Adaptation to contrasting habitats and Heterozygosity-fitness correlations in Eelgrass (*Zostera marina*)

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Katharina Keil

Kiel 2011

Referent: Prof. Dr. Thorsten Reusch
Korreferent: Prof. Dr. Frank Melzner

Tag der mündlichen Prüfung: 01.07. 2011
Zum Druck genehmigt: 01.07. 2011

gez. Prof. Dr. Lutz Kipp, Dekan

Contents

Summary

Seagrasses are ecosystem engineering species of outstanding importance for soft sediment coastal habitats. An ongoing decline of seagrass meadows in the last decades has caused rising interest in the ecology, genetics and evolution of these species. Eelgrass Zostera marina is distributed in the northern hemisphere in widely diverging habitats. In the Wadden Sea of the North Sea, contrasting habitat conditions have led to two differentiated forms of Z. marina. Plants situated on mudflats are exposed to higher environmental stress and show a reduced growth with narrower leaves and predominant sexual reproduction compared to the more robust subtidal form that also reproduces clonally via rhizomes. In order to detect habitat dependent selection and adaptation, I decided to conduct multilocus genome scans. Genome scans are a bottom-up approach using molecular markers, which make them suitable for non model organisms where no genomic information is available. In principle, genome scans apply statistical tests to detect outlier loci that fall outside a neutral distribution and are thus likely to be influenced by selection. As there is little congruence among the detected outliers when different models are compared, I used a combination of several models, a suitable approach for an explorative genome scan.

As a first step, I developed gene-linked microsatellite markers from an expressed sequence tags (EST) database and proved that they are easily identified by in silico search. Initial screening revealed that EST microsatellites proved to be almost as variable as anonymous markers. Information about the genes these markers are linked to make them a valuable tool for genome scanning as the putative functions of genes located close to outlier loci are known.

In a subsequent genome scan, I tested a first set of 25 microsatellites, 14 of which were EST- derived, for habitat dependent selection. I made use of a three times replicated habitat contrast that served to affirm the outliers detected by two different neutrality tests. Results showed that EST microsatellites were not more often detected as outliers than anonymous markers in our study. I found signs of divergent selection in three loci that were repeatedly detected in all three habitat contrasts. Two of them are EST

microsatellites, and one is linked to a putative nodulin gene building channels in cellular membranes suggesting a functional link with habitat differences.

In order to assess the validity of genome scans, I re-tested our three population pairs by adding single nucleotide polymorphisms (SNP), a novel marker type for Z. marina, adding a third neutrality test and by nearly doubling the number of markers assessed. Results were reassuringly consistent with the first scan as the three markers detected earlier were confirmed by the extended approach. Additionally, three other loci showed consistent signs of selection across habitat contrasts. Two newly detected loci are linked to an acid phosphatase gene (related to hyper-osmotic stress) and a seed maturation protein, respectively, both suggesting effects of habitat dependent selection in the Wadden Sea environment.

I also aimed to investigate how fitness and multilocus heterozygosity (MLH) are related in Z. marina. Using 37 microsatellites, I assessed the correlation of fitness measured as clone size and MLH in two Z. marina populations from the Baltic Sea. I showed for the first time that the intermediate heterozygosity principle, stating that offspring from parents with an intermediate level of relatedness will be have the highest fitness, applies in a clonal plant. In ecosystems that are structured and maintained by only one species, like seagrass beds, heterozygosity represents an important component of genetic variation that has rarely been assessed before.

Overall, the results of this work suggest that multilocus genome scans can be a useful tool to detect habitat dependent selection in exploratory studies. It has to be kept in mind, though, that inherent caveats of genome scans must be taken seriously and repeated population pairs represent the only way to handle this problem. Multilocus heterozygosity has rendered valuable insights into possible effects of inbreeding and outbreeding in natural populations, which is even more meaningful in ecosystems built by a single clonal species.

Zusammenfassung

Seegräser sind ‘Ökosystem-Ingenieure‘ mit herausragender Bedeutung für Flachwasser-Habitate auf Weichböden. Ein anhaltender Rückgang der Seegraswiesen in den letzten Jahrzehnten hat zu gesteigertem Interesse an der Ökologie, Genetik und Evolution dieser Arten geführt. Das große Seegras *Zostera marina* L. ist auf der gesamten nördlichen Hemisphäre und in sehr unterschiedlichen Habitaten verbreitet. Im Wattenmeer der Nordsee tritt *Z. marina* abhängig von kontrastierenden Habitaten in zwei verschiedenen Formen auf. Pflanzen auf Wattflächen sind größerem umweltbedingten Stress ausgesetzt und weisen eine verringerte Wuchshöhe und schmalere Blätter, sowie vorherrschend sexuelle Reproduktion auf. Dahingegen zeigt die sublitorale, robuste Form auch verbreitet klonale Reproduktion durch verzweigte Rhizome. Um habitatabhängige Selektion und Adaptation nachzuweisen, verwendete ich ‚Genome Scans‘ mit multiplen Loci. Genome Scans sind ein ‚Bottom-up‘ Ansatz, der molekulare Marker verwendet und demnach für Nicht-Modellorganismen geeignet ist, für die wenig genomische Information vorhanden ist. Das Prinzip der Genome Scans beruht auf der Anwendung statistischer Tests, um Ausreißer zu entdecken, die nicht innerhalb der Grenzen einer neutralen Verteilung liegen und demnach wahrscheinlich durch Selektion beeinflusst sind. Als passenden Ansatz für einen explorativen Genome Scan verwendete ich eine Kombination mehrerer Test-Modelle, da die gefundenen Ausreißer zwischen verschiedenen Modellen wenig überlappen.

Als ersten Schritt entwickelte ich gengekoppelte Mikrosatelliten- Marker aufgrund einer Datenbank von exprimierten Sequenzen (expressed sequence tags, EST) und zeigte, dass sie durch einen In-Silico Suchverfahren einfach identifiziert werden können. Erste Auswertungen zeigten dass EST- Microsatelliten in etwa genauso variabel wie herkömmliche anonyme Marker sind. Die Informationen über Gene, in deren naher Umgebung diese Marker lokalisiert sind, machen sie zu wertvollen Werkzeugen für Genome Scans, da mögliche Funktionen dieser Gene zum Teil bekannt sind.

In einem anschließenden Genome Scan untersuchte ich 25 Mikrosatelliten, von denen 14 gengekoppelt (aus EST-Sequenzen) waren, auf habitatabhängige Selektion. Ich

verwendete drei Replikate eines Habitatkontrasts um Ausreißer, die aufgrund zweier Neutralitätstests entdeckt worden waren, zu bestätigen. Meine Resultate zeigten, dass in meiner Studie EST-Mikrosatelliten nicht öfter als Ausreißer erkannt wurden als anonyme Marker. Drei Marker zeigten Spuren divergierender Selektion in allen drei Habitatkontrasten. Zwei davon sind EST-Mikrosatelliten, einer darunter ist an ein Gen gekoppelt, dass vermutlich für ein Nodulin-Protein kodiert, welches Wasserkanäle in Zellmembranen aufbaut. Dies legt einen funktionellen Zusammenhang mit Unterschieden in den Habitaten nahe.

Um die Validität von Genome Scans zu überprüfen, führte ich einen erneuten Test mit den drei Populationspaaren durch, wobei mit Single Nucleotide Polymorphisms (SNP) ein neuer Markertyp eingeschlossen wurde, zusätzlich ein dritter Neutralitätstest verwendet wurde und die Anzahl der Marker fast verdoppelt wurde. Der Ergebnisse waren weitgehend konsistent zu den des vorherigen Scans, indem die drei bereits erkannten Ausreißer bestätigt wurden. Darüber hinaus zeigten drei weitere Loci Anzeichen von Selektion in mehreren Habitatkontrasten. Zwei der neu endeckten Loci sind and ein Saure Phosphatase-Gen (steht in Zusammenhang zu hyperosmotischem Stress) bzw. an ein Gen für die Samenreifung gekoppelt, was in beiden Fällen auf einen Zusammenhang mit habitatabhängiger Selektion im Wattenmeer hinweist.

Eine weitere Untersuchung hatte zum Ziel, den Zusammenhang zwischen Fitness und Multi-Lokus Heterozygotie (MLH) bei Z. marina zu untersuchen. Ich verwendete 37 Mikrosatelliten und untersuchte die Korrelation der Fitness, gemessen als Klongröße, mit der MLH in zwei Z. marina-Populationen aus der Ostsee. Ich konnte zum ersten Mal zeigen dass das Prinzip der Mittleren Heterozygotie, welches besagt, dass die Nachkommen von Eltern, die eine mittlere genetische Verwandschaft aufweisen, die höchste Fitness besitzen, in einer klonalen Pflanze existiert. In Ökosystemen, die im Wesentlichen von einer einzigen Art aufgebaut und strukturiert werden, wie Seegraswiesen, ist Heterozygotie eine wichtige Komponente der genetischen Variation, die bisher noch unzureichend erforscht ist.

Die Ergebnisse dieser Arbeit zeigen, dass Multi-Lokus Genome Scans ein nützliche Werkzeug sein können, um habitatabhängige Selektion in explorativen Studien zu

entdecken. Die Nachteile von Genome Scans sollten jedoch nicht außer Acht gelassen werden, zu deren Umgehung nur die Verwendung wiederholter Populationspaare ermächtigt. Die Untersuchungen zur Multi-Lokus Heterozygotie haben wertvolle Einsicht in mögliche Effekte von Inzucht und Auszucht in natürlichen Populationen ermöglicht. Diese Aussage gewinnt umso mehr Bedeutung, wenn es sich um Ökosysteme handelt, die von einer einzigen klonalen Pflanzenart dominiert werden.

Introduction

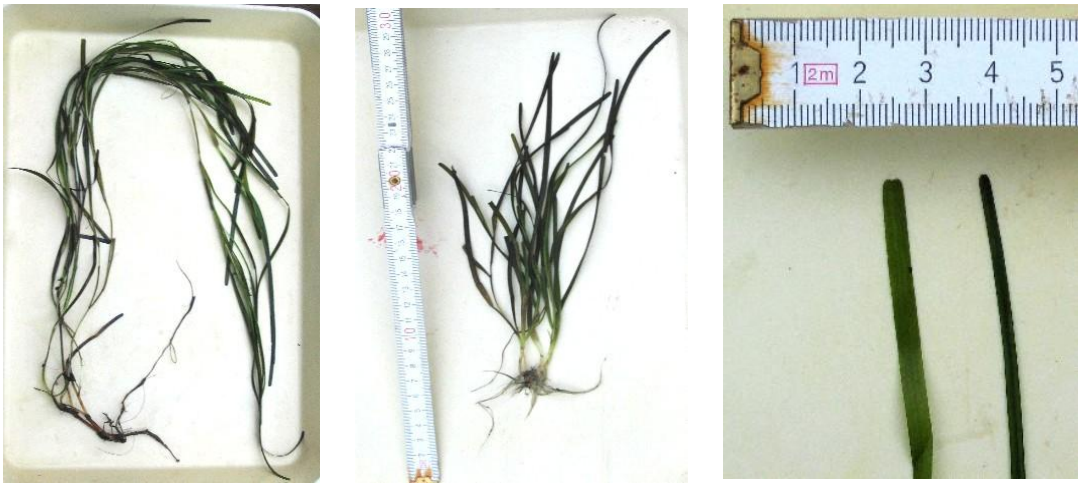
Life in contrasting habitats: Eelgrass Zostera marina

Ecosystem engineering species are of great importance in the terrestrial and marine environment as they maintain, structure and manipulate ecosystems (Jones et al. 1994). Forming extended stands along soft-sediment coasts, seagrasses belong to the most relevant species for ecosystem building and habitat provision in all water bodies except for the Polar Regions (den Hartog 1970). Seagrasses comprise a polyphyletic group of angiosperms that have returned to the marine environment from freshwater habitats (Les et al 1997). Seagrasses possess typical features of higher plants including flower stands, shoots consisting of a short stem and long ribbon-like leaves, aerenchym and rhizome with short roots. Reproduction can be clonally by new shoots growing out of rhizomes or sexually by subaqueous flowering and pollination. Although selfing is possible, several mechanisms exist to impede selfed progeny, the most important of which is the high prevalence of dioecy among the seagrasses. In monoecious species such as the genus Zostera, proterandry and cryptic self-incompatibility (Hämmerli and Reusch 2003a) hinder selfing.

Seagrass meadows belong to the world's most productive ecosystems and provide numerous functions, such as nutrient retention, sediment stabilization and provision of food and nursery habitat for various invertebrates and fish (Hemminga and Duarte 2000). Additionally, seagrasses regulate their environment in terms of physical and chemical characteristics (Enriquez et al. 2001; Madsen et al. 2001). Due to their economic relevance, seagrasses have been rated among the most valuable ecosystems worldwide (Costanza et al. 1997). Unfortunately, a decline of seagrass meadows has been reported in the last decades which gives rise to serious concern (Short and Wyllie-Escheverria 1996). Since the first quantitative records of seagrass areas have been taken, almost 30 % of seagrass meadows have disappeared and records of area losses at studied sites exceed those of gains 10-fold. Extrapolating these figures suggests that about 50,000 square km of seagrass meadows have been lost during the last 130 years (Waycott et al. 2009). This decline is attributed to a combination of anthropogenic and

natural factors, namely eutrophication, pollution, increased turbidity, mechanical disturbance, hydrodynamic changes and diseases (Hemminga and Duarte 2000).

Eelgrass *Zostera marina* is the most successful seagrass in temperate waters, making it the dominant seagrass in the northern hemisphere (den Hartog 1970). It is widely distributed along European coasts, inhabiting the Wadden Sea area in the southern and eastern part of the North Sea. During the 1930ies, the wasting disease caused by the protist Labyrinthula sp. caused almost all submerged eelgrass meadows too die off. Additionally, a combination of eutrophication and changes in water currents has drastically reduced potential areas suitable for seagrass meadows (Van Katwijk 2000a; Reise et al. 1989). The result was an almost complete loss of seagrass in the entire Wadden Sea area. Only intertidal populations remained, growing on mudflats where flooding and dry-falling alternate. Today, the largest portion of the remaining seagrass meadows is located in the North Frisian Wadden Sea North to the Eiderstedt peninsula (Germany), which comprises about 80% of the total seagrass meadow area in the Wadden Sea (Reise et al. 2005).



Phenotypic differences of two *Z. marina* forms. Robust, broad-leaved subtidal form (left), small and narrow-leaved intertidal form (middle), comparison of leaves (right)

There are marked differences in the phenology of intertidal vs. submerged plants (Jacobs 1982; Reusch 2002). While intertidal plants grow only up to 25 cm high and have narrow leaves, the subtidal form can reach heights up to 1 m with broader and more robust leaves. This is accompanied by differences in reproduction mode, as intertidal plants reproduce sexually be seeds mainly. In the subtidal form, asexual (clonal) reproduction occurs additionally, leading to spatially extended clones partly connected by rhizomes.

These differences in morphology and reproduction between intertidal and subtidal plants have been attributed to contrasting habitat conditions (McMillan and Phillips 1979; Backman 1991). Within small distances, the Wadden Sea offers strikingly different environmental conditions, for example regarding the situation on tidal flats compared to tidal creeks (a quasi-subtidal habitat as creeks are always filled with water). On tidal flats, plants are exposed to temperature changes that can reach extreme levels during spring and early summer (temperature difference of about 30°C between water-covered and sun-heated surface). Salinity fluctuations are considerable as heavy rain dilutes surface water on dry-fallen mudflats to almost freshwater conditions. Additionally, plants on tidal flats have to tolerate direct UV radiation not filtered by the water column, and they have to resist a certain degree of desiccation (although most of the Z. marina individuals are located in small water-filled pits where even at low tide, leaves seem to be covered most of the time).

Intertidal (left) and subtidal (right) habitat of contemporary Seagrass *Z. marina* in the North Frisian Wadden Sea

The genetic basis of the two forms is controversial. One assumption is that both forms are genetically identical and only morphologically differentiated (phenotypic plasticity, e.g. Gagnon et al. 1980). In line with this, a recent study from Brittany revealed that morphological differences between the two different *Z. marina* forms do not correspond to genetic differentiation (Becheler et al. 2010). Other explanations have assumed different levels of hypothetic separations between the forms: They have been regarded as different species (mainly historically, but still in the Netherlands and Great Britain, see e.g. http://www.kew.org/, the web site of the Royal Botanical Gardens of Kew) or subspecies. An alternative explanation that has not been discussed before is an 'adaptation-in spite of gene flow'-model. It implies that although the two forms are not genetically separated (no species barrier), genetic changes have taken place at few distinct gene loci in an adaptation process (reviewed in Orr 1998, Hendry et al. 2007). Such processes have been reported in plants by Antonovics and Bradshaw (1970, adaptation to heavy metal) and Goransson et al. (2009, adaptation to soil acidification), and in animals (e.g. Nacci et al. 1999 in fish).

While historically, large subtidal meadows are assumed to have inhabited the Wadden Sea shores; today subtidal eelgrass is lost apart from small remnant populations that occur in tidal creeks. Currently, three different locations of these remnant populations are known on the Halligen Hooge and Langeness and on the island of Sylt (Reusch 2002). Whether these populations are the descendants of the historical subtidal meadows remains to be shown. Obviously, these populations resemble the subtidal eelgrass regarding growth and reproduction as they reach a height of over 80 cm and feature clonal growth with extended rhizomes. At any rate, the populations inhabiting contrasting habitats can be used to study natural selection and adaptation in action.

When aiming at identifying the basis of adaptation and evolutionary change, molecular genetic methods offer suitable tools that complement ongoing ecological studies. In Zostera marina, molecular resources were already quite well developed prior to this study with an existing Expressed sequence tag (EST) database (publicly available at http://drzompo.uni-muenster.de/) and around 15 anonymous microsatellites already developed. Ecological genetic knowledge comprised marker-based population studies in the North Sea area (e.g. Reusch et al. 1999b; Reusch 2002), estimates of clone sizes

(Reusch et al. 1999a), and assessment of mating system, kinship structure and local adaptation (Hämmerli and Reusch 2002; Hämmerli and Reusch 2003b). Interspecific diversity and its relation to resilience of ecosystems have been studied by Reusch et al. (2005). All these studies have shown that microsatellites are suitable markers for eelgrass population genetics studies because of high resolution and variability. Recent development of Single Nucleotide Polymorphisms (SNP) has advanced the toolbox even further. These markers have proven to be valuable tools at large scale population genetics promoting phylogeographic studies of eelgrass populations distributed worldwide (Ferber et al. 2008).

In the light of the globally endangerment of seagrasses and especially envisaging the changes expected to come along with global change in the future, it is necessary to expand knowledge about the ability of species to adapt to changing habitat conditions. The ability of eelgrass to adapt to highly different habitat conditions make it an ideal study organism to assess the processes of natural selection and adaptation.

Detecting adaptive divergence in natural populations

The concept of Genome Scanning

The growing interest in ecological and evolutionary processes in natural populations has caused molecular marker-based approaches to develop at an explosive rate (reviewed e.g. by Luikart et al. 2003; Storz 2005; Stinchcombe and Hoekstra 2008). Selection being the key process of evolution (Darwin 1859) and the prerequisite of adaptational changes is the focus of evolutionary research (Feder and Mitchell-Olds 2003). To date, adaptive divergence in the wild is commonly studied with genetic markers (Vasemägi and Primmer 2005; Bonin 2008). The most obvious reason for this is that classical tools to study adaptation like controlled crossings and parent-offspring analyses are not applicable in most cases of natural populations. Marker-based approaches take advantage of genetic hitchhiking (Smith and Haigh 1974), which means that a (neutral) marker located close to a gene shows the same patterns of response to selection because of physical linkage. Thus, selection effects at genes can be tracked by

monitoring marker loci. When using molecular markers for the study of natural selection, two main principles can be distinguished (reviewed in Vasemägi and Primmer 2005): In top-down approaches, genes known to be involved in a special function, for example by comparison with closely related species, are tracked and surveyed in the target organism. This can be done by experimental manipulation of habitat conditions or – if possible- choice of natural populations that are subject to the respective selection pressure. Top-down approaches require deep knowledge concerning the genetics of the target species and are almost only applied in well-studied model organisms. The opposite is the bottom-up approach which implies testing of many (sometimes randomly chosen) gene or marker loci without a priori knowledge about function or location in the genome (Kauer et al. 2003; Schlötterer 2002b). From results of experimental or natural selection experiments, loci with conspicuous patterns are taken as selection candidates for further investigation. The best-known bottom up approach is termed genome scanning (e.g. Storz 2005), and is implies testing of multiple marker loci (from tens to ten-thousands).

Genome scan approaches, also termed population genomics (Luikart et al. 2003), or multilocus neutrality tests (e.g. Vasemägi and Primmer 2005) have gained increasing interest as a comparably easy-to use method in model as well as non-model organisms. The principle of genome scanning is as follows: In one or several populations or population pairs, multiple loci are tested regarding a variable, for example differentiation between populations (Wright's F_{ST}) or intra-population variability (lnRH). The measured variable is then compared to a distribution expected under neutrality, which can either be constructed by modelling or based on empirical data. Whereas demographic processes like genetic drift and migration affect all loci in the same way, selection affects only a few loci (Cavalli-Sforza 1966). A gene under disruptive selection in two populations spanning an environmental gradient shows more differentiation than selectively neutral loci. Stabilizing selection, on the other hand, is expected to make a locus show less differentiation than expected under neutrality. Loci that fall outside the neutral distribution are termed outliers and regarded as candidates for being influenced by selection (Beaumont and Nichols 1996). Whereas the robustness of an empirical distribution relies heavily on the amount of data points used to construct it, modelled

distributions are dependent on population genetic and demographic assumptions. As in most natural populations, parameters like the effective population size, the mutation rate of the markers, the migration rate or events in population history (like recent or past bottlenecks) are not known, models are error-prone by nature (see e.g. Storz 2005). The usual way to handle these difficulties is to construct models that are as robust as possible, which comes at the cost of being overly conservative. More sensitive models have been criticized as they report more outliers than is judged reasonable. Disturbingly, different models tend to detect different outliers (Vitalis et al. 2001; Vasemägi et al. 2005), which is no surprise given they are based on varying assumptions and based on diverse variables.

Because no model can be clearly prioritized over the others, a combination of models as diverse as possible and with little overlap in the detected outlier loci is a possible solution to make inference by genome scanning more robust. But the strongest enforcement for evidence for selection acting at certain loci is their detection in repeated population comparisons (Storz 2005, Vasemägi and Primmer 2005). This also solves the problem raised by multiple testing, as statistical errors are extremely unlikely to impact the same loci across different comparisons. Thus, any approach to detect selection in natural populations should always aim to use several pairs of populations spanning the same habitat contrast. A habitat contrast is by definition not necessary for a genome scan, and there have been approaches to omit such a setting. The reasons for this include testing genome scanning models in principle, finding out if loci with certain properties such as marker type or developing method are detected more often as outliers or drawing general conclusions about the ration of loci that are detected. However, if results of a genome scan are supposed to serve as a means to gain insights into natural selection processes, a habitat contrast cannot be omitted. As a first step, it renders hypotheses on abiotic or biotic factors that act as possible selective forces and permits the choice of suitable (for example gene-linked) markers.

Another problem of genome scans is multiple testing. As each locus that is screened for selective influences corresponds to one single test, the probability of some loci to lie outside of defined significance criteria rises with a growing number of loci. It is worth noting that many empirical studies report a fraction of candidate loci close to 5% that

fall outside the 95% confidence envelope (e.g. Wilding et al. 2001, 4.9%, Murray and Hare 2006, 1.4%; Campbell and Bernatchez 2004, 3.1%; Mäkinen et al. 2008, 4.9%). The classical way out of this problem that has long been known in ecological research, is to lower the significance criterion alpha in order to keep the desired significance level over all tests in total. While the risk to make an error of the first order (i.e. to erroneously detect loci that are not under selection) is indeed lowered, the risk for committing an error of the second order which implies not being able to detect loci that are subject selection is a high cost (Moran 2003). In the face of tens to thousands of loci that are assessed in a genome scan, it is obvious that application of stricter significance criteria is not an omnipotent tool. Different studies applied various approaches to deal with the multiple testing problem, ranging from ignoring it (e.g. Vigouroux et al. 2002, Schlötterer 2002a) to application of the rather strict Bonferroni correction (e.g. Storz and Nachman 2003). Most studies, however, tried to find their way between these two extremes, either by choosing a moderately lowered alpha (for example to 0.01, e.g. Wilding et al. 2001, Vasemägi et al. 2005) or by using the false discovery rate (Benjamini and Hochberg 1995). But as neither of these ways can provide an optimal solution, the statistical problem should always be kept in mind.

Tests of selective neutrality

The first test for loci under selection was developed by Lewontin and Krakauer (1973), who hypothesised that selection would be traceable in population comparisons when regarding parameters of population differentiation like Wright's F_{ST} at many loci. The main criticism of this test relates to the fact that F_{ST} can be influenced by demographic history and could be only applied with a sufficient number of populations without any sub-structuring (Robertson 1975). The most recent improvement of the test is based on a structured coalescence model and takes the relation of F_{ST} and heterozygosity into account, which is important (Beaumont and Nichols 1996). The model generates null distributions of F_{ST} dependent on heterozygosity based on the weighted mean F_{ST} of all loci. A set of natural mutation rates, drift-migration equilibrium and a symmetric island model of colonisation are used to model the distribution. Between two and as many

populations as present can be tested at the same time. As most of the parameters are roughly estimated, the models' simulations can only be approximate, but are robust against modification regarding population structure and drift-migration ratio (Beaumont and Nichols 1996). Based on similar preconditions, but with a different approach, the method of Vitalis et al. (2001) is restricted to comparison of population pairs. The underlying model is based on a pure drift scenario, where populations have divided at a past time point and have no further contact, i.e. no migrants are exchanged between populations. In this model also, parameters for population differentiation (F1 and F2) at every locus are regarded. Outliers are detected by judging from the joint distribution of F1/F1 estimates dependent on the number of alleles present in the populations. A third way to detect outlier loci does not regard differentiation between populations but intra-population variability of the marker loci. The lnRH test (Kauer et al. 2003) calculates the ratio of heterozygosity in two populations for all loci. As lnRH is approximately normally distributed under neutrality, one can construct a normal distribution based on the data obtained from a larger sample of loci. Loci that lie outside the thresholds of a chosen significance criterion are thus regarded as being subject to selection.

Microsatellites and SNP – comparing markers

The choice of a genetic marker is a central question in genome scans, as properties of loci such as mutation rate, mutational pattern and heterozygosity have direct implications on their ability to be detected in genome scans – regardless if they are influenced by selection or not. Obviously, markers have to show sufficient variability in the regarded populations to make a selection signal visible, which is why microsatellites are among the most used markers for these purposes. Microsatellites consist of small base motives that are repeated many times (for example di- tri- or tetranucleotides). They are distributed over the whole genome, in intronic as well as in extronic regions. The mutational process is a combination of a stepwise mutation model (SMM), leading to gain or loss of one repeat motif, and unusual mutation steps where bigger blocks of more motives are inserted or removed at a time. The second process is usually modelled with an infinite alleles model (IAM), which is based on the assumption that every

mutation leads to a new allele. Uncertainties about the mutation process are a caveat for interpretation of microsatellite data and for the use of some genome scan models, making critical evaluation and exact balance of advantages and disadvantages of one or the other mutational model indispensable. Microsatellites have been the most popular markers for population genetic studies within the last 20 years because they are easy to develop and reliable to genotype keeping some cautionary measures in mind, i.e. ensuring reliable intercalibration once data from different laboratories is merged. Their usually high mutation rate (compared to other markers or genomic regions) favours the formation of new alleles and application in population genetics have shown that they are ideally suited to define genetic differences between populations and individuals of one species. Microsatellites are assumed to be selectively neutral, although in some cases this is questionable, for example if located in promotor regions of a gene (Li et al. 2004). The classical way do develop microsatellites is to use bacterial libraries enriched for repetitive motives, and the resulting markers are termed anonymous or genomic microsatellites. Information about the vicinity of the marker or its location on a chromosome is lacking. An alternative way that has gained more importance recently is to search expressed sequence tag (EST) libraries for these motives. EST databases comprise information from DNA sequence that has been expressed (i.e. transcribed into mRNA) in live cells. After reverse-transcription the resulting cDNA strands can be sequenced, leading to shorter sequences. As relatively easy accessible sequence information, EST databases are constructed for a growing number of nonmodel organisms, for which they often represent the only source of genetic sequence information. Searching these databases for microsatellites is done in silico and often allows development of numerous new markers (reviewed in Bouck and Vision 2007) depending on the number of EST sequences. EST-microsatellites are by nature located close to an expressed gene, either in the 5’ or 3’ untranslated region (UTR) or even inside the open reading frame (ORF), which means within the part of the gene that is translated into a protein. This is why these markers are also called gene-linked microsatellites. The close linkage to a gene makes selective influences on EST microsatellites - either directly or by hitchhiking effect - more probable than in anonymous markers that can be located in any region on the genome (exons not excluded, but less frequent). Thus, the current assumption about EST markers is that

selection should play a more pronounced role in this marker type (Vasemägi et al. 2005). What comes as a disadvantage for population genetic studies on migration and demography requiring neutrality, is offering a promising perspective for genome scans. Additionally, comparison of EST sequences to genetic information from model organisms by BLAST search renders a more or less large fraction of the sequences annotated. One can thus directly connect selection effects visible at marker loci with the nearby gene of a known function. This does not exclude, however, that the candidate marker locus is linked to some other functional region a bit further apart than the EST coding region. Genome scans with gene-linked markers are at the moment one of the easiest ways to draw conclusions about the underlying genetics when natural populations are assessed across a habitat contrast.

Single nucleotide polymorphisms (SNP) represent the simplest genetic polymorphism, the exchange of only one base by another. Thus, every marker locus possesses mostly only two alleles (apart from 1-3% three-allelic SNP) which limits variability – a fact that is counterbalanced by SNP being extremely abundant in the genomes of higher organisms. This marker type is newer than microsatellites but applications are rapidly growing. Especially in silico search has amplified information about SNP loci in model- and to some extent also in non-model species. With a diverse set of methods including classical hands-on PCR and visualization steps, but also modern next generation sequencing technology, SNP are easily screened (e.g. extension-based Infinium assay (www.Illumina.com), real-time-PCR-based Taq-Man (www.appliedbiosystems.com), eventually combined with mass spectrometry (www.sequenom.com), pyrosequencing (Hyman 1988) –based, barcoded (e.g. Craig et al. 2008) or RAD-tag massive parallel SNP-typing (e.g. Hohenlohe et al. 2010)). With a growing number of samples and markers, SNP genotyping has recently become cost-effective also. In contrast to microsatellites, intercalibration between laboratories is usually unproblematic which makes it easy to merge datasets. In summary, SNP clearly represent the markers of choice for large scale genetic surveys now and in the near future, at least in model organisms. For non-models, sufficient genetic information has to be available and as the number of markers is often limited, cost-efficiency is not always granted. For population genetic analyses, it has to be kept in mind that one SNP holds less information than for example a

microsatellite marker providing easily 10-40 alleles even within moderate sample sizes. A rule-of-thumb declares that about 5 times more SNP loci are needed in order to obtain the same amount of population genetic information (Glaubitz et al. 2003).

Heterozygosity-fitness correlations in an ecosystem engineering species

Genetic variation reaches from species differentiation to variation between and within populations and individuals. Relations of genetic diversity at all levels to ecosystem functions, stability and resilience have been the subject of extensive research (reviewed in Loreau et al. 2002 and Hooper et al. 2005; see Tilman et al. 2006; Fontaine et al. 2006, Reusch et al. 2005; Duffy et al. 2001; Hughes and Stachowics 2011). The finest scale of genetic diversity is the intra-individual level of genetic variation, namely the heterozygosity at one locus in an individual. In ecosystems that consist only of one clonally growing species, like seagrass beds, the question arises how heterozygosity influences growth, reproduction and success – i.e. fitness in evolutionary terms. Studies about heterozygosity in clonal organisms are rare and clonality is in most cases not the central research interest (Pujol and McKey 2006, Orrellana et al. 2007, Szovenyi et al. 2009), but heterozygosity-fitness correlations have a long standing tradition in animals and plants (Reed and Frankham 2001; Coltman and Slate 2003; Szulkin et al. 2010 and refs therein). At first, the aim was to study the effect of inbreeding and outbreeding in natural and artificially bred populations. Inbreeding, the mating of two closely related individuals leads to reduced fitness of the offspring (Inbreeding depression, Charlesworth and Charlesworth 1987). One main reason for inbreeding depression is the increased likelihood of combination of recessive deleterious alleles that are otherwise hidden to selection. Fitness reduction is due to heterozygotes having an advantage as the dominant allele is able to conceal a disadvantage of the recessive allele (dominance). As the chance for identical alleles to be paired in a zygote is generally higher than randomly expected, the genome wide homozygosity is increased by inbreeding. Whereas dominance is nowadays regarded as the strongest force to cause inbreeding depression, overdominance describes a phenomenon of heterozygote advantage over

both homozygotes and is assumed to play a minor role (Lynch and Walsh 1998, Charlesworth and Willis 2009).

When non-related individuals mate, offspring is termed outbred and shows elevated heterozygosity. Outbreeding in its extreme form, namely cross-progeny from different groups (i.e populations, strains, subspecies) can lead to heterosis or hybrid vigor, which means the outbred individuals show either elevated (positive heterosis) or – if genetic dissimilarities become too large - reduced (negative heterosis) fitness levels. A negative heterosis effect, caused most likely by dominance, is also called outbreeding depression. One important factor that adds to outbreeding depression is the break-up of favourable allele combinations or disruption of co-adapted gene-environment complexes. In the light of inbreeding and outbreeding depression, the intermediate heterozygosity theory (or optimal outbreeding theory, Bateson 1983) has been formulated, stating that parents with intermediate genetic similarity will have the fittest offspring. Experimental evidence for this relationship is numerous, especially in plant species (Price and Waser 1979, Waser and Price 1989). For animals, parent-offspring analyses have often revealed this pattern (Marshall and Spalton 2000; Marr et al. 2002; Richard et al. 2009). But still, the relationship has not been assessed in clonal plants, and investigation of inbreeding and outbreeding effects in organisms that structure ecosystems are lacking.

Heterozygosity-fitness correlations are based on multilocus data from neutral genetic markers and have been investigated from the 1970s on by evolutionary biologists (e.g. Mitton and Grant 1984, Britten 1996). However, the mechanisms linking multilocus heterozygosity (MLH) to inbreeding are not always clearly differentiated. It is not per se possible to conclude from MLH to genome wide heterozygosity as the assessed markers will only give information about themselves in large and random-mating populations, if not one ore more of the following processes are occurring (reviewed in Szulkin et al. 2010): consanguineous matings, genetic drift or admixture / immigration. All of these processes cause variation in inbreeding and thus identity disequilibrium, a correlation in heterozygosity across loci, which makes marker loci reflect genome wide heterozygosity (i.e. HFC occur). Linkage disequilibrium, a non-random association of alleles of two loci in gametes, can add to these processes but is not the primary cause of them.

Although simulations and meta-analyses show that HFC will usually be weak (reviewed e.g. in Coltman and Slate 2003), many examples of this relationship have been reported (e.g. Coulsen et al. 1998, Coltman et al. 1998). In most cases, correlations are linear and positive, which means that increased heterozygosity is associated with higher fitness. One possible reason for this is that very high levels of heterozygosity have not been tested, so that outbreeding effects, if they exist, can not be seen. Nevertheless, confirmation for the intermediate heterozygosity theory is also rendered by a number of studies assessing MLH with genetic markers (e.g. Marshall and Spalton 2000, Greef et al. 2009). My goal is to provide one empirical example in a seagrass assessing HFC and testing for the intermediate heterozygosity principle by investigating a broad range of heterozygosity levels.

Thesis outline

This thesis is organized in four chapters, each of them written in manuscript form, i.e. containing the sections abstract, introduction, material and methods, results and discussion. Here I give a very short introduction of the main research questions each chapter deals with and formulate why we followed the respective approach.

Chapter I

The first chapter is a technical description of the development of expressed sequence tag (EST) microsatellites for eelgrass Z. marina. As my goal was to conduct a genome scan of contrasting habitats, I aimed to have a first set of gene-linked markers at hand. By the time the study was conducted, the EST library for Z. marina had only recently been constructed by T. B. H. Reusch and was unique for seagrasses - an important prerequisite for assessment of gene-liked markers in genome scans that opened new ways for evolutionary research in Z. marina. The questions I had were if in silico mined microsatellites would turn out as reliable, diverse and easily scorable as anonymous markers and if they could replace them in population genetic studies. For the first time, we were able to find out about the vicinity of microsatellites and if their close linkage to genes would influence them. The markers I developed were the first EST markers for this species, marking a turning point for easy access to novel marker loci. Subsequent establishment of PCR reactions and genotyping protocols were supposed to render ready-to-use markers. Being the starting point for all subsequent analyses, this methodological chapter introduces into how EST markers can be identified, developed and tested for their application in evolutionary or ecological genetic research.

Chapter II

In the second chapter, I present a genome scan, a novel approach in marine angiosperms. The central questions underlying the approach were if natural selection and adaptation could be traced in wild populations by genome scanning and how this approach would add to the methodological scope of ecological genetics. Our study organism, Z. marina, shows two distinct phenotypes related to habitat conditions. We

hypothesized that if genetic adaptation has taken place between these two forms, it would be possible to detect traces of selection. Nevertheless, we were aware that using only 25 markers are at the low end to detect candidate loci, and to produce a reasonable empirical null-distribution under neutrality. We took advantage of three independent habitat contrasts to confirm and evaluate the results of the genome scan. Our three field sites (Hallig Hooge, Hallig Langeness and Sylt island) provided three times identical habitat dependent Z. marina forms, a valuable 'natural experiment'. As we used two different types of microsatellites, we also aimed to assess if EST-derived or anonymous microsatellites would be detected as outliers more often. To find out if temporally fluctuating selection was playing a role in our system, we additionally conducted a temporal genome scan with older material from the sampling sites.

Chapter III

Based on the encouraging results from our first genome scan, we aimed to improve the approach of genome scanning by expanding the number of markers, including a novel marker type (namely SNPs) and adding a third model to detect selection. We concentrated on the three independent habitat comparisons (omitting temporal data) which were re-tested with 40 microsatellites (including 15 newly developed EST-derived markers). Additionally, we used 6 SNP that had proven to be polymorphic in our system as this marker type has rarely been used for population genetics and especially genome scans in non-model organisms. The main question was whether results of the first scan would be consistent, thus adding to information about the reliability of genome scans.

Chapter IV

In the fourth chapter I leave the subject of genetic differentiation between habitats and turn to genetic diversity on a smaller scale - the relation of heterozygosity and clone fitness of Z. marina in Baltic Sea populations. Using 26 EST-microsatellite markers I developed, combined with 11 anonymous microsatellites, we compared levels of multilocus heterozygosity to clone sizes. The first question was whether the intermediate heterozygosity principle applies to Z. marina clones, which is what is discussed below. Heterozygosity is a considerable component of genetic variation in monospecific clonal ecosystems which is probably as important for subsistence as inter-

individual genetic diversity. We aimed to point to this ecologically relevant but rarely assessed relation. The second question was whether EST microsatellites would show a different relation of heterozygosity and fitness compared to anonymous microsatellites. But as results from both marker types were qualitatively identical, we decided to exclude this point from the manuscript. Nevertheless, this issue will be revisited and discussed in the conclusions section.

CHAPTER I

Assessing *Zostera marina* on a tidal flat

CHAPTER I

Identification and characterization of 14 polymorphic EST-derived microsatellites in eelgrass (*Zostera marina*)

Abstract

Zostera marina, the dominant seagrass on the Northern Hemisphere, forms the basis of important but threatened marine ecosystems. Here, we report 14 microsatellite DNA markers derived from an expressed sequence tag library corresponding to a wide range of genes. All loci were moderately to highly polymorphic, with allele numbers ranging from three to eight in a single Wadden Sea population of 48 individuals. Observed heterozygosities ranged from 0.082 to 0.837. Reaction conditions for five pooled polymerase chain reactions are given. The markers will advance the population genetics of seagrasses because they allow indirect tests of selection on closely linked genes.

Seagrasses form one of the world's most productive ecosystems, harbouring a variety of invertebrate and fish species (den Hartog 1970). *Zostera marina* (eelgrass) is the dominant seagrass of the temperate northern hemisphere (Hemminga and Duarte 2000). Unfortunately, serious declines of seagrass beds, including *Z. marina*, have been reported in the past 15 years (Short and Wyllie-Escheverria 1996).

The loss of seagrass populations has led to increasing interest in the ecological genetics of seagrass ecosystems. *Z. marina* is currently being established as a half-model species in order to augment the knowledge about processes and interactions between organisms, genes and environment. Information about the population genetics of *Z. marina* (see Reusch et al. 2000) is currently expanded to genomic information. This includes expressed sequence tag (EST) libraries under varying environmental conditions, part of which contain 1103 genes after assembly and contig formation (TBH Reusch, unpublished data) and were used in the present study. The development of microsatellite markers from EST libraries has proven to be a rich source of novel, gene-linked molecular markers (Li et al. 2004; Vasemägi et al. 2005) that can be used for population genetic applications as well as other approaches, such as genome scans. Here, we report a first set of 14 EST-microsatellites for Z. marina that were developed from this database.

The EST database was searched for simple repeat motifs (microsatellites) using the software CODONCODE ALIGNER 1.2.0 (CodonCode) and the following settings: dinucleotide repeats had to occur at least seven times, tri-or tetranucleotide repeats at least five times. We found microsatellite motives in 22 sequences. We chose 17 of the sequences according to their blast search results that were all meaningful (e-values ranged from 1.00E-8 to 4.00E-75) and designed primers from them using the web program PRIMER3 (Rozen & Skaletzky 2000, http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). For one sequence, no primer pair was designable. In preliminary test runs with individuals from several populations, one primer pair did not amplify a product and one microsatellite locus proved to be monomorphic. The remaining 14 polymorphic microsatellites were tested in a sample of 48 *Z. marina* individuals from a population

located on the tidal flat near the North Frisian island Hallig Hooge (8°31’11’’E, 54°34’06’’N), Wadden Sea, Schleswig-Holstein, Germany. Total genomic DNA was extracted from silica-dried leaves using the Invisorb DNA Plant HTS 96-Kit/C (Invitek) according to the recommended protocols.

After initial testing, we multiplexed the 14 microsatellites in five PCRs (pools A-E, see Table I.1). We used 1µl DNA template, 0.1% bovine serum albumin (BSA), 1µl Invitek buffer (10XNH4-Reaction buffer: 166 mM (NH4)2SO4, 670 mM Tris-HCl, 4.5% TritonX-100) and 0.2 mM dNTPs in a 10µl reaction and ran amplifications on a DNA Engine thermocycler (BioRad). A typical PCR protocol included initial denaturation for 3 minutes at 94°C. One cycle consisted of denaturation at 94°C for 40 s, annealing at 54°C for 40 s, and extension at 72°C for 1 min. Final extension at 72°C for 30 min finished the program. We altered the number of cycles, primer concentrations and $MgCl_2$-concentrations between pools (Table I.1). Pool D was amplified in a hotstart PCR, using 6 µl Hot Start *Taq* MasterMix (Qiagen) and a modified PCR program (Table I.1).

Table I.1 Fourteen microsatellites pooled in five multiplex PCRs and corresponding PCR conditions

Multiplex PCR	Locus name	Primer concentration (µM)	*Taq** polymerase (U)	$MgCl_2$ concentration (mM)	No. of cycles
Pool A	ZMC19066	0.075	0.25	0.75	25
	CL32Contig2	0.125			
	ZMC19017	0.2			
	ZMA04093	0.25			
Pool B (touchdown PCR**)	CL853Contig1	0.15	0.5	1.25	10+16
	CL679Contig1	0.15			
	ZMC06073	0.15			
Pool C	ZMC02023	0.5	0.25	0.75	27
	ZMC19089	0.15			
	ZMC12075	0.15			
	CL734Contig1	0.175			
Pool D (hotstart PCR***)	ZMC13053	0.25	***	***	27
	CL412Contig1	0.25			
Pool E	ZMC01058	0.1	0.25	0.75	26

* Invitek; annealing temperature (T_a): 56°C except ** first 10 cycles of pool B: T_a decreases from 61°C to 56°C in 0.5°C-steps; ***Pool D: used QIAGEN Hot Start *Taq* MasterMix (6 µl per 10µl reaction) with initial denaturation for 10 min at 96°C.

Fragment length scoring was done on an ABI 3100 automated sequencer using an internal lane standard (Rox 350). Allele binning and calling was performed using the software GENESCAN 3.7 and GENOTYPER 3.7 (Applied Biosystems 1998, 2001).

The best GenBank hits of the EST sequences displayed a wide variety of other genes, including aquaporin to defensin genes (Table I.2). Alignment with homologous genes from GenBank revealed that six of the repeats were located in the 5' UTR, eight in the 3' UTR, while none was found in the coding region. All microsatellites tested proved to be medium to highly polymorphic in the tested sample (Table I.2). Allele numbers ranged from three to eight alleles per locus. The high ratio of polymorphic to non-polymorphic microsatellites is surprising as the chosen threshold length of 7 repeat motifs is rather short. We used the program GENEPOP version 3.4 on the web (Raymond and Rousset 1995, http://wbiomed.curtin.edu.au/genepop/index.html) to calculate expected and observed heterozygosities that ranged between 0.08 and 0.802, and 0.082 and 0.837, respectively. Tests for Hardy-Weinberg equilibrium and linkage disequilibrium were performed with GENEPOP as well. Three of the 14 loci (ZMC02023, ZMC12075 and CL412Contig1) showed deviation from Hardy-Weinberg-equilibrium. Observed heterozygosities of locus ZMC02023 and CL412Contig1 were lower than expected; H_o of ZMC12075 was higher than expected (Table I.2). Three out of 91 locus pairs revealed linkage disequilibria assuming a 95% significance level (ZMC19017/ZMC06073, p= 0.0029, CL853Contig1/ZMC12075, P = 0.0143 and ZMC06073/ZMC12075, p=0.0499) but none of the three pairs revealed linkage disequilibrium after Bonferroni correction for multiple testing (P =0.0005, 91 locus pairs). When using the markers for population genetic applications, these results should be regarded with caution unless further testing of other populations reveals that the underlying cause of linkage is not physical but caused by population genetic processes such as inbreeding. The marker loci presented may be a first step towards identifying markers under selection.

Table I.2 Characteristics of 14 EST-derived microsatellites in *Zostera marina* with best GenBank BLAST X hits

Locus name	Repeat motif	Primer sequence (5'-3')	GenBank BLAST-hit species, gene (Accession no.)	Primer location	Allele range size (bp)	No. alleles	H_E	H_O	GenBank Accession no.
ZMC19066	(ACC)7	F: FAM-GTCGCACGCTCTTCTTCC R: ATCCAGCTGTTGCAGTAGGC	*Arabidopsis thaliana* acid phosphatase (NP_194655)	5'UTR	67-82	4	0.140	0.146	AM408830
CL32Contig2	(AGG)9	F: FAM-AATCTGTTGCCACGAAGGAG R: TCACCTTCATCAAGCAGTCG	*Arabidopsis thaliana* unknown protein (AY143804)	3'UTR	89-107	5	0.389	0.347	AM408831
ZMC19017	(AAG)9	F: FAM-TCGTCGAGAAAGAGGAGGAA R: TGTTCTGATTCCGTTCTCCA	*Oryza sativa* bHLH-protein-like protein (NP_910407)	5'UTR	75-99	7	0.722	0.714	AM408843
ZMA04093	(GA)8	F: HEX-CGAACATGAATCTCCGAACC R: GGTAAATGCACCCAGCTCTC	*Nicotiana tabacum* purple acid phosphatase (BAC55157)	3'UTR	118-122	3	0.420	0.362	AM408832
CL853Contig1	(ATGG)5	F: *FAM*-CATTCCATTCAAGAGCAGCA R: CAACAAATCAATCAATCATTCACTC	*Nicotiana tabacum* aquaporin (AAL33585)	3'UTR	124-140	5	0.680	0.592	AM408833
CL679Contig1	(GT)9	F: HEX-ATAAAAACCGGCCTGATCG R: CACACACACAGACGATCGAA	*Castanea sativa* defensin (protease inhibitor) (AAL15885)	3'UTR	86-92	4	0.080	0.082	AM408834
ZMC06073	(TTC)8	F: NED-CGAATCCTCCTGCGTCTTT R: ACGCACCGGATTTTATGCT	*Arabidopsis thaliana* nucleotide sugar epimerase (NP_194773)	5'UTR	86-95	4	0.118	0.082	AM408835
ZMC02023	(TA)11	F: NED-TCGTTGGATACTGCACTTCTC R: GAATCCAACCAATTATTTAAATACC	*Zinnia elegans* lipid transfer protein TED4 (BAA06462)	3'UTR	134-148	6	0.421	0,333*	AM408836
ZMC19089	(GA)12	F: HEX-AGTGAAAAAACAAAGAAAGAAAGAGAAC R: CGTCGTCAGGTAGGCTCAA	*Arabidopsis thaliana* hypothetical protein (CAB81404)	3'UTR	88-96	3	0.503	0.490	AM408837
ZMC12075	(CT)5 (GT)5	F: HEX-CCTCTTTTTTCCTCTCTCTCTCTCT R: CTTCTGCGAATGATGCCATA	*Fritillaria agrestis* cytochrome c (O22642)	5'UTR	100-114	8	0.650	0,653*	AM408838
CL734Contig1	(AAG)10	F: NED-TGGTCTTCTTCACTCGCTCA R: AGCGACGATTCTTCAGCATT	*Oryza sativa* unknown protein (NP_913907)	3'UTR	75-96	8	0.757	0.816	AM408839
ZMC13053	(CT)13	F: FAM--CCCCATCTTTTGAGTTTGGA R: TCATCATTTCTTGCAATTTGAATC	*Arabidopsis thaliana* apaspory associated protein (NP_200543)	3'UTR	86-102	8	0.802	0.837	AM408840
CL412Contig1	(AT)8	F: HEX-GTACACCCGTTGCGTTCATT R: GATTCCGTAGACTTGCGTCTG	*Oryza sativa* putative nodulin 3 (XP_465955)	5'UTR	240-260	5	0.381	0,327*	AM408841
ZMC01058	(GA)8	F: NED-GAGAAGGCAGGAGAGAGAGAGA R: AACTTGTGCTTGCGGCTATT	*Gossypium hirsutum* vacuolar H+-ATPase (AAA82977)	5'UTR	122-132	6	0.592	0.571	AM408842

Expected (He) and observed heterozygosity (Ho) are calculated for 48 individuals from a Wadden sea population; *significant deviation from Hardy-Weinberg equilibrium ($P < 0.05$).
FAM, HEX, NED = fluorescent labels.

Acknowledgements

We thank I. Dankert and S. Carstensen for technical assistance and W. Lampert for continual support. This work was funded by Max-Planck society and the DFG (Re 1108/4).

CHAPTER II

Tidal creek and salt marshes

CHAPTER II

Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*

Abstract

Genome scans are a powerful tool to detect natural selection among populations in a larger sample of marker loci. We used replicated habitat comparisons to search for consistent signals of selection among contrasting populations of the seagrass Zostera marina, a marine flowering plant with important ecological functions. We compared two different habitat types in the North Frisian Wadden Sea, either permanently submerged (subtidal) or subjected to aerial exposure (intertidal). In three independent population pairs, each consisting of one tidal creek and one tidal flat population each, we carried out a genome scan with 14 expressed sequence tag (EST)-derived microsatellites situated in 5'- or 3'- untranslated regions of putative genes, in addition to 11 anonymous genomic microsatellites. By using two approaches for outlier identification, one anonymous and two EST-derived microsatellites showed population differentiation patterns not consistent with neutrality. These microsatellites were detected in several parallel population comparisons, suggesting that they are under diverging selection. One of these loci is linked to a putative nodulin gene, which is responsible for water channelling across cellular membranes, suggesting a functional link of the observed genetic divergence with habitat characteristics.

Introduction

A major goal of the rapidly expanding research field of 'ecological genomics' is to connect phenotypic adaptations with the underlying genetic architecture (Vasemägi & Primmer 2005). In order to achieve this goal, one strategy is to detect selection in a larger sample of marker loci among diverging populations (Luikart et al. 2003). Such multiple-marker-based neutrality tests, dubbed 'genome scans', have proven to be a promising approach for identification of genetic regions influenced by selective forces. Recent genome scans have revealed genes of agronomic importance in crop plants such as sorghum (Casa et al. 2005) and maize (Vigouroux et al. 2002) as well as traces of natural selection in man (Payseur et al. 2002; Kayser et al. 2003; Voight et al. 2006) and mouse (Ihle et al. 2006). Currently, genomic scans are increasingly expanded to nonmodel organisms (Storz 2005, mouse *Peromyscus*; Murray & Hare 2006, oyster *Crassostrea*) and have contributed to our understanding of the interplay of neutral genetic processes and selection by the environment in shaping the genetic structure of natural populations (Nielsen et al. 2006; Larsson et al. 2007).

Genome scans are a 'bottom-up' approach for identifying molecular marker loci that are linked to selectively relevant target loci through 'genetic hitchhiking' (Smith & Haigh 1974). While population processes such as genetic drift or migration are affecting all genetic regions in the same way, selection is only acting upon few gene loci that underlie specific traits (Orr 1998). Consequently, when applied to between-population samples, loci under divergent selection are expected to show higher population differentiation than loci subjected only to genetic drift and gene flow. In contrast, marker loci subjected to stabilizing selection will exhibit lower differentiation. The principle of detection involves the construction of a proper population genetic null model, with subsequent assessments of significant deviations from null expectations across a larger sample of marker loci.

Microsatellites or simple sequence repeats (SSRs) are amongst the most widely used genetic markers for population genetic and evolutionary studies. In this study, we compare two types of microsatellites in a genome scan. Anonymously developed microsatellites of unknown location within the genome are compared to expressed

sequence tag (EST) or gene-linked microsatellites in terms of the population structure they are revealing. Recently, EST databases are becoming a rich source for the development of microsatellite markers that are often situated in the 5'- or 3'-untranslated regions (UTR) of putative genes (Li et al. 2004; Bouck & Vision 2007). The close linkage between gene and marker locus in EST-based microsatellites potentially allows us to relate the results of genome scans to putative gene function in the immediate vicinity of the marker locus. We also intended to compare both marker types to add information on their respective utility to a growing body of literature in the past years (Cho et al. 2000; Vigouroux et al. 2002; Cherdsak et al. 2004; Vasemagi et al. 2005; Woodhead et al. 2005; Pashley et al. 2006; Yatabe et al. 2007).

Neutral genetic models form the basis of tests for detecting outlier loci in genome scans. When a null model is constructed by using empirical data, many marker loci (>100) are necessary (Akey et al. 2002). Therefore, when the number of markers assessed is smaller, a particular demographic scenario has to be assumed in order to deduce neutral expectations. One family of outlier tests is based on the Lewontin-Krakauer-Test (1973). An extension by Beaumont & Nichols (1996) uses coalescent simulations of numerous loci under different scenarios of population demography. Outlier loci are identified by comparing single-locus fixation indices (F_{ST}) with simulations based on the weighted mean F_{ST} over all loci. This method can identify loci that show lower and higher differentiation than expected. Another approach by Vitalis et al. (2001) uses the population differentiation parameters *F1* and *F2* derived from two diverging ancestral populations to construct a null model. This model can only be applied to population pairs and identifies outliers that do not fit into the expected distribution, regardless of whether they show higher or lower differentiation than expected. Although initially designed for complete population separation, simulations revealed that gene flow up to m = 0.1 does not alter the validity of inferences from this approach (Vitalis et al. 2001).

The few studies that have applied more than one analytical approach within the same data set yielded very inconsistent results, revealing different loci under selection depending upon the approach used for outlier detection (Vitalis et al. 2001; Vasemagi et al. 2005; Kane & Rieseberg 2007). This is not surprising, given that tests are based on widely divergent assumptions and demographic scenarios. An alternative approach that

has rarely been used (but see Wilding et al. 2001; Campbell & Bernatchez 2004) is to sample populations from replicated habitat contrasts allowing for replicated statistical tests of selection. If marker loci are detected as outliers in more than one habitat contrast, inferences on the potential role of selection are much stronger (Vasemagi et al. 2005).

In the present study, we compared populations of eelgrass (*Zostera marina*), a marine flowering plant that inhabits widely diverging habitat types along shores of the northern hemisphere, such as different water depths, wave exposures, sediment types and tidal elevations (den Hartog 1970; Jacobs 1982). By connecting a marker-based explorative approach with existing ecological knowledge, the present study aims at providing insights into the genetic basis of adaptation to environmental conditions of ecologically important species. Specifically, we compared several independent pairs of intertidal and subtidal eelgrass populations, to search for consistent signals of divergent selection among 25 marker loci. As we used two types of microsatellites, anonymously developed (genomic) ones and gene-linked ones, a second goal was to compare different marker types for the detection of selection. Specifically we expected to find more outlier loci among gene-linked vs. anonymous markers, because the former are by definition in close physical linkage to expressed genes that may be under selection. As some populations were sampled earlier, we were also able to assess any temporal shift in allele frequency at replicated sites, again comparing gene-linked and anonymous markers. As with the spatial analysis, we hypothesized that gene-linked microsatellites would reveal higher differentiation between years than anonymous microsatellites because, in addition to genetic drift, they are also subject to temporally fluctuating selection in a stressful habitat such as the Wadden Sea (Reise 1985).

Material and Methods

Study species and sampling sites

Our study species, eelgrass (*Zostera marina*) belongs to the seagrasses, a polyphyletic group of monocotyledoneous flowering plants (angiosperms) that returned to the

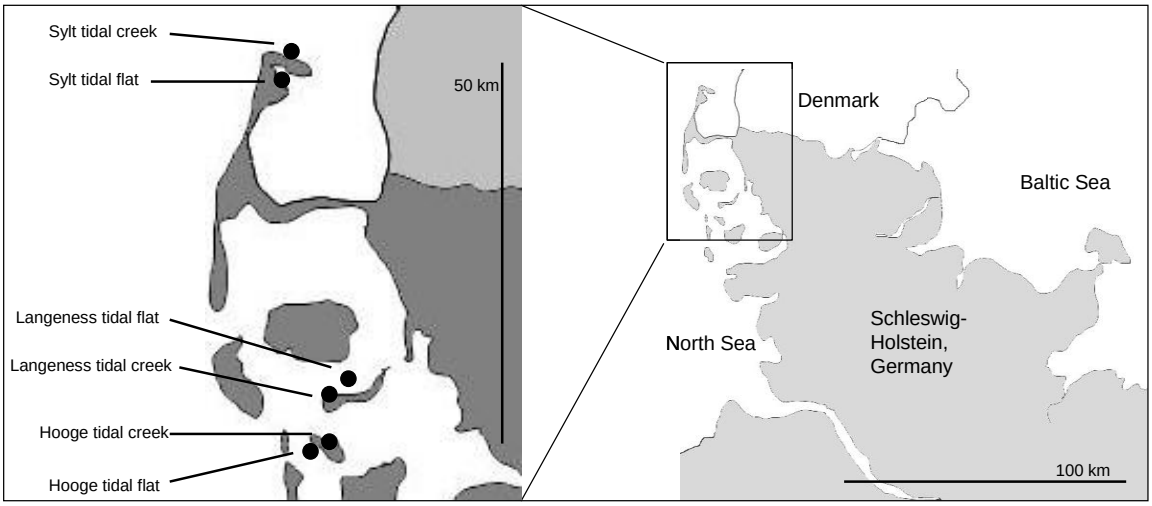
marine habitat three times independently within the group alismatidae (Waycott et al. 2006). In shallow waters, their extensive meadows play an important role as foundation species for productive ecosystems and for sediment stabilisation (Hemminga & Duarte 2000). Seagrass beds provide habitat and nursery for many fish and invertebrate species (den Hartog 1970), and have been rated one of the most valuable ecosystems (Costanza et al. 1997). During the last few decades, serious declines of seagrass meadows were recorded worldwide that are of concern to coastal conservation (Short & Wyllie-Escheverria 1996).

Zostera marina (eelgrass) is the dominant seagrass on the northern hemisphere, distributed in coastal areas of both the Atlantic and Pacific Ocean (den Hartog 1970). Its ecological importance along sedimentary coastlines of the North Sea as habitat-forming species cannot be overestimated. In this region, *Z. marina* has suffered from a combination of eutrophication and changing hydrodynamics leading to a conspicuous reduction of the area of seagrass meadows (Van Katwijk 2000; Reise et al. 1989, 2005). *Z. marina* exhibits a mixture of clonal (vegetative) and sexual reproduction via submarine pollination and seeds. The major dispersal agents are rafting reproductive shoots with ripe seeds (Reusch 2002). Where growing predominantly clonal, clones can attain a very old age (e.g. Reusch et al 1999a), but this is unlikely in the Wadden Sea area (see below).

In our study, we applied a genome scan to contrasting seagrass populations in the North Frisian Wadden Sea (Schleswig-Holstein, Germany). Most *Z. marina* populations are located on tidal flats that are subjected to aerial exposure twice a day during low tide. As second habitat type, we sampled tidal creeks that run between salt marsh areas. Eelgrass populations in this habitat are permanently submerged, possibly representing a less stressful habitat for a fully marine plant lacking any adaptation to desiccation. Also, plants on the tidal flat have to cope with fluctuation in salinity, when, for example, at low tide, heavy rain reduces the salinity level in this habitat (Reise 1985). Light attenuation differs between habitats as well, with the highest photosynthesis rates occurring at tidal flats during low tide.

The contrasting habitat situation causes phenotypic and reproductive differences between populations. The plants on the tidal flat only reach a height up to 25 cm and reproduce sexually by seeds. Rhizomes of one genet never run laterally but form short intermingled clusters with up to 25 vertical leaf shoots. Eelgrass plants from the tidal creeks grow up to 100 cm high, are perennial and reproduce asexually by horizontal rhizomes, in addition to sexual reproduction (Jacobs 1982; Reusch 2002; K. Oetjen, personal observation). Clonal reproduction is also apparent from identical genotypes sampled more than once in tidal creek populations but not on tidal flats (Reise et al 1989; Reusch 2002). Clones are relatively small within the creeks, seldom exceeding 1 m in lateral expansion. Note that we sampled ramets in this study at distances of > 1 m so as to prevent sampling identical ramets of the same genet.

Figure II.1 Eelgrass (*Zostera marina*) sampling sites, located at Sylt, Langeness and Hooge, in the North Frisian Wadden Sea, Schleswig-Holstein, Germany. Note that three population pairs were sampled in 2005, consisting each of one tidal flat habitat location and one tidal creek location that were situated geographically close together.



We sampled 3 population pairs from the Wadden Sea. In each pair, we combined one 'tidal flat' (F) site and one 'tidal creek' (C) site. One pair was located on the island of Sylt, one on Hallig Hooge and one on Hallig Langeness (Fig. II.1). Halligen are low-level islands

that are subjected to total flooding at storm high tides 5-10 times per year. Two sites forming a pair were situated geographically close together with distances between 1.30 and 4.02 km. We collected fresh plant material between 18 July and 2 August 2005 from a 20 x 40 m area by random sampling. Along the creeks that were 50 cm to 5 m wide, plants were sampled at random positions along a transect line at distances of > 1 m. Leaf material was dried and stored on silica gel. Total genomic DNA was extracted from approximately 2-10 mg leaf material using Invisorb DNA Plant HTS 96-Kit/C (Invitek) according to the manufacturer's protocols. Additionally, we used samples dating from previous years for temporal comparisons of the sampled populations. We used DNA from dried plant material sampled in 1999 at Hallig Hooge (tidal flat and tidal creek habitat) that was already analyzed previously (Reusch 2002). From the two other sites, we used extracted DNA collected 1997 (Sylt island) and 2001 (Hallig Langeness). The population 'Sylt island tidal flat' was not sampled in 1997. In total, we investigated five habitat contrasts and 11 populations (Table II.1). The term 'population' is used from now on to differentiate sampled groups that differ in some cases in the place they were sampled at, but in other cases only in sampling year. By comparing two populations, our analyses spanned either habitat contrasts (flat and creek populations from one site) or time contrasts (same locations in different years).

Table II.1 Study design, sampling sites and years of Z. marina sampling in the Wadden Sea. Same symbols indicate population pairs in habitat contrasts (HC) and time contrasts (TC).

Population code	HC	TC	Location (coordinates)	habitat type	year	*N*
H05W	*	*	Hallig Hooge (8°31'11"E, 54°34'06"N)	flat	2005	49
H05P	*	†	Hallig Hooge (8°32'26"E, 54°34'06"N)	creek	2005	49
L05W	†	§	Hallig Langeness (8°35'48"E, 54°38'59"N)	flat	2005	50
L05P	†	‡	Hallig Langeness (8°35'48"E, 54°38'59"N)	creek	2005	49
K05W	§		Sylt island (8°25'55"E, 55°1'37"N)	flat	2005	50
K05P	§	$	Sylt island (8°24'53"E, 55°2'51"N)	creek	2005	39
H99W	‡	*	same as H05F	flat	1999	47
H99P	‡	†	same as H05C	creek	1999	47
L01W	$	§	same as L05F	flat	2001	40
L01P	$	‡	same as L05C	creek	2001	38
K97P		$	same as K05C	creek	1997	27
					Total	485

F, flat, intertidal; C, creek, subtidal; *N* , number of unique genotypes.

Microsatellite analysis

In total, 25 microsatellites of two different marker types were used in this study. We developed a first set of 14 gene linked microsatellites based on a *Z. marina* EST-library containing sequences of 1103 unique genes (for details see Oetjen & Reusch 2007), with putative gene functions based on BLASTX search spanning from defence proteins to aquaporins (Table A-II.2, Appendix). They comprised eight di-, five tri- and one tetra-nucleotide repeat motifs. Additionally, we used 11 anonymous microsatellite loci that were developed previously using an enriched genomic library (Reusch et al. 1999b; Reusch 2000; EMBL nucleotide database, accession no. AJ249303-AJ249307, AJ009898, AJ009900, AJ009901 and AJ009904). These markers have proven to be highly polymorphic and easy to amplify in the study system and are thus suitable for population genetic approaches. PCR protocols followed Oetjen & Reusch (2007) and Reusch (2000). None of the newly developed nor already available markers were excluded from this study.

Microsatellite genotyping was carried out on an ABI 3100 automated sequencer according to the recommended protocols. Size calling and allele binning was done with the software GENESCAN 3.7 and GENOTYPER 3.7 (Applied Biosystems 1998, 2001). We succeeded in genotyping nearly all loci in all 489 samples. Only 4 samples had to be removed from the dataset because of missing locus information. We used 485 individuals for further analysis that had been genotyped for at least 24 out of 25 loci. In only 10 individuals, information at one locus was lacking.

Using microsatellites, the problem of possible null alleles has to be addressed. Null alleles do not amplify in the PCR due to DNA-sequence polymorphism in the primer binding region. Their presence can lead to false homozygote individuals (in case of a heterozygous locus), or to failed PCR amplification (in case of a homozygous locus). We tested for presence of null alleles using the software MICROCHECKER 2.2.3 (Van Oosterhout et al. 2004) at all single loci in all populations, using a 99% significance threshold for detecting homozygote excess to account for multiple testing. Calculations of null-allele frequencies were done following Brookfield's method (Brookfield 1, implemented in MICROCHECKER 2.2.3).

Population comparison based on genetic differentiation at microsatellite loci

We measured population differentiation using pair wise F_{ST}, estimated according to Weir & Cockerham (1984) implemented in the software MICROSATELLITE ANALYZER (MSA, Dieringer & Schlötterer 2003). We tested the significance of F_{ST}-values performing 10,000 permutations of genotypes among populations. F_{ST} was estimated for all microsatellite loci for population pairs and over all populations. Negative F_{ST}-values, if present, were set to zero before further analysis. Isolation-by-distance (IBD) based on geographic distances among sampling sites and F_{ST} values of population pairs estimated from the 2005 data was tested by Mantel's test. We used the shortest linear distance although such a distance measure certainly does not represent the major dispersal route of seeds or uprooted plants. However, we regarded this simplification as less error-prone than more complicated assessments of dispersal pathways under consideration of waterways that are essentially unknown. We also preferred direct distances because it is likely that seeds are transported by waterfowl.

In order to disentangle habitat and geographic distance effects, we also conducted partial Mantel tests. In an extension of the IBD test, we tested the correlation of geographic and genetic distance while controlling for habitat effects. Additionally, habitat effects in genetic distances were tested while controlling for geographic distance. All tests were done using the Web software IBDWS (Bohonak 2002; Jensen et al. 2005).

In order to compare anonymous and gene-linked microsatellites with respect to average population differentiation, we carried out two-way ANOVAs using the independent variables 'marker type' and 'population comparison' and the dependent variable 'pair wise F_{ST}'. We compared F_{ST} values of all loci from the five habitat contrasts sampled. Prior to all ANOVA we tested the assumption of variance homogeneity by Welch ANOVA. All statistic analyses were performed with the software JMP 5.01 (SAS 2002).

Tests of neutrality

We used two neutrality model approaches for outlier detection. The first approach was developed by Beaumont & Nichols (1996) and is implemented in the software package

FDIST 2 (http://www.rubic.reading.ac.uk/~mab/software/fdist2.zip). The model generates coalescent simulations of the distribution of F_{ST} conditioned on heterozygosity. Simulations are generated under a null hypothesis of neutral genetic drift with migration between populations. The model assumes mutation-drift equilibrium, while the mutation rate cannot be modified. Migration is also incorporated in the model without the possibility to change or define specific rates. We only compared population pairs because results become more reliable compared to approaches across several populations (Robertson 1975). To construct a null model, we ran simulations for 100 000 independent loci. These were performed under an infinite allele model because the alternative, a stepwise model of mutations is inconsistent with repeat type and allelic distribution at our loci. In order to explore the most suitable settings, we ran simulations both for 11 demes (the minimum number as we regarded 11 populations) and 100 demes (the maximum number). We decided to use the 100 demes model because the results did not show any differences except for the 100 demes approach being slightly more conservative. The program CPLOT, part of the software package FDIST 2, generated 0.95 quantiles of the expected distribution. Loci outside the null-distribution are outliers under a probability calculated by the software PV (also distributed with FDIST 2). Outliers with F_{ST} values of < 0 were excluded from further analyses because these F_{ST}-values are not defined but rather artefacts of the calculation procedure for F_{ST}.

The second approach was developed by Vitalis et al. (2001) and is implemented in the software DETSEL 1.0 (Vitalis et al. 2003). The model is based on the assumption that one population splits into two subpopulations that have no further contact and that genetic divergence is resulting from pure drift. Although this assumption is presumably violated in our situation, the model can be applied nevertheless as the authors have shown that it is robust to moderate migration (Vitalis et al. 2001). Parameter values on mutation rates and population size need to be a priori chosen, in order to simulate different scenarios of population demography. The software allows for parameter value sets to be chosen, including e.g. different mutation rates and population sizes in one simulation, thus mixing different scenarios to generate a broader and more conservative distribution. Based on mutation rates of *Z. marina* estimated for seven anonymous microsatellites from pedigree data that were in the range of 5×10^{-3} (TBH Reusch,

unpublished), we applied three mutation rates (μ = 0.001, 0.005 and 0.01) in an infinite-allele model. Population size before split (N_0) was set to 500 while we simulated ancestral population sizes (N_e) of 500, 1000 and 10,000 individuals. 100 000 coalescent simulations provided the expected distribution based on the parameters of population differentiation *F1* and *F2*. Other parameters were modulated in order to obtain an allele distribution similar to the real data, a procedure that is recommended by the authors. We set T0 ('time since an assumed bottleneck') to 50, 100 and 200 generations and t ('time since the population split') to 50 generations. To identify outlier loci, we compared the expected distribution of *F1* and *F2* conditioned on the number of alleles (k= 3, 4, 5 and ≥ 6) in the sample with the single locus *F1*/*F2* values. All loci outside the 99% confidence interval were defined as outliers, corresponding to a significance criterion of α = 0.01. Preliminary test runs had shown that the number of outliers under α = 0.05 was exceedingly large. This would mask loci with real deviations from neutral expectations and produce a high number of false positives, which is why we applied a stricter criterion in this test. The Beaumont model showed a very different result with only few loci being detected with the standard α = 0.05, which was therefore kept as is. Adjusting the criterion in single tests according to extremely high or low outlier rates instead of for example a Bonferroni correction is a procedure that has been used in explorative genome scans before (Vasemagi et al. 2005; Bonin et al. 2006) and is applied here. In both models, we simulated the expected distribution for pairwise comparisons of all five population pairs across habitat contrasts. For the Vitalis model, we also tested all four possible population pairs from the same habitat, in order to provide a test with a predicted negative outcome, i.e. no signs of selection.

For several reasons, we restricted our search to traces of diverging selection. First of all, we assumed that in comparisons of different habitat types, there should be at least some loci that respond to different selection pressures and show patterns of divergence between habitats. Consequently, we applied two methods that rather monitor exceptionally high differentiation between populations (Bonin et al. 2006). The model by Beaumont and Nichols seems to have difficulties in detecting stabilizing selection (Beaumont & Balding 2004) while it performs better under adaptive divergence of populations. Moreover, in our case, the mean divergence (F_{ST}) of populations was so

small that the lower limit of the simulated confidence interval (i.e. the 0.025 quantile) was sometimes located below the abscissa (table S1, Supplementary material). Note that the method by Vitalis et al. (2003) does allow the detection of stabilizing selection at all.

Nested Analysis of Molecular variance (AMOVA)

We used a nested analysis of molecular variance (AMOVA) to quantify genetic differentiation (as F_{ST}) due to the habitat contrast as an additional test to complement the outlier detection procedures. To do so, we partitioned genetic variation among sampling sites and among population pairs nested within sites. As an important advantage, nested AMOVA can summarize the differentiation across several population pairs that are putatively under divergent selection. Populations were grouped according to sampling site, with each group containing one population from the creek habitat and one from the tidal flat habitat type. Populations sampled at the same location in different years were assigned to different groups as high temporal variation occurred in the studied habitat. Nested AMOVAS were carried out with the software ARLEQUIN 3.11 (Excoffier et al. 2005). Then, levels of genetic variation at loci that had been identified as outliers previously were compared to those of non-outliers.

Results

Genetic population differentiation at microsatellite loci

All identified multilocus genotypes were unique, thus no multiply sampled clones had to be excluded prior to the analysis of genetic differentiation. The overall genetic differentiation among the six investigated populations in 2005 was low but highly significant (multi-locus F_{ST} = 0.017, P = 0.0001). F_{ST}-values for all possible population pairs ranged from 0.0 to 0.042 and were significant in most cases (Table A-II.1, Appendix). Negative F_{ST}-values were estimated for only three population pairs (H05C vs. H99C, H99F vs. L05F, and L05F vs. L01F) and were subsequently set to zero. Estimates of genetic differentiation between different years at the same site were also significant, attaining maximal multilocus F_{ST} estimates of 0.021 (L01C vs. L05C, P = 0.001). There was

no difference between multilocus F_{ST}-values estimated for pairs at the same site in different years and habitat contrast pairs (t-test, $\overline{X}_{\text{same site pairs}}$ = 0.0098, n=5, $\overline{X}_{\text{contrast pairs}}$ = 0.0129, n=5, *P* =0.025).

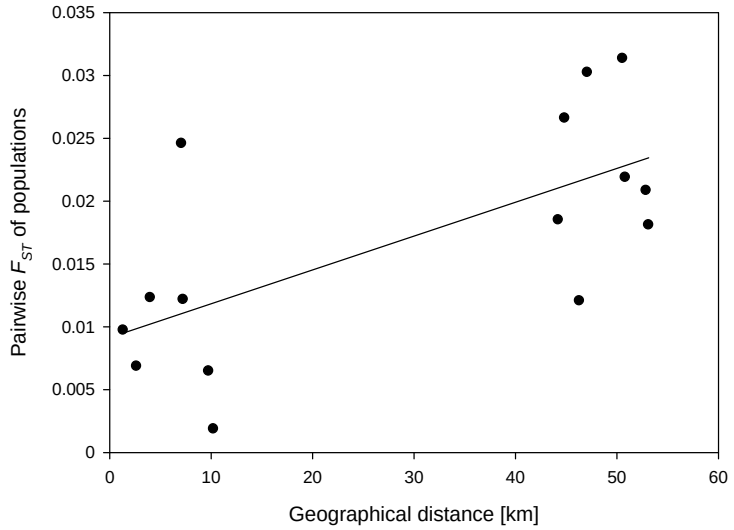


Figure II.2 Relationship of geographical distance (in km) and genetic distance (as standardized allelic variance F_{ST} estimated according to Weir and Cockerham 1984) of populations sampled at six sites in 2005. Isolation-by-distance is expressed by linear regression (F_{ST} = 0.0091493 + 0.000269 geographic distance, r2 = 0.44, P =0.0067). Mantel's test revealed a significant correlation of both distance matrices (P_{Mantel} = 0.046).

Although the overall genetic differentiation was low, we found a significant IBD pattern (Fig. II.2), indicating that increasing geographic distance between populations leads to greater genetic divergence (Mantel test: Z = 9.1798, r = 0.6657, P_{Mantel} = 0.0463 from 10 000 randomizations). Note that only the most recent sampling date of all sampling sites was included in this analysis. IBD patterns were driven by three populations from Sylt having the largest spatial distance and the highest genetic differentiation to all other populations (K05F, K05C and K97C). The partial Mantel test revealed that the correlation of genetic and geographic distance was only marginally significant when controlling for habitat effects (r = 0.643, P_{Mantel} = 0.054, 10 000 randomizations). No correlation of habitat type and genetic distance could be identified. The partial correlation of genetic

distance and habitat type controlling for geographic distance was not significant, either (P = 0.6). Note that these tests possess only low statistical power to detect habitat effects from our original data. Interestingly, the multilocus F_{ST}-values of comparisons of populations with different habitat types did not exceed the multi-locus F_{ST}-values of comparisons of the same habitat type.

Single-locus statistics revealed striking differences between loci with expected heterozygosities ranging from 0.10 to 0.9 (mean: 0.53), observed heterozygosities ranging from 0.02 to 0.96 (mean: 0.499) and allele numbers ranging from four to 31 (mean: 10.2, Table A-II.2, Appendix). A complete Table of all allelic data can be obtained from the authors upon request. The presence of null alleles was suggested for some loci, but only at two loci (GA35, ZMA04093) were there null-alleles detected in more than one of the 11 populations. Moreover, among the loci with putative null alleles, only locus GA35 was also identified by the outlier tests. This locus showed presence of null alleles in nine out of 11 populations with calculated null allele frequencies ranging from 0.09 to 0.16 (estimator: Brookfield 1). Global F_{ST} estimates ranged from negative values (locus ZMC06073: -0.00005) to a maximum of 0.042 (locus ZMC19089). Although the applied outlier tests account either for different heterozygosities of loci (Beaumont's approach) or for the number of alleles, respectively (Vitalis' approach), these differences might affect the results. The comparison of the two marker types revealed only weak differences based on population comparisons of contrasting habitats. There was no significant difference between F_{ST}-values of genomic and EST-microsatellites when taking into account five habitat contrasts from different years (Two-way-ANOVA, factors: habitat contrast, marker type, F_{ST}; P =0.84). There was also no significant interaction between the factors 'marker type' and 'habitat contrast'.

Additionally, we compared the differentiation of F_{ST}-values of both microsatellite types between two different years at one sampling site as a measure for temporal fluctuation in allele frequencies. This analysis, taking five population comparisons into account, revealed that F_{ST}-values of gene-linked microsatellites exceeded the F_{ST}-values of anonymous microsatellites (2-Way-ANOVA, $\overline{x}_{\text{GENE-LINK. MSAT}}$ = 0.015, n=70, SE=0.0035, $\overline{x}_{\text{ANONYM. MSAT}}$ = 0.008, n=55, SE=0.002, P =0.007). This supports our initial hypothesis, namely that gene-linked microsatellites are more strongly subjected to selection under

temporally fluctuating habitat conditions than anonymous ones, albeit only for average F_{ST} values.

Neutrality tests for detecting divergent selection

The approach by Beaumont & Nichols (1996) identified few outlier loci with higher differentiation than predicted under neutrality that would demonstrate divergent selection. In pairwise comparisons of different habitat types, no consistent patterns were apparent. One locus, ZMC19089, linked to a hypothetical A*rabidopsis thaliana* protein, showed higher differentiation (P = 0.01, Fig. II.3) in one comparison (populations L05F vs. L05C) while another anonymous genomic microsatellite (GA35) appeared to be an outlier in the comparison L01F vs. L01C (P = 0.02, Figure II.3). Note that the first locus, ZMC19089, reveals the highest F_{ST} value (0.042) of all loci in our study while having a medium heterozygosity (0.53) and a medium number of alleles (7). The other locus, GA35, has the highest heterozygosity (0.90) and the highest number of alleles (31). There were also some loci that were just within the confidence interval of the null-distribution, which is not a strong but a possible indication for selection (ZMA04093, purple acid phosphatase, H05F/H05C; CL853Contig1, aquaporin, L01F/L01C).

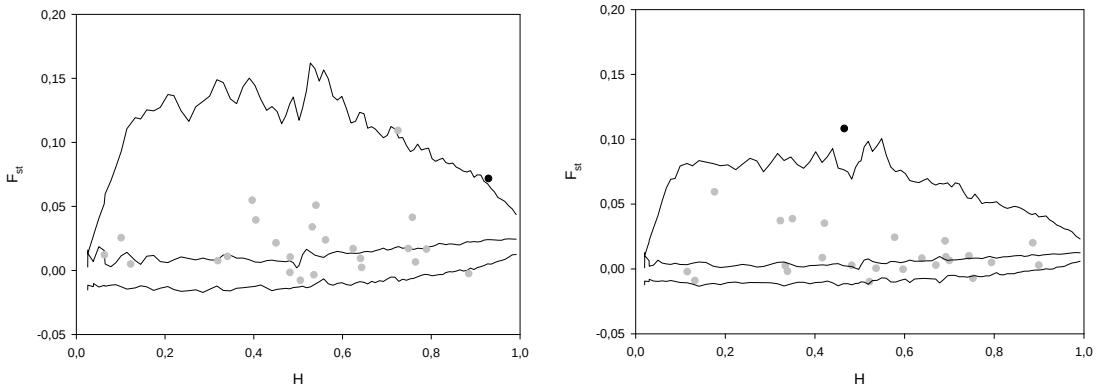


Figure II.3 Analysis of divergent selection at two loci in two different habitat contrasts (left: Langeness 2001; outlier: GA35, P = 0.02, right: Langeness 2005; outlier: ZMC19089, P = 0.01) detected with the model by Beaumont and Nichols (1996). Solid lines indicate 0.05, 0.5 and 0.95 quantiles of the expected distribution of F_{ST} conditioned on heterozygosity (H). Outlier loci are indicated by black dots.

In contrast, the analysis model by Vitalis et al. (2001) revealed many more outliers although we applied a stricter significance criterion of $\alpha = 0.01$ to all of the tests carried out with this model. Thirty-four outliers were identified in five habitat contrasts in total, on average 6.8 outlier loci per comparison (Table 2). We decided to use all five habitat comparisons although two of them were sampled at the same sites. Our rationale was that populations differed almost as much among different years than due to spatial distance. The results of DetSel confirm this view, because we did not find the same outliers in comparisons of populations at the same site in different years (Table II.2). Most of the 34 loci were identified as outliers more than once. Three Loci were detected in four out of five habitat comparisons: two EST- derived microsatellites (ZMC19066, linked to an acid phosphatase and CL412Contig1, linked to a nodulin gene) and one genomic microsatellite (GA20). Interestingly, these loci do not differ significantly from all other loci in heterozygosity or in the number of alleles (t-test, *P* =0.09 and 0.092) although they seem to have, on average, lower heterozygosities (0.10, 0.32 and 0.45) and fewer alleles (8, 7 and 9, respectively). Three additional loci were detected in three comparisons, the EST- microsatellites CL679Contig1 (linked to a defensin gene) and ZMC01058 (linked to a gene for vacuolar H+-ATPase) and the genomic microsatellite GA16. Using a re-sampling scheme, we assessed whether or not the occurrence of three out of 25 loci in four out of five comparisons can be explained by chance. A re-sampling with 20 000 random permutations revealed that this is unlikely (P =0.0065). Accordingly, we conclude that three loci (ZMC19066, CL412Contig1 and GA20) show signs of divergent selection that is consistent across habitats. As a further test, the same analysis was carried out in all four possible population pairs that belong to the same habitat type in order to provide a test under the alternative hypothesis. In all population pairs (H05F/L01F, H05C/L01C, H99F/K05F and H99C/L05C) without habitat contrast the detected outliers were not significantly different from random sampling (20 000 drawings, *P* =0.12). Over all, EST-microsatellites were not identified more often as outliers than genomic microsatellites (Chi-square-test, *P* = 0.31).

Table II.2 Divergent selection in eelgrass *Zostera marina*. Outlier loci detected in five pairwise comparisons of contrasting habitats with the method by Vitalis et al. (2003), unter application of a 99% confidence interval (i.e. significance criterion of α = 0.01)

Population comparison	H05 F / C	H99 F / C	L01 F / C	L05 F / C	K05 F / C
Identified outliers	ZMC19066	CL679Contig1	ZMC19066	ZMC19066	ZMC19066
	ZMA04093	ZMC06073	CL32Contig2	ZMC02023	CL32Contig2
	CL679Contig1	CL412Contig1	CL835Contrig1	ZMC19089	ZMC06073
	CL412Contig1	GA20	CL679Contig1	ZMC01058	CL412Contig1
	ZMC01058	GA16	CL412Contig1	GA12	ZMC01058
	GA20		GA12	GA20	GA12
				GA16	GA20
					GA17D
					GA16
					GA1
Number of outliers	6	5	6	7	10

Nested AMOVAS revealed that variation between populations of different habitat types accounted for 1.26 % of the total genetic variation while variation between sampling sites (groups in the AMOVA terminology) was only 0.68%. Thus, habitat type had twice the effect on genetic variation than geographic or temporal distances. When only the five loci formerly identified as outliers were included in the analysis, the percentage of variation due to habitat contrasts increased to 2.35%, confirming the influence of habitat differences. The remaining 20 loci accounted for a much smaller fraction of variance due to habitat differences, namely 1.03%. In an additional analysis, the percentage of variation due to habitat contrasts of every locus was estimated. The five loci identified as outliers possessed significantly higher values than the remaining loci (t-test, unequal variances, P =0.002). Qualitatively similar results were yielded by an analysis that included only the three strongest outliers, i.e. those identified in several habitat contrasts in Vitalis' method (data not shown).

Discussion

The first EST- microsatellite based genome scan carried out in a marine plant yielded encouraging results. Although the genetic differentiation among all populations was low, we found signals of divergent selection, at least for some population contrasts, in five microsatellite loci. In our study system, gene flow seems relatively high, as some of the populations could not be separated even based on 25 microsatellites used here, representing 255 alleles in total. Temporal variation between different years at one site was as high as variation between different sites, which confirmed our picture of the Wadden Sea as a highly variable ecosystem with rapidly changing conditions and spatially and temporarily fluctuating metapopulations of seagrasses. Our results are consistent with earlier research detecting high gene flow among Z. marina populations in the North Sea (Reusch 2002). Nevertheless, over the whole geographic range of our study area, a significant IBD pattern indicated genetic structuring that can be attributed to limited dispersal in this plant (Reusch 2002). Close genetic relationships between the sampled populations are recommended for outlier identification tests if the research hypothesis is divergent selection (Beaumont & Nichols 1996; Schlötterer 2002; Vitalis et al. 2001). Theory also predicts that while the major part of the genome is subject to admixture by gene flow, selection can maintain existing differences at single loci (Orr 1998). In the present study, the identification of similar outlier loci in repeated habitat contrasts, and a significant contribution of those outliers to a habitat effect among population pairs in AMOVA, mutually support each other. The latter finding is in line with other studies on average (but not locus-specific) genetic differentiation among contrasting habitats, for example in salt-marsh plants, fish and marine snails (Bockelmann et al. 2003; Reusch et al. 2001; Wilding et al. 2001).

Both analysis methods used here revealed high discrepancies concerning the number and identity of outlier loci, similar to other recent studies (Vasemagi et al. 2005; Vitalis et al. 2001). As expected, the number and affiliation of loci under selection are highly dependent upon the chosen method. The approach by Beaumont & Nichols (1996) detected few outliers and none of them appeared in more than one population comparison. This notwithstanding, three of the four loci that were significant or

marginally significant in Beaumont's test (ZMC19089, ZMA04093 and CL853Contig1) were also detected as outliers with the DETSEL test in one of the habitat contrasts. The method by Vitalis et al. (2001) revealed so many outliers on a pairwise basis that this questions the validity of the results, supporting earlier findings (Vasemagi et al. 2005). The underlying population-genetics model originally assumed complete separation of populations, but later simulations revealed that gene flow (i.e. the fraction of migrants) can be substantial (m = 0.1; nevertheless, inferences are still mostly correct. In our data set, such a high level of m would translate into equilibrium values of $F_{ST} < 0.01$ (assuming N_e to be in the range of 500-5000), thus substantially smaller than the average F_{ST} = 0.017 found in this study. An additional problem when interpreting genome scans is the issue of false positives due to multiple tests. By definition, a fraction of the loci examined will be outside the expected null distribution, revealing spurious deviation from neutrality. To separate these false positives from real outliers remains a critical point in genome scanning (Thornton & Jensen 2007). Adjusting the significance criterion by considering the number of tests performed, i.e. the number of loci involved, is an option but comes at the cost of increasing risk of committing an type-II error (Moran 2003). An alternative, in order to minimize detection of false positives, is to take only those loci into account that were repeatedly detected by several methods, or across several population comparisons. In this study, both candidates identified by Beaumont's FDIST2 approach occurred only in one of the habitat comparisons each. As both loci are also extreme in terms of their variability parameters compared to all other loci, such as extremely high heterozygosity, null allele presence and high allele number (GA35), or high population differentiation (ZMC19089), they cannot be regarded as candidates from our point of view. According to Vitalis' model, there are at least three loci that can be taken as strong candidates, namely the three loci that were detected in four out of five selection tests. Their variability estimators did not differ from those of all other loci. Two of them are gene-linked microsatellites. The locus ZMC19066 is linked to an acid phosphatase gene, coding for proteins with numerous functions in cell metabolism and signal transduction (Sheng 2003). The most interesting locus, CL412Contig1, is located in the 5' untranslated region (UTR) of a nodulin gene. These ubiquitous proteins belong to the aquaporin group and are involved in water channels across cell membranes, which make them important proteins in water regulation. Aquaporins appear in different

densities dependent on stress factors such as drought and high salinity (Luu & Maurel 2005). A relationship between the contrasting desiccation stress of the two habitats and some structural or expression differences in the associated nodulin gene appearance is a reasonable interpretation of the genetic divergence observed. The identified candidates can be used to formulate a priori hypotheses to be retested in further genomic scans of extended population samples, or in physiological experiments. Within our sample of 25 microsatellites, the identification of three candidate loci (12%) is a realistic result as similar proportions are reported in other studies. Whereas 12-21% of gene linked markers were detected in genomic scans in oak species and Atlantic salmon (Scotti-Saintagne et al. 2004; Vasemagi et al. 2005), 0-9% of anonymous loci showed signs of selection in whitefish, oak, salmon and snails (Wilding et al. 2001; Campbell & Bernatchez 2004; Scotti-Saintagne et al. 2004; Vasemagi et al. 2005). Whether or not both microsatellite types (gene-linked or anonymous) can equally be used for population genetics studies requiring neutrality is unresolved at present. It has been suggested that gene linked microsatellites are more often positive outliers in genome scans due to their close linkage to genes (Vasemagi et al. 2005) but presently there is no consensus to support or reject this notion (Cho et al. 2000; Vigouroux et al. 2002, Cherdsak et al. 2004; Vasemägi et al. 2005; Woodhead et al. 2005; Pashley et al. 2006; Yatabe et al. 2007). If gene-linked markers were indeed under selection more often than anonymous one, the outcome of studies requiring strict neutrality of markers would be biased. In our study, gene-linked microsatellites are not identified more often as outliers in genome scans than genomic ones, but they contributed a major fraction of the between-habitat genetic divergence in an AMOVA. An interesting result was that gene-linked microsatellites showed higher temporal differentiation than anonymous ones at the same site. We attribute this to rapid, year to-year fluctuation in the physical environment (Reise 1985). In addition, recent environmental change has taken place in the Wadden Sea habitat, as in other coastal areas, associated with global warming (Harley et al. 2006; IPCC 2001; Perry et al. 2005). These environmental changes might act as fluctuating selective forces with a stronger influence on gene linked markers because on average, they are more often linked to functional genes than anonymous ones.

EST-library derived microsatellite loci proved to be well suitable for genomic scanning in our study. The markers are appropriate tools for selection-detection studies because of their easy and efficient development. Microsatellites are assumed to be generally noncoding if not proven otherwise, but loci in close relationship to genes, in particular, are suspected of having regulatory functions, for example as promoter binding sites (Li et al. 2004). If true, selective neutrality of EST microsatellites would become questionable. We suggest avoiding EST microsatellites for population genetic purposes unless they are tested for neutrality in genome scans.

While genome scans have started with model organisms (man, Akey et al. 2002; fruit fly, Kauer et al. 2003; maize, Vigouroux et al. 2002), such approaches are increasingly applied to other non-models. Two studies that have recently conducted genome scans to compare habitats with different ecological conditions showed that specific genes correspond to different selection regimes that are in turn correlated with altitude (Bonin et al. 2006; Jump et al. 2006). We suggest that species with important ecological roles, in particular ecosystem engineering species such as *Z. marina* lend themselves to such approaches when distributed across a range of habitat types with different environmental regimes. Insights into the genetic basis of adaptation to environmental conditions are also useful for conservation and management by providing essential knowledge on local adaptation of possible donor populations when transplanting genotypes with specific traits is required, and such initiatives are planned or even ongoing among the seagrasses, included Z. marina (Van Katwijk 2000; Reise et al. 2005). From a point of view of study design in the field, subtidal– intertidal gradients represent one of the clearest habitat contrasts that are replicated over latitudinal scales, allowing for large-scale tests of habitat adaptation and its genetic basis. Earlier studies that have made use of intertidal gradients investigating snails (Wilding et al. 2001), salt-marsh plants (Bockelmann et al. 2003) and barnacles (Schmidt & Rand 2001) have shown the excellent suitability of marine coastal areas for searching for selectively maintained genetic polymorphism. We recommend expanding such approaches to even larger geographic scales while using several replicated habitat contrasts.

Acknowledgements

We thank I. Dankert and S. Carstensen for technical assistance and W. Lampert for continual support. We are grateful to Prof. Karsten Reise from AWI, Littoralstation Sylt for advice and facilitating access to field sites, as well as to Andreas Zipperle for field assistance. This work was funded by Max-Planck society and the DFG (Re 1108/4; and Re 1108/7-1).

CHAPTER III

Flowering *Z. marina* shoots in a tidal creek

CHAPTER III

New evidence for habitat-specific selection in Wadden Sea *Zostera marina* populations revealed by genome scanning using SNP and microsatellite markers

Abstract

Eelgrass Zostera marina is an ecosystem-engineering species of outstanding importance for coastal soft sediment habitats that lives in widely diverging habitats. Our first goal was to detect divergent selection and habitat adaptation at the molecular genetic level; hence, we compared three pairs of permanently submerged versus intertidal populations using genome scans, a genetic marker-based approach. We extend here a previous genome scan study by doubling the number of gene-linked microsatellites and adding SNP-markers, a novel marker for seagrasses. Three different statistical approaches for outlier identification revealed divergent selection at 6 loci among 46 markers (6 SNPs, 29 EST microsatellites and 11 anonymous microsatellites). These outlier loci were repeatedly detected in parallel habitat comparisons, suggesting the influence of habitat specific selection. A second goal was to test the consistency of the general genome scan approach by doubling the number of gene-linked microsatellites and adding single nucleotide polymorphisms (SNP) loci, a novel marker type for seagrasses, compared to a previous study. Reassuringly, results with respect to selection were consistent among most marker loci. Functionally interesting loci were linked to genes involved in osmoregulation and water balance, suggesting different osmotic stress, and reproductive processes (seed maturation), pointing to different life history strategies. The identified outlier loci are valuable candidates for further investigation into the genetic basis of natural selection.

Introduction

Seagrasses are a polyphyletic group of angiosperms that are ecosystem-engineering species by providing three-dimensional structure on soft sediment (Duarte 2002; Orth et al. 2006). Ecosystem services of seagrass beds include the provision of nursery habitat and food source for numerous fish and invertebrate species, the stabilization of sediments and carbon and nutrient storage (Hemminga and Duarte 2000). Because these ecosystems are distributed worldwide and range among the most valuable ecosystems (Costanza et al. 1997)), ongoing population declines during the last decades are of serious concern (Waycott et al. 2009). The responsible anthropogenic changes in coastal areas include changes in current regime (leading to increased suspension and turbidity and decrease of water transparency), eutrophication and mechanical disturbance (Hemminga and Duarte 2000).

Zostera marina L. (Eelgrass) is the dominant seagrass on the northern hemisphere and is widely distributed across European shallow waters. In the Wadden Sea area, eelgrass is subtidally almost extinct after the wasting disease in the 1930s (den Hartog 1970). Recovery never occurred, possibly due to ongoing pollution of coastal waters and changes in water movement and climate. However, some fragments in permanently submerged creeks remained, situated on the low lying islands of the northern Wadden Sea (the 'Halligen', Reusch 2002). In contrast, there are still substantial populations in intertidal areas (Reise et al. 2005). Due to its ecological importance, Wadden Sea populations of eelgrass have been the focus of recent genetic research provided valuable insights on genetic diversity, colonisation history and genetic exchange of populations (Reusch 2002; Olsen et al. 2004). The increasing genetic information available for *Z. marina*, including an EST (Expressed Sequence Tag) database available at http://drzompo.uni-muenster.de/ (Reusch et al. 2008; Wissler et al. 2009) now allows investigating the genetic basis of physiological adaptation to extreme natural environments such as tidal flats (Reise 1985).

Wadden Sea eelgrass populations offer an outstanding opportunity to gain insights into habitat-dependent selection using two habitat-specific phenotypes that are distinguished by their life cycle. The subtidal form lives permanently submerged in

creeks and reproduces asexually by forming perennial clones via rhizome expansion, in addition to sexual reproduction via seeds. An intertidal form growing on the extended tidal flats of the Wadden Sea area is subject to much higher yearly and daily fluctuation in temperature and salinity, because only a thin film of water remains when plants fall dry (Massa et al. 2009). Moreover, leaves are exposed to increased levels of UV light during low tide. Reproduction only takes place sexually via flowers and seeds, while asexual reproduction is rare (Reusch 2002).

We thus hypothesized that different habitat conditions have resulted in local adaptation visible not only at the phenotypic level, apparent as different life history strategies, but also at the molecular genetic level. To test this, genome scans recently became one of the most promising molecular genetic approaches. Genome scans extend traditional population genetics approaches to large marker numbers coupled with statistical tests in order to identify genetic loci influenced by selection (Schlötterer 2002b; reviewed by Luikart et al. 2003; Stinchcombe and Hoekstra 2008). The principle of detection is simple. If molecular markers are physically linked to functionally important and polymorphic genes, divergent selection acting on such genes can be detected via genetic hitchhiking (Smith and Haigh 1974; Lewontin and Krakauer 1973; Harr et al. 2003; Schlötterer 2003; Storz 2005). Whereas random or demographic effects affect all loci in the same way, effects of divergent selection lead to increase differentiation between populations at certain target loci (Nielsen 2005), quantified as Wright's F_{ST}. A critical step is to test the statistical significance of the numerical value to be outside neutral expectation. Here, coalescent simulation-based methods have made great progress that compare population differentiation at single loci with a model distribution and allow detection of patterns not consistent with neutral evolution (Beaumont and Nichols 1996; Vitalis et al. 2001). Another approach uses empirical data to construct the null distribution and is able to detect a reduction of genetic diversity at certain loci, which is caused by a 'selective sweep' (Schlötterer 2002a).

Here, we simultaneously use three model approaches to examine the mutual consistency of the different methods. Even more importantly, we assessed three repeated habitat contrasts in our system. If one locus is detected erroneously in one habitat contrast, it should not appear as an outlier in the other two contrasts (Wilding et

al. 2001; Luikart et al. 2003; Campbell and Bernatchez 2004; Storz 2005; Vasemagi et al. 2005; Bonin et al. 2006).

For our genome scan, we employed three classes of genetic markers: anonymous microsatellites, gene-linked microsatellites derived from an EST (expressed sequence tag) library, and SNPs (single nucleotide polymorphisms). Microsatellites are among the most used molecular markers for population genetic assays and combine high information content per marker locus with relatively easy and cost-effective development and screening. Anonymous markers are derived from standard enrichment libraries for microsatellite motifs and do not provide any information with respect to the vicinity of the marker. In contrast, gene-linked microsatellites are located close to a gene (often in the untranslated regions) and are typically derived from an EST library (Bouck and Vision 2007) In genome scans, gene linked markers represent a very useful means to assign functional genes possibly responsible for the effect seen at an outlier locus (Vasemagi et al. 2005; Namroud 2008).

SNPs represent the simplest genetic polymorphism possible, the substitution of one nucleotide by another. Such sites occur in very high numbers in the genomes of higher organisms. Recently, they have been regarded as the new genetic marker of choice, due to high frequencies, simple mutation mechanism and low error rates (Ryynanen et al. 2007). Information concerning the utility of SNPs for population genetic and evolutionary studies in nonmodel organisms is scarce, in particular in genome scans (e.g. Rengmark et al. 2006).

An important additional aim was to test the consistency of genome scan approaches via re-examining a given set of contrasting populations with an extended number of genetic markers. On one hand, neutral loci may be classified as being under selection due to incorrect models of demographic history (e.g. island vs stepping stone models, see Akey et al. 2004) or ascertainment bias (Thornton and Jensen 2007). On the other hand, some loci influenced by selection might be missed by genome scanning (false negatives, Teshima et al. 2006). We will thus for the first time an earlier genome scan in order to evaluate the usefulness of genome scans for detection of selection.

Material and methods

Collection of samples and DNA extraction

Three *Z. marina* population pairs from the North Frisian Wadden Sea (Northern Germany) were sampled. We extended the analysis of samples described in Oetjen and Reusch (2007) by reanalysing six populations sampled in 2005. Samples from three subtidal populations and three intertidal populations were reused, and no further sampling was done for this study. Sampling sites were located on three islands: Hallig Hooge, Hallig Langeness and Sylt island (for coordinates refer to Oetjen and Reusch 2007). While intertidal populations were located at the shores of the islands in the tidal flat area (Langeness: ca 1 km from shoreline, Hooge: ca 100 m, Sylt: ca 100 m), subtidal populations grew in tidal creeks inland the islands. Each population pair comprised one subtidal location and one intertidal location, thus spanning a habitat contrast while being maximally 4 km distant from each other. Maximal distances between sampling sites was 53 km. We collected samples at ≥ 1m random points by areal sampling on 40 m x 20 m on tidal flats, and by sampling transects along the creeks which was dictated by their elongated shape. By this sampling strategy, we avoided sampling clones more than once. As earlier research has shown, on tidal flats, plants grow mainly individually and not in clone patches and clones in creeks normally do not exceed 1 m in diameter (T. B. H. Reusch, personal observation). Sample sizes ranged between 39 (Sylt subtidal) and 50 individuals (Langeness intertidal). In total, 284 individuals were collected at the 6 locations, none of which were clone members. Seagrass samples were taken and total genomic DNA was extracted as described by Reusch (2002).

Genetic markers

In total, 37 Single nucleotide polymorphisms and 40 microsatellite markers were initially tested in a genomic scan approach. SNP markers were developed on a transocean-wide panel of genotypes including Pacific and Atlantic populations by S. Ferber. PCR conditions and genotyping procedure are described in Ferber et al. (2008). Initial testing revealed that of the 37 SNP markers, only 6 were polymorphic at the spatial scale of our study (Table III.1).

The microsatellite loci included 29 gene-linked and 11 anonymous markers (EMBL nucleotide database, accession no. AM408830-AM408843, AJ249303-AJ249307, AJ009898, AJ009900, AJ009901 and AJ009904) (Oetjen and Reusch 2007). Of the gene linked microsatellites (Table III.1, accession numbers FN435336-FN435350) 15 were newly developed from an EST database (http://drzompo.uni-muenster.de/). The database was screened for microsatellite motifs using the PERL script MISA (http://pgrc.ipk-gatersleben.de/misa/) under the prerequisite that dinucleotide motifs had to include at least 7 repeats, tri- and tetranucleotide motifs at least 5 repeats.

Genotyping

Taking SNP and microsatellite markers together, we assessed 46 genetic markers in total. Genotyping was carried out on a 3130xl ABI automated sequencer (Applied Biosystems). No clone was sampled twice, so we proceeded analyzing all samples. Of the 284 individuals genotyped, only 13 samples lacked information at one locus. For the microsatellites, we used standard fluorescent PCR with 6-FAM and HEX-labelled forward primers. For the SNP scoring, we used the SNaPshot kit (Applied Biosystems) following the protocols in Ferber et al. (2008). The size calling and allele binning were performed with the software GENEMAPPER (Applied Biosystems) against ROX 350 (microsatellites) and Liz 120 (SNP) size standards. The calculation of allele frequency, expected and observed heterozygosities and population differentiation (F_{ST}) was performed with the software MICROSATELLITE ANALYZER (MSA; Dieringer and Schlötterer 2003). Tests for null alleles were carried out using the MICROCHECKER 2.2.3 software (Van Oosterhout et al. 2004) with a 99% significance threshold and revealed no homozygote excess at any of the loci except for the anonymous microsatellite GA35 that was not detected as an outlier in any analysis.

Table III.1 Eelgrass *Zostera marina* : 15 new ly developed microsatellites and six SNPs loci first used in a genome scan

Locus name	type	Repeat motif	Primer sequences	Allele Size (range)	No. of alleles	H_E	PCR multiplex	Primer conc. (µM)	No. cycles
CL766Contig1	Msat	(GA)9	F: *FAM*-GAACGTTTCCCGGTCATTT R:GGAATCGGTCAAGCAAAAAC	122-128	4	0.066	Pool F	0.02	29
CL11Contig1	Msat	(AG)7	F:*FAM*-GTGGAGGAAAGTGTGGGTGT R:CTTGCATCCACCTTCATTTG	294-300	3	0.022	Pool G	0.5	30
CL559Contig1	Msat	(AG)10	F:*FAM*-CCACTTCCGTAGTTGCTGTT R:CGATGAGGACGATGAGGAAT	171-177	4	0.153		0.14	
ZME02125	Msat	(CCT)5(CTT)5	F:*FAM*-CGTTCAACTCAACACGCATT R:GGTGACGAAAAGAAGCGAAG	103-112	4	0.043		0.05	
ZMF02381	Msat	(ATC)8	F:*HEX*-GTGCAGGCGATCGAGTTATC R:AAATTCGAGCTCTCAACTTCAA	121-154	11	0.671		0.25	
CL202Contig1	Msat	(TA)7	F:*FAM*-TTGAAAAGATTAATTATTGGTGGTG R:TCAAGTCCGGATAAATTCGAT	200-206	4	0.121	Pool H	0.45	29
CL380Contig1	Msat	(CATC)5	F:*FAM*-CCGCCTTCTTCTTCGTTAGA R:TGTTGTTCTTGGAAAAGAATCAGT	121-133	5	0.069		0.1	
CL805Contig1	Msat	(AG)8	F:*HEX*-GGGGAGGTTTCCGAATACTTT R:TGGAAGATGTTGGACATGGA	184-208	4	0.034		0.35	
CL172Contig1	Msat	(TGGC)8	F:*FAM*-CTCCTGGACGCAGAAATATG R:GACAAACGATTAATTCAGAAACAAAA	184-208	7	0.297	Pool I	0.15	27
CL53Contig1	Msat	(GAT)6	F:*FAM*-AACTCCTGGCGCAACTACTG R:CTTCGTTTGCGGTTGCTT	90-93	2	0.011		0.15	
ZME06302	Msat	(GCA)5	F:*FAM*-TCTAGCTTGTCGATGGCTGA R:CCGTCAAATGTTTCCAAGGT	291-297	2	0.022	Pool J	0.2	28
ZMC05062	Msat	(CT)8	F:*HEX*-GAAGCCAACTTAATTCAACATCG R:TTAATATAAATCCGAGACACAGACTC	96-100	3	0.007	Pool K	0.225	29
ZMC19062	Msat	(GAC)5	F:*FAM*-CACTCTCCTCTTTCCGTTCG R:CAGGGGCCTTCCTCTTACTC	293-296	2	0.003		0.125	
ZME05315	Msat	(AC)8	F:*FAM*-AAACGAGATGGTGGTTCCAT R:TGCGAGCAGCTAACTAAGTCC	174-185	7	0.057		0.125	
ZME02369	Msat	(TTC)5	F:*FAM*-AAGTCGAAATGGGGATACCA R:TCGTCGGAAGAAAAAGAAGC	99-102	2	0.007	Pool L	0.15	30
287CT623*	SNP					0.089			
71CA329*	SNP					0.403			
20CT465*	SNP					0.403			
71TC178*	SNP					0.398			
71CT+	SNP					0.396			
95GT380*	SNP					0.492			
98AT518*	SNP					0.330			

For the microsatellites, the primer sequences and the repeat motifs are given. Two to five primer pairs were pooled. PCR were performed as follows: 0.2 µl DNA template in a 10 µL reaction; 1 µL 5x reaction buffer (Promega); 0.02 mM of each dNTP; 1.5 mM of MgCl2 and 0.25 units of Taq polymerase(Promega).
Thermocycling programme: initial denaturation 94°, 3 min, followed by 27-30 cycles of 1-min denaturation 94°, 1-min,annealing at 56° and 1 min extension at 72°, followed by afinal extension of 10 min at 72°.* Indicated SNP which were assessed according to Ferber et al.(2008).

Population structure

As baseline data prior to genome scan analyses, we carried out several nested analyses of molecular variance (AMOVA) to determine the component of genetic variation (estimated as F_{ST}) due to habitat contrasts. Genetic variation was partitioned among sampling sites and among population pairs nested within sites. Consequently, we grouped populations according to sampling sites, each group containing one subtidal and one intertidal population. In doing so, we summarized the effect of differentiation across the three population pairs sharing the same habitat contrast. Thus, nested AMOVA is capable of revealing a potential signal that might not be traceable else wise. Comparison of AMOVA results regarding the group of loci identified as outliers before and the remaining (neutral) loci gives insights into the effect of habitat differences on genetic variation in the two loci groups. All AMOVAs were carried out with the software Arlequin 3.11 (Excoffier et al. 2005).

Testing for divergent selection

We applied three different tests for divergent selection to the three population pairs on tidal flat or in subtidal creeks. Due to the spatial proximity of the population pairs, we concentrated on the detection of diversifying (divergent) selection which is defined here as selection that acts in contrasting directions in two populations. The first test was implemented in the software FDIST2 (http://www.rubic.reading.ac.uk/~mab/software/fdist2.zip) by Beaumont and Nichols (1996) and compares F_{ST} estimates of single loci in population pairs with a joint distribution based on weighted mean F_{ST} of all loci. The distribution is constructed using coalescent simulations. From this distribution, individual V values for every single locus are calculated. We regarded loci that lie outside the 0.95% quantile of the distribution as outliers, i.e. corresponding to a P-value of < 0.05 (one-sided test). We applied an infinite allele model that corresponds better to the allelic states and distribution of our data than the alternative, a stepwise mutation model. Settings of the model also comprised assumption of 100 demes (the maximum number), being the most conservative version after testing, number of populations was set to 2, and median sample size corresponded to the respective population pair (ranging from 98 to 85, as the model uses haploid data).

The second model approach is based on different assumptions concerning population history, albeit using population differentiation parameters to assess locus behaviour also. It was developed by Vitalis et al. (2001) and implemented in the software DetSel 1.0 (Vitalis et al. 2003). The coalescent simulations are based on the assumption that the two populations of a pair originate from one population that has split at a time point in the past. We applied three mutation rates (μ= 0.001, 0.005 and 0.01) that were realistic as former investigations of pedigree data had shown (TBH Reusch, unpublished data). We used the maximum population size before split (N_0=500) and simulated ancestral effective population sizes (N_e) of 500, 1000 and 10 000 individuals. Two other parameters were modulated until the achieved allele distribution resembled our original data, which is recommended by the authors of the model. T_0 (= time since an assumed bottleneck) was set to 50, 100 and 200 generations, and t (= time since the population split) to 50 generations. As allele numbers for all SNPs were restricted to two, this may distort the fit of the model data because microsatellites typically have > 5 alleles. Hence, we repeated all analyses with this model excluding the 6 SNPs from the data. 100,000 coalescent simulations created the distribution of the population differentiation parameters *F1* and *F2*. Identification of outlier loci was done by comparing the distribution conditioned on the number of alleles (k= 2,3,4,5 and >= 6) with the *F1/F2* values of single loci.

As a third approach we used the lnRH test that compares the variability of single loci within a larger sample of marker loci (Kauer et al. 2003). This is done by calculating the ratio of gene diversity (*H*, heterozygosity) in two populations of all loci. As such, the lnRH test should be most influenced by the specific group, and sample size of markers chosen. It has been shown that lnRH is approximately normally distributed under neutrality (Kauer et al. 2003). Constructing a normal distribution from the data obtained of our 47 loci, we tested whether the lnRH values of the single loci were located outside the 95% confidence interval.

As this study is an explorative one, we did not apply a strict Bonferroni correction to the significance levels of outlier tests, in line with other such studies in the field (Vasemagi et al. 2005; Bonin et al. 2006).

Functional assignment

The function of genes (according to BLASTX hits) linked to loci detected in outlier tests was categorized using gene ontology (GO) functional groups. The gene ontology system has been developed as a standardized vocabulary to characterize genetic functions and to allow comparison of genes across species, organisms, tissues or cells (Ashburner et al. 2000; Harris et al. 2006). The system comprises three main categories, 'molecular function', 'biological process' and 'cellular component', for all of which entries for a certain gene can exist. In each of the categories, subcategories with partly hierarchical order comprise bigger or smaller groups of genes under broad- to fine-scale terms.

Results

Population differentiation and genetic diversity

Among the six Wadden Sea *Z. marina* populations, the level of polymorphism varied considerably among the 46 loci. The expected heterozygosity of microsatellites ranged between 0.003 (locus ZMC19062) and 0.897 (locus GA35). As expected, heterozygosity values for the biallelic SNP loci were lower and ranged between 0.08 (287CT623) and 0.49 (95GT380). The number of alleles of microsatellites varied between 2 (CL53) and 29 (GA35). The global F_{ST} estimate according to Weir and Cockerham (1984) across all populations indicated low but significant genetic differentiation (F_{ST}= 0.017, $P < 0.0001$). Pair wise genetic differentiation between the three population pairs was also low but nevertheless significant. F_{ST} values for the contrasting population pairs on Hallig Hooge were 0.008, and 0.004 and 0.014 for Sylt island and Hallig Langeness (all markers). Excluding the six polymorphic SNP loci changed F_{ST} values very little.

A comparison of habitat types revealed slightly lower average allelic richness in subtidal (4.03) than in intertidal habitats (4.42), an effect was almost significant (two-tailed t-test, P = 0.08). Expected heterozygosity was similar between subtidal (0.36) and intertidal (0.37) habitats (ns, P = 0.37). In an AMOVA, most of the genetic variation was

explained by the within and among individual component (98.16%), while only 0.86% of variation was explained by habitat differences.

Tests for divergent selection

The three model approaches we used for identification of loci influenced by selection revealed partly consistent results (see Table III.2 for an overview on candidate loci). The approach by Beaumont and Nichols detected only few loci which is consistent with earlier results (Vitalis et al. 2001; Vasemagi et al. 2005; Oetjen and Reusch 2007). The DetSel model by Vitalis et al. detected several loci as being subject to selection. Reassuringly, all three loci that had been detected in an earlier study with far fewer loci (Oetjen and Reusch 2007; GA20, ZMC19066, ZMC01058) were also detected with the increased number of markers. The locus GA20, an anonymous microsatellite, was found to be an outlier in all three population pairs. Locus ZMC19066 (linked to acid phosphatase gene) was an outlier in the population pair on Hallig Langeness ($P < 0.01$) and on Hallig Hooge ($P < 0.03$) in our study. Locus ZMC01058 (linked to vacuolar H+-ATPase) proved to be an outlier in the population pair on Sylt island ($P < 0.01$), on Hallig Hooge ($P < 0.02$) and on Hallig Langeness ($P < 0.03$) also. One locus, that had been detected as an outlier in two population pairs (Hallig Hooge and Sylt island, both $P < 0.01$) earlier, CL412 (linked to putative nodulin 3), showed the same pattern in our recent study. A newly developed microsatellite (ZME05315, linked to putative ubiquitin-conjugating enzyme), was a significant outlier ($P < 0.01$) in population pair Hallig Hooge and on Sylt island ($P < 0.03$). Weaker evidence for selection was found for two other loci that were detected in only one of the three population pairs (newly developed microsatellite CL559, linked to putative seed maturation protein, and SNP 71TC178, both in Langeness population pair with $P < 0.01$). Excluding all SNPs from the data yielded qualitatively similar results.

The lnRH test detected eight outliers with *P* values significant or marginally significant at the 5% level. Most of them were detected in only one population pair (CL380 / $P = 0.0004$, CL679 / $P = 0.04$, ZME5315 / $P = 0.0188$ in pop pair Hallig Hooge, CL412 / $P = 0.0283$ and ZME2125 / $P < 0.0001$ in pop pair Hallig Langeness, CL766 / $P < 0.0002$ and ZMC19066 / $P = 0.018$ in pop pair Sylt island), except locus CL559, which occurred in two

comparisons (Hallig Hooge, *P* = 0.013, Sylt island, *P* = 0.0125). All of these loci were detected by DetSel in the respective population pair as well.

Table III.2 Microsatellite loci in eelgrass *Zostera marina* detected as statistical outliers in comparisons among tidal flat and subtidal populations either by more than one test procedure (Fdist, DetSel lnRH), or in more than one habitat contrast

Outlier locus	Habitat comparison/population pair									
	Hooge			Langeness			Sylt			
	Fdist	DetSel	lnRH	Fdist	DetSel	lnRH	Fdist	DetSel	lnRH	Homology / BLAST Hit
ZMC19089				**0.0140**	**<0.01**					*Arabidopsis thaliana* hypothetical protein CAB81404
CL734	**0.0190**	<0.02								*Oryza sativa* unknown protein NP_913907
GA20		**<0.01**			**<0.01**			**<0.01**		Anonymous microsatellite
ZMC01058		<0.02			<0.03			**<0.01**		*Gossypium hirsutum* vacuolar H+-ATPase AAA82977
ZMC19066		<0.03			**<0.01**	**0.0180**				*Arabidopsis thaliana* acid phosphatase NP_194655
CL412		**<0.01**						**<0.01**	0.0280	*Oryza sativa* putative nodulin 3 XP_465955
ZME05315		**<0.01**	**0.0188**					<0.03		*Arabidopsis thaliana* putative ubiquitin-conjugating enzyme 16
CL380		**<0.01**	**0.0004**							*Arabidopsis thaliana* nucleotide sugar epimerase-like protein Q9m0b6
CL559			**0.0130**			**0.0125**				*Oryza sativa* putative 24-kda seed maturation protein Q8s2k0
CL766					**<0.01**	**0.0002**				*Oryza sativa* putative ddt domain-containing protein Q6zi78
ZME2125								**<.0.1**	**0.0002**	*Oryza sativa* putative myb-related protein Q9aut3

Significant (bold figures) or marginally significant P values are reported, according to criteria given in 'Material and Methods' section (Fdist: P values < 0.025, DetSel P values <0.01, lnRH: P values <0.025). Those loci detected in at least two of three independent habitat comparisons are indicated in boldface

Compared to an earlier study, the results of the present study were mostly consistent. The locus ZMC19089 was newly identified as an outlier in the population pair (subtidal vs. intertidal) on Hallig Langeness (*P* = 0.014). Another locus with unknown function (CL734) became significantly different from neutral expectation (*P* = 0.019) in the population pair on Hallig Hooge in this study. A microsatellite locus (ZMA04093) linked to a purple acid phosphatase (BLAST hit: BAC55157 *Nicotiana tabacum*) is marginally

significant in the recent analysis (*P* = 0.0126) and is also closer to significance than the result of the previous scan (P = 0.0129).

An AMOVA revealed that when only the six outlier loci were regarded, the component of variance between contrasting habitat pairs approximately doubled from 0.86 to 1.64%. For the remaining 40 loci, variation between habitat contrasts decreased to 0.78 and the effect of variation between sampling sites increased to 1.09%.

Functional role of candidate loci

The strongest marker locus candidates were linked to a broad spectrum of genes. While GA20 was an anonymous microsatellite, ZMC01058 is linked to a vacuolar H+-ATPase, hence a gene that is involved in osmotic regulation. Even more interesting, ZMC19066 is located next to an acid phosphatase gene, which is associated to hyperosmotic stress according to gene ontology annotation. CL412 is located next to a putative nodulin gene which forms water channels in cell membranes and is hence also involved in osmoregulation (Luu and Maurel 2005). ZME05315 is linked to a putative ubiquitin-conjugating enzyme and CL559 to a putative seed maturation protein. The latter gene may be responsible for the much higher seed output of the intertidal populations compared to subtidal ones.

Discussion

The present analysis extends previous genome scans in *Z. marina* by doubling the number of markers employed and adding a third outlier test. For the first time in a genome scan in a marine plant, we included SNPs as a novel marker type. As a goal of this study, we assessed how 2 years of additional marker development would affect the reliability of results collected with relatively few markers. Several genome scan studies have reanalysed existing data sets with different outlier tests (Vitalis et al. 2001; Beaumont and Balding 2004) or assessed the validity of outlier tests by using more than one test statistic (Vasemagi et al. 2005; Kane and Rieseberg 2007; Tsumura et al. 2007; Egan 2008; Mäkinen et al. 2008). In contrast, extending a genome scan by doubling the

number of markers has been done here for the first time. Reassuringly, we found that earlier results of the genome scan were largely confirmed. Compared to previous analyses, the increased number of markers in our study was able to detect selection more reliably and to confirm most of the outliers detected before. All three loci that seemed to be influenced by habitat-dependent selection based on a smaller amount of data (ZMC19066, ZMC01058, GA20) were confirmed by the extended approach. Additionally, two new candidates could be detected. ZME05315, linked to a putative ubiquitin-conjugating enzyme, and CL559, linked to a putative seed maturation protein, revealed strong evidence for influence of selection.

It is well known that even under relatively high gene flow, selection gradients, if strong enough, will counteract gene flow (Silander 1979). Although the three population pairs are farther away from each other than the two habitat types, we cannot exclude that gene flow across similar habitats by tidal currents transports alleles under selection within a habitat type. Nevertheless, if true, homogenizing gene flow across the pairs at any one location is probably higher and would erase the signal of selection, with or without parallel gene flow.

Among the six SNPs loci tested in the genome scan, there were no further candidates. Only at one SNP (71TC178, detected in only one population pair), we found weak evidence for selection. This is most likely due to the low number of polymorphic SNPs included in the study. Although we originally targeted for approximately the same number of SNPs than microsatellites, most SNPs developed on a global panel of *Z. marina* genotypes (Ferber et al. 2008) were not polymorphic at the spatial scale chosen in this study. The reason for this is most likely the close proximity and genetic relationship of the populations that do not comprise a deeper phylogeographic signal that would be needed to detect polymorphism for this marker type. Note that the SNPs used here were detected based on a genotype panel that comprised widely distant populations from Atlantic and Pacific populations. Despite being valuable tools for population genetic analyses at larger geographic scales (Ferber et al. 2008), development of SNPs is costly where no in silico screening can be done and screening effort is higher as the information content per SNP is low. For example, about five times more SNPs than microsatellites are needed to determine relationships between

populations (Glaubitz et al. 2003). Nevertheless, it should be possible to use them for genome scanning in *Z. marina* also in the future, especially as a significant number of SNPs are ready to use with good amplification and screening properties. In silico search for novel SNPs is also easily feasible with a sound database, namely ESTs (http://drzompo.uni-muenster.de/; Wissler et al. 2009), already available and a *Zostera* genome project to be finished in the near future.

Despite the overall consistency of the two genome scans, we found some signals of selection to become weaker which sends a cautionary message to all studies using genome scans with relatively few marker loci, as results definitely have to be regarded as preliminary in such cases. This is even more important as the outlier test that detects the highest number of outlier loci (the DetSel test) has been criticised before, especially when the outlier findings are compared to the results by the Fdist test (see Vitalis et al. 2001; Vasemagi et al. 2005). Congruency of these two test statistics seems to be poor. Our study revealed close congruency of the DetSel test and the lnRH method, as all nine loci detected by lnRH were also detected by DetSel in the respective population pair. This is particularly encouraging because both approaches work differently, one by using an empirical null distribution (lnRH), the other a demographic model (DetSel).

The number of markers detected as outliers in this study was slightly higher (6 (13%) out of 46) compared to other recent studies (between 1.4 and 9.5%; Campbell and Bernatchez 2004; Vasemagi et al. 2005; Bonin et al. 2006; Kane and Rieseberg 2007). It is to be expected that for gene-linked (EST) markers, more signals of diverging selection can be detected than for anonymous markers (e.g. 21% in Scotti-Saintagne et al. 2004; 12% in Vasemagi et al. 2005). This is because for the latter type, chances are high that markers to reside in intronic and intergenic regions physically distant from a functionally important gene, thus missing the 'hitchhiking' signal.

Because the number of marker loci tested is still rather small, we cannot test whether or not the six outliers are over- or underrepresented in terms of their function (defined by GO – gene ontology categories) compared to the overall spectrum of genes. Notwithstanding, some loci merit attention, such as CL412, which is linked to a putative nodulin class 3. Nodulins belong to the protein class of aquaporins that form trans-

membrane water channels in cellular membranes (Luu and Maurel 2005). This locus, as well as ZMC01058 (linked to a vacuolar H+-ATPase) is likely to be involved in osmoregulation and water balance. It is conceivable that changes in either structure or expression of such genes are needed given the drastically different environmental conditions on tidal flats compared to a permanently submerged habitat (Reise 1985). Another interesting locus is CL559, linked to a putative seed maturation protein. As populations inhabiting both habitat types differ in their reproductive strategy – predominantly annual versus perennial, this gene points to adaptive divergence at the molecular level. The ripening process of seeds is an important factor in the life history of plants influencing perennial persistence or annual growth from new seeds. Above genes may serve as valuable starting points for further investigation into the genetic basis of natural selection and habitat adaptation in *Z. marina*.

For *Z. marina* as an engineering species with outstanding ecological importance, the prospects pro studying the genetic basis to ecological adaptation are promising, given a *Zostera* Genome Project has been initiated by the Joint Genome Institute. The limitations of current genome scan studies brought along by moderate marker numbers and restriction to EST sequences will be obsolete. In particular for SNP markers that allow a much higher throughput than microsatellites, tailored polymorphic sites in genes selected by their functionally relevance (e.g. osmoregulation) will revolutionize the possibilities to detect natural selection in natural settings. Because Z. marina grows under very different environmental conditions locally and spans a wide biogeographic range from the White Sea to Portugal in Europe alone, it will become a model for how genetic and ecological information can be merged in a common framework of ecological genetics and genomics (Ouborg and Vriezen 2007).

Acknowledgements

We thank B. Hasert for technical assistance. Many thanks to Prof. Karsten Reise from AWI Sylt for help and support, as well as to Andreas Zipperle for help collecting samples. This work was funded by Max-Planck society and the DFG (grant no. Re 1108/7-1).

CHAPTER IV

Sunrise over tidal flat

CHAPTER IV

Intermediate heterozygosity is associated with optimal fitness in the clonal plant Zostera marina

Abstract

The optimal outcrossing principle predicts that offspring from parents with intermediate genetic similarity will have a higher relative fitness compared to lower and higher heterozygosities. In long-living clonal organisms, the analysis of heterozygosity-fitness correlations (HFC) provides a powerful tool to asses this relationship. In this study, clone size serves as proxy for fitness in the seagrass *Zostera marina*. We assessed the correlation among clone area and multi-locus heterozygosity (MLH) based on 37 microsatellite markers most of which were gene-linked. For the first time in a clonal organism, we show that intermediate heterozygosity is associated with the largest clone sizes, confirming the hypothesis that not only inbreeding but also outbreeding depression account for fitness in a long living clonal organism.

Introduction

The optimal outbreeding theory predicts that offspring from parents with intermediate genetic similarity will possess maximum fitness (Bateson 1983). Mating of closely related individuals can cause inbreeding depression due to combination of deleterious recessive alleles, leading to a reduced fitness of the offspring (Charlesworth and Charlesworth 1987). Such inbred individuals possess a lower overall genetic heterozygosity than expected in Hardy-Weinberg-equilibrium. At the other extreme, dubbed outbreeding depression, fitness is also reduced, possibly due to the break-up of favorable allele combinations (positive epistasis). Here, heterozygosity values are higher than expected. Consequently, with genome wide heterozygosity representing the outbreeding level, individual fitness should be highest at intermediate levels of overall genetic heterozygosity (Thornhill 1993).

Examples of this relationship among plant and animal species have been demonstrated experimentally or revealed by parent-offspring analyses (Price and Waser 1979; Richard et al. 2009); albeit in other cases the hypothesis could not be confirmed (Robinson et al. 2009). Such experiments are difficult in long-lived species, as for example clonal organisms, where selection acts over long time scales and will accumulate even small fitness differences. One alternative over experiments is the assessment of heterozygosity fitness correlations using molecular genetic markers. Studies comparing multilocus marker data have revealed molecular data that support the intermediate heterozygosity idea (Marshall and Spalton 2000; Greeff et al. 2009; but see Blanchet et al. 2009). Although the method has been criticized because multilocus heterozygosity (MLH) does not always reflect inbreeding (Balloux et al. 2004), the correlation between MLH and genome wide heterozygosity holds true at least under certain conditions (Tsitrone et al. 2001).

Our question is whether the intermediate heterozygosity hypothesis holds for wild populations of clonal plants also. Our study system, the seagrass *Zostera marina* is an ecosystem engineering species forming submerged meadows on soft sediments. Clonal reproduction is frequent, leading to spatially extended *Z. marina* clones in the Baltic Sea region that have been shown to be extremely long-living (Reusch et al 1999a). Previous

work has shown that the area occupied by a *Z. marina* clone is positively correlated with reproductive output (Hämmerli and Reusch 2003b), which makes clone size a proxy for fitness. Also, it was shown before that clone size increased with increasing heterozygosity estimated from 9 microsatellite loci (Hämmerli and Reusch 2003b). A trend suggested that very high heterozygosities were associated with reduced clone size, but the intermediate heterozygosity hypothesis could not be tested for appropriately because of the limited number of markers. Consequently, we want to increase the number of markers drastically to obtain more heterozygosity classes for a high definition of the optimal heterozygosity signal.

Material and Methods

Study site and sampling

In the nontidal Baltic Sea, *Z. marina* grows clonally in patchy or dense beds. In the Kiel bay, the area covered by single clones ranges from less than one to over 30 square meters. We analyzed plant material from two closely located study sites at Falckenstein Beach, Kiel bay, northern Germany. Sampling grids were constructed and clones have been sampled and assigned to genotypes with microsatellite markers for a previous study (Hämmerli and Reusch 2003). Data of the two sampling sites were pooled as they were very similar regarding the measured variable. Clone sizes were assessed based on a 1x1m grid and clone sizes refer to the number of grid pixels, i.e. square meters where a clone was present.

Here, we examine 26 EST-derived microsatellites (Oetjen and Reusch 2007a, Oetjen and Reusch 2010), extended by two anonymous microsatellites (GA1, GA4, described in Reusch et al. (1999b)) and combine the results with existing data of 9 additional anonymous microsatellites. We performed our analyses with clones already analyzed in 2003, although the number of genotypes is slightly reduced due to loss of individual genotypes during DNA extraction or genotyping. For reason of comparability, we only included clones in our analysis that were genotyped at all loci. The removed individuals were a random selection of clone size classes. In the end, 167 clones were successfully genotyped at all 37 loci.

Genotyping and statistical analysis

The 26 EST- microsatellite markers were genotyped according to the protocols given in Oetjen and Reusch (2007a) and Oetjen et al. (2010). GA1 and GA4 were run in the multiplex 'Pool E' as described in Oetjen and Reusch (2007). Data from 9 previously genotyped anonymous microsatellites was included. Genotyping of the 28 new markers was done on a 3130xl genetic analyzer (Applied Biosystems); allele scoring was done using the software GeneMapper (Applied Biosystems). As a measure of individual heterozygosity in every clone, multilocus heterozygosity (MLH) was quantified as the proportion of the genotyped loci for which a clone was heterozygous. When used for estimation of inbreeding depression, MLH has advantages over alternative methods like the d^2 index specifically designed for microsatellites (Hansson 2010). Clones were grouped into heterozygosity classes derived from the fraction of heterozygous loci (which resulted in 13 different heterozygosity values).

We carried out a randomization test in order to produce a distribution of clone size – heterozygosity combinations under a true null hypothesis. In a second step, we assessed whether or not actual values of clone size / heterozygosity combinations differed from expected values. Based on 1000 randomizations, we calculated 95 % confidence intervals for the expected abundance of every clone size class. Statistical analyses were carried out in R (R development core team).

Results

Multilocus heterozygosities (MLH) among the 167 clones ranged from 0.135 to 0.486, with average heterozygosity of 0.29 (corresponding to 10.8 of 37 loci). The frequency distribution was unimodal. Among the six largest clones, five showed a MLH of 0.27, which was very close to the average over all clones (figure IV.1).

Mean clone sizes for all heterozygosity classes showed that the relation of MLH and clone size can be described as a curve with a maximum at intermediate heterozygosities and declining at both ends of the distribution (figure IV.2). A 1000 times randomization test showed that for low heterozygosities, clone sizes were below the expected mean (applying a 95% confidence interval). For the 0.27 heterozygosity class, clone size was

above the expected mean and for high heterozygosities, clone sizes were again lower than expected (see figure IV.2).

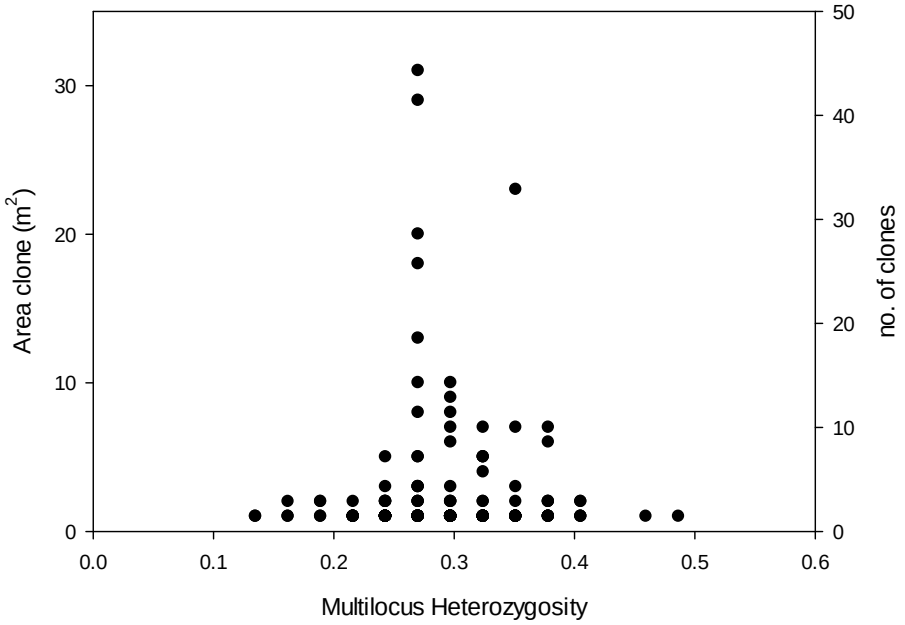


Figure IV.1 Distribution of clone area (n=167) as a function of multilocus heterozygosity. Average MLH = 0.29. Line: frequency distribution, i.e. number of clones per heterozygosity class

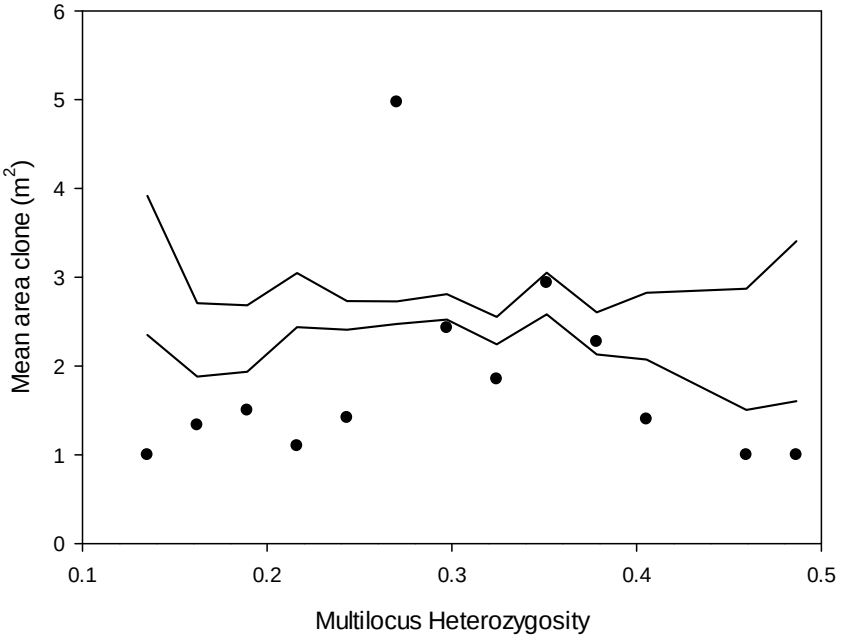


Figure IV.2 Mean clone areas for all classes of MLH (dots) and 95% confidence interval (lines) depicting the null-hypothesis of random matching of clone size and heterozygosity (alpha = 0.05).

Discussion

We showed that an intermediate level of heterozygosity is associated with the largest clone sizes in the ecosystem-engineering plant *Zostera marina*. To the best of our knowledge, this is the first assessment of the intermediate heterozygosity concept in a clonal organism. As clone size is a close proxy for fitness in any clonal organism, including seagrasses, we conclude that the highest relative fitness values correlate with intermediate heterozygosity, in line with the predictions Bateson 1983; Thornhill 1993). Our results confirm experimental crossing in *Z. marina* (Billingham et al. 2007), which however only assessed survival of seeds and seedlings.

Heterozygosity-fitness correlations (HFC) can be explained by three alternative mechanisms (see Hansson and Westerberg 2002), but only one includes effects of inbreeding: First, direct effects of the regarded loci account for fitness differences, which is arguable in our case as microsatellites are mostly selectively neutral (but see Li et al. 2004 and Sharopova 2008). The association of neutral markers and fitness traits can be caused by local effects (also termed associative overdominance), which implies marker loci and fitness loci are in linkage disequilibrium. As most of our markers have been developed from EST sequences, it is possible that close linkage to genes causes such an effect. Even if only few markers may be affected, a sufficiently strong signal could be created to shape the optimal heterozygosity curve. Alternatively, if identity disequilibrium between marker and fitness loci exists, the reason for fitness consequences would be inbreeding depression caused by homozygous loci (general effect). As local effects are unlikely to have severe impact on HFC regarding fitness traits like growth and survival (Szulkin et al. 2010), we assume that inbreeding effects or linkage disequilibrium – or a combination of both - has produced the fitness disadvantage of clones with low MLH in our case. The reduction of fitness with higher heterozygosity levels can be explained by disruption of favorable allele combinations or gene-environment combinations (Templeton 1986) caused by local adaptation existing in Baltic *Z. marina* populations (Hämmerli and Reusch 2002). Thus, a general outbreeding effect that invokes fitness costs at high MLH is a likely explanation for the reduced clone sizes at higher heterozygosity levels. In conclusion, we state that effects

of inbreeding and outbreeding account for a fitness optimum at an intermediate heterozygosity level in *Z. marina* clones.

The particular shape of the heterozygosity-fitness relation is also of interest, as the fitness optimum is very narrow. The reason for this might be that in plants that reproduce by selfing frequently, the costs for inbreeding are likely to be reduced but the fitness optimum will be smaller (Willi and van Buskirk 2005). As reports about selfing rates of *Z. marina* differ greatly, we cannot exclude that selfing occurs in our populations and contributes to HFCs.

The importance of clonal plants as ecosystem foundation species cannot be overestimated as they structure and build many ecosystems such as grasslands, coral reefs, forests and seagrass meadows. In these species, intra-specific genetic diversity has been found to influence the resilience of ecosystems (Reusch et al. 2005) and to affect clonal performance under stressful conditions in the field (Hughes and Stachowicz 2009). Our results show that not all clones are alike, with some possessing better long-term growth and persistence. Research in the future should combine diversity levels at the population level with the intra-individual level, in order to better predict the resilience of stands of seagrass and other clonal species.

Acknowledgements

We thank I. Rauch and B. Hasert for laboratory assistance. This work was funded by the DFG (grant no. Re 1108/7 to TBH Reusch).

Conclusions

Studying the genetic basis to ecological adaptation in *Zostera marina* by genome scanning has provided two main insights: First, genome scans in non-model organisms as an explorative approach can render valuable candidate loci for further investigation (chapter II). We have shown that based on a habitat contrast and phenotypic and life-cycle differences, consistent effects of divergent selection could be found at several marker gene loci (chapters II and III). One considerable advantage for interpretation of the genome scan was that a fraction of markers were EST-derived and thus, information about coding regions adjacent to the marker site was available. It is thus recommended to use EST markers, because of their fast and cost-effective development once an EST database exists (chapter I), and provided that the number of markers that can be employed is moderate (say ≤ 100). Nevertheless, with new approaches such as RAD genotyping, where 1000s of markers can be interrogated, such advantages are likely to disappear (e.g. Hohenlohe et al. 2010).

In our case, the functions of candidate loci point to habitat dependent selection regarding osmoregulation and reproduction (chapter III), which makes biological sense. Further ecophysiological investigation of phenotypes possessing the respective allelic variants at these loci is needed to reinforce the connection between environmental selection and genetic patterns. Genome scans can only be the first step towards uncovering selective effects in natural populations (Luikart et al. 2003; Storz 2005; Vasemägi and Primmer 2005). In *Z. marina*, interpretation of the genome scan results and subsequent analyses will proceed further as soon as the whole genome is sequenced and (mostly) annotated. Due to this missing link between marker-based pieces of genetic information and the whole genome, it cannot be said at the moment how the genome scan information will be used in the future.

The second insight is that genome scans must be handled cautiously. As my research has shown, genome scans can suffer from their inherent weaknesses, in particular the necessary simplification of underlying demographic scenarios when deriving null-models, and the multiple testing problem. Considering the first problem, one consequence is that models provide so different results that overlapping outliers sometimes do not exist at all (chapters II and III). Although making a virtue out of a necessity by combining different models is possible, we must ask ourselves whether what genome scans detect is really meaningful. Even if model construction will make progress in the future, and if several models would be used, evidence from one single genome scan cannot be totally convincing. A possible solution to alleviate the

above problem is to use repeated population comparisons (chapters II and III), and to regard only those loci as candidates that have been detected repeatedly in independent comparisons. The multiple testing problem will remain central in genome scans (e.g. Luikart et al. 2003). Careful evaluation of the candidate status of every detected locus is necessary unless – again – repeated comparisons serve to confirm the outlier status. Whenever application of repeated comparisons is not possible, genome scans are suitable for explorative analyses but in my opinion it would not be advisable to only rely on them.

Regarding the markers we tested, microsatellites hold a definite advantage over SNP in our system as they provide more variation per marker and proved to be much more polymorphic at small scales (chapter III). The prevailing opinion that SNP are needed in notably higher numbers for the same information content (Glaubitz et al. 2003) could not even be met in our case as most markers we tested were monomorphic at our scale. Compared to microsatellites, SNP seem to be more sensitive to scale-shifting (chapter III) although this problem is not unknown in microsatellites either. Because of this limitation, SNP hold more importance in well-studied organisms where sufficient genetic information is available to initially test a large number of markers. In model organisms, SNP have already taken over the predominant position of microsatellites (e.g. Wang et al. 1998, man; Cho et al 1990, Arabidopsis; Lindblad-Toh, mouse; Duran et al. 2009, Edwards and Batley 2010, crop plants) due to their obvious advantages like high abundances in almost every target gene or gene region and lack of intercalibration problems. But also in non-models, SNP will most likely replace microsatellites in the near future (Seeb et al. 2011) as high throughput methods are being established that allow massive SNP sequencing without whole genome information (e.g. RAD genotyping, Hohenlohe et al. 2010). This is especially important for evolutionary ecologists working on species with limited genetic resources, which concerns the larger part of the ecologically most relevant species.

Among microsatellites, differences between EST-derived and anonymous markers were not as pronounced as we expected. Although variation (i.e. number of alleles) seems to be slightly lower in EST-microsatellites, genome scan results suggest that they are equally applicable for population genetic purposes (chapters II and III). It is true that overall, we found a slightly higher fraction of outlier loci (chapter III) than comparable studies, which could be caused by the EST markers. But among EST microsatellites, the same fraction of outliers was detected than among anonymous markers (chapter III), which leads to the question: Why are these markers that are located in the close vicinity of expressed genes, obviously not affected by this? Or – if marker sample sizes in our studies were not sufficient to detect such a phenomenon – is it eventually

weaker than assumed? An answer will require knowledge on the exact location of the anonymous markers which may nevertheless be close to a coding region or to a site that is polymorphic for gene expression variation, i.e. within a promoter region (Li et al. 2004). Moreover, we currently have no linkage disequilibrium information that would inform us on the physical distance over which a marker can hitchhike with a selected genome location.

In a clonal plant like *Z. marina*, we showed for the first time that an intermediate level of heterozygosity favours large clone sizes (chapter IV). Intermediate genetic similarity of the parents being advantageous for offspring fitness is almost a standard assumption (Bateson 1983). As one possible explanation, mating of very far related individuals (as from different groups) may disrupt favourable gene-gene or gene-environment combinations. The important questions are therefore, which amount of dissimilarity is beneficial and when detrimental effects begin to overtake. *Z. marina* is a sessile organism with limited dispersal and resulting spatial genetic structuring, even at small scales (Billingham et al. 2007; Hämmerli and Reusch 2003c). One would not necessarily expect that individuals resulting from matings within a closer genetic neighbourhood (as in one seagrass meadow) display an intermediate heterozygosity pattern because events of extreme outbreeding should be seldom. But obviously, not only varied heterozygosity levels sufficiently but also differences in fitness were pronounced enough to produce such a pattern. Local adaptation as a possible reason appears less likely as individuals are located in close vicinity, unless micro-habitat changes have an unexpectedly large effect. Favourable gene combinations that are disrupted in (extremely) outbred individuals can explain the pattern, although a history of at least partly isolation would be helpful to explain the development of such combinations. Here, limited dispersal would be rather a promoting factor than an obstacle, which makes me favour the second explanation.

For assessment of multilocus heterozygosity-fitness correlations, EST-derived microsatellites proved to be equally well suited as were anonymous markers (chapter IV). Results did not change when only EST microsatellites were taken into account compared to a mixture of the two marker types (K. Keil, unpublished data). Taking only anonymous markers diluted the pattern, but this is most likely due to a smaller marker sample size of only 11 anonymous microsatellites that were assessed compared to 26 EST-derived markers. Here, again the question rises whether EST-microsatellites are eventually less affected by their direct vicinity than assumed (see discussion above). Of course, the mere fact that we found no evidence for EST markers representing genome wide heterozygosity less precise than anonymous markers, does not state this is generally true. That is to say, I cannot rule out a false negative result due to the rather low

number of markers in my study. Hence, I would not recommend to use EST and anonymous microsatellites interchangeably face value, at least unless their neutrality has been proven, for example by genome scanning.

Overall, marker-based approaches to study evolution and ecology in natural populations will stay an important complement to classical ecological research that will add enormously to the diversity of novel findings and new, emerging questions (Avise 2004). As for *Z. marina*, genetic information has greatly advanced our understanding of seagrass meadow dynamics, clonal performance and local adaptation (Procaccini et al. 2007). Insights into the genetic and genotypic diversity of seagrass meadows enable us to understand the respective importance of both compartments of genetic variation in single-species ecosystems (e.g. Reusch 2006, reviewed in Reusch and Hughes 2006). Seagrasses and especially *Zostera marina* are well suited for the study of adaptation because they inhabit a wide range of habitats under extremely varying conditions. Information about local adaptation (as in contrasting habitats) is an important prerequisite for the choice of donor populations in conservation projects.

Danksagung

Ich möchte mich bei vielen Menschen bedanken, ohne die diese Arbeit nie entstanden wäre und die mir geholfen haben, in den vergangenen Jahren nie mein Ziel aus den Augen zu verlieren.

Allen voran danke ich Thorsten Reusch für seine gute Betreuung, fruchtbare Diskussionen und wertvolle Expertise - und große akademischer Freiheit, die mir ermöglicht hat, eigene Ideen zu entwickeln und umzusetzen.

Bei Joachim Kurtz (stellvertretend für das IEB in Münster) möchte ich mich bedanken, dass ich nach Thorstens Umzug weiter im Institut für Evolution und Biodiversität arbeiten durfte. Gisep Rauch danke ich für die unbürokratische Aufnahme in seine Arbeitsgruppe (die nur zur Tarnung zeitweilig LSD- Lost souls department - genannt wurde). Ich danke Prof. Karsten Reise vom AWI List / Sylt für seine Unterstützung und die Möglichkeit, im Wattenmeer Proben zu nehmen. Andreas Zipperle danke ich für Seegrasproben, Steven Ferber von der RUG Groningen für Hilfe mit den SNPs. Hilde, Barbara und ganz besonders Ilka danke ich dafür, dass sie mich großartig unterstützt haben, sonst würde ich sicher heute immer noch PCRs machen. Den lieben IEB-und IFM-Leuten, die meinen Arbeitsalltag bereichern, meine Vorträge anhören und Ihr Urteil dazu abgeben, danke ich dafür. Gisep, August, Philine und Sarah danke ich dafür, dass sie mir in Fragen der Statistik beigestanden haben. Georg danke ich, der mir mit Rat und Tat geholfen hat, dass ich im Ringen mit dem Rechner nie den kürzeren gezogen habe. Philipp war bei der weltbesten Schweden-Fahrt dabei und bildet zusammen mit Nina meine Seegras-Diskussionsgruppe. Dolors, you are an awesome officemate, I am so glad to have you here! Big Hug!

Nina und Sascha, den besten Freunden dies gibt, danke ich dafür dass sie da sind. Ohne Euch wärs nicht gegangen!

Schließlich möchte ich mich bei meinen Eltern und meiner Familie dafür bedanken, dass sie an mich glauben und immer für mich da sind. Dir, Carsten, danke ich für deine Geduld, deine Liebe und die Fähigkeit, mich immer wieder aufzumuntern.

References

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12, 1805-1814.

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. Plos Biology 2: 1591-1599

Antonovics J, Bradshaw AD (1970) Evolution in closely adjacent plant populations. 8. Clinal patterns at a mine boundary. Heredity 25: 349-&

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (2000) Gene Ontology: tool for the unification of biology. Nature Genetics 25: 25-29

Avise JC (2004) Molecular markers, Natural History, and Evolution, Second edition. Sinauer Associates, Sunderland, MA 01375 USA

Backman TWH (1991) Genotypic and phenotypic variability of *Zostera marina* on the west coast of North America. Canadian Journal of Botany - Revue Canadienne De Botanique 69: 1361-1371

Balloux F, Amos W. & Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? Molecular Ecology 13: 3021-3031

Bateson P (1983) Optimal Outbreeding. Cambridge University Press, Cambridge

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology 13, 969-980.

Beaumont MA, Nichols RA (1996) Evaluating Loci for Use in the Genetic Analysis of Population Structure. Proceedings: Biological Sciences 263, 1619-1626.

Becheler R, Diekmann O, Hily C, Moalic Y, Arnaud-Haond S (2010) The concept of population in clonal organisms: mosaics of temporally colonized patches are forming highly diverse meadows of *Zostera marina* in Brittany. Molecular Ecology 19: 2394-2407

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289-300

Billingham MR, Simoes T, Reusch TBH & Serrao EA 2007 Genetic sub-structure and intermediate optimal outcrossing distance in the marine angiosperm *Zostera marina*. Marine Biology 152: 793-801

Blanchet S, Rey O, Berthier P, Lek S & Loot G (2009) Evidence of parasite-mediated disruptive selection on genetic diversity in a wild fish population. Molecular Ecology 18: 1112-1123

Bockelmann AC, Reusch TBH, Bijlsma R, Bakker JP (2003) Habitat differentiation vs. isolation-by-distance: the genetic population structure of *Elymus athericus* in European salt marshes. Molecular Ecology 12, 505-515.

Bohonak AJ (2002) IBD (Isolation By Distance): A program for analyses of isolation by distance. Journal of Heredity 93, 153-154.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Molecular Biology And Evolution 23, 773-783.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. Molecular Ecology 17: 3583-3584

Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. Molecular Ecology 16: 907-924, doi: 10.1111/j.1365-294X.2006.03195.x.

Britten HB (1996) Meta-analyses of the association between multilocus heterozygosity and fitness. Evolution 50: 2158-2164

Campbell D, Bernatchez L (2004) Genomic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Molecular Biology and Evolution 21, 945-956.

Casa AM, Mitchell SE, Hamblin MT et al. (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theoretical and Applied Genetics 111, 23-30.

Cavalli-Sforza L (1966) Population structure and human evolution. Proceedings of the Royal Society of London Series B-Biological Sciences 164: 362-379

Charlesworth D, Charlesworth B (1987) Inbreeding Depression and its Evolutionary Consequences. Annual Reviews of Ecology and Systematics 18: 237-268

Charlesworth D, Willis JH (2009) Fundamental Concepts in Genetics. The genetics of inbreeding depression. Nature Reviews Genetics 10: 783-796

Cherdsak L, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-tranferable microsatellite markers developed from loblolly pine ESTs. Theoretical and Applied Genetics 109, 361-369.

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney L, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in Arabidopsis thaliana. Nature Genetics 23: 203-207

Cho YG, Ishii T, Temnykh S et al. (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theoretical And Applied Genetics 100, 713-722.

Coltman DW, Bowen WD, Wright JM (1998) Birth weight and neonatal survival of harbour seal pups ape positively correlated with genetic variation measured by microsatellites. Proceedings of the Royal Society of London Series B-Biological Sciences 265: 803-809

Coltman DW, Slate J (2003) Microsatellite measures of inbreeding: A meta-analysis. Evolution 57: 971-983

Costanza R, d'Arge R, de Groot R et al. (1997) The value of the world's ecosystem services and natural capital. Nature 387, 253-260.

Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Slate J, Guinness FE, Clutton-Brock TH (1998) Microsatellites reveal heterosis in red deer. Proceedings of the Royal Society of London Series B-Biological Sciences 265: 489-495

Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ (2008) Identification of genetic variants using bar-coded multiplexed sequencing. Nature Methods 5: 887-893

Darwin C (1859) On the Origin of Species by Means of Natural Selection or the Preservation of favored races in the struggle for life. Murray, London

den Hartog C (1970) The seagrasses of the world. Verhandlingen Koninglijk Nederlandse Akademie Wetenschapen Afdeling Natuurkunde II 59, 1-275.

Dieringer D, Schlötterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3, 167-169.

Duffy JE, Macdonald KS, Rhode JM, Parker JD (2001) Grazer diversity, functional redundancy, and productivity in seagrass beds: An experimental test. Ecology 82: 2417-2434

Duran C, Appleby N, Clark T, Wood D, Imelfort M, Batley J, Edwards D (2009) AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. Nucleic Acids Research 37: D951-D953

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnology Journal 8: 2-9

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. Evolution 62: 1162-1181

Enriquez S, Marba N, Duarte CM, van Tussenbroek BI, Reyes-Zavala G (2001) Effects of seagrass *Thalassia testudinum* on sediment redox. Marine Ecology-Progress Series 219: 149-158

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

Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. Nature Reviews Genetics 4: 651-657

Ferber S, Reusch TBH, Stam WT, Olsen JL (2008) Characterization of single nucleotide polymorphism markers for eelgrass (*Zostera marina*). Molecular Ecology Resources 8: 1429-1435

Fontaine, C., I. Dajoz, J. Meriguet, and M. Loreau. 2006. Functional Diversity of Plant-Pollinator Interaction Webs Enhances the Persistence of Plant Communities. PLoS Biology 4:e1.

Gagnon PS, Vadas RL, Burdick DB, May B (1980) Genetic Identity of Annual and Perennial Forms of *Zostera marina* L. Aquatic Botany 8: 157-162

Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Molecular Ecology 12: 1039-1047

Goransson P, Andersson S, Falkengren-Grerup U (2009) Genetic adaptation to soil acidification: experimental evidence from four grass species. Evolutionary Ecology 23: 963-978

Greeff, J. M., van Vuuren, G. J. J., Kryger, P. & Moore, J. C. (2009) Outbreeding and possibly inbreeding depression in a pollinating fig wasp with a mixed mating system. Heredity 102: 349-356

Hämmerli A, Reusch TBH (2002) Local adaptation and transplant dominance in genets of the marine clonal plant *Zostera marina*. Marine Ecology Progress Series 242: 111-118

Hämmerli A, Reusch TBH (2003a) Flexible mating: cross-pollination affects sex-expression in a marine clonal plant. Journal of Evolutionary Biology 16: 1096-1105

Hämmerli A, Reusch TBH (2003b) Inbreeding depression influences genet size distribution in a marine angiosperm. Molecular Ecology 12: 619-629

Hämmerli A, Reusch TBH (2003c) Genetic neighbourhood of clone structures in eelgrass meadows quantified by spatial autocorrelation of microsatellite markers. Heredity 91: 448-455

Hansson B (2010) The use (or misuse) of microsatellite allelic distances in the context of inbreeding and conservation genetics. Molecular Ecology 19: 1082-1090

Hansson, B. & Westerberg, L. 2002 On the correlation between heterozygosity and fitness in natural populations. Molecular Ecology 11: 2467-2474

Harley CDG, Hughes AR, Hultgren KM et al. (2006) The impacts of climate change in coastal marine systems. Ecology Letters 9, 228-241.

Harr B, Kauer M, Schlötterer C (2003) Hitchhiking mapping: A population-based fine-mapping strategy for adaptive mutations in Drosophila melanogaster (vol 99, pg 12949, 2002). Proceedings of the National Academy of Sciences USA 100: 3004-3004

Harris MA, Clark JI, Ireland A, Lomax J, Ashburner M, Collins R, et al. (2006) The Gene Ontology (GO) project in 2006. Nucleic Acids Research 34: D322-D326

Hendry AP, Nosil P, Rieseberg LH (2007) The speed of ecological speciation. Functional Ecology 21: 455-464

Hemminga M, Duarte C (2000) Seagrass Ecology. Cambridge University Press, Cambridge.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6: e1000862

Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setala H, Symstad AJ, Vandermeer J, Wardle DA (2005) Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecological Monographs 75: 3-35

Hughes AR, Stachowicz JJ (2009) Ecological impacts of genotypic diversity in the clonal seagrass *Zostera marina*. Ecology 90: 1412-1419

Hughes AR, Stachowicz JJ (2011) Seagrass genotypic diversity increases disturbance response via complementarity and dominance. Journal of Ecology 99: 445-453

Hyman ED (1988) A new method of sequencing DNA. Analytical Biochemistry 174: 423-436

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.

IPCC (2001) Climate Change 2001: The Scientific Basis - Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge.

Jacobs RPWM (1982) Reproductive strategies of two seagrass species (*Zostera marina* and *Z. noltii*) along west European coasts. IN: Studies on Aquatic vascular plants (eds Symoens JJ, Hooper SS, Compere P), pp. 150-155. Royal Botanic Society of Belgium, Brussels

Jensen J, Bohonak AJ, Kelley SK (2005) Isolation by Distance Web Service. BMC Genetics 6, 13.

Jones CG, Lawton JH, Shachak M (1994) Organisms as Ecosystem Engineers. Oikos 69: 373-386

Jump AS, Hunt JM, Martinez-Izquierdo JA, Penuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. Molecular Ecology 15, 3469-3480.

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.

Kauer MO, Dieringer D, Schlotterer C (2003) A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of *Drosophila melanogaster*. Genetics 165, 1137-1148.

Kayser M, Brauer S, Stoneking M (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. Molecular Biology and Evolution 20, 893-900.

Larsson LC, Laikre L, Palm S et al. (2007) Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. Molecular Ecology 16, 1135-1147.

Les DH, Cleland MA, Waycott M (1997) Phylogenetic studies in alismatidae, II: Evolution of marine angiosperms (seagrasses) and hydrophily. Systematic Botany 22: 443-463

Lewontin RC, Krakauer J (1973) Distribution of Gene Frequency as a Test of Theory of Selective Neutrality of Polymorphisms. Genetics 74, 175-195.

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, function, and evolution. Molecular Biology and Evolution 21, 991-1007.

Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nature Genetics 24: 381-386

Loreau M, Downing AL, Emmerson M, Gonzalez A, Hughes J, Inchausti P, Joshi J, Norberg J, and Sala OE (2002) A new look at the relationship between diversity and ecosystem functioning. *in* Loreau M, Naeem S, and Inchausti P (eds) Biodiversity and ecosystem functioning. Oxford University Press, New York.

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.

Luu D-T, Maurel C (2005) Aquaporins in a challenging environment: molecular gears for adjusting plant water status. Plant, Cell and Environment 28, 85-96.

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc.

Madsen JD, Chambers PA, James WF, Koch EW, Westlake DF (2001) The interaction between water movement, sediment dynamics and submersed macrophytes. Hydrobiologia 444: 71-84

Mäkinen HS, Cano JM, Merilä J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback *Gasterosteus aculeatus* populations. Molecular Ecology 17: 3565-3582

Marr AB, Keller LF, Arcese P (2002) Heterosis and outbreeding depression in descendants of natural immigrants to an inbred population of song sparrows (*Melospiza melodia*). Evolution 56: 131-142

Marshall T C, Spalton JA 2000 Simultaneous inbreeding and outbreeding depression in reintroduced Arabian oryx. Anim. Cons. 3: 241-248

Massa SI, Arnaud-Haond S, Pearson GA, Serrao EA (2009) Temperature tolerance and survival of intertidal populations of the seagrass *Zostera noltii* (Hornemann) in Southern Europe (Ria Formosa, Portugal). Hydrobiologia 619: 195-201

McMillan C, Phillips RC (1979) Differentiation In Habitat Response Among Populations Of New World Seagrasses. Aquatic Botany 7: 185-196

Mitton JB, Grant MC (1984) Associations among protein heterozygosity, growth rate, and developmental homeostasis. Annual Review of Ecology and Systematics 15: 479-499

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos 100, 403-405.

Murray MC, Hare MP (2006) A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. Molecular Ecology 15, 4229-4242

Nacci D, Coiro L, Champlin D, Jayaraman S, McKinney R, Gleason TR, Munns WR, Specker JL, Cooper KR (1999) Adaptations of wild populations of the estuarine fish *Fundulus heteroclitus* to persistent environmental contaminants. Marine Biology 134: 9-17

Namroud M-C, Jean Beaulieu, Nicolas Juge, Jerome Laroche, Jean Bousquet (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. Molecular Ecology 17: 3599-3613

Nielsen EE, Hansen MM, Meldrup D (2006) Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms. Molecular Ecology 15, 3219-3229.

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genetics 39: 197-218

Oetjen K, Reusch TBH (2007) Identification and characterization of 14 polymorphic EST-derived microsatellites in eelgrass (*Zostera marina*). Molecular Ecology Notes 7: 777-780

Oetjen K, Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. Molecular Ecology 16: 5156-5157

Oetjen K, Ferber S, Dankert I, Reusch TBH (2010) New evidence for habitat-specific selection in Wadden Sea *Zostera marina* populations revealed by genome scanning using SNP and microsatellite markers. Marine Biology 157: 81-89

Olsen JL, Stam WT, Coyer JA, Reusch TBH, Billingham M, Bostrom C, et al. (2004) North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. Molecular Ecology 13: 1923-1941

Orr HA (1998) The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution 52, 935-949.

Orellana MR, Lopez-Pujol J, Blanche C, Bosch M (2007) Relationships between heterozygosity and fitness in the Iberian threatened larkspur *Delphinium bolosii* (Ranunculaceae). Flora 202: 161-168

Orth RJ, Carruthers TJB, Dennison WC, Duarte CM, Fourqurean JW, Heck KL, et al. (2006) A global crisis for seagrass ecosystems. Bioscience 56: 987-996

Ouborg NJ, Vriezen WH (2007) An ecologist's guide to ecogenomics. Journal of Ecology 95: 8-16

Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST databases as a source for molecular markers: Lessons from Helianthus. Journal Of Heredity 97, 381-388.

Payseur BA, Cutter AD, Nachman MW (2002) Searching for Evidence of Positive Selection in the Human Genome Using Patterns of Microsatellite Variability. Molecular Biology and Evolution 19, 1143-1153.

Perry AL, Low PJ, Ellis JR, Reynolds JD (2005) Climate Change and Distribution Shifts in Marine Fishes. Science 308, (5730), 1912-1915.

Price, M. V. & Waser, N. M. 1979 Pollen Dispersal and Optimal Outcrossing in Delphinium nelsoni. Nature 277: 294-297

Procaccini G, Olsen JL, Reusch TBH (2007) Contribution of genetics and genomics to seagrass biology and conservation. Journal of Experimental Marine Biology And Ecology 350: 234-259

Pujol B, McKey D (2006) Size asymmetry in intraspecific competition and the density-dependence of inbreeding depression in a natural plant population: a case study in cassava (Manihot esculenta Crantz, Euphorbiaceae). Journal of Evolutionary Biology 19: 85-96

R development core team 2008 R: A language and Environment for Statistical Computing. R foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and eucumenicism. Journal of Heredity 86: 248-249.

Reise K (1985) Tidal Flat Ecology: An Experimental Approach to Species Interactions. Springer Verlag, Berlin.

Reise K, Herre E, Sturm M (1989) Historical Changes in the Benthos of the Wadden Sea around the Island of Sylt in the North Sea. Helgoländer Wissenschaftliche Meeresuntersuchungen 43: 417-433

Reise K, Jager Z, De Jong D, Van Katwijk M, Schanz A (2005) Seagrass. In Wadden Sea Quality Status Report 2004. Wadden Sea Ecosystem No. 19 (eds Essink K, Dettmann C, Farke H, Laursen K, Lüerßen G, Marenic H, Wiersinga W), pp. 201-207. Common Wadden Sea Secretariat, Wilhelmshaven, Germany.

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55: 1095-1103

Rengmark AH, Slettan A, Skaala O, Lie O, Lingaas F (2006) Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. Aquaculture 253: 229-237

Reusch TBH (2000) Five microsatellite loci in eelgrass *Zostera marina* and a test of cross-species amplification in *Z. noltii* and *Z. japonica*. Molecular Ecology Notes 9, 371-373.

Reusch TBH (2002) Microsatellites reveal high population connectivity in eelgrass (*Zostera marina*) in two contrasting coastal areas. Limnology and Oceanography 47, 78-85.

Reusch TBH (2006) Does disturbance enhance genotypic diversity in clonal organisms? A field test in the marine angiosperm Zostera marina. Molecular Ecology 15: 277-286

Reusch TBH, Bostrom C, Stam WT & Olsen JL (1999a) An ancient eelgrass clone in the Baltic. Marine Ecology Progress Series 183: 301-304

Reusch TBH, Stam WT, Olsen JL (1999b) Microsatellite loci in eelgrass Zostera marina reveal marked polymorphism within and among populations. Molecular Ecology 8, 317-321.

Reusch TBH, Stam WT, Olsen JL (2000) A microsatellite-based estimation of clonal diversity and population subdivision in Zostera marina, a marine flowering plant. Molecular Ecology 9: 127-140

Reusch TBH, 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.

Reusch TBH, Ehlers A, Hämmerli A, Worm B (2005) Ecosystem recovery after climatic extremes enhanced by genotypic diversity. Proceedings of the National Academy of Sciences USA 102: 2826-2831

Reusch TBH, Hughes AR (2006) The emerging role of genetic diversity for ecosystem functioning: Estuarine macrophytes as models. Estuaries and Coasts 29: 159-164

Reusch TBH, Veron AS, Preuss C, Weiner J, Wissler L, Beck A, et al. (2008) Comparative analysis of expressed sequence tag (EST) libraries in the seagrass Zostera marina subjected to temperature stress. Marine Biotechnology 10: 297-309

Richard M, Losdat S, Lecomte J, de Fraipont M, Clobert J (2009) Optimal level of inbreeding in the common lizard. Proceedings of the Royal Society-B Biological Sciences 276: 2779-2786

Robinson SP, Kennington WJ, Simmons LW (2009) No evidence for optimal fitness at intermediate levels of inbreeding in *Drosophila melanogaster*. Biological Journal of the Linnean Society 98: 501-510

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Totowa, NJ, Humana Press.

Robertson A (1975) Gene frequency distributions as a test of selective neutrality. Genetics 81, 775-785.

Ryynanen HJ, Tonteri A, Vasemagi A, Primmer CR (2007) A Comparison of Biallelic Markers and Microsatellites for the Estimation of Population and Conservation Genetic Parameters in Atlantic Salmon (*Salmo salar*). Journal of Heredity 98: 692-704

Schlötterer C (2002a) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics 160, 753-763.

Schlötterer C (2002b) Towards a molecular characterization of adaptation in local populations. Current Opinions in Genetics & Development 12: 683-687

Schlötterer C (2003) Hitchhiking mapping - functional genomics from the population genetics perspective. Trends in Genetics 19: 32-38

Schmidt PS, Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of MPI genotypes. Evolution 55, 1336-1344.

Scotti-Saintagne C, Mariette S, Porth I et al. (2004) Genome scanning for interspecific differentiation between two closely related oak species *Quercus robur* L. and Q. petraea (Matt.) Liebl. Genetics 168, 1615-1626.

Sharopova N (2008) Plant simple sequence repeats: distribution, variation, and effects on gene expression. Genome 51: 79-90

Sheng L (2003) Protein phosphatases in plants. Annual Review Of Plant Biology 54, 63-92.

Short FT, Wyllie-Escheverria S (1996) Natural and human-induced disturbance of seagrasses. Environmental Conservation 25, 17-27.

Silander JA (1979) Micro-Evolution and Clone Structure in *Spartina Patens*. Science 203: 658-660

Smith JM, Haigh J (1974) Hitch-Hiking Effect of a Favorable Gene. Genetical Research 23, 23-35.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100: 158-170

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology 14, 671-688.

Storz JF, Nachman MW (2003) Natural selection on protein polymorphism in the rodent genus Peromyscus: Evidence from interlocus contrasts. Evolution 57: 2628-2635

Szovenyi P, Ricca M, Shaw AJ (2009) Multiple paternity and sporophytic inbreeding depression in a dioicous moss species. Heredity 103: 394-403

Szulkin M, Bierne N, David P (2010) Heterozygosity-Fitness Correlations: A time for Reappraisal. Evolution 64: 1202-1217

Templeton AR 1986 Coadaptation and Outbreeding Depression. in Soule ME (ed.) Conservation Biology: The Science of Scarcity and Diversity. XIII: 105-116, Sinauer Associates, Sunderland, Mass.

Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? Genome Res 16: 702-712

Thornhill NW 1993 The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives. University of Chicago Press, Chicago

Thornton KR, Jensen JD (2007) Controlling the false-positive rate in multilocus genome scans for selection. Genetics 175, 737-750.

Tilman D, Reich PB, Knops JMH (2006) Biodiversity and ecosystem stability in a decade-long grassland experiment. Nature 441:629-632.

Tsitrone, A., Rousset, F. & David, P. 2001 Heterosis, marker mutational processes and population inbreeding history. Genetics 159: 1845-1859

Tsumura Y, Kado T, Takahashi T, Tani N, Ujino-Ihara T, Iwata H (2007) Genome scan to detect genetic structure and adaptive genes of natural populations of *Cryptomeria japonica*. Genetics 176: 2393-2403

Van Katwijk MM (2000a) Possibilities for restoration of *Zostera marina* beds in the Dutch Wadden Sea. PhD Thesis, University of Nijmegen, The Netherlands.

van Katwijk MM, Hermus DCR (2000b) Effects of water dynamics on *Zostera marina*: Transplantation experiments in the intertidal Dutch Wadden Sea. Marine Ecology-Progress Series 208: 107-118

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4, 535-538.

Vasemagi 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.

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Molecular Ecology 14, 3623-3642.

Vigouroux Y, McMullen M, Hittinger CT et al. (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. PNAS 99, 9650-9655.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of Variation Across Marker Loci as Evidence of Selection. Genetics 158, 1811-1823.

Vitalis R, Dawson K, Boursot P, Belkhir K (2003) DetSel 1.0: A computer program to detect markers responding to selection. Journal of Heredity 94, 429-431.

Voight BF, Kudaravalli S, Wen XQ, Pritchard JK (2006) A map of recent positive selection in the human genome (vol 4, pg 154, 2006). Plos Biology 4, 659-659.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen NP, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale

identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082

Waser NM, Price MV (1989) Optimal Outcrossing in *Ipomopsis aggregata* - Seed set and offspring fitness. Evolution 43: 1097-1109

Waycott M, Procaccini G, Les D, Reusch TBH (2006) Seagrass evolution, ecology and conservation: a genetic perspective. In: Seagrasses: Biology, Ecology and Conservation (eds Larkum AWD, Orth RJ, Duarte CM), pp. 25-50. Springer, Dordrecht, The Netherlands.

Waycott M, Duarte CM, Carruthers TJB, Orth RJ, Dennison WC, Olyarnik S, Calladine A, Fourqurean JW, Heck KL, Hughes AR, Kendrick GA, Kenworthy WJ, Short FT, Williams SL (2009) Accelerating loss of seagrasses across the globe threatens coastal ecosystems. Proceedings Of The National Academy Of Sciences USA 106: 12377-12381

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. Evolution 38, 1358-1370.

Wilding CS, Butlin RK, Grahame J (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. Journal of Evolutionary Biology 14, 611-619.

Willi, Y. & Van Buskirk, J. (2005) Genomic compatibility occurs over a wide range of parental genetic similarity in an outcrossing plant. Proceedings of the Royal Society- B Biological Sciences 272: 1333-1338

Wissler L, Dattolo E, Moore AD, Reusch TBH, Olsen JL, Migliaccio M, Bornberg-Bauer E, Procaccini G (2009) Dr. Zompo: an online data repository for *Zostera marina* and Posidonia oceanica ESTs. Database 2009: bap009

Woodhead M, Russell J, Squirrell J et al. (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. Molecular Ecology 14, 1681-1695.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics 175, 1883-1893.

Appendix

The Appendix contains supplementary material that has been published online:

Chapter II: **Table A-II.1, Table A-II.2**

Chapter III: **Figure A-III.1, Table A-III.1**

Table AII.1. Pairwise population differentiation in Wadden Sea eelgrass (*Zostera marina*). F_{ST}-values (above diagonal) and P-values (below diagonal, tested by 10 000 rar of all pairwise population comparisons. n.s. = not significant after Bonferroni correction, n.s.* = not significant after Bonferroni correction, but P < 0.01.
Estimates include allele frequencies of all 25 microsatellite loci. Negative FST-values were set to zero.

	H05F	H05C	H99F	H99C	L01C	L01F	L05C	L05F	K97C	K05C	K05F
H05F		0.0097	0.0099	0.0121	0.0249	0.0058	0.0122	0.0019	0.038	0.0181	0.0219
H05C	0.0055		0.011	0	0.0303	0.004	0.0246	0.0065	0.0297	0.0208	0.0313
H99F	0.0055	0.0055		0.0124	0.0331	0.0037	0.0162	0	0.0328	0.0211	0.0252
H99C	0.0055	n.s.	0.0055		0.0283	0.0069	0.0248	0.0082	0.0333	0.0258	0.0342
L01C	0.0055	0.0055	0.0055	0.0055		0.0231	0.0206	0.0266	0.034	0.0372	0.0353
L01F	n.s.	n.s.	n.s.	n.s.*	0.0055		0.017	0	0.0299	0.0147	0.0209
L05C	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055		0.0123	0.0415	0.0302	0.0266
L05F	n.s.	n.s.*	n.s.	0.0165	0.0055	n.s.	0.0055		0.0302	0.0121	0.0185
K97C	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055		0.0185	0.0288
K05C	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055		0.0069
K05F	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	0.0055	n.s.*	

Table AII.2 Eelgrass Zostera marina in the Wadden Sea. Summary statistics of all microsatellite loci.

Locus	Type	$F_{ST, global}$	*P* (F_{ST})	H_{exp}	*A*	Species of best BLASTX-hit, GenBank accession no. (only for gene-linked microsatellites)
ZMC19089*	gene linked	0.041678	0.0001	0.53	7	*Arabidopsis thaliana* hypothetical protein CAB81404
ZMA04093	gene linked	0.039933	0.0001	0.52	5	*Nicotiana tabacum* purple acid phosphatase BAC55157
GA1	anonymous	0.027401	0.0001	0.67	14	
CL679Contig1	gene linked	0.027395	0.0002	0.12	4	*Castanea sativa* defensin (protease inhibitor) AAL15885
GA35**	anonymous	0.024837	0.0001	0.9	31	
GA23	anonymous	0.023713	0.0001	0.7	11	
GA17D	anonymous	0.022104	0.0001	0.59	8	
ZMC02023	gene linked	0.022012	0.0001	0.46	9	*Zinnia elegans* lipid transfer protein TED4 BAA06462.
ZMC19066	gene linked	0.021188	0.0001	0.1	8	*Arabidopsis thaliana* acid phosphatase NP_194655
CL32Contig2	gene linked	0.019533	0.0001	0.38	8	*Arabidopsis thaliana* unknown protein AY143804
CL853Contig1	gene linked	0.019468	0.0001	0.66	8	*Nicotiana tabacum* aquaporin AAL33585
GA19	anonymous	0.01864	0.0027	0.54	7	
CL734Contig1	gene linked	0.018461	0.0001	0.72	10	*Oryza sativa* unknown protein NP_913907
GA16	anonymous	0.018159	0.0003	0.38	11	
ZMC13053	gene linked	0.018128	0.0001	0.8	12	*Arabidopsis thaliana* apospory associated protein NP_200543
CL412Contig1	gene linked	0.01767	0.0005	0.32	7	*Oryza sativa* putative nodulin 3 XP_465955
GA20	anonymous	0.017633	0.0002	0.44	9	
ZMC19017	gene linked	0.016756	0.0001	0.7	9	*Oryza sativa* bHLH-protein-like protein NP_910407
ZMC01058	gene linked	0.015618	0.0001	0.52	9	*Gossypium hirsutum* vacuolar H+-ATPase AAA82977
GA4	anonymous	0.011592	0.0145	0.44	7	
ZMC12075	gene linked	0.009269	0.008	0.69	12	*Fritillaria agrestis* cytochrome c O22642
GA12	anonymous	0.006406	0.0455	0.3	7	
GA17H	anonymous	0.005936	0.0009	0.9	20	
GA2	anonymous	0.004126	0.0599	0.76	18	
ZMC06073	gene linked	-0.000054	0.448	0.13	4	*Arabidopsis thaliana* nucleotide sugar epimerase NP_194773
total		0.018755	0.0001		255	

A = number of alleles; genome scan *P*-values: * *P*=0.01, ** *P*=0.02 (*P*-values of significant outliers detected by the test by Beaumont & Nichols (1996))

Figure A-III.1 Map of sampling sites at Hallig Hooge, Hallig Langeness and Sylt island, North Frisian Wadden Sea, Schleswig-Holstein, Germany. Population pairs comprise a habitat contrast formed by one subtidal (tidal creek) and one intertidal (tidal flat) site. Distances between sampling sites: Hooge intertidal- Hooge subtidal 1.35 km, Langeness intertidal- Langeness subtidal 4.02 km, Sylt subtidal- Sylt intertidal 2.66 km. Maximum distance between Sylt subtidal and Hooge intertidal, 53.13 km.

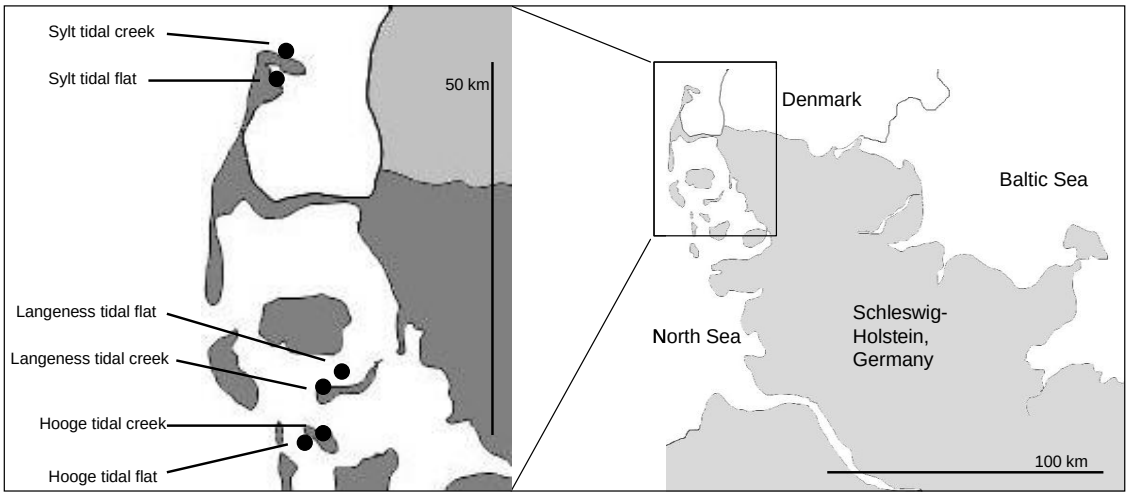


Table A-III.1 Complete list of all 46 genetic markers with their respective BLAST hits if available. 29 EST-derived microsatellites, of which 15 were newly developed for this study, 11 anonymous microsatellites and 6 SNPs (Ferber et al. 2008) were used.

Marker name	Acc. No.	new in genome scan	type	Blast hit (acc. No)
CL766Contig1	FN435336	yes	EST-Msat	Putative ddt domain-containing protein Q6zi78
CL11Contig1	FN435337	yes	EST-Msat	Metallothionein-like protein Q9au16
CL559Contig1	FN435338	yes	EST-Msat	Putative 24 kda seed maturation protein Q8s2k0
ZME02125	FN435339	yes	EST-Msat	Putative myb-related protein Q9aut3
ZMF02381	FN435340	yes	EST-Msat	Receptor-like kinase Q9lsu7
CL202Contig1	FN435341	yes	EST-Msat	Universal stress protein family Q2qnv2
CL380Contig1	FN435342	yes	EST-Msat	Nucleotide sugar epimerase-like protein Q9m0b6
CL805Contig1	FN435343	yes	EST-Msat	Hypothetical protein (putative acyl carrier protein-like) Q3hrx1
CL172Contig1	FN435344	yes	EST-Msat	Ubiquitin conjugating enzyme Q43821
CL53Contig1	FN435345	yes	EST-Msat	Btf3b-like transcription factor Q6dq93
ZME06302	FN435346	yes	EST-Msat	Holocarboxylase synthetase (fragment) Q7x9l1
ZMC05062	FN435347	yes	EST-Msat	Nadp-dependent malic enzyme 2 Q6pmi2
ZMC19062	FN435348	yes	EST-Msat	High mobility group protein 2 Q40094
ZME05315	FN435349	yes	EST-Msat	Putative ubiquitin-conjugating enzyme 16 Q9fwt2
ZME02369	FN435350	yes	EST-Msat	Putative serine carboxylase 2 Q5smv5
287CT623*		yes	SNP	DNA-binding protein MNB1B
71CA329*		yes	SNP	no hit
20CT465*		yes	SNP	Early response to dehydration
71TC178*		yes	SNP	no hit
95GT380*		yes	SNP	Anthranilate N-benzoyltransferase
98AT518*		yes	SNP	Pathogenesis-related protein PR-4A precursor
ZMC19066	AM408830	no	EST-Msat	Acid phosphatase NP_194655
CL32Contig2	AM408831	no	EST-Msat	Unknown protein AY143804
ZMC19017	AM408843	no	EST-Msat	BHLH-protein-like protein NP_910407
ZMA04093	AM408832	no	EST-Msat	Purple acid phosphatase BAC55157
CL853Contig1	AM408833	no	EST-Msat	Aquaporin AAL33585
CL679Contig1	AM408834	no	EST-Msat	Defensin (protease inhibitor) AAL15885
ZMC06073	AM408835	no	EST-Msat	Nucleotide sugar epimerase NP_194773
ZMC02023	AM408836	no	EST-Msat	Lipid transfer protein TED4 BAA06462.
ZMC19089	AM408837	no	EST-Msat	Hypothetical protein CAB81404
ZMC12075	AM408838	no	EST-Msat	Cytochrome c O22642
CL734Contig1	AM408839	no	EST-Msat	Unknown protein NP_913907
ZMC13053	AM408840	no	EST-Msat	Apaspory associated protein NP_200543
CL412Contig1	AM408841	no	EST-Msat	Putative nodulin 3 XP_465955
ZMC01058	AM408842	no	EST-Msat	Vacuolar H+-ATPase AAA82977
GA2	AJ249303-AJ249307, AJ009898, AJ009901, AJ009904	no	anonymous Msat	
GA35		no	anonymous Msat	
GA23		no	anonymous Msat	
GA17H		no	anonymous Msat	
GA12		no	anonymous Msat	
GA20		no	anonymous Msat	
GA17D		no	anonymous Msat	
GA19		no	anonymous Msat	
GA16		no	anonymous Msat	
GA4		no	anonymous Msat	
GA1		no	anonymous Msat	

*described in Ferber et al. 2008

Curriculum Vitae

Name	Katharina Keil, geb. Oetjen
Staatsangehörigkeit	deutsch
Geboren	15.09.1978 in Bremen

Ausbildung

2006 – 2011	Promotion am IFM-Geomar, Kiel über Genome Scans und Heterozygotie-Fitness-Korrelationen beim großen Seegras *Zostera marina*
12. 06. 2006	Diplomprüfungen in Ökologie, Zoologie, Landschaftsentwicklung, Universität Kiel (Note ‚mit Auszeichnung')
2002 – 2006	Hauptstudium Biologie, Universität Kiel
2002	Vordiplom (Note ‚gut')
2000 – 2002	Grundstudium Biologie, Universität Kiel
1998 – 2000	Studium der Medizin, Universität Kiel
1998	Allgemeine Hochschulreife, Kippenberg-Gymnasium Bremen

Publikationen

Oetjen K, Ferber S, Dankert I, Reusch TBH (2010) New evidence for habitat-specific selection in Wadden Sea *Zostera marina* populations revealed by genome scanning using SNP and microsatellite markers. Marine Biology: Volume 157, Issue 1, Page 81

Oetjen K, Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. Molecular Ecology, 16, 5156-5157.

Oetjen K, Reusch TBH (2007) Identification and characterization of 14 polymorphic EST-derived microsatellites in eelgrass (*Zostera marina*). Molecular Ecology Notes, 7, 777-780.

Description of the individual scientific contribution to multiple-author publications

The chapters of this thesis are published (chapters I, II and III) or submitted (chapter IV) to scientific journals with multiple authorship. This list serves as a clarification of my personal contributions on each publication.

Chapter I:

Identification and characterization of 14 polymorphic EST-derived microsatellites in eelgrass (*Zostera marina*)
Authors: Katharina Oetjen (= Katharina Keil), Thorsten B.H. Reusch
Published in Molecular Ecology Resources 7, 777-780 (2007)
Contributions: KK and TR developed the ideas for this study, KK conducted the experimental work, KK conducted data analyses; KK and TR discussed the results, KK wrote the manuscript.

Chapter II:

Genome Scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*
Authors: Katharina Oetjen (Keil), Thorsten B.H. Reusch
Published in Molecular Ecology 16, 5156-5167 (2007)
Contributions: KK and TR discussed the ideas for this study, KK conducted the study, KK conducted data analyses, KK and TR discussed the results, KK wrote the manuscript.

Chapter III:

New evidence for habitat-specific selection in Wadden Sea *Zostera marina* populations revealed by genome scanning using SNP and microsatellite markers
Authors: Katharina Oetjen (Keil), Steven Ferber, Ilka Dankert, Thorsten B.H. Reusch
Published in Marine Biology 157, 81-89 (2010)
Contributions: KK and TR developed the ideas for this study, SF, ID and KK conducted the

laboratory work, KK conducted data analyses; KK and TR discussed results, KK wrote the manuscript.

Chapter IV:

Intermediate heterozygosity is associated with optimal fitness in the clonal plant *Zostera marina*

Authors: Katharina Oetjen (Keil), August Hämmerli, Thorsten B.H. Reusch

Submitted to Biology Letters

Contributions: KK and TR developed the ideas for this study, KK conducted the study, KK and AH conducted data analyses; KK, AH and TR discussed data analysis and results, KK wrote the manuscript.

Erklärung

Hiermit versichere ich, dass diese Dissertation, abgesehen von der Beratung durch meinen Betreuer, selbständig von mir angefertigt wurde und dass sie nach Inhalt und Form meine eigene Arbeit ist. Ich habe keine als die angegebenen Hilfsmittel und Quellen verwendet und die Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft angefertigt. Die Arbeit wurde keiner anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegt. Dies ist mein einziges und bisher erstes Promotionsverfahren. Teile dieser Arbeit wurden als Manuskripte bei Zeitschriften veröffentlicht (Kapitel I in *Molecular Ecology Notes* mit Thorsten Reusch als Koautor, Kapitel II in *Molecular Ecology* mit Thorsten Reusch als Koautor, Kapitel III in *Marine Biology* mit Steven Ferber, Ilka Dankert und Thorsten Reusch als Koautoren) oder eingereicht (Kapitel IV bei *Biology Letters* mit August Hämmerli und Thorsten Reusch als Koautoren).

Kiel, den 05. Mai 2011 Katharina Keil