pattern of the origin of freshwater populations (Taylor and McPhail 2000; Raeymaekers et al. 2005; Malhi et al. 2006). In this context, Reusch et al.'s (2001) finding of a monophyletic origin of the river and lake populations in Northern Germany seems

Fig. 2 A

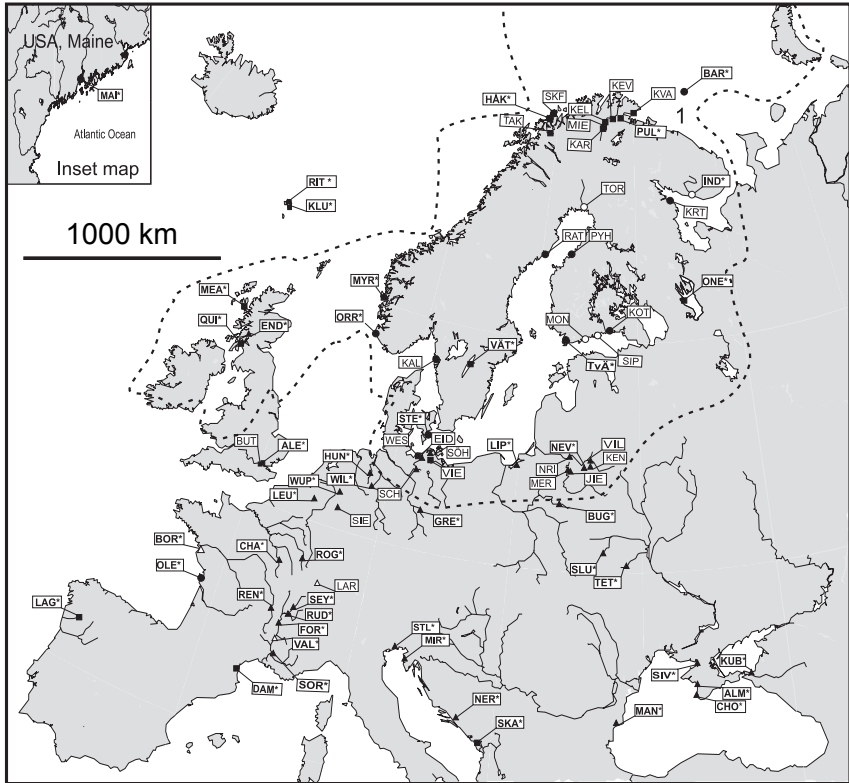
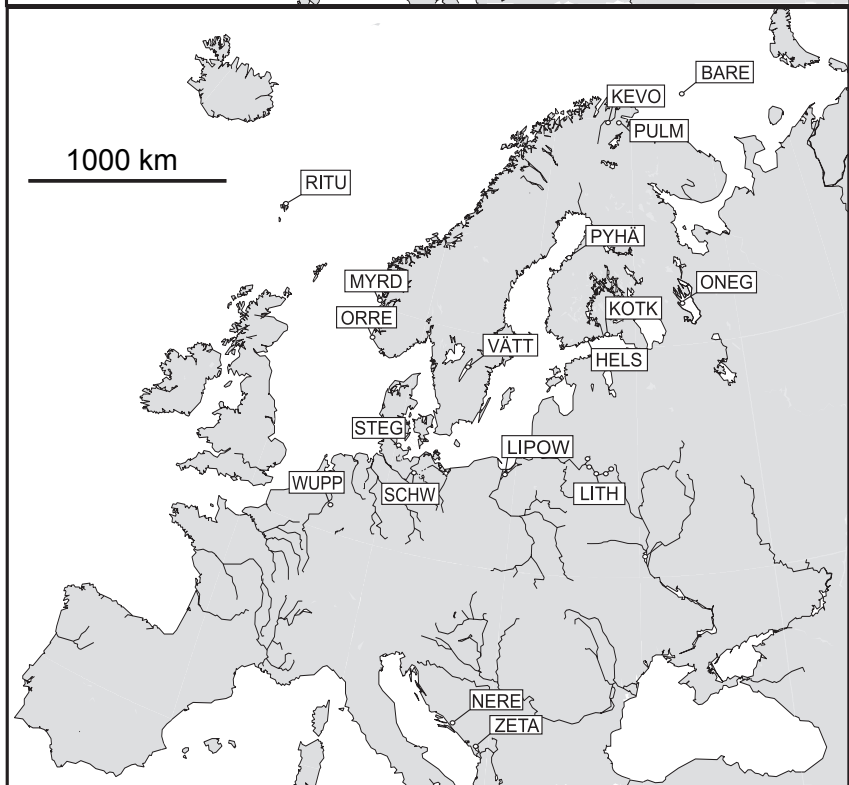


Fig. 2 B



to be an exception. Probably the specific historical hydrogeographic conditions partly explain the local pattern detected in Northern Germany (Reusch et al. 2001). The practical implication of these results are that the independently derived three-spined stickleback populations can be used as a model for studies in parallel evolution in phenotypic traits (McKinnon et al. 2004; Shapiro et al. 2004; Colosimo et al. 2005).

Furthermore, the patterns in population differentiation varied between geographical areas. A comparison of genetic differentiation between an allele frequency estimator (F_{ST}) and an estimator taking into account the contribution of stepwise mutations (R_{ST}) indicated differences in the age of freshwater populations in the latitudinal gradient. Among the Mediterranean lake populations significant differences between R_{ST} and F_{ST} (0.71 vs. 0.2) suggest a longer divergence time than e.g. in Scandinavian lake populations (R_{ST} = 0.05,

F_{ST} = 0.07). This may reflect a divergence originating from the late Pleistocene in the Mediterranean region whereas in Scandinavia freshwater populations are of postglacial origin. The geographical pattern in genetic differentiation makes sense also in the light of the geological history: the Mediterranean region was relatively unaffected by the Pleistocene glaciers in comparison to Scandinavia, which was covered by ice as late as ca. 10000 years ago (Dawson 1992; Eronen et al. 2001).

Mitochondrial DNA phylogeography of the three-spined stickleback in Europe

However, the microsatellite data did not resolve reliably the phylogenetic relationships of the three-spined sticklebacks and thus, the inferences of colonization history were not conclusive. In Chapter II, mtDNA sequence polymorphisms at the partial cytochrome b (965 bp) and control region (433 bp)

Fig. 2 C

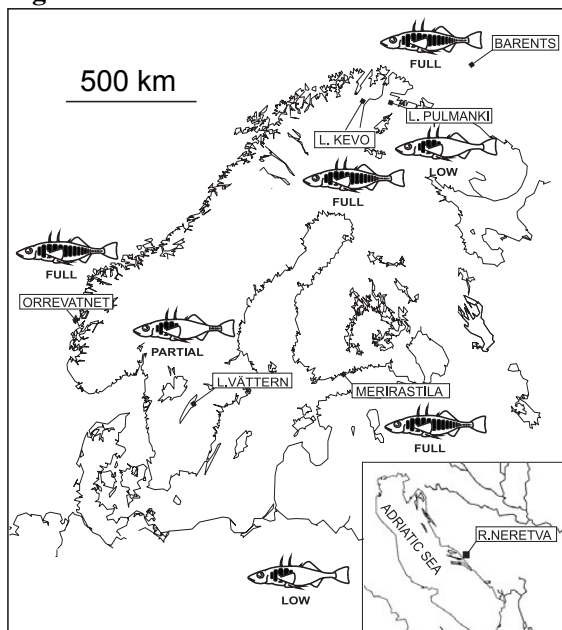


Fig 2 A. A map of sampling sites used in the large-scale phylogeographic studies (Chapter I and III). A subset of 49 sampling locations (labeled in bold and with an asterisks) were used the Chapter II). The dashed line indicated the margin the last glacial maximum ca. 20 Kya (Svendsen et al. 2004).

Fig 2 B. The geographical locations of the study populations to identify conservation units among the three-spined sticklebacks in Europe (Chapter III).

Fig 2 C. The sampling sites of the study populations used in the Chapters IV, V and VI. The plate morphs are indicated beside the population labels.

genes were used to further investigate the population history of three-spined sticklebacks in Europe. Altogether 172 individuals were sequenced from eight marine and forty-one freshwater locations; these samples were a subset from the Chapter I (Fig 1. A). For an outgroup comparison samples of *G. wheatlandi* were collected from North-Eastern North America.

The Bayesian phylogenetic (Ronquist and Huelsenbeck 2003) analysis revealed three major mtDNA lineages in Europe (Fig. 3 A). The basal trans-Atlantic lineage spanned from the East Coast of USA to the freshwaters of Scotland and to the Yonne-Saone basin in France. The European lineage covered populations from the Mediterranean region to Central, Western and Northern Europe whereas the Black Sea lineage was specific to the Black Sea drainages. A similar picture emerged also from the network analysis using parsimony criteria (Clement et al. 2000) to resolve the phylogenetic relationships among haplotypes (Fig. 3 B). The coalescent simulations assuming isolation with migration model (Hey and Nielsen 2004) indicated an effective isolation, but a recent divergence time of the main lineages. The coalescent simulations further suggested large differences in genetic diversity ($\theta = 4N_e\mu$) estimates between the geographical areas occupied by the main lineages. The European lineage harboured substantially larger (ca. 10-fold) genetic diversity in comparison to the trans-Atlantic and Black Sea lineages. In addition, significantly negative Tajima's D and Fu's Fs estimates together with a star-like mtDNA genealogy within the main lineages suggest relatively recent population expansions.

The geographical distribution of the main lineages and phylogenetic tree topology confirmed the earlier suggestion

of Orti et al. (1994) that the Atlantic three-spined sticklebacks have been recently derived from the Pacific. The trans-Atlantic lineage had a basal position relative to the other lineages in Europe. Although the coalescent simulations indicated isolation of the main lineages, the geographical distribution of haplotypes suggested recent trans-Atlantic migrations. Identical haplotypes were found in the East Coast of USA and in a lake population in Scotland and in R. Chamoux in the upper Yonne-Saone basin in France. Indications of long distance migrations have been observed also in the Pacific basin where haplotypes belonging to the Japanese clade have been found in the freshwaters of Alaska and British Columbia (Orti et al. 1994; Johnson and Taylor 2004). It is therefore likely, that long distance migration is occurring both in the Pacific and Atlantic and also historically between these basins.

The divergence time estimates between the main mtDNA lineages indicated a relatively recent 170 – 130 Kya divergence time but the estimates should be considered as tentative given the possible inaccuracies in molecular clock calibrations (Ho and Larson 2006). However, the divergence times seem to roughly coincide with the Saalian glacial maximum (ca. 160-140 Kya) and with a faster evolving molecular clock (10%/Mya) probably with the late Weichselian glacial maximum (ca. 20 Kya; Svendsen et al. 2004). The recent divergence times of the main lineages further support the recent origin of three-spined sticklebacks in the Atlantic. These molecular datings were in contrast to the fossil records in the Atlantic basin since the earliest fossils date back to ca. 1.9 millions of years ago (Bell and Foster 1994). This discrepancy is most likely a result of the extinction of the first Atlantic invaders whereas

the contemporary Atlantic population originates from a more recent colonization by Pacific *G. aculeatus* (Haglund et al. 1992; Orti et al. 1994).

The coalescent simulations indicated large differences in the genetic diversity among the different parts of the three-spined stickleback distribution area. The Black Sea and the Atlantic lineages harboured ca. ten fold less diversity than the populations in mainland Europe. These differences stem most likely from the regional differences in the severity of the Pleistocene climate conditions. During the last glaciation, ice masses covered larger land areas in Northern America in comparison to Europe (Dawson 1992). Furthermore, dramatic sea level changes in the Black Sea basin might have caused reductions in population size (Ryan et al. 2003). The difference in genetic diversities may also arise from a more rigorous sampling of the European populations in

comparison to the East Coast of USA and Black Sea populations.

Identifying conservation units

The results from the Chapters I and II suggested that in the Mediterranean region several freshwater populations have had long independent histories. Especially the populations in Lake Skadar and R. Neretva in the Balkan Peninsula had accumulated unique molecular characteristics and could be classified as ESUs according to the criteria of Moritz (1994). However, setting the conservation criteria to only neutral genetic markers might miss evolution in phenotypic traits and thus relevant characters in aiming to preserve important genetic variation (Crandall 2000; Fraser and Bernatchez 2001).

In order to evaluate the conservation status of the European three-spined stickleback populations, the molecular

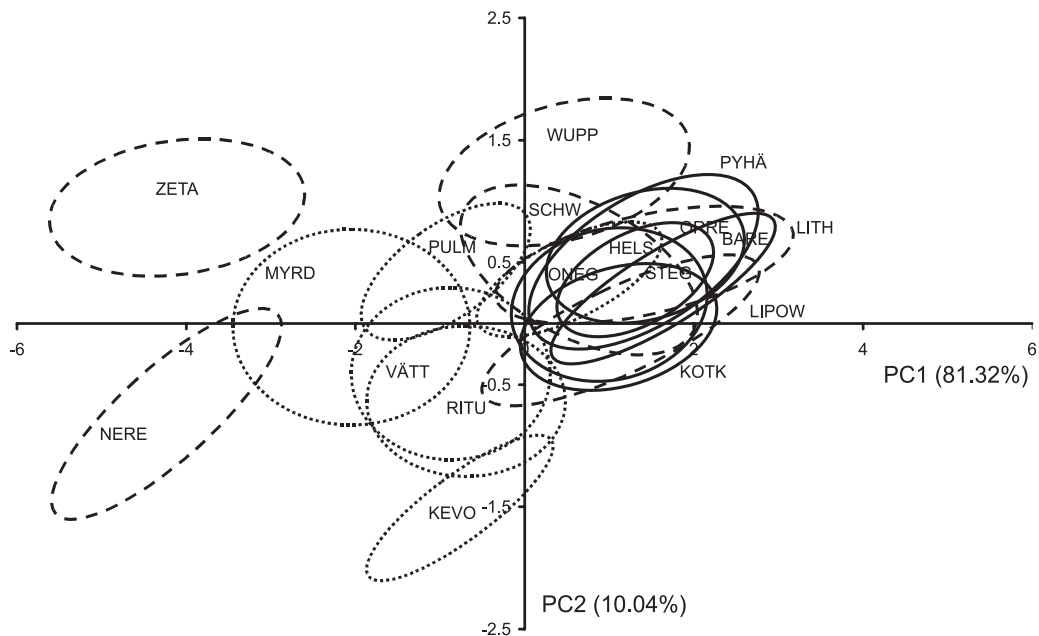


Fig 4. A principal component plot based on the measurements of the body armour traits in Chapter III. The Balkan Peninsula populations (Zeta and Nere) are distinct clusters due to the extreme reduction in the body armour traits.

data from chapters **I** and **II** together with morphological measurements were used to identify ESUs or valuable populations for conservation (**III**). The study populations comprised of a representative set of marine, lake and river populations (six populations from each habitat type) from northern Europe and the Balkan Peninsula (Fig. 2 B). The morphological measurements were taken from body armour traits (*viz.* length of first and second dorsal spines, pelvic spines, pelvic girdle and the standard length of the fish), which have adaptive value and also a strong genetic basis (reviewed in Bell and Foster 1994). The results indicate that the most divergent populations, both in neutral genetic markers and morphology, are the ones located in the Balkan Peninsula. Especially the river populations in R. Neretva and R. Zeta were distinct from all northern conspecifics in all marker classes. The river populations in the Balkan Peninsula had the most reduced bony armour traits in comparison to the northern European populations (Fig. 4). The crosshair classification of the Balkan populations rejected genetic and ecological exchangeability both in recent and historical time-scales, which would suggest managing them as different species (Crandall et al. 2000). However, due to limitations in our sampling, the taxonomical classification of R. Neretva and R. Zeta populations would not be warranted. Only one to six individuals were sequenced from these populations and the DNA samples for R. Zeta originated ~ 30 km apart from the sampling site of the morphological samples. Furthermore, the species delineation in three-spined sticklebacks is complicated due to extreme phenotypic diversification. A common practice has been to classify all divergent populations to belong to the

same species (*G. aculeatus*) to avoid too complex taxonomy (Bell and Foster 1994). Instead, our data strongly suggest that the R. Neretva and R. Zeta can be classified as distinct conservation units (Crandall 2000; Fraser and Bernatchez 2001) and even as evolutionary significant units *sensu* Moritz (1994). It is worth noting that the emphasis on conserving biodiversity in the three-spined stickleback, and also on the other Northern Hemisphere fish species, should not be only on the southern divergent lineages. Previous studies have reported valuable conservation units also in the evolutionarily young postglacial populations (Taylor 1999; Taylor and McPhail 2000; Fraser and Bernatchez 2001). The identification of conservation units in three-spined sticklebacks adds to the previous studies reporting unique fish populations in the Balkan Peninsula (e.g. Bohlen et al. 2003; Šlechtová et al. 2004; Susnik et al. 2007). It may appear that the Balkans is a biogeographically valuable region in Europe to preserve genetic diversity of various fish species and may deserve a special conservation policy.

All in all, the results in Chapters **I**, **II** and **III** support the general pattern emerging from the phylogeographic studies of fish species and other taxa; southern Europe harbours more divergent evolutionary lineages than Northern Europe, which is in agreement with the expectations born out from the differences in geological history (Bernatchez and Wilson 1998; Hewitt 2000). Three-spined sticklebacks were able to persist during the late Pleistocene glacial maximums in the Mediterranean region freshwaters and in the Black Sea. Furthermore, mtDNA data suggests other refugial areas in the western and eastern Atlantic. Probably the marine environment was the main refugial area given that marine

sticklebacks were probably able to migrate along the advancing ice-sheets (Bell and Foster 1994). However, it seems that the three-spined sticklebacks were not able to survive during the glacial maxima in the western European river systems, which are common refugial areas in other freshwater fishes (e.g. Englbrecht et al. 2000; Van Houdt et al. 2005; Østbye et al. 2005b). In addition, cold adapted species like bullhead (*Cottus gobio*) were able to survive in rivers of Southern England and in the Elbe river system, which were close to the last glacial margin (Hänfling et al. 2002). Such northern refugial areas are not likely in case of three-spined sticklebacks at least in the light of mtDNA divergence estimates.

GENETIC BASIS OF ADAPTIVE DIVERGENCE

The evolution of lateral plate number

Maybe the most striking and widespread adaptation with respect to the body armour traits in three-spined sticklebacks is the reduction in the number of lateral plates in the freshwater populations (Bell and Foster 1994). Marine sticklebacks usually have a complete set of lateral plates from head to tail (fully plated morph), but freshwater populations exhibit reduced number of lateral plates. Partially plated fish have ca. half of the plates whereas low plated fish possess only few anterior plates (Bell and Foster 1994). The reduction in the lateral plates may arise as an adaptation to low calcium level, salinity, temperature or different predation regimes

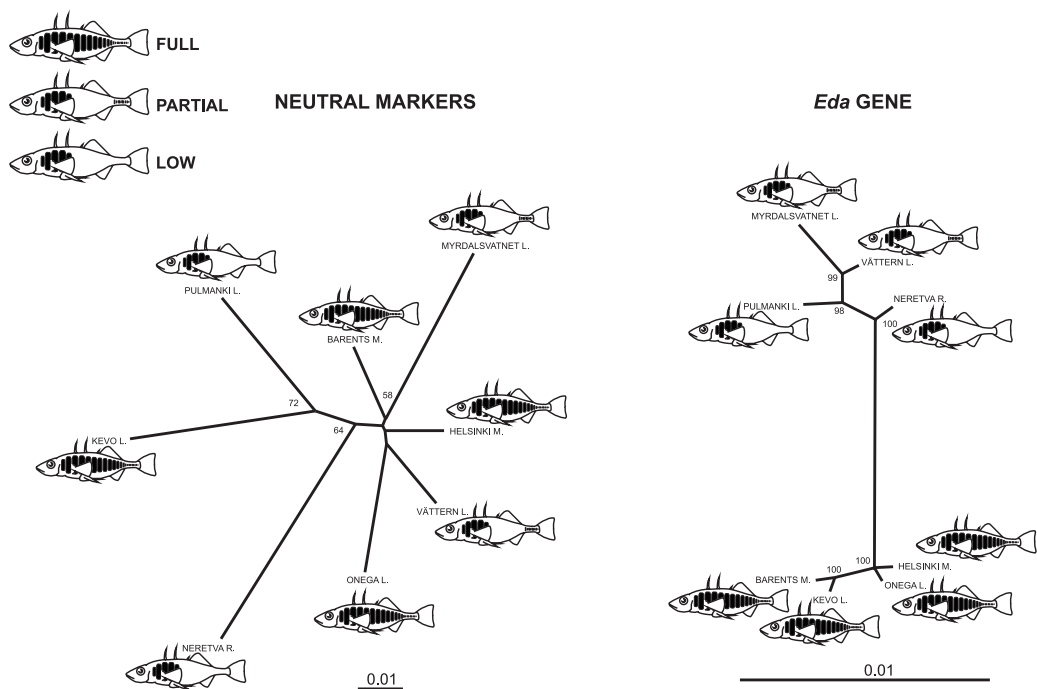


Fig. 5. Phylogenetic trees based on a putatively neutral set of microsatellites (left) and SNP variability at the *Eda*-gene (right). The plate morphs of the study populations are indicated along the population names (Chapter IV).

in the freshwater environment (e.g. Giles 1983; Hagen and Moodie 1982; Reimchen 1995). Several genetic models have been proposed to explain the variability in the lateral plate number. The simplest model suggested that the number of lateral plates is determined by a combination of two alleles; a dominant allele (A) and a recessive allele (a). The fully plated morph would be a result of genotype AA, partial plated Aa and the low plated with a combination of two recessive alleles aa (Münzing 1957). Later, more complicated models with two major loci coding for the number of lateral plates were suggested (Hagen and Gilbertson 1973).

Recent quantitative trait loci mapping studies have revealed one major locus and three modifier loci determining lateral plate variation (Colosimo et al. 2004). A more fine-scale QTL-mapping study identified the *Eda*-gene as the major locus and that the plate morphs were coded by full and low plated alleles segregating in the *Eda*-locus (Colosimo et al. 2005). Furthermore, the reduction in the number of lateral plates in freshwater populations was driven by natural selection changing the frequencies of the two alleles. In marine populations, the full plated allele was in high frequency compared to the low plated allele, but the frequencies were reversed in the freshwater populations (Colosimo et al. 2005).

In the Chapter IV the SNP variability at the *Eda*-locus was studied in eight European three-spined stickleback populations (Fig. 2 C) in comparison to variability in 18 putatively neutral microsatellite loci. The specific aim was to test whether the SNP variability in the *Eda*-gene - as detected on the global scale by Colosimo et al. (2005) - explains the variation in the lateral plates in a set of European populations. This was achieved by selecting freshwater

and marine populations with different plate morphs (low, full and partial). To formally test whether the reduction in the number of lateral plates was driven by natural selection, the quantitative genetic differentiation was estimated with a common garden experiment. Also, a microsatellite marker linked *Eda*-gene in an experimental QTL-cross (Colosimo et al. 2004) was used to investigate whether the divergence in this marker reflects selection in the *Eda*-gene.

The results are in agreement with the global pattern detected by Colosimo et al (2005). In the phylogenetic analysis based on the SNP variability at the *Eda*-gene, populations clustered to distinct clades according to their plate morphs (Fig. 5). Even the fully plated freshwater populations clustered together with the marine full plated morphs suggesting a similar origin of the full plated allele in the Scandinavian freshwaters. In contrast, the microsatellite tree showed a relatively random clustering of the same populations (Fig. 5). The partial plated morph seems to be heterozygote of full and low plated alleles, or homozygous for low plated allele and clustered together with low plated alleles. However, the attempt to demonstrate selection in a microsatellite locus linked to *Eda* failed; the genetic differentiation in this particular locus was not elevated in comparison to a set of putatively neutral microsatellite loci. On the other hand, the differentiation estimate (Q_{ST}) obtained from a common garden experiment clearly exceeded the divergence in neutral microsatellite markers (F_{ST}) suggesting that directional selection - rather than genetic drift - is the main evolutionary force behind the reduction in the lateral plate number.

A more comprehensive genotyping of marine populations showed that the low plated allele occurs in very low (0.2% -

3.2%) frequencies (Colosimo et al. 2005). Therefore, the reduction in the number of lateral plates evolves as a result of local selection favoring the low plated allele in freshwater populations carried by the marine colonizers (Colosimo et al. 2005). Thus, the low plated morph has evolved multiple times independently in many freshwater systems throughout the three-spined stickleback distribution area in the Pacific and Atlantic (Colosimo et al. 2005). In general, the reduction in lateral plates is a prime example of parallel evolution. Similar parallel evolution has been shown to be commonplace also in other body armour traits, such as pelvic reduction, albeit the genetic mechanisms for the latter have been shown to be different (Shapiro et al. 2004). Namely, the sequence variability at the *Pitx1* locus did not explain the loss of pelvic spines in freshwater populations, but was shown to be controlled by tissue specific regulation of transcription (Shapiro et al. 2004). In the populations with reduced pelvic spines, the *Pitx1* seemed to be ‘switched off’ during the development of the spines (Shapiro et al. 2004).

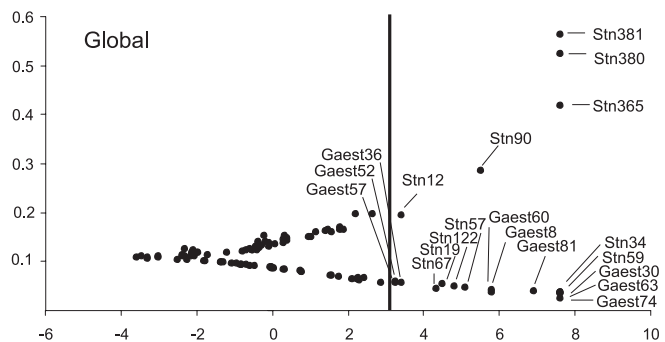
Detecting targets of natural selection at the molecular level

In recent years, the emphasis in evolutionary biology has been moving from the genetic marker based inferences

of population structure to the identification of the number and function of the actual genes that are coding for ecologically important traits (MacKay 2001; Luikart et al. 2003; Orr 2005a, b; Stinchcombe and Hoekstra 2007; Ungerer et al. 2007). This approach is becoming feasible due to the increase in the availability of genetic markers and also the possibility to develop genetic markers from the publicly available DNA sequence databases and from whole genome sequences (Benson et al. 2007; Storz and Hoekstra 2007; Hubbard et al. 2007). Identification of genes involved in evolutionary change might open a possibility to evaluate more specific questions such as how many genes code for a given phenotypic trait, and which evolutionary and ecological forces are involved in the evolution of these genes (Stinchcombe and Hoekstra 2007). A picture emerging from the available studies indicates that it is complicated to answer the above questions and would require a combination of detailed genomic and ecological studies (MacKay 2001; Feder and Mitchell-Olds 2003; Vasemägi and Primmer 2005; Ungerer et al. 2007).

In Chapters V and VI studies aiming to find signatures of natural selection at the molecular level were conducted. The objective was not only to perform an explorative genome scan, but also to narrow down the genomic region under selection, and further investigate the

Fig 6. The results of the Bayesian F_{ST} -test to detect loci subject to natural selection (Chapter V). Loci in the upper distribution were considered to indicate directional selection whereas the loci in the lower distribution indicated balancing selection. The vertical line shows the cut-off value for statistical significance at the 5% level.



possibility to identify the actual gene(s) involved in adaptation. In short, this approach is based on the population genetic theory, which predicts that natural selection affects the patterns of population differentiation and genetic variability at the nearby genomic regions of the selected locus – commonly known as genetic hitchhiking (Maynard-Smith and Haigh 1974). In standard hitchhiking mapping studies a large number of genetic markers are needed to distinguish the footprint of selection from the background level of population divergence (Schlötterer 2003; Storz et al. 2005). In comparison to QTL-mapping approach, no prior information of the phenotype is needed in the hitchhiking mapping. This might open a possibility to identify ecologically important genes that are not easily measurable at the phenotypic level, although the task of linking the selected gene to the ecological context might be more difficult (Schlötterer 2003).

An explorative genome scan in three-spined stickleback populations

In Chapter V, I conducted a generic genome scan with 103 microsatellite and two indel markers in three marine and four freshwater three-spined stickleback populations (Fig. 1). The marine populations originated from the Baltic Sea (Merirastila), North Sea (Orrevatnet) and from the pelagic area from the Barents Sea. The set of freshwater populations comprised of L. Vättern in southern Sweden, L. Pulmanki and L. Kevo from Finnish Lapland. For comparative purposes one distantly located river population (R. Neretva) was included in the analysis (Fig. 2 C). Specific care was taken in selecting the microsatellite markers to maximize the chances to detect ‘footprints’ of natural

selection. First, microsatellites linked to QTLs were chosen to identify selection operating on morphological traits. Second, to find novel candidate loci of natural selection, a large set of EST (expressed sequence tag) – associated microsatellites were developed and screened from the available EST-libraries at the NCBI. Microsatellite loci within ESTs are tightly linked to functional genes and provide a useful way to screen for gene-associated polymorphisms (Vasemägi et al. 2005). In order to identify the potential targets of natural selection – or outlier loci – both allele frequency (F_{ST} ; Beaumont and Balding 2004) and reduction in genetic diversity ($\ln RH$; Kauer et al. 2003) based tests were used. The logic behind these neutrality tests is that marker loci subject to balancing selection exhibit lower-than-average and directional selection loci higher-than-average degree of genetic differentiation (Beaumont 2005).

The overall pattern of selection reflected the predominant role of balancing selection over directional selection: 15 loci showed a footprint of balancing selection, but three loci indicated directional selection in a F_{ST} -based analysis (Fig. 6). This result is in agreement with the theory suggesting that balancing selection is more common than directional selection (e.g. Kimura 1981). However, to confirm the generality of this pattern more populations and loci would be needed given the low genomic coverage (ca. 0.0023% of the estimated gene number) and number of populations used in this study. Furthermore, most of the genome scan studies have concentrated on detecting directional selection loci and overlook balancing selection, which makes comparisons with other species difficult. The number of loci indicating directional selection (2.8%) was comparable to previous genome scan studies of various

taxa reporting 1.4-9.5% of the screened loci as outliers (reviewed in Stinchombe and Hoekstra 2007). Some authors have questioned the genome scan approach in finding targets of natural selection due to the low number of loci detected as putatively selected (Eyre-Walker 2006). However, it appears that the genome coverage of typical genome scans utilizing 100-200 genetic marker loci is extremely low (e.g Vasemägi et al. 2005; Kane and Rieseberg 2007) relative to the total gene number.

The strongest signal of directional selection emerged from three loci located in the intronic regions of the *Eda*-gene coding for the lateral plates. The other loci linked to the morphological QTLs did not show any indications of selection suggesting that the selection is not operating on these traits in our study populations. Alternatively, recombination and mutation may obscure the signal of selection unless the marker loci are not tightly linked to the causative gene like in the case of *Eda* (Raeymaekers et al. 2007). It is therefore likely that the attempt to demonstrate selection in a marker locus distantly located to *Eda*-gene in the chapter IV failed due to the recombination breaking down the linkage between *Eda* and the microsatellite locus informative in a QTL-cross.

The attempt to identify novel candidate loci underlying adaptive divergence was successful although only a handful of loci (Stn90, Stn12, Gaest84 and Gaest87) showed indications of directional selection. Locus Stn90 in chromosome VIII was the most potential outlier given the highly significant deviation ($P = 0.002$) from neutral expectations. In addition, the outlier status of this locus was confirmed in the Barents Sea and L. Pulmanki comparison using the Ln RH test (Ln RH = 2.01, $P < 0.05$). In the locus-

by-locus AMOVA-analysis 14.3% of the total variance of the Stn90 was explained by a comparison between marine and freshwater populations. Thus Stn90 might be associated to the adaptation to the freshwater.

A fine-scale mapping of a selective sweep

In the final Chapter (VI), a fine-scale hitchhiking mapping was performed in flanking regions of the candidate locus Stn90 identified in the previous study, which was possible due to the availability of the whole genome sequence of the three-spined stickleback (Hubbard et al. 2007). Microsatellite repeats have been annotated to the whole genome sequence of the three-spined stickleback (available at http://www.ensembl.org/Gasterosteus_aculeatus/index.html) and enables developing genetic markers in very specific genomic regions. The whole genome sequence allowed also a more detailed investigation of the genome content of the region flanking Stn90. In the processing of the whole genome sequences, protein and DNA sequence similarities from different species are used in the identification and prediction of genes (Hubbard et al. 2007). Once the genome sequence is in hand, the putative homologies of the genes in the candidate interval are easily available without requiring any time consuming genomic analysis at this stage.

By genotyping more loci in the nearby genomic regions of the candidate locus it would be possible to get further evidence for natural selection in this particular genomic region. If the signal of selection emerges from multiple loci, it is unlikely that the signal would be a false positive (Schlötterer 2003; Wiehe et al. 2007). In addition of confirming the action of selection on this genomic

region, the aim was also to identify the actual target of selection, i.e. the gene that is ecologically important. In order to assess the above objectives, twenty-four microsatellite markers were developed in ca. 800 kb region flanking the candidate locus *Stn90* (Fig. 7) and the patterns of genetic differentiation and diversity were compared to the putatively neutral set of markers from the previous Chapter (V) excluding the markers under balancing or directional selection. Here, the same set of study populations were used as in Chapter V (Fig. 2 C).

The results indicate a fairly large and continuous genomic region close to the candidate locus deviating from neutral expectations based on the differentiation in allele frequencies (F_{ST}) and the reduction in genetic diversity ($\ln RH$; Fig. 7). The most striking pattern in line with a selective sweep scenario was detected in the comparison between the Barents Sea and L. Pulmanki populations. Altogether, twelve loci in the candidate interval deviated from neutrality in this comparison in the $\ln RH$ test. A modified $\ln RH$ test (Wiehe et al. 2007), which accommodates information from multiple loci, excluded neutrality with a highly significant P-value ($P < 0.001$) in the candidate interval in the L. Pulmanki population. A similar type of result was achieved in the F_{ST} -test including all study populations but only seven loci were statistically significant outliers at the 2.5% cut-off level (Fig. 7). However, the overall pattern of selection in this genomic interval was rather complex. The identity and number of loci were not entirely the same between different neutrality tests, and also some loci outside of the continuous region displayed deviations from neutrality (Fig. 7). Furthermore, bottleneck analysis indicated that the genetic diversities at the

candidate interval deviated from the model of constant population size in at least the freshwater populations. This pattern was not observed in the putatively neutral set of markers suggesting that selection may have operated in the candidate interval also in other populations than in L. Pulmanki.

The occurrence of the selective sweep mainly in the L. Pulmanki population suggests that the genome interval contains genes important in the adaptation to the freshwater environment. This is in line with the general pattern of adaptive divergence in three-spined sticklebacks; most of the phenotypic diversification has evolved in freshwater populations (Bell and Foster 1994; McKinnon and Rundle 2002). Entering from the marine environment to freshwater exposes three-spined sticklebacks to entirely different conditions in terms of e.g. salinity, temperature and predator fauna. These are factors that might trigger divergence in ecologically important traits driven by natural selection and could be detected at the molecular level.

The chromosomal region subject to a selective sweep when comparing Barents Sea with L. Pulmanki was ca. 90 kb. Looking at the genescan gene predictions indicated that this region contained several genes, which had protein homologies to known genes such as *GAMT* and *DAZ* (Fig. 7). Both genes have putative biological functions in spermatogenesis according to the Gene Ontology categories (Harris et al. 2006). However, *GAMT* has been shown to be involved in body size regulation in mice (Vitarius et al. 2006). Body size has been shown to be under genetic control in Fennoscandian populations and has also adaptive value (McKinnon et al. 2004; Leinonen et al. 2006). Thus, *GAMT* would be an excellent starting point for a more detailed sequence analysis. Moreover, the

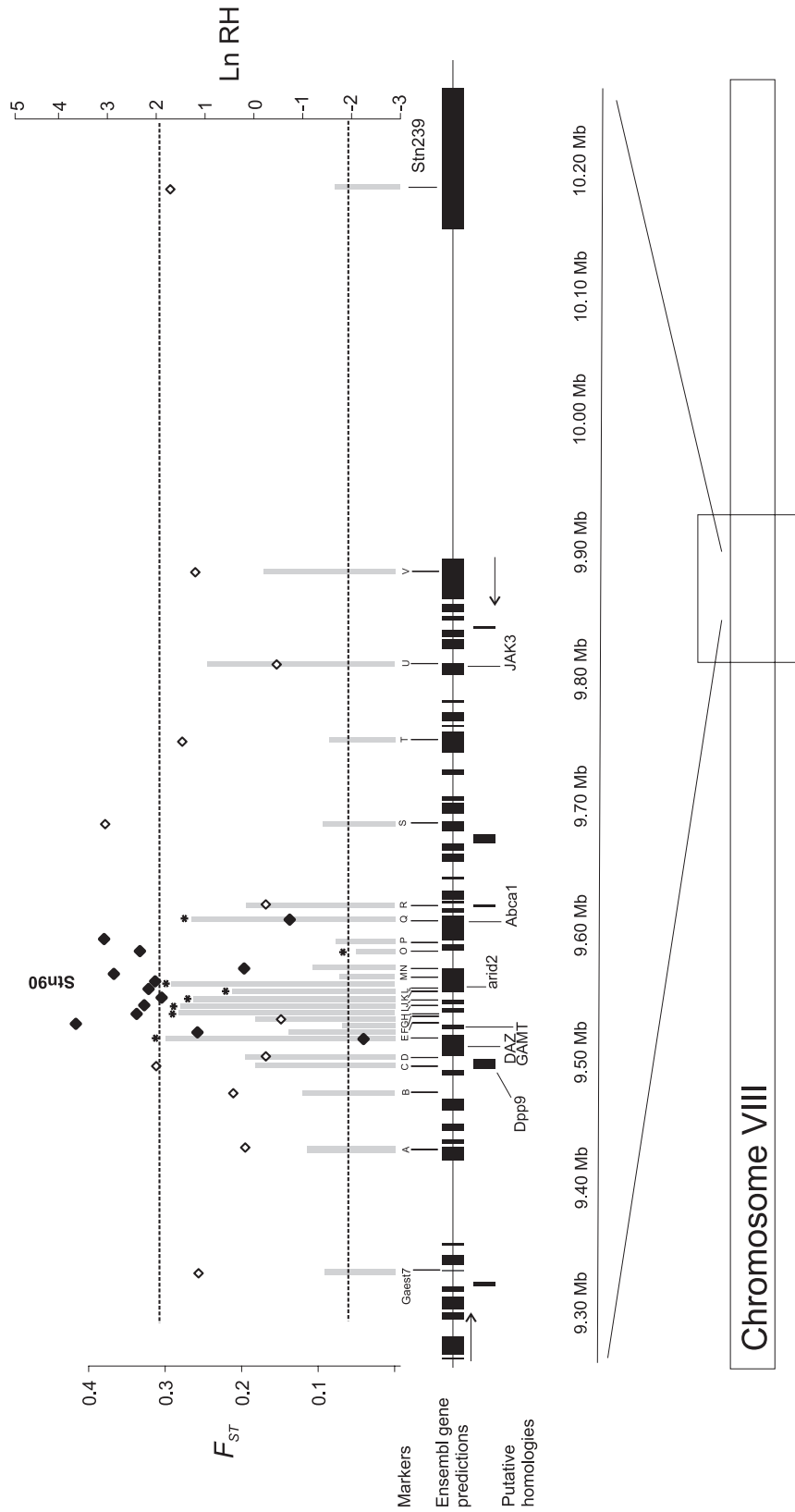


Fig. 7. The results of the fine-scale mapping study in the Chapter VI. The grey bars depict Bayesian F_{ST} -estimates (locus effects) and the asterisk at the top of the bar indicate significant deviations from neutrality at the 5%-level. The filled squares are the $\ln RH$ -values which were included in the multilocus test (Wiehe et. al. 2007). The horizontal dashed lines show the expected distribution $\ln RH$ -values under neutrality (-1.96 - 1.96). The black boxes are Genescan gene predictions.

genomic region might contain an array of genes with similar functions.

The results of the final Chapter (VI) demonstrates that the hitchhiking mapping approach would be useful in narrowing down genomic regions underlying adaptive divergence in wild three-spined

stickleback populations. Similar results have been obtained in natural populations of fruit fly (*Drosophila melanogaster*, Harr et al. 2002) and house mouse (*Mus musculus*, Ihle et al. 2006) due to natural selection at the molecular level.

CONCLUSIONS AND FUTURE DIRECTIONS

The major aim of this thesis was to investigate large-scale phylogeographic patterns and adaptive genetic differentiation among the three-spined stickleback populations in Europe. Based on the results of Chapter I, it was possible to conclude that freshwater populations were derived from multiple and independent colonizations by their marine ancestors. Future studies concentrating on phenotypic evolution might use this background information as a starting point to investigate how certain phenotypic characters have evolved. Studies similar to those as conducted on the reduction of the number of lateral plates (Colosimo et al. 2005), might reveal more cases of parallel evolution. The observation that similar mtDNA haplotypes were found in freshwater habitats in continental Europe and the Western Atlantic suggests that three-spined sticklebacks are capable of trans-Atlantic migrations. However, a more detailed sampling in European populations would probably shed light on the specific environmental factors promoting long distance migrations. For example, the frequency of Japanese clade haplotypes in the lakes of the Pacific basin was related to the lake elevation (Johnson and Taylor 2004).

The results in Chapters V and VI revealed some interesting patterns in natural selection at the molecular level. The ubiquity of balancing selection over directional selection is in agreement with the expectations of phenotypic evolutionary models and common sense: balancing selection should be the dominant

mode of selection. Nevertheless, this issue might benefit from future studies using a larger set of populations from different environments together with larger number of genetic markers. The fine-scale mapping study of Chapter VI shows that the hitchhiking approach would be feasible in wild three-spined stickleback populations. Further studies utilizing hitchhiking mapping would uncover genomic regions of adaptive divergence and would increase the knowledge of the functional genome evolution in three-spined sticklebacks. Perhaps the most intriguing results emerged from the final chapter (VI): the large genomic region in the chromosome VIII deviating from neutral expectations indicates that the genomic region may contain gene(s) important in adaptation to freshwater. With a more detailed sequence analysis it might be possible to identify the actual gene underlying the adaptive trait and estimate its fitness effects with experimental studies under different environmental conditions. Finally, this thesis proposes conservation guidelines to preserve valuable genetic diversity in the three-spined stickleback populations in Europe. The highly divergent freshwater populations in the Balkan Peninsula seem to deserve a special conservation status on the basis of results presented in the Chapter III. Conserving divergent freshwater populations is not important only for future evolutionary studies, but also contributes to the preservation of the biodiversity of the European ichthyofauna.

ACKNOWLEDGEMENTS

First of all, I want to thank my supervisor Juha Merilä for giving me a chance to start a PhD project in my thirties in the Ecological Genetics Research Group. Juha gave me relatively free hands to participate in designing my research and develop ideas, which gave confidence for a ‘young’ scientist.

I am also indebted to several hard-working post-docs in EGRU. I’m extremely grateful for José Manuel Cano having always time even for my naïve questions about statistical methods. Your contribution at the final stages of this thesis was substantial and is highly appreciated. Cim, your help was also crucial to get the job done, not only in laboratory work related issues but you were always willing to discuss of my work. Takahito Shikano must be the fastest genotyper in the history of MES-laboratory. Without Taka’s help it would not have possible to conduct the final chapter of this thesis.

I would also like to thank Tuomas Leinonen for the memorable field trips to L. Vättern, Baltic coast and Lapland in the summers of 2003 and 2004. Adventures in Kevo Biological Station and around Lake Pulmanki have engraved in my memory. Tuomas also allowed me to use his morphological data in my thesis, which is highly appreciated. Likewise, I got extensive help from Vilppu and Katja-Riikka to conduct the large scale genotyping projects. Thanks for the former and present members of EGRU for all the help and company during these years.

I would like to thank my collaborators in Europe and in North America for sending me such a comprehensive collection of sticklebacks. To be honest, Finnish sticklebacks are rare and not very well known, which we found out when

we were starting this project. Thus, the stickleback samples from you laid the foundation for this thesis. This is the list: Jörg Freyhof, Henry Persat, Per-Arne Amundsen, Aleksei Veselov, Arne Nolte, Teija Aho, Ari Haikonen, Ignacio Doadrio, Arne Levsen, Per Sjöstrand, Jaakko Paju, Anders Berglund, Mark Lazzari, Vladimír Kovács, Piotr Debowski, Dánjul Petur Højgaard, Volker Loeschcke, Ari Saura, Lennart Persson, Susie Coyle, Pat Monaghan, Dmitry Lajus, Thorsten Reusch and Audrius Steponėnas. Especially, Jörg Freyhof provided an immense set of samples from the Black Sea, Mediterranean region and from tributaries of river Dnieper, these places were totally inaccessible for us. Henry Persat collected loads of sticklebacks from France and was also interested in my results throughout these years.

The lively PhD community of the department deserves a virtual high five. Wednesday coffees were special occasions for me to put aside sometimes stressful work. Not only being good friends you helped me to sort out several every day problems related to computers and tricky software. In addition, I got substantial help from Raisa Nikula, Asta Audzijonyte and Louisa Orsini in phylogenetic analyses. Heikki, thanks for helping me in basically everything related to my work and being a good friend outside the office hours. A privilege has been to share morning coffees with Marjo and Anna-Liisa; what a perfect way to start a working day with two good humored colleagues. I already miss those moments.

It has been always nice to work in the fruitful atmosphere of the MES-laboratory. I have learned everything what I know from laboratory work during

these years from various people. Maybe the most influential years were with the Relict group, when I was working as a technician. Educational years with Risto Väinölä, Jukka Palo, Jouni Vainio and Tytti Kontula laid a solid ground to my understanding in population genetics and laboratory methods. Later, Jukka Palo has helped me a lot with phylogenetic analyses and he was also interested in reading my manuscripts beside his own work.

I wish to thank the Turku people: J-P, Anti, Anni, Laura, Paula and Craig for useful discussions on fish population genetics and always welcoming me to visit the exotic city of Turku.

I had also an opportunity to participate several courses in Finland and abroad. The stickleback molecular genetics course at Stanford University was an opportunity to meet all 'big names' in stickleback research. Lectures by Mike Bell, Donald McPhail, David Kingsley and Dolph Schluter gave me a lot of ideas for my own work and the intensive days in the lab taught me several cutting-edge methods in molecular genetics. The

Finnish Graduate School in Populations Genetics organized several useful courses at the Hailuoto Biological Station, which were educationally important and also a place meet other graduate students from Universities in Finland. During the EMBO course in 'statistical methods for identifying genes under natural selection' I was able to get hand-to-hand teaching from Jukka Corander, Christian Schlötterer, Mark Beaumont and Ziheng Yang, which was of crucial importance for analyzing the data in the final chapters.

This work has been supported by the Academy of Finland, The Finnish Graduate School in Population Genetics and travel grants from Chancellor of the University of Helsinki and Kymiyhtiön 100-vuotissäätiö.

Finally, I thank my loving mother for taking care of me and providing me a peaceful place in Kuusankoski to escape the sometimes hectic days of this PhD project. This thesis is dedicated to the memory of my father, who passed away at the early stages of this work.

LITERATURE CITED

- Asthana S, Schmidt S, Sunyaev S (2005) A limited role for balancing selection. *Trends in Genetics*. **21**, 30-32.
- Avise JC (2000) *Phylogeography – The History and Formation of Species*. Harward University Press. Cambridge, MA.
- Avise JC (2004). *Molecular Markers, Natural History, and Evolution*, Second Edition. Sinauer Sunderland, MA.
- Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science*. **312**, 570-572.
- Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution*. **20**, 435-440.
- Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*. **13**, 969-980.
- Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*. **5**, 251-261.
- Beckman JS, Weber JL (1992) Survey of human and rat microsatellites. *Genomics*. **12**, 627-631.
- Bell MA, Aguirre WE, Buck NJ (2004) Twelve years of contemporary armor evolution in a threespine stickleback population. *Evolution*. **58**, 814-824.
- Bell MA, Foster SA (1994) Introduction to the evolutionary biology of the threespine stickleback. In: *The Evolutionary Biology of the Threespine Stickleback* (Eds Bell MA, Foster SA). Oxford University Press, Oxford.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2007) GenBank. *Nucleic acids Research*. **35**, Database issue, D21-25.
- Bernardi G (2007) The neoselectionist theory of genome evolution. *Proceedings of the National Academy of Science*. **104**, 8385-8390.
- Bernatchez L, Wilson CC (1998) Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology*. **7**, 431-452.
- Bohlen J, Šlechtová V, Sanda R, et al. (2003) *Cobitis ohridana* and *Barbatula zetensis* in the River Moraca basin, Montenegro: distribution, habitat, population structure and conservation needs. *Folia Biologica*. **51** Suppl, 147-153.
- Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*. **13**, 3261-3273.
- Bromham L, Penny D (2003) The modern molecular clock. *Nature Reviews Genetics*. **4**, 216-224.
- Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*. **18**, 249-256.
- Charlesworth B (1994) The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research*. **63**, 213-227.
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*. **134**, 1289-1303.
- Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*. **2**, e64.
- Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*. **255**, 1-29.
- Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*. **15**, 1496-1502.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*. **9**, 1657-1659.

- Colosimo PF, Hosemann KE, Balabhadra S, et al. (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*. **307**, 1928-1933.
- Colosimo PF, Peichel CL, Nereng K, et al. (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology*. **2**, E109.
- Coyne JA, Orr HA (2004). *Speciation*. Sinauer Associates, Sunderland, MA.
- Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*. **15**, 290-295.
- Cresko WA, Amores A, Wilson C, et al. (2004) Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences, U.S.A.* **101**, 6050-6055.
- Cresko WA, McGuigan KL, Phillips PC, Postlethwait JH (2007) Studies of threespine stickleback developmental evolution: progress and promise. *Genetica*. **129**, 105-126.
- Darwin CR (1859) *The Origin of Species by Means of Natural Selection*. John Murray, London.
- Dawson AG (1992). *Ice Age Earth*. Routledge Press, London.
- Englbrecht CC, Freyhof J, Nolte A, Rassmann K, Schliewen U, Tautz D (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-pleistocene origin of the major central European populations. *Molecular Ecology*. **9**, 709-722.
- Eronen M, Glückert G, Hatakka L, Van De Plassche O, Van Der Plicht J, Rantala A (2001) Rates of Holocene isostatic uplift and relative sea-level lowering of the Baltic in SW Finland based on studies of isolation contacts. *Boreas*, **30**, 17-30.
- Estoup A, Jarne P, Cornuet J (2002) Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*. **11**, 1591-1604.
- Eyre-Walker A (2006) The genomic rate of adaptive evolution. *Trends in Ecology & Evolution*. **21**, 569-575.
- Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Reviews Genetics*. **4**, 651-657.
- Fisher RA (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- Fondon JW, 3rd, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proceedings of the National Academy of Sciences, U.S.A.* **101**, 18058-18063.
- Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*. **10**, 2741-2752.
- Harris MA, Clark JI, Ireland A, Lomax J, Ashburner J et al. (2006) The Gene Ontology (GO) project in 2006. *Nucleic Acids Research*. **34**, D322-326.
- Gibbard PL (1988) The history of the great northwest European rivers during the past three million years. *Proceedings of the Royal Society of London. Series B. Biological sciences*, **318**, 559-602.
- Giles N (1983) The possible role of environmental calcium levels during the evolution of phenotypic diversity in Outer Hebridean populations of the three-spined stickleback, *Gasterosteus aculeatus*. *Journal of Zoology*. **199**, 535-544.
- Hagen DW, Moodie GE (1982) Polymorphism for plate morphs in *Gasterosteus aculeatus* on the east coast of Canada and a hypothesis for their global distribution. *Canadian Journal of Zoology*. **60**, 1032-1042.
- Hagen DW, Gilbertson LG (1973) The genetics of plate morphs in freshwater threespine sticklebacks. *Heredity*. **31**, 75-84.
- Haglund TR, Buth DG, Lawson R (1992) Allozyme variation and phylogenetic relationships of Asian, North American, and European populations of the threespine stickleback, *Gasterosteus aculeatus*. *Copeia*. **1992**, 432-443.
- Harr B, Kauer M, Schlötterer C (2002) Hitchhiking mapping: a population-based

- fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences. U.S.A.* **99**, 12949-12954.
- Hänfling B, Hellemans B, Volckaert FA, Carvalho GR (2002) Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Molecular Ecology*. **11**, 1717-1729.
- Hedrick PW (2005) *Genetics of Populations* (Third Edition). Jones and Bartlett, Boston, MA.
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. **359**, 183-195.
- Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*. **10**, 537-549.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*. **405**, 6789-6907.
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*. **58**, 247-276.
- Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development*. **16**, 592-596.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*. **167**, 747-760.
- Ho SYW, Larson G (2006) Molecular clocks: when times are a-changin'. *Trends in Genetics*. **22**, 79-83.
- Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*. **14**, 599-612.
- Hubbard TJ, Aken BL, Beal K, et al. (2007) Ensembl 2007. *Nucleic Acids Research*. **35**, Database issue, D610-617.
- Hudson RR (1983) Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology*. **23**, 183-201.
- Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution*. **23**, 790-797.
- Johnson LS, Taylor EB (2004) The distribution of divergent mitochondrial DNA lineages of threespine stickleback (*Gasterosteus aculeatus*) in the northeastern Pacific Basin: post-glacial dispersal and lake accessibility. *Journal of Biogeography*. **31**, 1073-1083.
- Kane NC, Rieseberg LH (2007) Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics*. **175**, 1823-1834.
- Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics*. **22**, 253-259.
- Kimura M (1981) Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons. *Proceedings of the National Academy of Sciences*. **78**, 5773-5777.
- Kimura M (1968) Evolutionary rate at the molecular level. *Nature*. **217**, 624-626.
- Kingsley DM, Zhu B, Ozoegava K, De-Jong PJ, Schein J, Marra M et al. (2004). New genomic tools for molecular studies of evolutionary change in sticklebacks. *Behaviour*. **141**, 1331-1344.
- Knutsen H, Jorde PE, Andre C, Stenseth NC (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology*. **12**, 385-394.
- Kontula T, Väinola R (2001) Postglacial colonization of Northern Europe by distinct phylogeographic lineages of the bullhead, *Cottus gobio*. *Molecular Ecology*. **10**, 1983-2002.

- Koskinen MT, Haugen TO, Primmer CR (2002a) Contemporary fisherian life-history evolution in small salmonid populations. *Nature*. **419**, 826-830.
- Koskinen MT, Nilsson J, Veselov AJ, Potutkin AG, Ranta E, Primmer CR (2002b) Microsatellite data resolve phylogeographic patterns in European grayling, *Thymallus thymallus*, Salmonidae. *Heredity*. **88**, 391-401.
- Koskinen MT, Ranta E, Piironen J, et al. (2000) Genetic lineages and postglacial colonization of grayling (*Thymallus thymallus*, Salmonidae) in Europe, as revealed by mitochondrial DNA analyses. *Molecular Ecology*. **9**, 1609-1624.
- Kreitman M (1996) The neutral theory is dead. Long live the neutral theory. *BioEssays*. **18**, 678-683.
- Kristjánsson BK, Skúlason S, Noakes DLG (2002a) Rapid divergence in a recently isolated population of threespine stickleback (*Gasterosteus aculeatus* L.). *Evolutionary Ecology Research*. **4**, 659-672.
- Kristjánsson BK, Skúlason S, Noakes DLG (2002b) Morphological segregation of Icelandic threespine stickleback (*Gasterosteus aculeatus* L.). *Biological Journal of the Linnean Society*. **76**, 247-257.
- Lande R, Shannon S (1996) The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*. **50**, 434-437.
- Landry L, Vincent WF, Bernatchez L (2007) Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology*. **20**, 971-984.
- Leinonen T, Cano JM, Mäkinen H, Merilä J (2006) Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology*. **19**, 1803-1812.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*. **74**, 175-195.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*. **4**, 981-994.
- Lundy CJ, Moran P, Rico C, Milner RS, Hewitt GM (1999) Macrogeographical population differentiation in oceanic environments: a case study of European hake (*Merluccius merluccius*), a commercially important fish. *Molecular Ecology*. **8**, 1889-1898.
- Lynch M (1996) A quantitative genetic perspective on conservation issues. Pp 471-501. In: Avise JC & JL Hamrick: *Conservation Genetics*. Case Histories from Nature. Chapman & Hall, NY.
- Mackay TF (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics*. **35**, 303-339.
- Malhi RS, Rhet G, Bell AM (2006) Mitochondrial DNA evidence of an early Holocene population expansion of threespine sticklebacks from Scotland. *Molecular Phylogenetics and Evolution*. **40**, 148-154.
- Mangerud J, Jakobsson M, Alexanderson H, et al. (2004) Ice-dammed lakes and rerouting of the drainage of northern Eurasia during the last glaciation. *Quaternary Science Reviews*. **23**, 1313-1332.
- Martins W, de Sousa D, Proite K, Guimaraes P, Moretzsohn M, Bertoli D (2006) New softwares for automated microsatellite marker development. *Nucleic Acids Research*, **34**, e31.
- Mattern MY (2004) Molecular phylogeny of the Gasterosteidae: the importance of using multiple genes. *Molecular Phylogenetics and Evolution*. **30**, 366-377.
- Maynard Smith J, Haigh J (1974). The hitchhiking effect of a favourable gene. *Genetical Research*. **23**, 23-35.
- McKinnon JS, Mori S, Blackman BK, et al. (2004) Evidence for ecology's role in speciation. *Nature*. **429**, 294-298.
- McKinnon JS, Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution*. **17**, 480-488.

- Meiklejohn CD, Montooth KL, Rand DM (2007) Positive and negative selection on the mitochondrial genome. *Trends in Genetics*. **23**, 259-263.
- Merilä J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*. **14**, 892-903.
- Mills LS, Allendorf FW (1996) The one-migrant-per-generation rule in conservation and management. *Conservation Biology*. **10**, 1509-1518.
- Morin PA, Luikart G, Wayne RK, the SNP workshop group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*. **19**, 208-216.
- Moritz C. 1994. Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology & Evolution*. **9**, 373-375.
- Münzing J (1963) The evolution of variation and distributional patterns in European populations of the three-spined stickleback, *Gasterosteus aculeatus*. *Evolution*. **17**, 320-332.
- Münzing J (1959) Biologie, Variabilität und Genetik von *Gasterosteus aculeatus* L. (Pisces): Untersuchungen im Elbegebiet. *Int Revue Ges. Hydrobiol.* **44**, 317-382.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*. **39**, 197-218.
- Ohta T (2002) Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences. U.S.A.* **99**, 16134-16137.
- Ohta T, Gillespie JH (1996) Development of neutral and nearly neutral theories. *Theoretical Population Biology*. **49**, 128-142.
- Ohta T (1992) The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*. **23**, 263-286.
- Omland KE, Baker JM, Peters JL (2006) Genetic signatures of intermediate divergence: population history of Old and New World Holarctic ravens (*Corvus corax*). *Molecular Ecology*. **15**, 795-808.
- Orr HA (2005a) The genetic basis of reproductive isolation: insights from *Drosophila*. *Proceedings of the National Academy of Sciences, U.S.A.* **102** Suppl 1, 6522-6526.
- Orr HA (2005b) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*. **6**, 119-127.
- Orti G, Bell MA, Reimchen TE, Meyer A (1994) Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. *Evolution*. **48**, 608-622.
- Østbye K, Amundsen PA, Bernatchez L, et al. (2006) Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. *Molecular Ecology*. **15**, 3983-4001.
- Østbye K, Naesje TF, Bernatchez L, Sandlund OT, Hindar K (2005a) Morphological divergence and origin of sympatric populations of European whitefish (*Coregonus lavaretus* L.) in Lake Femund, Norway. *Journal of Evolutionary Biology*. **18**, 683-702.
- Østbye K, Bernatchez L, Naesje TF, Himberg KJ, Hindar K (2005b) Evolutionary history of the European whitefish *Coregonus lavaretus* (L.) species complex as inferred from mtDNA phylogeography and gill-raker numbers. *Molecular Ecology*. **14**, 4371-4387.
- Östlund-Nilsson S, Mayer I, Huntingford FC (2007) *The biology of the three-spined stickleback* (Eds. Östlund-Nilsson S, Mayer I, Huntingford FC). CRC press, NY.
- Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (*Ursidae*) populations. *Genetics*. **147**, 1943-1957.
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*. **5**, 568-583.
- Peichel CL (2005) Fishing for the secrets of vertebrate evolution in threespine sticklebacks. *Developmental Dynamics*. **234**, 815-823.

- Peichel CL, Nereng KS, Ohgi KA, et al. (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*. **414**, 901-905.
- Peters JL, Gretes W, Omland KE (2005) Late Pleistocene divergence between eastern and western populations of wood ducks (*Aix sponsa*) inferred by the 'isolation with migration' coalescent method. *Molecular Ecology*. **14**, 3407-3418.
- Pompanon F, Bonin A, Bellemain E, Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*. **6**, 847-859.
- Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution*. **16**, 37-45.
- Raeymaekers JA, Maes GE, Audenaert E, Volckaert FA (2005) Detecting Holocene divergence in the anadromous-freshwater three-spined stickleback (*Gasterosteus aculeatus*) system. *Molecular Ecology*. **14**, 1001-1014.
- Raeymaekers JA, Van Houdt JK, Larmuseau MH, Geldof S, Volckaert FA (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.
- Reimchen TE (1995) Predator-induced cyclical changes in lateral plate frequencies of *Gasterosteus*. *Behaviour*. **132**, 1079-1094.
- Reusch TB, Wegner KM, Kalbe M (2001) Rapid genetic divergence in postglacial populations of threespine stickleback (*Gasterosteus aculeatus*): the role of habitat type, drainage and geographical proximity. *Molecular Ecology*. **10**, 2435-2445.
- Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus* sp. Salmonidae) species pairs. *Molecular Biology and Evolution*. **24**, 1423-1438.
- Rogers SM, Isabel N, Bernatchez L (2007) Linkage maps of the dwarf and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. *Genetics*. **175**, 375-398.
- Rogers SM, Bernatchez L (2005) Integrating QTL mapping and genome scans towards the characterization of candidate loci under parallel selection in the lake whitefish (*Coregonus clupeaformis*). *Molecular Ecology*. **14**, 2, 351-361.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. **19**, 1572-1574.
- Ryan WBF, Major CO, Lericolais G, Goldstein SL (2003) Catastrophic flooding of the Black Sea. *Annual Review of Earth & Planetary Sciences*. **31**, 525-554.
- Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology & Evolution*. **1**, 9-10.
- Schlötterer C (2004) The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics*. **5**, 63-69.
- Schlötterer C (2003) Hitchhiking mapping - functional genomics from the population genetics perspective. *Trends in Genetics*. **19**, 32-38.
- Shapiro MD, Marks ME, Peichel CL, et al. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*. 717-723.
- Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.
- Šlechtová V, Bohlen J, Freyhof J, Persat H, Delmastro GB (2004) The Alps as barrier to dispersal in cold-adapted freshwater fishes? Phylogeographic history and taxonomic status of the bullhead in the Adriatic freshwater drainage. *Molecular Phylogenetics and Evolution*, **33**, 225-239.
- Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*. Special Issue on Ecological and Evolutionary Functional Genomics. Published early online.

- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*. **14**, 671-688.
- Storz JF, Hoekstra HE (2007) The study of adaptation and speciation in the genomic era. *Journal of Mammalogy*. **88**, 1-4.
- Susnik S, Snoj A, Wilson IF, Mrdak D, Weiss S (2007) Historical demography of brown trout (*Salmo trutta*) in the Adriatic drainage including the putative *S. letnica* endemic to Lake Ohrid. *Molecular Phylogenetics and Evolution*. **44**, 63-76.
- Svendsen JI, Alexanderson H, Astakhov VI, et al. (2004) Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews*. **23**, 1229-1271.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*. **7**, 453-464.
- Taylor EB, McPhail JD (2000) Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proceedings of the Royal Society of London. Series B. Biological sciences*, **267**, 2375-2384.
- Taylor (1999) Species pairs of north temperate freshwater fishes: Evolution, taxonomy, and conservation. *Reviews in Fish Biology and Fisheries*. **9**, 299-324.
- Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? *Genome Research*. **16**, 702-712.
- Ungerer MC, Johnson LC, Herman MA (2007) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity*. Special Issue on Ecological and Evolutionary Functional Genomics. Published early online.
- Van Houdt JK, De Cleyn L, Perretti A, Volckaert FA (2005) A mitogenic view on the evolutionary history of the Holarctic freshwater gadoid, burbot (*Lota lota*). *Molecular Ecology*. **14**, 2445-2457.
- Vasemägi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution*. **22**, 1067-1076.
- Vasemägi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*. **14**, 3623-3642.
- Vitarius JA, Sehayek E, Breslow JL (2006) Identification of quantitative trait loci affecting body composition in a mouse intercross. *Proceedings of the National Academy of Sciences, U.S.A.*, **103**, 19860-19865.
- Vucetich JA, Waite TA (2000) Is one migrant per generation sufficient for the genetic management of fluctuating populations? *Animal Conservation*. **3**, 261-266.
- Wiehe T, Nolte V, Zivkovic D, Schlötterer C (2007) Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. *Genetics*. **175**, 207-218.
- Wilson AJ, Gislason D, Skúlason S, Snorrason S, Adams CE, Alexander G, et al. (2004) Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular Ecology*. **13**, 1129-1142.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*. **16**, 97-159.
- Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*. **12**, 563-584.
- Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London. Series B, Biological Sciences*. **271**, 561-564.