The effect of climate change on marine brachyurans and the implication of shell disease

Von der Fakultät für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg zur Erlangung
des Grades und Titels eines

Doktors der Naturwissenschaften Dr. rer. nat.

angenommene Dissertation

von Herrn Philipp Krämer

geboren am 02.01.1982 in Salzkotten

1. Gutachterin:	Frau Prof. Gabriele Gerlach
2. Gutachter:	Herr Prof. Meinhard Simon
3. Gutachter:	Herr Prof. Dirk Albach

Tag der Disputation: 23.09.2015

Contents

Summary

In the marine ecosystem climate change is a major threat to crustacean biodiversity. Altered environmental conditions will physiologically challenge eukaryotes and prokaryotes. The microbiotic origin of crustacean shell disease and its dependence on the interaction of environmental factors with host susceptibility and local pathogens serves as a good model studying the response and potential resilience of crustaceans towards environmental alterations by climate change. Despite the broad distribution and long knowledge of shell disease as a syndrome, detailed current information are scarce for the North Atlantic. In my thesis I contribute to understand the effect of shell disease on the level of inter- as well as intra-specific interactions by using state of the art genetic methods combined with ecological data. I focus on the brown crab *Cancer pagurus* Linnaeus, 1758 as a species of high commercial value and ecological importance with generally high susceptibility to environmental factors and shell disease.

In chapter I, I study on the interaction of *C. pagurus* and its bacterial biofilm. I analyze whether temperature alone may alter bacterial communities in biofilms and whether long term differences in environmental factors between different populations may play a role in shell disease susceptibility and bacterial biofilm composition. I figure out that both temperature and population specific conditions may have an effect on bacterial biofilms. However, bacterial biofilms are relatively specific for individuals and stable for short term temperature alterations. I present the first broad overview of the bacterial community structure in *C. pagurus* biofilms. I detect many bacterial strains similar to those playing a role in shell disease inflammation in the American lobster *Homarus americanus* Milne Edwards, 1837. This shows the large geographic impact of those bacteria. The biofilm communities includes obligate anoxic and aerobic bacteria. Anoxic conditions may arise in biofilm communities in micro-habitats and may serve obligate anoxic pathogens as well as chitinolytic bacteria as starting point for lesion formation.

To test for the prevalence of shell disease in other crustaceans I compare shell disease in three decapod crabs of the North Sea. I discover differences between species which might have an effect on community structure in local marine benthic communities. A recent invaded species *Hemigrapsus penicillatus* (De Haan, 1835) seems to be more resistant to shell disease syndromes and may thus benefit in inter-specific competition with native crabs such as *Carcinus maenas* (Linnaeus, 1758) or *C. pagurus*. Body size of the host is a major factor influencing shell disease prevalence. This may lead to negative selective pressure for, not only large species, but also bigger sized individuals within species.

To study shell disease on the intra-specific level, population genetics is used as a tool to search for genetic signatures of adaptation on the distributional range of *C. pagurus*. In chapter II, I present two different approaches to understand local adaptation based on population genetics. I describe how population genetics data may be used to analyze relatedness structures in local populations. This is necessary to differentiate between local adaptation of populations and physiological similarities merely due to close relatedness of con-specifics. I used the R programming language to implement different statistics for this purpose and published the code in the R Network.

Furthermore, population genetics can be used to understand structures among populations.

I contribute to the recent discussion, which estimate, $\hat{G}_{ST}$ or D_{EST}, best describes population structure by modeling migration and mutation in an island population situation. Both estimators produce similar results in statistics, i.e. significance is congruent for both estimates. While $\hat{G}_{st}$ is heavily influenced by mutation, D_{EST} calculations are more robust in this situation.

In chapter III, I use a population genetics approach to analyze the genetic profile of different populations of *C. pagurus* on a latitudinal gradient from Norway to Spain. Variable sites in annotated genes of different populations were identified by 454 full transcriptomics. I find low differentiation among populations by neutral simple sequence repeat (SSR) markers implying high connectivity between different areas. Similarly, genetic markers derived from different expressed sequence tag (EST) sequences show weak structuring between populations. To find genes, which may be source for adaptation, two populations at environmental extreme positions are compared. Individuals in northern Norway are experimentally stressed with temperature conditions equivalent to the southern North Sea. Similarity in relative gene expression between populations suggests that similar high temperature conditions will be tolerated. Severity of shell disease is not increased by the treatment. Decreased transcripts during warm conditions are immune or metabolism related such as hemocyanines or gamma-butyrobetaine. Those genes may be important for acclimation over longer periods. Gene expression and population genetics showed a high degree of similarity implying a common gene pool among populations. High migration distance by larvae and adults connect populations over long distances and will help *C. pagurus* to persist during climate change induced environmental changes. Changes in population size during the last years especially at the edges of their distribution indicate that this species arlready underwent a northward shift. This may obscure former population structures. Significant differentiation between shell diseased and non shell diseased individuals are found in Northern Norway. A high degree of population admixture in shell diseased individuals in contrast to healthy individuals might suggest that diseased animals may originate from different populations. This indicates a subgroup of migrants in Northern Norway. The high degree of connectivity may be beneficial for migrants to invade into northern habitats. Being better adapted to more southern populations, they may more often develop shell disease in newly invaded habitats. This is also supported by two transcripts belonging to the hemocyanines, which show evidence for differences in expression between populations in contrast to ≈13,000 other transcripts studied. These transcripts will be promising to study to see if high connectivity and potential to shift already skews locally adapted populations in the wild and will further change the population structure of *C. pagurus* with climate change induced altered temperature conditions in the future.

Zusammenfassung

Die durch den Klimawandel verursachten Veränderungen von Umweltfaktoren im Ökosystem Meer bedrohen die Vielfalt der Krebse und wirken sich auf die Physiologie von Eukaryoten und Bakterien aus. Solche Veränderungen von Umweltbedingungen stehen in einigen Fällen in einem direkten Zusammenhang mit dem Auftreten der Schwarzfleckkrankheit. In meiner Doktorarbeit untersuche ich das Auftreten der Schwarzfleckkrankheit in lokalen Krebspopulationen, sowie die Konsequenzen auf ihre zukünftige Verbreitung. Der mikrobielle Ursprung dieser Krankheit und das übergreifende Auftreten bei verschiedenen Krebsen bieten ein gutes Modell, um die Reaktion und die potentielle Widerstandsfähigkeit des Ökosystems am Beispiel der Krebse auf zukünftige Klimaveränderungen abschätzen zu können. Obwohl die Schwarzfleckkrankheit weit verbreitet und lange bekannt ist, gibt es wenig detaillierte aktuelle Informationen für den Nordatlantischen Raum. In dieser Arbeit werden genetische Methoden genutzt, um zu einem besseren Ursachenverständnis der Schwarzfleckkrankheit sowie deren Auswirkung auf zwischen und innerartliche populationsübergreifende Interaktionen zu kommen. Exemplarisch wird dabei der Taschenkrebs *Cancer pagurus* Linnaeus, 1758 betrachtet, eine Art mit hohem kommerziellen Wert, ökologischer Relevanz und gleichzeitig starker Sensibilität für Temperaturveränderungen.

In Kapitel I werden Bakterien im Biofilm des Karapaxes und dessen Interaktion mit Umweltfaktoren betrachtet. Ich finde zahlreiche bakterielle Taxa, die eine Rolle bei unterschiedlichen Formen der Schwarzfleckkrankheit spielen können. Einige darunter mit großer Ähnlichkeit zu Bakterien der epizootischen Form der Schwarzfleckkrankheit. Diese ist bisher nur für den Amerikanischen Hummer *Homarus americanus* Milne Edwards, 1837 beschrieben und hat in Nordamerika zu verheerenden Auswirkungen geführt. Für kurzfristige Temperaturveränderungen zeigt sich die Biofilmgemeinschaft insgesamt relativ stabil. Bei erhöhten Temperaturbedingungen treten jedoch vermehrt anaerobe Bakterien auf. Begünstigt durch stark strukturierte Mikrohabitate auf dem Karapax, entstehen anoxische neben oxischen Bedingungen, so dass obligat anaeroben Pathogenen aber auch anaerob chitinolytischen Taxa ein Lebensraum geboten wird. Diese Morphologie der Außenhaut könnte zu einer verstärkten Anfälligkeitfür die Schwarzfleckkrankheit führen, vor allem in wärmeren Gewässern.

Betrachtet man das Auftreten der Schwarzfleckkrankheit bei verschiedenen Kurzschwanzkrebsen der Nordsee, zeigt sich, dass die Krankheit bei verschiedenen Arten in ähnlichen Habitaten unterschiedlich stark auftritt. Das lässt einen starken Einfluss auf deren interspezifische Konkurrenz vermuten. Im Gegensatz zu den beiden heimischen Arten, der Strandkrabbe *Carcinus maenas* (Linnaeus, 1758) und dem Taschenkrebs *C. pagurus*, ist bei der in jüngerer Vergangenheit eingewanderten Asiatischen Strandkrabbe *Hemigrapsus penicillatus* (De Haan, 1835) kein Krankheitsausbruch festzustellen. Auffällig ist, dass die Außenhaut der Asiatischen Strandkrabbe so gut wie keine Strukturen aufweist. Eine weitere Rolle scheint die Körpergröße zu spielen. Innerhalb einer Art sind besonders die großen Tiere betroffen, so dass die Gefahr einer negativen Selektion von großen Größenklassen innerhalb einer Populationen besteht.

Um den Effekt der Schwarzfleckkrankheit auf eine Art abschätzen zu können, wende ich populationsgenetische Methoden an. Im Kapitel II beschreibe ich, wie die Populationsgenetik für die Feststellung von Verwandschaftsverbänden genutzt werden kann. Das ist entscheidend, um fest-

stellen zu können, ob ähnliche genetische Ausprägungen innerhalb einer Population unabhängig von lokalen Familienverbänden bestehen. Die Methoden sind mit der Programmiersprache R implementiert und im R-Netzwerk publiziert. Im Folgenden geht es um populationsgenetische Ansätze, um Unterschiede zwischen Populationen zu untersuchen. Ich gehe der Frage nach, welche statistischen Schätzer für eine Beschreibung der Struktur zwischen Populationen geeignet sind. Operationalisiert wird dies, indem eine Inselpopulation unter verschiedenen Mutations- und Migrationsraten mit Hilfe der gängigen Schätzer $\hat{G}_{ST}$ and D_{EST} analysiert wird. Es zeigt sich, dass die Signfikanzen unter gleichen Bedingungen bei beiden Schätzern gleich sind. $\hat{G}_{ST}$ ist im Gegensatz zu D_{EST} stark durch Mutation in seiner maximal darstellbaren Höhe beeinflusst. Zusätzlich entwickelten und publizierten wir eines der ersten Programme als Paket innerhalb des R-Netzwerkes mit dem D_{EST} Schätzer mit Hilfe eines MMC Algorhythmus berechnet werden können.

In Kapitel III werden die Erkenntnisse zum Einfluss der Schwarzfleckkrankheit sowie die populationsgenetischen Ansätze kombiniert, um die Populationsstruktur von *C. pagurus* innerhalb Europas und im Kontext des Klimawandels und Temperaturänderungen zu untersuchen. Um variable Stellen im Transkriptom bei *C. pagurus* zu finden und Rückschlüsse auf die Anpassungsfähigkeit dieser Art zu ziehen, wird die 454 Next Generation Sequenzierung genutzt. Auf einem latitudinalen Gradienten der Populationen von Norwegen bis Spanien zeigen sich nur schwache genetische Unterschiede über weite Distanzen. Dieses Bild ist bei der Analyse von neutralen Markern sowie bei Markern im Transkriptom konsistent. Um Gene zu identifizieren die möglicherweise unter Selektionsdruck stehen, untersuche ich die Transkriptome zweier Populationen mit extrem unterschiedlichen Temperaturbedingungen. Die relative Expression von Transkripten in Norwegen bei einer experimentellen Anpassung an südlichere Bedingungen stellen eine erfolgreiche Akklimatisierung dar. Eine niedrige Expression unter warmen Bedingungen fand sich in Genen für Immunfunktionen und Metabolismus wie Hämozyanin oder Gamma-Butyrobetaine, deren Funktion für eine Akklimatisierung an wärmere Bedingungen von Bedeutung ist. Es zeigt sich keine erhöhte Schwarzfleckkrankheit während der Experimente. Die gute Anpassungsfähigkeit und die starke Konnektivität zwischen Populationen lässt vermuten, dass ein hoher genetischer Austausch über weite geographische Distanzen und ein nur geringer Grad an lokaler Anpassung stattfinden. Dies wird vermutlich getragen durch einen starken Larventransport und Wanderungen der adulten Tiere. Einige Hinweise für lokale Anpassung waren jedoch zu finden: Beim Vergleich kranker und gesunder Individuen fanden sich unerwartet hohe Unterschiede innerhalb der norwegischen Population. Die genetische Struktur innerhalb der kranken Sub-population spricht für eine gemixte Population aus verschiedenen Regionen. Die gesunden Tiere waren genetisch homogener, so dass es sich um ortstreue besser angepasste Individuen handeln könnte. Einen Hinweis, dass unter geänderten Temperaturen lokale Anpassungen zum Vorschein kommen können, stellt die Expression zweier zu den Hämozayninen gehörigen Transkripte dar. Entgegen der Erwartung änderte sich die Epxression nicht auf ein ähnliches Level im Vergleich zu Individuen aus wärmeren Populationen. Diese Transkripte sind für zukünftige Studien von besonderem Interesse. Sie bieten die Möglichkeit, lokale Anpassungen innerhalb von Populationen zu untersuchen und festzustellen, ob starke Populationsverbindungen und eine möglicherweise bereits vorhandene Bewegung von Individuen in den Norden auch in Zukunft die Populationsstruktur von *C. pagurus* ändert und sich ürsprüngliche lokale Strukturen mit potentiell angepassten Individuen vermischen.

Aims of the thesis

In my thesis I worked on the effect of anthropogenic influences in particular climate change mediated temperature changes on populations of the European brown crab *Cancer pagurus*. I used the commonly known shell disease as a syndrome to analyze the effect of environmental factors on the host. The term shell disease describes several forms of microorganism induced erosive ulcerations on the outer carapace of crustaceans. While the reasons for shell disease susceptibility are not fully understood, a multi-factorial interaction of a host and probably several pathogens in an environmental framework are most likely (Fig. 1). The model was originally developed to understand a special form of crustacean shell disease, the epizootic shell disease in the American lobster *Homarus americanus* (1). This form of the disease appears to be lethal, was recorded in 1997 (2) and since then heavily spread in some populations of North America (1).

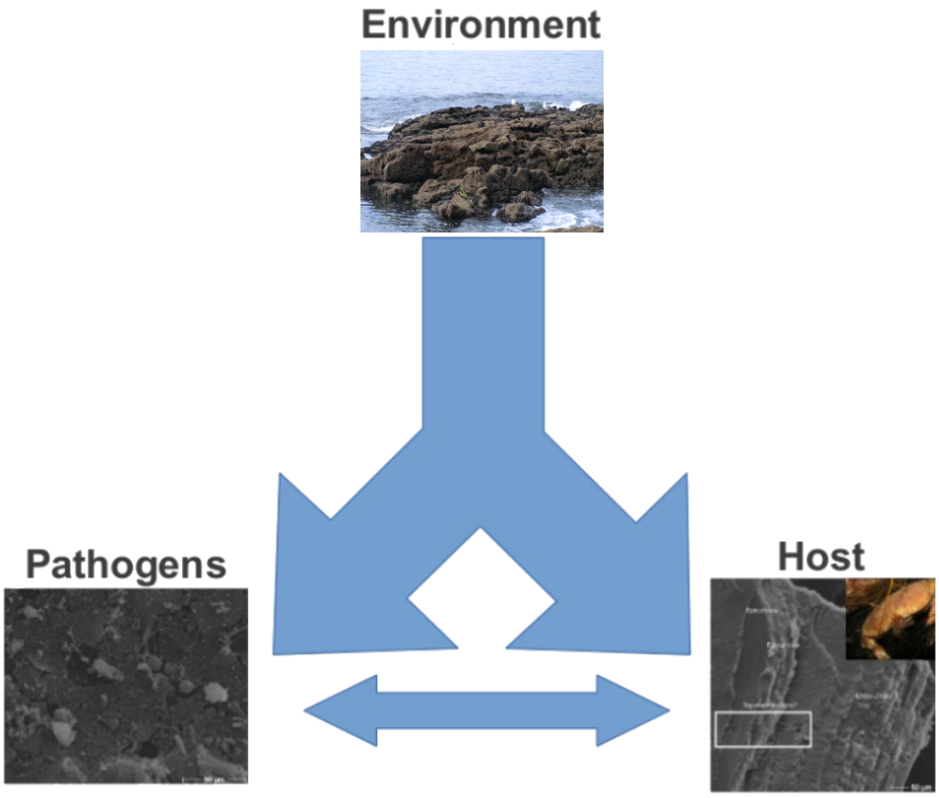


Figure 1: The model for shell disease shows the interaction of the pathogens and the host triggered by environmental conditions, which leads to the occurrence of shell disease in crustaceans (1).

In contrast to common forms of shell disease often called black spot disease, the epizootic form usually covers bigger areas of the carapace leading to vast erosion of the outer shell. The occurrence of epizootic shell disease led to massive ecological and economical consequences and is now intensively studied for the American lobster (1; 3; 4). It is still not known, if this form is caused by the normal epibiotic biofilm or if it is initiated by special pathogens (5; 6; 7). Analogous to the model of Castro and Sommer (1) and an earlier model of Tlusty et al. (8) I contribute to the understanding of the role of the interaction of all three factors - the host, the pathogen and the environment - by studying the effect of shell disease emergence on European brachyuran biodiversity.

Aims of the study:

1. Understanding the effects of shell disease on the European brown crab *Cancer pagurus* in local habitats, with regard to host pathogen interactions triggered by climate change (Manuscript I) and on inter-specific interactions as an effect of specific disease susceptibilities in brachyuran crabs (Manuscript II).

2. Providing tools for analyzing local population structure (Manuscript III and Manuscript IV) with emphasis to local adaptation and the implementation and understanding of recently developed estimates for population differentiation (Manuscript IV).

3. Analyzing the population structure of *Cancer pagurus* among European populations with regard to climate change related adaptation and disease susceptibility in local populations compared with neutral genetic variation of populations (Manuscript V).

Chapter 1

Host bacteria interactions in marine decapods - the effect of temperature and its ecological consequences

1.1 Introduction

1.1.1 Shell disease syndrome in crustaceans

Since the early 20^{th} century the syndrome of shell disease has been described in many wild crustacean populations (9). Examples for the disease are known for both major classes of crustacean diversity i.e. for malacostracan crustaceans such as *Homarus gammarus* (10), *Cancer pagurus* (11), *Scylla serrata* (Forskål, 1775) (12) or different crangonids (13) and maxillopod crustaceans such as copepods (14). The reports of shell disease in the wild are, however, heavily biased towards species of commercial interest, for instance big decapods for harvesting purposes (15) and sea lice (14). The visual appearance of black spots on the outer surface of crustaceans arises from the immune defense of the individual against penetration of the outer cuticle and the invading agents by shell blackening, i.e. melanization (16). Therefore it is more appropriate to discuss shell disease as a syndrome rather than a disease.

1.1.2 Bacterial biofilms as a source of shell disease

The broad taxonomic appearance of shell disease among crustacean taxa may bear from the similar structure of the cuticle as defense against external factors. The cuticle acts as an exoskeleton and as a protective barrier against physical injuries from con-specifics, predators, pathogens or abiotic environmental factors (17; 18). The cuticle in decapods consists at the outside of a thin epicuticle, followed by a thick exocuticle and a massive endocuticle at the inner surface (Fig. 1.1 (18)). Each of these layers has different properties for the organism. The outermost layer, the epicuticle is composed of lipids, proteins and calcium salts but lacks any chitin components (19; 20). The exocuticle in decapods is often a massive layer, which is responsible for pigmentation processes by melanin deposition (20). The endocuticle is the innermost layer and has a crucial role during the molting process (20). The absolute thickness of each layer varies between different species, different ontological stages and even between different regions of the cuticle of the same individual (20). Furthermore, the cuticle of crustaceans exhibits many structures connecting the inner hypodermis through all layers of the cuticle with the outermost layer, the epicuticle. These tegumental glands may vary in structure, function and density of appearance of the cuticle (20). Their function is discussed; besides its importance in the molting (21) and spawning cycle (22; 20), authors found that the secretion of the epicuticle (23; 20) and the tanning by phenols (24) and phenol oxidase (25; 26) may be driven by tegumental glands.

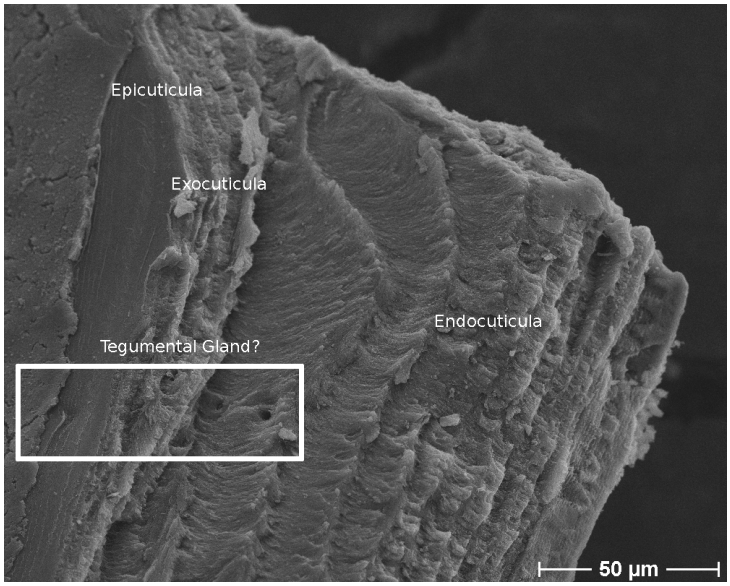


Figure 1.1: Scanning electron microscopy of a cross section through the outer carapace of *C. pagurus*. The thin epicuticle, a thicker exocuticle and the massive endocuticle are visible as different layers of the carapace. A channel through these layers is visible probably as a part of a tegumental gland.

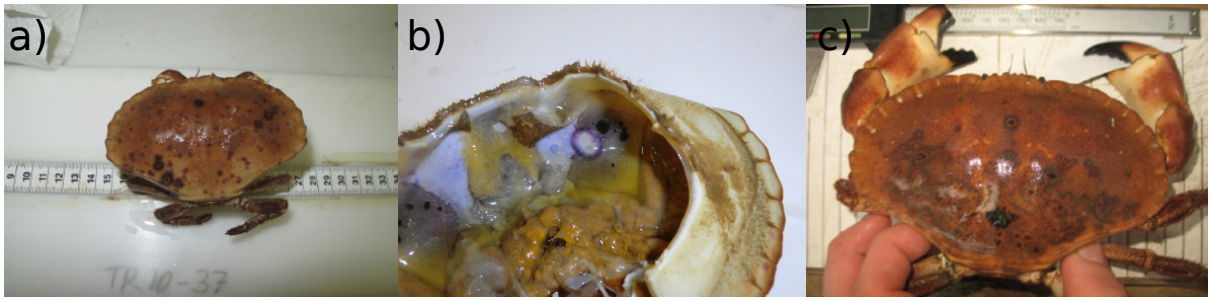


Figure 1.2: Shell disease occurrence in different individuals of *C. pagurus* in different stages of severity. (a) An individual with many small areas of outer shell disease. Shell disease can progress through the layers blackening soft tissues with often purple halos of melanization (b). Shell disease progressing through inner layers of the carapace with massive melanization all over the carapace (c).

Phenolic compounds (27) and phenol oxidase (28) may have antibacterial properties. Talbot and Zhao (21) speculate that in lobsters tegumental glands may have at least bacteriostatic properties.

The biofilm covering the carapace of healthy marine invertebrates consists of a pronounced community of different bacteria (29; 30) which interact with each other in numerous ways (31); for instance, several chemicals are released from biofilm prokaryots to prevent biofouling of alien species (32). These complex bacterial communities are constantly exposed to changing abiotic criteria such as temperature or pollution, which may alter their community structure (33). Biofilms on crustacean species are often also populated by pathogenic and chitinolytic bacteria (9). In the course of shell disease the outer shell of stressed hosts erodes by biofilm bacteria resulting in brown or black lesions of the carapace (34). In late stages, some bacteria may cause internal infections, eventually leading to the death of the host.

Despite discussing shell disease as a syndrome rather than a disease, authors classified different forms of shell disease (35; 12). Classification may arise from different etiologies or causative agents. Commonly known are the brown spot disease, enzootic disease, epizootic disease, rust spot disease and the cigarette burn syndrome, each of them more or less derived from the morphology of the disease spots (36; 37; 38). Furthermore, shell disease may be classified as, for example, diet induced shell disease (39) in cases for which the reason for the disease seems to be rather obvious. Intensive research was undertaken to identify specific bacterial strains, which may be the trigger for certain types of shell disease. As a result the list of known bacterial species within crustacean biofilms is relatively long and includes potential harmful *Vibrio*, *Flaveobacter*, *Pseudomonas* or *Pseudoalteromonas* (40). Recently, many forms of these different classifications had been closer analyzed with regard to their bacterial biofilm in lesions. Bacterial communities still overlapped in their composition, however, some more or less specific community compositions could be determined for different forms (41; 6). In trauma induced shell disease (TSD) remarkably high abundances of *Vibrio* and *Pseudoalteromonas* were found compared to other forms of shell disease (42). Enzootic shell disease (EnSD), characterized by irregular shaped orange lesions was dominated by *Tenacibaculum* sp. and *T. ovolyticum* (41). In diet induced shell disease lesions *Aquimarina homaria* and *Candidatus Kopriimonas aquarianus* were found among others (7). In epizootic shell disease (ESD), *Aquimarina homaria*, strains of *Pseudoalteromonas*, *Candidatus Kopriimonas aquarianus* as well as *Thalassobius sp. 131.I* are described (6). However, many bacterial taxa are common to different forms of lesions (43). Although, pinpointing down specific bacteria responsible for shell disease seemed to be promising it was already stated in 1970 that the shell disease syndrome may be the effect of several bacteria taxa interacting, rather than a specific strain (44). This was also stated in a

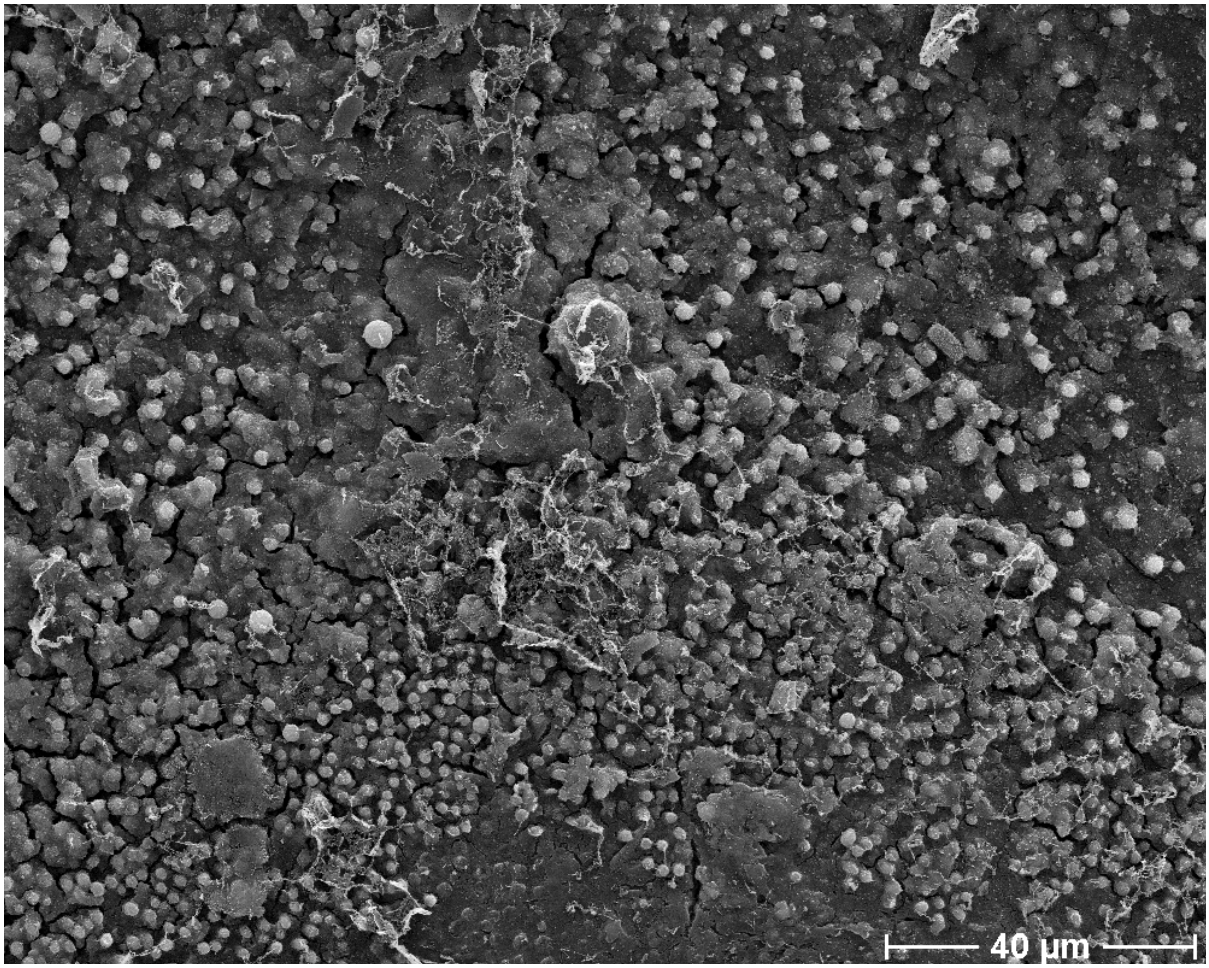


Figure 1.3: Scanning electron microscopy of the surface of *C. pagurus* outer carapace showing a dense biofilm with several coccoid bacteria on the surface.

more recent study for the American lobster *H. americanus* (5). The authors could not determine specific bacterial strains, but some differences on the community level between diseased and non diseased animals.

1.1.3 Effect of shell disease syndrome in crustacean populations

Despite its omnipresence in species, such as *C. pagurus*, shell disease is poorly understood in terms of effects on local communities and food webs. The different forms of shell disease may be caused by a combination of environmental factors, bacterial challenge and the internal state of the host (45). Finally, shell disease may cause internal infections which may lead to moribund individuals (46). The dependence of shell disease on different factors, such as environmental factors, may have strong effects on crustacean populations by inter-specific as well as intra-specific interactions. Cases where shell disease significantly affected fisheries are well known. This may be either due to high mortality in the adult population caused by shell disease and successive diseases or due to low market value of diseased individuals. Due to this the occurrence is very well documented in lobsters (*H. americanus*): in the US (the American lobster fishery in New England (northeastern US) it is observed since the 1990s; meanwhile up to 30 % of lobster in Long Island Sound at the East Coast of the US show symptoms and lobster fishery experienced a severe damage (47)). Prevalences of shell disease may be very different between

populations of a species. The American lobster for example suffered from shell disease especially in southern populations, while lobster populations at the coast of Maine and Massachusetts had lower prevalences (3). These latitudinal variation in shell disease prevalence may be connected to environmental factors such as temperature. However, anthropogenic mediated environmental factors may play a role in local shell disease prevalences. In 1997-98 an abnormal high prevalence of shell disease was recorded in Langland Bay (UK) compared to other populations in *C. pagurus* (11). This was assumed to be connected to levels of water pollution in this area. However, also other local factors may play a role. After decreased pollution at this location, prevalence remained high (48). A connection of shell disease outbreaks to pollution had been recorded earlier for the rock crab *Cancer irroratus* (49) as well as for deep sea crabs (50). In such areas a local decrease of the species may be the consequence and may lead to dramatic changes in local benthic communities by the loss of important species, such as *C. pagurus* or *H. americanus*. This may happen even faster if similar species in local communities may differ in shell disease prevalence or severity eventually competing for local resources. Especially in benthic crustacean communities prevalence or severity in shell disease had been shown to vary between species (51). If this has an effect on local communities greatly depends on the biology of either species and its role in the community. The role of shell disease as an omnipresent, environmentally dependent bacterial disease, however, may lead to dramatic changes in marine crustacean biodiversity in the face of anthropogenic mediated environmental changes, such as pollution, climate change or the global anthropogenic mediated connection of habitats.

1.2 Summary of findings and own contributions to manuscripts

Manuscript I: Bacterial biofilms on *Cancer pagurus* and its implication for future shell disease prevalence

Biofilms are mainly inhabited by Proteobacteria, Bacteroidetes and Actinobacteria. The compositions is surprisingly specific on the individual level. On the population level differences are more pronounced by rare taxa. The biofilm is inhabited by a diverse flora of pathogens and potential chitinolytic taxa, some of those bacterial strains are phylogenetically very similar to strains recently discussed as a causative agent of epizootic shell disease. Temperature treatment over 5 days does not change shell disease condition in either population. Bacterial community composition iss dependend on temperature conditions. Facultative abiotic taxa are found to be increased under high temperature conditions and are present in communities with facultative aerobic bacteria. In conclusion, *C. pagurus* shows a high prevalence of shell disease in populations. In relation to temperature connected abundance changes and the omnipresence of opportunistic pathogens future climate change may increase the severity of shell disease in *C. pagurus*, which may lead to even more and stronger forms of shell disease, such as epizootic shell disease.

Own contribution to the study: The idea for the manuscript was developed by me and Prof. Gerlach. I prepared the field trip to Helgoland, Germany and Bodø, Norway. I applied successfully at the European Science Foundation for funding for the work at the Lofoten. I planned the experiments as well as the sampling scheme and underwent the experiments at the field stations and sampling during research cruises at the *RV* Uthörn and *RV* Heincke and at Bodø by diving or by using traps in the local embayment. I prepared all samples for 454 sampling. Most of the 454 sequencing was done by me with assistance of technical staff of University of Oldenburg. I did all the data analysis and the writing. I discussed the manuscript with external scientists for specific analyses. Reviewing of the manuscript was done by Prof. Gerlach and Prof. Brinkhoff.

Manuscript II: Shell disease in crabs in the North Sea: Native *Carcinus maenas* and *Cancer pagurus* suffer but bioinvasive *Hemigrapsus penicillatus* flourishes

In this manuscript the prevalence of shell disease in three different crustaceans of the North Sea, two native, *C. pagurus* and *Carcinus maenas* and one introduced bioinvasive species, *Hemigrapsus penicillatus* is analyzed. The two native crabs are affected by shell disease. *C. pagurus* shows high prevalences. For *C. maenas* shell disease prevalence is medium in populations. *Hemigrapsus penicillatus* lacks any signs of shell disease in newly invaded habitats, however there is evidence that in native habitats these crabs may be challenged by shell disease. The lack of the disease may be advantageous for the bioinvasive crab and may be one factor leading to the enormous success in newly invaded European waters. Additionally, a difference in shell disease prevalence between hosts of different size is found. This indicates that especially bigger species and bigger individuals within species may have to cope with shell disease and may suffer from increased shell disease pathogens facilitated by climate change induced environmental alterations.

Own contribution to the study: The idea of the manuscript was developed during a master project with several scientists of my institution. I helped the bachelor student Marie Kaben-Dobbeck during the project, where finally shell disease was analyzed for *Hemigrapsus penicillatus* and for 50 % of the *Carcinus maenas* samples from the North Sea and few *C. pagurus*. I analyzed 100 % of *C. pagurus*. I added 50 % of the samples from *C. maenas* into the project by analyzing them for shell disease and reviewed all samples from *C. maenas* from the bachelors project. I did the preparations for scanning electron microscopy. I did the analysis of the data and the preparation of the manuscript. Reviewing of the manuscript was done by Dr. Wehrmann and Prof. Gerlach.

Manuscript I

Bacterial biofilms on *Cancer pagurus* and its implication for future shell disease prevalence

in form of a manuscript

Bacterial biofilms on *Cancer pagurus* in a changing climate and its implication for future shell disease prevalence

Running title: Bacterial biofilms in *Cancer pagurus*

Abstract

Bacterial biofilms on crustaceans are important due to their role in shell disease observed in for example copepods, shrimps, crabs or lobsters. Although, the brown crab *Cancer pagurus* is of great ecological and economical importance, a detailed analysis of the bacterial epibiotic communities associated to the shell of *C. pagurus* is still missing. We investigated 221 specimen and found shell disease present in most individuals indicating that shell disease is a normal state of the populations rather than abnormal. We contribute a detailed picture on the epibiotic bacterial biofilm by using next generation sequencing technology. We found bacterial communities as a combination of taxa known for other marine crustaceans and environmental taxa. The highest similarities were observed between samples taken from one individual. Communities showed strong similarities with biofilms studied in North American lobsters. We found many taxa with potential chitinolytic activity such as *Aquimarina* spp., *Kiloniella* spp., *Thalassobius* spp., *Leucothrix* spp. or Chitinophagaceae present in biofilms. To understand potential effects of climate change and temperature raises on the spread of the disease, we analyzed the biofilm composition on the carapace of brown crabs from different populations at the German area of the North Sea (Helgoland) and 1,500 km farther north in Norway (Bodø) under natural and experimentally changed temperature regimes. Short exposure (5 days) to changing temperature regimes had no effect on the prevalence of shell disease but on the diversity of epibiotic biofilm communities. We found temperature dependent differences in abundance of taxa such as the Flavobacteriia, Gammaproteobacteria or Deltaproteobacteria. This study represents the first overview of *C. pagurus* normal epibiotic biofilms and biofilms treated with different temperatures as a basis to study the potential spread of shell disease syndrome in *C. pagurus* and monitor epibiotic communities in populations considering future climate change.

Keywords: barcoding, 16S rRNA, marine biofilm, host-pathogen interaction

Introduction

Bacterial community interactions with an eukaryotic host can range from symbiosis, probiosis to antibiosis. Epibiotic biofilms on crustaceans get recently into the focus, due to the occurrence of biofilm induced shell disease causing heavy economical and ecological consequences in North American lobster populations [1]. Shell disease is a syndrome discussed controversially to emerge due to the presence of single pathogens [2–6] or even shifts in communities [7, 8]. *C. pagurus* is the biggest native brachyuran crab inhabiting the coastal regions from Marocco to the Northern Arctic at Europe [9]. The brown crab *C. pagurus* is facing various environmental conditions in its distributional range especially regarding water temperature. High abundances of *C. pagurus* are found in Norway with relatively cold water throughout the year [10]. They are also found in shallow waters of the southern North Sea with bottom water temperature in summer of up to 20 °C. The

size of this predator and the long inter-molt intervals allow even for pronounced eukaryotic fouling on carapaces. The mobile adults live close to the sea-floor interacting with substrates by burrowing and feeding. The brown crab is one of the most affected species by crustacean shell disease in Europe. Earlier investigations showed that the occurrence of shell disease is nearly omnipresent at bigger size classes in one population [5]. However, massive economical losses by shell disease were not recorded. Whether this is due to a stable interaction of shell disease bacteria with the host or the lack of occurrence of certain disease causing bacterial taxa is not known to date. Different forms of shell disease are described, among others the impoundment disease [11], the epizootic shell disease recorded in lobster [12] or the burn/rust spot disease [13]. All are characterized by chitinolytic degradation of outer cuticle by microorganisms followed by antimicrobial defense of hosts including melanization of tissues. Different forms of shell disease had been associated with different bacterial taxa [2–4, 8].

For the American lobster it has been shown that epizootic shell disease is correlated with bottom water temperature in North American coastal waters [14]. Accordingly temperature is a likely factor for increased pathogen abundance. In concert with species specific biology such as molting intervals it can trigger disease prevalence [15]. In *C. pagurus*, an increase in pink crab disease was postulated in Irish waters as a consequence of bottom water temperature [16]. Susceptibility to diseases related to the capability to cope with environmental changes will determine the fate of many crustaceans in the marine environment [17, 18]. Shell disease affects many different forms of crustaceans. For most marine crustaceans missing information on the normal conditions in biofilms constrains studies on the origin of such diseases. We contribute to the knowledge of normal bacterial biofilms associated to an economically and ecologically very important crustacean species in Europe, the brown crab *Cancer pagurus* Linnaeus, 1758. We aim at understanding which biofilm bacteria can be found under different conditions and whether temperature may play a role in bacterial community composition on carapaces. Our approach will contribute to the understanding, whether future climate change may increase shell disease in populations of *C. pagurus* or whether it even may lead to more dramatic forms such as epizootic shell disease which are at the moment not reported in Europe. Our results on normal biofilm composition and their stability under changed environmental conditions may thus contribute to the understanding, how the interaction of biofilms with crustacean host will influence crustacean biodiversity in a climate change scenario.

Analysis

Sampling and Experimental Conditions

C. pagurus was sampled in June 2010 close to the island of Helgoland, Germany (54° 10' 49.176" N, 7° 53' 20.197" E) and in November 2010 at Mørkvek Bugda near Bodø, Norway (67° 17' 2.299" N, 14° 32' 55.265" E) and from the "Tiefe Rinne" in the North Sea (54° 5' 15.0" N, 7° 32' 28.3" E; Suppl. S1 Text). At Helgoland and Mørkvek Bugda crabs were kept at ambient water temperature in tanks supplied with local water. Temperature was experimentally changed from ambient to either warm or cold conditions for over 5 days (Suppl. S1 Text). Epibacterial biofilms were sampled for either condition: Norway-Ambient-Temperature (7.5 °C), Germany-Ambient-Temperature (14.5 °C), Norway-Experimental-Warm (16 °C), Germany-Experimental-Cold (5 °C) and Germany-Wild (15 °C). Biofilm samples were taken and shell disease prevalence was recorded from carapaces (Fig. 1; Suppl. S1 Text).

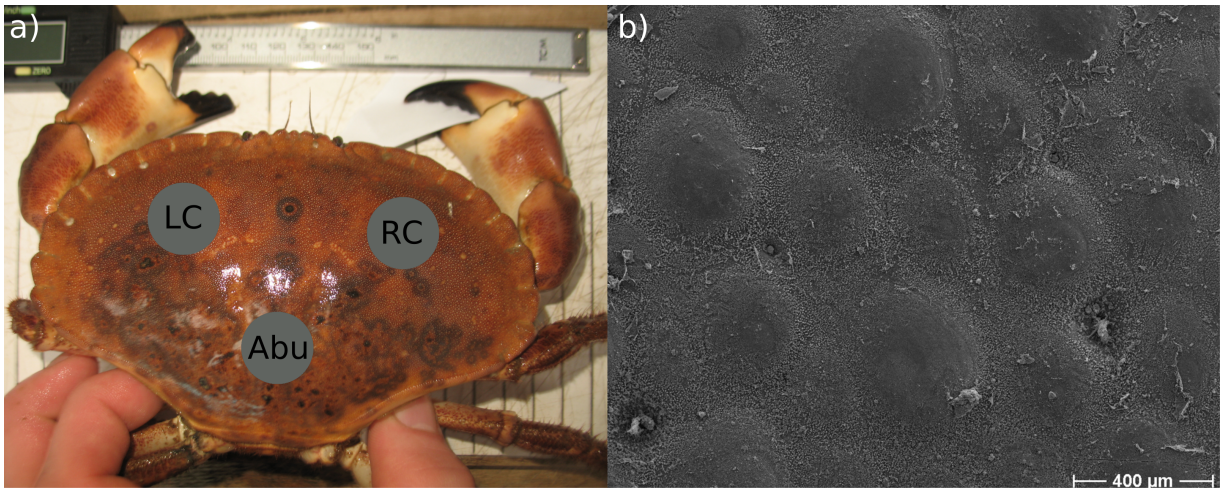


Figure 1. Sample areas for harvesting epibiotic biofilms. Sampling locations on carapaces of *C. pagurus* (a). Left carapace (LC), right carapace (RC) and anterior carapace (Abu) were sub-sampled using an area of 2 cm in diameter. Representative example of the outer surface of *Cancer pagurus* carapaces (b). Surface is covered with dense formations of extrusions and small depressions forming a high variety of micro-habitats for bacterial communities.

Shell disease analysis

Pictures from individuals from all populations and from each treatment were analyzed using gimp 2.7. External dorsal carapace surfaces were used as indication for shell disease prevalence. Each picture was sized to a stencil of equal sized areas [5]. The number of areas affected by shell disease on carapaces was determined. Discolorations were classified as follows: 1 - light brownish spots only on surface (not classified as shell disease), 2 - clear signs of shell disease up to heavy shell disease with erosion progressing through carapace layers, 3 - blackening due to injuries (Individuals were removed from the analysis). Non-parametric statistics was used to compare shell disease prevalence between populations as well as shell disease before and after temperature treatment. Light discolorations of category 1, which can be early signs of shell disease were omitted. This was done to avoid artifacts in population comparison due to weak shell disease signs from animals kept under tank conditions prior to sampling. Category 2 disease includes shell disease which cannot emerge in short timeframes.

Bacterial community analysis by 454 GS Junior Next Generation Sequencing

Bacterial community analysis was performed based on the primer pair 27f (TCAGAGTTTGATC-CTGGCTCAG) and 338r (CATGCTGCCTCCCGTAGGAGT) amplifying the V1 of bacterial 16S rDNA [19]. 454 Sequencing was performed with a 454 GS Junior Next generation sequencer (Roche) according to the standard protocol using Lib-A chemistry. Raw reads were analyzed using qiime ("Quantitative Insights Into Microbial Ecology") pipeline [20] including amplicon denoising [21] and cleaning for PCR chimeras using ChimeraSlayer [22]. De-multiplexed raw reads were used by qiime to build operational taxonomic units (OTUs), i.e. sequences which are 97 % similar by usearch clustering approach (version 6.1) [23]. Phylogenetic inference of microbial OTUs was calculated using PyNast aligner [20]. For each cluster a representative OTU was picked and used for assigning

the most precise taxonomic identifier possible using blast routines with cut-off of 90 % similarity and e-value of 10^{-3} [24] against SILVA111 database as reference [25]. In order to compare sequences found similar to shell disease pathogens on other species we used ARB neighbor joining method to build a tree on published 16s rRNA gene sequences for these groups including bootstrap analysis with 1000 replicates [26]. A more detailed description of the molecular analysis can be found in the supplement (Suppl. S1 Text).

Results

Shell disease prevalence

A total of 221 individuals of *C. pagurus* were analyzed for shell disease prevalence. We observed a high prevalence of shell disease in all populations with 80 % occurrence in the North Sea offshore samples of *C. pagurus*, 71 % at Helgoland (summer), 69 % at the "Tiefe Rinne" close to Helgoland (summer) and 65 % at Bodø in Norway (winter). Sampling time and geographic area had no effect on the prevalence of shell disease (Pearson's $x^2 - test$, $x^2 = 1.3137$, $df = 3$, $p = 0.72$). Similarly, the severity of shell disease was comparable among populations. In all populations we found several spots of shell disease on individual carapaces. Populations sampled from the North Sea showed higher median shell disease prevalence (Germany: median = 3; North Sea: median = 2; Deep ridge: median = 2). In contrast, median shell disease prevalence at Bodø was lower (Norway: median = 1). However, there was no significant difference between populations (Kruskal Wallis test, $x^2 = 4.5276$, $df = 3$, $p = 0.21$).

Epibiotic bacterial communities

Bacterial communities were harvested from 25 samples in total (Germany-Wild $N = 12$ in 6 individuals; Germany-Treatment $N = 6$ in 3 individuals; Norway-Treatment $N = 7$ in 4 individuals). A total of 409,775 sequences were obtained after checking for primer and MID as well as quality in reads considering all samples. The denoising process by AmpliconNoise and the removal of all reads with more than 50 low quality bases decreased the absolute number of reads to 315,777. The reads per sampled community ranged between 4,662 and 24,354 counts with a median of 11,428 and mean 12,125 read number. A total of 5,110 OTUs on a 97 % similarity level were calculated. 5,086 OTUs were successfully assigned using SILVA111 as reference [25]. After removal of PCR chimeras 5,017 OTUs were used for further analysis.

The core community was calculated from 12 samples of 6 wild caught individuals at Germany with a threshold of OTUs present in 80 % of the samples (Fig. 2). The core community consisted of three bacterial phyla; Actinobacteria (51.3 %), Bacteroidetes (11.4 %) and Proteobacteria (37.3 %) with a total of 30 different OTUs (Table 1). Bacterial genera within the Bacteroidetes present in more than 80 % of all samples were *Lutibacter* and *Pibocella*. For the Proteobacteria *Filomicrobium*, *Litoreibacter*, *Tateyamaria*, *Sphingopyxis*, *Granulosicoccus*, *Cocleimonas* and JTB255 marine benthic group were present. Within the Actinobacteria all taxa present at most of the samples were assigned to uncultured taxa.

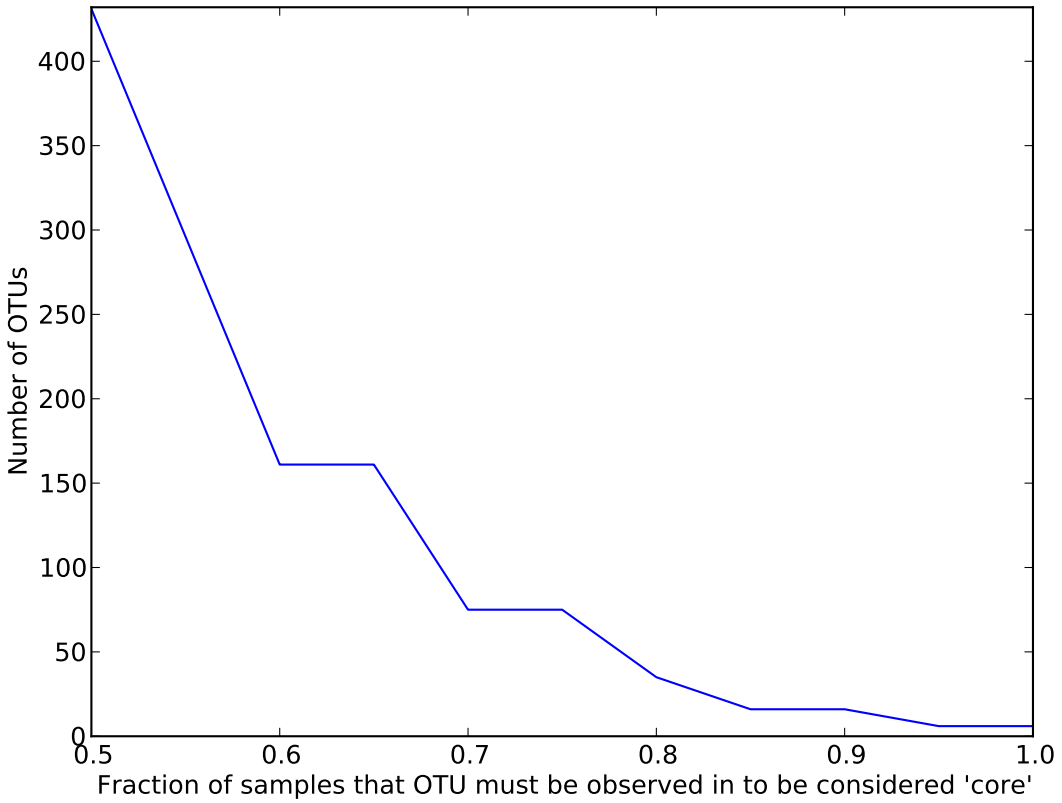


Figure 2. Number of OTUs present in the relative fraction of samples ($N = 12$) from wild caught individuals. In order to calculate a core community on samples the number of OTUs present in different fractions are plotted. The latest strong drop-off of OTU number observed in thee fraction of samples was observed after 75 %. The core community was therefore defined by all OTUs present in at least 80 % of the samples.

Table 1. Core community showing phylotypes present in at least 80 % of Germany wild samples. Threshold was defined visually from the plot showing number of OTUs present in at least 50 % to 100 % of the samples (Fig. 2). Presence of each taxon from the Germany wild ($N = 12$) core community is given for Germany Tank ($N = 6$) and Norway Tank ($N = 7$) population with the relative amount number of samples each taxon was observed in.

Taxon			Germany Wild	Germany Tank	Norway Tank
Actinobacteria					
	Acidimicrobiaceae	*Ilumatobacter*	83 %	83 %	57 %
		uncultured	92 %	100 %	100 %
Acidimicrobiales	Sva0996 marine group	uncultured actinobacterium	100 %	100 %	100 %
		uncultured bacterium	83 %	100 %	100 %
		uncultured organism	92 %	83 %	29 %
	uncultured	Gram positive bacterium	100 %	83 %	14 %
		actinobacterium	83 %	83 %	100 %
		bacterium	92 %	100 %	100 %
		organism	83 %	100 %	57 %
Bacteroidetes					
Cytophagales	Flammeovirgaceae	*Persicobacter*	83 %	100 %	71 %
Flavobacteriales	Flavobacteriaceae	BSD RB 42	83 %	100 %	57 %
		Lutibacter	83 %	100 %	100 %
		Pibocella	100 %	100 %	100 %
Sphingobacteriales	Saprospiraceae	uncultured	83 %	100 %	100 %
Proteobacteria					
OCS116 clade	uncultured bacterium	Other	92 %	83 %	100 %
Rhizobiales	Hyphomicrobiaceae	*Filomicrobium*	83 %	83 %	100 %
	Rhodobiaceae	*Anderseniella*	100 %	83 %	100 %
Rhodobacterales	Rhodobacteraceae	*Litoreibacter*	92 %	100 %	100 %
		Marinosulfonomonas	83 %	100 %	43 %
		Roseobacter CHAB-I-5 lineage	100 %	83 %	57 %
		Tateyamaria	92 %	100 %	100 %
		Thalassobius	100 %	50 %	100 %
		uncultured	92 %	100 %	100 %
Sphingomonadales	Erythrobacteraceae	*Altererythrobacter*	83 %	87 %	57 %
	Sphingomonadaceae	*Sphingopyxis*	92 %	100 %	100 %
Chromatiales	Granulosicoccaceae	*Granulosicoccus*	83 %	100 %	100 %
Thiotrichales	Thiotrichaceae	*Cocleimonas*	83 %	100 %	100 %
Xanthomonadales	Sinobacteraceae	JTB255 marine benthic group	83 %	100 %	100 %

Variability in bacterial communities on carapaces

Alpha variation was calculated on rarefied reads for each category of samples limited by the size of the pool of the smallest sample. The highest alpha variation was found in Germany warm water aquarium samples consistent for phylogenetic diversity and shannon index (Fig. 3, orange). The remaining pools showed smaller variation by phylogenetic diversity as well as shannon diversity (Fig. 3). On average the smallest variation by phylogenetic diversity was found for cold ambient Norway samples (Fig. 3 a, green) and for the shannon diversity for cold treated Germany samples (Fig. 3 b, red). However, in both cases variation overlapped with other pools except for warm ambient Germany samples (Fig. 3). All further statistics were applied on rarefied pools for each sample to 4,500 sequences, which was a read count representative for all categories in the non-exponential phase of OTU increase per sequencing depth.

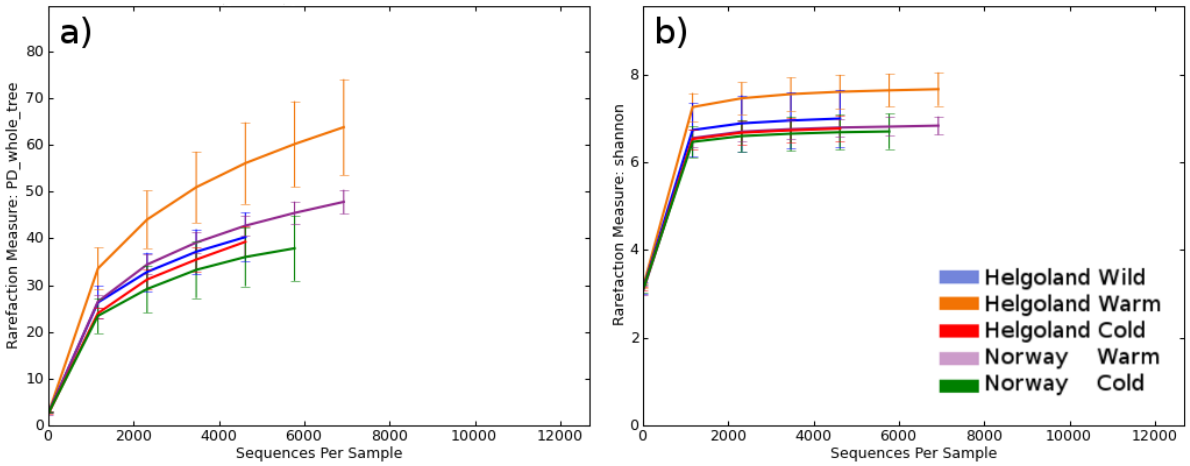


Figure 3. Rarefaction analysis of alpha diversity for each treatment. Results are shown for denoised reads by AmpliconNoise using a) phylogenetic diversity (PD) per sequencing depth and b) shannon index of diversity as indication of variation in samples. Different pools of samples are combined in different colored regression lines (red = Germany treated cold individuals $N = 3$; blue = Germany wild individuals $N = 12$; orange = Germany ambient warm tank samples $N = 3$; purple = Norway tank samples treated with warm temperature $N = 3$; green = Norway ambient cold temperature tank samples $N = 4$).

By comparing biofilms pooled for each geographic region we found a significant high correlation for this model by ANOSIM and the more robust Adonis analysis using non weighted UniFrac differentiation. Using weighted UniFrac analysis, which weights each OTU by its abundance in reads, differentiation between populations was significant but correlation by ANOSIM and Adonis was lower (Table 2). Principle Coordinate Analysis (PCoA) based on jackknifed community distances showed clustering for populations. Variation explained by principle coordinates was PC1 = 38 %, PC2 = 25 % and PC3 = 8 % showing a high variation in bacterial communities among samples within populations for weighted Unifrac similarity (Fig. 4). Germany tank population overlap with Norway population sampled from tank conditions. Germany wild population did overlap with few samples of Germany tank samples in PC1 vs. PC2 and PC1 vs. PC3 and additionally with Norway tank samples in PC2 vs. PC3 (Fig. 4). Clustering in populations for weighted UniFrac dissimilarities is resembled by the average distribution of higher taxa of bacteria to each biofilm comparing abundant taxa with

more than 1 % contribution to the average community (Fig. 5). Mean abundance of taxa from Proteobacteria was higher in samples from tank environment (Norway=53 %, Germany=42 %) than Germany wild conditions (33 %). Contrarily, Actinobacteria were less abundant in samples from tank environment (Norway=11 %, Germany=7 %) than Germany wild conditions (18 %). Bacteroidetes were lower abundant in Germany wild samples (38 %), more abundant in tank samples (45 %) and least abundant in Norway samples (30 %). Most prominent families contributing to the average community were Flavobacteriaceae similarly contributing to all pools of samples (Germany Wild=26 %, Germany Tank=32 % and Norway Tank=26 %; Fig. 6). Rhodobacteraceae were more present in biofilms on *C. pagurus* from tank environment (Germany=21 %, Norway=26 %) than in the Wild (Germany=8 %). Moraxellaceae were abundant only in Germany Wild samples (1 %). Families with low abundance but present over all pools of samples were Thiotrichaceae, Sinobacteraceae, Granulosicoccaceae, Phyllobacteriaceae, Saprospiraceae, Cryomorphaceae, Sva0996 marine group and Acidimicrobiaceae (Fig. 6). On the genus level most prominent were Maritiminonas with highest average abundance in Norway tank (14 %) and lesser abundance in samples from Germany (tank=8 %; wild=6 %). Several genera from the Rhodobacteraceae were present especially in tank samples from Norway e.g. *Leisingera*, *Roseobacter clade NAC* $11-7$, *Roseovarius*, *Shimia* and *Sulfitobacter*. At Germany tank samples only *Loktanella*, *Roseobacter NAC* $11-7$ and *Roseobacter clade CHAB*$-I-5$ were found in comparable high abundance. At Germany wild samples *Roseovarius*, *Pacificibacter* and *Loktanella* were present. Comparing different individuals within the Germany Wild population did show a considerable similarity among different biofilms from one individual compared to others as indicated by ANOSIM and Adonis models (Table 2). The presence of taxa on the phylum and family level between different individuals was similar with the exception of Moraxellaceae being heavily present in only one sample (Fig. 8). Similarity on the individual level was high in individual 1 with high abundance of Bacteroidetes, average abundance of Proteobacteria and low abundance of Actinobacteria. In individual 2 left and right carapace was similar with exceptional high abundance of Actinobacteria and low Bacteroidetes. High abundance of Proteobacteria and very low Acidobacteria was common for all samples from individual 3. Accordingly, similarities among samples of one individual were found in abundance in families as well. For example the abundance of Saprospiraceae and Phyllobacteriaceae was similar on individuals. Thiotrichaceae were nearly absent from all samples of individual 2, but common for all other samples (Fig. 8). Flavobacteriaceae were the most important taxon from Bacteroidetes and showed a similar pattern as on the phylum level (Fig. 8, 7).

In the epibiotic biofilms on *C. pagurus* phylotypes including bacteria capable of chitin decay were present in all populations. For example members of Clostridiaceae [27], Chitinophagaceae [28] or Shewanellaceae [29] were found. Moreover, some bacteria accused for different forms of shell disease in the American lobster *H. americanus* were phylogenetically very similar to those we found in *C. pagurus*. The uncultured *Aquimarina* species described for epizootic shell disease, proposed to be called *Aquimarina* 'homaria' I32.4 (gi | 372124983 | gb | $JN987166.1$ |) [2] was very similar to OTUs found at *C. pagurus* (Fig. 9). Additionally, four OTU's were assigned to *Kilionella*. This taxon was described for diet induced shell disease in *H. americanus* [2]. The specific strain was proposed to be called Candidatus *Kopriimonas aquarianus* (gi | 372124989 | gb | $JN987172.1$ |) [2].

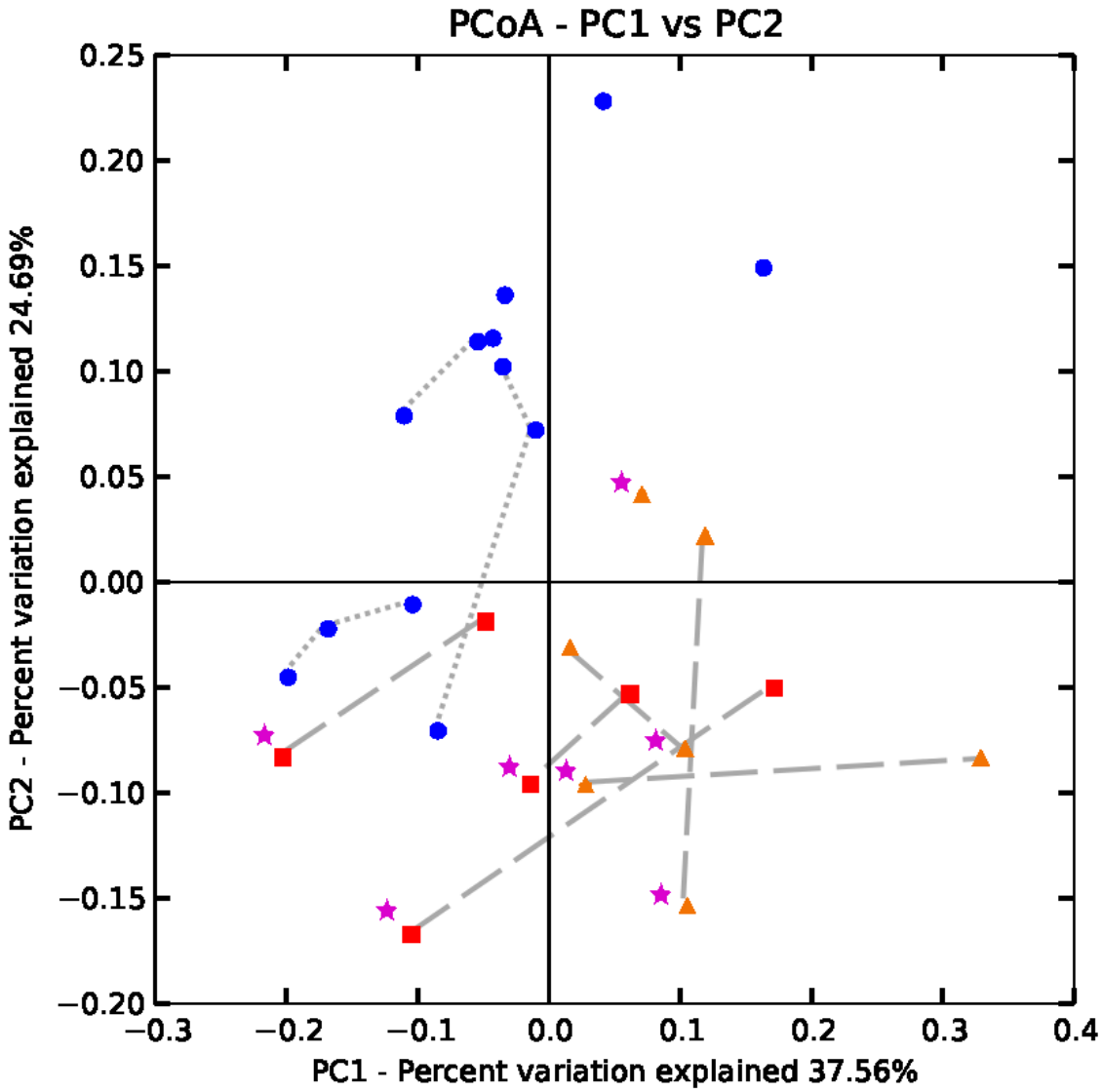


Figure 4. Multidimensional plot of PCoA of individual bacterial biofilms. PCoA is shown based on a) unweighted and b) weighted Unifrac dissimilarities for the first two principal coordinates PC1 and PC2. Different colors represent samples from different populations. Blue = Germany wild, red = Germany tank, orange = Norway tank. Samples connected by gray dotted or striped lines originate from one individual. Samples marked with a purple asterisk were held at cold temperatures (5 °C for Germany and 7.5 °C for Norway).

Table 2. Multivariate statistics of bacterial communities. ANOSIM and Adonis statistics for the comparison of bacterial communities between samples from Germany and Norway (Grouping: Population). Samples grouped for individual are analyzed at Germany wild population (Grouping: Variation on Carapaces). Results are given for treatments in each population (Grouping: Treatment) for Germany tank (Warm) vs Germany tank (Cold) and Norway tank (Cold) vs. Norway tank (Warm)). Estimates of similarity given for unweighted UniFrac and weighted Unifrac dissimilarity. Sample size (N) gives the absolute number of samples among groups. Significance level is given on 999 permutations and based on each pool rarefied to 4500 reads.

Grouping	Statistic	N	ANOSIM (R)	Adonis (R^2)
Population	unweighted UniFrac	25	0.700***	0.186***
	weighted Unifrac	25	0.310***	0.271***
Variation on Carapaces	unweighted UniFrac	12	0.583**	0.532***
	weighted Unifrac	12	0.583**	0.686**
Treatment Warm Norway	unweighted UniFrac	7	0.037	0.183
	weighted Unifrac	7	0.093	0.159
Treatment Cold Germany	unweighted UniFrac	6	0.000	0.250
	weighted Unifrac	6	0.000	0.306

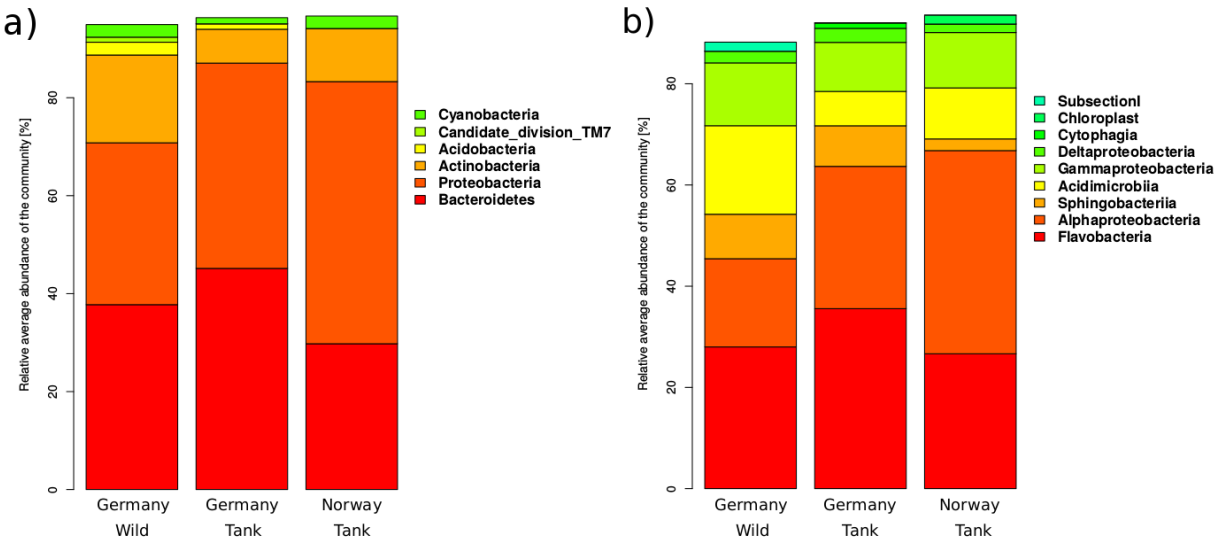


Figure 5. Relative average abundance of bacterial taxa in epibiotic biofilm of *C. pagurus*. Average contribution of different bacterial taxa with more than 1 % abundance from Germany wild samples, Germany tank samples and Norway tank samples assigned for the level of (a) phylum and (b) clade.

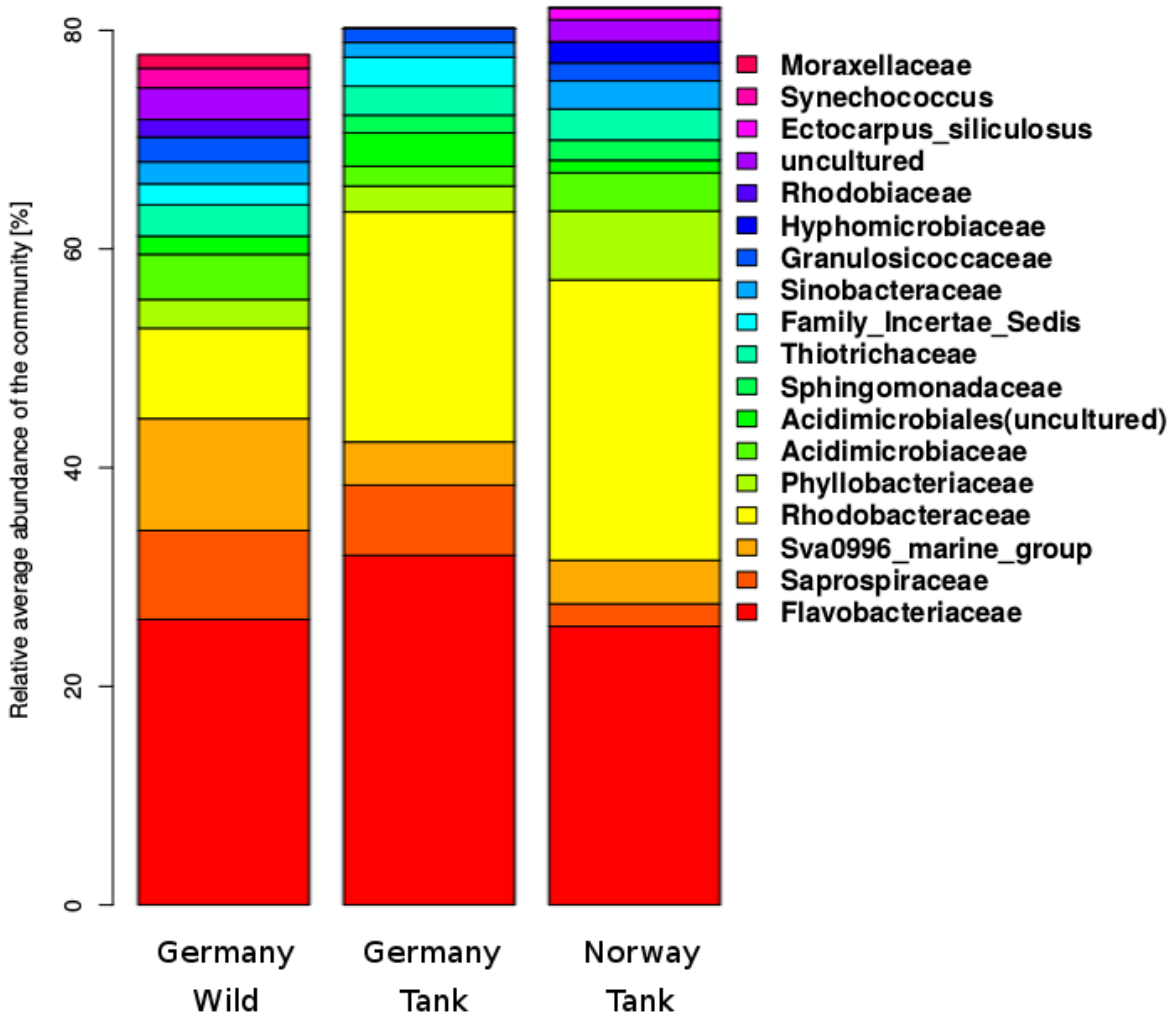


Figure 6. Relative average abundance of bacterial taxa in epibiotic biofilm of *C. pagurus*. Average contribution of different bacterial taxa with more than 1 % abundance from Germany wild samples, Germany tank samples and Norway tank samples assigned for the level of family.

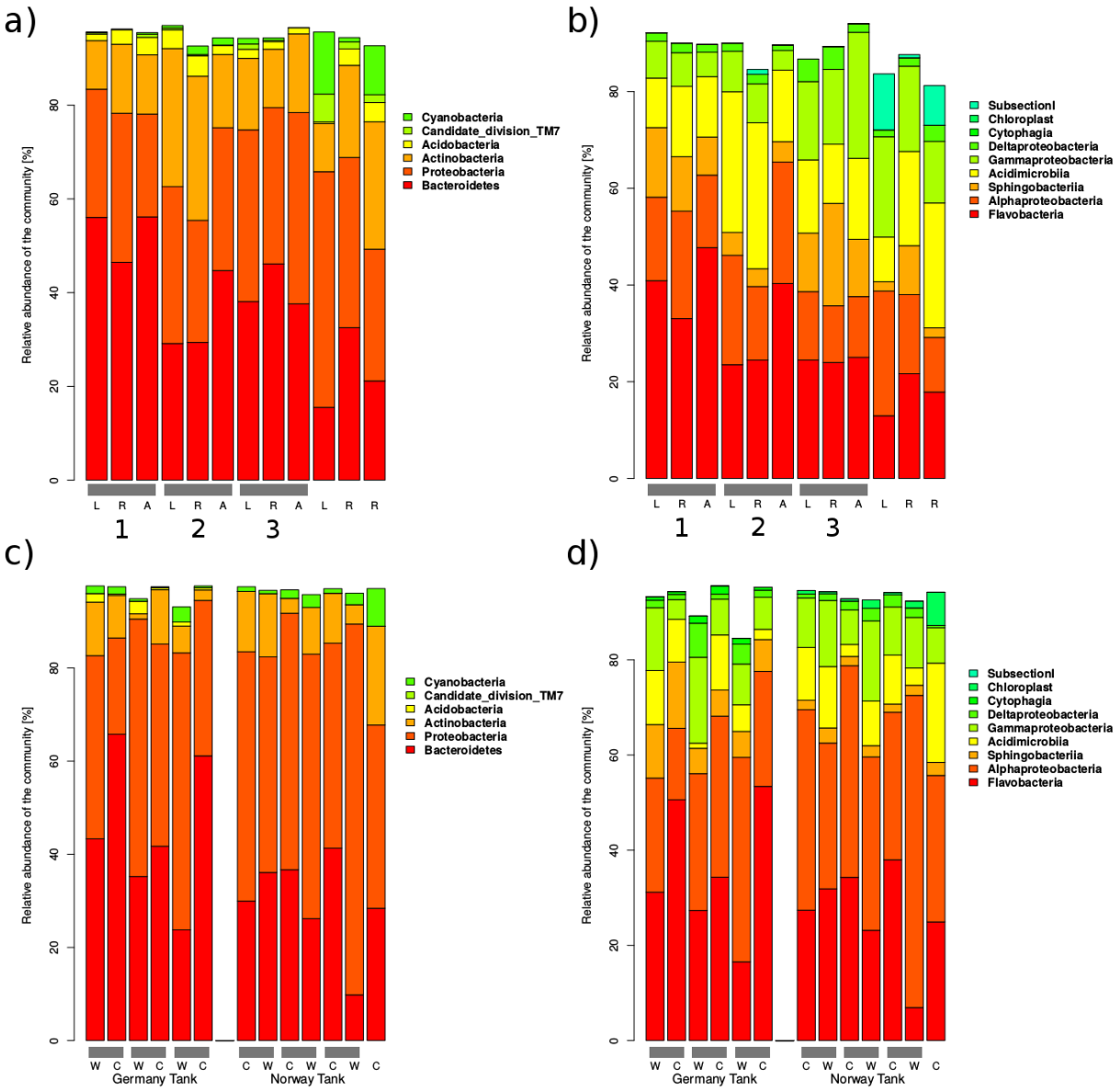


Figure 7. Individual variability in epibiotic biofilm of *C. pagurus*. Contribution of different bacterial taxa with more than 1 % abundance for each individual biofilm from Germany wild samples assigned for the level of (a) phylum and (b) clade and Germany tank samples compared to Norway tank samples assigned for the level of (c) phylum and (d) clade. Grey bars indicate samples originating from one individual at different carapace area according to sampling regime (Fig. 1). Samples marked with C or W originate either from left or right part of the carapace and indicate whether samples were taken during warm (W) or cold (C) experimental conditions.

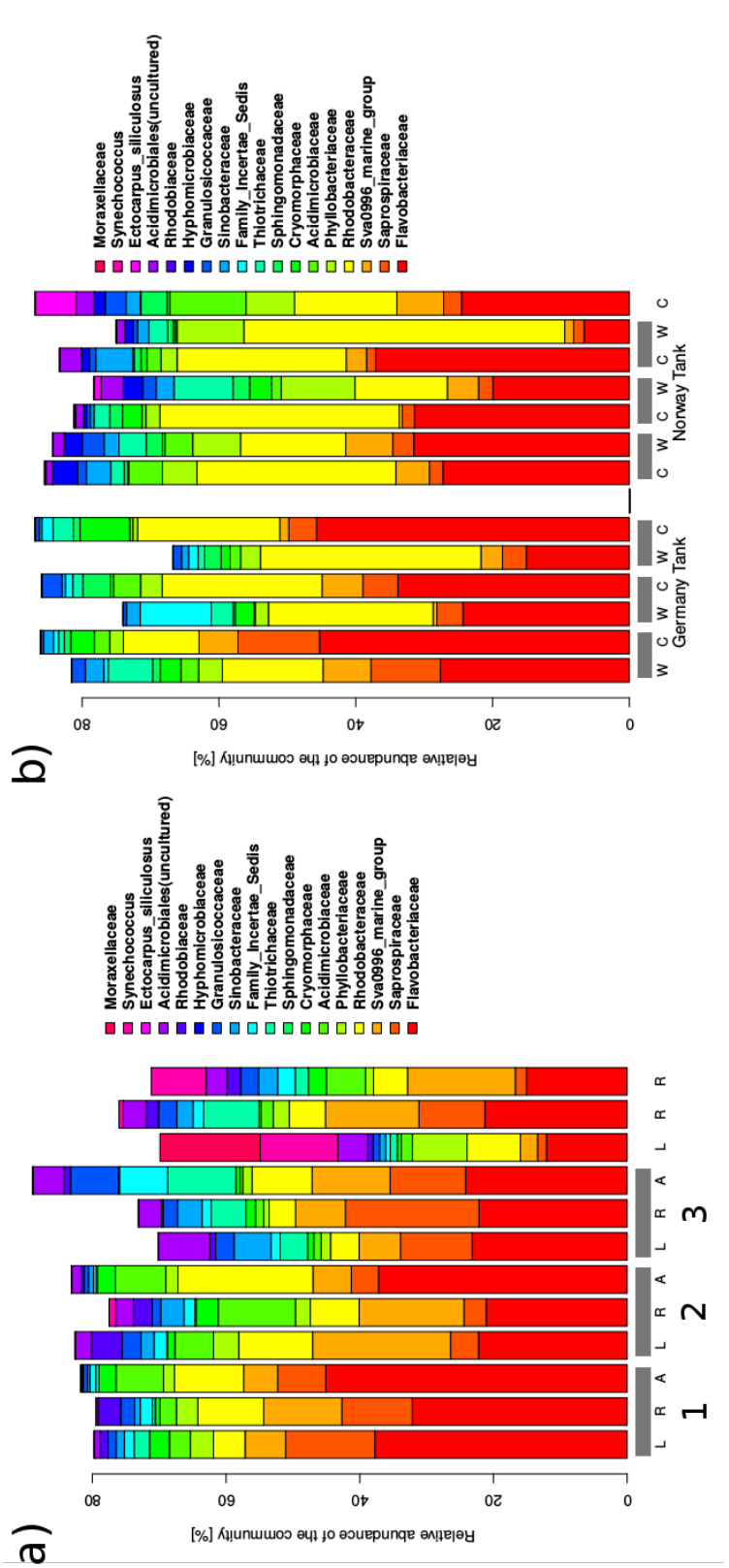
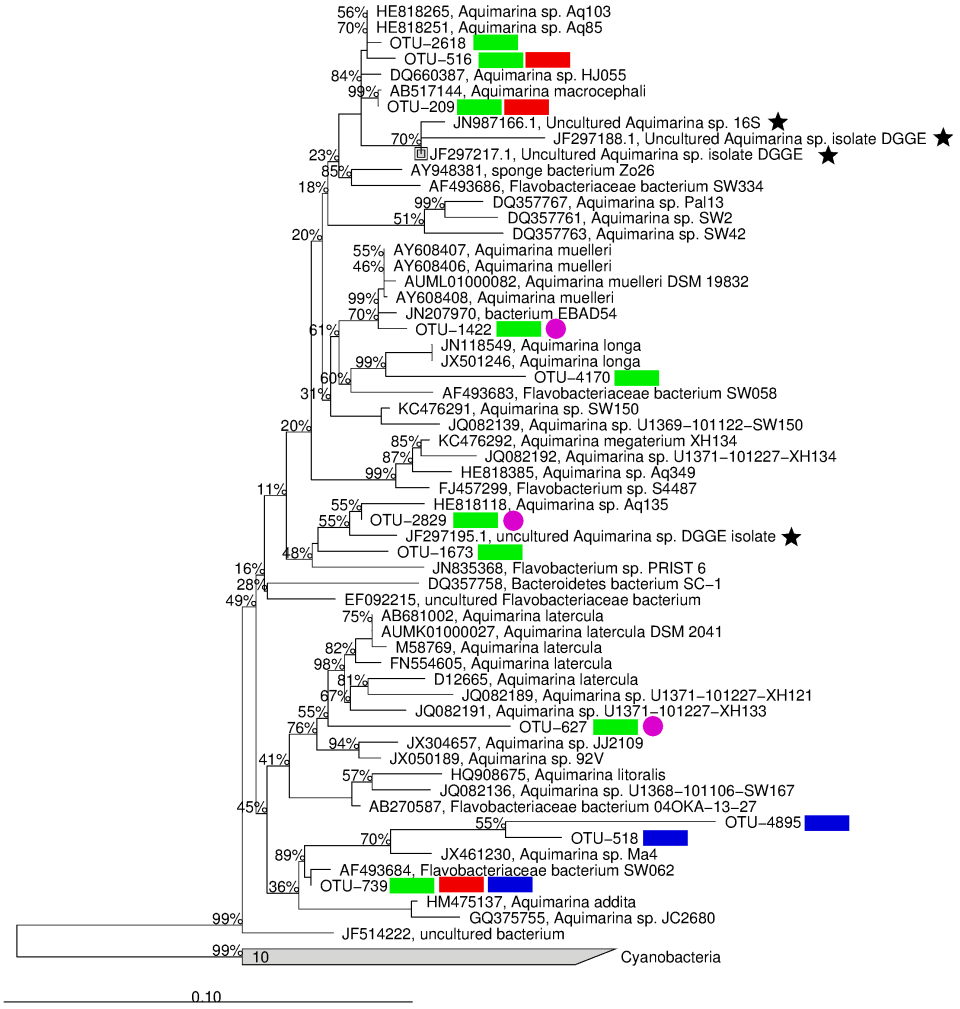
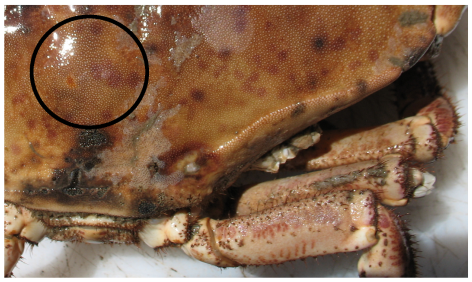


Figure 8. Individual variability in epibiotic biofilm of *C. pagurus*. Contribution of different bacterial taxa with more than 1 % abundance for each individual biofilm from Germany wild samples assigned for the level of (a) family and Germany tank samples compared to Norway tank samples assigned for the level of (b) family. Grey bars indicate samples originating from one individual at different carapace area according to sampling regime (Fig. 1). Samples marked with C or W originate either from left or right part of the carapace and indicate whether samples were taken during warm (W) or cold (C) experimental conditions.





Sampled area of one individual with shell disease similar to epizootic shell disease. OTUs found are indicated with ●

Figure 9. Neighbor Joining tree showing phylogenetic relationships based on 16S rRNA gene sequences of OTUs determined in this study and *Aquimarina* spp.. Bootstrap is based on 1000 replicated trees. Partial 16S rRNA gene sequences were included via maximum parsimony [26]. Sequences found on *H. americanus* carapaces [2,3] are marked with an asterisk. Sequences of our study are marked with a colored bar indicating the origin of the OTU; red = Germany tank, blue = Norway tank, green = Germany wild. Purple circle indicates OTUs found at one specimen of the study with strong shell disease.

Effect of temperature on biofilm and shell disease

Decreasing temperature in Germany for 120 h and increasing temperature for 120 h in Norway did not change shell disease prevalence in treated animals. We could not observe a shift in community composition by temperature treatment in either population (Table 2). ANOSIM and Adonis statistics did not reveal any similarities among treated or not treated biofilms (Table 2). Due to the specificity of bacterial biofilms on the individual level (Table 2), treated and non treated biofilm samples were compared paired for each individual. There was a strong decreased alpha diversity in cold water samples compared to warm water samples similarly for Germany and Norway samples. For all samples decreased alpha variation on rarefied reads was significant based on difference in phylogenetic diversity (Wilcoxon signed rank test (paired), $V = 21$, $p < 0.05$) and species richness (chao1; $V = 21$, $p < 0.05$), but not in evenness calculated by shannon diversity index (Wilcoxon signed rank test (paired), $V = 18$, $p = 0.16$). Accordingly, the contribution of taxa with abundance > 1 % to the total community was decreased in each sample taken from warm conditions compared to its counterpart from cold conditions (Fig. 8).

Comparing pairwise difference in abundance of dominant bacterial phyla between temperature showed similarities in community change (Fig. 10). Proteobacteria were higher in warm water samples compared to cold water. Bacteroidetes did show a strong change to increased abundances in cold environments for both Norway and Germany. Acidimicrobiia were on average more increased in warm environments (Fig. 10). The increase of Gammaproteobacteria was strongly influenced by an increase *Cocleimonas* spp. within the Thiotrichales at both treatments (Fig. 10). Within the Gammaproteobacteria Chromatiales were increased as well during warm water. Other groups increased under high temperature conditions were Clostridiales (*Coxiella* spp.), Nitrosomadales (*Nitrosomas* spp.) and Nitrospirales (*Nitrospira* spp.). Contrarily, Flavobacteriales within the Flavobacteriia consisted of genera which did show an increased abundance under cold water conditions such as *Maritimimonas*, *Tenacibaculum* or *Bacteroides* spp..

Discussion

Epibiotic biofilms of *C. pagurus*

We found high shell disease prevalence in *C. pagurus* in warm and cold water populations at the North Sea and Norway close to the Lofoten. Prevalence was comparable as for a local population at the British Isles [5, 30]. For this species shell disease seems to be the normal condition rather than an exception. The normal bacterial community composition observed was relatively different compared to other marine crustaceans in Europe. But it showed similarities to biofilms found in *H. americanus* in North America. We found strong specificity of biofilms on the individual level. Differences between populations were mainly derived from the presence of rare taxa, while the contribution of prominent taxa was more stable over regions. The diverse community was inhabited by many potential pathogens and several bacterial taxa strongly related to taxa accused for massive shell disease outburst in North American lobster populations.

The majority of bacteria in *C. pagurus* biofilms belonged to the phyla Proteobacteria and Bacteroidetes. In wild caught animals Actinobacteria played a major role as well. Those were underrepresented in samples derived from tank held individuals. However, the 80 % core community consisted of many Actinobacteria in contrast to their low abundance. Taxa such as Sva0996 marine group [31] and *Ilumatobacter* [32] were described for sediments and may thus be more

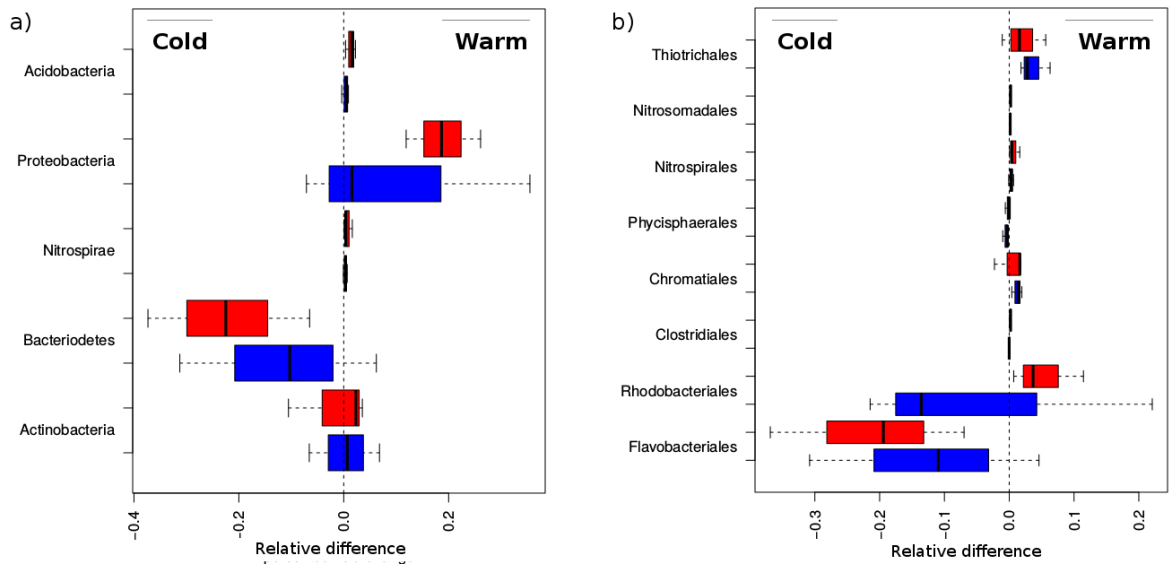


Figure 10. Average pairwise difference in abundance of dominant bacterial clades for epibiotic biofilms of *C. pagurus* kept under warm and cold conditions. Treatment was from ambient warm to cold conditions in Germany (red boxes) and from ambient cold conditions to warm conditions in Norway (blue boxes). Differences are shown for phyla a) and selected abundant orders b).

loosely connected to the biofilm. Most bacteria associated to planktonic marine copepod *Calanus finmarchicus* in Greenland for example were Proteobacteria. In contrast to *C. pagurus* associated biofilms, Actinobacteria were completely missing and Bacteroidetes were strongly underrepresented [33]. Gerds et al. [34] analyzed biofilms of four different copepod species in the same geographic region in Germany as we studied for *C. pagurus*. They found a high similarity among species and a high diversity for Proteobacteria. Contrarily to our findings, only few Bacteroidetes were present. Within the Bacteroidetes we found as part of the core community for example *Persicobacter* described to be associated with shellfish [35] and *Pibocella* described from marine algae [36]. Proteobacteria were a major part of the core community in *C. pagurus*. However, only two (*Marinosulfonomonas* spp., *Sphingopyxis* spp.) of 12 genera present in the core community were found on copepods in the same area as well [34]. This indicates that Proteobacteria seem to be an important taxon in epibiothic biofilms on crustaceans in general. However, the taxon composition within the phylum Proteobacteria seems to be more specific for different species of crustaceans. Additionally, Bacteroidetes and Actinobacteria seem to be more specific for benthic crustaceans than for planktonic species.

For biofilms of *H. americanus* in populations of New England the importance of Bacteroidetes and Proteobacteria in the biofilm community was similar to our findings for *C. pagurus* [8]. Several taxa were found in both species despite the huge geographic distance, such as *Thalassobius* spp., *Tenacibaculum* spp., *Pibocella* spp., *Aquimarina* spp. or *Ulvibacter* spp. [8]. Many of those taxa were described to be associated with benthic eukaryotes [36–38]. Thiotrichaceae found on *C. pagurus* are often found in marine biofilms on other benthic invertebrates such as on marine amphipods (*Thiotrix* spp.; [39]), on sand snails (*Cocleimonas* spp.; [40]) or on crustaceans algae and fish (*Leucothrix*

spp.; [41]). Still there were differences in dominant taxa as for example the most dominant taxon *Jannaschia* in *H. americans* was very rare in *C. pagurus* [8]. Thus the biofilm community of *C. pagurus* is a combination of taxa specific for biofilms on invertebrates on a regional scale (e.g. as found on copepods; [34]) and taxa more determined by the biology of the species (e.g. on as found benthic lobsters; [8]). On the individual level similarity based on occurrence of bacterial taxa as well as similarity based abundance was higher in individuals than between individuals. This indicates that even small scale habitat conditions as well as the individual state such as molting age, migration routes or immunocompetence may play a role for bacterial community composition. Similarity was driven mainly by the abundance of Flavobacteriaceae, Rhodobacteraceae and Thiotrichaceae, but was also visible on the phylum level.

Pathogen prevalence in *C. pagurus*

Chitinolytic taxa and potential pathogens were omnipresent in biofilms. Bacterial taxa which were in general characterized as chitinolytic such as Chitinophagaceae, members of *Photobacterium*, *Bacillus* [42] or *Clostridium* [43, 44] were present. We found sequences from bacteria, which had been recently discussed as source for different forms of shell disease of the genus *Thalassobius* [45], *Tenacibaculum* [4], very few *Vibrio*, *Pseudoalteromonas* [46] and *Kilionella* spp. [2]. We found several species of *Aquimarina* spp. some of those very similar to taxa suspected to cause several forms of shell disease including a strain similar to *Aquimarina 'homaria'* I32.4 discussed in relation to epizootic shell disease [2]. However, pathogens in the biofilm were not limited only to chitinolytic bacteria. We found on one individual a high abundance of *Psychrobacter* spp. which was discussed to cause disease in hemolymph of hypoxic *C. magister* [47]. On many individuals several strains of *Tenacibaculum* were found. Some of those strains cause diseases in fish aquaculture [48, 49]. Other strains are associated to crustaceans and are capable of chitin degradation [50]. This high number of genera with potential chitinolytic activity and the similarity to several cases of shell disease in other crustaceans shows that the epibiotic biofilm is populated by a vast amount of bacteria ready to erode the carapaces of *C. pagurus* during stressful environmental conditions. Biofilms may often harbor opportunistic pathogenic bacteria [51]. Under certain conditions for example high water temperature pathogens may arise with negative consequences for the host such as shown for coral bleaching derived from bacterial communities [52, 53].

Epibiotic biofilms under different temperature conditions

The severity of shell disease in North American lobster populations are discussed to be associated to high bottom water temperature in certain areas [1]. This was obviously not the case for *C. pagurus* in our study given the differences in normal temperature conditions at the Lofoten and the southern North Sea around Hegloland and the similarity in shell disease prevalence and severity. *C. pagurus* needs to be adapted to either environmental conditions. Although environmental conditions were very different, they have a similar likelihood of getting shell disease in both populations. We used a shift in temperature of 10 °C in the natural temperature range for *C. pagurus* changing the ambient temperature conditions in Germany to those in Norway and vice versa in order to see if this may affect shell disease or microbial communities. We stopped our experiments after 120 h. However, this had no effect on shell disease prevalence or severity in either population. Vogan et al. [54] grew bacteria harvested from shell disease lesions in *C. pagurus* at 15 °C on chitin media under 25 °C and found the saturation in growth for some bacteria between 5 and 8 days as well

as the highest chitinase production. They stated that under natural conditions growth as well as chitinase activity may be much more increased according to earlier publications on the interaction of chitinolytic bacteria with crustaceans [55, 56]. This fast utilization of chitin by microbes has also been shown in *Clostridium*, which was even increased by microbial community interactions [43, 44]. Still our experiments were probably too short for any lesion development or alteration of already present lesions. Although, we could not see an effect on the host itself, we observed a significant decrease in alpha diversity in bacterial communities from warm to cold conditions. These changes in communities ultimately did not progress to a high similarity among samples from one treatment group on the community level as shown by multivariate statistics. The reason for this was maybe the high individuality in biofilms at different hosts we already observed in wild samples. Another reason may be, that the treatment for 5 days probably was too short to lead to a stable fully shifted community structure. Temperature increase, however, led to changes in the community composition by changing abundance of taxa in both populations. Communities did change in a similar way by increased abundances of Gammaproteobacteria as well as a small increase of Deltaproteobacteria and contrarily a decrease in Flavobacteriaceae with higher temperature.

The changes were most prominent for Gammaproteobacteria. This was due to an increase of *Cocleimonas* spp. in warm conditions in Norway and in a similar way a decrease by cold treatment in Germany. *Cocleimonas* spp. is chemoheterotrophic and is capable of sulfur-oxidation at anoxic conditions [40]. This may indicate oxygen low environmental conditions during warm water conditions beneficial for this taxon. Decreased abundance during warm water temperature was found in the high abundant group of Flavobacteriaceae. *Tenacibaculum* spp. is a genus of bacteria with several candidates for heavy fish diseases [48, 57] and being discussed as one source for shell disease in the American lobster [2]. Those taxa were decreased in warm water conditions for both populations. Species found in our study were most similar to different species also found in the American lobster such as *T. maritinum* or *T. soleae*. Those species are strictly aerobic [58]. This may be a reason for a decreased abundance in high water temperature conditions. Deltaproteobacteria many of those adapted to anoxic conditions or even being strict anoxic [59] were found more abundant under warm conditions. Additionally, taxa from *Nitrospira* and *Nitrosomonas* were increased, both indicating the presence of ammonia. Ammonia originates from decomposition of animal or plant material [60–62] and due to physiological reasons by the host crab itself [63]. Anoxic conditions may lead to shell disease as it was discussed for the micro-flora in the biofilm of *H. americanus* [64]. This may be a reason for increased shell disease of species inhabiting waste disposal sites [65, 66]. Many different other bacterial taxa being partly or even strict anoxic were found in biofilm communities among others members of Firmicutes such as *Clostridium* spp. [67]. Within the Deltaproteobacteria, *Desulfothalea* spp., Deltasulfobacterales or Deltasufomonadales were present. Additionally, taxa with anoxic members such as Chromatiales [68] or Phycisphaeraceae [69] were part of the community. Strictly anoxic and aerobic species were found in every population and within one individual community. This may be a sign for very different micro-habitat conditions within the biofilm favored by thick biofilms and the specific external morphology of carapaces of *C. pagurus*. During warm water periods anoxic conditions may arise in biofilms due to increased oxygen consumption [70]. This may favor denitrifying bacteria [71] and sulfate reduction [70]. Dense biofilms between the extrusions under high nutrient and temperature conditions are very likely to become anoxic. In fact shell disease is discussed to often emerge around seta on the carapace [72, 73] which lay deep in those valleys in case of *C. pagurus*. The epibiotic biofilm may act as a harbor for many pathogens and chitinolytics. This may heavily trigger future shell disease outburst in the face

of environmental conditions, due to climate change.

The associated biofilm community of *C. pagurus* is a combination of biofilms similar to those found on other crustaceans and environmental bacteria from sediments. In the community many chitinolytic and potential pathogenic bacteria were present. In combination with high inter-molt intervals of the host, this may result in observed high prevalences of shell disease in populations regardless of the latitude and temperature conditions. Several authors stated that shell disease syndrome may be the result of the natural epibiotic biofilm [74, 75]. For *C. pagurus* in any case shell disease may easily emerge from natural biofilms, since even bacterial taxa very similar to those accused for epizootic shell disease in North American lobster are present. This may lead to similar scenarios as for the American lobster in North America if the right conditions are met [76]. Temperature seems to have the potential to alter the community structure in biofilms on *C. pagurus*, to what degree pathogens causing shell disease may benefit cannot be answered yet. In future studies the effect of environmental alterations will be important to monitor over longer periods than we did in our experiments. This should be done for individuals rather than for populations, as the individuality of biofilms may constrain the results. Furthermore, it may be very important to consider differences in micro-habitat conditions on carapaces. As we found many anoxic and aerobic taxa within the small area of 3.5 cm^2, we expect a structure of sub-communities probably reacting to alterations of environmental factors much more heavily than the total community. Furthermore, in these micro-habitats sampling shell diseased and non shell diseased areas may help to understand what environmental conditions may lead to heavier disease situations on the long run. Benthic crustaceans will face strong alterations in local environmental conditions by climate change [77]. According to our results biofilms on *C. pagurus* may react by a change on the community level and may act as refuge for pathogens. As a result, the prevalence of diseases is expected to increase in the near future.

Supporting Information

S1 Text

Bacterial community analysis by 454 GS Junior Next Generation Sequencing. Bacterial community analysis was performed with the primer pair 27f (TCAGAGTTTGATCCTGGCTCAG) and 338r (CATGCTGCCTCCCGTAGGAGT) amplifying the V1 of 16S bacterial rDNA [19]. Chimera primer for 454 GS Junior sequencing were designed by adding sequencing adapter A and B (Roche), the standard sequencing starter TCAG and a specific MID to each locus specific primer. Either of the specific developed MIDs from Roche was implemented for bar-coding different samples in one run. DNA was extracted from Q-tip by a method adapted from Kahn & Cerniglia [78]. A piece of the Q-tip was taken and added to 200 µl of 10 % of Chelex (Biorad) in a 1.5 ml Eppendorf bin by vortexing and then incubating at 56 °C for 10 minutes. 200 µl of 0.2 % of Triton-X-100 (Applichem) was added to the mixture of Chelex (Biorad) and Q-tip and was incubated for 10 minutes at 100 °C and finally cooled on ice. After centrifugation at 12,000 rpm for 3 minutes 2.5 µl of Chelex-Triton-DNA mixture was used as template for PCR. PCR was conducted using FastStart High Fidelity Taq (Roche) with a total volume of 50 µl containing 2.5 µl template DNA, 35.5 µl ddH2O, 0.5 µl DMSO, 5 µl FastStart Buffer including 18 mmol MgCl, 2.5 µl of either forward and reverse primer (10 µmol), 1 µl dNTP (10 µmol each) and 0.5 µl HighFidelityTaq (5 U/µl). PCR conditions were set as follows: Prior denaturation at 95 °C for 2 min followed by 30 cycles of denaturation at 95 °C for 1 min, annealing at 56 °C for 1 min and elongation at 72 °C for 1

min and a final elongation step of 72 °C for 6 min. PCR products were checked on integrity and on contamination in controls via 2 % agarose gel electrophoresis using GelGreen stain. Bands were excised from the gel and purified following the standard protocol in Quiagen Gel Extraction Kit (Quiagen). PCR products were finally harvested using 10 µl ddH2O. The amount of DNA was then measured using Picogreen stain compared with a molecular weight standard on Tecan infinite F200 Pro (Tecan). The PCR products were again purified using magnetic beads (Beckman Coulter) to avoid any contamination from primer dimers or remaining salts. The purified MID-tagged-PCR products were equimolarly pooled and diluted with TE-Buffer to a final concentration $1 * 10^{-7}$ $\frac{molecules}{\mu l}$ for 454 sequencing reaction. 454 Sequencing was performed at a 454 GS Junior Next generation sequencer (Roche) according to the standard protocol using Lib-A chemistry. Raw data were analyzed by GS-Suite (Roche) and extracted as raw reads with quality information, trimmed for sequencing adapters, low quality and short reads. Raw reads were analyzed using qiime ("Quantitative Insights Into Microbial Ecology") pipeline [20] including amplicon denoising [21] and cleaning for PCR chimeras using ChimeraSlayer [22]. Pooled reads were split for MIDs and trimmed for MIDs as well as locus specific primer on both sites. De-multiplexed raw reads were used by qiime to build operational taxonomic units (OTUs) i.e. sequences which are 97 % similar by usearch clustering approach (version 6.1) [23]. Phylogenetic inference of microbial OTUs was calculated using PyNast aligner [20]. In order to estimate whether the read number was appropriate to cover the community complexity in biofilms, a rarefaction analysis was performed calculating alpha diversity from OTU table using the increase of phylogenetic diversity by number of reads for each treatment. Samples were rarefied to 4500 reads per sample to remove sampling depth heterogeneity. Statistics comparing communities between treatments were based on ANOSIM and adonis statistics using qiime [22] either based on weighted or unweighted UniFrac dissimilarities. UniFrac estimates account for phylogenetic relations in OTUs when calculating dissimilarities [79] using either taxa information weighted by abundance (weighted UniFrac) and not weighted by abundance i.e. presence/absence data (unweighted UniFrac) [80]. For each cluster a representative OTU was picked and used for assigning the most precise taxonomic identifier possible using blast routines with cut-off of 90 % similarity and e-value of 10^{-3} [24] against SILVA111 database as reference [25]. In order to compare sequences found similar to shell disease pathogens on other species we used ARB neighbor joining method to build a tree on published 16s sequences for these groups including bootstrap support by 1000 replicates [26]. We then added the OTUs found in our study as well as the published sequences for shell disease pathogens to more accurately determine phylogenetic classification for these taxa [81].

Sampling and experimental setup. On board of the *RV* Uthörn samples from bacterial communities from wild animals were taken caught with traps at Germany Steingrund. Animals, either caught with traps or during diving operations, were kept in semi-closed recirculation systems at Germany marine station and Bodø marine station supplied with ambient water and temperature conditions for at least one week. After acclimation animals were used for temperature experiments. At Germany crabs were kept at ambient water temperature of 15 °C during June 2010, thereafter animals were slowly cooled down in an aerated 10 l aquarium within 5 h to 5 °C and were kept under this conditions for 120 h. Water was partly exchanged with pre-cooled water every 24 h. At Norway similar ambient water temperature compared to Germany experimental conditions temperature persisted. Water temperature was experimentally increased from 7.5 °C ambient temperature to 16 °C temperature within 10 h; animals were kept under these condition 120 h. Four individuals were sampled for either conditions Norway-Ambient-Temperature (7.5 °C), Germany-Ambient-

Temperature (15 °C), Norway-Experimental-Warm (16 °C), Germany-Experimental-Cold (5 °C), Germany-Wild (15 °C). Biofilm samples from the carapaces were taken from these four groups after each experiment as well as from control animals and wild animals. The surface of the upper left, the upper right and the central anterior area of the carapace was gently rinsed with filtered and sterilized marine water (Fig. 1. A sterilized Q-tip was taken to sample the surface in an area of roughly 2 cm diameter 1. Samples were frozen in −20 °C for storage. For each specimen prior to and after the experiment pictures were taken from the external carapace for later determination of shell disease prevalence. As a reference pictures of shell disease prevalence was additionally taken for more than 50 individuals caught during that time for Germany and Norway population, as well as from a population of *C. pagurus* sampled close to Germany at the "Tiefe Rinne" (54° 5' 15.0" N, 7° 32' 28.3" E) in early spring in 2010 after the long cold water period during winter. Several individuals were also sampled at different locations in the North Sea in December 2009 for shell disease analysis.

Acknowledgments

We thank the Alfred Wegener Institute for Polar and Marine Research at Germany for hosting during sampling *Cancer pagurus*. We also grateful thank Dr. Isabel Schmalenbach, Michael Jahnke and Prof. Dr. Franke for their support during the stay at Helgoland (Germany). Sampling *C. pagurus* would have not been possible without the great support of the crews of the research vessels *RV* Heincke and *RV* Uthörn. We also thank Jan Knott and Sebastian Fuhrmann for their assistance during sampling. For technical assistance we would like to thank Elke Frahmann and for technical assistance during 454 sequencing as well as many helpful discussions we would like to thank Eike Mayland-Quellhorst. For helpful discussions on the manuscript we would like to thank Prof. Thorsten Brinkhoff. For helpful discussions on statistics we would like to thank Dr. Bernd Wemheuer. We thank Prof. Galice Hoarau and Dr. Alexander Jüterbock for the opportunity to work at Norway on *C. pagurus*. We would like to thank the European Science Foundation funding parts of the project with a travel grant (Grant EX/3204).

References

1. Castro KM, Cobb JS, Gomez-Chiarri M, Tlusty M (2012) Epizootic shell disease in American lobsters *Homarus americanus* in southern New England: past, present and future. Diseases of Aquatic Organisms 100: 149–158.
2. Quinn RA, Metzler A, Tlusty M, Smolowitz RM, Leberg P, Chistoserdov AY (2012) Lesion bacterial communities in American lobsters with diet-induced shell disease. Diseases of Aquatic Organisms 98: 221–233.
3. Quinn RA, Metzler A, Smolowitz RM, Tlusty M, Chistoserdov AY (2012) Exposures of *Homarus americanus* shell to three bacteria isolated from naturally occurring epizootic shell disease lesions. Journal of Shellfish Research 31: 485–493.
4. Quinn RA, Cawthorn RJ, Summerfield RL, Smolowitz R, Chistoserdov AY (2013) Bacterial communities associated with lesions of two forms of shell disease in the American lobster (*Homarus americanus*, Milne Edwards) from Atlantic Canada. Canadian Journal of Microbiology 59: 380–390.
5. Vogan CL, Rowley AF (2002) Dynamics of shell disease in the edible crab *Cancer pagurus*: a comparative study between two sites on the gower peninsula, south wales, UK. Diseases of Aquatic Organisms 52: 151–157.
6. Costa-Ramos C, Rowley AF (2004) Effect of extracellular products of *Pseudoalteromonas atlantica* on the edible crab *Cancer pagurus*. Applied and Environmental Microbiology 70: 729–735.
7. Porter L (2004) The Microbiology and Pathology of Shell Disease in the Florida Spiny Lobster, *Panulirus argus* with a Comparison to Shell Disease in the American Lobster, *Homarus americanus*. Electronic theses, The Florida State University.

8. Meres NJ, Ajuzie CC, Sikaroodi M, Vemulapalli M, Shields JD, Gillevet PM (2012) Dysbiosis in epizootic shell disease of the American lobster (*Homarus americanus*). Journal of Shellfish Research 31: 463–472.
9. Food and Agriculture Organization of the United Nations (FAO). Species fact sheet *Cancer pagurus* (Linnaeus, 1758). URL `http://www.fao.org/fishery/species/2627/en`. Last visited 04.05.2015.
10. Woll AK, van der Meeren GI, Fossen I (2006) Spatial variation in abundance and catch composition of *Cancer pagurus* in norwegian waters: biological reasoning and implications for assessment. Ices Journal of Marine Science 63: 421–433.
11. Bullis R, Leibovitz L, Swanson L, Young R (1988) Bacteriologic investigation of shell disease in the deep-sea red crab, geryon-quinquedens. Biological Bulletin 175: 304–304.
12. Castro KM, Angell TE (2000) Prevalence and progression of shell disease in American lobster, *Homarus americanus*, from Rhode Island waters and the offshore canyons. Journal of Shellfish Research 19: 691–700.
13. Sindermann CJ (1991) Shell disease in marine crustaceans a conceptual approach. Journal of Shellfish Research 10: 491–494.
14. Glenn RP, Pugh TL (2006) Epizootic shell disease in American lobster (*Homarus americanus*) in Massachusetts coastal waters: Interactions of temperature, maturity, and intermolt duration. Journal of Crustacean Biology 26: 639–645.
15. Tlusty MF, Metzler A (2012) Relationship between temperature and shell disease in laboratory populations of juvenile American lobsters (*Homarus americanus*). Journal of Shellfish Research 31: 533–541.
16. Chualain CN, Hayes M, Allen B, Robinson M (2009) *Hematodinium* sp. in irish *Cancer pagurus* fisheries: infection intensity as a potential fisheries management tool. Diseases of Aquatic Organisms 83: 59–66.
17. Moullac GL, Haffner P (2000) Environmental factors affecting immune responses in Crustacea. Aquaculture 191: 121–131.
18. Philippart CJM, Anadon R, Danovaro R, Dippner JW, Drinkwater KF, Hawkins SJ, Oguz T, O'Sullivan G, Reid PC (2011) Impacts of climate change on european marine ecosystems: Observations, expectations and indicators. Journal of Experimental Marine Biology and Ecology 400: 52–69.
19. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Research 35: e120.
20. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R (2010) Pynast: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26: 266–267.
21. Reeder J, Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nature Methods 7: 668–669.
22. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methe B, DeSantis TZ, Consortium THM, Petrosino JF, Knight R, Birren BW (2011) Chimeric 16s rrna sequence formation and detection in sanger and 454-pyrosequenced pcr amplicons. Genome Research 21: 494-504.
23. Edgar RC (2010) Search and clustering orders of magnitude faster than blast. Bioinformatics 26: 2460–2461.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.
25. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Gloeckner FO (2013) The silva ribosomal rna gene database project: improved data processing and web-based tools. Nuclear Acids Research 41: D590-D596.
26. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar BA, Lai T, Steppi S, Jobb G, Foerster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Koenig A, Liss T, Luessmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH (2004) Arb: a software environment for sequence data. Nuclear Acids Research 32: 1363–1371.
27. Tishchenko G, Koppova I, Simunek J, Dohnalek J (2010) Extracellular complex of chitinolytic enzymes of clostridium paraputrificum strain j4 separated by membrane ultrafiltration. Folia microbiologica 55: 386-389.
28. Sangkhobol V, Skerman VBD (1981) Chitinophaga, a new genus of chitinolytic myxobacteria. International Journal of Systematic Bacteriology 31: 285-293.

29. Bowman JP, McCammon SA, Nichols DS, Skerratt JH, Rea SM, Nichols PD, McMeekin TA (1997) *Shewanella gelidimarina* sp. nov. and *Shewanella frigidimarina* sp. nov., novel antarctic species with the ability to produce eicosapentaenoic acid (20:5 omega 3) and grow anaerobically by dissimilatory fe(iii) reduction. International Journal of Systematic Bacteriology 47: 1040-1047.

30. Powell A, Rowley AF (2005) Unchanged prevalence of shell disease in the edible crab *Cancer pagurus* four years after decommissioning of a sewage outfall at Langland bay, UK. Diseases of Aquatic Organisms 68: 83–87.

31. Ravenschlag K, Sahm K, Pernthaler J, Amann R (1999) High bacterial diversity in permanently cold marine sediments. Applied and Environmental Microbiology 65: 3982-3989.

32. Matsumoto A, Kasai H, Matsuo Y, Omura S, Shizuri Y, Takahashi Y (2009) *Ilumatobacter fluminis* gen. nov., sp. nov., a novel actinobacterium isolated from the sediment of an estuary. Journal of General and Applied Microbiology 55: 201-205.

33. Dziallas C, Grossart HP, Tang KW, Nielsen TG (2013) Distinct communities of free-living and copepod-associated microorganisms along a salinity gradient in Godthabsfjord, west Greenland. Arctic, Antarctic and Alpine Research 45: 1–10.

34. Gerdts G, Brandt P, Kreisel K, Boersma M, Schoo KL, Wichels A (2013) The microbiome of North Sea copepods. Helgoland Marine Research 67: 757–773.

35. Muramatsu Y, Takahashi M, Kaneyasu M, Iino T, Suzuki Ki, Nakagawa Y (2010) *Persicobacter psychrovividus* sp. nov., isolated from shellfish, and emended descriptions of the genus *Persicobacter* and *Persicobacter diffluens*. International Journal of Systematic and Evolutionary Microbiology 60: 1735-1739.

36. Nedashkovskaya OI, Kim SB, Lee KH, Bae KS, Frolova GM, Mikhailov VV, Kim IS (2005) *Pibocella ponti* gen. nov., sp. nov., a novel marine bacterium of the family Flavobacteriaceae isolated from the green alga *Acrosiphonia sonderi*. International Journal of Systematic and Evolutionary Microbiology 55: 177-181.

37. Nedashkovskaya OI, Kim SB, Han SK, Rhee MS, Lysenko AM, Falsen E, Frolova GM, Mikhailov VV, Bae KS (2004) *Ulvibacter litoralis* gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the green alga *Ulva fenestrata*. International Journal of Systematic and Evolutionary Microbiology 54: 119-123.

38. Park SC, Choe HN, Baik KS, Seong CN (2013) *Aquimarina gracilis* sp. nov., isolated from the gut microflora of a mussel, *Mytilus coruscus*, and emended description of *Aquimarina spongiae*. International Journal of Systematic and Evolutionary Microbiology 63: 1782-1787.

39. Gillan DC, Ribesse J, de Ridder C (2004) The iron-encrusted microbial community of *Urothoe poseidonis* (Crustacea, Amphipoda). Journal of Sea Research 52: 21–32.

40. Tanaka N, Romanenko L, Iino T, Frolova G, Mikhailov V (2011) *Cocleimonas flava* gen. nov., sp. nov., a gammaproteobacterium isolated from sand snail (*Umbonium costatum*). International Journal of Systematic and Evolutionary Microbiology 61: 412-416.

41. Johnson WP, Sieburth JM, Sastry A, Arnold CR, Doty MS (1971) *Leucothrix mucor* infestation of benthic crustacea, fish eggs, and tropical algae. Limnology and Oceanography 16: 962–969.

42. Gooday GW (1990) Physiology of microbial degradation of chitin and chitosan. Biodegradation 1: 177–190.

43. Pel R, Hessels G, Aalfs H, Gottschal JC (1989) Chitin degradation by *Clostridium* sp. strain 9.1 in mixed cultures with saccharolytic and sulfate-reducing bacteria. FEMS Microbiology Letters 62: 191–200.

44. Pel R, Gottschal JC (1989) Interspecies interaction based on transfer of a thioredoxin-like compound in anaerobic chitin-degrading mixed cultures. FEMS Microbiology Letters 62: 349–357.

45. Chistoserdov A, Quinn RA, Gubbala SL, Smolowitz R (2009) Various forms and stages of shell disease in the American lobster share a common bacterial pathogen in their lesions. Journal of Shellfish Research 28: 689–689.

46. Chistoserdov AY, Smolowitz R, Mirasol F, Hsu A (2005) Culture-dependent characterization of the microbial community associated with epizootic shell disease lesions in American lobster, *Homarus americanus*. Journal of Shellfish Research 24: 741–747.

47. Scholnick DA, Haynes VN (2012) Influence of hypoxia on bacteremia in the dungeness crab, *Cancer magister*. Biological Bulletin 222: 56–62.

48. Wakabayashi H, Hikida M, Masumura K (1986) *Flexibacter maritimus* sp nov., a pathogen of marine fishes. Internaional Journal of Systematic Bacteriology 36: 396–398.

49. Hansen GH, Bergh O, Michaelsen J, Knappskog D (1992) *Flexibacter ovolyticus* sp. nov., a pathogen of eggs and larvae of Atlantic halibut, *Hippoglossus hippoglossus* L. International Journal of Systematic Bacteriology 42: 451-458.
50. Sheu SY, Lin KY, Chou JH, Chang PS, Arun AB, Young CC, Chen WM (2007) *Tenacibaculum litopenaei* sp nov., isolated from a shrimp mariculture pond. International Journal of Systematic and Evolutionary Microbiology 57: 1148–1153.
51. Wahl M, Goecke F, Labes A, Dobretsov S, Weinberger F (2012) The second skin: ecological role of epibiotic biofilms on marine organisms. Frontiers In Microbiology 3: 292.
52. Bourne D, Iida Y, Uthicke S, Smith-Keune C (2008) Changes in coral-associated microbial communities during a bleaching event. ISME Journal 2: 350–363.
53. Sussman M, Willis BL, Victor S, Bourne DG (2008) Coral pathogens identified for white syndrome (ws) epizootics in the indo-pacific. Plos One 3: e2393.
54. Vogan CL, Costa-Ramos C, Rowley AF (2002) Shell disease syndrome in the edible crab, *Cancer pagurus* - isolation, characterization and pathogenicity of chitinolytic bacteria. Microbiology-sgm 148: 743–754.
55. Dietrich MA, Hackney CR, Grodner RM (1984) Factors affecting the adherence of *Vibrio cholerae* to blue crab (*Callinectes sapidus*) shell. In Vibrios in the Environment. Wiley, New York. 601–611.
56. Hood MA, Meyers SP (1977) Microbiological and chitinoclastic activities associated with *Penaeus setijerus*. Journal of the Oceanographical Society of Japan 33: 235–241.
57. Avendano-Herrera R, Toranzo AE, Magarinos B (2006) *Tenacibaculosis infection* in marine fish caused by *Tenacibaculum maritimum*: a review. Diseases of Aquatic Organisms 71: 255–266.
58. Choi DH, Kim Y, Hwang CH, Yi H, Chun J, Cho BC (2006) *Tenacibaculum litoreum* sp. nov., isolated from tidal flat sediment. Internaional Journal of Systematic and Evolutionary Microbiology 56: 635–640.
59. Kuever J, Rainey FA, Widdel F (2005) Family II Desulfobulbaceae. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM (eds) Bergey's manual of systematic bacteriology, vol 2, 2nd edn, (The Proteobacteria), part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria). Springer, New York.
60. Waksman SA, Carey CL, Reuszer HW (1933) Marine bacteria and their role in the cycle of life in the sea I. decomposition of marine plant and animal residues by bacteria. Biological Bulletin 65: 57–79.
61. Watson SW, Bock E, Valois FW, Waterbury JB, Schlosser U (1986) *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. Archives of Microbiology 144.
62. Suzuki I, Dular U, Kwok SC (1974) Ammonia or ammonium ion as substrate for oxidation by *Nitrosomonas europaea* cells and extracts. Journal of Bacteriology 120: 556–558.
63. Durand F, Regnault R (1998) Nitrogen metabolism of two portunid crabs, *Carcinus maenas* and *Necora puber*, during prologned air exposure and subsequent recovery: a comparative study. The Journal of Experimental Biology 201: 2515–2528.
64. Bell SL, Allam B, McElroy A, Dove A, Taylor GT (2012) Investigation of epizootic shell disease in American lobsters (*Homarus americanus*) from Long Island Sound: I. characterization of associated microbial communities. Journal of Shellfish Research 31: 473–484.
65. Young RR (1991) Prevalence and severity of shell disease among deep-sea red crabs *Chaceon quinquedens* Smith 1879 in relation to ocean dumping of sewage sludge. Journal of Shellfish Research 10: 499–504.
66. Sawyer TK (1991) Shell disease in the Atlantic rock crab *Cancer irroratus* Say 1817 from the northeastern United States. Journal of Shellfish Research 10: 495–498.
67. Fendrich C, Hippe H, Gottschalk G (1990) *Clostridium halophilium* sp. nov. and *C. litorale* sp. nov., an obligate halophilic and a marine species degrading betaine in the Stickland reaction. Archives of Microbiology 154: 127–132.
68. Imhoff J (2005) Chromatiales ord. nov. In: Brenner D, Krieg N, Staley J, Garrity G, Boone D, De Vos P, Goodfellow M, Rainey F, Schleifer KH, editors, Bergey's Manual® of Systematic Bacteriology, Springer US. pp. 1-59.
69. Fukunaga Y, Kurahashi M, Sakiyama Y, Ohuchi M, Yokota A, Harayama S (2009) *Phycisphaera mikurensis* gen. nov., sp. nov., isolated from a marine alga, and proposal of Phycisphaeraceae fam. nov., Phycisphaerales ord. nov. and *Phycisphaerae classis* nov. in the phylum Planctomycetes. Journal of General and Applied Microbiology 55: 267–275.

70. Arnosti C, Jorgensen BB, Sagemann J, Thamdrup B (1998) Temperature dependence of microbial degradation of organic matter in marine sediments: polysaccharide hydrolysis, oxygen consumption, and sulfate reduction. Marine Ecology Progress Series 165: 59–70.

71. Kawagoshi Y, Fujisaki K, Tomoshige Y, Yamashiro K, Qiao Y (2012) Temperature effect on nitrogen removal performance and bacterial community in culture of marine anammox bacteria derived from sea-based waste disposal site. Journal of Bioscience and Bioengineering 113: 515–520.

72. Hsu AC, Smolowitz RM (2003) Scanning electron microscopy investigation of epizootic lobster shell disease in *Homarus americanus*. Biological Bulletin 205: 228–230.

73. Prince DL, Bayer RC, Loughlin M (1993) Etiology and microscopy of shell disease in impounded American lobsters, *Homarus americanus*. Bulletin of the Aquaculture Association of Canada 93: 87–89.

74. Porter L, Butler M, Reeves RH (2001) Normal bacterial flora of the spiny lobster *Panulirus argus* and its possible role in shell disease. Marine and Freshwater Research 5.

75. Rosen B (1970) Shell disease of aquatic crustaceans. Snieszko, Stanislas F (edited By) American Fisheries Society Special Publication, No 5 A Symposium On Diseases of Fishes and Shellfishes Viii + 526p Illus American Fisheries Society: Washington, DC, USA : 409–415.

76. Tlusty MF, Smolowitz RM, Halvorson HO, DeVito SE (2007) Host susceptibility hypothesis for shell disease in American lobsters. Journal of Aquatic Animal Health 19: 215–225.

77. Stocker T, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

78. Kahn A, Cerniglia C (1997) Rapid and sensitive method for the detection of *Aeromonas caviae* and *Aeromonas trota* by polymerase chain reaction. Lett Appl Microbiol 24: 233-239.

79. Lozupone C, Knight R (2005) Unifrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71: 8228-8235.

80. Lozupone AC, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental Microbiology 73: 1576–1585.

81. Shinzato N, Watanabe I, Meng XY, Sekiguchi Y, Tamaki H, Matsui T, Kamagata Y (2007) Phylogenetic analysis and fluorescence in situ hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. Microbial Ecology 54: 627–636.

Manuscript II

Shell disease in crabs in the North Sea: Native *Carcinus maenas* and *Cancer pagurus* suffer but bioinvasive *Hemigrapsus penicillatus* flourishes

in form of a manuscript

Shell disease in crabs in the North Sea: Native *Carcinus maenas* and *Cancer pagurus* suffer but bioinvasive *Hemigrapsus penicillatus* flourishes

Running title: Shell disease in North Sea brachyurans.

Abstract

Bacterial mediated shell disease has increased in many crustaceans such as lobster and has been reported worldwide for crustaceans as different as brachyurans, crangonids or even parasitic fish lice; although, its origin has not been clarified yet; ocean warming or pollution is discussed. To understand changes in future crustacean biodiversity we compared shell disease prevalence in native species *Carcinus maenas* and *Cancer pagurus* and the invasive Asian brushcrab *Hemigrapsus penicillatus*. Despite living under similar ecological conditions no shell disease was found in *H. penicillatus*, while medium and high shell disease prevalence was observed in *C. maenas*, and *C. pagurus*. Disease occurrence was additionally positively correlated with body size in *C. maenas* and *C. pagurus*. Specific characteristics of the outer cuticle may lead to differences in susceptibility. Scanning electron microscopy showed a pronounced irregular shape of the outer cuticle for *C. pagurus* and *C. maenas* benefiting fouling bacteria and metazoans. In contrast *H. penicillatus* had a smooth plane shape less colonized by fouling organisms. Differences in susceptibility for shell disease may be one reason for the success of the invasive *H. penicillatus* in the southern North Sea. Considering the severity of shell disease crustacean diversity in the North Sea may change with consequences for the community dynamics and food web interactions.

Keywords: competition, disease prevalence, black spot syndrome, biodiversity shifts

Introduction

Especially in the Northern Atlantic, numerous species of crustaceans suffer from different forms of shell disease [1]. Reports about the severity of shell disease in different species is biased by their economically importance; however, brachyuran crabs seem to be especially hit [2, 3, 1, 4, 5] as well as the American lobster (*Homarus americanus*) [6]. Various forms of the syndrome has been described beginning with brown spots to blackened areas or pits on carapaces and completely eroded cuticle areas especially in epizootic shell disease [7, 8]. These different stages of the disease are often difficult to distinguish and may represent different forms of shell disease. The disease outbreak is probably caused by different environmental factors [4, 9] as well as host conditions such as immunodeficiency, injuries, shell hardness and high susceptibility probably based on the genotypic background [10, 9, 7]. Blackening and erosion at carapaces appear to be a result of chitinolytic bacteria [11]. Many bacterial taxa harvested from lesions had been identified, i.e. strains from *Aeromonas*, *Pseudoaeromonas* or *Flaveobacter* [10, 12, 13]. The blackening of lesions itself is a host mediated defense against invading bacteria by melanization [14]. Finally, hosts die from internal infections [15] or by decreasing competitive ability towards con-specifics or defense against predators.

Susceptibility for shell disease is assumed to differ among species [1] and even among populations of the same species for instance in *C. pagurus* from different locations [4]. It has been shown that within populations of *C. maenas* prevalence of shell disease seems to be seasonally stable, but differs between sexes [1]. In lobster *H. americanus*, shell disease is prevalent in the southern part of lobster distribution (south of Cape Cod) but seems to spread northward along the US east coast [16, 9] probably due to shifts in environmental conditions. However, it is still unknown whether specific habitats or geographic regions are disease hot spots and whether different species are equally hit in similar habitats.

We compared the prevalence of the disease in three abundant benthic brachyuran crabs; two are indigenous species and widely distributed in the North Sea, *C. maenas* and *C. pagurus*. The third species *H. penicillatus* (synonymous *H. takanoi* [17]) has successfully invaded the North Sea first recorded in Europe in 1994 [18] and in the Wadden Sea in 2007 [19]. All three crabs share habitats at different live stages especially in coastal hard substrates and may compete on space or food in local habitats [20]. We included the invasive species to test the hypothesis that community alterations driven by competition often favors the invasive species since local predators, parasites and pathogens are more adapted to native species [21] than to invasive species (enemy release hypothesis [22]). In addition we aimed to understand whether different susceptibility might be caused by differences in the carapace morphology and different degree of biofouling.

Material and Methods

Sample sites

Cancer pagurus was collected at the rocky inter-tidal at Helgoland Island, at a sandy to boulder habitat east of Helgoland Island (Steingrund), at the deep muddy to shelly channel depression in the outermost Elbe estuary (Helgoländer Tiefe Rinne) and at different locations of the North Sea (Fig. 1). *Carcinus maenas* was collected from the Jade area close to Memmert Island and from swashes close to Juist and Langeoog Island (Fig. 1). *Hemigrapsus penicillatus* as well as *C. maenas* were collected from inter-tidal *Crassostrea*-reefs south of Juist Island and Langeoog (Fig. 1). Specimens of *C. maenas* and *H. penicillatus* were sacrificed by freezing immediately after catching specimens of *C. pagurus* were either frozen for later analysis or parameters were documented and pictures of carapaces were made. For each individual size, gender and presence of biofouling organisms were recorded.

Shell disease analysis

Analysis of the external carapaces was used to compare shell disease prevalence among species. We identified and quantified shell disease using a stencil of the outer carapace according to [23]. The stencil was subdivided into equal sized areas and prevalence of the disease was calculated from all individuals of either species [24] in order to find hot spots for shell disease emergence on the carapace. Blackening of carapaces was classified as follows: 1 - early to medium signs of shell disease (melanization of areas - black or sometimes reddish - on the surface of the carapace differing from background coloration often with root like or halo like melanization in inner layers of the carapace; distinct - point like - erosion of outer cuticula), 2 - heavy shell disease (massive erosion through carapace layers with inner infection of soft tissues), 3 - blackening due to injuries. Shell disease was analyzed from frozen specimen in the lab for *C. maenas*, *H. penicillatus* and *C.*

pagurus as well as from pictures for *C. pagurus*. Assessing shell disease prevalence from pictures and from frozen individuals was exemplary compared in two subsamples of the Area "Tiefe Rinne" and was found not significant by two-sample Wilcoxon test ($W = 764, p = 0.32$). Additionally, the gills and the inner site of carapaces were analyzed to determine whether external shell disease corresponded to inner infections of soft tissues. Within species, the effect of sex, size and location on the prevalence of outer shell disease was modeled using analysis of deviance including fitting the model by backwards selection with $\alpha = 0.05$ to the most parsimonious model [24].

Morphological analysis

To analyze the morphology of outer cuticle for each species and compare morphological characteristics among *C. pagurus*, *C. maenas* and *H. penicillatus* we used light microscopy and scanning electron microscopy (SEM). Similar sized pieces from the epibranchial region of carapace, one exemplary for each species, were excised from the carapace. Tissue fluid was replaced with ethanol using a series of dilution. Pieces of carapaces were placed into small perforated bins and incubated for ten minutes in different concentrations of ethanol as follows: 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %. Finally, samples were stored in 99 % ethanol over night. Using a BAL-TEC CPD 030 Critical Point Dryer at 10 °C the samples were dried in 96 % ethanol. Ethanol was substituted with CO_2 (31 °C, 73 bar) in $15 - 18$ similar steps. Samples were coated with a thin layer of gold by $2 - 4$ kV among high vacuum. Pictures were taken with a magnification range of $20x - 20000x$ in a scanning electron microscope (HITACHI S-3200N) for further analysis.

Results

Shell disease prevalence

To determine and compare the prevalence of shell disease in the three species we considered all animals except those of category 3 (blackening due to injuries). Shell disease prevalence was highest in populations of *C. pagurus* (85 % overall), medium in *C. maenas* (25 % overall) and not present in *H. penicillatus* (0 % overall). χ^2 probability test showed significant differences among species ($df = 2, \chi^2 = 186.95, p < 0.001$; Table 1). Outer shell disease occurrence correlated with size in *C. pagurus* and *C. maenas* i.e. bigger specimen showed more signs of shell disease than smaller ones and the smallest species of the study *H. penicillatus* did not show any sign of shell disease (Fig. 2). For further statistical analysis individuals were pooled in populations based on close sampling locations which most likely interact environmentally by tidal currents and genetically by panmictic populations (Fig. 1). We could not detect differences in shell disease in *C. pagurus* between the combined population at Helgoland and remaining North Sea samples using Kruskal Wallis test ($\chi^2 = 0.65, df = 1, p = 0.42$). *C. maenas* was differently diseased in the populations Memmert, Langeoog, Wilhelmshaven and Mellum; individuals from Langeoog and Mellum showed smaller median shell disease prevalence than population at Mellum and Wilhelmshaven using Kruskal Wallis test ($\chi^2 = 34.86, df = 3, p < 0.001$). Negative binomial distribution was fitted to describe shell disease prevalence in *C. pagurus* and in *C. maenas* at Wilhelmshaven and Memmert Island. Modeling revealed significantly more symptoms of shell disease in bigger size classes for *C. pagurus* (Table 2). In *C. maenas* for both populations the most parsimonious model included size as significant correlating with shell disease prevalence with bigger specimens showing more shell disease

(Table 2). Additionally, for Memmert Island female individuals showed more shell disease than males which contributed strongly to the model (Table 2).

A subsample of *C. pagurus* and *C. maenas* exemplary was analyzed for internal infections and gill blackening. In *C. pagurus* 75 % ($N = 12$) were found with internal infections of carapaces and 92 % with gill blackening ($N = 12$). In *C. maenas* internal infections (31.3 %, $N = 214$) and gill blackening (14.5 %, $N = 213$) were less frequent; in *H. penicillatus* no gill blackening and internal infections were observed. Epibionts were only present on *C. pagurus* and *C. maenas*, both showing a high prevalence of hydrozoans, balanids, bryozoans and polychaetes at carapaces (81 % [*C. pagurus*] $N = 87$; 58 % [*C. maenas*] $N = 269$); however, no macrobenthic epibiont was found on carapaces of *H. penicillatus*.

Morphology of outer cuticle

The outer shape of the epicuticle of all three brachyurans revealed an obvious difference between native *C. pagurus* and *C. maenas* in comparison to invasive *H. penicillatus*. While in both native crabs the outer cuticle was covered with dense projections (see Fig. 3) the outer layer of *H. penicillatus* appeared to be relatively smooth lacking any pronounced irregularities. In *C. pagurus* biofouling by hydrozoans and bryozoans on carapaces often grew root-like oriented between these projections. Scanning electron microscopy (SEM) of the epibranchial region of carapaces from frozen specimen showed differences between projections and valleys between projections. While the surface of projections itself was missing a pronounced biofilm, we found a dense mucus layer with bacteria between projections for *C. pagurus* and *C. maenas* (Fig. 3a,b). In *H. penicillatus* pronounced projections could not be observed even with bigger magnification by SEM. Rarely slight elevations were found and a thin mucus and bacterial layer covered most of the area of sampled carapace in *H. penicillatus*.

Discussion

By comparing the abundant native brachyuran crabs *Carcinus maenas*, *Cancer pagurus* and the invasive *Hemigrapsus penicillatus* in coastal habitats of the southern North Sea, we found strong differences in the prevalence of shell disease in the three crab species. While the bioinvasive *H. penicillatus* was not affected by any symptoms of shell disease, *C. pagurus* showed the highest prevalence of the disease; 86 % of animals in all populations were diseased showing a severity which is consistent with findings of [4, 25]. In addition to outer shell symptoms, we found 82 % of *C. pagurus* with additional inner infections and 92 % with blackening of gills. *C. maenas* seemed to be less affected with 22 % showing shell disease and 15 % with gill ulcerations. Shell disease in *C. maenas* had been described for other populations in the North Atlantic, however, gill ulcerations were rarely observed [1]. This shows that internal infections in *C. maenas* was lower than for *C. pagurus*; however, melanization as defense against bacteria was present in all populations of both species.

Different areas had no significant influence on the prevalence of disease in *C. pagurus*; animals sampled at different locations in the North Sea were equally diseased to those sampled at different positions around Helgoland. In *C. maenas* we found a remarkable difference between populations in shell disease prevalence. The population sampled close to Mellum was more healthy in contrast to other locations. We can only speculate if these healthy animals represented a subsample aggregated for reproduction since many individuals were bearing eggs. The fact that shell disease was observed

in different locations indicates that shell disease is not caused by a local pathogen but by commonly present opportunistic bacterial pathogens and may depend more on the situation of the host and its susceptibility to deal with bacterial pressure. Observations of biofilms on shell diseased lobster at the US coast support this conclusion [10, 26].

For both species smaller juvenile individuals were only found at tidal flats close to each island, while adults could only be caught at the sub-tidal areas. This was the reason for example for a very low overall shell disease prevalence in individuals of *C. maenas* from *Crassostrea*-reefs south of Juist Island where no bigger specimen were present. Due to the strong panmictic nature of populations over long distances in *C. maenas* [27] and *C. pagurus* [28], juveniles are most likely genetically not different from adults close to the islands. The strong differences in shell disease prevalence between juveniles and adults may be due to physiological factors alone or additionally due to differences in small scale habitat quality. The latter seems to be less likely, since the different size classes will share tidal habitats during high tides. For both species shell disease prevalence was significantly higher in larger size classes for populations only represented by similar sized big adults in Wilhelmshaven (*C. maenas*) and at Tiefe Rinne (*C. pagurus*) indicating host factors to be more important in shell disease prevalence. Additionally, female individuals of *C. maenas* at Memmert Island were significantly more diseased than males.

Longer inter-moult intervals in larger animals and in females, which moult dependent on mating timing may lead to a higher probability of getting shell disease. A recent publication by Tlusty and colleagues discussed moulting frequency as one major trigger for shell disease prevalence in populations using a modeling approach for *Homarus americanus* [29]. This is supported by our data for brachyurans showing that size of the host is a major factor in shell disease prevalence. Moulting frequency may additionally be a factor for the differences in shell disease prevalence between species showing that the smallest probably most often moulting species did not show shell disease, while the biggest species *C. pagurus* with the longest inter-moult intervals was most affected. However, it is known that under favorable conditions bacteria from shell disease lesions can show maximal chitinase activity after 120-196 h [25] causing chitin consisting carapaces to break down and initiate shell disease lesions within a short time. Although, small *C. maenas* moult frequently per year, we found small individuals with only 1.4 cm carapace width having symptoms of shell disease. This partly contradicts the theory that juveniles are physiologically more resistant towards changing environmental conditions and less susceptible towards diseases than adults [30]. Many animals of both native crab species with relatively small black spots on the outer cuticle had also a severe infection of the inner tissue with a halo of melanization at inner layers of the carapace and bigger areas of necrotic tissue under the carapace (Fig. 4). This observation shows that the health state of native crabs can be much worse than from what is visible at the outer carapace. Even inner organs may be affected [31]. Moulting may not necessarily lead to the loss of shell disease, symptoms at the outside of the carapaces may reappear fast after moulting, which has also been described for the epizootic shell disease in the American lobster *H. americanus* [32, 33, 34].

Differences in shell disease between the two native and the invasive species corresponds with differences of the carapace surface which may lead to differences in fouling activities. While we could not find any macrozootic fouling organisms on *H. penicillatus*, species from hydrozoa, bryozoa and balanids were commonly found on *C. maenas* carapaces and often highly abundant on carapaces of *C. pagurus*. Scanning electron microscopy showed a more structured surface of the shell in *C. pagurus* and *C. maenas* compared to *H. penicillatus*. Interestingly, biofouling by bacterial mats as well as root like growing metazoans such as hydrozoans were frequently found between these

structures of the shell surface. We found a pattern of channel like holes on the surface on all three crabs by scanning electron microscopy. Whether these structures have an effect on biofouling and shell disease in brachyuran crabs remains to be investigated to see if differences in susceptibility to external bacteria may be internally regulated by excreting bioactive compounds on carapaces via carapace channel systems or whether these pores may act as areas benefiting bacterial penetration and fouling. It is known that transcripts with antibacterial compounds play a role when new cuticles are built after moulting [35].

Differences in susceptibility to shell disease of these three benthic crab species may be a major trigger in benthic community dynamics of brachyuran crabs and may facilitate the bioinvasion by *H. penicillatus* as well as alterations in abundance of native *C. pagurus* and *C. maenas*. Species composition might change not only due to mortality per se but also due to decreasing competitiveness of native towards invasive and more healthy species. Bioinvasion of *H. penicillatus* nowadays might cause a decline of *C. maenas* in local hard substrate communities, since *H. penicillatus* appears to be a predator especially for small *C. maenas* [36]. Competition on food and habitat for brachyurans in benthic communities had been described in various publications for example for *C. maenas* and *H. americanus* [37], *C. maenas* and *Cancer irroratus* [20], *C. maenas* and *H. penicillatus* [36] as well as *C. maenas* and different species of *Hemigrapsus* [38]. The outcome of such competitive events on the population level depends certainly also on the health state of a species. The reason why *H. penicillatus* seems to be not affected by the disease remains unclear. However, since shell disease is a symptom which can be caused by bacterial communities rather than by specific disease agents an immunity towards shell disease by *H. penicillatus* is very unlikely. In the same way we do not expect an “immunity” towards metazoan fouling. It remains to be investigated whether this species is generally less susceptible to fouling organisms or if the enemy release hypothesis holds for the bioinvasion of *H. penicillatus*. This hypothesis suggests that invasive species may lack common diseases or parasites in new habitats [22]. [39] for example described that in *H. penicillatus* no parasites had been found at newly invaded areas. Similarly, no parasites were found on invading *C. maenas* in North America [40]. Interestingly, we found one individual of *H. penicillatus* with heavy signs of shell disease in a subsample of 13 individuals caught in their native habitats in Japan. However, since shell disease is present in newly invaded habitats in other crustacean species the ”enemy release” hypothesis would only be fully verified if certain pathogens would be the specific source of shell disease. Many investigations assume that an opportunistic chitinolytic community of bacteria is the disease agent rather than a single specific strain [41, 42]. However, there are investigations which found some bacteria highly abundant in shell disease lesions of different forms such as *Epsilonproteobacteria* and *Tenacibaculum* [26] as well as *Thalassobius* and *Aquimarinus* [10]. In principle shell disease should be more seen as a symptom - as an indicator of unfavorable conditions for the host (host susceptibility hypothesis [9]).

Shell disease as a globally appearing disease in crustaceans may have the potential in a fast changing environment with global environmental anomalies by climate change to dramatically change local crustacean communities with major impact on the food web. Monitoring local host susceptibility to such diseases is thus important in future investigations.

Acknowledgement

We thank the Alfred Wegener Institute for Polar and Marine Research at Helgoland for hosting during sampling *Cancer pagurus*. We also grateful thank Dr. Isabel Schmalenbach, Michael Jahnke

and Prof. Dr. Franke for their support during the stay at Helgoland and in sampling *C. pagurus*. Sampling *Carcinus maenas* and *C. pagurus* would have not been possible without the great support of the crews of the research vessels FK Senckenberg, *RV* Heincke, *RV* Uthörn *RV* Solea. We also thank Edith Markert, Torsten Janssen and Sebastian Fuhrmann for their assistance in sampling. We thank Dr. Thomas Glatzel from University of Oldenburg and David de Leeuw from David de Leeuw Muschelzucht GmbH for organizing and catching *Carcinus maenas* at the Jade Channel. We also thank Edith Kieselhorst from University of Oldenburg for her technical support during scanning electron microscopy. This is a contribution to the Biodiversity and Climate research Centre (BiK-F) funded by Loewe Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Excellenz of Hesses Ministry of Higher Education, research and Arts.

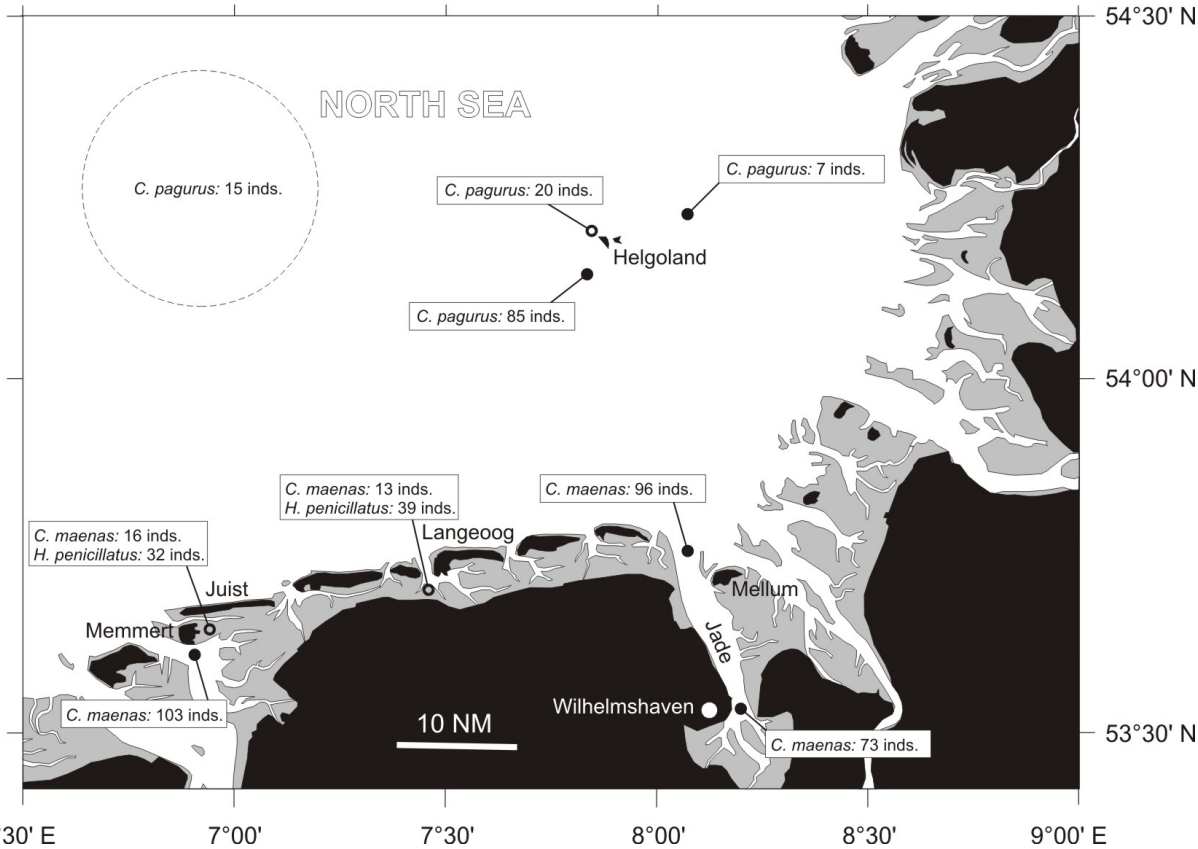


Figure 1. Map of sampling areas within the German Bight. Number of individuals sampled on each location is given for *Carcinus maenas*, *Cancer pagurus* and *Hemigrapsus penicillatus*. Open circles: inter-tidal reef sites, closed circles: sub-tidal sites.

References

1. Comely CA, Ansell AD (1989) The occurrence of black necrotic disease in crab species from the west of scotland. Ophelia 30: 95–112.
2. Joseph FRS, Ravichandran S (2012) Shell diseases of brachyuran crabs. Journal of Biological Sciences 12: 117–127.

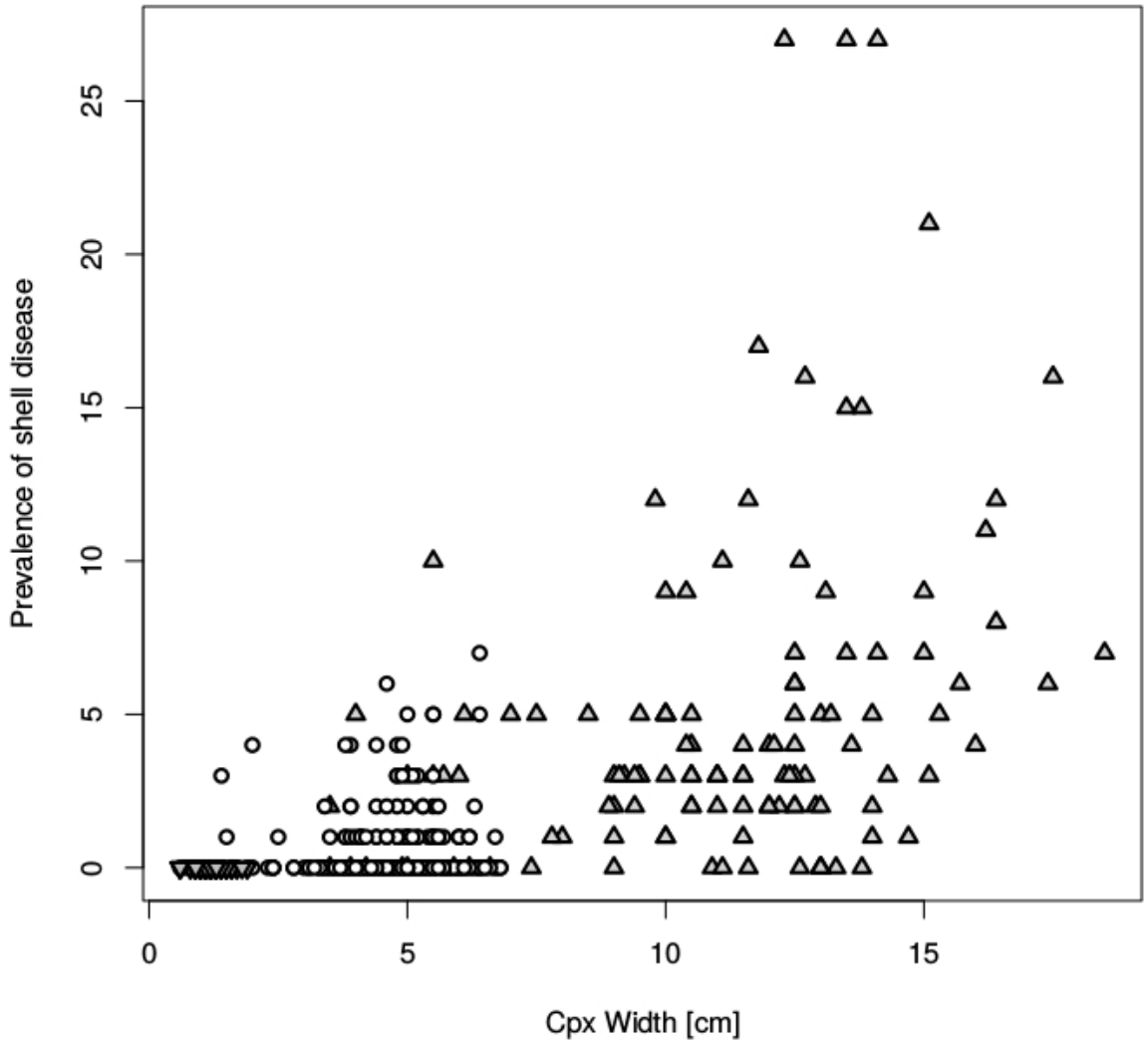


Figure 2. Prevalence of shell disease on external carapace correlated with size of host. Gray triangles upwards show *Cancer pagurus*, open circles indicate *Carcinus maenas* and *H. penicillatus* is given by gray triangles downwards.

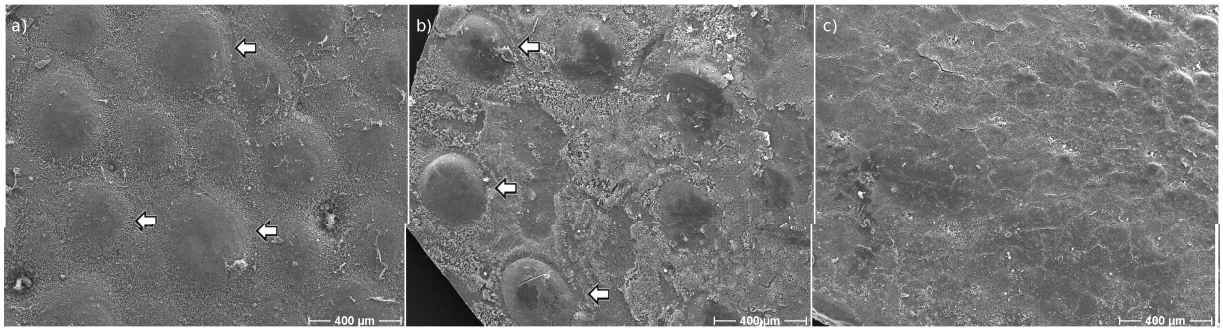


Figure 3. Scanning electron microscopy of the epibranchial region of outer cuticle for a) *Cancer pagurus*, b) *Carcinus maenas* and d) *Hemigrapsus penicillatus*. White arrows indicate projections on each carapace surface except for *H. penicillatus*.

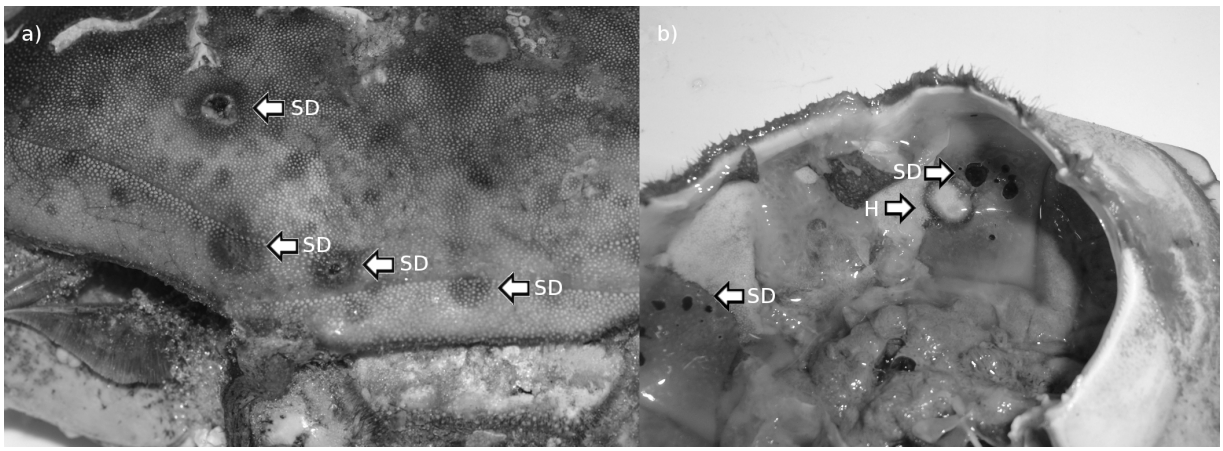


Figure 4. Shell disease is exemplary illustrated (white arrows; SD) for the outer cuticle at the anterior carapace of *C. pagurus* a). Illustration b) shows black necrotic ulcerations at shell diseased spots from the inner tissues (SD) of *C. pagurus* and a halo of melanization (H) originating from cuticle penetration which can be found frequently at severe stages.

Table 1. Summary of population parameters for *Cancer pagurus*, *Carcinus maenas* and *Hemigrapsus penicillatus*. Prevalence of shell disease, gender distribution and carapace width are shown for different types of habitat.

habitat	inter-tidal	sub-tidal	total
C. pagurus			
Carapace width [cm]	5.4 ± 1.3	11.9 ± 2.4	10.9 ± 3.3
Gender [m/f]	15/5	45/60	60/65
S.D. prevalence	44 %	91 %	85 %
C. maenas			
Carapace width [cm]	1.6 ± 0.7	4.9 ± 0.9	4.6 ± 1.3
Gender [m/f]	25/4	191/81	216/85
S.D. prevalence	14 %	26 %	25 %
H. penicillatus			
Carapace width [cm]	1.2 ± 0.3	NA	1.2 ± 0.3
Gender [m/f]	45/26	0/0	45/26
S.D. prevalence	0 %	NA	0 %

Table 2. Most parsimonious models based on analysis of deviance on number of shell disease occurrences on outer carapaces. Modeled are host size and gender for *Cancer pagurus* and *Carcinus maenas* at different locations.

Species	estimate	z	deviance	df	p
C. pagurus (overall)					
Intercept	-0.04	-0.11	163	121	0.90
Size	0.14	4.91	139	120	***
C. pagurus (Tiefe Rinne)					
Intercept	0,21	0.40	101	84	0.69
Size	0.11	2.50	95	83	*
C. maenas (Memmert)					
Intercept	-3.76	-3.62	123	118	***
Size	0.53	2.87	122	117	**
Gender	1.82	4.59	96	116	***
C. maenas (Wilhelmshaven)					
Intercept	-3.65	-2.33	60	71	*
Size	0.67	2.18	56	70	*

$*** = 0.001, ** = 0.01, * = 0.05$

3. Vazquez-Lopez H, Rocha-Ramirez A, Chazaro-Olvera S, Castillo-Baranco DR, Rodriguez-Valez AC, Cruz-Gomez A (2012) Shell disease in *Callinectes rathbunae* Contreras, 1930, parasitized by *Loxothylacus texanus* Boschma, 1933. Journal of Fisheries and Aquatic Science 7: 412–421.
4. Vogan CL, Llewellyn PJ, Rowley AF (1999) Epidemiology and dynamics of shell disease in the edible crab *Cancer pagurus*: a preliminary study of Langland Bay, Swansea, UK. Diseases of Aquatic Organisms 35: 81–87.
5. Andersen LE, Norton JH, Levy NH (2000) A new shell disease in the mud crab *Scylla serrata* from Port Curtis, Queensland (Australia). Diseases of Aquatic Organisms 43: 233–239.
6. Castro KM, Somers BA (2012) Observations of epizootic shell disease in American lobsters, *Homarus americanus*, in southern New England. Journal of Shellfish Research 31: 423–430.
7. Homerding M, McElroy A, Taylor G, Dove A, Allam B (2012) Investigation of epizootic shell disease in American lobsters (*Homarus americanus*) from Long Island Sound: II. immune parameters in lobsters and relationships to the disease. Journal of Shellfish Research 31: 495–504.
8. Smolowitz R, Chistoserdov AY, Hsu A (2005) A description of the pathology of epizootic shell disease in the American lobster, *Homarus americanus*, H. Milne Edwards 1837. Journal of Shellfish Research 24: 749–756.
9. Tlusty MF, Smolowitz RM, Halvorson HO, DeVito SE (2007) Host susceptibility hypothesis for shell disease in American lobsters. Journal of Aquatic Animal Health 19: 215–225.
10. Quinn RA, Metzler A, Tlusty M, Smolowitz RM, Leberg P, Chistoserdov AY (2012) Lesion bacterial communities in American lobsters with diet-induced shell disease. Diseases of Aquatic Organisms 98: 221–233.
11. Sindermann CJ (1989) Shell disease in marine crustaceans a conceptual approach. Journal of Shellfish Research 8: 461–461.
12. Allam B, Dove A, McElroy A, Taylor G, Fast M (2009) Characterization of microbial communities and immune system responses associated with shell disease in the American lobster, *Homarus americanus*. Journal of Shellfish Research 28: 677–678.
13. Bell SL, Allam B, McElroy A, Dove A, Taylor GT (2012) Investigation of epizootic shell disease in American lobsters (*Homarus americanus*) from Long Island Sound: I. characterization of associated microbial communities. Journal of Shellfish Research 31: 473–484.
14. Soederhaell K (1982) Phenoloxidase activating system and melanization a recognition mechanism of arthropods? a review. Developmental and Comparative Immunology 6: 601–611.
15. Vogan CL, Costa-Ramos C, Rowley AF (2001) A histological study of shell disease syndrome in the edible crab *Cancer pagurus*. Diseases of Aquatic Organisms 47: 209–217.
16. Glenn RP, Pugh TL (2006) Epizootic shell disease in American lobster (*Homarus americanus*) in Massachusetts coastal waters: Interactions of temperature, maturity, and intermolt duration. Journal of Crustacean Biology 26: 639–645.
17. Markert A, Raupach MJ, Segelken-Voigt A, Wehrmann A (2014) Molecular identification and morphological characteristics of native and invasive asian brush-clawed crabs (Crustacea: Brachyura) from Japanese and German coasts: *Hemigrapsus penicillatus* (de Haan, 1835) versus *Hemigrapsus takanoi* Asakura & Watanabe 2005. Organism, Diversity and Evolution 14: 369–382.
18. Dauvin JC, Rius AT, Ruellet T (2009) Recent expansion of two invasive crabs species *Hemigrapsus sanguineus* (de Haan, 1835) and *H. takanoi* Asakura and Watanabe 2005 along the Opal Coast, France. Aquatic Invasion 4: 451–465.
19. Obert B, Herlyn M, Grotjahn M (2007) First records of two crabs from the North West Pacific *Hemigrapsus sanguineus* and *H. takanoi* at the coast of Lower Saxony, Germany. Wadden Sea Newsletter 1: 21–22.
20. Matheson K, Gagnon P (2012) Effects of temperature, body size, and chela loss on competition for a limited food resource between indigenous rock crab (*Cancer irroratus* Say) and recently introduced green crab (*Carcinus maenas* L.). Journal of Experimental Marine Biology and Ecology 428: 49–56.
21. Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, McCauley DE, O'Neil P, Parker IM, Thompson JN, Weller SG (2001) The population biology of invasive species. Annual Review of Ecology and Systematics 32: 305–332.
22. White TA, Perkins SE (2012) The ecoimmunology of invasive species. Functional Ecology 26: 1313–1323.

23. Vogan CL, Rowley AF (2002) Dynamics of shell disease in the edible crab *Cancer pagurus*: a comparative study between two sites on the gower peninsula, south wales, UK. Diseases of Aquatic Organisms 52: 151–157.

24. R Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

25. Vogan CL, Costa-Ramos C, Rowley AF (2002) Shell disease syndrome in the edible crab, *Cancer pagurus* - isolation, characterization and pathogenicity of chitinolytic bacteria. Microbiology-sgm 148: 743–754.

26. Quinn RA, Cawthorn RJ, Summerfield RL, Smolowitz R, Chistoserdov AY (2013) Bacterial communities associated with lesions of two forms of shell disease in the American lobster (*Homarus americanus*, Milne Edwards) from Atlantic Canada. Canadian Journal of Microbiology 59: 380–390.

27. Domingues CP, Creer S, Taylor M, Queiroga IH, Carvalho GR (2010) Genetic structure of *Carcinus maenas* within its native range: larval dispersal and oceanographic variability. Marine Ecology Progress Series 410: 111-123.

28. Ungfors A, McKeown NJ, Shaw PW, Andre C (2009) Lack of spatial genetic variation in the edible crab (*Cancer pagurus*) in the Kattegat-Skagerrak area. ICES Journal of Marine Science 66: 462–469.

29. Tlusty M, Kim A, Castro K (2014) Modeling shell disease in American lobster (*Homarus americanus*) as individual-based health trajectories. Canadian Journal of Fisheries and Aquatic Sciences 71: 808-813.

30. Melzner F, Gutowska MA, Langenbuch M, Dupont S, Lucassen M, Thorndyke MC, Bleich M, Poertner HO (2009) Physiological basis for high CO2 tolerance in marine ectothermic animals: pre-adaptation through lifestyle and ontogeny? Biogeosciences 6: 2313–2331.

31. Comeau M, Benhalima K (2009) Internal organ pathology of wild American lobster (*Homarus americanus*) from eastern Canada affected with shell disease. New Zealand Journal of Marine and Freshwater Research 43: 257–269.

32. Castro KM, Cobb JS, Gomez-Chiarri M, Tlusty M (2012) Epizootic shell disease in American lobsters *Homarus americanus* in southern New England: past, present and future. Diseases of Aquatic Organisms 100: 149–158.

33. Landers DF (2005) Prevalence and severity of shell disease in American lobster *Homarus americanus* from eastern Long Island Sound, Connecticut. Aquatic Forim Series, New England Aquarium, Boston, MA, 94-97 pp.

34. Castro KM (2005) Update on shell disease in American lobsters in southern New England waters. Journal of Shellfish Research 24.

35. Terwilliger NB, Dangott L, Ryan M (1999) Cryptocyanin, a crustacean molting protein: Evolutionary link with arthropod hemocyanins and insect hexamerins. Proceedings of the National Academy of Sciences of the United States of America 96: 2013-2018.

36. van den Brink AM, Wijnhoven S, McLay CL (2012) Competition and niche segregation following the arrival of *Hemigrapsus takanoi* in the formerly *Carcinus maenas* dominated Dutch delta. Journal of Sea Research 73: 126–136.

37. Haarr ML, Rochette R (2012) The effect of geographic origin on interactions between adult invasive green crabs *Carcinus maenas* and juvenile American lobsters *Homarus americanus* in Atlantic Canada. Journal of Experimental Marine Biology and Ecology 422: 88–100.

38. Jensen GC, McDonald PS, Armstrong DA (2002) East meets west: competitive interactions between green crab *Carcinus maenas*, and native and introduced shore crab *Hemigrapsus* spp. Marine Ecology Progress Series 225: 251–262.

39. McDermott JJ (2011) Parasites of shore crabs in the genus *Hemigrapsus* (Decapoda: Brachyura: Varunidae) and their status in crabs geographically displaced: a review. Journal of Natural History 45: 2419–2441.

40. Blakeslee AMH, Keogh CL, Byers JE, Kuris AM, Lafferty KD, Torchin ME (2009) Differential escape from parasites by two competing introduced crabs. Marine Ecology Progress Series 393: 83–96.

41. Porter L (2004) The Microbiology and Pathology of Shell Disease in the Florida Spiny Lobster, *Panulirus argus* with a Comparison to Shell Disease in the American Lobster, *Homarus americanus*. Electronic theses, The Florida State University.

42. Meres NJ, Ajuzie CC, Sikaroodi M, Vemulapalli M, Shields JD, Gillevet PM (2012) Dysbiosis in epizootic shell disease of the American lobster (*Homarus americanus*). Journal of Shellfish Research 31: 463–472.

Chapter 2

Statistical backgrounds to analyze local adaptation based on population genetics

2.1 Introduction

Population genetic theory offers some basic models to study the genetic profile of populations in order to estimate ecological and evolutionary processes. Many estimates are derived directly from allele frequencies. For diploid organisms the genotype distribution in populations is given by the Hardy-Weinberg equilibrium (52; 53; 54) and for a diallelic locus the probabilities can be calculated based on allele frequencies p and q as follows:

$$(p+q)^2 = p^2 + 2pq + q^2 = 1 \tag{2.1}$$

p = frequency of allele p; q = frequency of allele q

This equation can be generalized for multiallelic loci by adding more frequencies to the equation $[(p+q+r+s...)^2 = 1]$. The model, however, assumes an ideal population being in panmixia; with no mutation, selection or migration altering the allele distribution. Additionally, the size of the population is indefinitely large though we expect no effects of genetic drift on allele frequencies over generations (52). If all assumptions of the model are matched, the Hardy-Weinberg equilibrium will be stable over generations (52). However, whether genotype frequencies may follow these expected probabilities or not may have different meanings and may be central for any estimation diversity in or between demes and their biological meaning.

2.1.1 Allele diversity in demes

In diploid organisms one allele of each locus derives from a single parent. Being homozygous or heterozygous in loci is dependent on the genetic composition of parents as described by classical Mendelian genetics. On the population level the proportion of heterozygotes and homozygotes in demes can be used to estimate different population parameters by F-statistics. This was described by Sewell Wright (55; 56) and further developed in particular by Masatoshi Nei (57; 58). Inbreeding can cause a deficiency of heterozygotes in a population in relation to the expected frequency of these genotypes given by Hardy-Weinberg equilibrium. This is more generally described by F-statistics as the inbreeding coefficient or fixation index:

$$F = 1 - \frac{O(pq)}{E(pq)}. \tag{2.2}$$

F = inbreeding coefficient; $O(pq)$ = observed frequency of genotype pq; $E(pq)$ = expected frequency of genotype pq assuming Hardy-Weinberg equilibrium.

$O(pq)$ is the observed genotype frequency and $E(pq)$ the expected one. This forms the basis of calculations on fixation and diversity of demes (55; 56). In order to analyze population structure, the inbreeding coefficient can be divided into three dimensions. F_{IT} as the inbreeding of an individual relative to the total population, F_{IS} as inbreeding of the individual in relation to its subpopulation and F_{ST} as the relation of inbreeding of the population in consideration to the total population (55; 56).

$$1 - F_{IT} = (1 - F_{IS}) * (1 - F_{ST}) \tag{2.3}$$

F = inbreeding coefficient; subscripts I, S and T determine the relation of inbreeding. This is done by comparing two dimensions in ascending order either the individual (I) with sub-population (S) or the total-population (T) or the sub- (S) with the total-population (T).

The extension of F-statistics for multiallelic situations were further developed by Nei (57) and Nei and Chesser (58) based on heterozygosities. An alternative approach was developed by

Weir and Cockerham (59) as an estimate based on variance components. However, Cockerham (60; 61) had shown that this approach is similar to Wrights equations.

$$f = \frac{F - \theta}{1 - \theta} \tag{2.4}$$

$F \hat{=} F_{IT}$ and describes the correlation of genes within individuals; $\theta \hat{=} F_{ST}$ and describes the correlation of genes between individuals; $f \hat{=} F_{IS}$ describes the correlation of genes in individuals of a population (59).

A deficiency of heterozygotes in demes may have different biological reasons. By using F_{IS} or F_{IT} it is possible to calculate the relation of the observed frequency of heterozygotes with the heterozygotes which we expect in one deme or over all demes from allele frequencies. A deficiency in one deme might indicate kin structures in this population. Strong population structures may then emerge, due to co-ancestry in populations, i.e. the probability to be related by decent (62). In order to further estimate intra-population structure by relatedness several estimates had been developed for example by Blouin et al. (63) or Queller and Goodnight (64). The former refers to the Mendelian codominant inheritance of alleles. Whether sharing alleles indicates common ancestors and true relatedness depends on the frequency of alleles in other demes, i.e. the reference. For example, if a reference population shows a high frequency of alleles C and D and only low frequencies of A and B, the sampled patch of individuals with alleles A and B may represent a kin group. Blouin's approach models mean relatedness of full and half siblings over loci based on the reference population, in order to differentiate between identity by state and identity by descent in pairwise individual comparisons (63). More referring to allele diversity, Queller and Goodnight (64) developed an equation for estimating kinship accounting for a bias, due to variability in allele frequencies. Comparing individuals with shared and different alleles in their genotypes may lead to a biased estimated relation derived from total allele diversity in populations. Thus, for an estimate on relatedness for infinite populations the amount of differently related individuals may be modeled based on reference populations, or directly implementing a bias-correction for the estimator, as had been done in Queller and Goodnight's r_{xy} (64).

$$r_{xy,l} = \frac{0.5(I_{ac} + I_{ad} + I_{bc} + I_{bd}) - p_a - p_b}{1 + I_{ab} - p_a - p_b} \tag{2.5}$$

I_{a-d} = allele identities I of individual alleles $a - d$ in locus l; p_{a-b} = frequencies of allele a or b in reference populations.

Analyzing inter-individual genetic relatedness can be very important to understand the evolution of cooperative behavior or explain fine spatial structure of populations. For instance, in marine species with larval dispersal it has been shown that larvae disperse in kin groups, which may be one reason for genetic patchiness in adult populations (65). Empirical studies from different aquatic species underpin the importance of relatedness on genetic population structure such as for perch (66), lobsters (67) or barnacles (68).

2.1.2 Allele diversity between demes

In order to estimate connectivity among populations we face different situations when analyzing allele diversities. Alleles derived from loci may have different frequencies, dependent on genetic properties of each locus. In detail, loci may be monomorphic with just one allele or polymorphic with different alleles. These characteristics may mirror the selection pressure on the locus in a continuum of highest to lowest selection pressure with increasing allele diversity or the amount

of fixation in alleles, due to genetic drift. Whether a locus in consideration is under selection is important to understand the biological meaning of population structures in a species.

Studying markers assuming a neutral model Neutral markers have the potential to show population structure. In many cases we expect a certain degree of isolation between demes, which may lead to sympatric, parapatric, peripatric or allopatric populations (69). These terms describe the degree of isolation in a geographic way. Sympatric are two demes, which are isolated but share the same geographical area. Parapatric are demes which are still genetically and geographically connected, but do differ over a continuum with highest differences between the edges. Allopatry and peripatry refers to demes which are geographically distinct. While allopatry describes two similar sized demes, peripatry describes a situation where one deme is much smaller than the other (69). Those definitions may additionally describe secondary processes. For example, demes, which got geographically isolated and got differentiated, are described as allopatric. If those populations may later overlap in their distribution this would be called secondary sympatric. By theory each deme on its own remains in Hardy-Weinberg equilibrium, but with increasing time alleles might vary in frequency due to genetic drift or selection processes.

For neutral markers, the variation in allele frequencies is truly random and can be described by several models. The Wright-Fisher model (70) is used to calculate the frequency of alleles given that there is no overlap between generations, population size is finite and the allele frequencies are multinomially distributed (71). In principal this model will describe a decrease in allele diversity in demes (71). Since the direction of the drift should be random in case of neutrality, a difference between allele composition between demes, i.e. fixation in demes for different alleles, is most likely. To describe the differences between demes the simplest way is to calculate the absolute mean difference in allele frequencies (equation (2.6)).

$$D = \frac{\sum_{k=1} |p_{ik} - p_{jk}|}{2} \tag{2.6}$$

p_{ik} = allele frequency of allele k in population i; p_{jk} = allele frequency of allele k in population j

However, the equation 2.6 may only describe the situation correctly when finite populations are considered. Any estimation done for infinite populations is biased and may produce a first category error. Given that there are two finite genetically identical populations with 20 alleles of the same frequency and we randomly draw a sample from either population of size X, then we may observe a bias relative to the size of the sample *vs.* the size of the finite population. In Fig. 2.1 we see an exponential fit of this relation with bias decreasing with increasing sample size. Additionally, the absolute number of alleles per populations is important; with decreasing allele diversity the bias decreases. This has been pointed out by Rogers (72) describing the sampling bias dependent on heterozygosity for different distance estimates. Manhattan distance, which is related to equation 2.6 ($\sum_{k=1} |p_{ik} - p_{jk}|$), is heavily biased with heterozygosity.

Furthermore, the authors describe that the probability of two alleles being identical in populations as $\sum p^2$. Their expected distance is then $1 - \sum p^2$, which equals the heterozygosity in this allele (72). To deal with the bias in differentiation estimates based on heterozygosity, it is better described as a difference in heterozygosities in populations. This follows the general rule of dealing with differentiation as an estimate based on alpha and beta diversities as done for several estimators used in ecology (73). In population genetics heterozygosities can be divided into mean intra-population (H_S) and inter-population heterozygosity (H_T) and its bias corrected estimates ($\hat{H}_S$, $\hat{H}_T$) (58).

There are different estimators using these classical estimates of heterozygosity. The most famous one is F_{ST} (55) and its derived estimator for multiple alleles G_{ST} (57). Those estimates

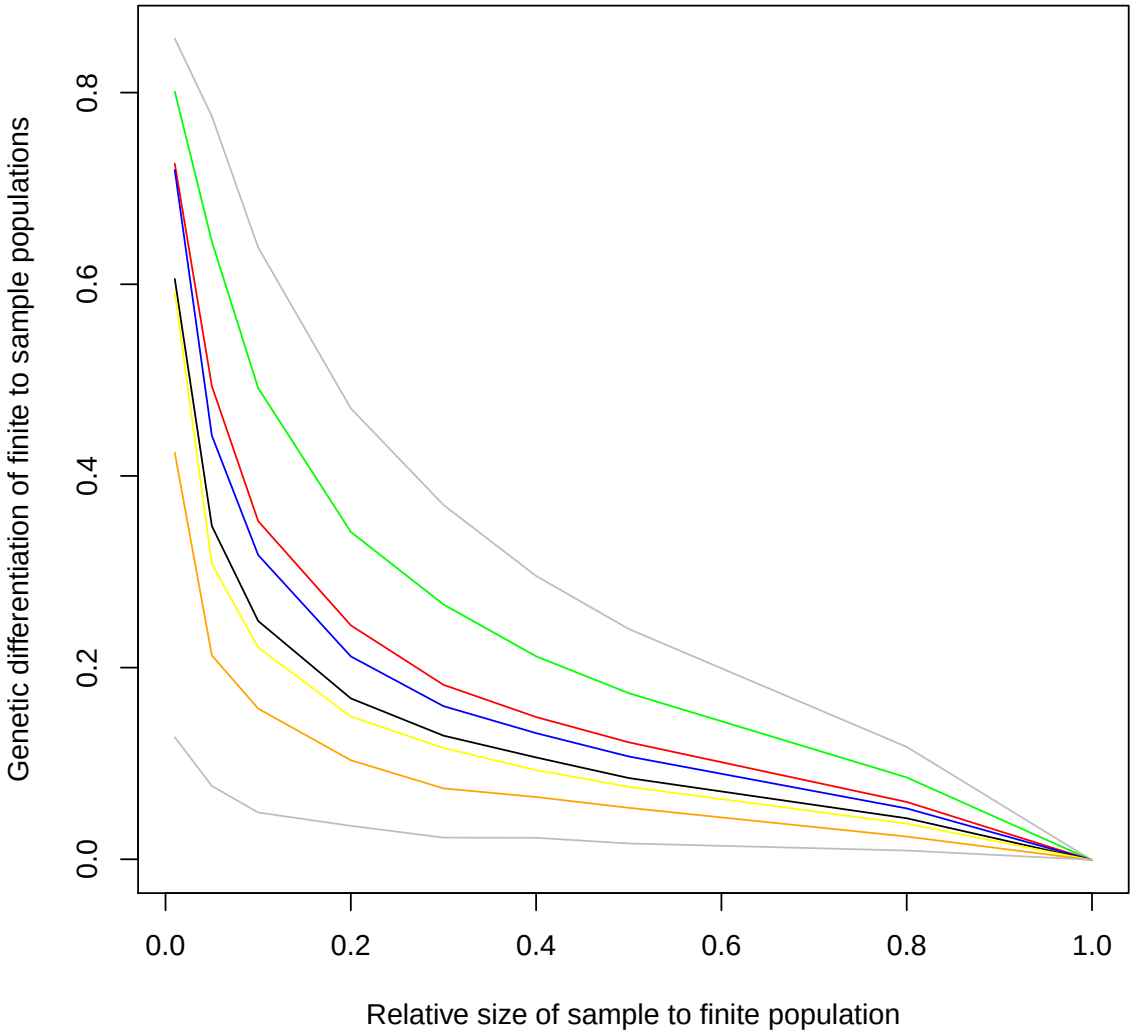


Figure 2.1: Bias of absolute mean differentiation in relation to relative sample size. On the y-axis the difference of the absolute mean differentiation calculated from two finite populations to the difference calculated based on a sample taken from the same populations is plotted. On the x-axis the relative size of the sample taken from the finite population is indicated. Line colors indicate different numbers of alleles in the finite ancestral population *cyan* = 200, *green* = 100, *red* = 50, *blue* = 40, *black* = 25, *yellow* = 20, *orange* = 10, *gray* = 2.

$$H_S = \sum_{j=1}(1 - \sum_{i=1} p_{ij}^2)$$
$$H_T = 1 - \sum_{i=1} p_i^2 \tag{2.7}$$

p_{ij} = frequency of allele i in population j; p_i = frequency of allele i in all populations

$$D_{ST} = H_S - H_T \tag{2.8}$$

H_S = heterozygosity over subpopulations; H_T = heterozygosity among all populations (57)

$$\hat{H}_S = H_S[\frac{2\bar{N}}{(2\bar{N}-1)}]$$
$$\hat{H}_T = H_T + [\frac{H_S}{(2\bar{N}n)}] \tag{2.9}$$

H_S = heterozygosity over subpopulations; H_T = heterozygosity among all populations; $\bar{N}$ = harmonic mean of population size; n = number of populations

have a long tradition and were originally developed for differences in protein sequences. Quite recently, there started an intense discussion on these estimates. F_{ST} and G_{ST} do not provide a stable estimate of population differentiation; they are dependent on the total allele diversity in populations and will decrease in maximal mathematically possible value as allele diversity increases (74; 75; 76). However, this can also be advantageous since F_{ST}, respectively G_{ST}, is an estimator for fixation of allele composition between demes (57). Due to this, it is a good estimate for analyzing genetic drift or differences in selection. During genetic drift allele diversities decrease differently in demes, while during events of selection only few alleles will be positively selected and remain in the population. Both processes will decrease absolute allele diversity in a population (55; 56). However, scientists claim for estimators of population differentiation, which may be used as an estimate for a percentage genetic difference of populations. Recent attempts proposed two promising approaches. In 2005 an estimator called G'_{STmax}, which is derived from G_{ST}, was developed by (75). The principal idea is that the estimated G_{ST} between demes is divided by the absolute mathematically possible G_{ST} based on total allele diversity. This approach will bring the value in a scale of $0-1$, and, according to Hedrick (75), performs well as an estimate for the absolute differentiation in the finite population.

$$\hat{G}_{ST} = \frac{\hat{H}_T - \hat{H}_S}{\hat{H}_T} \tag{2.10}$$

$\hat{H}_S$ = bias corrected heterozygosity over subpopulations; $\hat{H}_T$ = bias corrected heterozygosity among all populations

$$\hat{G}'_{STmax} = \frac{\hat{G}_{ST}}{\hat{G}_{STmax}} \tag{2.11}$$

$\hat{G}_{ST}$ = bias corrected multilocus equivalent of inbreeding coefficient F_{ST}; $\hat{G}_{STmax}$ = maximum possible $\hat{G}_{ST}$ (75)

In equation 2.10 we see that the difference of H_T and H_S divided by H_T may lead to extreme low values in case of high heterozygosities (76). Equation 2.11 deals with this upper limit of

G_{ST} by rescaling the estimator with the maximum possible G_{ST} for each situation G_{STmax}. Another approach was proposed by Jost (74), which is called D. Still, this estimate is based on alpha and beta diversity, but with different effects than for F_{ST} derived estimators.

$$D_{EST} = [\frac{\hat{H}_T - \hat{H}_S}{1 - \hat{H}_S}][\frac{n}{n-1}] \tag{2.12}$$

$\hat{H}_S$ = bias corrected heterozygosity over subpopulations; $\hat{H}_T$ = bias corrected heterozygosity among all populations; n = number of populations (74)

The $1 - H_S$ in the denominator scales the difference of alpha and beta diversity, i.e. the difference of H_S and H_T by the maximal possible value of $H_T - H_S$ which is $1 - H_S$. Given the absolute differentiation between finite demes, all estimates G_{ST}, G'_{STmax} and D_{EST} behave their own way. There are still discussions about which estimator to use in general or in certain situations. Leng and Zhang (77) for example compared both methods for a finite island model and neutral loci and found that D is relative sensitive regarding different mutation models, while G_{ST} has its limitations when it comes to high rates of mutation and gene-flow. In practice, Leng and Zhang (77) propose to combine both methods. Wang (78) for example disagrees with Jost's approach to use D as a measure of differentiation. He stated, that it is heavily dependent on the gene diversity of markers and does not correctly describe genetic differentiation dependent on divergence, time or genetic drift as shown by a modeling approach. Wang (78) recommends to still use G_{ST} with caution in situations of high polymorphic markers until a better estimate is developed. Other authors agree to use G_{ST} and F_{ST}. They argued, that both estimates cannot be replaced neither by recently developed D_{EST} (79; 80) nor by G'_{STmax} (80). All these estimates may be appropriate to analyze neutral effects on genetic diversity such as isolation by distance or distinct geographical barriers in certain situations. Still, estimates have to be used with caution and may behave more or less appropriate in specific situation. Additionally, population effects such as bottle necks or founder effects can be analyzed using these estimates. However, not all markers are necessarily neutral. To find out whether markers are really neutral and what they tell about population structure further analyzes are necessary.

Using markers to study selection with emphasis on geographic differentiation In order to analyze local adaptation an approach to quantify phenotype variation between demes is given by the Q_{ST} estimate (81). By using Q_{ST} any phenotype variation can be compared and correlated with genetic variation using F_{ST} derived estimators to study adaptation in demes as described by Edelaar and Bjorklund (82). However, if no phenotype data are available to study adaptation one may use the combination of markers under selection and markers which are neutral to study adaptational effects on genetic variation. Classical F_{ST} derived estimators are very useful when it comes to the analysis of genetic adaptation. Fixation of neutral markers by genetic drift will decrease allele diversity within generations, however, it does not have a directional component towards a certain allele. In contrast, genetic adaptation on local environmental conditions may lead to fixation in a locus under selection or by hitchhiking in linked sequence regions. Fixation due to genetic adaptation may have different dynamics than fixation by genetic drift. G_{ST} values depend on absolute allele diversity. For fixed alleles G_{ST} values can be significantly larger and appear as an outlier compared to those for neutral markers. Different loci with different degrees of fixation will show similar D_{EST} if their differences in allele frequencies between demes are similar. One major theoretical background was again developed by Sewell Wright, known as Wright's Shifting Balance Theory (SBT) (83; 84; 85). According to Wright's theory, there are three phases in the process of selection: I - random genetic drift may deliver genetic profiles affected by selection, II - selection may fix genetic profiles differently in demes, III - interdeme migration from high to low fitness may alter gene frequencies in demes

(83). Wade (85) studied the effect of selection under island model migration and Wrights SBT phase III migration in neutral genetic differentiation. The phase III migration increases F_{ST} even for neutral loci while for island models genetic variation remains low on neutral markers, for example due to individual selection and founder effects (85). Founder effects or bottlenecks represent a situation where effective population size is relatively small, leading to a loss of allele richness (86), most likely a loss of rare alleles (87). In case of loss of allele richness we may expect a decrease of adaptive ability (88). Santos et al. (88) showed that founder effects and bottlenecks by altering gene diversity may have strong effects on the adaptation of the first generations after the event. However, they also stated that laboratory experiments do not provide good estimates for wild populations (88). To study adaptation based on genetic variation it is necessary to understand population history of the species.

2.2 Summary of findings and own contributions to manuscripts

Manuscript III: Demerelate: an R package to calculate inter-individual relatedness for kinship analysis based on codominant diploid genetic markers

In order to analyze intrademe genetic structure I developed Demerelate, an R package to calculate inter-individual relatedness and inbreeding coefficients. The package provides estimates of inbreeding based on F-statistics from (57) and (59). Furthermore, kin structure can be estimated using three different statistics for genetic relatedness. These can be compared with geographic distance to give a first overview on the correlation of genetic relatedness with geographic patchiness.
SOFTWARE: *Demerelate latest version 0.8-1* (89)

Own contribution to the study: The idea to implement the statistics provided by Blouine et al. (63) was given by Prof. Gerlach. The idea to further implement statistics from Nei (57), Weir and Cockerham (59) and Queller and Goodnight (64) was given by me. The package, the manual and the manuscript was written by me. Prof. Gerlach reviewed the text of the manuscript as well as the manual on Demerelate and tested the R-code.

Manuscript IV: Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics!

To further analyze the population structure of species using genetic diversity between populations several estimators can be used. Which estimator serves best in a given situation is still rigorously discussed. We participated to this discussion by modeling neutral genetic differences in populations using a finite island model for different mutation and migration rates. The performance of G_{ST} and D_{ST} is compared for certain situations. We found that while significance is the same for both estimates the maximal possible value of G_{ST} dramatically decreases with mutation rate, while D_{ST} can be any value between 0 and 1, dependent on absolute differentiation. Furthermore, we developed one of the first tools to calculate Jost's D - the R-package DEMEtics. The modeling approach is published in Molecular Ecology (76) in combination with the package which is also published under the R-project.
SOFTWARE: *DEMEtics latest version 0.8-5* (90).

Own contribution to the study: I contributed to the manuscript Gerlach et al., 2010 (76) by writing R tools for the modeling approach, preparing results, reviewing the manuscript, discussing on statistics and the manuscript. I did part of the coding for DEMEtics and reviewed most of the package.

Manuscript III

Demerelate: an R package to calculate inter-individual relatedness for kinship analysis based on codominant diploid genetic markers

in form of a manuscript
R code Demerelate version 0.8-2 published at R CRAN

Demerelate: calculating inter-individual relatedness for kinship analysis based on co-dominant diploid genetic markers using R

Abstract

The *Demerelate* package offers algorithms to calculate different inter-individual relatedness measurements (allele sharing indices [M_{xy}, B_{xy}] and pairwise relatedness within groups [r_{xy}]) to analyze kinship structures within populations based on datasets of co-dominant diploid genetic markers. Statistics are based on randomization tests; modeling relatedness coefficients by logistic regression, modeling relatedness with geographic distance by mantel correlation and comparing mean relatedness between populations using pairwise t-tests. Demerelate provides an advance on previous similar software packages, due to the ability to calculate pairwise relatedness between individuals along with F_{IS}, as well as combining analysis of relatedness and spatial structuring. An UPGMA tree visualizes genetic relatedness among individuals. Additionally, *Demerelate* summarizes information on datasets (allele vs. genotype frequencies; heterozygosity; F_{IS}-values). *Demerelate* is to our knowledge the first R-package implementing different relatedness estimators including Blouin's M_{xy} combined with geographic information. The R environment enable users to better understand relatedness within populations due to the flexibility of *Demerelate* of accepting different datasets as empirical data, reference data, geographical data and by providing intermediate results. Each statistic and tool can be used separately, which helps to understand the suitability of the data for relatedness analysis, and can be easily implemented in custom pipelines.

Keywords: allele sharing, mantel test, genotype sharing, heterozygosity

Running title: Demerelate - calculating pairwise relatedness

Introduction

The rapid development of new sequencing technologies has brought much hope to identify the allelic variants explaining local adaptation and divergence in natural populations. However, it is crucial to differentiate between genetic similarity due to local adaptation or inter-individual relatedness. Kin biased dispersal and settlement in juvenile stages can be one major reason for genetic patchiness in adult populations (Broquet *et al.* 2013). Famous examples for kin structured populations are found in social mammals such as primates (Wikberg *et al.* 2012) or rodents (Stephens *et al.* 2013). Empirical studies from different aquatic species underpin the importance of relatedness on genetic population structure for fish and invertebrates such as for perch (Gerlach *et al.* 2001), reef fish (Buston *et al.* 2009), lobsters (Iacchei *et al.* 2013) or barnacles (Veliz *et al.* 2006). By analyzing pairwise relatedness, we differentiate between genetic similarity by state and genetic similarity by descent (Blouin 2003). Genetic similarity by state is given by measures of similarity based on genotype (M_{xy}) or allele identities (B_{xy}). While the former is strictly based on the Mendelian rule of genetic inheritance, the latter simply counts allele identities between individuals. Both measures may be heavily biased when simply applied to empirical data. In order to be useful as an estimate of genetic similarity by descent, both statistics need to be bias corrected. M_{xy} is corrected with thresholds randomly generated from a reference population (Blouin *et al.* 1996). These thresholds show which degree of similarity can be securely assigned as similarity by descent in contrast to limited

variability in genotypes in certain loci. The allele identity measurement B_{xy} is directly corrected for the influence of allele frequencies by co-analyzing a reference population to calculate r_{xy} as bias corrected measurement for relatedness (Queller and Goodnight 1989). Most important for calculating a good estimate of relatedness is an appropriate reference population. An infinite big parental population would be optimal; however, the reference population often has to be based on the empirical data. If the reference population fully fitted the Hardy-Weinberg-Principle (Stern 1943), and this can be tested in `Demerelate` by calculating F_{IS} (Weir and Cockerham 1984; Nei and Chesser 1983) with corresponding statistics (randomization tests), each generation would be equally appropriate as a reference as long as they are infinitely big and all preconditions for a normal distribution are fulfilled to generate stable allele frequencies over generations. Because this is virtually never the case, all samples in the reference populations, which may bias the estimate, need to be removed. To obtain a neutral reference the population data need to be manually cleaned for any hidden kin or population structure. This can be done by omitting the population which should be tested or any known close relatives of the reference population (Queller and Goodnight 1989). The bigger the reference the less severe is the consequence of removing single samples; however, in small populations existing kin structures may heavily bias the estimate. For a good estimate of relatedness many genetic loci are needed which show some variability in number of alleles. Calculations of relatedness based on loci with small number of alleles may lead to a high similarity by state and high correction of estimates during statistical analysis. These loci have little power in detecting the difference between similarity by state and by descent. Vice versa, loci with many alleles may have high power in detecting relatives as long as they do not violate Hardy Weinberg Equilibrium between generations. Blouin *et al.* (1996) briefly discussed the power of relatedness estimates M_{xy} and r_{xy} in discriminating relatedness categories. They described for r_{xy} the dependence of misclassification of estimates depending on average heterozygosity for loci. When working with numbers of loci less than 20 this may lead to substantial overestimation of the number of siblings (Blouin *et al.* 1996). Therefore, an appropriate selection of markers and a good reference population are key factors to get a reliable estimate on relatedness in populations. In comparison with already available software for estimating relatedness, the strength of `Demerelate` is its flexibility to combine different analyses. Using `Loci.test` combined with `emp.calc` it is possible to get a good overview of the power of the empirical data prior to statistical evaluation. It is crucial to understand if marker variability and the degree of diversity by state may allow for a good estimate of relatedness in populations. The reference population should be analyzed for indications of relatedness as well. By using standard R routines it is easy to omit samples biasing the reference population. `Demerelate` can also analyze the data for relatedness including F statistics, geographic information and plotting the results in one simple step. Furthermore, data can be combined with custom reference populations or information on geographic distance to illustrate relatedness on a spatial scale. Functions may be easily included in pipelines and can be combined with parallel computation routines in R. `Demerelate` analyzed the included example with 1000 bootstrap mutations in two minutes on an i5-3320M CPU @ 2.60GHz 4 Thinkpad T430 with 8 GB RAM.

Overview of functions

Demerelate calculates relatedness (Blouin *et al.* 1996, M_{xy}) based on the mean number of allele positions shared between individuals and r_{xy} an estimator of inter-individual relatedness derived from mean number of alleles shared (B_{xy}) (Queller and Goodnight 1989; Li and Horwitz 1953;

Oliehoek *et al.* 2006).

$$r_{xy,l} = \frac{0.5(I_{ac} + I_{ad} + I_{bc} + I_{bd}) - p_a - p_b}{1 + I_{ab} - p_a - p_b} \quad (1)$$

Where $r_{xy,l}$=pairwise estimator of relatedness for individual x and y for locus l, I_{ac}=Identity of alleles in position x_a and y_c, p_a=frequency of allele a in the genepool. Standard input format is used from the program *DEMEtics* (Gerlach *et al.* 2010). The estimator B_{XY} is also implemented in `Demerelate`; however, it is not recommended to use it in combination with statistics, since it may not differentiate correctly between identity by state and by descent (Oliehoek *et al.* 2006). B_{XY} should be used in `emp.calc()` omitting statistics and in `Demerelate` you may should use r_{XY} instead. Missing data are omitted when flagged as `NA`. Alternatively, the parameter `NA.rm` can be set as `FALSE` to skip the removal of samples with `NA`-values. This is not recommended, however calculations are stopped if certain `NA` combinations may lead to strange results. *Demerelate* is executed with default parameters as follows: `tab.pop`, `tab.dist`, `ref.pop`, `pairs`, `iteration`, `value`, `Fis`, `object`, `p.correct`, `dis.data`, `file.output` and `NA.rm`. F_{IS}, a summary of the input data, r_{XY}, B_{XY} or M_{XY} are calculated on either a R object or a .txt file (`object=TRUE/FALSE`). To correct for the bias between identity by state and identity by descent from a reference population (Blouin 2003) for M_{XY} or r_{XY} calculations, reference populations are calculated based either on locus information integrated over all populations from empirical data (`ref.pop="NA"`) or using custom reference populations. Random pairs of unrelated individuals are created defined by `pairs`. Defined by the same parameter, pairs of full siblings are created as randomized offspring descending from common parents and half siblings are created as randomized offspring from two parental pairs with only one shared parent (Blouin *et al.* 1996). For M_{XY}, thresholds for the empirical pairings are subsequently calculated based on likelihood ratios modeled by a logistic regression. Using a χ^2 probability test, the observed frequency of full and half siblings in empirical data is compared to the expected frequency of full and half siblings in a randomly generated population of non-related individuals. For r_{XY}, randomly generated populations are directly compared with empirical populations using a t-test, while the correction from allele frequencies in the reference population is directly implemented in the estimator calculation. F_{IS}-values are calculated based on Nei and Chesser (1983) and Weir and Cockerham (1984) including bootstrapping statistics. The output files summarize information for each population: allele/genotype frequencies, heterozygosity, F_{IS}-statistics, relatedness metrics (r_{XY} or M_{XY}), relatedness thresholds, χ^2 statistics and the probability for each individual pair being non-related, full siblings or half siblings. For each population, a single pdf file can be created containing: histograms of relatedness among individuals with corresponding thresholds of full and half siblings for each locus and all loci combined. Additionally, an UPGMA tree visualizes inter-individual relationships (Fig. 2). The tree is built by transforming similarities from relatedness metrics into distances by either calculating $1 - M_{xy}$ or $1 - r_{xy}$. In order to compare populations using r_{xy}, a pairwise t-test is calculated on multiple comparisons of means between populations and reference populations. When adding data on geographic distances, for each population a correlation between genetic relatedness and geographic distance is calculated and plotted. The statistics are based on mantel correlations (Mantel 1967; Mantel and Valand 1970; Oksanen *et al.* 2013). The output filenames include the name of the population and a bar-code. In order to evaluate the feasibility of data for relatedness analysis, `Loci.test(tab.pop, ref.pop, value, bt)` can be used to calculate mean differences in relatedness (`value`) from bootstrapped iterations (`iteration=1000`), each with a different random combination of loci. The resulting correlation between mean difference of bootstrap relatedness and number of randomized loci is

plotted and gives an overview of the influence of different number of loci on the relatedness analysis.

Example

Each input needs to be formatted as dataframe object (`object=TRUE`) or has to be readable by `read.table` as dataframe (`object=FALSE`). The first column of each dataframe contains a unique sample name, the second column implies the population as given in `data(demerelpop)`. Following column can be used for distance calculation consisting of either two columns with geographic positions for parameter `tab.dist` or genetic information given by appending pairwise columns for each diploid locus (`inputdata`, `ref.pop`). When geographic information is used `tab.dist` and `inputdata` need to be exactly the same in sample and population structure. No `NA`'s should be present in `tab.dist`. Example data `demerelpop`, `demerelref` and `demereldist` were randomly generated and are provided with the package. `demerelpop` is composed of three different populations. Pop−FS−HS consists of 10 full siblings (FS: 45 pairswise comparisons) combined with 10 half siblings (HS: 45 pairswise comparisons). Pop−FS−Non consists of 10 full sibilings (FS: 45 pairswise comparisons) and 10 random individuals (Non: 45 pairswise comparisons). Pop−Non is a population of 20 randomly drawn individuals (Non: 190 pairswise comparisons). `demerelref` is the parental population of all three populations and consists of 1000 randomly generated individuals. `demereldist` consists of geographic information for each sample in `demerelpop`. In order to obtain a first overview of the dataset `Loci.test()` provides a bootstrapping approach to visualize the variability in pairwise relatedness relative to randomly drawn loci information for each population.

```
>data(demerelpop)
>Loci.test(demerelpop, bt=1000, ref.pop=NA, object=TRUE,
 value="rxy", file.output=TRUE)
```

All information of the dataset is used for analysis by `Loci.test()`. Datasets have to be split manually if for example populations should be analyzed separately or only few loci should be considered. The function will randomly choose loci information from the dataset and calculates the statistic given by `value` (M_{Xy}, B_{Xy} or r_{Xy}) for each pairwise comparison as a mean over loci. A list object is created given pairwise similarities for randomly selected loci and a plot correlating mean differences in similarities of replicates with number of randomly chosen loci for each calculation. Comparing different estimators for relatedness by `Loci.test()` will help to decide, which number of loci would be appropriate for relatedness estimation. The example in Figure 1 illustrates how the variability of relatedness estimation decreases with number of loci taken into account. In general, the approach by M_{Xy} is conservative by dividing pairwise similarity into classes of similarity within the range of 0 to 1. Consequently, variability in relatedness estimation is increased with high number of alleles per locus and may lead to misclassification of relatedness (Blouin *et al.* 1996). Using r_{Xy} in those cases may be more precise, due to internal correction for allele diversity (Queller and Goodnight 1989). In any case `Loci.test()` should be considered to decide if the number of loci are sufficient for estimation of relatedness. To calculate each similarity coefficient (M_{Xy}, B_{Xy} or r_{Xy}) for a dataset the function `emp.calc` is provided.

```
>data(demerelpop)
>pop.relate <- emp.calc(demerelpop, value="Bxy", ref.pop="NA")
```

`emp.calc` returns a matrice of similarities for all pairwise comparisons in the dataset omitting any statistics. These results can be easily used for downstream analysis using R routines such as plotting the results or different statistics.

Considering an empirical dataset the function `Demerelate` is used as a framework to calculate different similarity values, their statistics, F-statistics of single populations and the relationships with geographic distances via mantel correlation.

```
>data(demerelpop)
>data(demerelref)
>data(demereldist)
>Demerelate(demerelpop, ref.pop=demerelref, tab.dist=demereldist, Fis=TRUE,
 object=TRUE, pairs=1000, iteration=1000, value="Mxy", file.output=TRUE)
```

`demerelpop` is splitted based on information of populations organized in column two and each population is treated separately for further statistics. The thresholds of different similarity estimates are calculated based on the information given by the parameter `ref.pop`. If set as `NA` all information from `inputdata` are taken as reference. A bootstrap sample of 1000 pairs is taken in order to define thresholds of relatedness accounting for full siblings and half siblings using logistic regression analyses. Randomly generated populations of the size of the empirical ones are treated with the same thresholds and defined as expected frequency of siblings as the null hypothesis. Frequencies of each population and the randomly generated population are compared either with a χ^2 probability test (M_{xy}) or a t-test (r_{xy}) depending on the estimator used. If `Fis=TRUE` an output is created which summarizes genetic information for each population and locus (allele vs. genotype frequencies; heterozygosity; F_{IS}-values) and calculates F_{IS}-values over loci based on Nei and Chesser (1983) and Weir and Cockerham (1984). Finally, if `tab.dist` is used correlation between genetic similarity and geographic distance will be visualized and mantel statistics are provided (Figure 3).

Notes on implementation and documentation

The program *Demerelate* is written in R and thus applicable for any operational system supporting the java environment (Windows, MacOS, Linux). The package requires a standard R base installation and the packages `sfsmisc` (Maechler *et al.* 2012), `mlogit` (Croissant 2011) and `fts` (Armstron 2012). If mantel statistics are used for correlating geographic distance with genetic distance `vegan` package is required (Oksanen *et al.* 2013). The convenient output format (.txt/.pdf and R object) allows further processing of results. The package and documentation is available through the Comprehensive R Archive Network (CRAN): http://cran.r-project.org/web/packages/-Demerelate/index.html.

Acknowledgement

We thank Dr. Peter Harmand for helpful discussions on statistics and advice for programming in R.

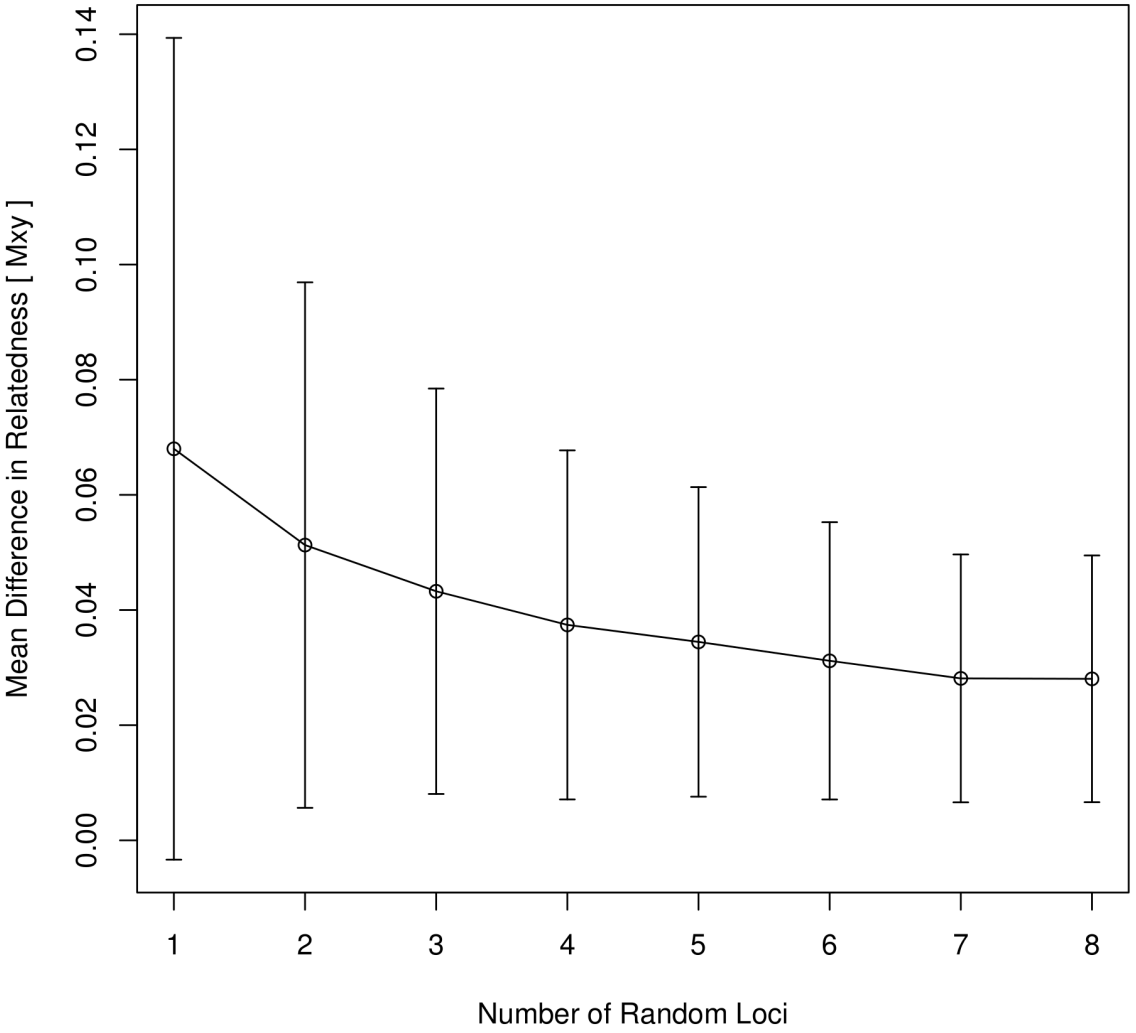


Figure 1: Effect of used number of loci on mean difference in relatedness (M_{Xy}). Different numbers indicated by the x-axis are randomly resampled from the dataset and each pairwise similarity is calculated and compared to the similarity given by all available data. Mean difference in pariwise similarity (M_{Xy}) among individuals dependent of used number of loci is calculated and plotted for visualization.

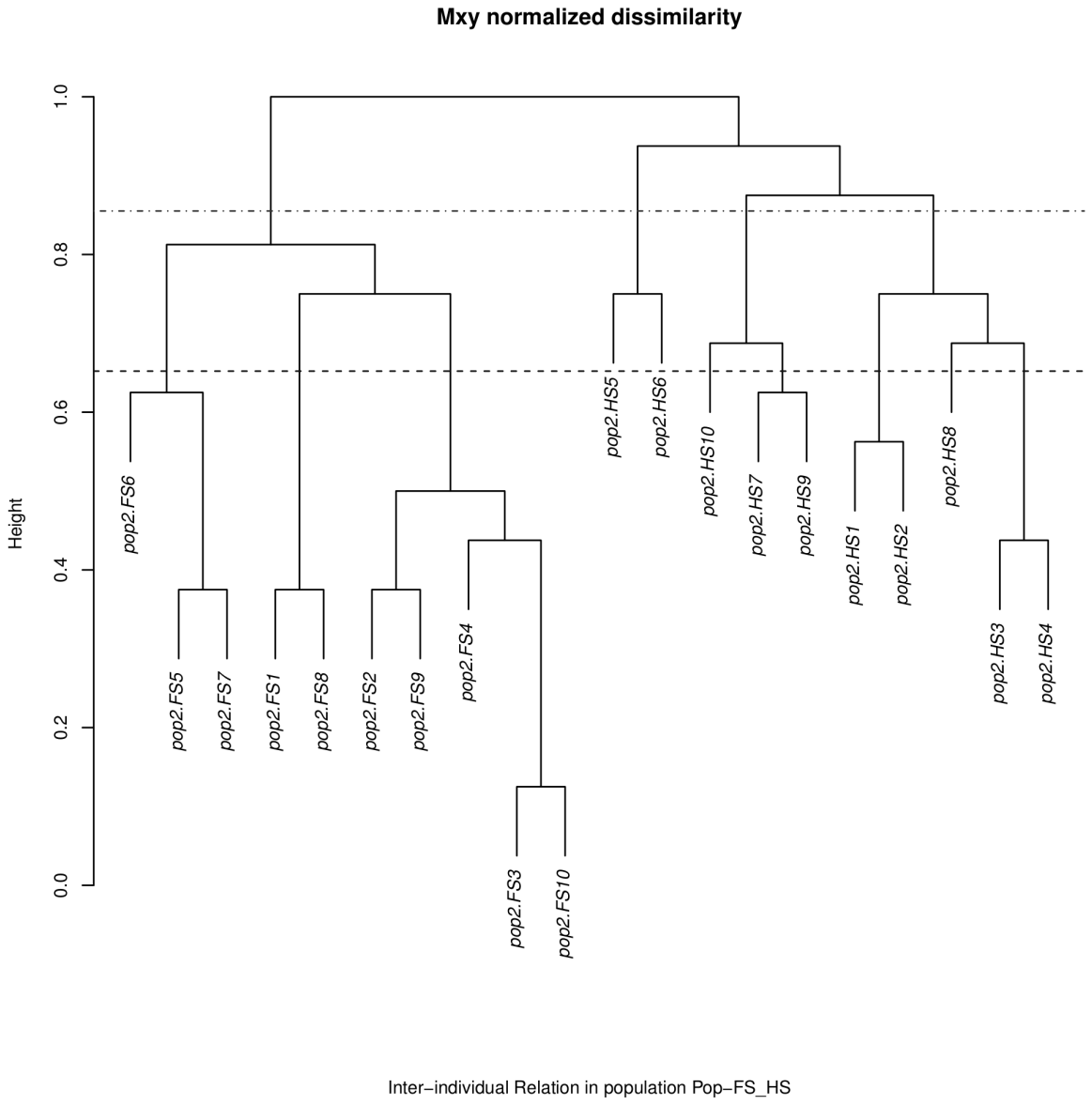


Figure 2: UPGMA cluster of genetic relatedness for population Pop−FS−HS. The plot is based on M_{xy} values transformed into dissimilarities by $D = 1 - M_{\mathrm{xy}}$. Lines indicate thresholds of relatedness calculated by logistic regression based on `demerelref` (dashed = full siblings, dottdashed = half siblings).

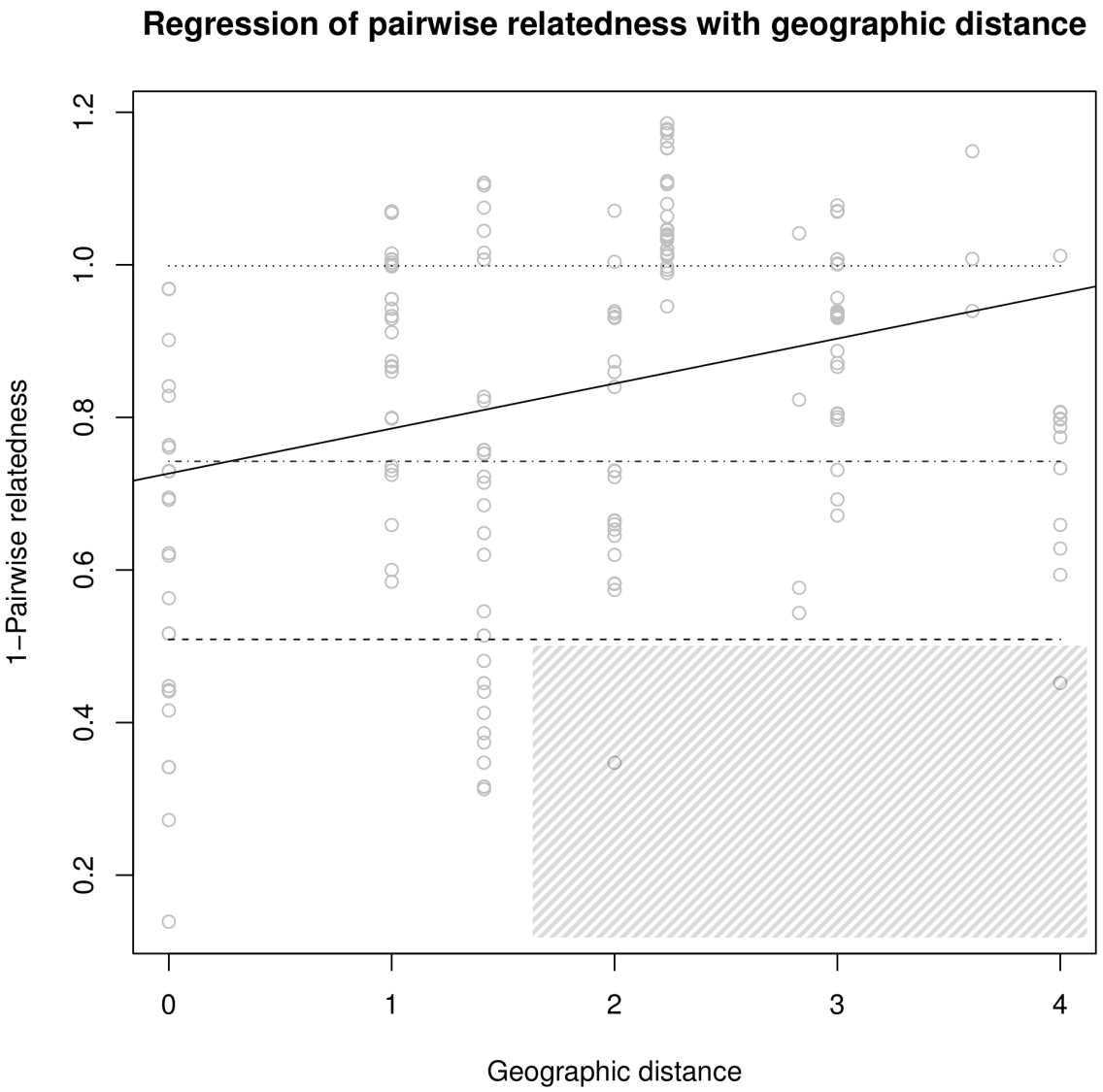


Figure 3: Spatial structure of relatedness. For population Pop−FS−HS linear correlation of empirical distance of individuals from genetic data (r_{xy} transformed into dissimilarities by $D = 1-$ pairwise relatedness) and geographic distances derived from relative geographic positions. Pearson's product-moment correlation calculated by Mantel Statistics (0.274, p = 0.0039) is based on 1000 permutations. Lines in the plot indicate modelled thresholds of relatedness (dashed = full siblings, dottdashed = half siblings, random non related = dotted grey). The example shows that at a geographic distance of > 1.5 very few full siblings could be found (indicated by the striped rectangle).

Table 1: In empirical populations the number of full (N_{full}) and half siblings (N_{half}) as well as non-related pairs (N_{non}) are determined by using estimates for pairwise relatedness. M_{xy} relatedness thresholds (T_{half}, T_{full}) were modeled by logistic regression based on a reference population of 1000 individuals. p_f and p_{fh} are based on χ^2 statistics comparing proportions of either full siblings or full siblings+half siblings in empirical data compared to randomly drawn pairs. $** < 0.01$.

Population	T_{full}	T_{half}	N_{full}	N_{half}	N_{non}	p_f	p_{fh}
Pop−FS−HS							
Observed	0.35	0.14	38	60	92	**	**
Expected			0	11	179		
Pop−FS−Non							
Observed	0.35	0.14	44	12	134	**	**
Expected			0	18	172		
Pop−Non							
Observed	0.35	0.14	0	15	175	NA	0.71
Expected			0	17	173		

References

Armstrong W (2012) *fts: R interface to tslib (a time series library in C++)*. R package version 0.7.7.

Blouin M (2003) Dna-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution*, **18**, 503–511.

Blouin M, Parsons M, Lacaille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology*, **5**, 393–401.

Broquet T, Viard F, Yearsley JM Genetic drift and collective dispersal can result in chaotic genetic patchiness. *Evolution*, **67**, 1660–1675.

Buston PM, Fauvelot C, Wong MYL, Planes S (2009) Genetic relatedness in groups of the humbug damselfish Dascyllus aruanus: small, similar-sized individuals may be close kin. *Molecular Ecology*, **18**, 4707–4715.

Croissant Y (2011) *mlogit: multinomial logit model*. R package version 0.2-2.

Gerlach G, Schardt U, Eckmann R, Meyer A (2001) Kin-structured subpopulations in eurasian perch (*Perca fluviatilis* l.). *Heredity*, **86**, 213–221.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on Gst and D: forget Gst but not all of statistics! *Molecular Ecology*, **19**, 3845–3852.

Iacchei M, Ben-Horin T, Selkoe KA, Bird CE, Garcia-Rodriguez FJ, Toonen RJ (2013) Combined analyses of kinship and Fst suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. *Molecular Ecology*, **22**, 3476–3494.

Maechler M *et al.* (2012) *sfsmisc: Utilities from Seminar fuer Statistik ETH Zurich*. R package version 1.0-20.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Mantel N, Valand RS (1970) A technique of nonparametric multivariate analysis. *Biometrics*, **26**, 547–558.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253–259.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2013) *vegan: Community Ecology Package*. R package version 2.0-8.

Oliehoek P, Windig J, Arendonk J, Bijma P (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics*, **173**, 483–496.

Li CC, Horvitz DG (1953) Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics*, **5**, 107-117.

Queller D, Goodnight K (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Stephens HC, Schmuki C, Burridge CP, OReilly-Wapstra JM (2013) Habitat fragmentation in forests affects relatedness and spatial genetic structure of a native rodent, (*Rattus lutreolus*). *Austral Ecology*, **38**, 568–580.

Stern C (1943) The Hardy-Weinberg law. *Science*, **97**, 137–138.

Veliz D, Duchesne P, Bourget E, Bernatchez L (2006) Genetic evidence for kin aggregation in the intertidal acorn barnacle (*Semibalanus balanoides*). *Molecular Ecology*, **15**, 4193–4202.

Wahlund S (1928) Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, **11**, 65–106.

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

Wikberg EC, Sicotte P, Campos FA, Ting N (2012) Between-group variation in female dispersal, kin composition of groups, and proximity patterns in a black-and-white colobus monkey (*Colobus vellerosus*). *PLoS ONE*, **7**, e48740.

Author contributions

P.K. developed and coded the R package Demerelate. P.K. drafted the manuscript. G.G. assisted in drafting the manuscript. All authors read and approved the final manuscript.

Manuscript IV

Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics!

published in Molecular Ecology

MOLECULAR ECOLOGY

Molecular Ecology (2010)

NEWS AND VIEWS

COMMENT

Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics!

GABRIELE GERLACH, ALEXANDER JUETERBOCK, PHILIPP KRAEMER, JANA DEPPERMANN and PETER HARMAND

Institute of Biology and Environmental Sciences, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany

G_{ST}-values and its relatives (F_{ST}) belong to the most used parameters to define genetic differences between populations. Originally, they were developed for allozymes with very low number of alleles. Using highly polymorphic microsatellite markers it was often puzzling that G_{ST}-values were very low but statistically significant. In their papers, Jost (2008) and Hedrick (2005) explained that G_{ST}-values do not show genetic differentiation, and Jost suggested calculating D-values instead. Theoretical mathematical considerations are often difficult to follow; therefore, we chose an applied approach comparing two artificial populations with different number of alleles at equal frequencies and known genetic divergence. Our results show that even for more than one allele per population G_{ST}-values do not calculate population differentiation correctly; in contrast, D-values do reflect the genetic differentiation indicating that data based on G_{ST}-values need to be re-evaluated. In our approach, statistical evaluations remained similar. We provide information about the impact of different sample sizes on D-values in relation to number of alleles and genetic divergence.

Keywords: D, genetic differentiation, G_{ST}, population genetics

Received 8 July 2009; revision received 15 June 2010; accepted 26 June 2010

Recently, Jost (2008) argued that G_{ST} and its relatives are inappropriate measures of genetic differentiation between populations. He showed convincingly that when using highly polymorphic microsatellite markers the G_{ST}-value cannot reach its maximum value of 1. Even when populations share no alleles at all, G_{ST}-values remain low. These low G_{ST}-values can be misinterpreted as low genetic differentiation between populations leading to the false assumption of high connectivity. However, G_{ST} remains an appropriate indicator for calculating migration rates for populations that are consistent with the finite island model (discussed in Jost 2009; Ryman & Leimar 2009). To calculate 'real' genetic differences between populations, he suggested a different method using the differentiation index D and the bias-corrected estimator D_{est}. These indices are based on the effective number of alleles resulting in a more meaningful perception of differentiation (see also Heller & Siegismund 2009; Jost 2009; Ryman & Leimar 2009). The intent of our analysis is to calculate D_{est} and G_{ST_est}-values for assessing the differentiation between populations for which the 'true' divergence and number of present alleles are known. This would enable us to find out whether D_{est} or/and G_{ST_est}-values are better estimators of population differentiation for populations of different sample sizes and markers with different number of alleles.

Correspondence: Gabriele Gerlach, Fax: +49 441 798 3937;
E-mail: gabriele.gerlach@uni-oldenburg.de

Genetic diversity values

Based on artificial and randomized data sets, we calculated G_{ST}, G_{ST_est}, D, D_{est}, H_T, H_S, H_{T_est} and H_{S_est} using our newly developed package '*DEMEtics*' in R (R Development Core Team, 2009); we tested statistical significance using the null hypothesis of zero differentiation. D-values were based on equation 11 (Jost 2008):

$$D = [(H_T - H_S)/(1 - H_S)]\,[n/(n-1)].$$

The bias-corrected D_{est}-values were based on equation 12 (Jost 2008):

$$D_{est} = [(H_{T\_est} - H_{S\_est})/(1 - H_{S\_est})]\,[n/(n-1)].$$

$G_{ST} = (H_T - H_S)/H_T$ and the bias-corrected G_{ST_est}-values were calculated according to Nei & Chesser (1983):

$$G_{ST\_est} = [(H_{T\_est} - H_{S\_est})/H_{T\_est}]\,[n/(n-1)].$$

For comparison, we also calculated G_{ST_est}-values and corresponding P values using the program FSTAT 2.93 (Goudet *et al.* 1996). However, no difference was detected between values calculated by our programme and by FSTAT 2.93.

Generating test data with theoretical allele frequencies

To compare different methods to calculate genetic diversity (G_{ST} and D), we created artificial data sets with therefore known divergence. These data consisted of two populations with 40 diploid individuals each. To show the effect of increasing number of alleles (i.e. *mutational* effects), we calculated genetic diversity for populations with N_{a_total} = 2, 4, 8, 10, 12, 16, 20, 26, 32 and 40 alleles total. All alleles had the same frequencies. For each number of alleles, we started with the assumption that populations shared no alleles (100% divergence); e.g. for N_{a_total} = 8, population A had alleles 1, 2, 3 and 4 and population B had alleles 5, 6, 7 and 8 at equal frequencies. Then we decreased divergence, simulating *migrational* gene flow by exchanging 5% of alleles in

both directions: population A obtained 5% alleles from population B (four alleles: one time alleles 5, 6, 7 and 8) and population B obtained 5% alleles from population A (four alleles: one time alleles 1, 2, 3 and 4), resulting in a divergence of 90%. In terms of frequencies for the eight alleles, this means passing from the distribution (20/80, 20/80, 20/80, 20/80, 0, 0, 0, 0) and (0, 0, 0, 0, 20/80, 20/80, 20/80, 20/80) to (19/80, …, 19/80, 1/80, …, 1/80) and (1/80, …, 1/80, 19/80, …, 19/80) in the first step and to (18/80, …, 18/80, 2/80, …, 2/80) and (2/80, …, 2/80, 18/80, …, 18/80) in the second step. With this process, we decreased genetic divergence in 10% steps from 100% to 0%. Alleles that had been introduced from the other population were not exchanged in this process. At 0% divergence, populations were identical having the same number and frequency of alleles and were considered finite populations with theoretical allele frequencies.

Comparison of G_{ST} and D in finite populations with theoretical allele frequencies

We then calculated the genetic difference between these two finite populations of 40 individuals each based on G_{ST}- and D-values; results are shown in Fig. 1a,b. The maximum possible G_{ST} fell rapidly as the number of alleles per population increased (see Fig. 1a). In contrast, D-values always equalled 1 when populations shared no alleles (100% divergence) independent of the number of alleles (see Fig. 1b). G_{ST}-values only equalled 1 when population A had allele 1 and population B had allele 2 and did not share any alleles. The maximum G_{ST}-value already dropped to 0.328 when 2 different alleles occurred in each population (four alleles total) even when populations did not share any alleles (100% divergence; Fig. 1a). By using four different unique alleles in each population (eight alleles total; 100% divergence), the maximum G_{ST}-value was 0.137 (Fig. 1a). For 5–20 different alleles in each population, the maximum G_{ST}-value ranged from 0.105 to 0.019 at 100% divergence. These calculations support Jost's conclusion that G_{ST} is not an appropriate metric for genetic difference while D reflects the known divergence between populations much better. G_{ST} ranked the divergence between populations in the finite island model correctly (Fig. 1a) when the number of alleles remained the same but divergence decreased (simulating *migrational* effects). For *mutational* effects (increasing number of alleles), our results confirm Jost (2009) that the ranking is a problem when calculating G_{ST}-values. With increasing number of alleles, G_{ST}-values can decrease even when the real divergence between populations increases. At equal allele frequencies, D-values ranked genetic distances correctly. The main advantage of D over G_{ST} is that it is based on the complete partitioning of heterozygosity into independent within- and between-group components, whereas G_{ST} confounds the two components.

Limitations of D and G_{ST} based on different H_S and H_T

While in the aforementioned calculations, allele frequencies in both populations were maintained equal, we next evaluated how G_{ST}- and D-values behave in populations with different H_S and H_T combinations caused by different allele frequencies. The results are shown in Fig. 2a,b.

The triangle gives the domain of definition of G_{ST} and D as a function of H_S and H_T. Because H_S is always less than or equal to H_T, points below the triangle are not possible. Points above the triangle can only be generated for more than two populations.

Each line within the triangle shows G_{ST}- (Fig. 2a) and D-values (Fig. 2b) ranging from 0 to 1 in steps of 0.1. Diamonds reflect G_{ST}- (Fig. 2a) and D-values (Fig. 2b) for a given number of alleles (N_{a_total}) and decreasing divergence between populations from the left ($D = 1$, divergence = 100%) to the right side ($D = 0$, divergence = 0%). Figure 2a shows that with increasing allele numbers the possible range of G_{ST}-values becomes rapidly smaller. With 16 alleles, G_{ST}-values were constrained to < 0.1, even when the two populations were completely different. $G_{ST} = 1$ can only be reached when $H_S = 0$ which is realized when each of both populations bears a single allele. In contrast, D-values continued to range from 0 to 1 independent of the number of alleles present. In this respect, D-values provide a clearer metric for population differentiation than G_{ST}-values.

Different H_S and H_T values can result in the same D-values as well as in the same G_{ST}-values. Table 1 gives an example for populations with two alleles (columns 1–2) and also three alleles (columns 3–5) at different frequencies. The two populations in the first column are more differentiated than those in the second column; and populations in column 3 are more differentiated than those

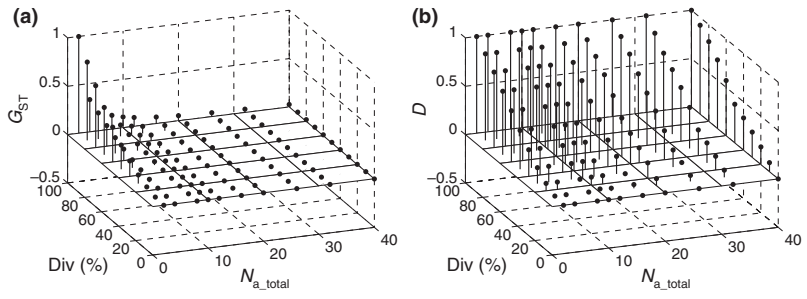


Fig. 1 (a,b) Genetic differences between two populations analysed by (a) G_{ST}- and (b) D-values. N_{a_total} = total number of alleles in two populations. Div = divergence [%]. For further explanation see text.

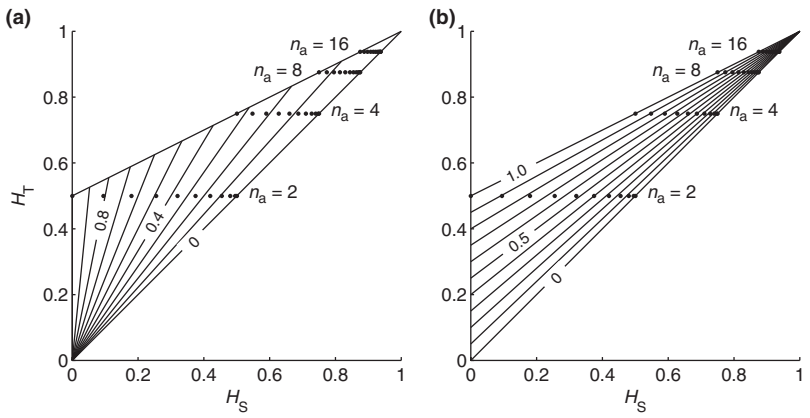


Fig. 2 (a,b) The triangle gives the domain of definition of G_{ST} and D as a function of H_S and H_T. Each line within the triangle shows (a) G_{ST}- and (b) D-values ranging from 0 to 1 in steps of 0.1. Diamonds reflect G_{ST}- and D-values for a given number of alleles (N_{a_total}) and decreasing divergence between populations from the left (H_S = 0; 100% divergence) to the right side (H_S = 1; 0% divergence).

Table 1 D- and G_{ST}-values for populations with alleles with different frequencies (F_a)

Allele	F_a [%]	F_a [%]	F_a [%]	F_a [%]	F_a [%]
Population 1					
1	62.9	75.4	10.0	15.0	30.0
2	37.1	24.6	4.4	9.2	7.8
3			85.6	75.8	62.2
Population 2					
1	1.6	17.5	10.0	25.0	33.0
2	98.4	82.5	59.2	53.5	55.2
3			30.8	21.5	14.8
D	0.5	0.5	0.5	0.5	0.5
G_{ST}	0.43	0.34	0.27	0.20	0.17

shown in column 4 and 5; however, all give the same value D = 0.5. In all these cases, the differences were reflected in different G_{ST}-values. This observation is an example of ranking problems with D. Similarly, one can construct examples with different allele frequencies giving the same G_{ST}-value but different D-values. These examples elucidate that there is not one encompassing metric to describe genetic differences between populations. However, the advantage of the D-metric is that it reflects genetic divergence between populations better than G_{ST}, because the latter is strongly limited in range with increasing allele frequency.

Comparison of G_{ST_est}- and D_{est}-values in randomly sampled populations

Effect of number of alleles and sample size

We evaluated the bias correction in G_{ST_est}- and D_{est}-values based on calculations for randomly sampled populations. We used the artificial data set described above and drew 30 random subsamples of alleles (with replacement) from each of these two populations. We then calculated the mean G_{ST}-, G_{ST_est}-, D- and D_{est}-values. These randomly generated subsamples consisted either of 10, 20, 40 or 60 individuals per population. This approach should simulate the field situation of sampling different numbers of individuals from two populations and calculating genetic difference between them. Different from the natural situation, in our artificial samples, the genetic divergence and number of alleles in the original populations were known. Therefore, we could compare how accurately the calculated G_{ST_est}- and D_{est}-values responded to a sampling regime of populations with a different number of alleles as well as to differences in sample size. It is also important to notice the standard deviation of G_{ST_est}- and D_{est}-values correlated to sample size. By calculating D_{est}-values according to equation 12 (Jost 2008), the bias resulting from differences in allele frequencies generated by random sampling should be corrected. The same calculations were performed for G_{ST} and the bias-corrected G_{ST_est}-values (see Table 2).

For both, D_{est} and G_{ST_est}, we calculated P values and confidence intervals to find out whether they differed significantly from a random distribution based on bootstrapping (500 iterations).

Considering P values in almost all cases, D_{est}-values were statistically significant when G_{ST_est}-values were too. We can conclude that G_{ST_est}-values do not reflect genetic differentiation between populations when there are more than two alleles present, but at least the calculated statistical significance remains meaningful. Therefore, old data sets need to be reanalysed to find out how different populations were, but it is probably that the conclusions based on statistical significance are the same.

By using the G_{ST_est} method, a great error seemed to have been to ignore statistically significant G_{ST_est}-values simply because the values themselves were so low. For instance, G_{ST_est} was very low (0.005, Table 2, line 142) when 60 individuals were sampled per population with a total number of alleles n = 32 and a 'real' divergence of 40%, but this small number was still significantly different from the bootstrap distribution (P = 0.009). The corresponding D_{est}-value showed a similar level of significance but a much larger value to reflect the difference between populations (D_{est} = 0.268, P = 0.008, Table 2). D_{est}-values are clearly a better indicator for population structure than G_{ST_est}.

4 NEWS AND VIEWS: COMMENT

Table 2 Mean values of genetic difference and heterozygosity between two different populations based on 30 random samples drawn from populations with alleles of equal frequency; div: genetic divergence between finite populations (see text for detailed information); sample size: number of individuals sampled per population; N_a total number of alleles in both populations; H_{S_est} mean heterozygosity in each population with bias correction, H_{T_est} mean total heterozygosity with bias correction, G_{ST} biased mean G_{ST}-value, G_{ST_est} bias-corrected mean value; SD standard deviation of G_{ST_est}; CI confidence interval of G_{ST_est} (percentile method); P statistical significance based on bootstraps with 500 repeats; D mean D-value according to equation 11, Jost (2008); D_{est} bias-corrected mean D_{est}-values (equation 12 Jost 2008); SD standard deviation of D_{est}-values; CI confidence interval (percentile method); P statistical significance of D_{est}-values based on bootstraps with 500 repeats (bold values indicate statistical significance)

Line	Div [%]	Sample size	N_a	H_{S_est}	H_{T_est}	G_{ST}	G_{ST_est}	SD	P	(−CI) perc	(+CI) perc	D	D_{est}	SD	P	(−CI) perc	(+CI) perc
1	100	10	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
2	80	10	2	0.187	0.499	0.636	**0.621**	0.17	**0.002**	**0.354**	**0.870**	0.760	**0.756**	0.12	**0.002**	**0.499**	**0.949**
3	60	10	2	0.29	0.496	0.432	**0.412**	0.19	**0.002**	**0.162**	**0.708**	0.573	**0.562**	0.17	**0.002**	**0.256**	**0.829**
4	40	10	2	0.397	0.492	0.213	**0.188**	0.12	**0.027**	**0.002**	**0.464**	0.322	**0.297**	0.15	**0.025**	**0.023**	**0.618**
5	20	10	2	0.481	0.497	0.054	0.029	0.06	0.421	−0.047	0.205	0.094	0.050	0.11	0.423	−0.086	0.317
6	0	10	2	0.486	0.495	0.041	0.015	0.06	0.443	−0.048	0.185	0.071	0.025	0.10	0.452	−0.086	0.283
7	100	20	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
8	80	20	2	0.182	0.499	0.64	**0.633**	0.12	**0.002**	**0.432**	**0.824**	0.770	**0.769**	0.09	**0.002**	**0.595**	**0.908**
9	60	20	2	0.341	0.498	0.325	**0.313**	0.09	**0.002**	**0.130**	**0.529**	0.478	**0.471**	0.10	**0.002**	**0.242**	**0.691**
10	40	20	2	0.425	0.492	0.147	**0.135**	0.07	**0.014**	**0.015**	**0.309**	0.243	**0.228**	0.11	**0.012**	**0.038**	**0.461**
11	20	20	2	0.477	0.496	0.05	0.037	0.03	0.124	−0.022	0.155	0.091	0.070	0.06	0.117	−0.035	0.266
12	0	20	2	0.498	0.499	0.013	0.001	0.02	0.486	−0.025	0.079	0.026	0.001	0.03	0.474	−0.049	0.143
13	100	40	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
14	80	40	2	0.19	0.499	0.622	**0.619**	0.08	**0.002**	**0.473**	**0.763**	0.761	**0.760**	0.07	**0.002**	**0.637**	**0.866**
15	60	40	2	0.319	0.498	0.364	**0.358**	0.07	**0.002**	**0.218**	**0.511**	0.525	**0.522**	0.08	**0.002**	**0.358**	**0.674**
16	40	40	2	0.433	0.497	0.135	**0.129**	0.07	**0.002**	**0.041**	**0.245**	0.229	**0.222**	0.10	**0.003**	**0.083**	**0.384**
17	20	40	2	0.48	0.499	0.042	0.036	0.04	0.125	−0.004	0.108	0.079	0.068	0.07	0.121	−0.006	0.190
18	0	40	2	0.497	0.498	0.006	0.000	0.00	0.416	−0.013	0.041	0.012	0.000	0.01	0.422	−0.024	0.074
19	100	60	2	0	0.5	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
20	80	60	2	0.181	0.5	0.639	**0.637**	0.07	**0.002**	**0.517**	**0.754**	0.776	**0.776**	0.06	**0.002**	**0.678**	**0.862**
21	60	60	2	0.322	0.499	0.358	**0.355**	0.07	**0.002**	**0.240**	**0.481**	0.522	**0.520**	0.08	**0.002**	**0.390**	**0.647**
22	40	60	2	0.425	0.499	0.151	**0.147**	0.05	**0.002**	**0.067**	**0.243**	0.257	**0.252**	0.07	**0.002**	**0.129**	**0.387**
23	20	60	2	0.474	0.499	0.053	**0.049**	0.04	**0.017**	**0.007**	**0.114**	0.099	**0.092**	0.06	**0.017**	**0.015**	**0.198**
24	0	60	2	0.499	0.498	0.003	−0.001	0.00	0.561	−0.008	0.022	0.006	−0.002	0.01	0.555	−0.017	0.044
25	100	10	4	0.498	0.746	0.351	**0.328**	0.02	**0.002**	**0.289**	**0.419**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
26	80	10	4	0.584	0.738	0.229	**0.205**	0.06	**0.002**	**0.101**	**0.326**	0.757	**0.737**	0.15	**0.002**	**0.474**	**0.934**
27	60	10	4	0.654	0.738	0.134	**0.109**	0.05	**0.018**	**0.017**	**0.231**	0.520	**0.468**	0.18	**0.017**	**0.151**	**0.766**
28	40	10	4	0.699	0.742	0.08	0.054	0.05	0.142	−0.022	0.166	0.339	0.251	0.21	0.136	−0.057	0.586
29	20	10	4	0.73	0.74	0.037	0.011	0.03	0.365	−0.041	0.103	0.175	0.053	0.14	0.376	−0.197	0.401
30	0	10	4	0.734	0.736	0.026	0.000	0.02	0.521	−0.043	0.083	0.123	−0.009	0.13	0.524	−0.227	0.328
31	100	20	4	0.494	0.746	0.346	**0.335**	0.01	**0.002**	**0.312**	**0.385**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
32	80	20	4	0.589	0.746	0.22	**0.208**	0.03	**0.002**	**0.133**	**0.289**	0.775	**0.766**	0.08	**0.002**	**0.582**	**0.910**
33	60	20	4	0.65	0.748	0.141	**0.129**	0.05	**0.002**	**0.062**	**0.208**	0.570	**0.548**	0.15	**0.002**	**0.327**	**0.750**
34	40	20	4	0.702	0.745	0.07	**0.057**	0.03	**0.012**	**0.004**	**0.126**	0.326	**0.284**	0.12	**0.010**	**0.057**	**0.527**
35	20	20	4	0.73	0.746	0.033	0.020	0.02	0.155	−0.016	0.077	0.168	0.109	0.09	0.124	−0.071	0.350
36	0	20	4	0.746	0.746	0.011	−0.001	0.01	0.539	−0.022	0.039	0.062	−0.009	0.06	0.536	−0.126	0.188
37	100	40	4	0.5	0.749	0.337	**0.331**	0.00	**0.002**	**0.322**	**0.353**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
38	80	40	4	0.59	0.749	0.217	**0.211**	0.03	**0.002**	**0.161**	**0.262**	0.778	**0.774**	0.07	**0.002**	**0.655**	**0.876**
39	60	40	4	0.658	0.746	0.124	**0.118**	0.02	**0.002**	**0.070**	**0.171**	0.528	**0.517**	0.06	**0.002**	**0.351**	**0.670**
40	40	40	4	0.711	0.747	0.054	**0.005**	0.02	**0.003**	**0.014**	**0.092**	0.271	**0.248**	0.08	**0.004**	**0.090**	**0.422**
41	20	40	4	0.738	0.747	0.018	0.012	0.01	0.149	−0.007	0.041	0.100	0.067	0.06	0.145	−0.035	0.212
42	0	40	4	0.747	0.747	0.005	−0.001	0.00	0.519	−0.011	0.019	0.029	−0.007	0.02	0.531	−0.066	0.102
43	100	60	4	0.497	0.748	0.339	**0.335**	0.01	**0.002**	**0.327**	**0.352**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
44	80	60	4	0.59	0.749	0.215	**0.211**	0.03	**0.002**	**0.170**	**0.253**	0.776	**0.773**	0.06	**0.002**	**0.677**	**0.857**
45	60	60	4	0.657	0.749	0.126	**0.122**	0.02	**0.002**	**0.083**	**0.166**	0.542	**0.535**	0.07	**0.002**	**0.402**	**0.662**
46	40	60	4	0.707	0.748	0.058	**0.054**	0.02	**0.002**	**0.025**	**0.089**	0.289	**0.275**	0.07	**0.002**	**0.143**	**0.417**
47	20	60	4	0.739	0.748	0.016	0.011	0.01	0.070	−0.003	0.033	0.087	0.066	0.04	0.066	−0.016	0.175
48	0	60	4	0.747	0.748	0.005	0.001	0.00	0.422	−0.007	0.016	0.030	0.006	0.02	0.421	−0.041	0.087
49	100	10	8	0.74	0.866	0.166	**0.141**	0.02	**0.002**	**0.107**	**0.196**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**

Table 2 (Continued)

Line	Div [%]	Sample size	N_a	H_{S_est}	H_{T_est}	G_{ST}	G_{ST_est}	SD	P	(−CI) perc	(+CI) perc	D	D_{est}	SD	P	(−CI) perc	(+CI) perc
50	80	10	8	0.777	0.86	0.118	**0.093**	0.03	**0.004**	**0.038**	**0.163**	0.798	**1.000**	0.13	**0.003**	**0.501**	**0.936**
51	60	10	8	0.812	0.863	0.081	**0.055**	0.03	**0.041**	**0.004**	**0.124**	0.623	**0.525**	0.22	**0.036**	**0.187**	**0.800**
52	40	10	8	0.841	0.866	0.051	0.025	0.03	0.173	−0.019	0.087	0.444	0.266	0.23	0.156	−0.101	0.611
53	20	10	8	0.856	0.862	0.029	0.004	0.02	0.421	−0.033	0.060	0.284	0.031	0.19	0.429	−0.326	0.417
54	0	10	8	0.862	0.865	0.026	0.001	0.01	0.480	−0.033	0.053	0.263	−0.008	0.20	0.491	−0.358	0.376
55	100	20	8	0.745	0.871	0.154	**0.142**	0.01	**0.002**	**0.124**	**0.170**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
56	80	20	8	0.787	0.87	0.106	**0.093**	0.02	**0.002**	**0.060**	**0.131**	0.808	**0.788**	0.07	**0.002**	**0.617**	**0.921**
57	60	20	8	0.826	0.869	0.06	**0.048**	0.02	**0.003**	**0.016**	**0.086**	0.549	**0.491**	0.13	**0.003**	**0.253**	**0.708**
58	40	20	8	0.842	0.866	0.039	**0.026**	0.01	**0.028**	−0.002	0.063	0.380	**0.291**	0.11	**0.020**	**0.037**	**0.551**
59	20	20	8	0.862	0.871	0.02	0.008	0.01	0.272	−0.013	0.037	0.229	0.098	0.13	0.249	−0.131	0.358
60	0	20	8	0.865	0.867	0.014	0.001	0.01	0.422	−0.017	0.028	0.156	0.010	0.08	0.434	−0.197	0.268
61	100	40	8	0.749	0.873	0.148	**0.141**	0.00	**0.002**	**0.133**	**0.155**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
62	80	40	8	0.792	0.874	0.099	**0.093**	0.01	**0.002**	**0.071**	**0.116**	0.802	**0.793**	0.04	**0.002**	**0.675**	**0.892**
63	60	40	8	0.827	0.872	0.057	**0.051**	0.01	**0.002**	**0.029**	**0.076**	0.553	**0.525**	0.06	**0.002**	**0.357**	**0.680**
64	40	40	8	0.849	0.871	0.031	**0.025**	0.01	**0.004**	**0.007**	**0.047**	0.335	**0.287**	0.09	**0.004**	**0.113**	**0.468**
65	20	40	8	0.867	0.873	0.012	0.006	0.01	0.211	−0.005	0.022	0.150	0.078	0.08	0.207	−0.057	0.246
66	0	40	8	0.873	0.873	0.005	−0.001	0.00	0.598	−0.009	0.011	0.067	−0.016	0.04	0.603	−0.122	0.130
67	100	60	8	0.747	0.873	0.147	**0.143**	0.00	**0.002**	0.137	0.153	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
68	80	60	8	0.792	0.873	0.096	**0.092**	0.01	**0.002**	0.074	0.111	0.789	**0.782**	0.04	**0.002**	**0.686**	**0.866**
69	60	60	8	0.829	0.873	0.054	**0.050**	0.01	**0.002**	0.032	0.069	0.536	**0.517**	0.07	**0.002**	**0.382**	**0.645**
70	40	60	8	0.852	0.873	0.028	**0.024**	0.01	**0.002**	0.010	0.040	0.315	**0.282**	0.06	**0.002**	**0.144**	**0.426**
71	20	60	8	0.866	0.873	0.011	**0.007**	0.00	**0.045**	−0.002	0.019	0.142	**0.094**	0.06	**0.041**	−0.014	0.222
72	0	60	8	0.873	0.873	0.004	0.000	0.00	0.531	−0.006	0.009	0.054	−0.002	0.04	0.540	−0.076	0.104
73	100	10	12	0.816	0.903	0.118	**0.093**	0.01	**0.002**	**0.063**	**0.141**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
74	80	10	12	0.848	0.903	0.084	**0.058**	0.01	**0.006**	**0.021**	**0.107**	0.822	**0.764**	0.11	**0.003**	**0.503**	**0.937**
75	60	10	12	0.867	0.905	0.064	0.038	0.02	0.056	**0.001**	**0.089**	0.686	**0.550**	0.23	**0.037**	**0.212**	**0.803**
76	40	10	12	0.889	0.904	0.039	0.014	0.01	0.203	−0.020	0.060	0.490	0.233	0.21	0.181	−0.159	0.580
77	20	10	12	0.895	0.903	0.031	0.006	0.02	0.406	−0.026	0.053	0.400	0.063	0.29	0.414	−0.339	0.438
78	0	10	12	0.897	0.906	0.032	0.006	0.01	0.359	−0.025	0.052	0.419	0.099	0.26	0.360	−0.306	0.477
79	100	20	12	0.825	0.91	0.104	**0.092**	0.01	**0.002**	**0.076**	**0.115**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
80	80	20	12	0.855	0.91	0.072	**0.059**	0.01	**0.002**	**0.036**	**0.087**	0.813	**0.783**	0.08	**0.002**	**0.611**	**0.915**
81	60	20	12	0.88	0.912	0.047	**0.034**	0.01	**0.026**	**0.011**	**0.062**	0.618	**0.537**	0.19	**0.023**	**0.296**	**0.745**
82	40	20	12	0.896	0.909	0.025	0.012	0.01	0.198	−0.006	0.037	0.366	0.209	0.22	0.204	−0.046	0.468
83	20	20	12	0.902	0.91	0.02	0.008	0.01	0.259	−0.010	0.032	0.313	0.134	0.17	0.228	−0.130	0.411
84	0	20	12	0.912	0.912	0.011	−0.002	0.00	0.579	−0.016	0.018	0.191	−0.041	0.09	0.615	−0.283	0.235
85	100	40	12	0.828	0.913	0.098	**0.092**	0.00	**0.002**	**0.084**	**0.103**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
86	80	40	12	0.86	0.914	0.064	**0.058**	0.01	**0.002**	**0.042**	**0.074**	0.793	**0.777**	0.05	**0.002**	**0.654**	**0.880**
87	60	40	12	0.884	0.914	0.037	**0.031**	0.01	**0.002**	**0.017**	**0.047**	0.549	**0.504**	0.09	**0.002**	**0.328**	**0.662**
88	40	40	12	0.898	0.913	0.023	**0.016**	0.01	**0.007**	**0.004**	**0.032**	0.371	**0.298**	0.10	**0.006**	**0.121**	**0.481**
89	20	40	12	0.909	0.913	0.01	0.004	0.00	0.216	−0.005	0.015	0.179	0.071	0.08	0.218	−0.086	0.252
90	0	40	12	0.912	0.913	0.006	0.000	0.00	0.535	−0.007	0.010	0.118	−0.003	0.07	0.543	−0.140	0.168
91	100	60	12	0.83	0.914	0.096	**0.091**	0.00	**0.002**	**0.086**	**0.099**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
92	80	60	12	0.862	0.914	0.061	**0.057**	0.01	**0.002**	**0.045**	**0.070**	0.780	**0.768**	0.05	**0.002**	**0.669**	**0.857**
93	60	60	12	0.883	0.914	0.037	**0.033**	0.01	**0.002**	**0.021**	**0.046**	0.553	**0.524**	0.08	**0.002**	**0.383**	**0.654**
94	40	60	12	0.899	0.914	0.02	**0.016**	0.01	**0.002**	**0.006**	**0.027**	0.341	**0.290**	0.08	**0.003**	**0.148**	**0.438**
95	20	60	12	0.909	0.914	0.009	0.005	0.00	0.077	−0.002	0.013	0.165	0.093	0.06	0.075	−0.029	0.236
96	0	60	12	0.913	0.914	0.004	0.000	0.00	0.539	−0.005	0.007	0.080	−0.004	0.05	0.544	−0.100	0.118
97	100	10	20	0.882	0.936	0.08	**0.054**	0.01	**0.003**	**0.030**	**0.092**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
98	80	10	20	0.905	0.939	0.058	**0.033**	0.01	**0.020**	**0.005**	**0.070**	0.860	**0.772**	0.13	**0.005**	**0.519**	**0.915**
99	60	10	20	0.918	0.937	0.043	0.017	0.01	0.111	−0.009	0.053	0.697	0.468	0.21	0.069	**0.108**	**0.727**
100	40	10	20	0.921	0.936	0.038	0.013	0.01	0.209	−0.014	0.049	0.634	0.320	0.31	0.178	−0.067	0.618
101	20	10	20	0.93	0.937	0.03	0.004	0.01	0.377	−0.021	0.039	0.542	0.112	0.31	0.364	−0.299	0.452
102	0	10	20	0.932	0.936	0.027	0.001	0.01	0.440	−0.024	0.036	0.494	−0.013	0.31	0.464	−0.453	0.357
103	100	20	20	0.894	0.945	0.064	**0.052**	0.00	**0.002**	**0.040**	**0.068**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
104	80	20	20	0.908	0.943	0.049	**0.036**	0.01	**0.002**	**0.021**	**0.055**	0.850	**0.807**	0.08	**0.002**	**0.633**	**0.924**
105	60	20	20	0.924	0.944	0.032	**0.019**	0.01	**0.008**	**0.004**	**0.039**	0.653	**0.531**	0.11	**0.004**	**0.277**	**0.738**

Table 2 (Continued)

Line	Div [%]	Sample size	N_a	H_{S_est}	H_{T_est}	G_{ST}	G_{ST_est}	SD	P	(−CI) perc	(+CI) perc	D	D_{est}	SD	P	(−CI) perc	(+CI) perc
106	40	20	20	0.933	0.944	0.024	0.011	0.01	0.090	−0.003	0.029	0.528	0.334	0.17	0.078	**0.058**	**0.584**
107	20	20	20	0.936	0.943	0.019	0.006	0.00	0.188	−0.008	0.024	0.431	0.187	0.15	0.169	−0.102	0.458
108	0	20	20	0.944	0.943	0.011	−0.002	0.00	0.585	−0.013	0.013	0.282	−0.075	0.14	0.609	−0.361	0.214
109	100	40	20	0.897	0.947	0.058	**0.052**	0.00	**0.002**	**0.046**	**0.060**	1.000	**1.000**	0.05	**0.002**	**1.000**	**1.000**
110	80	40	20	0.913	0.947	0.042	**0.035**	0.00	**0.002**	**0.026**	**0.046**	0.827	**0.803**	0.08	**0.002**	**0.683**	**0.897**
111	60	40	20	0.926	0.946	0.028	**0.021**	0.00	**0.002**	**0.012**	**0.032**	0.631	**0.570**	0.09	**0.002**	**0.409**	**0.717**
112	40	40	20	0.938	0.947	0.015	**0.009**	0.00	**0.009**	**0.001**	**0.019**	0.411	**0.290**	0.09	**0.006**	**0.096**	**0.480**
113	20	40	20	0.941	0.947	0.011	0.005	0.00	0.066	−0.002	0.015	0.322	0.175	0.05	0.056	−0.020	0.373
114	0	40	20	0.945	0.945	0.006	−0.001	0.00	0.583	−0.006	0.007	0.171	−0.023	0.06	0.590	−0.194	0.170
115	100	60	20	0.897	0.947	0.057	**0.053**	0.00	**0.002**	**0.049**	**0.059**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
116	80	60	20	0.913	0.948	0.04	**0.036**	0.00	**0.002**	**0.028**	**0.044**	0.823	**0.807**	0.05	**0.002**	**0.711**	**0.883**
117	60	60	20	0.928	0.948	0.024	**0.020**	0.00	**0.002**	**0.013**	**0.028**	0.593	**0.547**	0.08	**0.002**	**0.410**	**0.673**
118	40	60	20	0.938	0.948	0.014	**0.010**	0.00	**0.003**	**0.004**	**0.017**	0.400	**0.321**	0.09	**0.002**	**0.170**	**0.469**
119	20	60	20	0.943	0.948	0.009	**0.005**	0.00	**0.035**	−0.001	0.011	0.267	**0.161**	0.08	**0.031**	**0.013**	**0.318**
120	0	60	20	0.947	0.948	0.004	0.000	0.00	0.540	−0.004	0.005	0.127	−0.012	0.06	0.545	−0.135	0.134
121	100	10	32	0.918	0.954	0.06	**0.034**	0.01	**0.008**	**0.014**	**0.065**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
122	80	10	32	0.932	0.954	0.046	0.020	0.01	0.062	−0.002	0.051	0.865	**0.725**	0.17	**0.020**	**0.446**	**0.865**
123	60	10	32	0.939	0.953	0.037	0.012	0.01	0.167	−0.010	0.043	0.758	0.462	0.31	0.125	**0.116**	**0.691**
124	40	10	32	0.946	0.954	0.031	0.005	0.01	0.308	−0.016	0.034	0.677	0.195	0.45	0.277	−0.203	0.484
125	20	10	32	0.949	0.954	0.028	0.002	0.01	0.413	−0.019	0.031	0.627	0.003	0.43	0.426	−0.415	0.323
126	0	10	32	0.952	0.955	0.025	0.000	0.01	0.492	−0.021	0.028	0.597	−0.181	0.56	0.526	−0.624	0.159
127	100	20	32	0.929	0.962	0.045	**0.033**	0.00	**0.002**	**0.022**	**0.047**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
128	80	20	32	0.939	0.961	0.033	**0.021**	0.00	**0.004**	**0.009**	**0.036**	0.832	**0.755**	0.09	**0.002**	**0.563**	**0.888**
129	60	20	32	0.945	0.96	0.026	**0.014**	0.01	**0.018**	**0.001**	**0.030**	0.705	**0.547**	0.12	**0.007**	**0.296**	**0.742**
130	40	20	32	0.955	0.961	0.017	0.005	0.00	0.194	−0.006	0.019	0.541	0.228	0.18	0.165	−0.076	0.492
131	20	20	32	0.958	0.961	0.014	0.002	0.00	0.322	−0.009	0.016	0.475	0.090	0.16	0.310	−0.215	0.370
132	0	20	32	0.96	0.961	0.013	0.000	0.00	0.446	−0.010	0.014	0.438	0.002	0.22	0.437	−0.314	0.293
133	100	40	32	0.932	0.965	0.039	**0.033**	0.00	**0.002**	**0.028**	**0.039**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
134	80	40	32	0.943	0.964	0.027	**0.020**	0.00	**0.002**	**0.013**	**0.028**	0.794	**0.747**	0.07	**0.002**	**0.608**	**0.858**
135	60	40	32	0.951	0.964	0.019	**0.013**	0.00	**0.002**	**0.006**	**0.021**	0.637	**0.540**	0.08	**0.002**	**0.360**	**0.695**
136	40	40	32	0.959	0.964	0.011	0.005	0.00	0.059	−0.001	0.012	0.422	**0.230**	0.10	**0.042**	**0.031**	**0.428**
137	20	40	32	0.961	0.964	0.009	0.003	0.00	0.126	−0.003	0.010	0.369	0.151	0.09	0.106	−0.055	0.357
138	0	40	32	0.963	0.964	0.007	0.001	0.00	0.442	−0.005	0.007	0.285	0.022	0.14	0.444	−0.176	0.231
139	100	60	32	0.932	0.965	0.038	**0.034**	0.00	**0.002**	**0.030**	**0.038**	1.000	**1.000**	0.00	**0.002**	**1.000**	**1.000**
140	80	60	32	0.945	0.965	0.024	**0.020**	0.00	**0.002**	**0.015**	**0.026**	0.782	**0.749**	0.06	**0.002**	**0.643**	**0.842**
141	60	60	32	0.952	0.965	0.017	**0.013**	0.00	**0.002**	**0.008**	**0.019**	0.613	**0.544**	0.09	**0.002**	**0.404**	**0.672**
142	40	60	32	0.96	0.965	0.009	**0.005**	0.00	**0.009**	**0.001**	**0.010**	0.397	**0.268**	0.09	**0.008**	**0.112**	**0.425**
143	20	60	32	0.963	0.965	0.006	0.002	0.00	0.126	−0.002	0.007	0.282	0.114	0.07	0.116	−0.046	0.280
144	0	60	32	0.965	0.966	0.005	0.000	0.00	0.371	−0.003	0.005	0.219	0.026	0.06	0.369	−0.127	0.192

With increasing number of alleles, the standard deviation of D_{est}-values became very large for small sample sizes. When original populations were identical (0% divergence) and consisted of 32 different alleles total, D_{est}-values ranged widely and could reach 0.49 when only 10 individuals were sub-sampled per population (Table 2, line 126: mean D_{est} = −0.181 ± 0.56 SD, single values ranged from −1.47 to 0.49). This result illustrates clearly how large errors can occur when sample sizes are too small in relation to a large number of alleles. Even for a sample size of 20 individuals per population (with 20 alleles, and 40% divergence between populations), mean D_{est} = 0.228 ± 0.178 SD (Table 2, line 130), but single D_{est}-values ranged from −0.267 to 0.54. Here, the calculation of statistical significance is necessary to show whether D_{est}-values are significantly different from a random distribution. Independent of the method for calculating genetic differences, our example shows that such a sampling regime cannot produce reliable results. We suggest taking negative D_{est}-values as a potential warning signal that sample sizes might be too small compared to the number of alleles. So far it was also suggested to replace negative values as zero [see description of the computer program SMOGD which calculates D_{est}-values and CI (Crawford 2010)]. G_{ST_est} values could also become negative, but because the values of G_{ST_est} themselves will be much smaller, the standard deviation of G_{ST_est} will be much smaller than those of D_{est}-values.

The question remains whether P values or confidence intervals should be used to decide about statistically significant differences between populations. Statistically

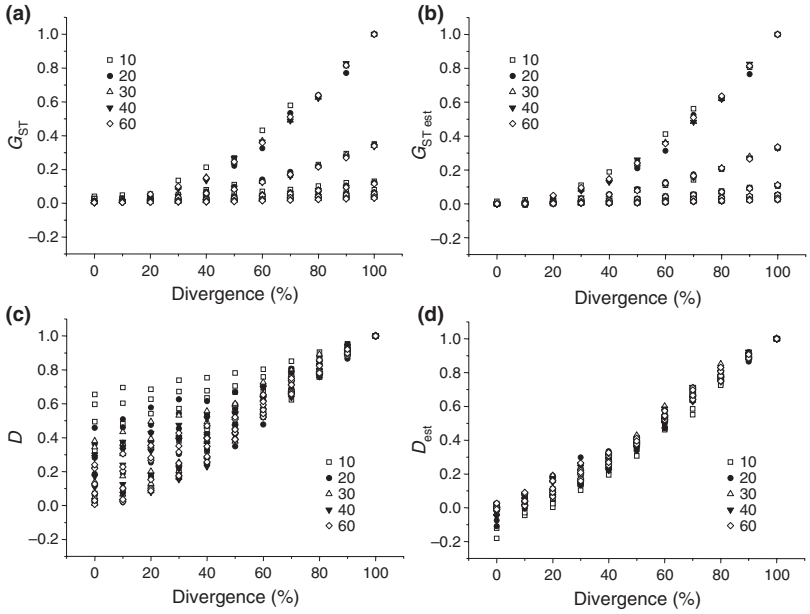


Fig. 3 (a,b,c,d) Comparison of bias corrections calculated for populations with increasing divergence, different numbers of alleles and different sample sizes. (a) Uncorrected G_{ST}- and (b) bias-corrected G_{ST_est}- (c) uncorrected D- and (d) bias-corrected D_{est}-values. Each of the five symbols occurs six times reflecting genetic differences calculated for six different numbers of alleles (N_{a_total} = 2, 4, 8, 12, 20 or 32). Because of almost equal values, symbols overlap. Box shows symbols according to different sample sizes per population (10, 20, 30, 40 and 60 individuals).

significant P values show whether genetic differences between populations are larger than generated by chance. By using very large sample sizes, P values will become significant even when differentiation between populations is small; however, considering past studies in population genetics, the real problem tends to be sample sizes that are too small rather than too large. Calculating genetic distances based on small sample sizes without considering statistical significance can create serious misinterpretation.

There are several methods to calculate confidence intervals considering the asymmetry of the bootstrap distribution that occurs especially when sample sizes are small [normal, percentile, basic, accelerated, studentized or bias corrected, accelerated (BCA) bootstrap] (DiCiccio & Efron 1996). To our knowledge, there is no uniformly superior and universally accepted method to calculate bootstrap confidence intervals. We used two different methods: bootstrapping estimated half of the width of the 95%-confidence interval as 1.96 * SD (normal method), centred at the D_{est} estimator. Shown in Table 2 we used a second method (similar to the program SPADE, Chao & Shen 2009; see also Chao *et al*. 2008), which calculates confidence intervals following the percentile method. This seems appropriate when distributions of bootstrap values are asymmetrical; however, the problem of asymmetry still persists for some distributions.

Table 2 shows that in most cases P values were consistent with confidence intervals based on the percentile method. Only for few cases, this last method generated statistically significant results for D_{est}-values when P values did not (Line 99, 106, 123 Table 2). For instance, at low sample sizes (n = 10) and a high number of alleles (N_{a_total} = 20 or 32), P values indicated a significant difference between populations when the 'real' divergence was 80%; confidence intervals indicated statistical significant differences when the 'real' divergence was 60%.

Bias correction of G_{ST} and D

Drawing random samples from infinite populations produces a bias in calculating G_{ST}- and D-values (see Table 2 and Fig. 3) (differences between real divergence between populations and D-values). Bias corrections for G_{ST}-values had been suggested by Nei & Chesser (1983). A similar bias correction was suggested by Jost (2008) for D-values. In Fig. 3a–d, we show the effect of bias corrections of G_{ST}- and D-values. While the bias correction was small for G_{ST}-values, D_{est}-values differed much more from D-values. We compared the relative change for G_{ST} and D introduced by the bias correction and found that the one for D is more pronounced. With a slight underestimation, D_{est} correctly showed the true divergence that existed in the populations from which samples were drawn with larger samples sizes showing a better result.

Summary and conclusions

By using artificial data sets with known divergence between two populations, we could test the accuracy of different methods of calculating genetic differences. We found that for more than two alleles, G_{ST_est}-values do not clearly reflect genetic differences while D_{est}-values do. Our simulations support the use of G_{ST} as a fixation index that increases with the decrease of allele diversity caused by genetic drift or selection. The maximum value of G_{ST} is dependent on the averaged homozygosity over populations which was more generally described by Nei (1973) and Hedrick (2005). Thus, G_{ST} is only appropriate to measure population differentiation for two alleles (see Table 2). This reflects that G_{ST} measures fixation and D differentiation.

In randomly generated data sets with different allele frequencies, bias corrections were smaller for G_{ST}- than for D-values, and D_{est}-values reflected the 'real' divergence better

than G_{ST_est}-values. However, negative D_{est}-values could occur especially when sample size was too small for the given number of alleles.

Calculations of statistical significance and confidence intervals (e.g. based on bootstrapping) are essential for estimating if genetic differences are meaningful. We could show that statistical significance did not differ when genetic difference was calculated with G_{ST_est} or D_{est}. Therefore, while prior calculations of G_{ST_est}-values do not show differentiation between populations, prior calculations of statistical significance between populations are probably to remain the same. To calculate real differences between populations, analyses which used G_{ST} to rank data sets need to be redone. We clearly want to state that D_{est}-values in combination with calculations of statistical significance will give a much better and appropriate description of genetic difference between populations.

The package '*DEMEtics*' to calculate D_{est}, G_{ST_est}-values as well as confidence intervals and *P* values can be requested from the authors or downloaded (http://cran.r-project.org/web/packages/DEMEtics/index.html).

Acknowledgements

We thank Lou Jost and Anne Chao for very helpful comments and Jelle Atema for reviewing the manuscript and improving the English.

References

Chao A, Shen T-J (2009) Program **SPADE** (Species Prediction and Diversity Estimation). Program and user's guide at http://chao.stat.nthu.edu.tw.

Chao A, Jost L, Chiang SC, Jiang Y-H, Chazdon R (2008) A two-stage probabilistic approach to multiple-community similarity indices. *Biometrics*, **64**, 1178–1186.

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals (with Discussion). *Statistical Science*, **11**, 189–228.

Goudet J, Raymond M, Demeeüs T, Rousset F (1996) Testing genetic differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Hedrick P (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Heller R, Siegismund R (2009) Relationship between three measures of genetic differentiation GST, DEST and G'ST: how wrong have we been? *Molecular Ecology*, **18**, 2080–2083.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Jost L (2009) D vs. GST: response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology* **18**, 2088–2091.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, **70**, 3321–3323.

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, **47**, 253–259.

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ryman N, Leimar O (2009) GST is still a useful measure of genetic differentiation – a comment on Jost's *D*. *Molecular Ecology*, **18**, 2084–2087.

Gabriele Gerlach is broadly interested in evolutionary ecology, particularly in the application of molecular markers to ecological questions. Her current attention concentrates on the impact of behavior and adaptations on the genetic structure of populations. Alexander Jueterbock, Philipp Kraemer and Jana Deppermann are graduate students working on the genetic structure, dispersal and genomics of different marine organisms in Gerlach's lab. Peter Harmand is a mathematician specialized in analysis and statistics.

doi: 10.1111/j.1365-294X.2010.04784.x

Chapter 3

Population structure with emphasis to local adaptation in *Cancer pagurus*

3.1 Introduction

3.1.1 Population genetics in marine crustaceans

Studying marine crustaceans is of major importance, due to their role in ecosystem functioning. Marine crustacean species such as high abundant small pelagic copepod species serve as a link between primary ocean production and predatory fish (91). Benthic decapods are important for reintroducing biomasses into the foodweb as scavengers (92) or as benthic predators (93; 94). Furthermore, they may act as important habitat engineers (95; 96). These broad functions of crustaceans are often maintained in different areas by many precisely adapted species to local conditions. In the "light of evolution" (97) this established the enormous diversity of crustacean species - for the ocean roughly 68,000 crustacean species are estimated (98). A bigger diversity is found at lower latitudes (99). This pattern is discussed as the latitudinal biodiversity gradient (LDG) (100). However, diversity in higher latitudes is probably often underestimated. From highly distributed species, new species emerge from cryptic diversity between populations (101). Over time favorable adaptations may lead to a selective sweep in a population or even in the complete distributional range of a species (102). On a geographic scale especially in species with a high distribution and high population structuring this may lead to differently adapted populations in different areas, for example discussed for salmonid fishes (103). In the course of recently developed sequencing technologies many studies discovered genetic differences between populations of crustaceans in temperature and arctic zones (104). Some studies detected a correlation with phenotype variances as well (105; 106). Studying diversity in species is important to understand ecosystem functioning and the effect of future selection pressure for ecological as well as economical reasons. Climate change triggers the distributional range of species with considerable effects of species including local adaptations in populations (107). Examples for crustacean species showing different adaptations to different local conditions are rare up to now. From controlled experiments comparing fitness in different populations by either common garden experiments or transplant experiments only 11 studies could describe adaptation for crustaceans (108). Only two cases were described for decapod crustaceans. Most of these sources for adaptation are temperature conditions. However, this may be heavily biased due to the proxy nature of temperature being connected to several biotic as well as abiotic selection pressures. There are several possible sources for selection known for marine crustaceans described either for a geographic or for a time scale.

3.1.2 Processes under selection in marine crustaceans

Crustaceans belong to the invertebrates characteristically possessing a poikilothermic physiology. Thus, to deal with different water temperatures is not just important for maintaining the energy budget of individuals, but also to directly deal with altered conditions on the level of cell physiology to keep homeostasis (109). A very vulnerable phases for crustaceans is the larval phase (110). On the one hand water temperature triggers timing and speed of ontogeny (111) limiting the distributional range of larvae. On the other hand tolerance towards temperature conditions of planktonic larvae may directly limit the distributional capacity of the species (112). Temperature can even lead to unsuccessful larval development by increased mortality (113). This temperature sensitivity may be stronger in larvae than for the adult form (114). There is evidence that genetic variation within species leads to differently adapted populations in geographic regions with different water temperatures. For example for the fiddler crab transplant experiments did show differently adapted individuals, hence there was no fixed population differentiation for this crab (115). In crustaceans mating is often linked to molting cycles since only priorly molted females get buried (116). If altered temperature conditions disturb timing of spawning this may lead in the end to massive mortalities, due to not optimal environmental

conditions for larvae. The optimal range of temperature to successfully finish larval development may be relatively small, even smaller than the ambient temperatures for the adult (113). This range may thus be heavily under selective pressure and may have the potential to be differently selected in different areas (115).

Direct effects of temperature on crustaceans of any ontological stage are buffered by well known proteins on the cell physiological level (117; 109). During temperature stress proteins in cytoplasm may become instable. Several very conserved protein cascades have been developed to maintain cell functioning even during long periods of wrong environmental conditions. These so called chaperons actively re-fold structures of proteins, which may become unstable and thus lose their activity during wrong temperature conditions. Heat shock proteins, such as Hsp70 or Hsp90 are not limited in their activity to heat stress, but may be increased in their expression during different forms of environmental stress (118; 119; 120). One protein of central importance for cell functioning during stress response is ubiquitin (Nobel Prize in Chemistry in 2004 (121)). This conserved protein may be affected by several stressors and is, for example, included in the ubiquitin-proteasome pathway to remove damaged proteins after temperature stress (122) or oxidative stress (123). Temperature may indirectly be selective on crustaceans as well. The concept of "oxygen and capacity limited thermal tolerance" (124) describes how temperature and its effect on oxygen availability and utilization limits the distributional range of species. With increasing water temperature the capacity of O_2 which can be solved in water is decreased. Moreover, increasing water temperature leads to increased biological activity and, hence, oxygen consumption. Decreasing levels of O_2 are called hypoxia and may lead to anoxic conditions in cell physiology for marine invertebrates. Hypoxic zones, the so called "oxygen minimum zones", are "likely" to be increased by climate change especially in temperate and sub-polar regions (125). This heavy selection pressure is already present in local embayments and will further spread more independent from latitudes than temperature driven by hydrographic and anthropogenic effects (125). Reactions to hypoxia are known to be increased ventilation activity or migration into more favorable habitats (126). Long term skewed levels of CO_2 and O_2 may however have considerable effects on the cell physiology. Recent studies found that stenotherm antarctic fish acclimated to higher temperatures switch from lipid based to carbohydrate based metabolism, probably to be prepared for hypoxic conditions by anoxic metabolism (127). It is known that several metabolic answers to anaerobiosis persist in invertebrates such as phosphokinases buffering acidosis and low ratios of ATP/ADP, aspartate and glycogen based metabolism or lactat and opin creation by glycogen fermentation (126). These metabolic pathways consist of several enzymes with specific activities. In contrast to the omnipresence of such pathways among invertebrates, early studies showed genetic variance within species and between species of crustaceans (128), which may indicate its importance for adaptation and speciation under different environmental conditions. One of the important facts is that temperature as well as oxygen stress may interfere with the specific immune system of crustaceans. The crustacean immune system, the "innate immune system", is conserved among the phylum crustacea and well characterized on the molecular level (129). There are several studies about aquaculture and wild crustacean's immune and stress related response on the molecular level (130; 47). According to them, higher temperature is known to suppress the immune system of crustacean species and may weaken their antimicrobial resistance. The central players are the hemocytes, which react on microbial pathogens with sequestration or direct killing of pathogens and secretion of different types of antibacterial proteins (131). A central protein cascade activated in the presence of alien substances is the prophenoloxidase cascade (ProPO) (132). Affiliated transcripts may be up-regulated in case of pathogenic challenge and are itself dependent on activity of several other proteins, for example on the b-glucan-binding protein activity, which is one of the first steps in activating hemocyte mediated defense reactions (132). Antimicrobial immune reactions of crustaceans were determined for several organs of the host such as the

hepatopancreas and the outer shell (133; 134). Temperature driven environmental effects may thus be selective on three important levels of crustacean biology the level of protein stability, the level of cell metabolism and the level of immune defense.

3.2 Summary of findings and own contributions to the manuscript

Manuscript V: Characterization of transcripts of *Cancer pagurus* acclimatized to different temperature regimes and the population structure on a latitudinal gradient

I use a combined approach of finding candidate genes as potential indicator for adaptation in brown crabs from two different latitudes Norway and Germany (Helgoland). Experimentally exposed crabs to different temperature regimes are studied. By 454 sequencing I analyze a pool of 13,709 transcripts, 42 differently expressed EST's and 5 polymorphic loci to study local adaptation in *C. pagurus* to climate change mediated temperature changes in local habitats. To understand whether *C. pagurus* populations show differences in adaptation in the wild I compare 5 new polymorphic EST-linked loci with published neutral SSRs and analyzed the genetic structure among several European populations of *C. pagurus* from Portugal to Norway. The occurrence of shell disease as a maladaptation to environmental conditions is examined. The data indicate that probably due to large population size and long larval dispersal the probability to maintain differences in adaptation between populations is low. Small D_{EST} and G_{ST} values indicate high connectivity and the potential to shift northwards towards colder climate; however the existence of shell disease even in northern populations shows strong pathogen stress in local populations. There is evidence that different sub-populations in Norway are differently affected by shell disease indicating differences in adaptation of residents and possible immigrants towards local conditions.

Own contribution to the study: The idea of the manuscript was developed by me and Prof. Gerlach. I took the samples with help of different local scientific authorities or fishermen, except for the population at Bergen, Norway. I developed the experimental setups at both marine stations in Norway and Helgoland and conducted the animal care and experiments during several weeks at scientific stays at the field station. I prepared the samples for massive parallel sequencing. Sequencing was conducted at the Institute of Clinical Molecular Biology, Kiel. I did all the statistical analyses for developing the transcriptome information, gene expression analysis and development of new microsatellite markers. I developed the protocols for the markers and run 50 % of the microsatellites on the sequencer. Remaining lab work was done by a technical assistant under my supervision. I conducted all analyses on transcriptome analysis, population genetics and shell disease analysis for each population. I wrote the manuscript, which was reviewed by Prof. Gerlach.

Manuscript V

Characterization of transcriptomes of *Cancer pagurus* acclimatized to different temperature regimes and the population structure on a latitudinal gradient

in form of a manuscript

Characterization of transcripts of *Cancer pagurus* acclimatized to different temperature regimes and the population structure on a latitudinal gradient

Running title: Population structure of *Cancer pagurus*

Abstract

Studies of local adaptation provide important insights into the power of natural selection relative to gene flow. But correlations of environmental factors and phenotype variations are difficult since the latter often bear uncertainty of the origin of variation. Here we aim to disentangle variation due to phenotype plasticity and genetically based local adaptation in the brown crab *Cancer pagurus* to understand the impact of genetic diversity and genetic adaptation on survival of marine crustaceans in a changing climate. We used a combined approach of finding candidate genes as potential indicators for adaptation in brown crabs from two different latitudes Norway and Germany (Helgoland). We experimentally exposed crabs to different temperature regimes and obtained via 454 pyrosequencing a pool of 13,709 transcripts, 42 differently expressed EST's and 5 polymorphic SSR loci. To understand whether *C. pagurus* populations show differences in adaptation in the wild we compared transcriptome derived polymorphic EST-linked loci with neutral SSRs and analyzed the genetic structure among several European populations of *C. pagurus* from Portugal to Norway. We examined the occurrence of shell disease as a maladaptation to environmental conditions. Our study shows that probably due to large population size and long larval dispersal local adaptation seems to be low. Instead small D_{EST} and G_{ST} values in both marker types indicate high connectivity and gene flow among populations suggesting a northward shift towards colder climate driven by northward current regimes; However, the existence of shell disease even in northern populations indicates stress related immune deficiencies in local populations. We found evidence that different sub-populations in Norway were differently affected by shell disease indicating differences in adaptation of residents and possible immigrants towards local conditions.

Keywords: Local adaptation, shell disease, hemocyanine, decapoda

Introduction

Future climate change will affect the distribution of marine species and may have the potential to alter migration and the fitness of local populations [1]. The ability to cope with future climate change will depend on the species' plasticity to local environmental conditions or its ability to shift into more suitable habitats. For example predictions on a distributional shift of *Zostera noltii* showed that during the 21^{st} century a shift of several hundreds of kilometers northward is very likely as a consequence of increased temperature and sea-level rise [2]. Such northward shifts were described as a general pattern for many species [1, 3, 4]. From paleontological data, several mollusc's were found stable in their tolerated conditions and did react to environmental change by shifts in distribution or extinction [5]. Whether the alternative - local adaptation to rapidly changing environmental conditions - might happen in the wild is assumed but hardly understood [6]. We studied the brown crab *Cancer pagurus*, which has a wide distribution at the coast of Europe from the Sub-Arctic to the Mediterranean and is a very abundant species of the family Cancridae within the brachyurans in this area. The species is facing various environmental conditions and

different water temperature in its habitats. This caused the question whether the species is adapted to a wide variance of environmental conditions by phenotype plasticity or whether locally adapted populations persist. To date, there is very limited information on population structure [7] and nothing is known on local adaptations to environmental conditions due to a lack of suitable genetic marker. *C. pagurus* is an economically important species with high catch rates in the northern Atlantic Ocean [8, 9]. Due to its abundance and size the species is also an important ecological player in local habitats, e.g. as scavenger [10], habitat engineer by pit digging activities [11,12] or as a predator of key organisms of local food webs such as sea urchins or limpets [13,14]. Understanding the population structure and the adaptive potential of *C. pagurus* will provide important knowledge to assess future survival of this species and may be a good model for other marine organisms.

Next Generation Sequencing studies gives us the opportunity to increase our knowledge in genetic information for non-model species. We aimed to find candidate genes and genetic variable sites in ESTs for studying population structure and potential genes under selection in comparison to neutral SSRs in *C. pagurus*. Since temperature is important for this poikilothermic species we analyzed transcriptomes of animals experimentally exposed to different temperature ranges.

Several studies focused on temperature effects on poikilothermic species [15–17]. In most experiments animals experienced a sudden shock to high or low temperature conditions, providing information on stress related transcripts, which act as chaperone's on the cell physiological level. Chaperone's are well characterized on the genetic and physiological level in crabs [18] and lobsters [19, 20]. Differential expression of those transcripts might indicate whether animals are stressed by temperature or whether they can acclimate to treatment temperature conditions in short range experiments. Acclimatization experiments rather than "shock" experiments may serve as a better source for interesting transcripts and their quantitative expression for adaptation to environmental conditions. Temperature treatments within a non-lethal range may differently affect gene expression in crustaceans. For example hemocyanines, the main oxygen carriers in crustaceans, are heavily dependent on environmental temperature conditions and their gene transcription may differ in habitats with different temperature conditions [21]. Different types of hemocyanines are additionally important for the immune defense in crustaceans [22]. The expression of many immune related genes are affected by heat stress, for example in penaeid shrimp antimicrobial peptides such as crustins and paenaedins [23], serine protease or prophenoloxidase among others [23] showed different expression levels depending on temperature. A broad overview on transcripts related to temperature conditions are important to understand the physiological temperature acclimation on the molecular level. For example, in fish temperature acclimatization had been shown to lead to a changed metabolism and was discussed to be dependent on only few transcripts [24]. Temperature might be only one factor that animals have to adapt to when global climate changes [25]. Many of those are directly related to local temperature conditions such as oxygen levels in habitats and oxygen consumption of animals. Recently, the effect of hypercabnia was analyzed on the transcriptome level leading to a transcript map of *Hyas arenarius*, the spider crab [26]. In the wild, maladaptation to environmental conditions may directly lead to an increased susceptibility to diseases such as the common shell disease. Shell disease is known to differ significantly among populations of North American lobster *Homarus americanus* living under different local temperature regimes [27]. However, several other environmental conditions such as increased pollution were suspected to favor susceptibility to diseases such as shell disease [28, 29]. The disease status in populations may often be linked to the ability to express immune related genes and thus may be one important class of genes being under strong selection in animals living under different environmental conditions. In the

course of global climate change these genetic variants may be essential for a successful adaptation to future climate change scenarios.

In this study we aimed to understand the fate of the brown crab as a crustacean model in a changing climate. We performed temperature acclimation experiments, transcriptome analysis, examined the genetic structure of wild populations, determined the degree of shell disease as an indicator for maladaptation to understand whether the brown crab can cope with temperature increase, adapt locally or shift to more suitable habitats.

Methods

Temperature treatment Adult female *Cancer pagurus* were caught during diving operations or with baited traps from local populations in Helgoland, Germany in July 2010 and Bodø, Norway in October 2010 (Table 1). Crabs were kept in semi-circulated tanks supplied with local sea water. Specimen were acclimatized to laboratory conditions for two weeks. Temperature was set to normal ambient conditions, 14.5 °C Helgoland (HE.A) and 7.5 °C Bodø (NO.A). Two crabs from Norway were slowly exposed to 16 °C (1 °C / 100 min) (NO.W) and crabs were held under these conditions for 120 h. Crabs were not fed during treatment. Thereafter, all crabs from Helgoland and Norway were sacrificed by disconnecting ventral and dorsal neural ganglions and a piece of the hepatopancreas was dissected and incubated in RNA-later (Qiagen) and frozen at −20 °C for storage until RNA extraction.

Sample preparation RNA was extracted from hepatopancreas from 6 animals (2 samples for each treatment: Helgoland (normal) Warm (HE.A), Norway (normal) Cold (NO.A), Norway (treatment) Warm (NO.W)) using MasterPure Complete DNA and RNA Purification Kit (Epicentre) following the standard protocol of the instructor for RNA extraction including removal of contaminating DNA (digestion time of 30 min). RNA quality was subsequently checked on stained agarose gels and a RNA Chip using Bioanalyzer. cDNA was reverse transcribed by SMART cDNA kit (Clontech Laboratories). cDNA of animals from the same temperature treatments was pooled and ligated to either standard Roche MIDs RL1, RL2, RL9, RL10, RL11 and RL12 for later identification. Barcoded and pooled samples were sequenced on 454 GS FLX Sequencer (Roche) at Institute for klinische Medizin Kiel.

Assembly In order to detect Expressed sequence tags (EST) from 454 sequencing we used MIRA [30] as an assembler outperforming many others and being stricter in contig assembly i.e. not omitting variation by contig alignments [31] and the Newbler GS de Novo Assembler (ver. 2.6; Roche) approach, which gathers reads into contigs in a more liberal way [31]. Reads were assembled by Newbler according to standard parameters, trimming points for MIDs and SMART cDNA primer sites as well as on information of quality. In a similar way we used MIRA assembler to calculate a second batch of contigs. This raw reads were trimmed for MIDs and SMART cDNA primer sites. SnoWhite (Version 1.1.4) [32] increased the quality in reads prior to assembly with MIRA [33]. Resulting contigs from both approaches were combined using CD-HIT to calculate a common pool of EST contigs corrected for technical redundancy [34]. Redundant reads were gathered in clusters with a similarity of at least 95 %. Remaining contigs were merged into one pool of transcripts, which was the basis for further analysis. Annotation was performed by mapping in Blast2Go against the several protein databases including UniProt using blastx routine with the cut-off value of 10^{-3} [35].

Contigs which did not annotate sufficiently were annotated manually against nucleotide and EST db using blastn with a cut-off value of 10^{-3}.

Quantitative genetic In order to analyze differences in gene expression in animals of different treatments, we assembled the pool of reads of all samples using non-redundant merged contigs by CD-HIT. Against this assembly a mapping of RAW reads was performed using gsMapper (Roche). In order to understand differences in expression of transcripts between samples we used the edgeR package [36]. Pools were normalized for absolute read counts. Dispersion was estimated and for each pairwise treatment exact negative binomial tests were performed [37]. Significance values were corrected for multiple comparisons using false discovery rate approach according to Benjamini and Hochberg [38].

Genetic variation among wild populations of *Cancer pagurus* The full set of contigs from animals of Norway and Helgoland was filtered for Retroelements (LINES, SINES, LTR elements), DNA transposons, Rolling circles, Small RNAs, Satellites, Simple Repeats and sites of low complexity using RepeatMasker [39]. MISA was used [40] for a more accurate look on the distribution of SSR motifs in contigs. The full set of contigs served as a reference for posterior read mapping using Newblers GsMapper (ver. 2.6; Roche). Contigs, which showed sequence variation between samples, were selected to develop loci for population genetic analysis. Primer were developed using Primer3 [41, 42]. Primer were tested using tissue samples of *C. pagurus* preserved in ethanol 96 %. For DNA extraction, a small piece of tissue was mixed with 200 µl of 1 % of Chelex (Biorad) and incubated at 56 °C for 10 minutes. PCR was run using FastStart High Fidelity taq (Roche) with a total volume of 10 µl containing 0.5 µl template DNA, 6.6 µl ddH_2O, 0.1 µl DMSO, 1 µl FastStartBuffer including 18 mmol MgCl, 0.5 µl of either forward and reverse primer (10 µmol), 0.2 µl dNTP (2.5 mM) and 0.1 µl HighFidelityTaq (5 U/µl, Quiagen). PCR conditions were set as follows: Prior denaturation at 95 °C for 2 min followed by 35 cycles of denaturation at 95 °C for 30 sec, annealing at 60 °C for 30 sec and elongation at 72 °C for 30 sec and a final elongation step of 72 °C for 6 min. PCR products were checked on integrity and on contamination in controls via 2 % agarose gel electrophoresis using GelGreen stain (Biotium). Length polymorphism was checked using polyacrlyamid gel electrophoresis. 10 % polyacrylamid gels were prepared using 10 ml acrylamid (40 % [29:1] Roth), 10 x TBE buffer, 25.3 ddH_2O, 670 µl APS (10 %, Roth) and 33.3 µl TEMED (Roth). Gels were polymerized for 60 min and casted in a vertical electrophoresis chamber (Biometra) and run at 20 °C, 160 V for 12 h with 1 x TBE buffer. Gels were stained with GelGreen (Biotium) and visualized on a blue transilluminator imaging system (peqlab). To summarize loci information images were analyzed using Gimp (Version 2.7) and Demerelate [43]. In order to determine the population structure among European populations of *C. pagurus*, individuals from ten populations were sampled (Table 1) and tissues were preserved in 96 % Ethanol until used in the lab.

We determined fragment lengths of 12 SSR loci on Beckman CEQ 2000 XL capillary sequencer. PCR was conducted with forward primers labeled with blue, green or black WellRED dye at the 5' end (D2-PA, D3-PA, D4-PA; Sigma Aldrich). PCR was either based on redtaq polymerase or on Quiagen Faststart polymerase with different PCR conditions specific for each locus summarized in Table 3. PCR conditions were as follows: 6.25 µl ddH_2O, 0.25 µl of either forward and reverse primer (10 µmol), 1 µl dNTP, 1 µl red taq buffer and 0.25 red taq polymerase adding 1 µl DNA template. Quiagen Faststart recipe was as follows: 2.95 µl ddH_2O, 2 µl Q-solution * 5, 2 µl MgCl (25 mM). 0.25

Table 1. Summary of sampling locations with GPS positions and samples size (N) and average F_{IS} estimates. Asterisks indicate statistical significance ($p < 0.05$) when the proportion of F_{IS} giving a smaller or bigger value than observed was < 0.95 after 2400 randomizations. Bold values are significant after adjusted for false discovery rate in multiple comparisons (FDR) [44].

Population	N	GPS position (N)	GPS position (E)	F_{IS}
Denmark (DK)	63	56°10' 13.342"	9°30' 6.425"	0.062*
Spain (ESP)	18	42°14' 26.156"	8°43' 14.616"	0.052
Great Britain (GB)	50	50°30' 13.068"	4°39' 8.993"	0.045
Germany (HE)	42	54°10' 49.176"	7°53' 20.197"	0.032
Norway Bodø (NO)	68	67°17' 2.299"	14°32' 55.265"	**0.110***
Norway Bergen (BE)	78	60°23' 28.546"	5°19' 19.395"	0.020
France (RO)	49	48°43' 34.316"	3°59' 7.169"	0.069*
Germany (TR)	50	54°10' 49.176"	7°53' 20.197"	0.026
Germany (WHV)	20	53°37' 50.53"	8°1' 32.869"	0.096

of either forward and reverse primer (10 µmol), 1 µl dNTP (2.5 mM), 1 µl 10 * buffer (15mM $MgCl_2$) and 0.05 Hot Start taq polymerase (5 U/µl) adding 1 µl DNA template. Fragment length of alleles were scored by eye using default parameters from Beckman CEQ workbench. All loci were analyzed for linkage disequilibrium using Fstat version 2.9.3 [44]. In order to test neutrality in "Neutral"-SSRs [45] EST-SSRs we compared allele diversity by F_{ST} with the expected heterozygosity for each locus H_e applying MMC statistics using the software Lositan [46]. Population genetic structure was analyzed based on markers not affected by LD or selection. Pairwise genetic difference between populations were estimated by D_{EST} and G_{ST} [47, 48]. F_{IS} values were calculated using Fstat routines version 2.9.3 [44]. Exact test for differences in allele composition between populations and effektive population sizes were estimated using Genepop routines version 4.2 [49, 50]

Determination of shell disease Occurrence and severity of melanization due to shell disease was analyzed in animals from Norway and Germany using a stencil on pictures of the outer carapace [51]. The stencil was subdivided into equal sized areas and severity of the disease was calculated for each individual by counting the numbers of affected carapace areas based on pictures from each animal. Blackening of carapaces was assessed as follows: 1 - light brownish spot only on surface (not classified as shell disease), 2 - clear signs of shell disease up to heavy shell disease with erosion progressing through carapace layers, 3 - blackening due to injuries (Individuals were removed from the analysis). Shell disease is analyzed as a potential maladaptation in *C. pagurus* to bacterial challenge. Therefore only clear signs of shell disease which most likely have consequences for the internal health of the host were recorded (category two).

Results

Deep sequencing of *C. pagurus* transcriptome - assembly and annotation Total non normalized RNA content was sequenced from hepatopancreas on $\frac{1}{2}$ FLX run (408,628 reads, 118,761,176 bases). Each pool was sequenced twice with following read numbers HE.A (RL1): 32,411 and 31,001 reads; HE.A (RL2): 41,959 and 40,823 reads; NO.W (RL9): 36,633 and 34,932 reads; NO.W (RL10) 47,353 and 44,723; NO.A (RL11) 26,541 and 25,506 reads; NO.A (RL12) 24,478 and 22,268 reads.

We combined all pools using the parallel approach with MIRA and Newbler assembler followed by CD-HIT clustering leading to 13,709 contigs with a minimum length of 11 bps a maximum length of 8,914 bps and an average length of 646 bps with 13,414 contigs > 100 bps (Suppl. 5).

Quantitative genetic markers We used a total of 1,166 contigs based on a cut-off approach keeping only those contigs with at least 100 counts per million reads and present in at least two biological samples. Samples from each treatment were very similar in their tagwise expression as indicated by MDS scaling of log fold change in expression level among pools (Fig. 1). The common dispersion in reads was 0.29 and the biological coefficient of variation (BCV) was 0.54 (Fig. 1). We compared animals from Germany, Helgoland (HE) and Norway (NO); since animals of both populations lived under different temperature when we caught them; we named them Helgoland and Norway ambient HE.A. and NO.A, but note that HE.A. is 14.5 °C and NO.A. is 7.5 °C. Animals from NO.A. were shifted to NO.W (16 °C). For each treatment two samples were sequenced. By using the exact test of edgeR [37] including correction for false discovery rate a significant fold change (logFC) between treatments was found for a total of 43 contigs. For the comparison of animals from both ambient conditions Helgoland warm water and Norway cold water 10 contigs were up-regulated and 7 were down-regulated in Norway cold water (Fig. 2). Comparing animals HE.A to NO.W 32 contigs were differently expressed, 20 with lower expression in NO.W and 12 with higher (Fig. 2). Comparing NO.A with NO.W one contig was down regulated by the treatment, while 9 were up regulated (Fig. 2). The remaining 36 contigs were annotateable to genebank using blastx or blastn on EST and nucleotide database. They represented different conserved transcripts responsible mainly for immune response, metabolism or reproduction. Several contigs were less frequently expressed at individuals from NO.W and HE.A maintained in warm water. This was the case for a contig annotated to ferritin and a predicted protein consisting mainly of repetitive elements (Table 2). Sodium dependent nutrient amino acid transporter and C-type lectin were decreased only by warm water treatment (Table 2). Many contigs were significantly decreased in Norway cold conditions (NO.A) compared to Helgoland warm conditions (HE.A) and followed the same trend when Norway warm (NO.W) compared to Norway cold (NO.A) such as serine carboxypeptidase (Table 2). Vitellogenin variants were consistent in fold change between treatments (Table 2), showing considerable higher abundance in Norway samples than in Helgoland samples. PKc2385 was similarly expressed in animals NO.A and HE.A conditions, but was strongly affected by temperature treatment in Norway (NO.W). Cryptocyanines were annotated by 10 contigs showing a significant fold change between treatments (Table 2). Fold change differences were consistent between contig variants, abundance was not affected by warm treatment in Norway, but was significantly different between populations (HE.A vs NO.A). One exception was cryptocianine 2, which was also negatively affected by warm water treatment in Norway (Table 2). All variants of gamma-butyrobetaine dioxygenase were decreased in NO.W to a similar level as in Helgoland warm water conditions (Table 2). Hemocyanines were found to be significantly different between treatments in three different contigs hemocyanine subunit 5 and subunit 1 were strongly decreased by warm water treatment in Norway (Table 2). Those hemocyanines were the only transcripts affected by temperature treatment and differently abundant in Helgoland warm and Norway warm conditions. The variants of different genes were relatively different, not exceding 91 % similarity. Exact values for each pairwise comparison can be found in the supplement (Suppl. 8, 9, 10, 11).

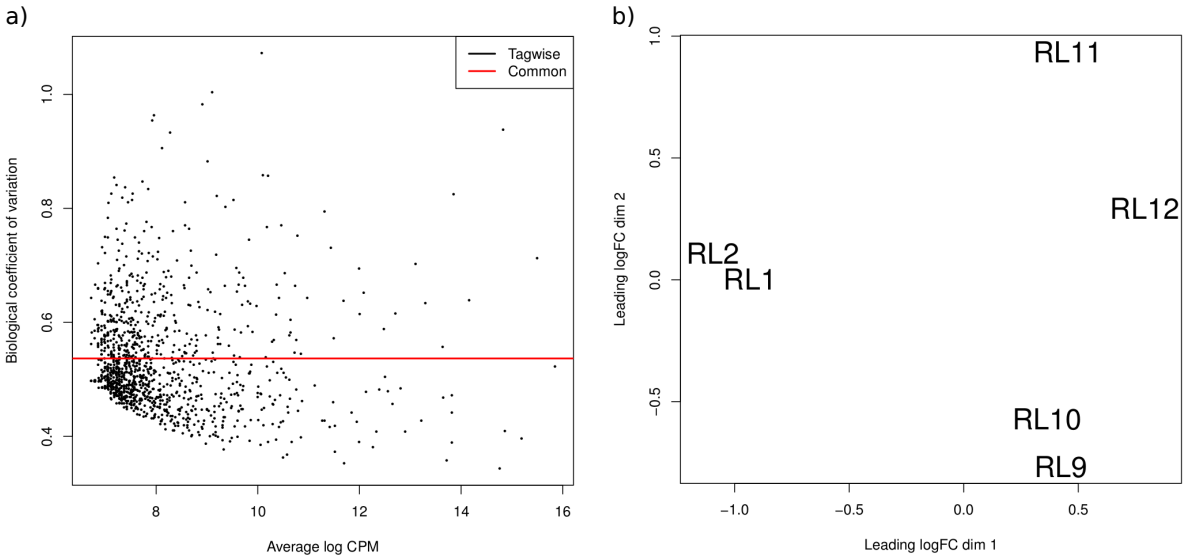


Figure 1. Variation in expression levels among samples (a). The average log transformed number for counts per million reads for each contig is compared to the biological variation for each contig (BCV). (b) Multidimensional scaling of samples according to normalized expression in transcripts. RL1 and RL2 are samples from ambient Helgoland conditions (14.5 °C). RL9, RL10 were samples during temperature increased conditions at Norway (NO.W; 16 °C). RL11 and RL12 originate from ambient temperature conditions at Norway (NO.C; 7.5 °C).

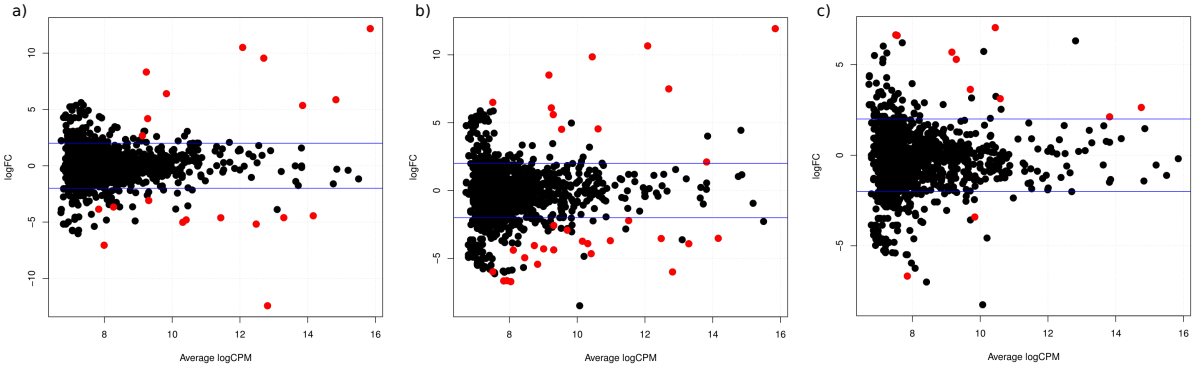


Figure 2. Negative binomial differences in expression level (logFC) between different treatments for *C. pagurus* by the average of counts per millions for each contig. Red dots indicate significant differences corrected according to Benjamini and Hochberg [38] in expression levels normalized with number of reads for each sample. Pairwise comparisons of differential expression levels are compared for a) warm adapted *C. pagurus* from Helgoland and from Norway, b) warm adapted *C. pagurus* from Helgoland and from Norway cold ambient temperature and c) ambient cold *C. pagurus* from Norway and warm treated animals of the same population.

Table 2. Comparison of transcript expression. Differently expressed transcripts between treatments are marked with $-, +$ dependent on the averaged difference in expression. An absolute fold change > 1 is marked with $--$ *or* $++$. Exact fold change calculated for each comparison can be found in the supplementary material. Annotation was made according to blastx [52]. Asterisks indicate statistical significance (p) between treatments with $* = 0.05, ** = 0.01, *** < 0.001$ given by exact negative binomial test [37]. False discovery rates to account for multiple comparisons is given by bolt values for each comparison with $p < 0.05$ corrected according to Benjamini and Hochberg [38]

ID	Annotation	Functional Group	HE.A − > NO.W	HE.A − > NO.A	NO.W − > NO.A
PKc39	XP001662579.1	sugar transporter	++	++**	++
PKc1815	EFX83258.1	sodium-dependent nutrient amino acid transporter	–	++**	++***
PKc414	XP005570452.1	intracellular protein transport protein USO1-like	–**	–***	–
PKc757	ADM26622.1	ferritin	+	++***	++***
PKc4043	ADM26622.1	ferritin	++	++***	++**
PKc2506	XP002737769.1	serine carboxypeptidase 1-like	-	–***	–*
PKc2514	AFJ11395.1	microsomal prostaglandin E synthase	–	–**	–*
PKc2690	EGI58309.1	Oligosaccharyltransferase complex subunit OSTC	–**	–***	–
PKc2863	ABK29471.1	CHK1 checkpoint-like protein	++***	++**	–
PKc2877	XP001641190.1	predicted protein (repeats)	+	++***	++***
PKc4047	NP872442.1	cytochrome c oxidase subunit I	++**	++***	++*
PKc4470	ADZ96217.1	JHE-like carboxylesterase 1	-	–***	–*
PKc4597	AFJ59945.1	C-type lectin 1	–	++*	++***
PKc159	ACO36035.1	vitellogenin	++***	++***	-
PKc2244	AEI59132.1	vitelline percursor	++***	++***	+
PKc7528	AAD09762.1	cryptocyanin	–***	–***	+
PKc2499	ABB59714.1	cryptocyanin 2	–***	–***	++**
PKc4083	XP005499390.1	gamma-butyrobetaine dioxygenase	0	++***	++***
PKc2878	XP005168960.1	gamma-butyrobetaine dioxygenase isoform X1	+	++***	++***
PKc3690	XP002599964.1	gamma-butyrobetaine dioxygenase, partial	++	++***	++***
PKc4038	AAW57893.1	hemocyanin subunit 5	–**	++	++***
PKc77	EFX68242.1	heme-binding protein 2	++***	++**	-
PKc537	AAW57889.1	hemocyanin subunit 1	–**	++*	++***

Marker development Based on 454 sequencing 6,668 simple sequence repeats were detected accounting for 4.28 % of the total transcriptome (Table 5). Small RNAs and retroelements were rarely observed in the transcriptome ($N = 43$, $N = 53$), but accounted for an absolute length of > 18.000 bp. 6,925 SSRs could be detected by using MISA [40], searching for di-, tri-, tetra-, penta-, hexa-, hepta- and octa- repeat motifs with minimum length of 5 repetitions. The majority consisted of di-nucleotide (4,960) and tri-nucleotide motifs (1,627) (Fig. 3). Longer motifs were rarer and motifs bigger than septa- repeat motifs with at least 5 repetitions were not present in contigs. By mapping the single RAW reads against the total assembly a further estimate on polymorphism in repetitive elements could be made. After checking for possible primer sites 37 primer pairs were selected for validation of polymorphisms in 454 assembly. Finally, 5 primer pairs were developed for fragment length comparison in different populations of *C. pagurus* (Table 3) using a capillary sequencer (Beckman CEQ 2000).

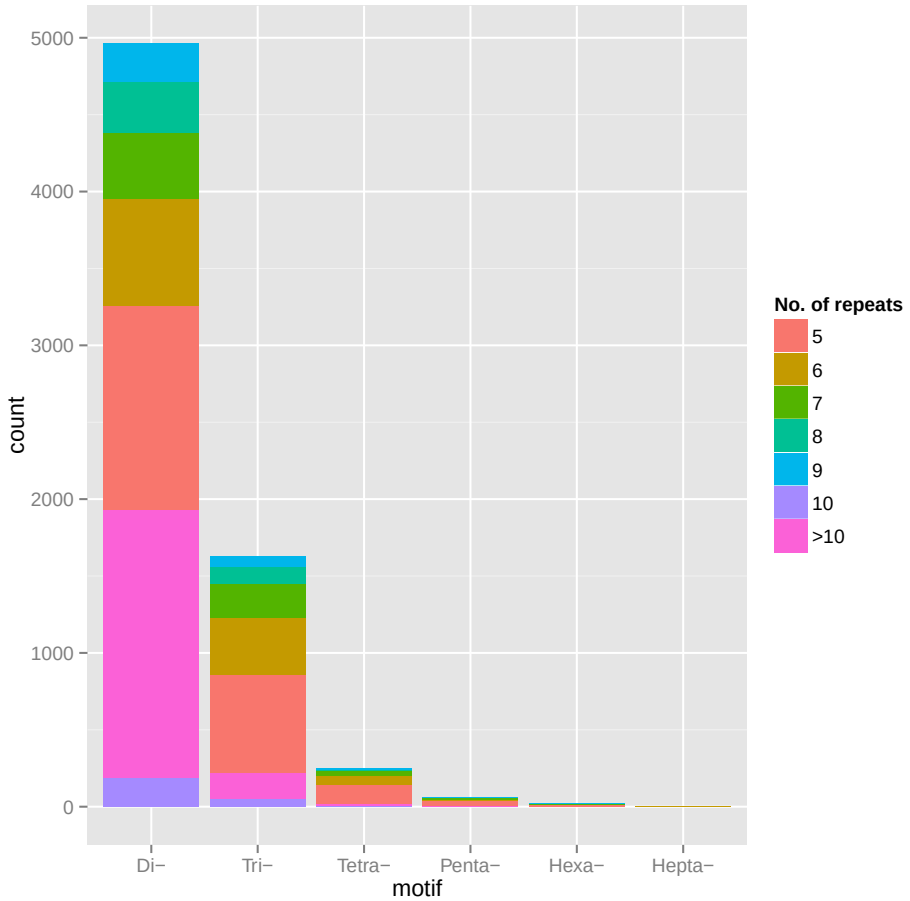


Figure 3. Distribution of SSR motifs for the total assembly calculated by MISA [40].

Table 3. Variable sites in expressed sequence tag sequences in the brown crab *C. pagurus*. For each locus based on number of individuals N the number of alleles (N_a), expected population heterozygosity (H_e) and observed population heterozygosity (H_o) and F_{IS} were calculated [44]. Annotation of contig is given based on best hit of blastx annotation [52].

ID	motif	T_a	Primer pair	N	N_a	H_e	H_o	p	Annotation/Reference
PKc2794	$(TGACC)_8$	60 °C	accagacgatttagcagcaa tcacgatcaaaaggattcagg	44	6	0.485	0.477	0.017	protein mitochondrial-like
PKc2798	$(GT)_{10}$	60 °C	caactcctggaggctttgtc accacatgtgaaccttgacg	50	8	0.748	0.780	-0.043	redox-regulatory protein fam213a
PKc4524	$(AC)_8$ $(ACAT)_7$	60 °C	taagcggcaggtagacaggt tttgcagttttaccgtacgtatt	43	4	0.323	0.349	-0.081	not annotateable
PKc5440	$(CCAGC)_7$	60 °C	gcattgttccagcctctagc accctaaaaggaaggcctga	49	3	0.302	0.347	-0.149***	not annotateable
PKc3295	$INDEL$	60 °C	gatagggacagggagcgagt atctcacccgaacgaggatt	50	2	0.213	0.200	0.063	not annotateable

Genetic structure Overall, genetic differentiation among European populations of *C. pagurus* was low D_{EST} values ranged from -0.005 to 0.035; concluding that a high connectivity exists among populations. All loci were tested for linkage disequilibrium using FSTAT showing no significant linkage between loci either within populations nor overall populations [44]. Neutrality tests did not show a significant outlier on either neutral nor EST-SSRs using the geographic demes as reference [46]. For further population genetic analysis we used the full set of loci ($n = 12$). On average, estimated difference between populations based on D_{EST} and on G_{ST} were low between populations (Table 4). Significant population structuring was found between population from Spain and Denmark in comparison to most other populations with only two exceptions. The population at Denmark was not significantly different to both other populations sampled in the North Sea at Helgoland and at the "Tiefe Rinne" and the population at France was not different to neither Denmark nor Spain. Between northern and southern population from Norway a difference in genetic composition was determined (Table 4). We used a second method to evaluate genetic differences among populations, the fisher exact test for differences in allele composition based on genepop [49,50]. Similarities in allele frequencies were found between populations at Great Britain, Spain and Norway and in a second cluster among populations around the North Sea (Fig. 4). However, this method is based on allele frequencies, which may be problematic if populations are not in Hardy-Weinberg equilibrium. Population significantly deviating from Hardy-Weinberg equilibrium were Northern Norway, France and Denmark 1. Effective population size based on single populations were estimated as infinite effective size with infinite upper confidence level, due to too small sample sizes in relation to N_e. Only the population at Denmark did show a smaller effective population size of $N_e = 2542; CI : 139 - Infinite$. The Atlantic cluster had a smaller N_e ($N_e = 1174; CI : 175 - Infinite$) than the North Sea cluster ($N_e = -1756; CI : 697 - Infinite$).

Table 4. Pairwise genetic differences between populations. Average values of D_{EST} and G_{ST} calculated from the full set of loci for all pairwise comparisons. The lower triangle shows D_{EST} values the upper triangle G_{ST}. Statistical significance after 1000 bootstrap iterations is indicated by $*$ for $p < 0.05$.

	DK	ESP	GB	HE	NO	BE	RO	TR
DK	0	0.004	0.002	0.001	0.003*	0.001	0.003	0.002
ESP	0.023*	0	0.005	0.009	0.006*	0.008	0.007	0.008*
GB	0.016*	0.024*	0	-0.000	0.001	0.000	0.002	0.002
HE	0.010	0.035*	-0.005	0	0.002	0.001	0.001	-0.000
NO	0.008	0.019*	0.002	0.002	0	0.002	0.003	0.003*
BE	0.009*	0.027*	0.003	0.007	0.008*	0	0.004*	0.004*
RO	0.011	0.017	-0.004	-0.000	0.001	0.002	0	0.001
TR	0.009	0.027*	0.004	0.003	0.010*	0.002	-0.000	0

Phenotype genotype correlation We analyzed shell disease in brown crab populations as a potential indication for maladaption due to temperature stress. Prevalence of shell disease was analyzed for Northern Norway ($N = 65$) and populations at the German North Sea (Helgoland, "Tiefe Rinne" and some random samples in the southern North Sea area; $N = 157$). Average severity of shell disease was similar in all populations. In Norway 65 % and at the North Sea 71 % of individuals were infected with shell disease. Of the 27 equally sized carapace areas symptoms of the disease

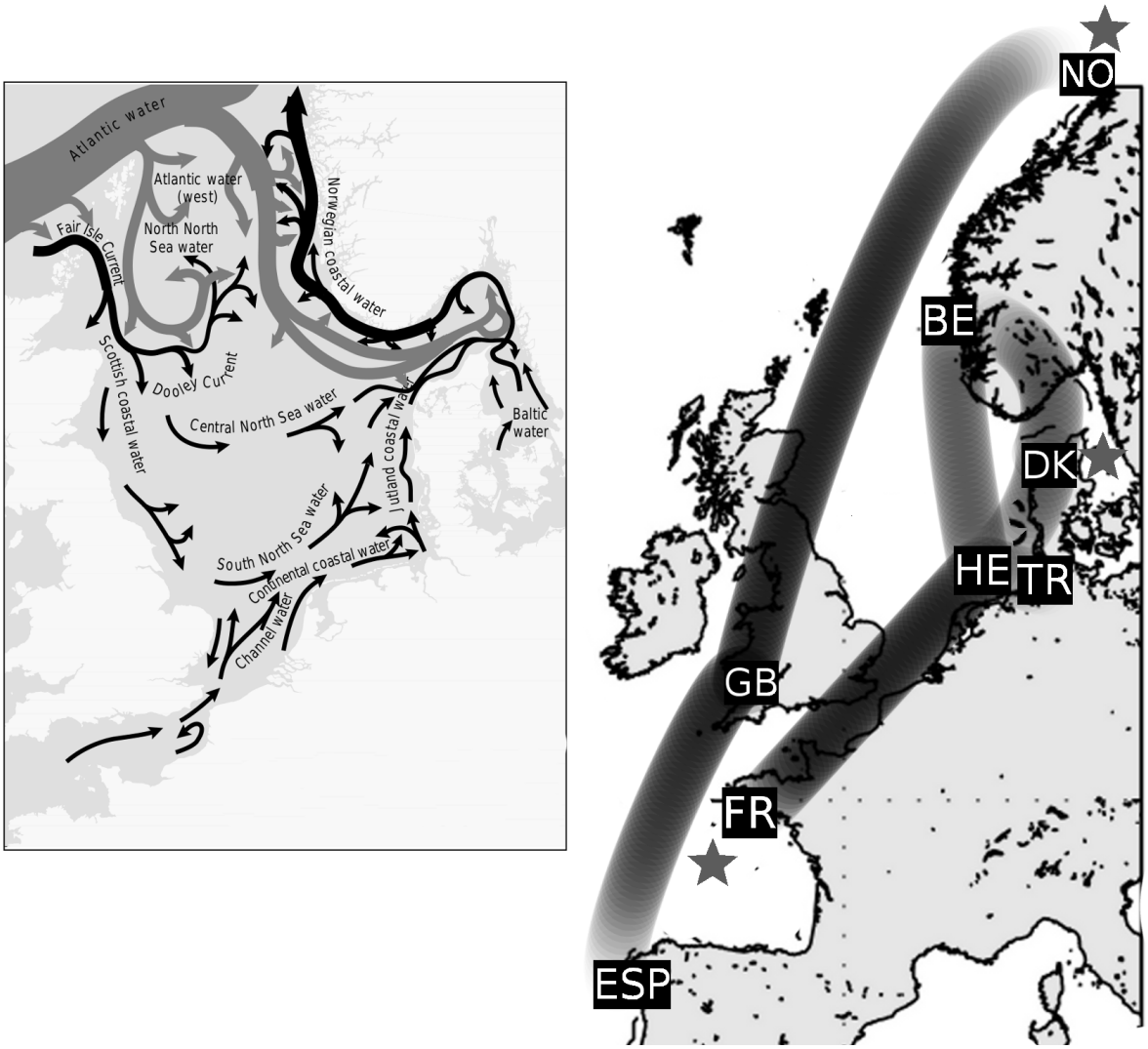


Figure 4. Geographic distribution of sampled populations. Small map showing main currents in the central area around the North Sea adapted from [53]. Brighter currents are mainly influenced by the main Atlantic current of the Gulf Stream. Each population is abbreviated with NO = Norway (Bodø), BE = Norway (Bergen), DK = Denmark, HE = Germany (Helgoland), TR = Germany ("Tiefe Rinne"), FR = France (Roscoff), GB = Great Britain (Cornwall) and ESP = Spain (Vigo). Populations significantly deviating from Hardy Weinberg Equilibrium ($p < 0.05$) are highlighted with a grey star. Populations showing no significant differences in allele distribution calculated by Fisher's exact test ($p < 0.05$) are connected with gray shades [49, 50].

were found, on average, in three carapax areas in Norwegian population, and four areas in the North Sea population. Shell disease was found also in the population at Great Britain, France and Spain. We found genetic differences between healthy and diseased animals. Shell diseased individuals were less heterozygotic than healthy animals. Deviation from Hardy-Weinberg equilibrium indicated by F_{IS} values in total and infected subpopulations suggested different population histories of infected compared to healthy individuals. For further genetic analysis Norway and North Sea populations were split in individuals affected by shell disease and those showing no signs of the disease. Both differed in F_{IS} values; healthy animals had an $F_{IS} = 0.016$ at Norway and $F_{IS} = 0.032$ at Helgoland; diseased animals had an $F_{IS} = 0.153$ at Norway and $F_{IS} = 0.043$ at Helgoland. In the diseased Norwegian population $F_{IS} = 0.153$ was significantly higher than expected ($p < 0.01$). 6 out of the 12 loci showed a strong deficiency of heterozygotes ($p < 0.05$). Differences between shell diseased and not shell diseased individuals was significant in Norway ($D_{EST} = 0.03, p < 0.05$), but not for the North Sea ($D_{EST} = 0.006, p < 0.17$).

Discussion

The study shows that high phenotype plasticity towards a range of bottom water temperature might help the brown crab *Cancer pagurus* to persist even under unfavorable conditions. Norwegian animals exposed to experimentally induced stressful temperature conditions were able to alter gene expression to maintain homeostasis and did not show altered shell disease severity. Such ability to acclimate to temperature differences might explain the wide geographic distribution of the species. We found weak evidence for population structure. Differences in allele composition between two groups of populations were found that were probably caused by main current regimes (Fig. 4), one connected from Spain to North Great Britain and North Norway and a second one from France to the North Sea and South Norway. Differentiation by more robust D_{EST} and on G_{ST} analysis was present only for single populations. We could not find evidence for genetic adaptation in local populations comparing neutral versus our transcriptome derived EST linked markers. Based on population genetic analysis, high connectivity caused by larval dispersal, adult migration and very high effective population sizes seem to prevent obvious local adaptation. However, this has to be interpreted with caution. Evidence for differently expressed genes of the hemocyanines indicate that they may be differently selected in different populations; however this needs to be studied more thoroughly. Crustacean shell disease as a potential indicator of maladaptation is prevalent in *C. pagurus* populations across the whole distributional range. When comparing populations in Germany (North Sea) and Norway (Lofoten) the similar prevalence (diseased animals: 69 % Norway; 71 % Germany) and severity of shell disease could be observed in both populations. We could not observe that shell disease is selective i.e. lethal under given environmental conditions but shell disease may lead to a reduced fitness on the long run [28]. Furthermore, we found evidence that shell diseased and healthy animals in local populations in Norway may originate from different locations. While healthy animals in Norway were genetically different from populations in the south, diseased animals shared alleles with other populations indicating connectivity. Whether diseased animals from more southern populations migrated northwards perhaps to escape unsuitable conditions needs to be proven further.

In the American lobster [54] high susceptibility to shell disease and possible mortality were discussed to be dependent on environmental effects influencing homeostasis. Individuals with shell disease have more internal infections in *C. pagurus* [55]. The high abundance of shell disease indi-

cates that *C. pagurus* has to face severe bacterial challenge. This indicates that under certain local conditions similar consequences as for the shell disease outburst in *Homarus americanus* in New England [27] may arise in *C. pagurus*. In this study we provide the first temperature dependent and population specific transcriptome information for *C. pagurus* originating from populations at different environmental conditions. Weiss et al. [56] found out that *C. pagurus* larvae successfully mature in the North Sea only when kept in a very narrow temperature range of 14 °C ±3 °C. The tolerated temperature range of *C. pagurus* was determined in populations from Norway [57]. Temperatures above 20 °C are not tolerated [58, 59] and below 5 °C individuals become inactive and stop feeding [57]. We increased the temperature experimentally in *C. pagurus* from Norway to 16 °C. Such temperatures are not experienced in the wild in this geographic area during the year. Thus we can compare abnormally challenged individuals from Norway with individuals from Germany at similar ambient temperature in order to see if individuals from Norway can acclimate to unusual conditions within the non lethal range of temperature conditions [57–59]. Quantitative analysis of expressed genes using next generation sequencing showed strong similarities in expression profiles among samples of different temperatures and some genes significantly differently expressed under different temperatures. The expression of important genes for immune functions such as ferritin [60] or C-type lectin [61] changed with temperature: we observed a lower expression during warm temperature treatment in Norway. Decreased immunity during warmer periods may be one important factor influencing the future distributional range of *C. pagurus* since local populations are more susceptible for shell disease and following internal bacterial infections. Especially, hemocyanines were discussed to be one important factor in adaptation to shell disease infections [62, 63]. We found hemocyanin 5 as well as 1 significantly decreased by warm water treatment in Norway. Other transcripts such as the serine carboxypeptidase gene was differently expressed when exposed to different temperature conditions. It is known to be expressed as an immune response to White Spot Syndrome Virus (WSSV) infection in other crustaceans [23]. Suppression of immune activity due to climate change was described for *Nephrops norvegicus* inhabiting similar water depths and geographic areas as *C. pagurus* [64] as well as for other crustaceans such as penaeid shrimp [65]. However, apart from direct altered gene expression in immune related transcripts, immune functions may be indirectly affected by stress hormones which adjust several other physiological conditions to maintain optimal immunity of the animal [66]. We found nutrient uptake proteins decreased during warm treatments such as sodium dependent nutrient amino acid transporter and gamma-butyrobetaine. Especially, gamma-butyrobetaine dioxygenase is essential step to maintain the energy budget (fatty acid transport [67]). It was also found increased after a short heat shock and subsequent recovery in a cold adapted fish [68]. Other increased transcripts were mostly connected to reproduction such as vitellogenin, microsomal postaglandin (ovary development [69]) or JHE-like carboxylesterase (regulation of reproduction, metamorphosis [70]). This shows that higher temperatures can act on reproduction [71] and can severely influence the larval stage. Especially, skewed time of spawning may dramatically influence the probability of survival of pelagic migratory larvae [72]. Due to the high variability of larval duration dependent on temperature [73] this may at least lead to altered dispersal routes and duration in *C. pagurus*. Cryptocyanines involved in molting [74] were rarely affected by temperature treatment but significantly different between populations indicating differences in molting time among populations during sampling. Regular heat shock proteins were not differently expressed although we challenged animals from Northern Norway with temperatures which were higher than what they experience as bottom water temperature throughout the year. This shows that homeostatic effects on the

long term might play a role; however, individuals are not strictly adapted to local conditions and may buffer high temperature conditions by phenotype plasticity indicated by altered gene expression of several important transcripts. These results imply that despite the huge distance between northern Norway and Helgoland in the North Sea populations are genetically similar. This pattern was revealed by population genetic analysis. Only weak evidence for differences between Atlantic offshore populations and populations around the North Sea in allele composition could be determined. However, lack of Hardy-Weinberg equilibrium in populations at northern Norway, France and Denmark may interfere with comparing allele compositions at these sites. The more robust D_{EST} and G_{ST} showed only weak population structure with the highest difference in Spain to the more northern populations. The connectivity among populations and thus the migratory potential of *C. pagurus* seems to be very high. This was also described for populations at the baltic area [7]. *C. pagurus* populations maintained a huge effective population size and a strong connectivity over long geographic ranges for example between northern Norway and more southern populations. This will help *C. pagurus* to shift over these wide ranges and adapt via migration into more favorable habitats rather than to adapt locally. The distributional ability is given by long duration of pelagic larval stages (up to 140 days; [73]). Furthermore, there is evidence that adults migrate over long distances of more than 200 km [75]. This might lead to the extinction of *C. pagurus* populations in areas with unfavorable conditions and explaining the nearly disappearance of *C. pagurus* in some areas of Portugal (pers. comm Vitor Almada). It also may be an indication for the increased population sizes in Norway [8] following a northward shift into these habitats.

Within some populations a deviation from Hardy-Weinberg equilibrium was found. This was strongest in the northern Norway population with 11 out of 12 loci showing heterozygote deficencies. Among markers a significant positive F_{IS} value indicates a probable Wahlund effect in this population [76]. This might be explained by an admixture of *C. pagurus* of the local population with immigrating *C. pagurus* by long distant connectivity through main currents from the southern populations.

The high connectivity among populations from different geographic areas, the high effective population sizes and high genetic variation makes animal migration and shifting as the most likely mode of adaptation to environmental conditions. However, selection on different genotypes is not unlikely. Based on our data selection is not restricted to local habitats and populations but to the whole distributional range. The occurrence of shell disease provides further evidence of ongoing selection or at least an omnipresent source of selection under raising temperature conditions. Strong deviations from Hardy-Weinberg equilibrium was found in the pool of infected animals for Norway in contrast to healthy animals. This indicates a Wahlund effect only in the diseased group. This may show that animals originating from different areas may be more likely to develop shell disease under non ambient conditions. This was not the case for animals at Helgoland. Furthermore, allele frequencies between infected and non infected animals were significantly different in Norway but not in Helgoland. This further indicates that this effect of immigration from several other origins may be stronger for northern Norway than for Helgoland. *C. pagurus* seems to be another species following the general pattern of a northward shift triggered by climate change environmental alterations [1, 3, 4].

Maladaptation in single genes and selection by high pathogen pressure as given by high prevalence in shell disease may lead to altered genetic composition in local populations. It has been shown that SSR and SNPs in functional ESTs may be linked to adaptation against bacterial pathogens in shrimp [77]. We could not find genes affected by such events of selection. However, with 37

variable markers at ESTs and over 6,000 possible sites for further variable SSRs our dataset may provide several loci, which may be hitchhiking variants to close selected genes. Furthermore, genes found differently expressed at treatments may serve as interesting candidates for further studies on differences in genetic profiles between populations and should be studied for their genetic variations among populations. Especially, hemocyanines may be one major candidate to be differently expressed between populations and should be studied in *C. pagurus* more thoroughly. Our data may thus provide a good pool to find genes under selection which may be important in future climate change scenarios and may help to understand the effect of genetic variation on the shift and probably disappearance of populations of *C. pagurus* in Europe.

Acknowledgement

We would like to thank Elke Frahmann, Anke Müller for their great technical assistance in the lab. We would like to thank Aaron Jäschke, Sebastian Fuhrmann, Michael Jahnke for assistance during sampling. We would like to thank the European Science Foundation funding parts of the project with a travel grant (Grant EX/3204). We thank Dr. Eva Philipp and Dr. Philipp Rosenstiehl for the RNA-seq at IKBM Kiel. We thank as well Prof. Galice Hoarau, Dr. Alexander Jüterbock for the opportunity to work at the field station in Norway and Prof. Heinz-Dieter Franke and Isabel Schmallenbach for the opportunity to work the AWI in Helgoland.

References

1. Philippart CJM, Anadon R, Danovaro R, Dippner JW, Drinkwater KF, Hawkins SJ, Oguz T, O'Sullivan G, Reid PC (2011) Impacts of climate change on european marine ecosystems: Observations, expectations and indicators. Journal of Experimental Marine Biology and Ecology 400: 52–69.
2. Valle M, Chust G, del Campo A, Wisz MS, Olsen SM, Garmendia JM, Borja A (2014) Projecting future distribution of the seagrass *Zostera noltii* under global warming and sea level rise. Biological Conservation 170: 74-85.
3. Helmuth B, Mieszkowska N, Moore P, Hawkins SJ (2006) Living on the edge of two changing worlds: Forecasting the responses of rocky intertidal ecosystems to climate change. Annual Review of Ecology, Evolution, and Systematics 37: 373-404.
4. Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421: 37-42.
5. Saupe EE, Hendricks JR, Portell RW, Dowsett HJ, Haywood A, Hunter SJ, Lieberman BS (2014) Macroevolutionary consequences of profound climate change on niche evolution in marine molluscs over the past three million years. Proceedings of the Royal Society B 281.
6. Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. Evolutionary Applications 7: 1-14.
7. Ungfors A, McKeown NJ, Shaw PW, Andre C (2009) Lack of spatial genetic variation in the edible crab (*Cancer pagurus*) in the Kattegat-Skagerrak area. ICES Journal of Marine Science 66: 462–469.
8. Woll AK, van der Meeren GI, Fossen I (2006) Spatial variation in abundance and catch composition of *Cancer pagurus* in norwegian waters: biological reasoning and implications for assessment. Ices Journal of Marine Science 63: 421–433.
9. Tully O, Robinson M, Cosgrove R, OKeeffe E, Doyle O, Lehane B (2006) The brown crab (*Cancer pagurus* L.) fishery: Analysis of the resource in 2004-2005. Fisheries Resource Series 4: 48.
10. Ramsay K, Kaiser MJ, Moore PG, Hughes RN (1997) Consumption of fisheries discards by benthic scavengers: utilization of energy subsidies in different marine habitats. Journal of Animal Ecology 66: 884–896.

11. Hall SJ, Robertson MR, Basford DJ, Fryer R (1993) Pit-digging by the crab *Cancer pagurus*: A test for long-term, large-scale effects on infaunal community structure. Journal of Animal Ecology 62: 59-66.
12. Hall SJ, Basford DJ, Robertson MR, Raffaelli DG, Tuck I (1991) Patterns of recolonisation and the importance of pit-digging by the crab *Cancer pagurus* in a subtidal sand habitat. Marine Ecology Progress Series 72: 93-102.
13. Silva ACF, Hawkins SJ, Boaventura DM, Thompson RC (2008) Predation by small mobile aquatic predators regulates populations of the intertidal limpet *Patella vulgata* (L.). Journal of Experimental Marine Biology and Ecology 367: 259–265.
14. Sivertsen K (2006) Overgrazing of kelp beds along the coast of Norway. Journal of Applied Phycology 18: 599–610.
15. Zhao L, Jones WA (2012) Expression of heat shock protein gene in insect stress responses. ISJ 9: 93-101.
16. Wagner EJ, Bosakowski T, Intelmann S (1997) Combined effects of temperature and high ph on mortality and the stress responsne of rainbow trout after stocking. Transactions of the American Fisheries Society 126: 985-998.
17. Barton BA (2002) Stress in fishes: A diversity of responses with particular reference to changes in circulating corticosteroids. Integrative and Comparative Biology 42: 517-525.
18. Frederich M, O'Rourke M, Towle D (2006) Differential increase in AMPK and HSP70 mRNA expression with temperature in the rock crab, *Cancer irroratus*. FASEB Journal 20: A827.
19. Chang ES (2005) Stressed-out lobsters: Crustacean hyperglycemic hormone and stress proteins. Integrative and Comparative Biology 45: 43–45.
20. Towle DW, Smith CM (2006) Gene discovery in *Carcinus maenas* and *Homarus americanus* via expressed sequence tags. Integrative and Comparative Biology 46: 912–918.
21. Decker H, Hellmann N, Jaenicke E, Lieb B, Meissner U, Markl J (2007) Minireview: Recent progress in hemocyanin research. Integrative and Comparative Biology 47: 631–644.
22. Garcia-Carreno FL, Cota K, Navarrete Del Toro MA (2008) Phenoloxidase activity of hemocyanin in white-leg shrimp *Penaeus vannamei*: Conversion, characterization of catalytic properties, and role in postmortem melanosis. Journal of Agricultural and Food Chemistry 56: 6454–6459.
23. Tassanakajon A, Somboonwiwat K, Supungul P, Tang S (2013) Discovery of immune molecules and their crucial functions in shrimp immunity. Fish & Shellfish Immunology 34: 954–967.
24. Windisch HS, Kathoever R, Poertner HO, Frickenhaus S, Lucassen M (2011) Thermal acclimation in antarctic fish: transcriptomic profiling of metabolic pathways. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology 301: 1453-1466.
25. Poertner HO, Langenbuch M (2005) Synergistic effects of temperature extremes, hypoxia, and increases in CO2 on marine animals: From Earth history to global change. Journal of Geophysical Research 110: 1–15.
26. Harms L, Frickenhaus S, Schiffer M, Mark FC, Storch D, Held C, Poertner HO, Lucassen M (2014) Gene expression profiling in gills of the great spider crab *Hyas araneus* in response to ocean acidification and warming. BMC Genomics 15: 1–17.
27. Glenn RP, Pugh TL (2006) Epizootic shell disease in American lobster (*Homarus americanus*) in Massachusetts coastal waters: Interactions of temperature, maturity, and intermolt duration. Journal of Crustacean Biology 26: 639–645.
28. Vogan CL, Llewellyn PJ, Rowley AF (1999) Epidemiology and dynamics of shell disease in the edible crab *Cancer pagurus*: a preliminary study of Langland Bay, Swansea, UK. Diseases of Aquatic Organisms 35: 81–87.
29. Powell A, Rowley AF (2005) Unchanged prevalence of shell disease in the edible crab *Cancer pagurus* four years after decommissioning of a sewage outfall at Langland bay, UK. Diseases of Aquatic Organisms 68: 83–87.
30. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Research 14: 1147–1159.
31. Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PGD (2012) Evaluating characteristics of de novo assembly software on 454 transcriptome data: A simulation approach. Plos One 7: e31410.

32. Dlugosch KM, Lai Z, Bonin A, Hierro J, Rieseberg LH (2013) Allele identification for transcriptome-based population genomics in the invasive plant *Centaurea solstitialis*. G3 3: 359-367.
33. Barker MS, Dlugosch KM, Dinh L, Challa RS, Kane1 NC, King MG, Rieseberg LH (2010) Evopipes.net: Bioinformatic tools for ecological and evolutionary genomics. Evolutionary Bioinformatics 6: 143-149.
34. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658–1659.
35. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.
36. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.
37. Robinson MD, Smyth GK (2008) Small-sample estimation of negative binomial dispersion, with applications to sage data. Biostatistics 9: 321–332.
38. 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 57: 289-300.
39. Smit AFA, Hubley R, Green P (1996-2010). Repeatmasker open-3.0.
40. Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 106: 411–422.
41. Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research 40: e115.
42. Koressaar T, Remm M (2007) Enhancements and modifications of primer design program primer3. Bioinformatics 23: 1289-1291.
43. Kraemer P, Gerlach G (2013) Demerelate: Functions to calculate relatedness on diploid genetic data. URL `http://CRAN.R-project.org/package=Demerelate`. R package version 0.8-1.
44. Goudet J (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86: 485–486.
45. McKeown NJ, Shaw PW (2008) Polymorphic nuclear microsatellite loci for studies of brown crab, *Cancer pagurus* L. Molecular Ecology Resources 8: 653—655.
46. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F(st)-outlier method. BMC Bioinformatics 9: 323.
47. Jost L (2008) G(st) and its relatives do not measure differentiation. Molecular Ecology 17: 4015–4026.
48. Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on G(st) and D: forget G(st) but not all of statistics! Molecular Ecology 19: 3845–3852.
49. Raymond M, Rousset F (1995) Genepop (version-1.2) - population-genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248–249.
50. Rousset F (2008) Genepop ' 007: a complete re-implementation of the genepop software for windows and linux. Molecular Ecology Resources 8: 103–106.
51. Vogan CL, Rowley AF (2002) Dynamics of shell disease in the edible crab *Cancer pagurus*: a comparative study between two sites on the gower peninsula, south wales, UK. Diseases of Aquatic Organisms 52: 151–157.
52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.
53. OSPAR Commission (2000) Quality Status Report 2000, Region II – Greater North Sea. OSPAR Commission, London. 136 pp.
54. Tlusty MF, Smolowitz RM, Halvorson HO, DeVito SE (2007) Host susceptibility hypothesis for shell disease in American lobsters. Journal of Aquatic Animal Health 19: 215–225.
55. Vogan CL, Costa-Ramos C, Rowley AF (2001) A histological study of shell disease syndrome in the edible crab *Cancer pagurus*. Diseases of Aquatic Organisms 47: 209–217.

56. Weiss M, Thatje S, Heilmayer O, Anger K, Brey T, Keller M (2009) Influence of temperature on the larval development of the edible crab, *Cancer pagurus*. Journal of the Marine Biological Association of the United Kingdom 89: 753–759.
57. Karlsson K, Christiansen ME (1996) Occurrence and population composition of the edible crab (*Cancer pagurus*) on rocky shores of an islet on the south coast of Norway. Sarsia 81: 307–314.
58. Hallbaeck H (1972) Ekologiska och etologiska studier av krabbtaskan (*Cancer pagurus*) på svenska västkusten. Fauna och Flora 67: 65-66.
59. Hallbaeck H (1993) 'hela havet är fullt av krabbor...'. Yrkesfiskaren (Goeteborg) 17: 8-10.
60. Zhang D, Jiang K, Zhang F, Ma C, Shi Y, Qiao Z, Ma L (2011) Isolation and characterization of a ferritin cDNA from the mud crab *Scylla paramamosain*. Journal of Crustacean Biology 31: 345-351.
61. Geijtenbeek TBH, Gringhuis SI (2009) Signalling through c-type lectin receptors: shaping immune responses. Nature Reviews Immunology 9: 465–479.
62. Terwilliger NB (2007) Hemocyanins and the immune response: defense against the dark arts. Integrative and Comparative Biology 47.
63. Vogan CL, Costa-Ramos C, Rowley AF (2002) Shell disease syndrome in the edible crab, *Cancer pagurus* - isolation, characterization and pathogenicity of chitinolytic bacteria. Microbiology-sgm 148: 743–754.
64. Hernroth B, Skold HN, Wiklander K, Jutfelt F, Baden S (2012) Simulated climate change causes immune suppression and protein damage in the crustacean nephrops norvegicus. Fish & Shellfish Immunology 33: 1095–1101.
65. Moullac GL, Soyez C, Saulnier D, Ansquer D, Avarre JC, Levy P (1998) Effect of hypoxic stress on the immune response and the resistance to vibriosis of the shrimp *Penaeus stylirostris*. Fish & Shellfish Immunology 8: 621–629.
66. Adamo SA (2012) The effects of the stress response on immune function in invertebrates: an evolutionary perspective on an ancient connection. Hormones and Behavior 62: 323–330.
67. Vaz F, van Gool S, Ofman R, Ijlst L, Wanders R (1998) Carnitine biosynthesis: identification of the cdna encoding human gamma-butyrobetaine hydroxylase. Biochemical and Biophysical Research Communications 250: 506-510.
68. Buckley BA, Somero GN (2009) cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish *Trematomus bernacchii* to alter gene expression in response to heat stress. Polar Biology 32: 403-415.
69. Reddy PS, Reddy PR, Nagaraju GPC (2004) The synthesis and effects of prostaglandins on the ovary of the crab *Oziotelphusa senex senex*. General and Comparative Endocrinology 135: 35-41.
70. Lee SO, Jeon JM, Oh CW, Kim YM, Kang CK, Lee DS, Mykles DL, Kim HW (2011) Two juvenile hormone esterase-like carboxylesterase cDNAs from a pandalus shrimp (*Pandalopsis japonica*): Cloning, tissue expression, and effects of eyestalk ablation. Comparative Biochemistry and Physiology B-biochemistry & Molecular Biology 159: 148–156.
71. Nagaraju GP (2010) Reproductive regulators in decapod crustaceans: an overview. Journal of Experimental Biology 214: 3-16.
72. Tlusty MF, Myers A, Metzler A (2008) Short- and long-term dietary effects on disease and mortality in American lobster *Homarus americanus*. Diseases of Aquatic Organisms 78: 249–253.
73. Nichols J, Thompson B, Cryer M (1982) Production, drift and mortality of the planktonic larvae of the edible crab (cancer pagurus) off the north-east coast of england. Netherlands Journal of Sea Research 16: 173-184.
74. Terwilliger NB, Dangott L, Ryan M (1999) Cryptocyanin, a crustacean molting protein: Evolutionary link with arthropod hemocyanins and insect hexamerins. Proceedings of the National Academy of Sciences of the United States of America 96: 2013-2018.
75. Ungfors A, Hallback H, Nilsson PG (2007) Movement of adult edible crab (*Cancer pagurus* L.) at the swedish west coast by mark-recapture and acoustic tracking. Fisheries Research 84: 345–357.
76. Dharmarajan G, Beasley JC, Fike JA, Raizman EA, Wu CC, Pogranichniy RM, Rhodes OEJ (2012) Effects of kin-structure on disease dynamics in raccoons (*Procyon lotor*) inhabiting a fragmented landscape. Basic and Applied Ecology 13: 560–567.
77. Zhao X, Guo L, Zhang Y, Liu Y, Zhang X, Lun J, Chen J, Li Y (2012) SNPs of hemocyanin c-terminal fragment in shrimp *Litopenaeus vannamei*. Febs Letters 586: 403–410.

Supplementary information

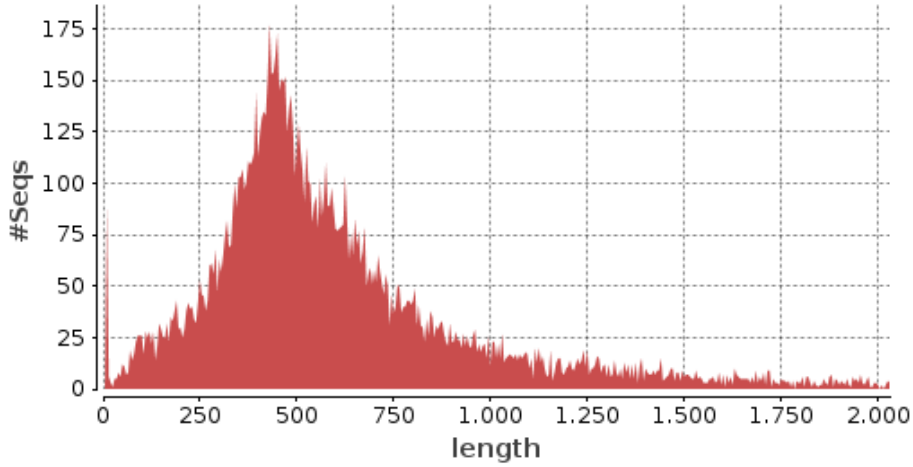


Figure 5. The histogram shows the number of sequences of contigs from the 454 transcriptome of *Cancer pagurus*. The length of the sequence is given in basepairs after quality filtering and contig assembly.

Table 5. Summary of genetic motifs identified by RepeatMasker version open-4.0.1, sensitive mode [39]. Total length of analyzed sequences were 8,849,553 bp, which were assembled into 13,709 contigs with GC level of 41.85 %. All contigs were filtered for Retroelements, LINEs, LTR elements, DNA transposons, Small RNAs, Satellite sequences, Simple Repeat sequences (SSRs) and sequences with low complexity. For each element the total number (total number), the total length of all hits concatenated (total length) and the relative proportion of concatenated sequences to the total length of all contigs (total amount) was calculated.

Type	total number	total length	total amount [%]
Retroelements	53	4,892 bp	0.06 %
LINEs	33	3,126 bp	0.04 %
LTR elements	20	1,766 bp	0.02 %
DNA transposons	1	82 bp	0.00 %
Small RNA	43	13,417 bp	0.15 %
Satellites	3	903 bp	0.01 %
Simple repeats	6,668	378,726 bp	4.28 %
Low complexity	878	63,483 bp	0.72 %

Table 6. Fold change of differently expressed transcripts between treatments. Annotation was made according to blastx [52]. Asterisks indicate significance level of change between treatments with $* = 0.05, ** = 0.01, *** < 0.001$ given by exact negative binomial test [37]. False discovery rates to account for multiple comparisons is given by bolt values for each comparison with $p < 0.05$ corrected according to Benjamini and Hochberg [38]

ID	Annotation	Functional Group	HE.A − > NO.W	HE.A − > NO.A	NO.W − > NO.A
PKc39	XP001662579.1	sugar transporter	2.05	**3.35****	1.31
PKc124	DY656749.1	cDNA *C. maenas*	**-3.60****	-2.15*	1.44
PKc1815	EFX83258.1	sodium-dependent nutrient amino acid transporter	-2.41	4.27**	**6.60*****
PKc414	XP005570452.1	intracellular protein transport protein USO1-like	-2.33**	**-3.67*****	-1.35
PKc415	NA	NA	-2.64*	**-6.64*****	-4.16
PKc757	ADM26622.1	ferritin	0.06	**2.16*****	**2.10*****
PKc4043	ADM26622.1	ferritin	2.04	**4.55*****	2.52**
PKc969	XP001662795.1	phosphatidylethanolamine-binding protein	-2.28**	**-4.35*****	-2.08
PKc2506	XP002737769.1	serine carboxypeptidase 1-like	-0.29	**-2.16*****	-1.88*
PKc2514	AFJ11395.1	microsomal prostaglandin E synthase	-1.37	**-3.74****	-2.37*
PKc2651	NA	NA	-1.28	**-4.33****	-3.07*
PKc2660	NA	NA	-1.42	**-4.00****	-2.59
PKc2690	EGI58309.1	Oligosaccharyltransferase complex subunit OSTC	-3.65**	**-4.98*****	-1.35
PKc2740	NA	NA	-2.53*	**-4.36****	-1.85
PKc2863	ABK29471.1	CHK1 checkpoint-like protein	**5.49*****	**4.09****	-1.40
PKc2877	XP001641190.1	predicted protein (repeats)	0.35	**5.60*****	**5.27*****
PKc4047	NP872442.1	cytochrome c oxidase subunit I	2.68**	**4.56*****	1.90*
PKc4279	NA	NA	**-3.02*****	**-2.53****	0.49
PKc4470	ADZ96217.1	JHE-like carboxylesterase 1	-0.86	**-2.85*****	-1.99*
PKc4597	AFJ59945.1	C-type lectin 1	-1.75	1.86*	**3.60*****
PKc3986	NA	NA	-3.79**	**-6.63*****	-2.97

Table 7. Fold change of differently expressed transcripts between treatments. Annotation was made according to blastx [52]. Asterisks indicate significance level of change between treatments with $* = 0.05, ** = 0.01, *** < 0.001$ given by exact negative binomial test [37]. False discovery rates indicated by bolt values corrected according to Benjamini and Hochberg [38] with $p < 0.05$.

a)

ID	Annotation	Functional Group	HE.A − > NO.W	HE.A − > NO.A	NO.W − > NO.A
PKc159	ACO36035.1	vitellogenin	**12.22*****	**11.95*****	-0.22
PKc2244	AEI59132.1	vitelline percursor	**10.53*****	**10.68*****	0.39
PKc2362	ACO36035.1	vitellogenin	**9.63*****	**7.54*****	-2.04*
PKc2385	AGM75775.1	vitellogenin	4.42**	-2.21	**-6.70*****
PKc4056	ACO36035.1	vitellogenin	**8.35*****	**6.11****	-2.00*
PKc54	ABX89617.1	vitellogenin	**6.46*****	5.03**	-1.39

b)

ID	Annotation	Functional Group	HE.A − > NO.W	HE.A − > NO.A	NO.W − > NO.A
PKc1224	AAD09762.1	cryptocyanin	-2.20	**-6.72****	-4.65
PKc4042	AAD09762.1	cryptocyanin	-3.53**	**-5.36*****	-1.86
PKc3693	AAD09762.1	cryptocyanin	**-4.56****	-2.84*	1.72
PKc4198	AAD09762.1	cryptocyanin	**-7.06*****	-2.98**	3.94
PKc7528	AAD09762.1	cryptocyanin	**-4.77*****	**-4.66*****	0.09
PKc11799	AAD09762.1	cryptocyanin	**-4.99*****	**-3.92*****	1.06
PKc1984	AAD09762.1	cryptocyanin	**-4.40*****	**-3.54****	0.87
PKc7650	AAD09762.1	cryptocyanin	**-5.14*****	**-3.54*****	1.59
PKc846	AAD09762.1	cryptocyanin	**-4.54*****	**-3.90*****	0.63
PKc2499	ABB59714.1	cryptocyanin 2	**-12.43*****	**-6.00*****	6.30**

c)

ID	Annotation	Functional Group	HE.A − > NO.W	HE.A − > NO.A	NO.W − > NO.A
PKc4083	XP005499390.1	gamma-butyrobetaine dioxygenase	0.00	**6.51*****	**6.63*****
PKc2878	XP005168960.1	gamma-butyrobetaine dioxygenase isoform X1	3.08	**9.87*****	**7.01*****
PKc3690	XP002599964.1	gamma-butyrobetaine dioxygenase, partial	3.08	**8.53*****	**5.68*****

Table 8. Percentage similarity of contigs annotated to vitellogenin.

	PKc4056	PKc2362	PKc2385	PK54	PKc159
PKc2362	59 %				
PKc2385	75 %	78 %			
PKc54	69 %	71 %	91 %		
PKc159	35 %	25 %	35 %	40 %	
PKc2244	78 %	72 %	90 %	88 %	41 %

Table 9. Percentage similarity of contigs annotated to cryptocyanin.

	PKc3693	PKc4198	PKc4042	PKc7528	PKc7650	PK1224	PKc846	PKc2499	PKc11799
PKc4198	69 %								
PKc4042	70 %	84 %							
PKc7528	69 %	82 %	81 %						
PKc7650	72 %	78 %	80 %	87 %					
PKc1224	66 %	77 %	83 %	77 %	75 %				
PKc846	54 %	61 %	68 %	58 %	61 %	67 %			
PKc2499	18 %	11 %	13 %	13 %	15 %	17 %	17 %		
PKc11799	67 %	76 %	75 %	68 %	73 %	67 %	60 %	17 %	
PKc1984	23 %	16 %	18 %	19 %	21 %	22 %	23 %	72 %	25 %

Table 10. Perceptual similarity of contigs annotated to gamma-butyrobetaine dioxygenase.

	PKc3690	PKc4083
PKc4083	29 %	
PKc2878	35 %	45 %

Table 11. Perceptual similarity of contigs annotated to hemocyanine.

	PKc4038	PKc77
PKc77	38 %	
PKc537	46 %	31 %

Table 12. Fold change of differently expressed transcripts between treatments. Annotation was made according to blastx [52]. Asterisks indicate significance level of change between treatments with $* = 0.05, ** = 0.01, *** < 0.001$ given by exact negative binomial test [37]. False discovery rates indicated by bolt values corrected according to Benjamini and Hochberg [38] with $p < 0.05$.

ID	Annotation	Functional Group	HE.A $->$ NO.W	HE.A $->$ NO.A	NO.W $->$ NO.A
PKc4038	AAW57893.1	hemocyanin subunit 5	-2.08**	1.02	**3.10*****
PKc77	EFX68242.1	heme-binding protein 2	**4.24*****	3.48**	-0.74
PKc537	AAW57889.1	hemocyanin subunit 1	-1.54**	1.07*	**2.61*****

Synthesis

Shell disease - local problem or generally affecting crustacean biodiversity? *Cancer pagurus* was found heavily affected by shell disease in the North Sea and northern Norway. Bacteria were found with strong similarity to potential causative agents of epizootic shell disease. *C. pagurus* is likely to experience shell disease in wild populations similar to the severity of epizootic shell disease in the American lobster *Homarus americanus*. As a consequence *C. pagurus* would disappear from local habitats, due to omnipresent pathogen and chitinolytic bacteria, which were present in biofilms. To date phenotype plasticity or migration into more favorable habitats helps *C. pagurus* to persist. However, changed environmental conditions, altered pressure by bacterial pathogens or intra-specific competition may have already altered population structures of *C. pagurus* by changed fitness or migration rates leading to low genetic diversification between populations. This high connectivity may connect locally adapted populations of brown crabs which may consequently disappear.

For Europe epizootic shell disease is not documented. However, other forms of shell disease are often described in European populations of crustaceans (135; 10; 11; 12; 13; 136; 137). In this study one of the biggest benthic crustaceans of the Atlantic, the brown crab *C. pagurus* was found heavily affected by shell disease in two distant populations in the German North Sea and at the coast of Norway. Furthermore, strongly diseased animals were found in the other edge of the species distribution at the border of Portugal and Spain and signs of shell disease were found in several other populations among Europe. Prevalence was similarly high in warm water populations at the German North Sea as well as cold water populations in the north of Norway. In Germany > 80 % of individuals show signs of shell disease. Earlier publications (11; 136; 48) calculated a similar prevalence for adults at local populations at the British Isles. The strong differences in environmental conditions between populations and the similar degree of shell disease among populations may indicate, that local environmental conditions are not the most important factor for shell disease prevalence in *C. pagurus*. Accordingly, the model described by Castro et al. (1) to understand the epizootic form of shell disease shows the importance of the environment, the state of the host and the pathogens in shell disease emergence. This model may be applicable to other forms of shell disease as well. But what maintains a similar and high prevalence of shell disease over vast distances and different environmental conditions? To contribute answering such questions in my thesis I present an overview of the variability of the genetic diversity of hosts and the variability in epibiotic biofilm communities.

Epibiotic bacterial biofilms of *C. pagurus* The bacterial community of the biofilm found on carapaces of *C. pagurus* are very different to other planktonic crustaceans found in the same area (138). Stronger similarities were found with communities described for *H. americanus* (5) inhabiting similar habitats at North America as *C. pagurus*. This shows that environmental conditions in local habitats are more important for bacterial community compositions than global differences in presence of bacterial taxa. This follows the general rule ”everything is everywhere, but, the environment selects” (139), which in this case can probably hold true for shell disease associated bacteria in the biofilms of *C. pagurus* as well. There was a remarkable

similarity of samples taken from one individual in contrast to others of the same area. This shows that even small scale environmental conditions in concert with host factors will trigger the epibiotic bacterial community on *C. pagurus*. These two factors may be probably much more important for the occurrence of shell disease than an infection by potential harmful chitinolytic bacteria or other pathogens. This is most likely since those bacterial taxa were omnipresent in biofilms.

Different forms of shell disease and potential pathogens were described for *H. americanus*, *C. pagurus* and other crustaceans (6; 42; 10; 140; 46; 141; 40). I could determine sequences from bacteria very similar to those, which had been recently discussed causing different forms of shell disease in *H. americanus* of the genus *Thalassobius* (43), *Tenacibaculum* (42) and very few *Vibrio* and *Pseudoalteromonas* (40). *Kiloniella* strains found in *C. pagurus* were most similar to *Kiloniella laminariae* LD81, which had been described to be very similar to a causative shell disease strain (6). Several species of *Aquimarina* spp. were found, some of those were very similar to *Aquimarina "homaria"* suspected to cause epizootic shell disease (6). Following these results, different pathogens are constantly present over vast geographic distances and are even found in North America. Under favorable conditions shell disease may emerge due to one of these pathogens.

Changing temperature conditions may affect prevalence of shell disease (3; 142), most likely accompanied by a change to increased shell disease pathogens. In the temperature change experiments neither of both could be clearly observed. This is interesting, since temperature changes of 10°C in Norway lead to conditions, which cannot be experienced in the wild in this area. Therefore this shows the stability of the natural epibiotic biofilm for changes in temperature over 5 days and the plasticity of the host to temperature changes. However, there were alterations on the biofilm community level. A decreased diversity under cold conditions for each community was found indicating a beginning shift in bacterial communities. Increased abundance of taxa at warm conditions were found for *Nitrospira* spp. and *Nitrosomonas* spp.. Both taxa play an important role in the nitrogen cycle. Nitrate is transfered from ammonia, which originates from decomposition of animal or plant material (143; 144; 145) and due to physiological reasons by the host crab itself (146). This gives an indication that increased temperature conditions may favor anoxic conditions in the biofilm of *C. pagurus*. Furthermore, an increase in taxa such as Chromatiales favoring anoxic conditions were found. Anoxic conditions may lead to shell disease as it was discussed for *H. americanus* (147) and may be a reason for increased shell disease of species inhabiting waste disposal sites (50; 49). Under future climate change scenarios oxygen levels may decrease due to warmer water temperature in marine benthic habitats (125). Anaerobic areas under high nutrient and high temperature conditions are very likely to emerge under dense biofilms (148). In fact shell disease is discussed to often occur around seta on the carapace (149; 150), which lay deep in valleys of the outer carapace in the case of *C. pagurus*. We found strictly aerobic strains in one community with strictly anoxic bacteria on *C. pagurus*. Such small scale micro-habitat conditions on carapaces may favor communities with many different forms of pathogens. Therefore the condition of the host is one central factor in order to cope with a constant threat of pathogens in the context of future climate change environmental alterations increasing the likelihood of different pathogens to accumulate. Environmental conditions or biological properties of the host for *C. pagurus* at the moment do not lead to heavy outbursts of shell disease such as in the American lobster populations of New England (1). To date the situation of shell disease in *C. pagurus* probably does not lead to fatal losses in populations. However, the presence of similar pathogens and the occurrence of symptoms and bacterial taxa indicative for epizootic shell disease in *C. pagurus* shows that epizootic shell disease may be one candidate to trigger future biodiversity in benthic crustaceans in Europe.

Effects of shell disease on inter-specific interactions Strong differences in shell disease prevalence were observed between species in local habitats. While *Hemigrapsus penicillatus* the invading Asian brush crab did not show any symptoms of shell disease in the North Sea, the native green shore crab *Carcinus maenas* and the brown crab *C. pagurus* showed medium to heavy prevalence of shell disease. For populations I found the general pattern, that in both native species smaller individuals did show less shell disease, while bigger animals were more heavily affected. Shell disease occurrence in different species and different size of animals are probably most affected by differences of molting frequency. While smaller crustaceans molt more frequently, they may faster dispose unhealthy biofilms than bigger individuals. Depending on the state of the host, physiological factors may play a role as well. It was found that even after molting shell disease may reappear (151; 152). Several subsequent infections of gills and the internal tissues were apparent, which most likely would not disappear after molting. This supports the idea that shell disease is a symptom of a broad disease affecting other tissues in an individual rather than a single disease (153; 34). Probably, molting will have a bigger effect on shell disease in smaller size classes compared to other factors. The frequency of renewal may avoid accumulation of chitinolytic or other harmful bacteria in biofilms. At a specific size molting becomes rare in *C. pagurus* (less than once per year (154; 155)). During a time-frame of more than one year epibiotic biofilms will have enough time for a mature structure and accumulation of several potential pathogens. This may result in a higher importance of immune reactions to invading pathogens. In *C. pagurus* most of the individuals with shell disease have also infected hemolymphs (156). This shows the relationship of the internal immune reactions with invading bacteria - especially in bigger size classes of brown crabs. If smaller animals have a better chance to avoid shell disease inflammations, this will lead to a positive selection for smaller size classes or even smaller species. This is additionally supported by recent findings, where molting frequency was modeled as one important factor for shell disease prevalence (157). Smaller animals or species may be more resistant towards environmental changes (158). Thus stronger selection pressure due to climate change is expected in those cases to be more pronounced for larger animals in local populations. For the future fate of large crustaceans three scenarios are possible: local extinction, rapid adaptation to new local conditions or a shift into new habitats.

Population genetics as a tool to study adaptability in North Sea brachyurans Population genetics theory delivers several tools to understand the connectivity of populations of a species and may help to draw inferences on other population history factors or even adaptation. F-statistics and especially F_{ST} (55) and G_{ST} (57) here have traditionally been used in order to understand the structure of populations and to study regions of the genome, which are putatively under selection. Recently debates arose on the use of alternative estimators to understand neutral genetic variation in terms of population structure on a geographic scale (75; 80; 74; 159; 79). The D_{EST} estimate was developed in 2008 (74) to give a better estimate on the genetic difference between populations. We studied the behavior of D_{EST} in comparison to G_{ST} in order to understand their application for neutral genetic variation. The biggest problem of F-statistics is its dependence on absolute diversity. A highly diverse locus may never reach a difference between populations of one, even if the populations are completely different. This bears some problems if markers are truly neutral and should serve to interpret the absolute genetic difference between populations. We found that D_{EST} scales better with diversity in markers and is thus not limited in its absolute maximum in any given situation. For both estimators, D_{EST} and G_{ST}, statistical evaluations give the same results. But the disadvantage of F-statistics, being dependent on total gene diversity, is an advantage in certain situations. Even in early publications (55; 160; 161) on F-statistics it was noted that F-statistics does measure fixation rather than differentiation. This means that for neutral markers G_{ST} may be informative to describe how fixed the structure in populations may be. In our study both

statistics were compared using a finite neutral genetic model; any allele appearing in only one deme is determined as a new mutation, while changes in the frequency of already existing alleles is interpreted as migration. The most extreme case of such a model would be, that two populations are isolated from each other, sharing no allele at all. In both populations random genetic drift will then lead to the disappearance of many (most likely rare) alleles and may thus produce high G_{ST} values, due to lowered overall genetic diversity over time. Those populations are called fixed and will become even more fixed the longer this situation persists. Comparing the genetic population structure to the one of an earlier state G_{ST} values had become higher, however, D_{EST} were constant over time. This is a reason why D_{EST} values are limited in measuring genetic drift or migration (79). It is more useful to estimate mutation (77; 159) and when working with high variable markers (79). Under natural conditions true fixation in neutral markers will be opposed through mutation and migration between demes, however fixation in single genetic motifs may appear if the genomic region is under selection. Since G_{ST} and F_{ST} are sensible to this genetic differences it is often used as a measure for selection in loci (162). In general it should be considered to combine both statistics, which was also claimed by other authors (77), for a comprehensive approach to study population history in species dependent on neutral and non neutral markers.

Evolutionary effects of shell disease The situation of *C. pagurus* populations in Europe is under heavy alterations during the last decades. Southern populations at Portugal and Spain show a decline in abundance of the species (pers. comm Vitor Almada), while northern populations at Norway are increasing (163). A reason for this may be climate change induced temperature conditions leading to a shift in the distribution. Such a scenario is predicted for a marine sea-grass, *Zostera noltii* (164). To analyze the situation in *C. pagurus* two populations most extreme in their temperature regime were selected for comparing plasticity of *C. pagurus* towards different temperatures. In both populations high shell disease prevalence was found. Increased temperature at northern Norway did not lead to fatal losses or an increase of shell disease severity during experiments. An increase of shell disease severity was expected, since maximum chitinase activity from bacteria of shell disease lesions was measured after few days (46). However, shell disease may emerge at non-diseased animals in the lab in less than two weeks (7). Further, no increased level of heat shock protein or related transcripts were found implying that animals successfully acclimatized to these conditions. Increased expression was found for transcripts involved in metabolism and decreased expression of genes involved in the immune system. However, the expression was comparable to the expression in animals from Helgoland samples at similar temperature conditions. This indicates that animals from northern habitats can acclimatize to temperatures only present in southern populations. The survivability of individuals during experiments and the similar gene expression between populations under similar environmental conditions indicate low genetic adaptation in populations. High genetic diversity and effective population sizes in combination with high genetic connectivity over latitudes indicated strong population admixture and common gene pools over wide geographic ranges. High migratory potential of adults (165; 166; 167) and long duration of larval dispersal at low water temperature conditions (113) result in a high distributional potential of *C. pagurus* (168). This points to population shifting as the most likely mode of adaptation to future environmental conditions. This was observed as a general pattern for many species termed as a northward shift as a reaction to climate change (169; 170; 171). Such migrations may have already obscure population structures in *C. pagurus*.

There were some indications for local adaptation in *C. pagurus* as well. Two transcripts of the hemocyanines (subgroup 1 and subgroup 5) were affected differently under temperature increase. They showed significant differences between Helgoland ambient and Norway high temperature samples and were decreased by temperature increase in Norway. Hemocyanines are responsible

for respiration and immune reactions in crustaceans (172). Adaptation in the expression of these genes may be essential in order to adapt to altered temperature conditions and temperature related pathogen pressure. Those genes may be interesting candidates for further studies on local adaptation in *C. pagurus* populations. They may indicate a better performance in terms of immunity and oxygen utilization in Helgoland animals at high temperature conditions compared to northern Norway animals. Further, the genetic composition showed differences between animals from Helgoland and Norway with and without signs of shell disease. Strong differences in Hardy-Weinberg equilibrium were found in the pool of diseased animals for northern Norway. This may be due to a Wahlhund effect showing local admixture of different subpopulations and may indicate that diseased animals originated from different other areas. Furthermore, allele frequencies between diseased and non-diseased animals were significantly different in Norway but not in Helgoland. This might indicate that immigration from several other origins is more frequent in northern Norway than in Helgoland. Speculatively, the local population is better adapted to small temperature variance in the northern Norway population than immigrating individuals. The high exchange of genetic material over long distances in combination with a climate change scenario may, however, on the long run be more beneficial for immigrants. This would lead to a slow shift in populations and may be a reason for mixed populations, due to successful immigration and mixing with resident populations throughout the distributional range of *C. pagurus*.

References

1. Castro KM, Cobb JS, Gomez-Chiarri M, Tlusty M (2012) Epizootic shell disease in American lobsters *Homarus americanus* in southern New England: past, present and future. Diseases of Aquatic Organisms 100: 149–158.

2. Castro KM, Angell TE (2000) Prevalence and progression of shell disease in American lobster, *Homarus americanus*, from Rhode Island waters and the offshore canyons. Journal of Shellfish Research 19: 691–700.

3. Glenn RP, Pugh TL (2006) Epizootic shell disease in American lobster (*Homarus americanus*) in Massachusetts coastal waters: Interactions of temperature, maturity, and intermolt duration. Journal of Crustacean Biology 26: 639–645.

4. Gomez-Chiarri M, Cobb JS (2012) Shell disease in the American lobster, *Homarus americanus*: A synthesis of research from the New England lobster research initiative: Lobster shell disease. Journal of Shellfish Research 31: 583–590.

5. Meres NJ, Ajuzie CC, Sikaroodi M, Vemulapalli M, Shields JD, Gillevet PM (2012) Dysbiosis in epizootic shell disease of the American lobster (*Homarus americanus*). Journal of Shellfish Research 31: 463–472.

6. Quinn RA, Metzler A, Tlusty M, Smolowitz RM, Leberg P, Chistoserdov AY (2012) Lesion bacterial communities in American lobsters with diet-induced shell disease. Diseases of Aquatic Organisms 98: 221–233.

7. Quinn RA, Metzler A, Smolowitz RM, Tlusty M, Chistoserdov AY (2012) Exposures of *Homarus americanus* shell to three bacteria isolated from naturally occurring epizootic shell disease lesions. Journal of Shellfish Research 31: 485–493.

8. Tlusty M (2005) New in vivo methods to study shell formation and possible implications for shell disease. In: Lobster Shell Disease Workshop Forum. New England Aquarium, Boston, MA, 5-1, pp. 68–71.

9. Sindermann CJ (1991) Shell disease in marine crustaceans a conceptual approach. Journal of Shellfish Research 10: 491–494.

10. Roald SO, Arusjo J, Hastein T (1981) Occurrence of shell disease in lobsters *Homarus gammarus* in the southern part of Oslo Fjord Norway. Fiskeridirektoratets Skrifter Serie Havundersokelser 17: 153–160.

11. Vogan CL, Llewellyn PJ, Rowley AF (1999) Epidemiology and dynamics of shell disease in the edible crab *Cancer pagurus*: a preliminary study of Langland Bay, Swansea, UK. Diseases of Aquatic Organisms 35: 81–87.

12. Andersen LE, Norton JH, Levy NH (2000) A new shell disease in the mud crab *Scylla serrata* from Port Curtis, Queensland (Australia). Diseases of Aquatic Organisms 43: 233–239.

13. Dyrynda EA (1998) Shell disease in the common shrimp *Crangon crangon*: variations within an enclosed estuarine system. Marine Biology 132: 445-452.

14. Rushtonmellor SK, Whitfield PJ (1993) Transmission and scanning electron-microscopic studies of crustacean shell disease in fish lice of the genus *Argulus* (crustacea, branchiura). Journal of Zoology 229: 397–404.

15. Behringer DC, Butler MJ, Stentiford GD (2012) Disease effects on lobster fisheries, ecology, and culture: overview of dao special 6. Diseases of Aquatic Organisms 100: 89–93.

16. Soederhaell K (1982) Phenoloxidase activating system and melanization — a recognition mechanism of arthropods? a review. Developmental and Comparative Immunology 6: 601–611.

17. Moret Y, Moreau J (2012) The immune role of the arthropod exoskeleton. Isj-invertebrate Survival Journal 9: 200–206.

18. Nagasawa H (2012) The crustacean cuticle: structure, composition and mineralization. Frontiers in Bioscience (Elite Edition) 1: 711–720.

19. Stevenson RJ (1985) Dynamics of the integument. In: The Biology of Crustacea, volume 9. Academic Press, New York, 1-42 pp.

20. Felgenhauer BE (1991) External Anatomy and Integumental Structures. In: Microscopic Anatomy of Invertebrates, volume 10. Wiley-Liss Inc., New York, 3–43 pp.

21. Talbot P, Zao P (1991) Secretion at molting by the pleopod tegumental glands of the lobster *Homarus americanus* (Milne Edwards). Journal of Crustacean Biology 11: 1–9.

22. Herrick RH (1893) Cement glands and origin of egg membranes in the lobster. John Hopkins University Circular I2: 103.

23. Yonge CM (1932) On the nature and permeability of chitin. 1. the chitin lining the foregut of decapod crustacea and the function of the tegumental glands. Proceedings of the Royal Society of London Series B 111: 298-329.

24. Patane L (1959) Attivita elaboratrice e granulopessia nel cosiddeto "tessuto pleonale" degli isopodi terrestri. Bulletino Sedute Della Accademia Gioenia Di Scienze Naturali In Catania 5: 158-164.

25. Stevenson J (1961) Polyphenol oxidase in tegumental glands in relation to molting cycle of isopod crustacean *Armadillidium vulgare*. Biological Bulletin 121: 554–560.

26. Stevenson RJ, Schneider P (1962) Tyrosinase activity of organs containing tegumental glands in crayfish. Journal of Experimental Zoology 150: 17–25.

27. Greenberg M, Dodds M, Tian M (2008) Naturally occurring phenolic antibacterial compounds show effectiveness against oral bacteria by a quantitative structure-activity relationship study. Journal of Agricultural and Food Chemistry 56: 11151–11156.

28. Terwilliger NB, Ryan MC (2006) Functional and phylogenetic analyses of phenoloxidases from brachyuran (*Cancer magister*) and branchiopod (*Artemia franciscana*, *Triops longicaudatus*) crustaceans. Biological Bulletin 210: 38–50.

29. Armstrong E, Yan LM, Boyd KG, Wright PC, Burgess JG (2001) The symbiotic role of marine microbes on living surfaces. Hydrobiologia 461: 37–40.

30. Porter L, Butler M, Reeves RH (2001) Normal bacterial flora of the spiny lobster *Panulirus argus* and its possible role in shell disease. Marine and Freshwater Research 5.

31. Hall-Stoodley L, Costerton JW, Stoodley P (2004) Bacterial biofilms: From the natural environment to infectious diseases. Nature Reviews Microbiology 2: 95–108.

32. Kristensen JB, Meyer RL, Laursen BS, Shipovskov S, Besenbacher F, Poulsen CH (2008) Antifouling enzymes and the biochemistry of marine settlement. Biotechnology Advances 26: 471–481.

33. Anderson-Glenna MJ, Bakkestuen V, Clipson NJW (2008) Spatial and temporal variability in epilithic biofilm bacterial communities along an upland river gradient. Fems Microbiology Ecology 64: 407–418.

34. Vogan CL, Costa-Ramos C, Rowley AF (2001) A histological study of shell disease syndrome in the edible crab *Cancer pagurus*. Diseases of Aquatic Organisms 47: 209–217.

35. Smolowitz R, Chistoserdov AY, Hsu A (2005) A description of the pathology of epizootic shell disease in the American lobster, *Homarus americanus*, H. Milne Edwards 1837. Journal of Shellfish Research 24: 749–756.

36. Joseph FRS, Ravichandran S (2012) Shell diseases of brachyuran crabs. Journal of Biological Sciences 12: 117–127.

37. Sindermann C (1989) The shell disease syndrome in marine crustaceans. NOAA Technical Memorandum NMFS-F/NEc-64, 51 pp.

38. Castro KM, Factor JR, Angell T, Landers DFJ (2006) The conceptual approach to lobster shell disease revisited. Journal of Crustacean Biology 26: 646-660.

39. Tlusty MF, Myers A, Metzler A (2008) Short and long-term dietary effects on disease and mortality in american lobster, homarus americanus. Diseases of Aquatic Organisms 78: 249-253.

40. Chistoserdov AY, Smolowitz R, Mirasol F, Hsu A (2005) Culture-dependent characterization of the microbial community associated with epizootic shell disease lesions in American lobster, *Homarus americanus*. Journal of Shellfish Research 24: 741–747.

41. Quinn RA, Smolowitz R, Chistoserdov AY (2013) Culture-independent analysis of bacterial communities in hemolymph of American lobsters with epizootic shell disease. Diseases of Aquatic Organisms 103: 141–148.

42. Quinn RA, Cawthorn RJ, Summerfield RL, Smolowitz R, Chistoserdov AY (2013) Bacterial communities associated with lesions of two forms of shell disease in the American lobster (*Homarus americanus*, Milne Edwards) from Atlantic Canada. Canadian Journal of Microbiology 59: 380–390.

43. Chistoserdov A, Quinn RA, Gubbala SL, Smolowitz R (2009) Various forms and stages of shell disease in the American lobster share a common bacterial pathogen in their lesions. Journal of Shellfish Research 28: 689–689.

44. Rosen B (1970) Shell disease of aquatic crustaceans. Snieszko, Stanislas F (edited By) American Fisheries Society Special Publication, No 5 A Symposium On Diseases of Fishes and Shellfishes Viii + 526p Illus American Fisheries Society: Washington, DC, USA : 409–415.

45. Tlusty MF, Smolowitz RM, Halvorson HO, DeVito SE (2007) Host susceptibility hypothesis for shell disease in American lobsters. Journal of Aquatic Animal Health 19: 215–225.

46. Vogan CL, Costa-Ramos C, Rowley AF (2002) Shell disease syndrome in the edible crab, *Cancer pagurus* - isolation, characterization and pathogenicity of chitinolytic bacteria. Microbiology-sgm 148: 743–754.

47. Howell P, Benway J, Giannini C, McKown K, Burgess R, Hayden J (2005) Long-term population trends in American lobster (*Homarus americanus*) and their relation to temperature in Long Island Sound. Journal of Shellfish Research 24: 849–857.

48. Powell A, Rowley AF (2005) Unchanged prevalence of shell disease in the edible crab *Cancer pagurus* four years after decommissioning of a sewage outfall at Langland bay, UK. Diseases of Aquatic Organisms 68: 83–87.

49. Sawyer TK (1991) Shell disease in the Atlantic rock crab *Cancer irroratus* Say 1817 from the northeastern United States. Journal of Shellfish Research 10: 495–498.

50. Young RR (1991) Prevalence and severity of shell disease among deep-sea red crabs *Chaceon quinquedens* Smith 1879 in relation to ocean dumping of sewage sludge. Journal of Shellfish Research 10: 499–504.

51. Comely CA, Ansell AD (1989) The occurrence of black necrotic disease in crab species from the west of scotland. Ophelia 30: 95–112.

52. Stern C (1943) The Hardy-Weinberg Law. Science 97: 137-138.

53. Hardy GH (1908) Mendelian proportions in a mixed population. Science 28: 49-50.

54. Weinberg W (1908) Über den Nachweis der Vererbung beim Menschen. Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg 64: 368-382.

55. Wright S (1951) The genetical structure of populations. Annals of Eugenics 15: 323–354.

56. Wright S (1969) Evolution and the Genetics of Populations, Volume 2, Theory of Gene Frequencies, volume 2. University Of Chicago Press, 520 pp.

57. Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the United States of America 70: 3321–3323.

58. Nei M, Chesser RK (1983) Estimation of fixation indexes and gene diversities. Annals of Human Genetics 47: 253–259.

59. Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population-structure. Evolution 38: 1358–1370.

60. Cockerham CC (1969) Variance of gene frequencies. Evolution 23: 72–84.

61. Cockerham CC (1973) Analyses of gene frequencies. Genetics 74: 679–700.

62. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics 105: 767-779.

63. Blouin MS, Parsons M, Lacaille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology 5: 393–401.

64. Queller DC, Goodnight KF (1989) Estimating relatedness using genetic-markers. Evolution 43: 258–275.

65. Broquet T, Viard F, Yearsley JM (2013) Genetic drift and collective dispersal can result in chaotic genetic patchiness. Evolution 67: 1660–1675.

66. Gerlach G, Schardt U, Eckmann R, Meyer A (2001) Kin-structured subpopulations in Eurasian perch (*Perca fluviatilis* L.). Journal of Heredity 86: 213–221.

67. Iacchei M, Ben-Horin T, Selkoe KA, Bird CE, Garcia-Rodriguez FJ, Toonen RJ (2013) Combined analyses of kinship and fst suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. Molecular Ecology 22: 3476–3494.

68. Veliz D, Duchesne P, Bourget E, Bernatchez L (2006) Genetic evidence for kin aggregation in the intertidal acorn barnacle (*Semibalanus balanoides*). Molecular Ecology 15: 4193–4202.

69. Coyne JA, Orr HA (2004) Speciation. Sinauer Associates Inc., Sunderland, MA, USA,. ISBN 0-87893-091-4.

70. Tran TD, Hofrichter J, Jost J (2013) An introduction to the mathematical structure of the Wright-Fisher model of population genetics. Theory In Biosciences 132: 73–82.

71. Ewens WJ (1979) Testing the generalized neutrality hypothesis. Theoretical Population Biology 15: 205–216.

72. Rogers JS (1991) A comparison of the suitability of the Rogers, modified Rogers, Manhattan, and Cavalli-Sforza and Edwards distances for inferring phylogenetic trees from allele frequencies. Systematic Zoology 40: 63-73.

73. Jost L (2009) Partitioning diversity into independent alpha and beta components. Ecology 88: 2427–2439.

74. Jost L (2008) G(st) and its relatives do not measure differentiation. Molecular Ecology 17: 4015–4026.

75. Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59: 1633–1638.

76. Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on G(st) and D: forget G(st) but not all of statistics! Molecular Ecology 19: 3845–3852.

77. Leng L, Zhang DX (2011) Measuring population differentiation using G(st) or D? a simulation study with microsatellite dna markers under a finite island model and nonequilibrium conditions. Molecular Ecology 20: 2494–2509.

78. Wang J (2012) On the measurements of genetic differentiation among populations. Genetics Research 94: 275–289.

79. Ryman N, Leimar O (2009) G(st) is still a useful measure of genetic differentiation - a comment on jost's D. Molecular Ecology 18: 2084–2087.

80. Whitlock MC (2011) G '(st) and D do not replace F-st. Molecular Ecology 20: 1083–1091.

81. Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. Journal of Evolutionary Biology 14: 892–903.

82. Edelaar P, Bjorklund M (2011) If F-st does not measure neutral genetic differentiation, then comparing it with Q(st) is misleading. or is it? Molecular Ecology 20: 1805–1812.

83. Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc 6th Int Cong Genet 1: 356-366.

84. Wade MJ (1992) Sewall Wright: gene interaction in the shifting balance theory. Pp. 35–62 in J. Antonovics and D. Futuyma, eds. Oxford surveys of evolutionary biology VI. Oxford Univ. Press, New York.

85. Wade MJ (2013) Phase iii of Wright's Shifting Balance Process and the variance among demes in migration rate. Evolution 67: 1591–1597.

86. Watterson GA (1984) Allele frequencies after a bottleneck. Theoretical Population Biology 26: 387–407.

87. Maruyama T, Fuerst PA (1985) Population bottlenecks and nonequilibrium models in population-genetics. 2. Number of alleles in a small population that was formed by a recent bottleneck. Genetics 111: 675–689.

88. Santos J, Pascual M, Simoes P, Fragata I, Lima M, Kellen B, Santos M, Marques A, Rose MR, Matos M (2012) From nature to the laboratory: the impact of founder effects on adaptation. Journal of Evolutionary Biology 25: 2607–2622.

89. Kraemer P, Gerlach G (2013) Demerelate: Functions to calculate relatedness on diploid genetic data. URL `http://CRAN.R-project.org/package=Demerelate`. R package version 0.8-1.

90. Jueterbock A, Kraemer P, Gerlach G, Deppermann J (2010) DEMEtics: Evaluating the genetic differentiation between populations based on Gst and D values.

91. Sommer U, Stibor H, Katechakis A, Sommer F, Hansen T (2002) Pelagic food web configurations at different levels of nutrient richness and their implications for the ratio fish production: primary production. Hydrobiologia 484: 11-20.

92. Groenewold S, Fonds M (2000) Effects on benthic scavengers of discards and damaged benthos produced by the beam-trawl fishery in the southern north sea. ICES Journal of Marine Science 57: 1395-1406.

93. Breen PA, Mann KH (1976) Changing lobster abundance and the destruction of kelp beds by sea urchins. Marine Biology 34: 137-142.

94. Mann KH, Breen PA (1972) The relation between lobster abundance, sea urchins, and kelp beds. Journal of the Fisheries Research Board of Canada 29: 603-609.

95. Hall SJ, Robertson MR, Basford DJ, Fryer R (1993) Pit-digging by the crab *Cancer pagurus*: A test for long-term, large-scale effects on infaunal community structure. Journal of Animal Ecology 62: 59-66.

96. Hall SJ, Basford DJ, Robertson MR, Raffaelli DG, Tuck I (1991) Patterns of recolonisation and the importance of pit-digging by the crab *Cancer pagurus* in a subtidal sand habitat. Marine Ecology Progress Series 72: 93-102.

97. Dobzhansky T (1973) "Nothing in biology makes sense except in the light of evolution". American Biology Teacher 35.

98. Martin J, Davis G (2006) Historical trends in crustacean systematics. Crustaceana 79: 1347-1368.

99. Abele LG (1974) Species diversity of decapod crustaceans in marine habitats. Ecology 55: 156-161.

100. Dowle EJ, Morgan-Richards M, Trewick SA (2013) Molecular evolution and the latitudinal biodiversity gradient. Heredity 110: 501-510.

101. Loerz AN, Linse K, Smith PJ, Steinke D (2012) First molecular evidence for underestimated biodiversity of *Rhachotropis* (crustacea, amphipoda), with description of a new species. PLoS ONE 7: e32365.

102. Schaffner S, Sabeti P (2008) Evolutionary adaptation in the human lineage. Nature 1: 14.

103. Primmer CR (2011) Genetics of local adaptation in salmonid fishes. Heredity 106: 401-403.

104. Sotelo G, Moran P, Fernandez L, Posada D (2008) Genetic variation of the spiny spider crab *Maja brachydactyla* in the northeastern Atlantic. Marine Ecology Progress Series 362: 211-223.

105. 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 Genetics 6.

106. Brian JV, Fernandes T, Ladle RJ, Todd PA (2006) Patterns of morphological and genetic variability in UK populations of the shore crab, *Carcinus maenas* Linnaeus, 1758 (Crustacea: Decapoda: Brachyura). Journal of experimental marine biology and ecology 329: 47-54.

107. Atkins KE, Travis JMJ (2010) Local adaptation and the evolution of species' ranges under climate change. Journal of Theoretical Biology 266: 449–457.

108. Sanford E, Kelly MW (2011) Local adaptation in marine invertebrates. Annual Review of Marine Science 3: 509-535.

109. Guschina IA, Harwood JL (2006) Mechanisms of temperature adaptation in poikilotherms. FEBS Letters 580: 5477–5483.

110. Nichols J, Thompson B, Cryer M (1982) Production, drift and mortality of the planktonic larvae of the edible crab (cancer pagurus) off the north-east coast of england. Netherlands Journal of Sea Research 16: 173-184.

111. Wear RG (1974) Incubation in british decapod crustacea and the effects of temperature on the rate and success of embryonic development. Journal of the Marine Biology Association of the United Kingdom 54: 745-762.

112. Storch D, Santelices P, Barria J, Cabeza K, Poertner HO, Fernandez M (2009) Thermal tolerance of crustacean larvae (zoea I) in two different populations of the kelp crab *Taliepus dentatus* (Milne-Edwards). The Journal of Experimental Biology 212: 1371-1376.

113. Weiss M, Thatje S, Heilmayer O, Anger K, Brey T, Keller M (2009) Influence of temperature on the larval development of the edible crab, *Cancer pagurus*. Journal of the Marine Biological Association of the United Kingdom 89: 753–759.

114. Anger K, Thatje S, Lovrich G, Calcagno J (2003) Larval and early juvenile development of *Paralomis granulosa* reared at different temperatures: tolerance of cold and food limitation in a lithodid crab from high latitudes. Marine Ecology Progress Series 253: 243-251.

115. Sanford E, Holzman SB, Haney RA, Rand DM, Bertness MD (2006) Larval tolerance, gene flow, and the northern geographic range limit of fiddler crabs. Ecology 87: 2882–2894.

116. Templeman W (1934) Mating in the American lobster. Contributions to Canadian Biology and Fisheries 8: 422-432.

117. Richter K, Haslbeck M, Buchner J (2010) The heat shock response: Life on the verge of death. Molecular Cell 40: 253–266.

118. Hartl FU (1996) Molecular chaperones in cellular protein folding. Nature 13: 571-579.

119. Parsell DA, Lindquist S (1993) The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Annu Rev Genet 27: 437-496.

120. Nollen EAA, Morimoto RI (2002) Chaperoning signaling pathways: molecular chaperones as stress-sensing 'heat shock' proteins. Journal of Cell Science 115: 2809-2816.

121. Hershko A, Ciechanover A, Rose I (2004). Nobel prize in chemistry. URL http://nobelprize.org/chemistry/laureates/2004/. Last visited 03.05.2015.

122. Lyzenga WJ, Stone SL (2012) Abiotic stress tolerance mediated by protein ubiquitination. Journal of Experimental Botany 63: 599–616.

123. Shang F, Taylor A (2011) Ubiquitin-proteasome pathway and cellular responses to oxidative stress. Free 51: 5-16.

124. Poertner HO (2010) Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating climate-related stressor effects in marine ecosystems. Journal of Experimental Biology 213: 881-893.

125. Stocker T, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

126. Grieshaber MK, Hardewig I, Kreutzer U, Poertner HO (1994) Physiological and Metabolic Responses to Hypoxia in Invertebrates., volume 125 of *Reviews of Physiology, Biochemistry and Pharmacology*. Springer, 118 pp.

127. Windisch HS, Kathoever R, Poertner HO, Frickenhaus S, Lucassen M (2011) Thermal acclimation in antarctic fish: transcriptomic profiling of metabolic pathways. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology 301: 1453-1466.

128. Bulnheim HP, Scholl A (1981) Genetic variation between geographic populations of the amphipods *Gammarus zaddachi* and *G. salinus*. Marine Biology 64: 105-115.

129. Vazquez L, Alpuche J, Maldonado G, Agundis C, Pereyra-Morales A, Zenteno E (2009) Review: Immunity mechanisms in crustaceans. Innate Immunity 15: 179-188.

130. Dove AD, Allam B, Powers JJ, Sokolowski MS (2005) A prolonged thermal stress experiment on the American lobster, *Homarus americanus*. Journal of Shellfish Research 24: 761-765.

131. Tincu JA, Taylor SW (2004) Antimicrobial peptides from marine invertebrates. Antimicrobial Agents and Chemotherapy 48: 3645-3654.

132. Cerenius L, Soderhall K (2004) The prophenoloxidase-activating system in invertebrates. Immunological Reviews 198: 116-126.

133. Haug T, Kjuul A, Stensvag K, Sandsdalen E, Styrvold O (2002) Antibacterial activity in four marine crustacean decapods. Fish and Shellfish Immunology 12: 371–385.

134. Noga EJ, Engel DP, Arroll TW, McKenna S, Davidian M (1994) Low serum antibacterial activity coincides with increased prevalence of shell disease in blue crabs callinectes-sapidus. Diseases of Aquatic Organisms 19: 121–128.

135. Young RR (1989) Prevalence and severity of shell disease among deep-sea red crabs of the middle Atlantic bight in relation to ocean sewage sludge dumping. Journal of Shellfish Research 8: 462–462.

136. Vogan CL, Rowley AF (2002) Dynamics of shell disease in the edible crab *Cancer pagurus*: a comparative study between two sites on the gower peninsula, south wales, UK. Diseases of Aquatic Organisms 52: 151–157.

137. Sindermann CJ (1984) Disease in marine aquaculture. Helgoländer Wissenschaftliche Meeresuntersuchungen 37: 505-532.

138. Gerdts G, Brandt P, Kreisel K, Boersma M, Schoo KL, Wichels A (2013) The microbiome of North Sea copepods. Helgoland Marine Research 67: 757–773.

139. Becking LB (1934) Geobiologie of inleiding tot de milieukunde. The Hague, the Netherlands: WP Van Stockum & Zoon (in Dutch) .

140. Vogan CL, Powell A, Rowley AF (2008) Shell disease in crustaceans - just chitin recycling gone wrong? Environmental Microbiology 10: 826–835.

141. Costa-Ramos C, Rowley AF (2004) Effect of extracellular products of *Pseudoalteromonas atlantica* on the edible crab *Cancer pagurus*. Applied and Environmental Microbiology 70: 729–735.

142. Tlusty MF, Metzler A (2012) Relationship between temperature and shell disease in laboratory populations of juvenile American lobsters (*Homarus americanus*). Journal of Shellfish Research 31: 533–541.

143. Waksman SA, Carey CL, Reuszer HW (1933) Marine bacteria and their role in the cycle of life in the sea I. decomposition of marine plant and animal residues by bacteria. Biological Bulletin 65: 57–79.

144. Watson SW, Bock E, Valois FW, Waterbury JB, Schlosser U (1986) *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. Archives of Microbiology 144.

145. Suzuki I, Dular U, Kwok SC (1974) Ammonia or ammonium ion as substrate for oxidation by *Nitrosomonas europaea* cells and extracts. Journal of Bacteriology 120: 556–558.

146. Durand F, Regnault R (1998) Nitrogen metabolism of two portunid crabs, *Carcinus maenas* and *Necora puber*, during prologned air exposure and subsequent recovery: a comparative study. The Journal of Experimental Biology 201: 2515–2528.

147. Bell SL, Allam B, McElroy A, Dove A, Taylor GT (2012) Investigation of epizootic shell disease in American lobsters (*Homarus americanus*) from Long Island Sound: I. characterization of associated microbial communities. Journal of Shellfish Research 31: 473–484.

148. Colon-Gonzalez M, Mendez-Ortiz MM, Membrillo-Hernandez J (2004) Anaerobic growth does not support biofilm formation in *Escherichia coli* k-12. Research in Microbiology 155: 514–521.

149. Hsu AC, Smolowitz RM (2003) Scanning electron microscopy investigation of epizootic lobster shell disease in *Homarus americanus*. Biological Bulletin 205: 228–230.

150. Prince DL, Bayer RC, Loughlin M (1993) Etiology and microscopy of shell disease in impounded American lobsters, *Homarus americanus*. Bulletin of the Aquaculture Association of Canada 93: 87–89.

151. Landers DF (2005) Prevalence and severity of shell disease in American lobster *Homarus americanus* from eastern Long Island Sound, Connecticut. Aquatic Forim Series, New England Aquarium, Boston, MA, 94-97 pp.

152. Castro KM, Angell T, Somers B (2005) Lobster shell disease in southern New England: monitoring and research. Aquatic Forim Series, New England Aquarium, Boston, MA, 165-172 pp.

153. Comeau M, Benhalima K (2009) Internal organ pathology of wild American lobster (*Homarus americanus*) from eastern Canada affected with shell disease. New Zealand Journal of Marine and Freshwater Research 43: 257–269.

154. Bennett DB (1974) Growth of the edible crab (*Cancer pagurus* L.) off south-west England. Journal of the Marine Biological Association of the United Kingdom 54: 803-823.

155. Smith M, Laurans M, Scott F (2008). Evaluating the potential impact of current minimum landing sizes (mls) in english channel and associated edible crab (*Cancer pagurus*) fisheries under different growth and mortality assumptions - a data poor simulation modelling case study. Sixth framework programme Research for Policy Support Priority 8 Specific Targeted Research EC Contract No. 0022745.

156. Vogan CL, Rowley AF (2002) Effects of shell disease syndrome on the haemocytes and humoral defences of the edible crab, *Cancer pagurus*. Aquaculture 205: 237–252.

157. Tlusty M, Kim A, Castro K (2014) Modeling shell disease in American lobster (*Homarus americanus*) as individual-based health trajectories. Canadian Journal of Fisheries and Aquatic Sciences 71: 808-813.

158. Melzner F, Gutowska MA, Langenbuch M, Dupont S, Lucassen M, Thorndyke MC, Bleich M, Poertner HO (2009) Physiological basis for high CO2 tolerance in marine ectothermic animals: pre-adaptation through lifestyle and ontogeny? Biogeosciences 6: 2313–2331.

159. Ryman N, Leimar O (2008) Effect of mutation on genetic differentiation among nonequilibrium populations. Evolution 62: 2250-2259.

160. Wright S (1965) The interpretation of population structure by f-statistics with special regard to systems of mating. Evolution 19: 395-420.

161. Wright S (1978) Evolution and the Genetics of Populations, volume 4 of *Variability Within and Among Natural Populations.* University Of Chicago Press, Chicago, Illinois., 590 pp.

162. Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? Trends in Ecology and Evolution 20: 435–440.

163. Woll AK, van der Meeren GI, Fossen I (2006) Spatial variation in abundance and catch composition of *Cancer pagurus* in norwegian waters: biological reasoning and implications for assessment. Ices Journal of Marine Science 63: 421–433.

164. Valle M, Chust G, del Campo A, Wisz MS, Olsen SM, Garmendia JM, Borja A (2014) Projecting future distribution of the seagrass *Zostera noltii* under global warming and sea level rise. Biological Conservation 170: 74-85.

165. Ungfors A, Hallback H, Nilsson PG (2007) Movement of adult edible crab (*Cancer pagurus* L.) at the swedish west coast by mark-recapture and acoustic tracking. Fisheries Research 84: 345–357.

166. Edwards E (1979) The Edible Crab and its Fishery in British Waters. Fishing News Book Ltd, Farnham Surrey, England. 142 pp.

167. Hunter E, Eaton D, Stewart C, Lawler A, Smith MT (2013) Edible crabs "go west": Migrations and incubation cycle of *Cancer pagurus* revealed by electronic tags. PLoS ONE 8: 1-9.

168. Ungfors A, McKeown NJ, Shaw PW, Andre C (2009) Lack of spatial genetic variation in the edible crab (*Cancer pagurus*) in the Kattegat-Skagerrak area. ICES Journal of Marine Science 66: 462–469.

169. Helmuth B, Mieszkowska N, Moore P, Hawkins SJ (2006) Living on the edge of two changing worlds: Forecasting the responses of rocky intertidal ecosystems to climate change. Annual Review of Ecology, Evolution, and Systematics 37: 373-404.

170. Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421: 37-42.

171. Philippart CJM, Anadon R, Danovaro R, Dippner JW, Drinkwater KF, Hawkins SJ, Oguz T, O'Sullivan G, Reid PC (2011) Impacts of climate change on european marine ecosystems: Observations, expectations and indicators. Journal of Experimental Marine Biology and Ecology 400: 52–69.

172. Terwilliger NB, Dangott L, Ryan M (1999) Cryptocyanin, a crustacean molting protein: Evolutionary link with arthropod hemocyanins and insect hexamerins. Proceedings of the National Academy of Sciences of the United States of America 96: 2013-2018.

Abbreviations

ANOSIM	Analysis of Similarity
BCV	Biological Content of Variation
bp	base-pair
CRAN	Comprehensive R Archive Network
DNA	Deoxyribonucleic Acid
EnSD	Enootic Shell Disease
ESD	Epizootic Shell Disease
EST	Expressed Sequence Tags
FDR	False Discovery Rate
HE	Helgoland
IBD	Isolation by distance
INDEL	Insertion Deletion Mutation
LC	Left Carapace
LDG	Latitudinal Biodiversity Gradient
LINE	Long Interspersed Elements
LTR	Long Tandem Repeats
CPM	logarithmic counts per million
MDS	Multi Dimensional Scaling
MgCl	Magnesium Chlorid
MID	Multiplex Identifier
MMC	Marcov Monte Carlo
NO	Norway
OTU	Operational Taxonomic Unit
PAGE	Polyacrylamid gel electrophoresis
PCoA	Principle Coordinate Analysis
PCR	Polymerase Chain Reaction
Q-PCR	Quantitative Polymerase Chain Reaction
RC	Right Carapace
RNA	Ribonucleic acid
RV	Research Vessel
SBT	Shifting Balance Theory
SEM	Scanning Electron Microscopy
SINE	Short Interspersed Nuclear Elements
SSR	Simple Sequence Repeat
TSD	Trauma Shell Disease
UPGMA	Unweighted Pair Group Method with Arithmetic mean

Danksagung

Bleibt zu sagen

Ich danke Frau Prof. Gabriele Gerlach für ihre ständige Unterstützung, viele anregende Diskussionen und für die wissenschaftlichen Freiheiten, die ich während meiner Arbeit hatte.

Herrn Prof. Simon danke ich für die Übernahme der Zweitgutachtertätigkeit.

Herrn Dr. Peter Harmand danke ich für interessante Diskussionen über Biostatistik und spannende Einblicke in die Mathematik.

Auch Herr Prof. Thorsten Brinkhoff stand mir hilfreich zur Seite: Mit ihm habe ich die mikrobiologischen Fragestellungen meiner Arbeit diskutiert. Vielen Dank.

Für Zusammenhalt, Mut machen und mentalen Beistand danke ich Kristin Tietje, Alexandra Segelken-Voigt, Cornelia Hinz und Jana Deppermann.

Anke Müller und besonders Elke Frahman haben viele Stunden mit mir im Labor gestanden und mir oft aus der Bredouille geholfen. Danke!

Herrn Dr. Thomas Glatzel danke ich für so manchen väterlichen Rat.

Für gute Diskussionen und immer mal wieder einfach eine Kaffeepause danke ich Eike Mayland Quellhorst.

Viele haben geholfen Proben zu sammeln: Michael Jahnke, Achim Wehrmann, Aaron Jäschke, Sebastian Fuhrmann, Jan Knott, David de Leeuw, Edith Markert, Torsten Janssen und ein unbekannter Fischer. Danke.

Die European Science Foundation hat diese Arbeit finanziell unterstützt und meinen Auslandsaufenthalt in Norwegen ermöglicht (Grant EX/3204).

Ich danke außerdem Herrn Prof. Galice Hoarau für seine freundliche und hilfreiche Unterstützung bei meinem Aufenthalt an der Universität Nordland in Norwegen.

Herrn Dr. Alexander Jüterbock danke ich für die nette Unterkunft und gute Gesellschaft in Norwegen.

Frau Dr. Isabel Schmallenbach und Herrn Prof. Franke danke ich für die Möglichkeit, in der Biologischen Anstalt Helgoland des Alfred Wegener Instituts zu arbeiten.

Herr Dr. Matthias Labrenz vom Leibniz Institut für Ostseeforschung in Warnemünde hat mir erste Einblicke in die Bakteriengenetik ermöglicht. Dafür danke ich ihm.

Meinen Eltern bin für ihre Begleitung und ihren Rückhalt dankbar. Beides sind für mich wichtige Stützen gewesen.

Ich danke meiner Frau, Lena Hochstein, für die wunderbare Unterstützung in den letzten Jahren, für unglaublich viel Geduld und die Bereitschaft, mit mir auch noch den größten Quatsch durchzumachen.

Meiner Tochter Carla danke ich, dass sie mir gezeigt hat, was wichtig ist im Leben.

Annex

Package 'Demerelate'

May 24, 2013

Version 0.8-0

Date 2012-05-24

Title Functions to calculate relatedness on diploid genetic data

Author Philipp Kraemer and Gabriele Gerlach

Maintainer Philipp Kraemer <Philipp.Kraemer@uni-oldenburg.de>

Depends R (>= 2.15.0), fts, mlogit, sfsmisc

Suggests MASS

Description Functions to calculate pairwise relatedness on diploid genetic datasets. Different estimators for relatedness can be combined with information on geographical distances. Information on heterozygosity, allele- and genotype diversity as well as genetic F-statistics are provided for each population.

License GPL (>= 2)

URL http://www.r-project.org

BugReports http://Demerelate.bugtracker.url

R topics documented:

`Demerelate-package` *Demerelate — Algorithms to estimate pairwise relatedness within populations based on allele sharing*

Description

The package Demerelate provides several functions to calculate relatedness of individuals based on diploid genetic markers. Following indices can be calculated:

1. B_{xy} (number of alleles shared) as described in Li and Horvitz 1953.
2. M_{xy} (genotype sharing) as described in Blouin et al. 1996. Sharing rate is calculated according to shared allele positions i.e. 0, 1 or 2 shared allele postions for diploids. A sharing rate of 0 is calculated if no alleles are shared, a rate of 0.5 if one allele position is equal in individuals and a rate of 1 if individuals match in both allele positions.
3. The estimator r_{xy} based on Queller and Goodnight 1989 adapted to pairwise comparisons as described in Oliehoek et al. 2006 is calculated as follows:

$$r_{xy,l} = \frac{0.5(I_{ac}+I_{ad}+I_{bc}+I_{bd})-p_a-p_b}{1+I_{ab}-p_a-p_b}$$
$$I_{a-d} = \text{allele identities of individual I in locus l}$$
$$p_{a-b} = \text{frequencies of allele a or b in reference populations}$$

For each estimator randomized offsprings and randomized non related populations are created from a reference population. By default as reference all populations will be used, however functions can be forced to use specific data as reference. Be careful when using several populations or specific references. All populations need to be free of any pronounced population structure. References must contain every allele of the tested populations! Any violation of this may lead to strange results. Additionally, geographic distances can be used combined with genetic data to find out whether geographic distance has an effect on relatedness.

Allele sharing indices as B_{xy} or M_{xy} are analysed using a logistic regression model. Thresholds for individuals being full-siblings, half-siblings or non-related for each population are calculated. χ^2 statistics are used to calculate whether populations contain more siblings than expected in randomly drawn populations of the same number of individuals as the empirical population. The pairwise

estimator r_{xy} is calculated for different sets of references (full siblings, half siblings and non-related). T-tests are used to calculate whether significantly more siblings are present in populations than expected. Relatedness matrices for empirical populations as well as randomized populations are passed to the local directory and can be used for downstream analyses. Several plots from each analysis are exported as .pdf files for visualization of relatedness within populations. Additionally, basic information on datasets can be calculated for example allele and genotype frequencies, F_{is} values within populations or analysis on relative informativity of loci for relatedness analysis.

Details

Package: Demerelate
Version: 0.8-0
Date: 2013-05-24
Depends: R (>= 2.15.0), fts, mlogit, sfsmisc
Suggests: MASS
License: GPL (>= 2)
URL: http://www.r-project.org
BugReports: http://Demerelate.bugtracker.url

Demerelate	– Head function to calculate pairwise relatedness
Loci.test	– Function to analyse datasets of diploid genetic information
F.stat	– Function to calculate F_{is} values for single populations
demrelpop	– Example data set from randomized diploid markers for three populations
demrelref	– Example data set from randomized diploid markers as reference
demreldist	– Example data set from randomized relative geographic information

Author(s)

Philipp Kraemer and Gabriele Gerlach

Maintainer: Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Armstrong, W. (2012) fts: R interface to tslib (a time series library in c++). by R package version 0.7.7.
Blouin, M.S. et al. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Croissant, Y. 2011 mlogit: multinomial logit model R package version 0.2-2.
Li C.C. and Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5, 107-17.
Maechler, M. and myn others. (2012) Utilities from Seminar fuer Statistik ETH Zurich. R package version 1.0-20.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with

a focus on their use in conservation programs. Genetics, 173, 483-496.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.
Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.

`allele.sharing` *Calculates allele sharing rates or similarity estimators for two populations*

Description

Internal function of Demerelate to calculate different allele sharing indices or estimators for any pair of population. Following indices can be calculated: B_{xy} (number of alleles shared) as described in Li and Horvitz 1953. M_{xy} (genotype sharing) as described in Bluoin et al. 1996. The estimator r_{xy} based on Queller and Goodnight 1989 adapted to pairwise comparisons as described in Oliehoek et al. 2006.

Usage

```
allele.sharing(pop1, pop2, allele.column, onlypairs = FALSE,
               value = NA, ref.pop)
```

Arguments

`pop1`	Specific dataframe of type inputformat. Population one used for calculations. Individuals passed to rows of resulting matrix. Inputformat needs to be standard three digits mode for Demerelate.
`pop2`	Specific dataframe of type inputformat. Population two used for calculations. Individuals passed to columns of resulting matrix. Inputformat needs to be standard three digits mode for Demerelate.
`allele.column`	Numeric value which needs to be an odd number. It equals the number of the first column in the dataframe containing allele information. Order of loci in both populations needs to be exactly equal.
`onlypairs`	If set as TRUE allele sharing is only calculated for diagonal comparisons in the resulting pairwise relatedness matrix. If set as FALSE lower triangle is calculated from the pairwise relatedness matrix.
`value`	Character defining method to calculate allele sharing or similarity estimates. Can be set as "rxy", "Bxy" or "Mxy".
`ref.pop`	Reference population used for relatedness calculations.

Details

The function uses column one of each population dataframe to build a matrix with column names from samples of `pop2` and row names from samples of `pop1`. Similarity indices can be calculated for each pair. Indices are calculated using allelic information of one locus which first column must be indicated by `allele.column`. `allele.column+1` will be used automatically as second allele information.

Value

Function returns the object tab.all as matrix of rows passed from population one and columns from population two.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Blouin, M.S. et al. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Li C.C. and Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5, 107-17.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.

See Also

`inputformat` `Demerelate` `emp.calc` `queller`

Examples

```
## internal function not intended for direct use
```

`Demerelate` *Demerelate — Algorithms to estimate pairwise relatedness within populations based on allele sharing*

Description

Head function of Demerelate. This function should be called if any estimation of relatedness is intended. Additionally, some F statistics can be calculated. Default parameters are set for convenient usage. Only an input dataframe containing allelic information is necessary. Geographical distances, reference populations or alterations on statistics can be set by adapting parameters.

Usage

```
Demerelate(inputdata, tab.dist = "NA", ref.pop = "NA",
           object = FALSE, value = "Mxy", Fis = FALSE,
           file.output = FALSE, p.correct = FALSE,
           iteration = 1000, pairs = 1000,
           dis.data = "relative")
```

Arguments

`inputdata`	R object or external file to be read internally with standard Demerelate inputformat. Dataframe will be split by population information and calculations will run separately. If no reference population information is specified (reference.pop = "NA") all information on loci are used as reference by omitting population information.
`tab.dist`	R object or external file to be read internally with standard Demerelate inputformat. Geographic distances can be defined and will be analysed combined with genetic data. Column three and four of standard inputformat are used for x and y coordinates.
`ref.pop`	R object or external file to be read internally with standard Demerelate inputformat. Custom reference populations will be loaded for the analysis. Population information of reference file will be omitted so that allele frequencies are calculated from the whole dataset.
`object`	Information whether inputdata are objects or should be read in as files.
`value`	String defining method to calculate allele sharing or similarity estimates. Can be set as "rxy", "Bxy" or "Mxy".
`Fis`	*logical*. Should F_{is} values be calculated for each population?
`iteration`	Number of bootstrap iterations in F_{is} calculations.
`pairs`	Number of pairs calculated from reference populations for randomized full siblings, half siblings and non related individuals.
`file.output`	*logical*. Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.
`p.correct`	*logical*. Should Yates correction from `prop.test(...)` be used in χ^2 statistics when calculating p-values on differences between empirical and randomized relatedness in populations.
`dis.data`	The kind of data to be used as distance measure. Can be "relative" - relative x and y coordinates should be given in tab.dist or "decimal" for geographic decimal degrees.

Details

Pairwise relatedness is calculated from inputdata. Be sure to fit exactly the inputformat. Missing values are omitted when flagged as `NA`. If no additional reference populations are defined, inputdata

omitting population information are used to calculate references. If no good reference populations are available you need to take care of bias in calculations. In any case you should consult for example Olihoek et al. 2006 to get an idea of bias in relatedness calculations.
Geographic distances between individual pairs are calculated when tab.dist = ... Distances calculated from x-y coordinates by simple Pythagorean mathematics can be applied to any metrical positions in sampling. Geographic coordinates from e.g. GPS need to be transformed to decimal GPS coordinates. Be sure to have positions for each individual or remove missing values from inputdata.
Each calculation will have its unique bar-code and is named with the date and population name. Calculations are performed for each population in the inputdata.

Value

Function returns files in a folder named with a bar-code and date of analysis as follows if file.output is set as TRUE:

`Empirical.relatedness.Population.txt`
: Matrix of relatedness values for each population.

`Geographic.distances.Population.txt`
: Matrix of geographic distances for each population.

`Relate.mean.Populationout.name.txt`
: Depends on selected estimators and mode of analysis. Either a summary of correlation of relatedness with geographic distance for each population or a summary of tests for relatedness within populations compared to reference populations is written to the file.

`Random.Halfsib.distances.overall.txt`
: Matrix of relatedness values calculated from randomized reference population for half siblings.

`Random.NonRelated.distances.overall.txt`
: Matrix of relatedness values calculated from randomized reference population for non related individuals.

`Random.Fullsib.distances.overall.txt`
: Matrix of relatedness values calculated from randomized reference population for full siblings.

`Cluster.Populationout.name.pdf`
: Containing an UPGMA cluster dendogram of relatedness values and a histogram of relatedness values per locus and for loci overall.

`Total-Regression.Population.pdf`
: Containing regression plot and linear fit for geographic distance and genetic relatedness.

`Summary.Populationout.name.txt`
: Summary of analysis of F statistics and allele/genotype frequencies.

Function returns via `return` following objects as one list:

`dem.results[[1]]`
: Settings of the calculation are passed to this list object.

`dem.results[[2]]`
: Mean relatedness for empirical population over all loci.

`dem.results[[3]]`
: Summarized relatedness statistics with thresholds and randomized populations from the dataset.

`dem.results[[4]]`
: Statistical analysis of the number of siblings found for each population.

`dem.results[[5]]`
: F_{is} values and statistics for each population if `Fis==TRUE`

`dem.results[[6]]`
: Summary of linear regression of distance data are provided.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Blouin, M.S. et al. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Li C.C. and Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5, 107-17.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.

See Also

`inputformat` `emp.calc` `stat.pops` `F.stat`

Examples

```
## Data set is used to calculate Blouins allele sharing index on
## population data. Pairs are set to 10 for convenience.
## For statistical reason for your final results you may want to
## use more pairs to model relatedness (1000 pairs are recommended).

data(demrelpop)

dem.results <- Demerelate(demrelpop[,1:6], value="Mxy",
               file.output=FALSE, object=TRUE, pairs=10)


## Demerelate can be executed with values Bxy, rxy and Mxy you
## should consult the references to decided which estimator may
## be useful in your case.
## Be careful with Bxy this estimator may be biased and should be
## used with caution. You may want to use rxy instead.
```

demreldist *Example dataset for package Demerelate to calculate inter individual pairwise genetic relatedness.*

Description

The dataset gives randomized relative positions for each individual in the dataset.

Usage

```
data(demreldist)
```

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

inputformat Demerelate

Examples

```
## Consult examples from Demerelate
```

demrelpop *Example dataset for package Demerelate to calculate inter individual pairwise genetic relatedness.*

Description

The dataset gives randomized loci information of populations with differently related individuals. *Pop-FS-HS* consists of 10 full siblings (45 pairs) combined with 10 half siblings (45 pairs). *Pop-FS-Non* consists of 10 full sibilings (45 pairs) and 10 random individuals (45 pairs). *Pop-Non* is a population of 20 random drawn individuals (190 pairs). Datasets are based on information of 8 diploid loci with total number in alleles indicated by column names, i.e. number of alleles = 5, 10, 15, 20, 25, 30, 35, 40.

Usage

```
data(demrelpop)
```

Details

Standard inputformat is given for all applications in a similar way. Table is formatted as dataframe with headers set as TRUE. Headernames are not necessary but recommended.
First column contains sample IDs (mode=character), which should be unique. However, no proof-reading is implemented yet, but double named individuals may lead to errors or strange results. Column two contains population information (mode=factor). Pairwise relatedness is only calculated within populations. If you want to compare pairwise individual comparisons of several populations you need to build a new fictive population. You should do that with caution, since hidden population structures may produce strange results. Column three and four containing two alleles of a diploid marker for the sample defined in column one. Each marker should be appended in pairwise columns from then on.
Allele size should be sorted in columns with the small allele in the odd numbered column and the bigger in even numbered column for convenience. However, Demerelate will take care of not sorted allele sizes and handle it correctly. The same inputformat is used for reference populations.
All populations of inputdata are used as reference if no reference is defined by `reference.pop`. Additionally, the distance data are given (`tab.dist`) by the same inputformat. In column three and four you will need to define the x and y coordinate of each individual either by relative x-y coordinates or by decimal degrees of geographic coordinates.

Example input. Every dataframe in Demerelate should be organized in this way.

Sample-ID	Population	locus.1.a	locus.1.b	locus.2.a	locus.2.b	...
Ind.Norway.01	Norway	001	002	001	002	...
Ind.Norway.02	Norway	001	003	002	005	...
Ind.Norway.03	Norway	001	004	003	004	...
Ind.Norway.04	Norway	002	005	006	008	...
Ind.Germany.01	Germany	001	001	001	006	...
Ind.Germany.02	Germany	002	002	001	007	...
Ind.Germany.03	Germany	001	006	001	004	...
Ind.Germany.04	Germany	003	004	001	002	...

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

`Demerelate` `demreldist` `demrelref`

Examples

```
# Please consult examples from Demerelate
```

`demrelref` *Example dataset for package Demerelate to calculate inter individual pairwise genetic relatedness.*

Description

The dataset gives randomized loci information of a population with 1000 individuals. Dataset is based on information of 8 diploid loci with total number in alleles indicated by column names, i.e. number of alleles = 5, 10, 15, 20, 25, 30, 35, 40.

Usage

```
data(demrelref)
```

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

`Demerelate` `demreldist` `demrelpop`

Examples

```
## Consult examples from demrelpop
```

`emp.calc` *Function to calculate pairwise relatedness within populations based on allele sharing.*

Description

Allele sharing and relatedness estimators can be calculated for one population as a mean over several loci.

Usage

```
emp.calc(tab.pop.pop, value="rxy", ref.pop="NA")
```

Arguments

`tab.pop.pop`	Object with information for one population and loci. Object needs to be formatted according to inputformat
`value`	Either "rxy","Bxy","Mxy" are available to calculate mean relatedness among loci using allele.sharing
`ref.pop`	Reference population needs to be specified for rxy calculations

Details

The function calculates pairwise relatedness for all individuals in the dataframe. The output is a matrix of similarities by relatedness values in the population.

Value

`empirical.list` Object containing matrix of mean individual pairwise relatedness.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Blouin, M.S. et al. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Li C.C. and Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5, 107-17.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.

See Also

`allele.sharing` `queller`

Examples

```
## demrelpop needs to be an object of class data.frame formatted
## according to inputformat. For each pairwise comparison in
## in the data.frame empirical estimates of relatedness are
## calculated omitting statistics.
## Bxy, rxy and Mxy are possible to chose as estimator.

 data(demrelpop)

 # As example Mxy is calculated for one population of demrelpop

 demrelpop.sp <- split(demrelpop,demrelpop[,2])

 empirical.result <- emp.calc(demrelpop.sp[[1]], value="Mxy",
```

```
ref.pop="NA")
```

`F.stat` *Head function to calculate F statistics.*

Description

Calculation of F_{is} values for a single population and several loci. Calculations are based on statistics from Weir and Cockerham 1984 and Nei 1977.

Usage

```
F.stat(tab.pop, object = TRUE, iteration = 1000,
       directory.name = "NA", out.name = "NA")
```

Arguments

`tab.pop`	A file with a dataframe containing genetic information with format inputformat.
`object`	logical - whether data are object or file.
`iteration`	Number of bootstrap replicates
`directory.name`	Directory name where files should be exported. Directory must be present when running the function. If set as "NA" results are only printed on screen.
`out.name`	Name of file for exporting summary of statistics.

Details

This function is normally executed by Demerelate when F_{is} set as TRUE. However, F_{is} values can be calculated on single populations without the need of doing a whole new relatedness analysis. This may be useful if F_{is} values were not calculated in the first place. Be sure that the path already is present prior to analysis. However, if set as NA results are printed on screen or exported to an object. Inputdata will be splitted by population names and calculations will be made for each single population. Note that mean F_{is} values calculated by Weir and Cockerham's method are either given as arithmetic mean of single loci F_{is} and as weighted average over loci, which should be unbiased to a first aproximation according to Weir and Cockerham 1984. Since there are many ways to calculate mean F_{is} and all may serve in different situations as 'good' estimate you should spent some thoughts on this issue.

Value

Function returns a list for each population containing the following information:

`output.fis[[1]]`
: Empirical F_{is} values according to Nei 1977.

`output.fis[[2]]`
: Empirical F_{is} values according to Weir and Cockerham 1984.

`output.fis[[3]]`
: p values for significance for F_{is} values according to Nei 1977.

`output.fis[[4]]`
: p values for significance for F_{is} values according to Weir and Cockerham 1984.

Additionally, a file is generated containing all these information if file.output is set as TRUE or alternatively when directory.name and out.name are given.

`SummaryPopulationout.name.txt`
: Combined output with different F_{is} metrices and allele\genotype frequencies

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.

See Also

`Fis.calc`

Examples

```
## Fis statistics are calculated on demrelpop. Weir and Cockerham
## and Nei estimates are provided and either written to an object
## or to a file.
## For end results iteration of 1000 is recommended.


data(demrelpop)
fstat.results <- F.stat(demrelpop, iteration = 10,
                        directory.name = "NA",
                        out.name = "NA")
```

`Fis` *Calculates allele and genotype frequencies*

Description

Internal function to prepare allele and genotype frequencies for F statistics. Rows with `NA` values are removed from the calculation.

Usage

```
Fis(tab.pop, allele.column)
```

Arguments

`tab.pop`	Object with information for one or several populations and loci. Format needs to be in accordance with inputformat.
`allele.column`	Loci information from column `allele.column` and `allele.column`+1 are used for the calculations.

Details

Be sure to remove all missing data in your data file. Prior to calculation of F_{is} values every sample containing `NA` values in any column will be removed. It may be better to remove complete loci columns to get good results. In any case you should check this.

Value

`tab.freq`	Table of allele frequencies
`tab.freq.gen`	Table of genotype frequencies
`Homo`	Sum of homocygotes
`pop.size`	Population size
`Fis`	F_{is} values calculated according to Nei 1977
`Fis.weir`	F_{is} values calculated according to Weir and Cockerham 1984

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.

16 *Fis.calc*

See Also

weir Fis.calc F.stat

Examples

```
## internal function of F.sat not intended for direct use
```

`Fis.calc` *Calculation of Fis empirical and bootstrapped values*

Description

Internal function of Demerelate and F.stat to calculate F_{is} values from empirical datasets. Performing randomization statistics and preparing output.

Usage

```
Fis.calc(tab.pop, iteration, number.loci, object,
         directory.name, out.name)
```

Arguments

`tab.pop`	Data.frame following format of inputformat only column three and four are used for calculations
`iteration`	Number of replicates for bootstrap statistics
`number.loci`	Number of loci in tab.pop
`object`	Whether tab.pop is an object or file
`directory.name`	Name of the directory results send to
`out.name`	Filename of the output

Value

Function returns a list containing the following information:

`output.fis[[1]]`
: Empirical F_{is} values according to Nei 1977.

`output.fis[[2]]`
: Empirical F_{is} values according to Weir and Cockerham 1984.

`output.fis[[3]]`
: p values for significance for F_{is} values according to Nei 1977.

`output.fis[[4]]`
: p values for significance for F_{is} values according to Weir and Cockerham 1984.

Additionally, a file is generated containing all these information if file.output is set as TRUE or alternatively when directory.name and out.name are given.

`SummaryPopulationout.name.txt`
: Combined output with different F_{is} metrices and allele\genotype frequencies

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-Statistics for the analysis of population structure. Evolution, 38, 1358-1370.
Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.

See Also

`weir` `F.stat` `Fis`

Examples

```
## internal function of F.stat not intended for direct usage
```

`geo.dist` *Calculates geographic distances.*

Description

Internal function to prepare geographic distances for the combined analysis with genetic relatedness.

Usage

```
geo.dist(pop1, pop2, onlypairs = FALSE, value)
```

Arguments

`pop1`	Specific type of dataframe as in `inputformat`. Population one used for calculations. Individuals passed to rows of resulting matrix. Inputformat should be standart with x and y coordinate mode for Demerelate.
`pop2`	Specific type of dataframe as in `inputformat`. Population two used for calculations. Individuals passed to columns of resulting matrix. Inputformat should be standard with x and y coordinate mode for Demerelate.

onlypairs	If set as TRUE geographic distances is calculated only for diagonal comparisons in matrix. If set as FALSE lower triangel is calculated.
value	String defining method to calculate geographic distances. Can be set as "relative" or "decimal".

Details

Two different methods of distance calculations are implemented in Demerelate. If using "relative", distances will be calculated from x-y coordinates using normal pythagoratic mathematics. When working with geographic positions value needs to be set to "decimal". x and y coordinates need to be given as geographic positions in decimal degrees.

Value

matrix.share	Object containing geographic distances.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

Demerelate allele.sharing inputformat

Examples

```
## internal function not intended for direct use
```

---

glm.prep *Preparation for multinomial logistic regression.*

Description

Function combines randomized full and half siblings with empirical values and non related random individuals to one multinomial logistic regression to calculate relatedness thresholds.

Usage

```
glm.prep(empirical.list, offfull.list,
         offhalf.list, offnon.list)
```

Arguments

`empirical.list`	Mean relatedness for empirical population over all loci. Population information from inputdata are ommited.
`offfull.list`	Mean relatedness of populations randomized as full siblings over all loci from the reference population.
`offhalf.list`	Mean relatedness of populations randomized as half siblings over all loci from the reference population.
`offnon.list`	Mean relatedness of populations of randomized non relatives over all loci from the reference population.

Details

The function uses the package mlogit from Croissant 2011 to combine all the information from reference populations to one multilogistic regression model using mlogit(..). Thresholds are calculated and used for downstream calculations.

Value

`sumlmr`	Summary of mlogit model
`half`	Thresholds for full and half siblings

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Croissant, Y. 2011 mlogit: multinomial logit model R package version 0.2-2.

See Also

`offspring` `random.pairs`

Examples

```
## internal function not intended for direct use
```

`input.txt` *Reads different types of tables and returns an object.*

Description

Internal function to prepare inputdata for Demerelate.

Usage

```
input.txt(tab.txt, mod)
```

Arguments

`tab.txt`	Can be either a filename which will be passed to an object or an object in the R workspace
`mod`	Can be either "dist", "pop" or "ref.pop". Different informations are printed

Value

`tab`	Object containing dataframe as inputdata.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

`Demerelate` `inputformat`

Examples

```
## internal function not intended for direct use
```

`Lin.reg.distance` *Linear regression of empirical genetic relatedness with geographic distance.*

Description

Function caluclates CI intervals and exports plots and matrices for linear regression of empirical genetic relatedness with geographic distance.

Usage

```
Lin.reg.distance(dist.m, emp.dist, pairs, tab.pop.pop,
                 offhalf.list, offfull.list,
                 relate.off.non.Mxy.mean, file.output,
                 directory.name, out.name, inputdata,
                 object, value, iteration)
```

Arguments

`dist.m`	Object provided by `geo.dist`.
`emp.dist`	Empirical relatedness of population.
`pairs`	Number of bootstrap replicates to calculate confidence interval for linear regression.
`tab.pop.pop`	Population information to model linear fit.
`offhalf.list`	`Object` - matrix of mean pairwise relatedness of each randomized comparison of half siblings.
`offfull.list`	`Object` - matrix of mean pairwise relatedness of each randomized comparison of full siblings.
`relate.off.non.Mxy.mean`	`Object` - matrix of mean pairwise relatedness of each randomized comparison of non relatives..
`file.output`	*logical* - Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.
`directory.name`	Name of the directory results are send to.
`out.name`	Filename of the output.
`inputdata`	R object or external file to be read internally with standard Demerelate inputformat. Dataframe will be split by population information and calculations will run separately. If no reference population information is specified (reference.pop = "NA") all information on loci are used as reference by omitting population information.
`object`	Information whether inputdata are an object or should be read in as file.

`value` String defining method to calculate allele sharing or similarity estimates. Can be set as "rxy", "Bxy" or "Mxy".

`iteration` Number of bootstrap iterations in F_{is} calculations.

Details

Regression model is calculated by lm(...) for datasets. Be aware that your data need to fit assumptions for linear regression statistics. You should consult more statistics to check for e.g. normality or heteroskedastics in your data. Linear regression may be in most cases useful if calculating r_{xy} as estimator. The B_{xy} and M_{xy} may need considerable sample sizes to fit general assumptions for tests.

Value

`Total-Regression.Population.pdf`
Containing regression plot and linear fit for geographic distance and genetic relatedness.

`Relate.mean.Populationout.name.txt`
A summary of correlation of relatedness with geographic distance.

Summary statistics from linear regression of geogrphical distance and mean relatedness over loci are exported as list object.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

`geo.dist` `allele.sharing` `Demerelate`

Examples

```
## internal function not intended for direct use
```

`Loci.test` *Analysis on differences in mean relatedness based on number of loci used for calculations*

Description

Random samples are drawn from populations specified in `tab.pop`. Pairwise relatedness is calculated by either *rxy*, *Mxy* or *Bxy*. Pairwise relatedness is calculated for different number of loci beginning with only one up to the maximum number of loci in the dataset. Reference populations can be set for the calculations if `ref.pop="NA"`. `tab.pop` is used as an estimate of overall allele frequencies. Distance between pairwise relatedness estimates are calculated using `dist()` with euclidean calculation.
Graphically, a .pdf file is exported and the object `Random.loci.differences` contains each pairwise difference for usage in further statistics.

Usage

```
Loci.test(tab.pop, ref.pop = "NA", object = FALSE,
          value = "rxy", bt = 1000,
          file.output = FALSE)
```

Arguments

`tab.pop`	Specific dataframe of type inputformat. All populations which should be analyzed.
`ref.pop`	Specific dataframe of type inputformat. Population information which should be used as a reference for either threshold or allele frequency calculation.
`object`	*logical* - is tab.pop object.
`value`	String defining method to calculate allele sharing or similarity estimates. Can be set as "rxy", "Bxy" or "Mxy".
`bt`	Number of bootstrap replicates.
`file.output`	*logical* - should a .pdf file be sent as standard output in your working directory.

Value

`Loci.test.Sys.Date().pdf`
: Plot of linear regression of used loci and calculated mean estimate of relatedness from each bootstrapped replicate defined by `bt`.

`Random.loci.differences`
: List of differences in relatedness derived from bootstrap replicates. The x'th list element contains a matrix of means from replicates each over x random loci.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Blouin, M.S. et al. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Li C.C. and Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5, 107-17.

Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.
Maechler, M. and many others. (2012) Utilities from Seminar fuer Statistik ETH Zurich. R package version 1.0-20.

See Also

`inputformat` `allele.sharing`

Examples

```
## Loci.test to analyse Loci distribution in example data
## Bxy, rxy and Mxy are possible as estimators.

data(demrelpop)
demrelpop.sp <- split(demrelpop,demrelpop[,2])

Loci.results <- Loci.test(demrelpop.sp[[1]][,1:8],
                ref.pop = "NA", object = TRUE,
                value = "rxy", bt = 10)
```

`offspring` *Mendelian offpsring generator*

Description

Random generation of offspring from two parent individuals for one locus.

Usage

```
offspring(parent1, parent2, allele.column, pairs)
```

Arguments

`parent1`	data.frame of parent one.
`parent2`	data.frame of parent two.
`allele.column`	Loci information in column `allele.column` and `allele.column+1`.
`pairs`	Number of offspring generated from parent1 and parent2.

Details

Mendelian generator of random offspring.

Value

`offspr`	Dataframe of offpsring generated with length pairs.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

`random.pairs`

Examples

```
## internal function not intended for direct usage
```

`queller` *Queller and Goodnight estimator of relatedness*

Description

Calculation of symetric allele sharing using the RE-RAT method and allele frequencies for correction based on a reference population.

Usage

```
queller(pop1, pop2, allele.column, row.position,
        col.position, matrix.share, ref.pop)
```

Arguments

`pop1`	Population data containing allele information. All individual names are found in row names of pairwise relatedness matrix.
`pop2`	Population data containing allele information. All individual names are found in column names of pairwise relatedness matrix.
`allele.column`	Column number where loci information is taken from i.e. `allele.column` and `allele.column+1`.
`row.position`	Row position of the pairwise relatedness matrix which contains the name of one individual from population data `pop1`.
`col.position`	Column position of the pairwise relatedness matrix which contains the name of one individual from population data `pop2`.
`matrix.share`	Pairwise relatedness matrix containing individual names of `pop1` and `pop2`.
`ref.pop`	Reference population to calculate allele frequencies necessary for the calculation of the estimator.

Details

Be aware that unlike in some other programs single individuals are not removed from the calculation to correct for any bias. If you want to use custom reference population to calculate overall allele frequencies, which may be corrected for single individuals or parts of the population, these population need to be constructed manually and indicated by ref.pop. Otherwise it may be feasible to use the whole population information to calculate allele frequency distributions.

Value

`share` value of r_{xy}. NA is returned if estimator could not be calculated.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.

See Also

`random.pairs`

Examples

```
## internal function not intended for direct usage
```

`random.pairs` *Random pairs generator*

Description

Generating of random pairs from a single table with format inputformat.

Usage

```
random.pairs(tab.all, allele.column, pairs)
```

Arguments

`tab.all`	Dataframe containing genetic information with format inputformat.
`allele.column`	Loci information in column `allele.column` and `allele.column+1`.
`pairs`	Number of pairs to be generated.

Value

`rand.pairs`	List of random pairs.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

See Also

`offspring`

Examples

```
## internal function not intended for direct usage
```

`relate.calc` *Coordinates internally reference and empirical datasets and statistics*

Description

Internal function of Demerelate to combine different populations of randomized offspring and empirical populations to calculate thresholds and statistics for later use as reference for each empirical population.

Usage

```
relate.calc(tab.pop, pairs, file.output, value,
            directory.name, ref.pop)
```

Arguments

`tab.pop`	Dataframe following format of inputformat used as reference for randomizations.
`pairs`	Number of pairwise comparison for each randomization.

`file.output` *logical*. Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.

`value` String defining method to calculate allele sharing or similarity estimates. Can be set as "rxy", "Bxy" or "Mxy".

`directory.name` Name of the directory results send to.

`ref.pop` R object or external file to be read internally with standard Demerelate inputformat. Custom reference populations will be loaded for the analysis. Population information of reference file will be omitted so that allele frequencies are calculated from the whole dataset.

Details

The function internally calls any procedure of randomization and offspring generation. Finally, it coordinates the preparation of multinomial logistic regression for threshold calculation.

Value

`relate.return[[1]]`
`Object` - matrix of mean pairwise relatedness of all loci of each comparison in tab.pop.

`relate.return[[2]]`
`Object` - matrix of mean pairwise relatedness of each randomized comparison of full siblings.

`Random.Fullsib.relatedness.overall.txt`
`file` - matrix of mean pairwise relatedness of each randomized comparison of full siblings.

`relate.return[[3]]`
`Object` - matrix of mean pairwise relatedness of each randomized comparison of half siblings.

`Random.Halfsib.relatedness.overall.txt`
`file` - matrix of mean pairwise relatedness of each randomized comparison of half siblings.

`relate.return[[4]]`
`Object` - matrix of mean pairwise relatedness of each randomized comparison of non relatives.

`Random.NonRelated.relatedness.overall.txt`
`file` - matrix of mean pairwise relatedness of each randomized comparison of non relatives.

`relate.return[[5]]`
`Object` - calculated thresholds for relatedness from reference populations.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Blouin, M.S. et al. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Li C.C. and Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5, 107-17.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.

See Also

`emp.calc` `offspring` `random.pairs` `glm.prep`

Examples

```
## internal function not intended for usage without Demerelate
```

`stat.pops` *Calculation of Fis empirical and bootstrapped values*

Description

Internal function of Demerelate to use prepared thresholds of relatedness to calculate state of relatedness for empirical data. Additionally, it combines information in different plots and prepares each single output.

Usage

```
stat.pops(Thresholds, tab.pop.pop, pairs, p.correct,
          directory.name, out.name, file.output,
          inputdata, object, value, iteration,
          ref.pop)
```

Arguments

`Thresholds`	Thresholds of relatedness to be used for statistics on empirical data.
`tab.pop.pop`	Dataframe following format of inputformat only column three and four are used for calculations.
`pairs`	Number of randomized pairings used for reference populations.
`p.correct`	*logical* - should Yates correcture be used for χ^2 statistics.
`directory.name`	Name of the directory results send to.
`out.name`	Filename of the output.

`file.output` *logical*. Should a cluster dendogram, histograms and .txt files be sent as standard output in your working directory. In some cases (inflating NA values) it may be necessary that this value has to be set as FALSE due to problems in calculating clusters on pairwise NA values.

`inputdata` R object or external file to be read internally with standard Demerelate inputformat. Dataframe will be split by population information and calculations will run separately. If no reference population information is specified (reference.pop = "NA") all information on loci are used as reference by omitting population information.

`object` Information whether inputdata are an object or should be read in as file.

`value` String defining method to calculate allele sharing or similarity estimates. Can be set as "rxy", "Bxy" or "Mxy".

`iteration` Number of bootstrap iterations in F_{is} calculations.

`ref.pop` R object or external file to be read internally with standard Demerelate inputformat. Custom reference populations will be loaded for the analysis. Population information of reference file will be omitted so that allele frequencies are calculated from the whole dataset.

Details

Values for logical operators and data are given by Demerelate.

Value

`ClusterPopulationNameOutName.pdf`
Combined information on dataset by different types of plots.

`Relate.meanPopulationNameOutName.txt`
Combined information on dataset regarding calculated thresholds and number of different types of relatives in population. Summary of χ^2 statistics.

`out.stat[[1]]` Empirical relatedness in populations.

`out.stat[[2]]` Summary statistics for χ^2 statistics.

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

Blouin, M.S. et al. (1996) Use of microsatellite loci to classify individuals by relatedness. Molecular Ecology, 5, 393-401.
Li C.C. and Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5, 107-17.
Oliehoek, P. A. et al. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics, 173, 483-496.
Queller, D.C. and Goodnight, K.F. (1989) Estimating relatedness using genetic markers. Evolution, 43, 258-275.

See Also

weir F.stat

Examples

```
## internal function not intended for usage without Demerelate
```

`weir` *Calculation of F_is based on Weir and Cockerham 1984*

Description

Function used internally by Demerelate and F.stat to calculate F_{is} values based on Weir and Cockerham (1984) statistics.

Usage

```
weir(tab.pop, tab.freq, popsize)
```

Arguments

`tab.pop`	Dataframe following format of inputformat only column three and four are used for calculations.
`tab.freq`	Vector of allele frequencies with name of each allele in column name.
`popsize`	Population size.

Details

The function provides the alternative methods to Nei's F_{is} calculation. F_{is} values are calculated according to Weir and Cockerham 1984. Consult references for details

Value

`fis.weir`	F_{is} value calculated by method of Weir and Cockerham 1984

Author(s)

Philipp Kraemer, <philipp.kraemer@uni-oldenburg.de>

References

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

See Also

Fis.calc F.stat Fis

Examples

```
## internal function of F.stat not intended for direct usage
```

Index

Curriculum Vitae

PHILIPP KRÄMER

PERSÖNLICHES

Geboren in Salzkotten, 2. Januar 1982

Adresse Am Stadion 8, 27798 Hude

E-Mail philipp.kraemer@uni-oldenburg.de

Familienstand Verheiratet, 1 Kind

Beruf Diplom Biologe

ARBEITSERFAHRUNG

seit 10/2013 Carl von Ossietzky Universität Oldenburg

Wissenschaftlicher Mitarbeiter Biological timing in a changing environment: Clocks and rhythms in polar pelagic organisms

seit 10/2010 datadiving GmbH & Co. KG

Gesellschafter Gründungsmitglied der datadiving GbR seit 2013 GmbH & Co. KG mit Sitz in Hamburg. Angewandtes Forschungstauchen, Einsatzleitung, verschiedene Projekte im Bereich Meeresbiologische Datenerfassung und Gutachten.

09/2008-02/2009 Carl von Ossietzky Universität Oldenburg

Wissenschaftlicher Mitarbeiter Populationsgenetische Studien am Atlantischen Kabeljau *Gadus morhua*, Georges Bank.

AUSBILDUNG

seit 05/2009 Carl von Ossietzky Universität Oldenburg

Promotions-student "Implication of climate change on the susceptibility of the brown crab (Cancer pagurus) Evolutionary effects of host-disease interaction."

08/2007-07/2008 Carl von Ossietzky Universität Oldenburg

Diplomarbeit "Feeding of fish in different habitats of the North Sea" im MarFood Projekt

04/2005-07/2008 Carl von Ossietzky Universität Oldenburg

Hauptdiplom Studium der Biologie, Diplomprüfungen in Zoologie, Ökologie und Genetik

11/2005-10/2006 Carl von Ossietzky Universität, Oldenburg

Geprüfter Forschungstaucher Studiumsbegleitende Ausbildung zum Forschungstaucher

10/2002-04/2005 Georg August Universität, Göttingen

Vordiplom Studium der Biologie; Vordiplomsprüfungen in Zoologie, Genetik, Chemie und Physik

08/1991-06/2001 Goerdeler Gymnasium Paderborn

Allg. Hochschulreife Schulische Ausbildung, Abitur

FÖRDERUNG / PREIS

Poster Preis Für das Poster: "Local adaptation of brown crab (*Cancer pagurus*): immune and stress response in a changing climate." New Model Systems for Linking Evolution and Ecology. European Molecular Biology Laboratory; 2013 May 1 - 4, Heidelberg, Germany.

ESF Travel Grant European Science Foundation Grant EX/3204; 10/2012 - 12/2012: Thermal adaptation in ectotherms: Linking life history, physiology, behaviour and genetics. Title: Population genetics of brown crabs: Local thermal adaptation and its implications on host-pathogen interactions. Host: Hoarau, G.

PUBLIKATIONEN

Patent US 8646390 B2 Device for the use of technical equipment underwater; Krone, R. and Kraemer, P., 2014.

Kraemer, P. and Gerlach, G., 2013. Demerelate: Functions to calculate relatedness on diploid genetic data. R package version 0.8-0. http://cran.r-project.org/web/packages/Demerelate/index.html

Kraemer, P. and Gerlach, G., 2013. Local adaptation of brown crab (*Cancer pagurus*): immune and stress response in a changing climate. Poster session presented at: New Model Systems for Linking Evolution and Ecology. European Molecular Biology Laboratory; 2013 May 1 - 4, Heidelberg, Germany.

Patent US 8413609 B2 Device for colonizing and harvesting marine hardground animals; Krone, R. and Kraemer, P., 2013.

Patent WO 2011147400 A2 Device for developing habitats in the underwater area of an offshore construction; Kraemer, P., Krone, R. and Schroeder, A., 2013.

Patent US 8291863 B2 Transportable device for colonizing and harvesting invertebrates, and use of said device; Krone, R. and Kraemer, P., 2012.

Kaben-Dobbeck, M., Kraemer, P., Markert, A., Wehrmann, A. and Gerlach, G., 2011. Significance of shell disease between the native shore crab *Carcinus maenas* and the invasive Asian brush crab *Hemigrapsus takanoi*. 104th Annual DZG Meeting 2011, 9th-12th September, Saarbrücken, Germany

Gerlach, G., Jueterbock, A., Kraemer, P., Deppermann, J. and Harmand, P., 2010. Calculations of population differentiation based on GST and D: Forget GST but not all of statistics! Molecular Ecology, 19: 3845-3852.

Krämer, P., Dannheim, J. and Schröder, A., 2007. Eat whats on your plate! Feeding of demersal fish in different habitats. 42nd European Marine Biology Symposium, 27th-31st August, Kiel, Germany.

Erklärungen gemäß §10 der Promotionsordnung

Hiermit erkläre ich gemäß §10 der Promotionsordnung, dass ich mit dieser Dissertation den Titel des Dr. rer. nat. (Doktor) anstrebe.

Hiermit erkläre ich gemäß §10 der Promotionsordnung, dass ich die Arbeit selbstständig verfasst und nur angegebenen Hilfsmittel benutzt habe.

Hiermit erkläre ich gemäß §10 der Promotionsordnung, dass ich meine Dissertation weder in ihrer Gesamtheit noch in Teilen einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegt habe.

Oldenburg, den