AWI
Alfred-Wegener-Institut
für Polar- und Meeresforschung
in der Helmholtz-Gemeinschaft

The recent evolutionary history of Antarctic and Subantarctic benthic isopods: Development and analysis of fast evolving molecular markers

Die jüngere evolutive Vergangenheit ausgewählter Isopoden im Benthos der Antarktis und Subantarktis: Entwicklung und Analyse schnell evolvierender molekularer Marker

Dissertation zur Erlangung des Grades
eines Doktors der Naturwissenschaften der
Fakultät für Biologie und Biotechnologie
der Ruhr-Universität Bochum

angefertigt am
Lehrstuhl für Spezielle Zoologie, Ruhr-Universität Bochum und am
Alfred-Wegener Institut für Polar- und Meeresforschung, Bremerhaven

vorgelegt von
Florian Leese (*18.07.1978),
aus Lauterbach (Hess.)

Bochum und Bremerhaven, April 2008

1. Gutachter: Prof. Dr. J.-W. Wägele (RUB, ZFMK Bonn)

2. Gutachter: Prof. Dr. D. Begerow (RUB)

Erklärung

Hiermit erkläre ich, dass ich die Arbeit selbstständig verfasst und bei keiner anderen Fakultät eingereicht und dass ich keine anderen als die angegebenen Hilfsmittel verwendet habe. Es handelt sich bei der heute von mir eingereichten Dissertation um fünf in Wort und Bild völlig übereinstimmende Exemplare.
Weiterhin erkläre ich, dass digitale Abbildungen nur die originalen Daten enthalten und in keinem Fall inhaltsverändernde Bildbearbeitung vorgenommen wurde.

Bochum, den 15.04.2008

Florian Leese

Für meine Familie, insbesondere Rinske und Frida (*09. April 2008)

Contents

Chapter 1

General Introduction

„WHY ARE THERE SO MANY DIFFERENT KINDS OF ANIMAL SPECIES?"
Hutchinson, 1959

Background

High biodiversity has traditionally been associated primarily with the old tropical ecosystems, where high productivity and the great variety of different niches allow the coexistence of many different species. It is therefore relatively new to science that the marine ecosystems of the cold and inhospitable Antarctic are much more diverse in terms of species numbers than previously thought. While taxonomists and systematic biologists are still in the process of discovering the unknown treasures to make them available for science, evolutionary biologists and ecologists alike adopt the frequently asked, central question in biology in the Antarctic and ask "why are there so many species in the marine Antarctic benthos?"

1.1 The Antarctic

The Antarctic is a uniquely extreme region on Earth. Antarctica is the coldest, windiest, driest, highest and most isolated continent on Earth. The surrounding Southern Ocean is characterized by extraordinarily strong current systems and constantly low water temperatures within a range of -1.86 to +2.0°C. Formerly, the Antarctic continent was part of the large southern supercontinent Gondwana. After its break-up, approximately 180 Myr ago, the East and West Antarctic plates drifted polewards and collided, forming the continent Antarctica, surrounded by a continuous ocean system (Lawver et al. 1992). In the late Eocene (37-33 Myr BP) the separation of the Australian plate from Antarctica triggered the severe cooling of the Antarctic (Kennett 1977). Reduction in global CO_2 concentration is discussed as another major triggering mechanism of global cooling (Barrett 2003, DeConto and Pollard 2003, Barker and Thomas 2004), although this is still controversial (Veizer et al. 2000, Shaviv and Veizer 2003 but see Royer et al. 2004). Between 30 and 24 Myr ago, the opening of the Drake Passage (Barker and Burrell 1977, Barker et al. 1982, Lawver et al. 1992, Livermore et al. 2005), the last continental bridge between Antarctica and South America, enabled the development of a very strong continuous current system: the Antarctic Circumpolar Current (ACC), also termed West Wind Drift with respect to the underlying forces (Fig. 1). The Southern Ocean was formed as an independent, circumpolar Ocean, thermally isolated from the upper water layers of all other oceans (but see Clarke et al. 2005). The mild 'greenhouse' conditions prevailing in the Cretaceous shifted to harsh 'icehouse' conditions during the Eocene (Fig. 2). A permanent ice cover was established over major areas of the continent in the early Oligocene with sea ice and shelf ice recurrently covering major parts of the marine habitats (Barron et al. 1991). For Antarctic terrestrial and marine life, this major shift in environmental conditions demanded fundamental physiological and behavioural adaptations as 'life in Earth's largest freezer' (Thatje et al. 2008) is extremely challenging on all organizational levels (Clarke 1983, Pörtner and Playle 1998, Poulin et al. 2002, Pörtner et al. 2007). Many species, in particular terrestrial ones, could not cope with the profound climate change and went extinct (Dayton 1990, Knox 1994, Crame 1997 for review). However, despite the harsh climatic conditions physical disturbances by ice scouring (Gutt 2001, Gutt et al. 2004, Barnes and Conlan 2007) the marine Antarctic realm houses a very diverse biota, in particular in the benthic faunal communities on the Antarctic shelf, which contrasts the extreme paucity of Antarctic terrestrial life (Convey 2001, 2003). Due to the strong isolation and the distinct current regimes, the Antarctic shelf with its benthic inhabitants provides a unique natural laboratory to study and test evolutionary hypotheses concerning colonization of and speciation.

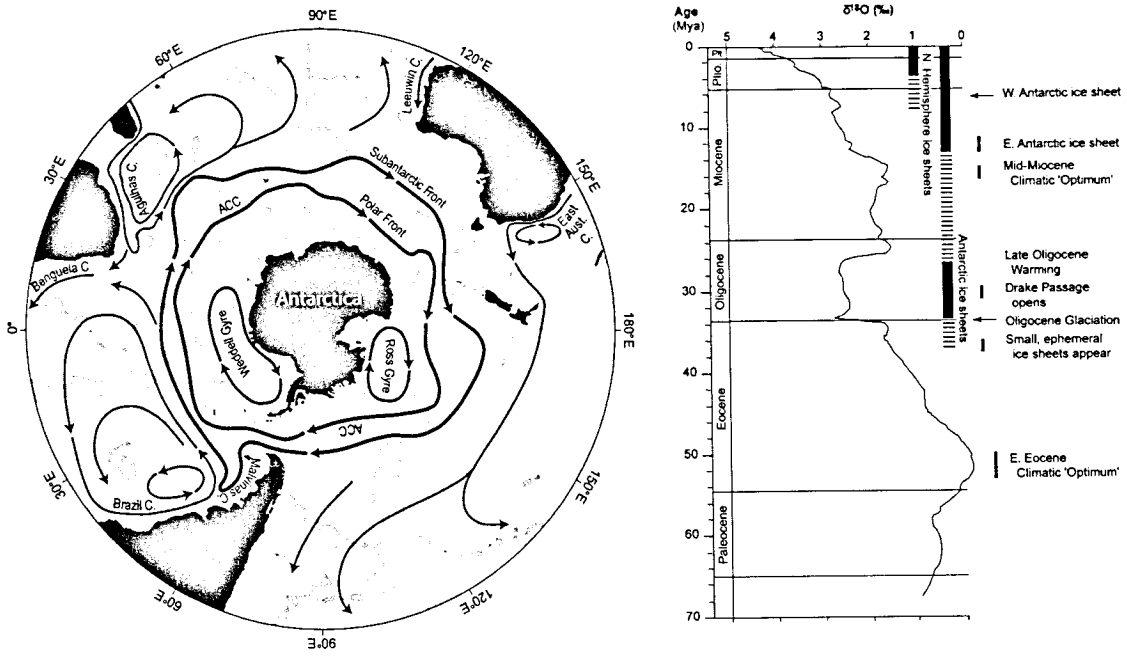


Figure 1: (*left*) Map of the major current systems in the Southern Ocean and the neighbouring ocean basins in the southern hemisphere, south of 20°S. The Antarctic Circumpolar Current (ACC) is the strongest current system of the world's oceans characterized by a fast eastward drift of the surface water layers between the Subantarctic Front and the Polar Front. Depths less than 3500 m are shaded grey. Modified after Rintoul et al. (2001). **Figure 2:** (*right*) Change of the climate during the last 70 million years as inferred from oxygen isotopes. Adapted from Poulin et al. (2002).

1.2 The marine Antarctic benthos

The marine Antarctic benthos reveals a unique and diverse fauna, which differs significantly from other known benthic communities (Knox and Lowry 1977, Dayton 1990, Clarke 1992, Arntz et al. 1994, 1997, Knox 1994, Clarke and Johnston 2003, Gutt et al. 2004). Many of the benthic taxa produce a relatively small number of lecitotrophic eggs, which they fertilize and brood in the interior of their body (Dell 1972, Picken 1980, Luxmoore 1982, Clarke 1991, Arnaud 1992, Hain and Knox 1994, Poulin and Feral 1996). This predominance of K-strategists is interpreted as a result of adaptive convergence in evolutionary time scales as a response to polar conditions with strong seasonality (White 1977, Clarke 1979, 1980).

Recent biodiversity analyses estimate species number as high as 4100 species (Clarke and Johnston 2003), which is much higher than previously thought. Species numbers are however systematically underestimated because of two fundamental problems: (1) the still poor sampling effort around the Antarctic, in particular the East Antarctic, leaves many species go unnoticed. All expeditions to unexplored regions detect animal species, which are new to science. (2) Many species that were assumed to have a broad distribution simply consist of a flock of morphologically indistinguishable, i.e. cryptic species (Allcock et at al. 1997, Page and Linse 2002, Held 2003, Lee et al. 2004, Held and Wägele 2005). As a consequence, species number in the marine Antarctic benthos is probably underestimated by an order of magnitude (Gutt et al. 2004).

Precise estimates of species numbers are important for biodiversity studies and analyses of the underlying ecological and evolutionary processes in general (Rosenzweig 1995). For the

Antarctic region in particular, the recent findings of many closely related cryptic species with smaller, allopatric or parapatric distribution ranges in the Antarctic benthic taxa, however, challenge what had become an almost unquestioned biogeographic paradigm in Antarctic benthic research for many decades: The concept of circumantarctic distribution of many species (Dayton 1990, Arntz et al. 1994, Clarke 1996, Clarke and Johnston 2003). For the notothenioid fishes (Eastman and McCune 2000) and the Antarctic serolid isopods (Held 2000) radiations on the Antarctic continent after thermal isolation have been documented. The recent findings add additional support that many species do not solely represent the remnants from the warm Eocene (Crame 1994) that were able to physically adapt to the harsh conditions, but new, Antarctic lineages. Recent phylogenetic studies provided evidence against the commonly held opinion that evolution might be retarded in the Antarctic (Held 2001, Lanfear et al. 2007). While these data continuously add to our knowledge on Antarctic biodiversity, the central evolutionary question about the origin of this high biodiversity, i.e. the nature of the 'diversity pump' (Clarke and Crame 1992), remains subject of speculations.

Recurrent large-scale glacial disturbances

The lack of obvious physical barriers on the shelf at present coupled with the strong current regimes are seen to promote gene flow counteracting genetic differentiation which is necessary for speciation. While knowledge about other microevolutionary processes promoting speciation even in parapatric or sympatric conditions increases (Knowlton 1993, 2000, Palumbi 1994) the most parsimonious explanation for speciation in the Antarctic have often been the scenarios invoking habitat fragmentation by glacial advances in the Pleistocene (2 Myr – 20 Kyr BP, e.g. Clarke and Crame 1989, 1992, Crame 1997, Aronson and Blake 2001, Thatje et al. 2005). During the recurrent glacial periods on Milankovitch periods (Imbrie et al. 1993) grounded shelf ice advanced as far as to the outer continental shelf in most but not all parts of the Antarctic shelf (Anderson et al. 2002, Huybrechts 2002, Evans et al. 2004, Ingolfsson 2004) (Fig. 3). While the high biodiversity of endemic sessile or largely immobile benthic taxa on the shelf indicates that the shelf fauna must have survived the glacial periods in some sheltered regions it remains an open question where and how they survived glacial periods when most of the habitat was eradicated by massive glaciers (Fig. 4). It is still untested, whether these conditions are ultimately responsible for speciation by historical allopatric fragmentation by means of massive grounded ice caps for several ten thousands of years. In particular, as the mobility of many brooding species is more restricted than that of species with pelagic larvae (Palumbi 1994, Bohonak 1999), even minor glaciations might have sufficiently reduced gene flow thus allowing for speciation even under para- or peripatric conditions. If population sizes underwent a severe bottleneck during the harsh glacial periods this further explains rapid divergence and speciation on the Antarctic shelf. Demographic histories of Antarctic fish species were shown to be different from those of pelagic fishes, which has been interpreted as a consequence of the glacial influences on population structure (Janko et al. 2007). Consequently, a scenario of historical allopatric fragmentation of populations and lineage sorting within these during glacial maxima are regarded a plausible mechanism of frequent allopatric speciation of benthic taxa on the Antarctic shelf. The few studies conducted on benthic isopods support this scenario (Held 2000, 2003, Held and Wägele 2005). However, direct, molecular genetic evidence for such a scenario is rare and only from results of a pioneer study on benthic fishes (Janko et al. 2007).

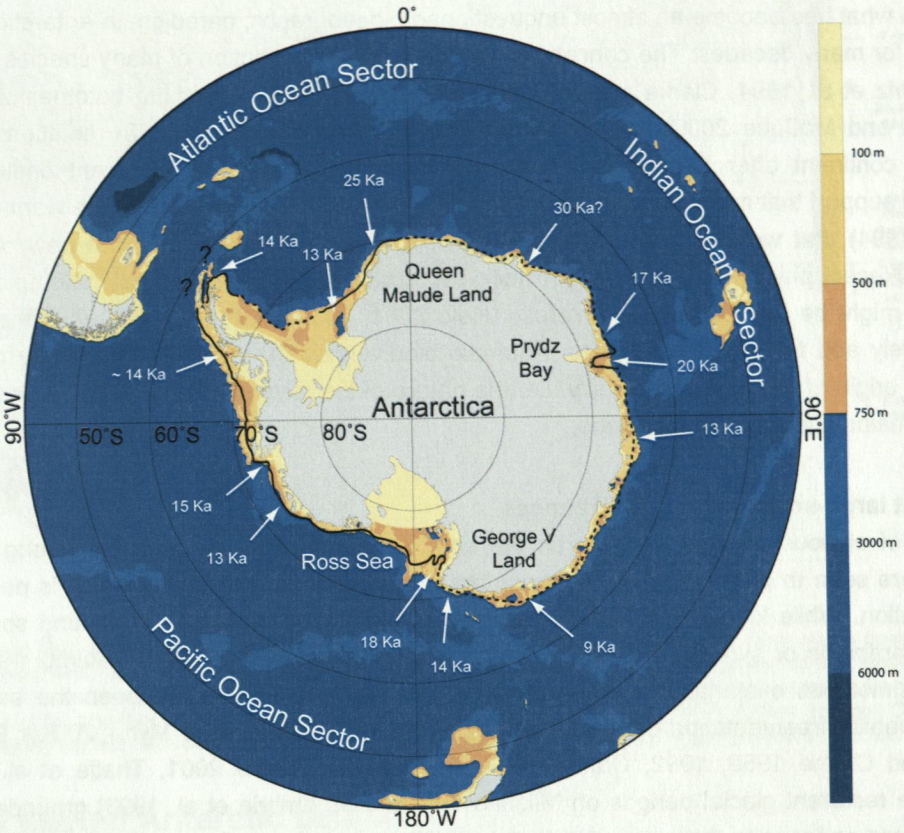


Figure 3: Maximum extent of grounding shelf ice around the Antarctic continent during the last glacial maximum (black line) according to Anderson et al. (2002), Evans et al. (2004). Dashed parts of the line refer to estimates of the extent of grounded ice. White numbers are estimates of the minimum age (in 1,000 years) of ice-sheet retreat from the Antarctic shelf during deglaciation based on radiocarbon dating (Anderson et al. 2002).

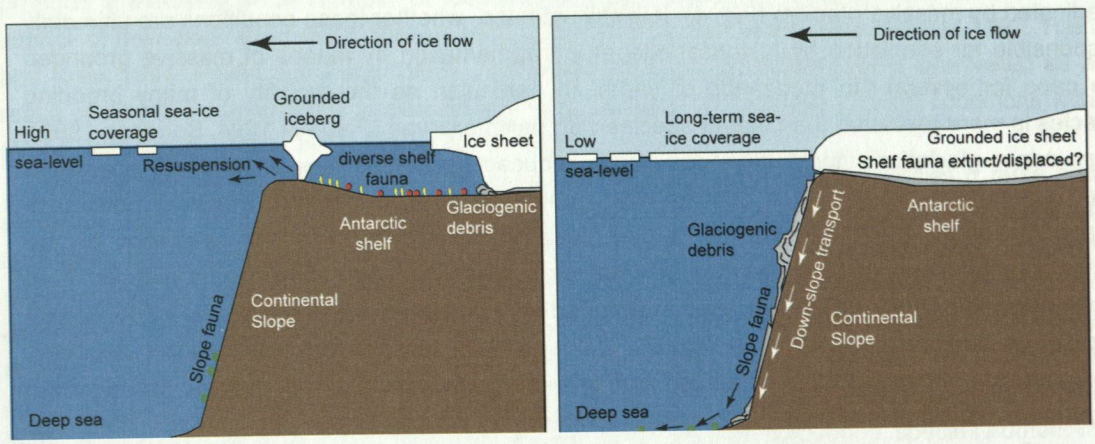


Figure 4: Vertical profile of the Antarctic shelf region at interglacial periods (left) and glacial maxima (right). Grounded shelf ice advanced as far as to the continental shelf break during glacial maxima, erasing most of the habitat available for the shelf fauna. Modified after Thatje et al. (2005), see also Brey et al. (1996).

1.3 Study case: serolid isopods (Isopoda, Serolidae Dana, 1852)

The Serolidae Dana, 1852 are a family of marine benthic isopods predominantly distributed on softbottom habitats on the continental shelves of the Southern Hemisphere (Wägele 1994, see Martin and Davies 2001 for taxonomic positioning). Some species also colonized deep-sea habitats and few extended their ranges into the northern hemisphere. There are currently 88 described species in 21 genera (retrieved March, 11th, 2008, from the Integrated Taxonomic Information System online database, http://www.itis.gov.) This number must be regarded as provisional due to recent findings of cryptic species (Held 2003, Leese and Held 2008, Held in prep.). The Antarctic serolid isopods were demonstrated to represent a monophyletic clade with closest relatives found on the shelves of South America (Held 2000). This makes this group the second animal group after the notothenioid fishes (Eastman and McCune 2000) for which the existence of an Antarctic species flock has been demonstrated. Like all other peracarid isopods the Serolidae brood their offspring in a ventral brood pouch, the marsupium, and lack free-swimming distribution stages.

Serolids are a model taxon for studying microevolutionary patterns in the Antarctic as (1) they are a particularly successful taxon group with many endemic Antarctic taxa (Brandt 1991, Wägele 1994, Held 2000), (2) there are numerous studies on their ecological, physiological and biogeographical characteristics (Clarke 1982, Luxmoore 1982, 1984, Wägele 1986, 1987, 1994, Brandt 1988, 1991, Held 2000, Held 2001, Held 2003), and (3) members of the Serolidae with very similar life strategies inhabit also Subantarctic islands and cold-temperate shelf areas, which allows a comparative analysis to distinguish between population genetic features that are due to their life strategy (low mobility, few descendants) and those that are a result of the history of the inhabited niches (Antarctic shelf with recurrent disturbances). In the context of Southern Ocean history, the three aspects in the center of analysis in this thesis, are (1) the population genetic structure of species and the amount of gene flow between populations, (2) the population genetic diversity in different regions of a species and (3) patterns of historical reductions or expansion of the effective population size as a response to fluctuating conditions such as glaciations. These three aspects will be investigated using one model species from the High Antarctic shelf (*Ceratoserolis* n. sp. 1), one from remote Antarctic and Subantarctic islands (*Septemserolis septemcarinata*), and, for comparison, one from the Magellan Region (*Serolis paradoxa*) (Fig. 5).

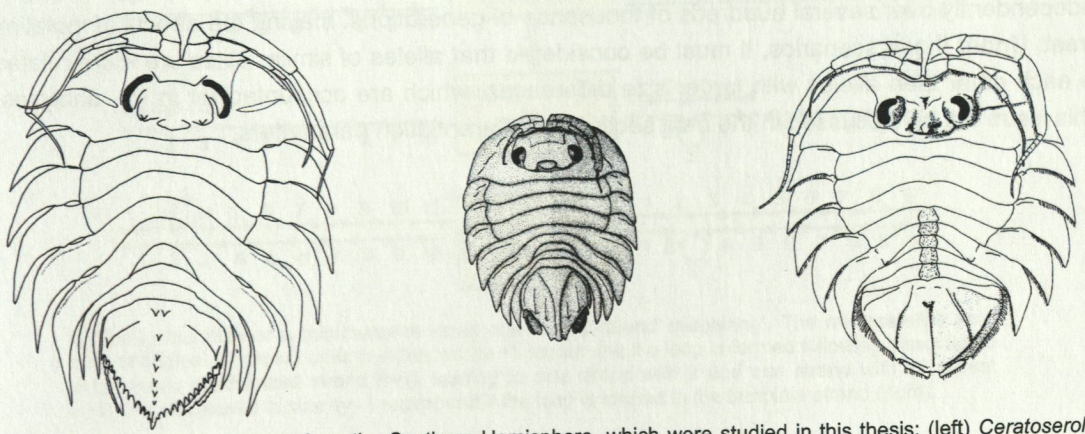


Figure 5: Three serolid isopods from the Southern Hemisphere, which were studied in this thesis: (left) *Ceratoserolis trilobitoides sensu* Wägele 1986 from the Antarctic (drawing: Dr. Christoph Held), (middle) *Septemserolis septemcarinata* from Subantarctic islands (drawing: Brandt, 1991) and *Serolis paradoxa* from the Patagonian shelf (drawing: Wägele, 1994). Not drawn to scale.

1.4 Molecular markers for Antarctic population genetics

Molecular markers for population genetic or phylogenetic studies differ in their resolution, the time and costs needed for setup and their reproducibility (Avise 1994, Sunnucks 2000, Schlötterer 2004 for reviews). The choice of a molecular marker system has frequently been made on the basis of availability or was 'just a matter of fashion' (Schlötterer 2004) rather than being the well-considered selection based on the resolution needed for the scientific question to be addressed.

For phylogenetic studies, which address questions on splitting events of major taxonomic lineages (genera, families, classes etc.), conserved single-locus DNA markers, which suffer little from multiple substitutions, are the matter of choice in general. For studies aimed at describing or validating species status ('Barcoding'), fast evolving markers are preferred (e.g. COI, 16S, 18S, 28S rDNA). For population genetic questions dealing with the distribution of genetic variability among populations of a single species, even more variable, preferably independent and co-dominant markers should be applied (Held and Leese 2007). In Antarctic population genetic research such variable markers were previously only applied to few vertebrate or economically important invertebrate species (see Held and Leese 2007 and literature cited within). For the Antarctic benthos, most studies utilized gene sequencing applying universal primers (e.g. Simon et al. 1994). No study so known to the author isolated most informative microsatellite markers, which on the other hand are the most variable and informative marker system for population genetics and therefore the method of choice for the present thesis.

Microsatellites are genomic regions with a tandemly repeated nucleotide motif with a unit length of 2-6 bp, e.g. 5'-ACACACACACACAC-3' $(AC)_7$ (see Chambers and MacAvoy (2000) for an overview over the different definitions). They are particularly variable (Weber and Wong 1993, Schlötterer 2000), which makes them extremely useful for population genetic or individual assignment studies, where much variation at taxonomically low levels is needed (Selkoe and Toonen 2006, see Fig. 1 in Held and Leese 2007). Their application for more exploratory studies is still growing (Fig. 6), in particular for more exploratory studies.

The source of the uniquely high polymorphism of most microsatellites is the result of errors in the replication machinery of the cell (see Fig. 7). The most frequent mutational process is the so-called slip-strand mispairing, SSM (Levinson and Gutman 1987). The dominance of this mutational mechanism has profound consequences on the inference of differentiation between closely related populations, as mutations cannot be neglected if populations diverged independently over several hundreds of thousands of generations, making the effects of mutation great. Under these scenarios, it must be considered that alleles of similar sizes are more related to each other than alleles with larger size differences, which are accounted for in R_{ST} analyses. This issue will be discussed in the next section on differentiation parameters.

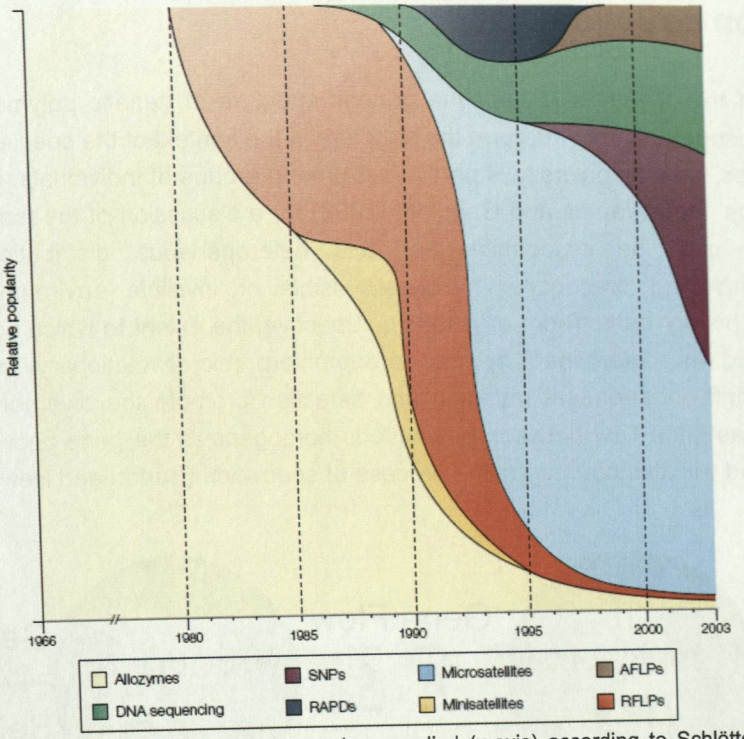


Figure 6: Popularity of molecular markers applied (x-axis) according to Schlötterer (2004). The enormous expansion in SNP popularity is based on their frequent application in studies on model organisms such as *Homo sapiens*, *Drosophila melanogaster* or *Mus musculus*. The number of studies based on microsatellites is shrinking on the relative scale but increases exponentially in absolute numbers (see records in GenBank, Primer Notes in Molecular Ecology Notes and Conservation Genetics).

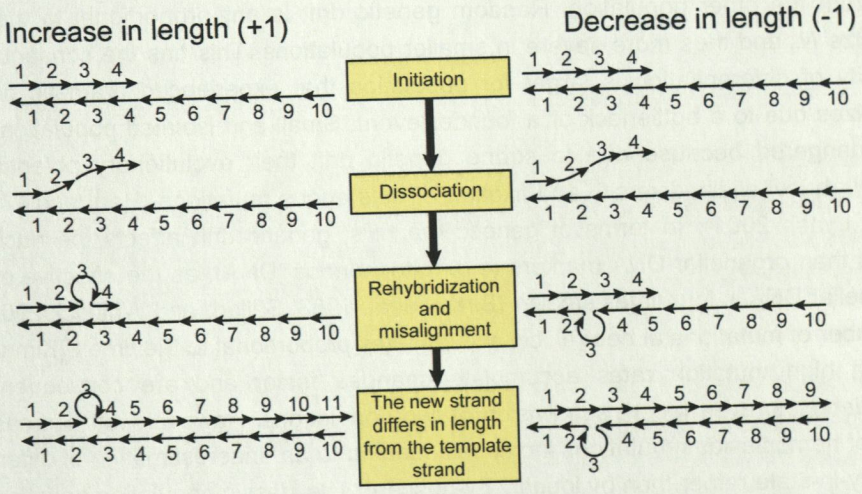


Figure 7: Mutation at a microsatellite locus due to 'slip-strand mispairing'. The microsatellite core motif of original 10 repeat units is extended by +1 repeat unit if a loop is formed following dissociation in the newly synthesized strand (left), leading to one strand with 9 and one strand with 10 repeat units. It is decreased in size by -1 repeat unit if the loop is formed in the template strand (right).

1.5 Population structure

The main topic of this thesis is to infer the genetic structure of genetic polymorphisms within species of serolid isopods and to interpret the structure in the context of the species' history.

In most species, genetic diversity is partitioned among groups of individuals, which are often termed 'populations' (see Waples and Gaggiotti (2006) for a discussion of the term 'population'). Manifold causes may be responsible for this heterogeneous distribution of genetic polymorphisms, including present or historical, visible or invisible environmental barriers, behavioural or life history traits. From a genetic perspective, the extent to which populations may become subdivided, is determined by four predominant microevolutionary forces (Fig. 7): Random genetic drift, independent mutation and selection promote the divergence of isolated populations whereas gene flow between populations homogenizes the gene pool and precludes local adaptation and will also counteract the process of speciation (Barton and Hewitt 1985).

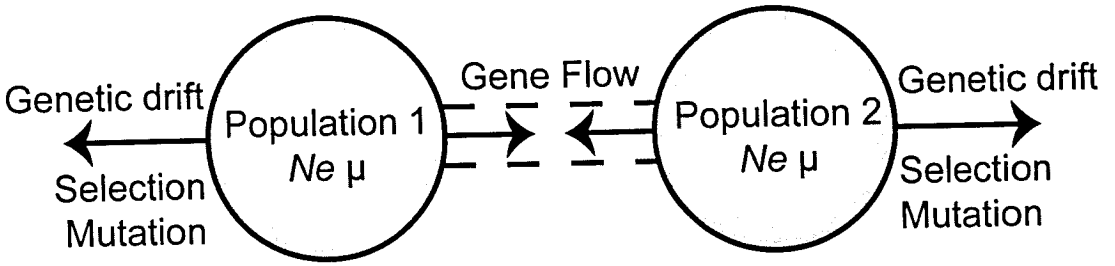


Figure 8: Microevolutionary processes that drive and counteract the disruption of a once contiguous gene pool of two populations.

Gene flow is proportional to the number of migrants *m*, which successfully mate with individuals from the other population. Random genetic drift is anti-proportional to the effective population size N_e and thus more severe in smaller populations. This has the consequence that the probability of differentiation is larger for population that experienced dramatic declines in population sizes due to a bottleneck or a founder event. Small and isolated populations may be severely endangered because due to strong genetic drift their evolutionary potential will be affected through low allelic diversity and fixation of deleterious mutations (Hartl and Clark 1997, Higgins and Lynch 2001). In terms of genetic markers, genetic drift affects the nuclear DNA markers less than organellar DNA markers (e.g. mitochondrial DNA), as the effective population size of organellar DNA is four times smaller (Birkey et al. 1983, Ballard and Whitlock 2004).

The number of mutations at neutral loci accumulates proportional to the time *t* (Kimura 1983). Markers with high mutation rates accumulate changes faster and are consequently more sensitive in detecting even recent subdivision of two populations. However, with increasing time the number of homoplasious mutations increases, leading to an underestimation of differentiation due to identity-in-state rather then by identity-by-descent. The classic population genetic statistics derived by Wright (1931, 1951, 1965) does not account for this. However, this 'homoplasy-trap' (Wägele 2001) must be considered in particular when the time of independent evolution in subpopulations is not negligible, in particular for fast evolving markers such as microsatellites. The two most common used differentiation statistics, F- and R-statistics, differ in particular in this respect of incorporating mutation in their assumptions:

For classical F_{ST} estimates (Wright 1951, 1965) assuming an island model of migration, the amount of population differentiation is simply a function of the number of migrants, if mutation is: negligible.

$$F_{ST} = \frac{1}{4N_e + m(+\mu)}$$

For a bi-allelic locus, Wright introduced an approach to partition the genetic variation in a subdivided population that is undoubtedly the most commonly used one until today. However, the statistical background was slightly modified to account for multiple alleles and to allow a statistical analysis of the significance of differences in an Analysis of Variance Framework, e.g. Weir and Cockerham (1984). They defined F_{ST} as a correlation coefficient that estimates the probability that two alleles from two subpopulations share a common ancestry:

$$F_{ST} = \frac{Q_S - Q_T}{1 - Q_T}$$

where, Q refers to the probability of identity-in-state (IIS) for pairs of genes either between individuals within a subpopulation (Q_S) and between subpopulations (Q_T). All calculations rely on calculating variances within and between subpopulations for the alleles observed. The drawback of F-statistics is that it assumes that mutation is small compared to mutation ($m >> \mu$) and therefore genetic drift is the dominating microevolutionary force to partition genetic variability. While this scenario is appropriate for population genetic analyses on a small temporal scale with fast evolving markers, the assumption of $m >> \mu$ is strongly violated when using microsatellites to infer population structure that might have built up over a long time. Therefore, Slatkin (1995) proposed to implement the mutation model of microsatellites, the stepwise mutation model (see

$$R_{ST} = \frac{W_T - W_S}{W_T}$$

Fig. 7) to this formula by:

where W_T is the mean squared difference in allele size between the total number of pairs of alleles. W_S refers to the mean sum of squares of differences in allele size within each subpopulation. Consequently, R_{ST} is the fraction of the total variance of allele size that is between populations. The two allele models that underlie the two different statistics are (1) the Infinite Allele Model, IAM, and (2) the Stepwise Mutation Model, SMM:

(1) The IAM was proposed by Kimura and Crow (1964). This mutation model assumes that each mutation creates a new allele not previously present in the gene pool at a given mutation rate μ. Allele sizes are not predictable and relatedness of alleles cannot be inferred. Similar alleles in this model share the same ancestry and are identical-by-descent. This model is assumed for the calculation of F-statistics following the assumption that mutation is great of mutation ($m>>\mu$), i.e. differentiation is mainly caused by random genetic drift.

(2) The SMM was proposed by Otha and Kimura (1973). The model assumes each mutation creates a new allele that differs by +1 or - 1 repeat unit from the ancestral allele for a microsatellite locus. The probabilities of adding or deleting are μ/2. Consequently, this model allows for homoplasy by stating that identical alleles are not necessarily identical-by-descent. In particular when divergence between two populations is large, neglecting mutation may severely underestimate the true amount of divergence. Under a SMM, the allele size keeps information about the relatedness of alleles, i.e., alleles of very different sizes are expected to be more distantly related than alleles of similar size. This is the underlying model assumed in R-statistics (Slatkin 1995, Rousset 1996).

Both models assume that there are an infinite number of possible allelic states. For microsatellite loci empirical evidence supports that there is an upper limit of microsatellite lengths (Nauta and Weissing 1996). Although several microsatellites with a length of several thousands of nucleotides were detected (Mayer et al. in prep.), most microsatellites are in the range of 5 to 50 repeat units. For populations that were separated for a long period, high mutation rates will at a certain point start to deflate R_{ST} estimates. F_{ST} estimates are much less affected by this phenomenon (Balloux et al. 2000, Balloux and Lugon-Moulin 2002). The same effect is observed when mutations deviate from a strict SMM (Balloux et al. 2000). All the points raised make clear that neither F- nor R-statistics are 'better' in general. If little is known about the degree of isolation among populations, the mutation rate of the markers and the mutational mechanisms, both methods should always be applied and compared (Balloux and Lugon-Moulin 2002). To make F_{ST} values comparable among studies using different markers and most important to make an understanding of the relative amount of the fixation of different alleles, F_{ST} values should be standardized, as they often do not reach the theoretical maximum of 1. This approach was suggested by Hedrick (2005) and extended in a variance framework by Meirmans (2006) based on the principle of estimating the maximum possible F_{ST} for a sample set (no alleles shared between populations) and subsequently dividing the observed F_{ST} by $F_{ST\,max}$.

Beside the summary statistics based analyses method, molecular genetic toolbox however encompasses more sophisticated analytical methods, in particular Maximum Likelihood and Bayesian approaches, which should be considered.

Bayesian inference of population structure

The drawback of F- and R-statistics for detecting population structure is that both approaches are summary statistics (F_{ST}: variances of allele frequencies, R_{ST}: variances of allele lengths) and a descision has to be made *a priori* about which specimens to treat as one population. Especially, when working on largely unknown marine species that were caught from largely unknown habitats such as Antarctic marine bottom sediments, it cannot be assumed that all organisms collected by bottom trawling over several kilometres belong to a single panmictic population. Although approaches exist that utilize the inbreeding coefficient F_{IS} to detect cryptic subpopulations (Goudet et al. 1994) there are now more powerful Bayesian approaches implemented to account for this problem. Bayesian inference methods utilize not only summary statistics but all genotypic information under various models (admixture/no admixture, allele frequencies correlated/not correlated; see Pritchard et al. 2000, Falush et al. 2003) to calculate the posterior probability of the data observed for a range of parameters such as the putative

number of populations *K* assuming Hardy-Weinberg equilibrium within populations (Pritchard et al. 2000, Falush et al. 2003, Evanno et al. 2005, Chen et al. 2007).

$$\ln \Pr(D \mid K)_{K1 \to n}$$

These estimates allow facilitate to detect even cryptic population structure and in addition they allow characterization of hybrids or migrants (Falush et al. 2003).

Haplotype networks phylogenetic trees and Nested-Clade Analysis

For sequence data of the mitochondrial DNA fragments investigated in this thesis, pairwise genetic distances were calculated within and between populations to assess the amount of sequence differences between populations, and to test for typical bimodal distributions in the pairwise differences indicating population structure. In addition, haplotype networks of intraspecific polymorphisms (Posada and Crandall 2001) and Neighbor-Joining trees (Saitou et al. 1987) were calculated to visualize the partitioning of intraspecific genetic polymorphisms. No attempt to estimate the true number of substitution events based on a substitution model was performed due to the low taxonomical level investigated. Models generally perform an upward correction of the substitutions, while at population levels may generally lead to much greater genetic distances due to in particular effects of genetic drift that bias the phylogenetic signal of sequence data at population level (Maddison and Knowles 2006). To distinguish between population structure and population history, Nested-Clade Analysis (Posada et al. 2006) was conducted based on mitochondrial DNA haplotype networks.

1.6 Population diversity

Several approaches exist to comparatively estimate the genetic diversity of a population. For multiallelic loci, the allelic diversity, i.e., the number of different alleles in a population was often used as a measure. However, this measure depends highly on sample size and in particular in Antarctic research, sample sizes vary greatly and are generally low. A better estimator for multiallelic data, which is not sample size dependent, is the amount of expected heterozygosity (H_e), also called gene diversity (Nei 1987). H_e refers to the chance that two randomly chosen allels from a population are different and is expressed by the formula

$$H_e = 1 - \Sigma p_i^2$$

where p_i is the frequency of the *i*th allele. Values range from 0 to 1. This measure was used to assess population genetic diversity of populations based on microsatellite data in this thesis. For sequence data, the haplotype diversity and pairwise nucleotide diversity were calculated according to Nei (1987).

1.7 Population demography (present and historical)

Effective population size (N_e) is a central concept in population genetics, introduced by Sewall Wright in 1931 (Wright 1931). While the census population size (N_c) refers to the total number of

individuals in a population, N_e is the number of individuals of a population that contribute to the gene pool of the next generation. Studies have shown that N_e is often several orders of magnitude smaller than N_c, in particular for marine species (Hedgecock 1994, Turner et al. 2002, Hutchinson et al. 2003). Small effective population size increases the rates of inbreeding, loss of genetic variation and fixation of slightly deleterious alleles (Nei et al. 1975, Maruyama and Fuerst 1985, Hartl and Clark 1997 for review). Thereby, small N_e reduces the evolutionary important adaptive potential of a species, increasing the probability of extinction (Vrijenhoek 1994, Frankham 1995, Lynch et al. 1995). The finding that abundant (in terms of N_c) species may have a very small effective counterpart (N_e), i.e. a low N_e/N_c ratio, has the counterintuitive consequence that even millions of individuals in a populations may not suffice to maintain genetic integrity. Thus, the effective counterpart of population size that determines evolutionary capacity of a population must be determined in this context.

Genetic methods are increasingly used to estimate present N_e. Several theoretical approaches toward estimating N_e have been proposed (see Wang 2005 for review). The most popular, temporal method is based on estimating N_e utilizing information on the rate of change in allele frequencies between samples of one population taken at two or more different points in time under the assumption that only genetic drift is responsible for the change in allele frequencies (Krimbas and Tsakas 1971, Nei and Tajima 1981, Waples 1989). The linkage disequilibrium (LD) method that utilizes information from one sample for estimating N_e has received little attention since its theoretical implementation by Hill (1981), primarily as it was found to be biased when sample size is low (England et al. 2006). However, Waples (2006) introduced a correction procedure that accounts for this bias, which now makes this method attractive in particular for studies where no temporal sampling is possible. This criterion for a method is in particular important for Antarctic studies as for different sampling sites in general there are no systematic temporal sampling schemes. This method is consequently used with the bias correction method (Waples 2006) to calculate present N_e for populations of the serolid species investigated in this thesis.

While estimates of contemporary N_e is of considerably interest to predict the populations future evolutionary capacity, evolutionary biologists are also interested in particular to make inferences on the past demographic history as this may often directly be linked to processes that shaped the present day distribution of the population (Hewitt et al. 2000). For allele frequency data it was shown that severe reductions in N_e lead to a much faster reduction of the number of alleles than in the expected heterozygosity (Nei et al. 1975, Maruyama and Fuerst 1985). Consequently, a low number or alleles and a significantly larger heterozygosity than expected for the number of alleles under equilibrium conditions indicate a recent bottleneck (Cornuet and Luikart 1996, Luikart et al. 1998). This method is applied in this study to test for recent bottlenecks in the populations of the serolid isopods investigated.

In addition, information from mitochondrial sequence data, which are less variable but have a fourfold smaller population size (Birkey et al. 1983) and may be therefore more susceptible to random genetic drift are also obtained by sequencing a fragment of the COI or 16S rDNA. Information on population demography from sequence data are assessed by calculating indices originally developed to detect genes that deviate from the expectations of neutrality, i.e. deviations from mutation-drift equilibrium: Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997). Tajima's D depends on the magnitude of differences between the number of differences between the number of segregating sites (S) and the average pairwise nucleotide differences (π).

Population expansions after bottleneck or founder events inflate S relative to π. Consequently, this statistic tends to be negative when there is an excess of recent mutations (or rare alleles). Fu's F_S statistic is based on the probability of finding a number of alleles greater or equal to the observed number in a sample drawn from a stationary population and is even more sensitive in detecting population expansions from smaller samples (Ramos-Onsins and Rozas 2002). In addition, pairwise mismatch analyses were performed by plotting pairwise distances among specimens and testing the distribution of the function against a function predicted by a sudden-expansion model (Rogers and Harpending 1992, Scheider and Excoffier 1999).

1.8 Objectives of this thesis

Despite a growing number of studies on phylogeny, biogeography and physiology of Antarctic benthic taxa, little is known about microevolutionary patterns. No study known to the author applied high resolving nuclear markers such as microsatellites or AFLPs to study population structure, gene flow, gene diversity and demographic processes in Antarctic benthic invertebrates to provide direct evidence for the realized dispersal capability of the low mobile brooding species or for evidence of refuge areas on the shelf.

The finding of broadly distributed brooding species along the Antarctic shelf and on remote Antarctic islands is largely based on morphological evidence. While for some species, e.g. *Ceratoserolis* n. sp. 1 (group 1 in Held 2003), one cryptic species of the *Ceratoserolis trilobitoides* (Eights, 1833) complex, molecular studies revealed that specimens from both sides of the Weddell Sea in fact belong to one species based on distance criteria defined by Held (2003), such empirical evidence is lacking for brooding species from the remote Antarctic Islands such as *Septemserolis septemcarinata* (Miers, 1875). In view of the strong isolation between the islands it seems highly implausible that gene flow between islands can be maintained. While in the Antarctic glaciations might have imposed a severe fragmentation of habitats the other case can be expected from the Magellan Region: Shallow water isopods of the species *Serolis paradoxa* (Fabricius, 1775) from the Patagonian coast and the Falkland Islands are today isolated by several hundreds of kilometres sea exceeding the species vertical distribution. However, during glacial maxima sea level was much lower and might thus have promoted gene flow in this region.

Central aspects of this thesis are:

1. *Development of a laboratory and computational strategy to isolate microsatellites from anonymous genomes with little effort in short time to make microsatellites more attractive for exploratory studies working with genetically relatively unknown species.* Target species:
    a. *Ceratoserolis* n. sp. 1 from the High Antarctic shelf
    b. *Septemserolis septemcarinata* from remote Antarctic and Subantarctic islands
    c. *Serolis paradoxa* from the Strait of Magellan and the Falkland Islands

2. *Analysis and discussion of patterns of genetic structure, gene flow, genetic diversity and demographic processes in these species in context of the climatological and geological history.* The questions are in particular:
    a. *Ceratoserolis* n. sp. 1

i. Do we find population genetic evidence that *Ceratoserolis* n. sp. 1 survived glacial advances in two independent refugia?
ii. Did the glacial advances lead to severe bottleneck events?
iii. How mobile are species?
iv. Are species further to the south genetically less diverse due to more severe environmental conditions?

b. *Septemserolis septemcarinata*
   i. Do populations on the highly remote islands represent populations of one species or are they morphologically indistinguishable, cryptic species with now overlap in allele spectra?
   ii. Are the small and young islands less diverse than old and large islands as would be expected based on the Island Biogeography Theory (MacArthur and Wilson 1967)?

c. *Serolis paradoxa*
   i. To what extent do populations from the Falkland Islands and the Strait of Magellan presently exchange genes?
   ii. Did the recurrent glaciations have an influence on population structure?
   iii. Did the recurrent glaciations that resulted in much lower sea levels allow dispersal of this species to maintain genetic integrity between regions?

1.9 Outline of this thesis

The thesis consists of this introduction (**chapter 1**), eight chapters corresponding to scientific publications and manuscripts on the development and analysis of molecular markers on serolid isopods (**chapters 2-9**) and a chapter that summarizes the results and points out their importance for present and future Antarctic research (**chapter 10**).
The eight manuscript chapters discuss the necessity of a microevolutionary framework to study Antarctic biodiversity (**Chapter 2**) and develop an innovative laboratory and analytical framework to isolate appropriate markers from anonymous genomes (**Chapter 3** and **Chapter 4**). The development of three highly informative marker sets for three serolid isopods is discussed in **chapter 5** and **chapter 6**. In **chapters 7-9**, the novel markers are used together with a mitochondrial marker to analyse the population genetic structure, diversity and the present and past demography of three serolid isopods: from the high Antarctic shelf (**chapter 7**), from remote Subantarctic Islands (**chapter 8**) and from the Strait of Magellan and the Falkland Islands (**chapter 9**).

References

Allcock, A. L., A.S. Brierley, J.P. Thorpe, and P.G. Rodhouse. 1997. Restricted gene flow and evolutionary divergence between geographically separated populations of the Antarctic octopus *Paraledone turqueti*. Marine Biology **129**: 97-102.

Anderson, J. B., S.S. Shipp, A.L. Lowe, J.S. Weller, and A.B. Mosola. 2002. The Antarctic ice sheet during the last glacial maximum and its subsequent retreat history: a review. Quaternary Science Reviews **21**: 49-70.

Arntz, W. E., J. Gutt, and M. Klages. 1997. Antarctic marine benthos: an overview, p. 3-14. *In* B. Battaglia, J. Valencia, and D.W. Walton [eds.], Antarctic communities: Species, structure and survival. Cambridge University Press, Cambridge.

Arntz, W. E., T. Brey, and V.A. Gallardo. 1994. Antarctic zoobenthos. Oceanography and Marine Biology: An Annual Review **32**: 241-304.

Aronson, R. B., and D.B. Blake. 2001. Climate change and benthic community structure in Antarctica. American Zoologist **41**: 27-39.

Avise, J. C. 1994. Molecular markers, natural history, and evolution. Chapman and Hall, New York.

Ballard, J. W., and M.C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology **13**: 729-744.

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

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

Barker, P. F., and E. Thomas. 2004. Origin, signature and palaeoclimatic influence of the Antarctic circumpolar current. Earth Science Reviews **66**: 143-162.

Barker, P. F., and J. Burrell. 1977. The opening of Drake Passage. Marine Georesources & Geotechnology **25**: 15-34 .

Barker, P. F., and J. Burrell. 1982. The influence upon Southern Ocean circulation, sedimentation, and climate of the opening of the Drake Passage, p. 377-385. *In* C. Craddock [ed.], Antarctic Geoscience. University of Wisconsin Press, Madison.

Barnes, D. K. A., and K.E. Conlan. 2007. Disturbance, colonization and development of Antarctic benthic communities. Philosophical Transactions of the Royal Society of London B Biological Sciences **362**: 11-38.

Barret, P. 2003. Cooling a continent. Nature **421**: 221-223.

Barron, J., B. Larson, and J.A. Baldauf. 1991. Evidence for Late Eocene to Early Oligocene Antarctic glaciation and observation on late Neogene glacial history of Antarctica: results from Leg 119. *In* J. Barron, B. Larson et al. [eds.], Proceedings of the Ocean Drilling Program: Scientific Results, **119**: 869-891.

Barton, N. H., and G.M. Hewitt. 1985. Analysis of hybrid zones. Annual Review of Ecology and Systematics **16**: 113-148.

Birky, C. W., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts and some results. Genetics **103**: 513-527.

Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. Quarterly Review of Biology **74**: 21-43.

Brandt, A. 1988. Antarctic Serolidae and Cirolanidae (Crustacea: Isopoda): New genera, new species and rediscriptions. Koeltz Scientific Books, Königstein.

Brandt, A. 1991. Colonization of the Antarctic shelf by the Isopoda (Crustacea, Malacostraca). Berichte zur Polarforschung **98**: 1-240.

Brey, T., C. Dahm, M. Gorny, M. Klages, M. Stiller, and W.E. Arntz. 1996. Do Antarctic benthic invertebrates show an extended level of eurybathy? Antarctic Science **8**: 3-6.

Chambers, G. K., and E.S. MacAvoy. 2000. Microsatellites: consensus and controversy. Comparative Biochemistry and Physiology B Comparative Biochemistry **126**: 455-476.

Chen, C., E. Durand, O. Forbes, and O. Francois. 2007. Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. Molecular Ecology Notes **7**: 747-756.

Clarke, A. 1979. On living in cold water: K-strategies in Antarctic benthos. Marine Biology **55**: 111-119.

Clarke, A. 1980. A reappraisal of the concept of metabolic cold adaptation in polar marine invertebrates. Biological Journal of the Linnean Society **14**: 77-92.

Clarke, A. 1982. Temperature and embryonic development in polar marine invertebrates. International Journal of Invertebrate Reproduction and Development **5**: 71-82.

Clarke, A. 1983. Life in cold water: The physiological ecology of polar marine ectotherms. Oceanography and Marine Biology: An Annual Review **21**: 341-453.

Clarke, A. 1991. What is cold adaptation and how should we measure it? American Zoologist **31**: 81-92.

Clarke, A. 1996. Marine benthic populations in Antarctica: patterns and processes. Antarctic Research Series **70**: 373-388.

Clarke, A., and A.J. Crame. 1989. The origin of the Southern Ocean marine fauna, p. 253-268. *In* A.J. Crame [ed.], Origin and evolution of the Antarctic biota. The Geological Society, London.

Clarke, A., and J.A. Crame. 1992. The Southern Ocean benthic fauna and climate change: a historical perspective. Philosophical Transactions of the Royal Society of London B Biological Sciences **338**: 299-309.

Clarke, A., and N.M. Johnston. 2003. Antarctic marine benthic diversity. Oceanography and Marine Biology: An Annual Review **41**: 47-114.

Clarke, A., D.K.A. Barnes, and D.A. Hodgson. 2005. How isolated is Antarctica? Trends in Ecology and Evolution **20**: 1-3.

Convey, P. 2001. Antarctic ecosystem, p. 171-184. *In* S.A. Levin [ed.], Encyclopedia of Ocean Sciences Biodiversity. Academic Press, San Diego.

Convey, P. 2003. Maritime Antarctic climate change: signals from terrestrial biology. Antarctic Research Series **79**: 145-158.

Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144**: 2001-2014.

Crame, J. A. 1994. Evolutionary history of Antarctica, p. 188-214. *In* G. Hempel [ed.], Antarctic science. Springer Verlag, Berlin.

Crame, J. A. 1997. An evolutionary framework for the polar regions. Journal of Biogeography **24**: 1-9.

Dayton, P. K. 1990. Polar benthos, p. 631-685. *In* W.O. Smith [ed.], Oceanography Part B: Chemistry, biology and geology. Academic Press, London.

DeConto, R. M., and D. Pollard. 2003. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO_2. Nature **421**: 245-249.

Dell, R. K. 1972. Antarctic benthos, p. 1-216. *In* F.S. Russell, and M. Yonge [eds.], Advances in marine biology. Academic Press.

Eastman, J. T., and A.R. McCune. 2000. Fishes on the Antarctic continental shelf: evolution of a marine species flock? Journal of Fish Biology **57**: 84-102.

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

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

Evans, J., J.A. Dowdeswell, and C. O Cofaigh. 2004. Late Quaternary submarine bedforms and ice-sheet flow in Gerlache Strait and on the adjacent continental shelf, Antarctic Peninsula. Journal of Quaternary Sciences **19**: 397-407.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**: 1567-1587.

Frankham, R. 1995. Conservation genetics. Annual Review of Genetics **29**: 305-327.

Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**: 915-923.

Goudet, J., DeMeeus, A.J. Day, and C.J. Gliddon. 1994. The different levels of population structuring of the dogwhelk, *Nucella lapillus*, along the south Devon coast, p. 81-95. *In* A.R. Beaumont [eds.], Genetics and Evolution of Aquatic Organisms. Chapman and Hall, London.

Gutt, J. 2001. On the direct impact of ice on macrobenthic communities, a review. Polar Biology **24**: 553-564.

Gutt, J., B.I. Sirenko, I.S. Smirnov, and W.E. Arntz. 2004. How many macrozoobenthic species might inhabit the Antarctic shelf? Antarctic Science **16**: 11-16.

Hain, S., and P.M. Arnaud. 1992. Notes on the reproduction of high-Antarctic molluscs from the Weddell Sea. Polar Biology **12**: 303-312.

Hartl, D. L., and A.G. Clarke. 1997. Principles of population genetics. Sinauer Associates, Sunderland MA.

Hedgecock, D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms? , p. 122-134. *In* A.R. Beaumont [ed.], Genetics and Evolution of Aquatic Organisms. Chapman and Hall, London.

Hedrick, P. 2005. A standardized genetic differentiation measure. Evolution **59**: 1633-1638.

Held, C. 2000. Phylogeny and biogeography of serolid isopods (Crustacea, Isopoda, Serolidae) and the use of ribosomal expansion segments in molecular systematics. Molecular Phylogenetics and Evolution **15**: 165-178.

Held, C. 2001. No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). Polar Biology **24**: 497-501.

Held, C. 2003. Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda), p. 135-139. *In* A.H. Huiskes, W.W. Gieskes, J. Rozema, R.M. Schorno, S.M. van der Vies, and W.J. Wolff [eds.], Antarctic biology in a global context. Backhuys Publishers, Leiden.

Held, C., and F. Leese. 2007. The utility of fast evolving molecular markers for studying speciation in the Antarctic benthos. Polar Biology **30**: 513-521.

Held, C., and J.-. Wägele. 2005. Cryptic speciation in the giant Antarctic isopod *Glyptonotus antarcticus* (Isopoda: Valvifera: Chaetiliidae). Scientia Marina **69**: 175-181.

Hewitt, G. M. 2000. The genetic legacy of the quaternary ice ages. Nature **405**: 907.

Higgins, K., and M. Lynch. 2001. Metapopulation extinction caused by mutation accumulation. Proceedings of the National Academy of Sciences of the USA **98**: 2928-2933.

Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. Genetical Research **38**: 209-216.

Hutchinson, W. F., C. van Oosterhout, S.I. Rogers, and G.R. Carvalho. 2003. Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proceedings of the Royal Society B Biological Sciences **270**: 2125-2132.

Huybrechts, P. 2002. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. Quaternary Science Reviews **21**: 203-231.

Imbrie, J., A. Berger, E.A. Boyle, S.C. Clemens, A. Duffy, W.R. Howard, G. Kukla, J. Kutzbach, D.G. Martinson, A. McIntyre, A.C. Mix, B. Molfino, J.J. Morley, L.C. Peterson, N.G. Pisias, W.L. Prell, M.E. Raymo, N.J. Shackleton, and J.R. Toggweiler. 1993. On the structure and origin of major glaciation cycles. 2. the 100,000-year cycle. Paleoceanography **8**: 699-735.

Ingolfsson, O. 2004. Quaternary glacial and climate history of Antarctica, p. 3-43. *In* J. Ehlers, and P.L. Gibbard [eds.], Quaternary Glaciations - Extent and Chronology, Part III. Elsevier, Amsterdam.

Janko, K., G. Lecointre, A. Devries, A. Couloux, C. Cruaud, and C. Marshall. 2007. Did glacial advances during the Pleistocene influence differently the demographic histories of benthic and pelagic Antarctic shelf fishes?--Inferences from intraspecific mitochondrial and nuclear DNA sequence diversity. BMC Evolutionary Biology **7**: 220.

Kennett, J. P. 1977. Cenozoic evolution of Antarctic glaciation, the Circum-Antarctic Ocean, and their impact on global palaeoceanography. Journal of Geophysical Research **82**: 3843-3860.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press.

Kimura, M., and J.F. Crow. 1964. The number of alleles that can be maintained in a finite population. Genetics **49**: 725-738.

Knowlton, N. 1993. Sibling species in the sea. Annual Review of Ecology and Systematics **24**: 189-216.

Knowlton, N. 2000. Molecular genetic analyses of species boundaries in the sea. Hydrobiologia **420**: 73-90.

Knox, G. A. 1994. The biology of the Southern Ocean. Cambridge University Press, Cambridge.

Knox, G. A., and J.K. Lowry. 1977. A comparison between the benthos of the Southern Ocean and the North Polar Ocean with special reference of the Amphipoda and the Polychaeta, p. 423-462. *In* M.J. Dunbar [ed.], Polar Oceans. Arctic Institute of North America, Calgary.

Krimbas, C. B., and S. Tsakas. 1971. The genetics of *Dacus oleae* V. Changes of esterase polymorphism in a natural population following insecticide control: selection or drift? Evolution **25**: 454-460.

Lanfear, R., J.A. Thomas, J.J. Welch, T. Brey, and L. Bromham. 2007. Metabolic rate does not calibrate the molecular clock. Proceedings of the National Academy of Sciences of the USA **104**: 15388-15393.

Lawver, L., L. Gahagan, and M. Coffin. 1992. The development of paleoseaways around Antarctica. , p. 7-30. *In* J. Kennett, and D. Warnke [eds.], The Antarctic Paleoenvironment: A perspective on global change. American Geophysical Union.

Lee, Y. H., M. Song, S. Lee, R. Leon, S.O. Godoy, and I. Canete. 2004. Molecular phylogeny and divergence time of the Antarctic sea urchin (*Sterechinus neumayeri*) in relation to the South American sea urchins. Antarctic Science **16**: 29-36.

Levinson, G., and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution **4**: 203-221.

Livermore, R., A. Nankivell, G. Eagles, and P. Morris. 2005. Paleogene opening of Drake Passage. Earth and Planetary Science Letters **236**: 459-470.

Luikart, G., F.W. Allendorf, J.M. Cornuet, and W.B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity **89**: 238-247.

Luxmoore, R. A. 1982. Moulting and growth in serolid isopods. Journal of Experimental Marine Biology Ecology **56**: 63-85.

Luxmoore, R. A. 1984. A comparison of the respiration rate of some Antarctic isopods with species from lower latitude. British Antarctic Survey Bulletin **62**: 53-65.

Lynch, M., J. Conery, and J. Burger. 1995. Mutation accumulation and the extinction of small populations. Genetics **146**: 489-518.

Maddison, W. P., and L.L. Knowles. 2006. Inferring phylogene despite incomplete lineage sorting. Systematic Biology **55**: 21-30.

Maruyama, T., and P.A. Fuerst. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics **111**: 675-689.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution **60**: 2399-2402.

Nauta, M. J., and F.J. Weissing. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics **143**: 1021-1032.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and F. Tajima. 1981. Genetic drift and estimation of effective population-size. Genetics **98**: 625-640.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution **29**: 1-10.

Otha, T., and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetical Research **22**: 201-204.

Page, T. J., and K. Linse. 2002. More evidence of speciation and dispersal across the Antarctic Polar Front through molecular systematics of Southern Ocean *Limatula* (Bivalvia: Limidae). Polar Biology **25**: 818-826.

Palumbi, S. R. 1994. Genetic divergence, reproductive isolation and marine speciation. Annual Review of Ecology and Systematics **25**: 547-572.

Picken, G. B. 1980. Reproductive adaptations of Antarctic benthic invertebrates. Biological Journal of the Linnean Society **14**: 67-75.

Posada, D., and K.A. Crandall. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology and Evolution **16**: 37-45.

Posada, D., K.A. Crandall, and A.R. Templeton. 2006. Nested clade analysis statistics. Molecular Ecology Notes **6**: 590-593.

Poulin, E., A.T. Palma, G. Leiva, E. Hernandez, P. Martinez, S.A. Navarrete, and J.C. Castilla. 2002. Temporal and spatial variation in the distribution of epineustonic competent larvae of *Concholepas concholepas* along the central coast of Chile. Marine Ecology-Progress Series **229**: 95-104.

Poulin, E., and J.P. Feral. 1996. Why are there so many species of brooding Antarctic echinoids? Evolution **50**: 820-830.

Pörtner, H.-O., L.S. Peck, and G.N. Somero. 2007. Thermal limits and adaptation in marine Antarctic ectotherms: an integrative view. Philosophical Transactions of the Royal Society of London B Biological Sciences **362**: 2233-2258.

Pörtner, H.-O., and R.C. Playle. 1998. Cold ocean physiology. Cambridge University Press, Cambridge.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**: 945-959.

Ramos-Onsins, S. E., and J. Rozas. 2002. Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution **19**: 2092-2100.

Rintoul, S. R., S.R. Hughes, and D. Olbers. 2001. The Antarctic Circumpolar Current system, p. 271-302. *In* G. Siedler, J. Church, and J. Gould [eds.], Ocean circulation and climate. Academic Press, New York.

Rogers, A. R., and H.C. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution **9**: 552-569.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambrige University Press, Cambridge.

Rousset, F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics **142**: 1357-1362.

Royer, D. L., R.A. Berner, I.P. Montanez, N.J. Tabor, and D.J. Beerling. 2004. CO_2 as a primary driver of Phanerozoic climate. GSA Today **14**: 4-10.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4** : 406.

Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma **109**: 365-371.

Schlötterer, C. 2004. The evolution of molecular markers - just a matter of fashion. Nature Reviews Genetics **5**: 63-69.

Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics **152**: 1079-1089.

Selkoe, K., and R.J. Toonen. 2006. Microsatellites for Ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters **6**: 615-629.

Shaviv, N. J., and J. Veizer. 2003. Celestial driver of Phanerozoic climate? GSA Today **13**: 4-10.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. Annals of the Entomological Society of America **87**: 651-701.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**: 457-462.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology and Evolution **15**: 199.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585-595.

Thatje, S., C.-D. Hillenbrand, A. Mackensen, and R. Larter. 2008. Life hung by a thread: Endurance of Antarctic fauna in glacial periods. Ecology **89**: 682-692.

Thatje, S., C.D. Hillenbrand, and R. Larter. 2005. On the origin of Antarctic marine benthic community structure. Trends in Ecology and Evolution **20**: 534-540.

Turner, T. F., J.P. Wares, and J.R. Gold. 2002. Genetic effective size is three orders of magnitude smaller than adults census size in an abundant, estuarine-dependant marine fish (*Scianops ocellatus*). Genetics **162**: 1329-1339.

Veizer, J., Y. Godderis, and L.M. François. 2000. Evidence for decoupling of atmospheric CO_2 and global climate during the Phanerozoic eon. Nature **408**: 698-701.

Vrijenhoek, R. C. 1994. Genetic diversity and fitness in small populations, p. 37-53. *In* J. Loeschcke, S.K. Jain, and J. Tomiuk [eds.], Conservation Genetics. Birkhäuser Verlag, Basel.

Wang, J. 2005. Estimation of effective population sizes from data on genetic markers. Philosophical Transactions of the Royal Society of London B Biological Sciences **360**: 1395-1409.

Waples, R. S. 1989. A generalised approach for estimating effective population size from temporal changes in allele frequency. Genetics **121**: 379-391.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics **7**: 167-184.

Waples, R. S., and O. Gaggiotti. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology **15**: 1419-1439.

Wägele, J.-W. 1986. Polymorphism and distribution of *Ceratoserolis trilobitoides* (Eights, 1833) (Crustacea, Isopoda) in the Weddell Sea and synonymy with *C. cornuta* (Studer, 1879). Polar Biology **6**: 127-137.

Wägele, J.-W. 1987. On the reproductive biology of *Ceratoserolis trilobitoides* (Crustacea: Isopoda): Latitudinal variation of fecundity and embryonic development. Polar Biology **7**: 11-24.

Wägele, J.-W. 1994. Notes on Antarctic and South American Serolidae (Crustacea, Isopoda) with remarks on the phylogenetic biogeography and a description of new genera. Zoologische Jahrbücher Abteilung für Systematik Ökologie und Geographie der Tiere **121**: 3-69.

Wägele, J.-W. 2001. Grundlagen der phylogenetischen Systematik. Dr. Friedrich Pfeil, München.

Weber, J. L., and C. Wong. 1993. Mutation of human short tandem repeats. Human Molecular Genetics **2**: 1123-1128.

Weir, B. S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**: 1358-1370.

White, M. G. 1977. Ecological adaptations by Antarctic poikilotherms to the polar marine environment, p. 197-208. *In* G.A. Llano [ed.], Adaptations within antarctic ecosystems. Gulf Publ. Co., Houston.

Wright, S. 1931. Evolution of Mendelian populations. Genetics **16**: 167-176.

Wright, S. 1951. The genetical structure of populations. Annual of Eugenics **15**: 323.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution **19**: 395-420.

Chapter 2

The utility of fast evolving molecular markers for studying speciation in the Antarctic benthos

Christoph Held and Florian Leese*

Polar Biology, 2007

*(both authors contributed equally to the paper)

Abstract

The Southern Ocean is surprisingly rich in species that coexist in one of the most extreme environments on Earth yet the processes leading to speciation in this ecosystem are not well understood. To remedy this, tools that measure the genetic connectedness within a species are needed. Although useful for phylogenetic purposes, the readily available mitochondrial markers (e.g. 16S, COI) suffer from numerous shortcomings for population genetics. Therefore, molecular markers are needed that are sufficiently variable, unlinked, bi-parentally inherited, and distributed over the whole genome. We argue that microsatellites are suitable markers that have not been widely used in exploratory studies due to their difficult initial set-up. Working with the *Ceratoserolis trilobitoides* species complex (Isopoda), we demonstrate that using a novel protocol many microsatellites can be identified quickly. An increased availability of these highly sensitive markers will be useful for studies addressing the origin of species in the Southern Ocean and their response to future climate change.

Introduction

The fauna of the Southern Ocean differs substantially from those of all other oceans due to extreme environmental conditions (Aronson and Blake 2001, Clarke 1983, Peck 2002). Despite these, an extraordinarily high species diversity within the zoobenthos is now well established (Gray 2001) and molecular methods continue to add to our knowledge about the number of species and their phylogenies. The identification of species is particularly difficult in the Southern Ocean and, perhaps not surprisingly, molecular data have identified cryptic species in many taxa (Bernardi and Goswami 1997; Held 2003; Held and Wägele 2005; Raupach and Wägele 2006). It is unclear at this point whether this represents the Antarctic share of a more general phenomenon that currently receives attention in the scientific community (cryptic species and DNA barcoding; see Hebert et al. 2003; Tautz et al. 2003) or whether cryptic speciation in the Southern Ocean is driven mainly by processes unique to the Antarctic.

Until now, the taxa receiving most attention from molecular biologists have been fishes (Bargelloni et al. 2000, Lecointre et al. 1997, Near et al. 2003, Near et al. 2004, Ritchie et al. 1997), molluscs (Page and Linse 2002), echinoderms (Lee et al. 2004), and crustaceans (Baltzer et al. 2000, Clarke et al. 2004, Held 2000, Held 2003, Held and Wägele 2005, Jarman 2001, Jarman et al. 2000, Jarman et al. 2002, Loerz and Held 2004, Patarnello et al. 1996, Raupach et al. 2004, Zane et al. 1998, Zane and Patarnello 2000). Although primarily aimed at resolving phylogenetic questions, interpreting the molecular trees in the light of additional sources of information (mainly biogeography and geology) has allowed answering questions broader in scope concerning colonization of habitats, origins of distributions, evolution of adaptations, and speed of molecular substitution on an evolutionary timescale (Held 2001, Near 2004, Stankovic et al. 2002). This phylogenetically motivated approach has proven fruitful and we continue to see a place for it in the future.

However, while our knowledge on the diversity and distribution of modern species in the Southern Ocean continues to grow, our understanding of what led to the high number of species in the first place is less comprehensive. Typical adaptations in life cycle of Antarctic taxa (e.g. few descendants, no pelagic distributional stages, and high immobility of adults) are assumed to decrease gene flow thus enhancing the possibility of speciation. On the other hand, the consequences of frequent periodical glaciations of major parts of the Antarctic continent can counteract this by eradicating most of the benthic communities through large-scale spread of grounded ice (Thatje et al. 2005). This could also have structured genetic diversity by lineage sorting in the isolated, ice-free refugia.

In order to understand which of the above mentioned processes or adaptations may have led to the disruption of a previously contiguous gene pool and ultimately led to the generation of new species in the evolutionary past, a more comprehensive knowledge of what is structuring the present-day distribution of genetic polymorphisms within a species becomes necessary. Molecular methods and population genetics in particular can provide a valuable contribution to building a microevolutionary framework that will enable us to finally measure population interconnections.

Markers with higher rates of change than single locus coding genes (including the mitochondrial genome) are required to measure population genetic parameters such as gene flow through migration and dispersal. Despite the fact that the vast majority of the Antarctic ecosystem is composed of benthic invertebrate animals (Gutt et al. 2004), such markers are only available

for a handful of vertebrates and pelagic invertebrates in the Southern Ocean (Hoelzel et al. 2002, Reilly and Ward 1999, Roeder et al. 2001, Shaw et al. 2004, Valsecchi et al. 1997). To our knowledge, the Antarctic benthos is completely unstudied in this respect.

The aim of this study is twofold. On the one hand, we would like to emphasize the emerging necessity to develop more fine-scaled molecular markers in an attempt to understand the genesis of the surprising diversity of species in the Antarctic benthos. On the other hand, we demonstrate that even in almost unknown genomes suitable molecular markers, which are sufficiently variable to allow for microevolutionary studies of intraspecific patterns of DNA polymorphism and gene flow, can be isolated in a standard molecular lab in a relatively short time. Our study object is a new species found in the isopoda genus *Ceratoserolis* Brandt 1988, a taxon serving as a model for the Antarctic zoobenthos.

Material and methods

Here we outline the strategy of obtaining microsatellite markers from a relatively unknown genome (in this case from the *Ceratoserolis trilobitoides* species complex).

In short, the genome of the species under study (target genome) is cut into shorter fragments. In order to identify the fragments containing short tandem repeats (microsatellites), they are hybridized against immobilized genomic fragments of an unrelated species (reporter genome). The stringency of the subsequent washing steps is such that only fragments with higher similarity than what can be expected by chance are retained. Since the two genomes show little similarity by descent, the greatest similarity between them occurs often between regions of low complexity such as microsatellites. The retained fragments of the target genome will therefore likely be enriched in microsatellites.

For more technical details and an evaluation of the properties of this and other protocols for the isolation of microsatellites, the reader is referred to another paper (Leese et al. accepted) and the original papers describing the method (Nolte et al. 2005, Zane et al. 2002).

Sampling

Specimens belonging to the *C. trilobitoides* complex were collected by hand-sorting bottom trawled gear on three expeditions to the Antarctic: ANT XIII/3 and ANT XIV/2 in 1996 with RV Polarstern, and ICEFISH 2004 with RV Nathaniel B. Palmer. Specimens were preserved in cold 80-96% ethanol shortly after sorting the catch. Genomic DNA was extracted with the Qiagen DNeasy Kit according to the manufacturer's protocol. For details concerning preservation and DNA extraction methods see Held (2000).

Molecular genetic techniques

Three different hybridization filters with single stranded reporter genomes of *Mus musculus domesticus*, *Drosophila melanogaster*, and *Homo sapiens* were prepared according to Nolte et al. (2005). DNA from 18 specimens from six populations of the *C. trilobitoides* species complex was pooled. Digestion of this template (1 µg) and ligation to the AFLP adaptors (5'-TACTCAGGACTCAT-3'/5'-GACGATGAGTCCTGAG-3') were carried out simultaneously with an

excess of 10–20 U of the *MseI* isoschizomer *TruI* and 10–30 U T4-DNA ligase in a 1× Buffer-R reaction mix (100 µl). Both enzymes have different optimal working temperatures (65 and 22°C). In accordance with the suggestions of the supplier, the reaction was incubated at an intermediate temperature (37°C) for 6 h. Genomic fragments were subjected to a 10-min incubation at 72°C in a complete PCR mix without primers (1× *Taq* Buffer (1.5 mM Mg2+), 0.2 mM dNTPs, 2 U *Taq* Polymerase, all products Eppendorf) to eliminate nicks remaining from the adapter ligation. Of this mixture, 5 µl was PCR amplified in a total volume of 50 µl with the AFLP adaptor-specific primer MseI-N (5'-GATGAGTCCTGAGTAAN-3') in a Techne thermocycler (94°C 4 min, 25 cycles of 94°C 30 s, 53°C 1 min, 72°C 1 min, and a final elongation step of 72°C for 7 min). Fragments from this preamplification were size-selected (400–800 bp) on a 1.5% agarose gel. Reactions were purified from the gel using the Eppendorf Perfectprep Gel Cleanup Kit. To eliminate shorter fragments trapped in the gel matrix, electrophoresis and subsequent purification of the PCR products from the gel were repeated twice.

Amplified fragments containing microsatellites were selectively enriched by hybridizing 300–500 ng against immobilized single-stranded DNA fragments of the reporter genomes bound to nylon membranes as described in Nolte et al. (2005). Three washing steps were performed to keep only those fragments of the target genome that formed stable duplexes with the immobilized reporter genome, indicating high similarity in low-complexity regions.

Thereafter, the hybridized fragments were eluted using 30 µl of TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 7.5).

Enriched fragments were amplified again as above in 25 µl reaction volumes to obtain a sufficient concentration for cloning (10–100 ng) and purified from the gel as above. The purified enriched fragments were now cloned into a pCR®2.1 TOPO® TA vector (Invitrogen), and transformed into competent E. coli (Promega JM 109, TOP10F').

Initially, the colonies were PCR screened for microsatellites using a combination of primers located in the multiple cloning site of the vector and a third primer anchored in the repeat sequence of the insert, if present (Lunt et al. 1999). In subsequent runs, this confirmation step was omitted since the majority of the inserts (approximately 90%) contained microsatellites so that all inserts were sequenced without further screening. Sequencing using M13 forward and reverse vector primers was either conducted in-house on a LiCor 4200 or ABI 3130xl automated sequencer, or outsourced to Macrogen (Seoul, Korea). To improve signal quality for sequences containing long microsatellites, DMSO was added to a final concentration of 5%. BLAST searches were carried out for all sequences to exclude the possibility of contamination with mobilized fragments from the reporter genome.

Scoring criteria

Microsatellites with repeat units ranging from 2 – 6 nucleotides in length were identified in the target genome using an improved version of the repeat finding program Sputnik (Abajian 1994, Morgante et al. 2002) with modifications (see Leese et al. accepted for details). We rejected all microsatellites with less than 90% identity and shorter than ten nucleotides (e.g. four repeat units of a dinucleotide repeat). Microsatellites with internal imperfections were thus scored as two independent loci of the same type if they were separated by a stretch of nucleotides if its inclusion had led to rejection of the whole section. All other settings were according to the recommendations given by Chambers and MacAvoy (2000).

Results

In total, 371 colonies with inserts enriched for microsatellites using the cross-hybridization technique were screened. Among these, two colonies contained redundant fragments (i.e. identical to other inserts) and were excluded from analysis. Of the remaining 369 colonies, 326 colonies contained inserts with at least one microsatellite (88.3% positives). The remaining 43 colonies either contained no microsatellites or only repeats shorter than our scoring threshold outlined above (11.7% negatives). According to these criteria, a total of 781 microsatellites were detected within the 326 positive inserts (Ø 2.4 ± 1.6). A maximum number of ten independent microsatellites were detected within a single 652 bp insert.

Among the different repeat families, dinucleotide repeats were the most frequent (66.6%), followed by tetranucleotide (16.2%), pentanucleotide (11.4%), hexanucleotide (2.9%), and trinucleotide repeats (2.8%). As many as 54 different repeat types were found in total. Dinucleotide microsatellites were typically composed of more repeat units per microsatellite than microsatellites with longer repeat units (tri- to hexanucleotide repeats), although length frequency distribution varies between repeat motifs of identical unit length, too (Table 1).

The amount of perfect microsatellites, characterized by arrays of a certain repeat unit that are not interrupted by nucleotide substitutions, was 51.5%.

Table 1: Types, total number and relative frequency of microsatellites detected within the genome of *Ceratoserolis trilobitoides sensu lato* screening 369 non-redundant inserts. The *Ceratoserolis* genome was hybridized against immobilized genome fragments of *Mus musculus, Drosophila melanogaster*, and *Homo sapiens* according to the reporter genome protocol by Nolte et al. (2005). Further details in Leese et al. accepted.

Type of microsatellite	Frequency of microsatellite types		Average number of repeats (mean ± SD)	Number of different motifs
	[n]	[% of total]		
Dinucleotide	484	67.9	25.37 ± 27.11	4
Trinucleotide	21	3.0	8.05 ± 9.00	6
Tetranucleotide	122	17.1	11.07 ± 15.55	16
Pentanucleotide	66	9.3	11.85 ± 16.34	13
Hexanucleotide	20	2.8	7.35 ± 7.07	13

Discussion

Identification of microsatellites is a constant matter of discussion since no consensus has been reached about what a microsatellite actually is (Chambers and MacAvoy 2000, Ellegren 2004). Consequently, the absolute number and length distribution of microsatellite arrays vary with the scoring criteria used to detect microsatellites. This topic is further discussed elsewhere (Leese et al. accepted).

Shortcomings of mitochondrial genes as molecular markers in studies of speciation and gene flow

Traditionally, molecular phylogenies and interpretations derived from these trees including phylogeography (Avise 2000) have relied extensively on relatively conserved genes because these were easiest to target because even in unknown genomes conserved regions that allow the construction of the so-called universal primers can be relied upon (Simon et al. 1994). However, a corollary of the requirement for conserved stretches of nucleotides serving as universally applicable priming sites is that there is both an upper limit for the variability of the molecular markers located between them as well as a limit in the number available. In addition, the most frequently used genes in the mitochondrial genome suffer from additional limitations. In summary, the utility of mtDNA for population genetic purposes is limited because it is short, all of its genes behave as a single, linked locus due to the lack of recombination, it cannot *a priori* be assumed to evolve strictly neutral, it is generally maternally inherited, effective population size for mtDNA is only one-quarter of nuclear DNA, and its rate of change is limited (Ballard and Whitlock 2004). This in turn limits the application of genes located in the mitochondrial genome approximately to the level of species and the phylogenetic relations between them (Figure 1).

Populations, gene flow and speciation

In evolutionary biology the term "population" is loosely used to describe a geographical region of origin of individuals of a species, or more specifically, non-random, variable patterns of genetic connectivity inside a species (Waples and Gaggiotti 2006). According to the latter, two individuals are likely to be more similar in some respect if they belong to the same population and more different if they belong to different populations, either as a result of adaptation to similar environments (selection) or as the result of stronger isolation (genetic drift) between them. It is noteworthy that even if populations are defined on a geographical basis, only the degree of (genetic) isolation between them is often implied. Either way, populations that are sufficiently isolated from one another have the potential to accumulate local traits (morphological, behavioural, physiological or genetic) that are different from the ones found in other populations (Coyne and Orr 1998).

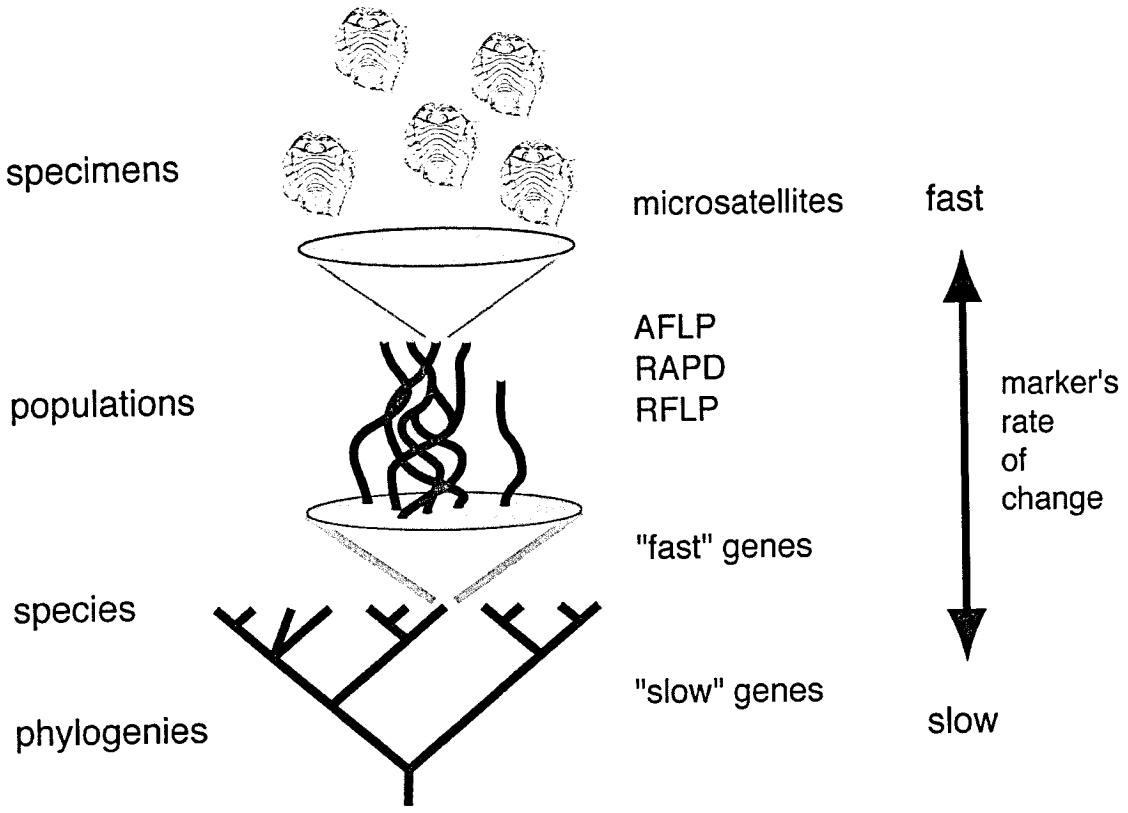


Figure 1: Different types of molecular markers possess different rates of change and hence resolve different time windows. While slower genes are required to resolve the deeper bifurcations in a phylogenetic tree, faster and more finely resolving markers are better suited for more recent events such as identification of species. In order to study the level of populations or individuals, which are necessary to understand speciation, molecular markers with rates of change that exceed the rates of coding genes are needed. Each class of molecular markers covers a range of substitution rates and may differ between loci. The order at which the classes of markers appear in the right half of this figure should be taken as an approximation.

Most species concepts postulate that for speciation to occur the once contiguous gene pool must be disrupted, either as a primary requirement or as a consequence. The formation of species can thus be interpreted as the accumulation of differences beyond a stage that ensures reproduction and the continuation of a common gene pool for all populations.

However, high levels of gene flow between populations synchronize genetically mediated traits between them and consequently oppose the process of differentiation (drift or selection) and ultimately speciation. Therefore, the amount of gene flow within a species is a critically important baseline information in the light of which all scenarios concerning the evolution of this and closely allied species must be discussed.

Excluding false negatives

Although some of the faster mitochondrial loci (e.g. 16S rRNA, COI, mitochondrial control region) do conserve intraspecific genetic polymorphisms which under favourable conditions can be interpreted in the context of population genetics (e.g. Roman and Palumbi; Zane et al. 1998),

these loci are mostly informative over larger distances (e.g. intercontinental). The main limitation of loci with limited substitution rates becomes obvious when dealing with cases of perceived homogeneity between samples. Homogeneity can be the result of two fundamentally different reasons. Either two populations are indeed genetically indistinguishable as a result of high levels of gene flow between them or short times since their separation. On the other hand, they may just appear so because the marker's rate of change is inadequate to detect existing genetic differences (false negative: no evidence for differentiation when in reality there is). The impossibility of proving homogeneity is a fundamental limitation and not a property of a particular class of molecular markers. It is advisable to interpret results indicating homogeneity between samples with caution. The absence of a proof (for differentiation between populations) is no proof for the absence of such differentiation for the reasons outlined above. By analysing many independent loci covering a broad range of substitution rates (including very high ones), however, the risk of falling victim to false negatives can be greatly reduced.

Fast evolving markers to the rescue

A better marker for population level studies would consist of many, unlinked, neutral loci distributed across the genome. This would avoid the potentially misleading effects of mitochondrial genes. It would also ideally consist of loci with different levels (including very high levels) of substitution rates (Gaffney 2000).

There are several classes of markers that come close to these requirements (see Figure 1). Random Amplified Polymorphic DNA or RAPDs have lost much of their appeal because it has been shown that the method is unreliable and has a tendency to produce artefactual bands not inherited from the parents (Ellsworth et al. 1993, Perez et al. 1998). Restriction Fragment Length Polymorphism (RFLP) is a more reproducible method, but when high rates of change are desirable yields result that are not easy to interpret due to the multitude of bands. If homologizing beyond mere electrophoretic mobility is desirable or necessary, excising and sequencing or hybridising of bands becomes necessary which offsets the ease of use of RFLP.

Amplified Fragment Length Polymorphism (AFLP) and Microsatellites are perhaps closest to the ideal marker for population studies (Bensch and Akesson 2005). AFLPs tend to be easier to set up initially whereas microsatellites as co-dominant, multiallelic markers have more analytical power in certain contexts (information per locus 4 – 10x higher). A further disadvantage of the AFLP technique is that reproducibility of AFLP patterns can be strongly biased by DNA quality and other methodological parameters (Bensch and Akesson 2005). Other fast evolving markers with genome-wide distribution have become available recently (SNPs, microarrays). A detailed comparison with microsatellites is beyond the scope of this article, however, they do not offer advantages over microsatellites in terms of analytical power and especially the effort required for initial set-up (Schlötterer 2004). We are therefore only considering microsatellites (Figure 2) for the remainder of this paper.

In contrast to the universal priming sites that function across wide taxonomic groups, microsatellites are not usually flanked by highly conserved regions so that in general microsatellites have to be isolated for every species of interest *de novo*. Since there can be a significant difference in the frequency and types of repeats between the genomes of different species which in ecological and systematic studies are generally only poorly known, isolating microsatellites can be a tedious and time-consuming process.

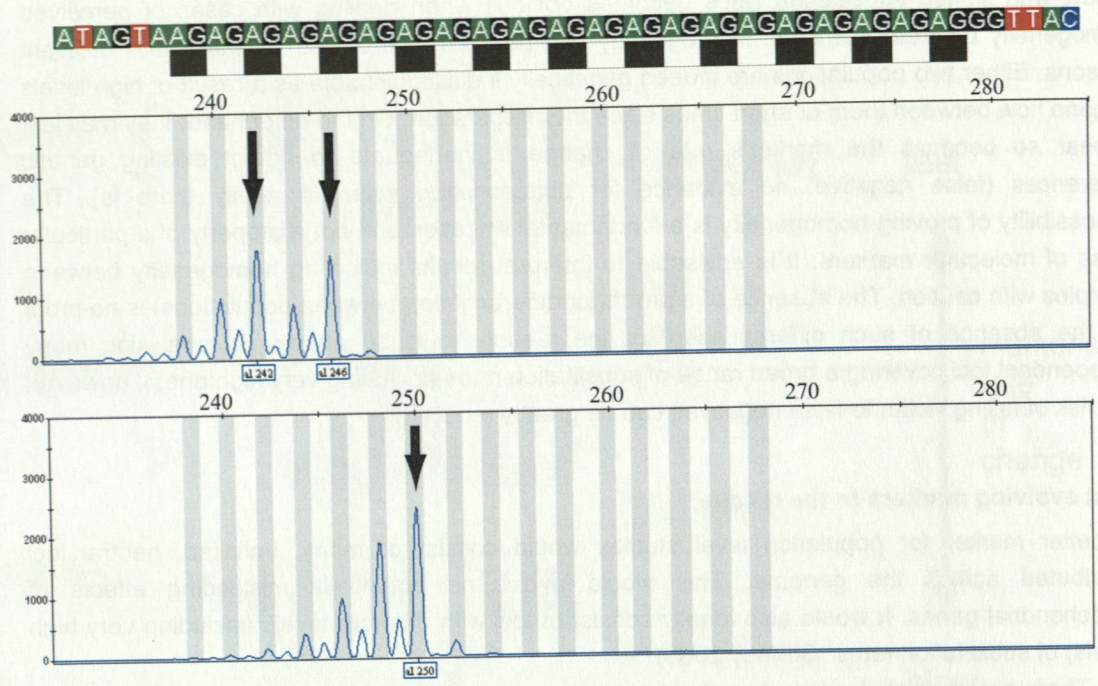


Figure 2: A microsatellite locus from the genome of *Ceratoserolis trilobitoides sensu lato*. Top: Nucleotide sequence of an $(AG)_n$ dinucleotide microsatellite with 21 repeat units, each repeat unit composed of two nucleotides is represented by a black or white box. Bottom: Fragment analysis showing a heterozygote specimen with two alleles of different length (242 and 246 bp) and a homozygote with a single allele (250 bp) (black arrows). The less intense peaks indicate fragments that are a result of in vitro artefacts (slippage) during amplification and are not scored.

Although the mutational processes of microsatellite arrays are still not fully understood and selection on some repetitive motifs (e.g. if located within a coding region or upstream the promoter region) has been reported (Kashi and Soller 1999, Li et al. 2002), the majority of microsatellite loci are assumed to be selectively neutral. In contrast to the mitochondrial genes that act as a single linked locus, the null hypothesis is that microsatellites are unlinked and act as independent loci.

In summary, microsatellites are in almost perfect agreement with the requirements of a better marker for fine-scaled studies at the population level. The initial effort required for their initial set-up is perhaps the main reason why microsatellites have not been used more widely. In that sense, there is a trade-off between ease of use, reliability, and analytical power of molecular markers.

Simplified ways to obtain suitably fast markers

A variety of protocols exist for the identification and isolation of microsatellite loci. These differ markedly in their requirements (time and equipment) as well as their efficiency (Leese et al. accepted, Zane et al. 2002).

The case study presented in this paper demonstrates that it has become now possible to isolate a large number of microsatellite loci from an unknown genome with reasonable effort (Leese et al. accepted). The use of a modified protocol (Nolte et al. 2005) shows that many obstacles previously associated with the isolation of microsatellites may no longer be a problem.

One of the main advantages of the reporter genome protocol is that no assumptions have to be made concerning the type of repeats to be screened. While it is entirely possible in traditional protocols to miss even abundant repeat types by not including the corresponding reporter oligomer in the screening, the reporter genome protocol will report all repeat types that are present in both reporter and target genome, a most valuable property when dealing with unknown genomes (Leese et al. 2008, accepted).

Another welcome property of microsatellites as molecular markers is that they typically require the amplification of short (100-250 bp) fragments only. This facilitates amplification of partly degraded material not collected and preserved with molecular studies in mind. The potential inclusion of museum material fixed in formalin would greatly alleviate one of the main problems of working in the Antarctic: the notoriously difficult and unpredictable acquisition of samples.

Conclusions

Fast evolving molecular markers such as microsatellites allow a far more detailed look on the processes at the population level than more conserved markers. Analyses of microevolutionary processes structuring the Antarctic benthos are needed for a better understanding of how locally characteristic traits evolve in a species in this unique environment. This opens new perspectives on the origin of biodiversity and the evolution of adaptations to the environment in the Southern Ocean. Collecting data on fast evolving loci in the Antarctic benthos is also a prerequisite for detecting a response of Antarctic benthic communities to future environmental change.

We provide evidence from one Antarctic crustacean that a new isolation protocol has brought suitably fast evolving microsatellite markers into reach for a standard molecular lab working on an unknown genome within a short time.

Acknowledgments: Arne Nolte (University of Cologne) provided his microsatellite protocol ahead of print and offered valuable assistance during the initial test phase in the lab. For computational assistance and the development of the powerful computer programs "Phobos" and "ssr-stat" we thank Christoph Mayer (University of Bochum). Wolfgang Wägele (Museum Koenig, Bonn) discussed various aspects of the project with us. We also thank Joseph Eastman, Sven Thatje, and one anonymous reviewer for valuable comments and suggestions on the manuscript. This work was supported by the DFG grant He 3391/3 to Christoph Held (AWI Bremerhaven) and in part by NSF grant OPP 01-32032 to H. William Detrich (Northeastern University, Boston, USA). This is publication number 11 from the ICEFISH Cruise of 2004.

References

Abajian C (1994) Sputnik beta version 29-juli-1994. http://espressosoftware.com/pages/sputnik.jsp

Aronson RB, Blake DB (2001) Global climate change and the origin of modern benthic communities in Antarctica. Am Zool 41: 27-39

Avise JC (2000) Phylogeography. The history and formation of species. Harvard University Press, Cambridge

Ballard JW, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol Ecol 13: 729-744

Baltzer C, Held C, Waegele J-W (2000) *Furcarcturus polarsterni* gen. nov., sp. nov., a large deep-sea arcturid isopod from the Drake Passage, with a preliminary molecular characterization. Polar Biol 23: 833-839

Bargelloni L, Zane L, Derome N, Lecointre G, Patarnello T (2000) Molecular zoogeography of Antarctic euphausiids and notothenioids: from species phylogenies to intraspecific patterns of genetic variation. Antarct Sci 12: 259-268

Bensch S, Akesson M (2005) Ten years of AFLP in ecology and evolution: why so few animals? Mol Ecol 14: 2899-2914

Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Physiol B Biochem Mol Biol 126: 455-476

Clarke A (1983) Life in cold water: The physiological ecology of polar marine ectotherms. Oceanography and Marine Biology: An Annual Review 21: 241-453

Clarke A, Aronson RB, Crame JA, Gil JM, Blake DB (2004) Evolution and diversity of the benthic fauna of the Southern Ocean continental shelf. Antarct Sci 16: 559-568

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. Philos Trans R Soc Lond B Biol Sci 353: 287-305

Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. Nat Rev Genet 5: 435-445

Ellsworth LE, Rittenhouse KD, Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. BioTechniques 14: 214-217

Gaffney PM (2000) Molecular tools for understanding population structure in Antarctic species. Antarct Sci 12: 288-296

Gray JS (2001) Antarctic marine benthic biodiversity in a world-wide latitudinal context. Polar Biol 24: 633-641

Gutt J, Sirenko BI, Smirnov IS, Arntz WE (2004) How many macrozoobenthic species might inhabit the Antarctic shelf? Antarct Sci 16: 11-16

Held C (2000) Phylogeny and biogeography of serolid isopods (Crustacea, Isopoda, Serolidae) and the use of ribosomal expansion segments in molecular systematics. Mol Phylogenet Evol 15: 165-178

Held C (2001) No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). Polar Biol 24: 497-501

Held C (2003) Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda)

Antarctic biology in a global context. In: Huiskes AHL, WWC Gieskes, J Rozema, RML Schorno, SM van der Vies and WJ Wolff (eds). Backhuys Publishers, Leiden, pp 135-139

Held C, Wägele J-W (2005) Cryptic speciation in the giant Antarctic isopod *Glyptonotus antarcticus* (Isopoda: Valvifera: Chaetiliidae). Sci Mar 69: 175-181

Hoelzel AR, Natoli A, Dahlheim ME, Olavarria C, Baird RW, Black NA (2002) Low worldwide genetic diversity in the killer whale (*Orcinus orca*): implications for demographic history. Proc R Soc Lond B Biol Sci 269: 1467

Jarman SN (2001) The evolutionary history of krill inferred from nuclear large subunit rDNA sequence analysis. Biol J Linn Soc 73: 199-212

Jarman SN, Elliott NG, Nicol S, McMinn A (2000) Molecular phylogenetics of circumglobal *Euphausia* species (Euphausiacea: Crustacea). Can J Fish Aquat Sci 57: 51-58

Jarman SN, Elliott NG, Nicol S, McMinn A (2002) Genetic differentiation in the Antarctic coastal krill *Euphausia crystallorophias*. Heredity 88: 280-287

Jarman SN, Nicol S (2002) Sources of variance in studies of krill population genetics. Ccamlr Science 9: 107-116

Kashi Y, Soller M (1999) Functional role of microsatellites and minisatellites. In: Goldstein DB and C Schlötterer (eds) Microsatellites. Evolution and applications. Oxford University Press, New York, pp 10-23

Lecointre G, Bonillo C, Ozouf-Costaz C, Hureau J-C (1997) Molecular evidence for the origins of Antarctic fishes: Paraphyly of the Bovichtidae and no indication for the monophyly of the Notothenioidei (Teleostei). Polar Biol 18: 193-208

Leese F, Mayer C, Held C (accepted) Isolation of microsatellites from unknown genomes using known genomes as enrichment templates. Limnol Oceanogr Methods

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11: 2453-2465

Loerz AN, Held C (2004) A preliminary molecular and morphological phylogeny of the Antarctic Epimeriidae and Iphimediidae (Crustacea, Amphipoda). Mol Phylogenet Evol 31: 4-15

Lunt DH, Hutchinson WF, Carvalho GR (1999) An efficient method for PCR-based isolation of microsatellite arrays (PIMA). Mol Ecol 8: 891-893

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30: 194-200

Near TJ (2004) Estimating divergence times of notothenioid fishes using a fossil-calibrated molecular clock. Antarct Sci 16: 37-44

Near TJ, Pesavento JJ, Cheng CHC (2003) Mitochondrial DNA, morphology, and the phylogenetic relationships of Antarctic icefishes (Notothenioidei: Channichthyidae). Mol Phylogenet Evol 28: 87-98

Near TJ, Pesavento JJ, Cheng CHC (2004) Phylogenetic investigations of Antarctic notothenioid fishes (Perciformes: Notothenioidei) using complete gene sequences of the mitochondrial encoded 16S rRNA. Mol Phylogenet Evol 32: 881-891

Nolte AW, Stemshorn KC, Tautz D (2005) Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Mol Ecol Notes 5: 628-636

Page TJ, Linse K (2002) More evidence of speciation and dispersal across the Antarctic Polar Front through molecular systematics of Southern Ocean *Limatula* (Bivalvia: Limidae). Polar Biol 25: 818-826

Patarnello T, Bargelloni L, Varotto V, Battaglia B (1996) Krill evolution and the Antarctic ocean currents: Evidence of vicariant speciation as inferred by molecular data. Mar Biol 126: 603-608

Peck LS (2002) Ecophysiology of Antarctic marine ectotherms: limits to life. Polar Biol 25: 31-40

Perez T, Albornoz J, Dominguez A (1998) An evaluation of RAPD fragment reproducibility and nature. Mol Ecol 7: 1347-1357

Raupach MJ, Held C, Wagele JW (2004) Multiple colonization of the deep sea by the Asellota (Crustacea: Peracarida: Isopoda). Deep-Sea Research Part II-Topical Studies In Oceanography 51: 1787-1795

Reilly A, Ward RD (1999) Microsatellite loci to determine population structure of the Patagonian toothfish *Dissostichus eleginoides*. Mol Ecol 8: 1753-1754

Ritchie PA, Lavoue S, Lecointre G (1997) Molecular phylogenetics and the evolution of Antarctic Notothenioid fishes. Comp Biochem Physiol A Comp Physiol 118: 1009-1025

Roeder AD, Marshall RK, Mitchelson AJ, Visagathilagar T, Ritchie PA, Love DR, Pakai TJ, McPartlan HC, Murray ND, Robinson NA, Kerry KR, Lambert DM (2001) Gene flow on the ice: genetic differentiation among Adelie penguin colonies around Antarctica. Mol Ecol 10: 1645-1656

Schlötterer C (2004) The evolution of molecular markers - just a matter of fashion. Nat Rev Genet 5: 63-69

Shaw PW, Arkhipkin AI, Al Khairulla H (2004) Genetic structuring of Patagonian toothfish populations in the Southwest Atlantic Ocean: the effect of the Antarctic Polar Front and deep-water troughs as barriers to genetic exchange. Mol Ecol 13: 3293-3303

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87: 651-701

Stankovic A, Spalik K, Kamler E, Borsuk P, Weglenski P (2002) Recent origin of sub-Antarctic notothenioids. Polar Biol 25: 203-205

Thatje S, Hillenbrand CD, Larter R (2005) On the origin of Antarctic marine benthic community structure. TREE 20: 534-540

Valsecchi E, Palsboll P, Hale P, Glockner-Ferrari D, Ferrari M, Larsen F, Mattila D, Sears R, Sigurjonsson J, Brown M, Corkeron P, Amos B (1997) Microsatellite genetic distances between oceanic populations of the humpback whale (*Megaptera novaeangliae*). Mol Biol Evol 14: 355-362

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15: 1419-1439

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11: 1-16

Zane L, Ostellari L, Maccatrozzo L, Bargelloni L, Battaglia B, Patarnello T (1998) Molecular evidence for genetic subdivision of Antarctic krill (*Euphausia superba* Dana) populations. Proc R Soc Lond B Biol Sci 265: 2387-2391

Zane L, Patarnello T (2000) Krill: a possible model for investigating the effects of ocean currents on the genetic structure of a pelagic invertebrate. Can J Fish Aquat Sci 57: 16-23

Chapter 3

Isolation of microsatellites from unknown genomes using known genomes as enrichment templates

Florian Leese, Christoph Mayer and Christoph Held

Limnology and Oceanography Methods, in press

Abstract

This study analyses the performance of a method to isolate microsatellites from largely unknown target genomes. This 'reporter genome protocol' (*RGP*) utilizes naturally occuring repeat motifs in genomes of distantly related organisms as hybridization probes. The *RGP* proved very successful in all 13 enrichment reactions from eight marine target species (including a dinoflagellate, a diatom and six arthropod species), yielding on average >85% positive colonies. The *RGP* screens for all repeat types occurring in the reporter genomes and is therefore less biased than standard protocols that typically test few, short motifs only. Using the genomes of *Mus musculus*, *Drosophila melanogaster* and *Homo sapiens* as reporter genomes in this study, 133 different di- to hexanucleotide repeat types were obtained. Success of the *RGP* did not depend on overall microsatellite density in the reporter genome but increased with genetic distance between target and reporter genome. Relative abundance of repeat types in the reporter genome had a significant effect on repeat type frequencies in the target library. Altogether, the *RGP* is very efficient, easy to use and reports a large number of different repeat types. It greatly simplifies the isolation of microsatellites from unknown genomes and makes microsatellite markers more attractive for a wide range of studies.

Introduction

Microsatellites are stretches of tandemly repeated DNA motifs typically 2-6 nucleotides long and frequently found in eukaryotic genomes. Their high analytical power (unlinked, codominant, single locus, multiallelic markers) has led to their continued use (Schlötterer 2004) in population genetics (Bowcock et al. 1994; Paetkau et al. 1995; Zeller et al. 2006), forensic analyses (Balding 1999), assignment studies (Jones et al. 2003) and genetic mapping (Dib et al. 1996; Weissenbach et al. 1992). However, microsatellites are still often neglected as they suffer from the drawback that their initial setup is unpredictably difficult (e.g. Weetman et al. 2007). As primer sites are generally not conserved across species boundaries, microsatellite loci have to be identified *de novo* for most species. This often led researchers to choose less powerful but more easily obtainable markers such as RAPD, ISSR, mtDNA, AFLP or RFLP. Despite a growing number of efficient microsatellite isolation protocols in recent years (Glenn and Schable 2005; Zane et al. 2002) there is a continuous urge for more efficient strategies to facilitate the establishment of microsatellite markers.

In model organisms whose genomes are almost completely known, the task of finding microsatellite loci is reduced to scanning sequence databases with a suitable search tool (Benson 1999, Phobos this study). In more exploratory studies, focusing on organisms with almost unknown genomes, the aim is to find informative markers with as little effort as possible. There are two main strategies to achieve this (see Zane et al. 2002 for further details):

The first, more traditional strategy involves building a non-enriched, partial genomic library of the genome of the organism under study (from now on referred to as *target genome*). Those fragments in the library that include microsatellites are subsequently identified by hybridizing against labelled synthetic oligomer probes consisting of typical nucleotide repeat motifs such as $(AC)_8$, $(AG)_8$ etc. The major drawback of screening a non-enriched library for microsatellites is its inefficiency, i.e. many fragments without repeats must be screened in order to identify fragments with microsatellites.

The second, now *de facto* standard strategy aims to increase the efficiency by including an enrichment step, selectively favouring microsatellite-containing fragments, prior to the construction of the partial genomic library of the target genome. This is particularly important for genomes for which low densities of microsatellites have been reported The enrichment is typically achieved by letting the repeat-containing genomic fragments hybridize to synthetic repeat oligomers and removing those without repeat elements during subsequent washing steps (Armour et al. 1994; Kandpal et al. 1994; Karagyozov et al. 1993; Kijas et al. 1994; and Zane et al. 2002 for a review).This approach of enriching microsatellites has one major drawback: The motifs of the microsatellites searched for have to be defined *a priori*, introducing what we call a *selection bias* in the genomic library created. There are 4, 10, 33, 102, and 350 different and independent patterns for di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively (see material and methods for details). Unfortunately, in the absence of knowledge about the most frequent repeat types in an anonymous target genome the choice of synthetic repeat probes is largely arbitrary and often based only on known frequency patterns from distantly related model species. This increases the risk that even abundant repeat types in the anonymous target genome may be overlooked completely due to the choice of hybridization oligomers that are rare or absent in the target genome.

The need for highly successful protocols becomes even more obvious when taking into

account that after the initial identification, often a significant proportion of the microsatellites will have to be excluded for several resons (Fig. 5). Candidate microsatellites with excessive stutter bands, no suitable flanking region for primer placement and no or too many amplicons have to be discarded. The same holds true for microsatellites that would bias the subsequent data analysis steps due the presence of null alleles, allelic dropout, linkage disequilibrium, etc. (Selkoe and Toonen 2006).

In contrast to all other protocols a novel selective enrichment protocol proposed by Nolte et al. (2005) does not rely on the use of synthetic probes, but uses genomes of unrelated organisms (from now on referred to as *reporter genomes*) as hybridization templates (Fig. 1). The concept behind this technique is the assumption that evolutionary distant genomes will show a high similarity by state in genomic regions of low complexity such as microsatellite loci. These regions act as 'non-synthetic universal hybridization probes' (Fig. 1).

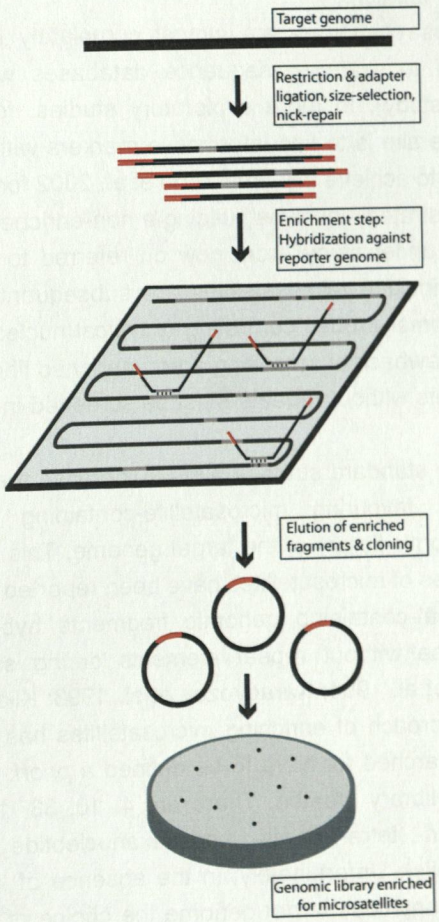


Figure 1: The principle of the reporter genome protocol after Nolte et al. (2005). Fragments of DNA of a target species (*target genome*) are ligated to AFLP-adaptors and hybridized against single-stranded DNA of a distantly related taxon (*reporter genome*). As a result, only fragments with a high identity-by-state are retained, which are often simple sequence repeats (represented by light-grey areas in the DNA fragments). These fragments are eluted, cloned and transformed into *E.coli*.

Consequently, the technique will preferentially detect repeat types that are abundant in both the target and the reporter genome and is thus potentially less susceptible to the effects of selection bias. Although it holds several advantages over standard protocols, its utility has not been tested systematically across different taxonomic groups since its inception (Nolte et al. 2005).

In this study, eight marine target species from three very different taxonomic groups (arthropods, diatom, dinoflagellate, Tab. 1) are used to systematically test the applicability and performance of the reporter genome protocol.

Table 1: Species names, taxonomic position and origin of samples, from which DNA was extracted.

Species	Taxonomic position	Origin of samples
Ceratoserolis trilobitoides (Eights, 1833)	Animal, isopod crustacean	Southern Ocean, Elephant Island
Serolis paradoxa (Fabricius, 1775)	Animal, isopod crustacean	Strait of Magellan, Patagonia, Chile
Glyptonotus antarcticus (Eights, 1853)	Animal, isopod crustacean	Southern Ocean, South Sandwich Islands
Septemserolis septemcarinata (Miers, 1875)	Animal, isopod crustacean	Southern Ocean, Bouvet Island
Munida gregaria (Fabricius, 1793)	Animal, decapod crustacean	South Atlantic Ocean, Falkland Islands
Pseudo-nitzschia multistriata Takano, 1993	Diatom algae	Provided by A. Lüdeking, Napels
Alexandrium tamutum Montresor, Beran & John, 2004	Protist, dinoflagellate	Provided by T. Alpermann, Bremerhaven

Material and procedures

Samples

Samples of the eight marine target species were obtained from different sources (Tab. 1). DNA was extracted using either the Qiagen DNeasy Tissue Mini or the Plant Mini Kit. *Mus musculus domesticus, Drosophila melanogaster* (CantonS), and *Homo sapiens* were used as reporter genomes. The *Ceratoserolis* species analysed in this study is one species from a known complex of cryptic species, which is referred to as "group 1" in Held (2003) and *Ceratoserolis* n. sp. 1 in Leese and Held (2008). The species is not identical to the type species *Ceratoserolis trilobitoides sensu strico*. A formal description is in preparation, but until its publication the species' name *C. trilobitoides* (Eights, 1833) is valid and will be used here.

Reporter genome protocol

Hybridization chips with single stranded DNA of *Mus musculus domesticus, Drosophila melanogaster* (CantonS), and *Homo sapiens* were prepared as follows: 5x5 mm pieces of Hybond N+ (GE Healthcare) were incubated in a 1:1 mixture of 100 ng/ul reporter genome DNA and 1M NaOH for five minutes at room temperature. Then the chip was transferred to 1M Tris-CL (pH 5.0) for 2 minutes, following a 2 minutes incubation step in 2xSSC at 50°C. Excess of liquid was removed and the DNA baked to the membrane for 12 hours at 80°C.

A small-insert genomic library enriched for microsatellites was created using total genomic DNA of the target species. Digestion of 500 ng template DNA and ligation to standard AFLP adaptors 5'-TACTCAGGACTCAT-3' / 5'GACGATGAGTCCTGAG-3' (Vos et al. 1995) were carried out simultaneously with 20 U of the *MseI* isoschizomer *TruI* and 20 U of T4-DNA ligase in a 1X Buffer-R reaction mix (50 µl; all products Fermentas Lifesciences). Both enzymes have different optimal working temperatures (65°C and 22°C). In accordance with the suggestions of the supplier, the reaction was incubated at an intermediate temperature (37°C) for 6 hours and both enzymes were added in excess. Ligase was inactivated by a 10 minutes hold step at 65°C. The reaction was run twice on a 1.5% TAE agarose gel and the 400-800 bp fraction was excised and purified (Eppendorf Perfectprep® Gel cleanup). Of the 30 µl of eluted DNA, 5 µl were amplified in a total volume of 50 µl with 0.5 µM of the AFLP adaptor-specific primer *Mse*I-N (5'-GATGAGTCCTGAGTAAN-3') and 0.03 U/µl HotMaster® *Taq* (Eppendorf). PCR conditions were 65°C for 10 min to repair nicks remaining from adapter ligation, an initial denaturation step at 94°C for 2 min followed by 25 cycles (*P. multistriata*, *A. tamutum* 40 cycles) of 30 sec at 94°C, 45 sec at 52°C, 80 sec at 65°C and a final elongation step of 10 min at 65°C. The reaction was purified using the Qiaquick PCR Kit (Qiagen).

The hybridization chips were pre-incubated for 15 min at 50°C in 500 µl hybridization buffer (5x SSC, 0.5% SDS, 1x Denhardt's, 50 µg/ml Heparin). Then 400 ng template DNA was added and the tube heated to 95°C for 5 minutes to denature template DNA. For hybridization, the mix was kept for 30 minutes at 50°C. Subsequently, the hybridization chip was washed three times in hybridization buffer (50°C) and then again thrice in washing buffer (0.1x SSC, 0.1% SDS, 50°C). Finally, DNA was eluted from the hybridization chip by transferring it into 500 µl TE buffer (pH 8.0, 5 min at 80°C). DNA was precipitated using a standard isopropanol sodium-acetate protocol (Sambrook et al. 1989).

The enriched fragments were amplified as above (25 µl reaction). Purified fragments were cloned into pCR2.1-TOPO® TA vector and transformed into competent TOP10F' or JM109 *E.coli* (Invitrogen, Promega). Cultures of positive colonies according to blue-white selection were grown overnight in LB-medium containing 100 µg/ml ampicillin. Plasmid preparation was performed using the Eppendorf FastPlasmid® Mini Kit or outsourced to GATC-Biotech (Konstanz, Germany). Shotgun sequencing using standard M13-forward and/or reverse primers was either conducted in-house on an ABI 3130xl or LI-COR 4200 automated sequencer, or outsourced to Macrogen (Seoul, Korea) and GATC-Biotech (Konstanz, Germany). The addition of DMSO to a final concentration of 5% improved the quality of reads with microsatellites. BLAST searches were carried out for all sequences to exclude the possibility of contamination with mobilized DNA from the reporter genome.

In this study, eight target genomes were hybridized against some or all of three reporter genomes, yielding a total of thirteen libraries (see Tab. 2 for a matrix of combinations).

Non-enriched genomic library

A non-enriched partial genomic library was created for *C. trilobitoides* according to Rassmann et al. (1991). In total, 3456 clones were dotted onto four nitrocellulose membranes and screened with the radioactively labeled γ-^{33}P oligomer probes $(AC)_8$. Pre-hybridization and hybridization were carried out in rotating glass tubes in an incubator at 43°C. The filters and intensifying screen were exposed overnight to X-Ray films (Kodak) at -20°C and developed according to the recommendations of the manufacturer. Colonies with strong signal intensity on the film were

traced back to the 96-well plates based on their position in the spatial array on the nitrocellulose membranes. Positive colonies were grown overnight, extracted (Chelex) and sequenced (conditions see above).

Data analysis

Sequence data analysis was performed in a semi-automated workflow based on the Staden software package, version 1.70 (Staden 1996) into which the novel microsatellite search tool Phobos (Mayer, www.rub.de/spezzoo/cm) and Primer3 (Rozen and Skaletsky 2000) had been integrated (Beszteri et al. in preparation). Self-similarity dotplots were used to visualize intrinsic structural features of repeat regions using the Staden tool Spin or dotlet (Junier and Pagni 2000) and to help avoid placing primers in regions that were not unique.

First, sequence data were processed using the Staden pre-processing tool Pregap with integrated programs Phred (Ewing and Green 1998; Ewing et al. 1998) for base calling, Cross-match for cloning vector and adaptor identification and Phobos for masking of repeats. Normal shotgun assembly of the processed reads was conducted with the Staden assembly tool Gap4, allowing for a maximum of 2% mismatch between two reads. All non-redundant contigs were entered into the Staden database for subsequent analysis. Phobos was used to search for all perfect and imperfect microsatellites present in the genomic libraries and the three different reporter genomes, i.e. *Mus musculus* (NCBI Build 36.1, May 2006: ftp://ftp.ncbi.nih.gov/genomes/M_musculus, taxonomic ID 10090), *D. melanogaster* as annotated by the FlyBase consortium (FlyBase Release 5.1, February 2005; ftp://ftp.ncbi.nih.gov/genomes/Drosophila_melanogaster), and *Homo sapiens* (NCBI Build 36.2, March 2006).

The parameter-settings used in Phobos in the present analysis were as follows: The scoring parameters (match, mismatch, N, gap) used were (1, -6, 0, -6), where the first repeat unit was not scored. No more than two consecutive N's were allowed. For computing a percentage perfection, N's were counted as mismatch positions. The *"recursion depth"* parameter used was 5. Microsatellites were reported if they achieve a minimum score of 8 and additionally had a minimum length of three perfect repeat units (minlength_b=2). Consequently, only dinucleotide repeat arrays with a minimum length of 5 repeat units, trinucleotide with a minimum length of 4, tetra-, penta-, and hexanucleotide repeats with a minimum length of 3 repeat units were included in the analysis. Microsatellite motifs were reported according to Chambers and MacAvoy (2000), where a minimal alphabetic representation is determined by allowing cyclic permutations and computing the reverse complement of the nucleotides in the repeat unit, e.g. GA, AG, TC, CT are all represented by AG. Applying this reporting scheme, as many as 4 dinucleotide, 10 tri, 33 tetra, 102 penta, and 350 hexanucleotide motifs exist.

After the microsatellites had been detected by Phobos, they were analysed with the newly developed program sat-stat, version 1.0.1. This analysis tool is capable of filtering and sorting microsatellites according to their properties such as repeat type, unit length, total length, score, and perfection and to analyse these properties statistically.

For this study, only repeats with a unit size of 2-6 bp and ≥90% percentage perfection were considered, compound microsatellites were not considered separately (see Weber (1990) for a definition and Fig. 2f).

To statistically assess whether the repeat class or repeat type content of two partial genomic libraries differs significantly, permutation tests (100,000 permutations) were performed with a self-written computer program (Christoph Mayer, available on request).

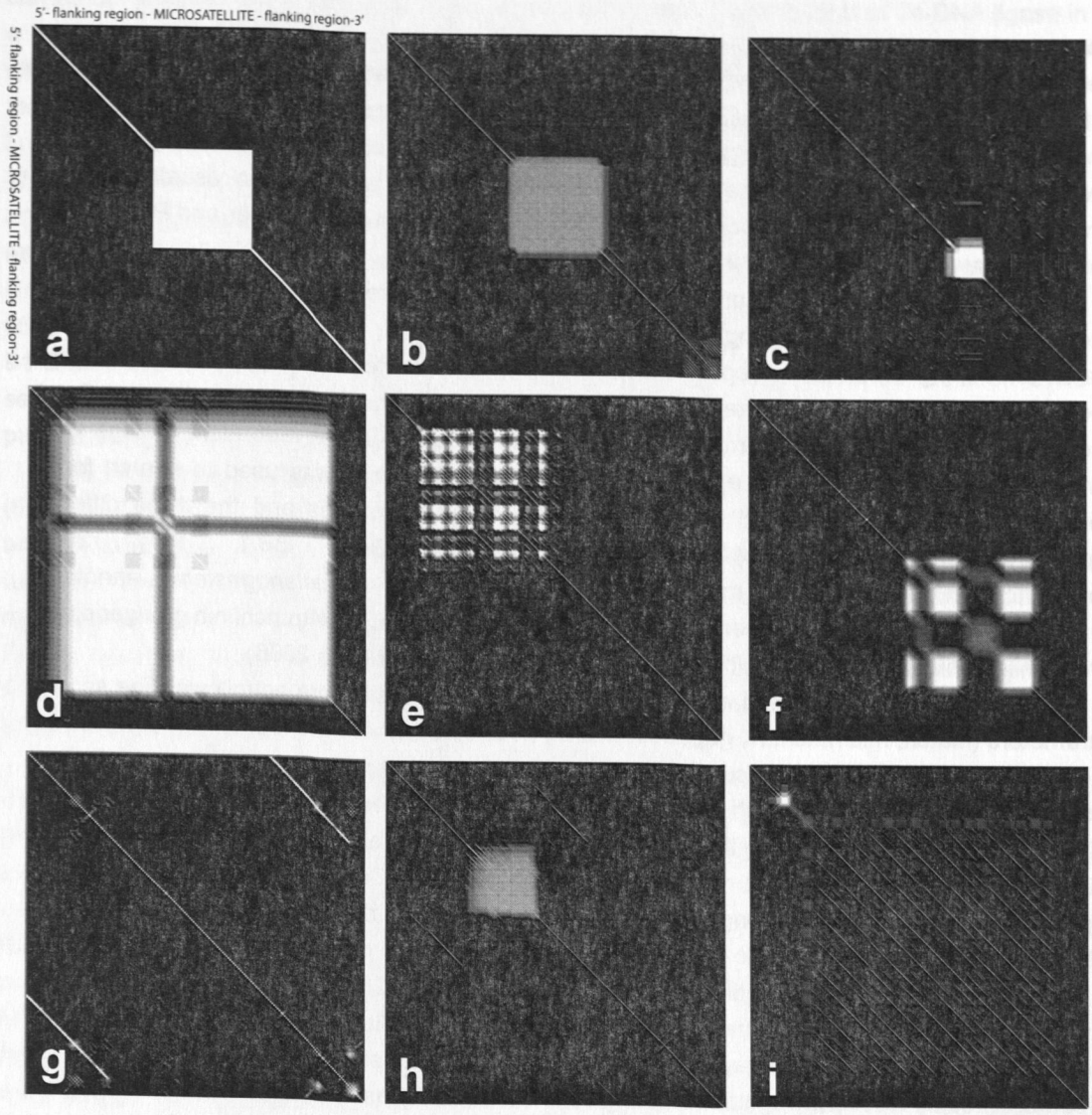


Figure 2: Dot-plots are showing the self-similarity of a sequence by plotting it against itself. (a) Bright areas indicate a high self-similarity resulting from the presence of a microsatellite. For population genetic analyses, highly perfect repeat loci (b, c) should be preferred. Many microsatellite-containing fragments are in fact imperfect with one (d) or several (e) interruptions. f gives an example of a complex microsatellite consisting of five microsatellites (two different types. Flanking regions can reveal duplications (g, h) that may complicate marker setup. Also, higher-order repeat structures are found in sequences when plotted against itself (h, i). GenBank accession numbers: b) EU056273, c) EU056276 , d) EU234060, e) EU234059, f) EU234055, g) EU234058, h) EU234056, i) EU234057.

Assessment

Reporter genome protocol

With the reporter genome protocol 1370 clones from thirteen genomic libraries and eight different target species (Tab. 1) were sequenced. Redundant samples were found in all but the genomic library derived from the bopyrid isopod enriched using *M. musculus* as reporter genome (Ø=13.1%, standard deviation 13.3). All redundant sequences were removed for analysis yielding a total number of 1207 unique sequences (Tab. 2).

BLASTn searches were performed to check whether sequences in the enriched library originated from the reporter genome rather than the target genome. No contaminant fragments of the reporter genome DNA were detected indicating that mobilization of reporter genome fragments was not a problem. In the genomic library from *C. trilobitoides* hybridized against *D. melanogaster*, a BLAST search revealed 19 inserts originating from regions of the 18S or 28S rRNA gene. These inserts belonged to five almost identical groups, of which only one group consisting of two inserts contained a microsatellite.

The amount of perfect repeats ranged from 45.9% to 77.3.

Table 2: Number of non-redundant clones analysed, cloning efficiency, total number and number of different microsatellite motifs detected in the fifteen genomic libraries enriched for microsatellites using the reporter genome protocol.

Library	Target genome	Reporter genome	Clones analysed	Clones with microsatellites	Number of msats	Number of motifs
1	*C. trilobitoides*	*M. musculus*	180	171 (95%)	395	35
2	*C. trilobitoides*	*D. melanogaster*	111	91 (82%)	226	36
3	*C. trilobitoides*	*H. sapiens*	64	62 (97%)	146	16
4	*S. paradoxa*	*M. musculus*	146	132 (90%)	351	41
5	*S. paradoxa*	*D. melanogaster*	131	110 (84%)	229	36
6	*G. antarcticus*	*M. musculus*	63	62 (98%)	179	25
7	*G. antarcticus*	*D. melanogaster*	72	61 (84%)	130	25
8	*Bopyridae*	*M. musculus*	28	25 (89%)	92	11
9	*Bopyridae*	*D. melanogaster*	45	40 (89%)	114	19
10	*S. septemcarinata*	*M. musculus*	146	37 (25%)	63	20
11	*M. gregaria*	*D. melanogaster*	9	8 (89%)	21	6
12	*Pseudo-nitzschia multistriata*	*M. musculus*	16	16 (100%)	44	7
13	*A. tamutum*	*M. musculus*	196	171 (87%)	546	88

Efficiency of enrichment

Taking the percentage of clones with at least one microsatellite as a measure enrichment success, twelve of the thirteen genomic libraries were highly enriched for microsatellites with an average success rate of 85.5% and a standard deviation of 18.9% (Tab. 2). Enrichment efficiency was lowest for *S. septemcarinata* (25%) using *M. musculus* as reporter genome (library 10) and highest for *P. multistriata* and *G. antarcticus* using *M. musculus* as reporter genome (100% and 98.4%, libraries 12 and 6, Tab. 2). Excluding the outlier library 10, the average success rate was 90.5% ± 5.8%.

For target genomes that were enriched with more than one reporter genome, the influence of different reporter genomes on enrichment success was evaluated (libraries 1-9). Success of the

enrichment depended significantly on the reporter genome. The mean enrichment efficiency of *M. musculus* was significantly (t-test, P=0.016) higher than using *D. melanogaster* as reporter genome (93.3% ±4.2, and 84.9% ±2.9, respectively). For *C. trilobitoides*, using the reporter genomes of *H. sapiens* or *M. musculus* was almost equally efficient (Tab. 2).

For all thirteen genomic libraries, as many as 133 different repeat motifs were detected (Tab. 4). The major proportion of microsatellites were dinucleotide repeats, except for *P. multistriata* (Fig. 3), the proportions of longer repeat classes differed considerably between taxonomic groups (Tab. 4).

Table 3: The table heading shows the total number and density of microsatellites in the three reporter genomes. The table rows show the densities, relative amounts, and length characteristics (number of repeat units) of the five repeat families in these genomes. Detection parameters in *Phobos* see material and methods.

Reporter genomes	***M. musculus*** (2590 Mbp) 2424254 microsatellites 476 different units			***D. melanogaster*** (122 Mbp) 96779 microsatellites 467 different units			***H. sapiens*** (2940 Mbp) 1651071microsatellites 454 different units		
	density [bp/Mbp]	relative amount	length [Ø±SD]	density [bp/Mbp]	relative amount	length [Ø±SD]	density [bp/Mbp]	relative amount	length [Ø±SD]
dinucleotide	10044	41.9	28.1±24.5	3869	29.9	15.8±8.1	3302	33.0	18.2±12.2
trinucleotide	2891	12.1	20.7±26.7	4061	31.3	14.8±8.3	1554	15.5	14.9±9.9
tetranucleotide	7326	30.6	24.3±25.5	2494	19.2	15.1±7.9	3500	35.0	19.8±16.1
pentanucleotide	2261	9.4	29.8±30.2	1110	8.6	19.3±7.6	1204	12.0	22.8±14.8
hexanucleotide	1447	6.0	38.7±44.0	1425	6.0	24.3±28.5	449	4.5	25.2±18.2

Influence of the reporter genome characteristics on the microsatellites detected

The genomes of the three taxa used as reporter genomes differ considerably in total genome size as well as in the density, relative abundance, and lengths of their microsatellites (Tab. 3, Fig. 4). Density of microsatellites is highest in the genome of *M. musculus* (23733 bp/Mbp), intermediate in *D. melanogaster* (12924 bp/Mbp), and lowest in *H. sapiens* (9927 bp/Mbp).

The relative amounts of di- to hexanucleotide repeat classes and the length characteristics differ between the three genomes (Tab. 3, Fig. 4). Dinucleotide repeats are abundant in all three genomes, but are relatively more abundant in *M. musculus* and *H. sapiens* than in *D. melanogaster*. *D. melanogaster* is comparatively rich in trinucleotide repeats, whereas tetranucleotide repeats occur at higher densities in the genomes of *M. musculus* and *H. sapiens*. Penta- and hexanucleotide repeats constitute only a minor proportion of repeats in the genomes (Tab. 3).

The enriched genomic libraries (except 8, 11 and 12) contained microsatellites from all five different repeat classes (Fig. 4). However, dinucleotide repeats were the most abundant repeat class in ten of the 13 libraries.

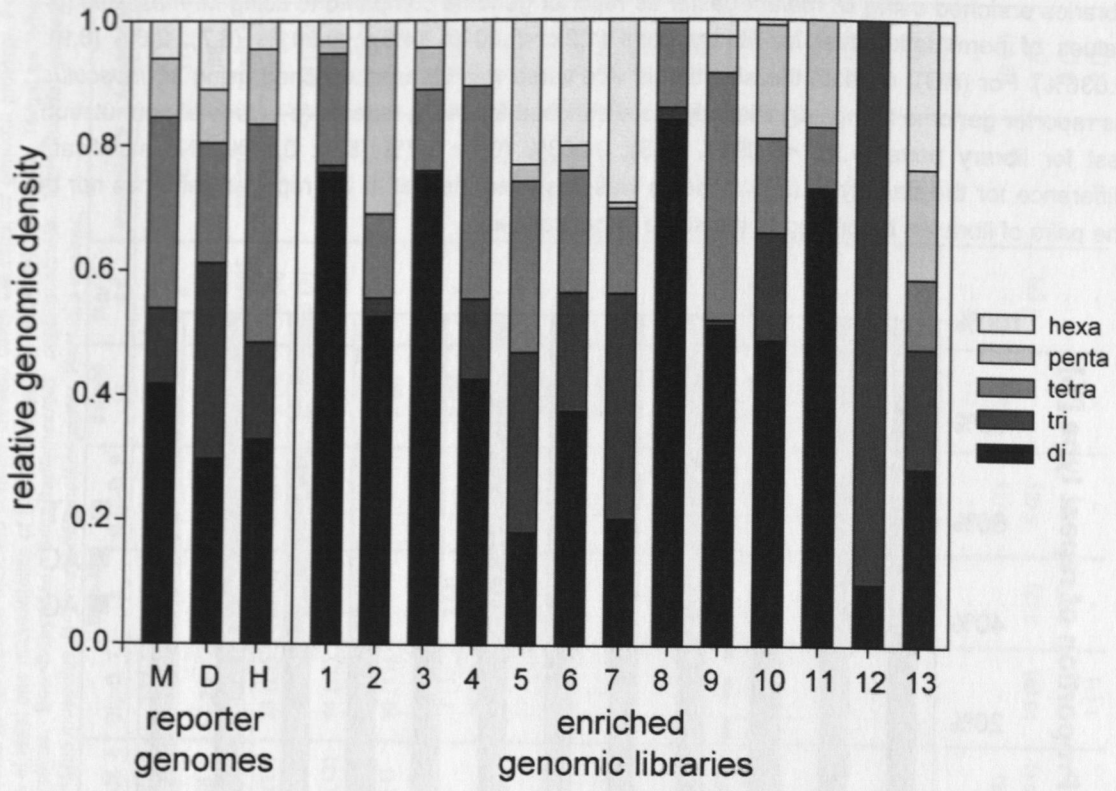


Figure 3: Relative amounts of di- to hexanucleotide repeats in the reporter genomes (M=*Mus musculus*, D=*Drosophila melanogaster*, H=*Homo sapiens*) and the 13 genomic libraries enriched using the reporter genome protocol. Many different repeat classes were found in all but library 12 (*Pseudo-nitzschia multistriata*).

Using a statistical permutation test (see material and methods) it has been analysed whether the high relative density of dinucleotide repeats in the reporter genome of *M. musculus* as compared to *D. melanogaster* and the high relative density of trinucleotide repeats in *D. melanogaster* as compared to *M. musculus* (Fig. 4) finds a significant reflection in the enriched genomic libraries. The library pairs with the same target genome being enriched for microsatellites using different reporter genomes that have been compared are (1,2), (4,5), (6,7), and (8,9). It can be observed (Fig. 4) that indeed, in each library pair, the relative density of dinucleotide repeats is always higher in the library produced with *M. musculus*. The probabilities (p-values) that in these pairs a higher density of dinucleotide repeats than in libraries 1, 4, 6, and 8 (created with *M. musculus*) could have arisen by chance are, respectively, 0.02%, 0.002%, 1.6%, 1.5%, which is highly significant. Similarly, in the library pairs (1,2), (4,5), (6,7), the relative densities of trinucleotide repeats are higher in the libraries created using *D. melanogaster* as reporter genome. The probabilities that an even higher relative trinucleotide density could have arisen by chance are 21%, 0.6%, 5.8%, showing a significant level only for pair (4,5) with *S. paradoxa* as target species.

Furthermore, it has been tested, whether the repeat type densities of $(AC)_n$, $(AG)_n$, and $(AT)_n$ repeats in the reporter genome had an influence on the enriched libraries:

The relative densities of $(AC)_n$ repeats were significantly higher for the set of genomic

libraries enriched using *D. melanogaster* as reporter genome compared to using *M. musculus* (p-values of permutation test for library pairs (1,2): <0.001% (4,5): <0.001% (6,7): 2.6% (8,9): 0.036%). For $(AG)_n$ repeats the situation is vice versa with libraries enriched using *M. musculus* as reporter genome being significantly more enriched for $(AG)_n$ repeats (p-values of permutation test for library pairs (1,2): <0.001% (4,5): 0.003% (6,7): 1.7% (8,9): 0.04%). No significant difference for the density of $(AT)_n$ repeats was observed, neither in the reporter genomes nor in the pairs of libraries belonging to the same target species.

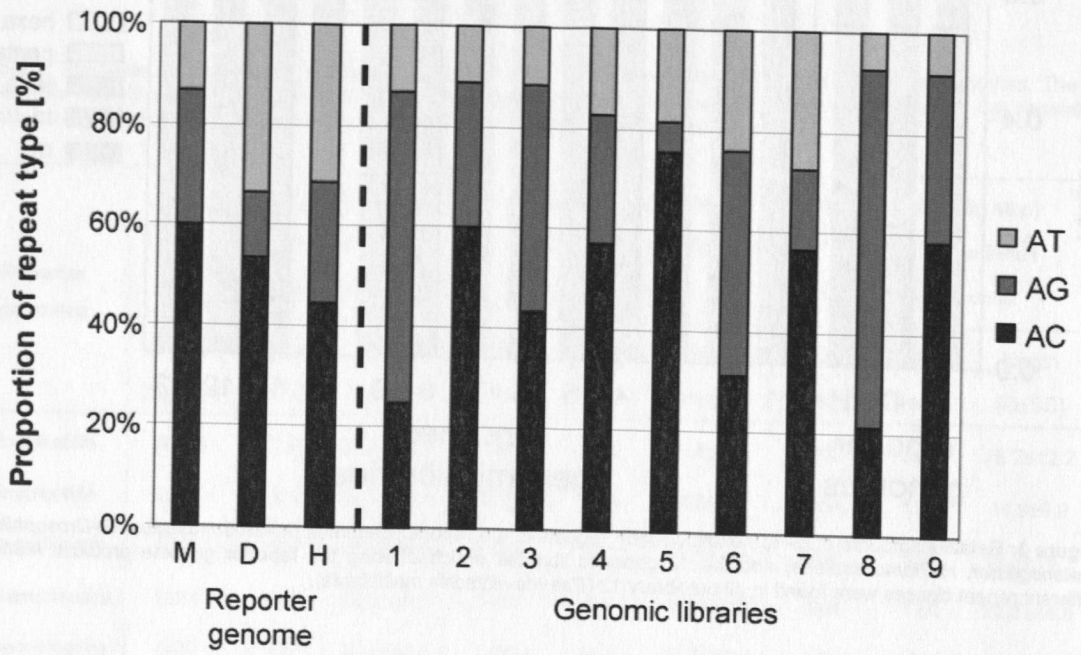


Figure 4: Relative densities of the three major dinucleotide repeats $(AC)_n$, $(AG)_n$ and $(AT)_n$ in the genomic libraries of the four isopods *Ceratoserolis trilobitoides* (1-3), *Serolis paradoxa* (4,5), *Glyptonotus antarcticus* (6,7) and the bopyrid isopods (8,9) enriched using the reporter genomes of M=*Mus musculus*, D=*Drosophila melanogaster*, H=*Homo sapiens*.

Table 4: Efficiency characteristics and total number of each particular repeat type detected in the thirteen enriched libraries created using the reporter genome protocol. For each occurring repeat type the number of microsatellites and the number of inserts with at least one microsatellite of that repeat type (in brackets) are shown. Target genomes of: Ctr=*Ceratoserolis trilobitoides*, Spa= *Serolis paradoxa*, Gly=*Glyptonotus antarcticus*, Bop=Bopyridae, Sse=*Septemserolis septemcarinata*, Mgr=*Munida gregaria*, Pmu=*Pseudo-nitzschia multistriata*, Ata=*Alexandrium tamutum*. Reporter genomes of M=*Mus musculus*, D=*Drosophila melanogaster*, H=*Homo sapiens*.

Target-Reporter Genomic library	Ctr-M 1	Ctr-D 2	Ctr-H 3	Spa-M 4	Spa-D 5	Gly-M 6	Gly-D 7	Bop-M 8	Bop-D 9	Sse-M 10	Mgr-D 11	Pmu-M 12	Ata-M 13
# microsatellites	395	226	146	351	229	179	130	92	114	63	21	44	546
# positive inserts	171	91	62	132	110	62	61	25	40	37	8	16	171
# units	35	36	16	41	36	25	25	11	19	20	6	7	88
density [bp/Mbp]	284610	202313	265765	302243	348958	410562	347445	206999	300053	30274	231505	288330	246011
AC	69 (44)	79 (54)	48 (30)	123 (63)	91 (53)	33 (17)	33 (25)	10 (6)	34 (21)	6 (5)	6 (5)	3 (3)	179 (83)
AG	190 (110)	37 (26)	53 (31)	59 (28)	7 (4)	53 (24)	12 (7)	70 (20)	45 (12)	11 (9)	5 (3)	-	28 (22)
AT	27 (24)	11 (10)	9 (8)	23 (19)	13 (13)	16 (13)	14 (12)	2 (2)	3 (3)	11 (9)	-	-	2 (2)
CG	1 (1)	-	1 (1)	1 (1)	1 (1)	-	-	-	-	-	-	-	23 (20)
AAC	1 (1)	1 (1)	1 (1)	6 (6)	23 (23)	5 (4)	16 (11)	-	-	2 (2)	-	11 (11)	21 (12)
AAG	1 (1)	2 (1)	-	4 (3)	-	6 (5)	9 (6)	-	2 (1)	2 (2)	-	3 (3)	42 (18)
AAT	8 (6)	3 (3)	-	6 (5)	1 (1)	-	1 (1)	2 (2)	-	5 (4)	-	-	5 (4)
ACC	1 (1)	1 (1)	-	-	1 (1)	-	1 (1)	1 (1)	-	-	-	3 (3)	10 (6)
ACG	-	1 (1)	-	8 (4)	2 (2)	-	-	-	-	-	-	-	2 (2)
ACT	-	-	-	4 (4)	-	-	-	-	1 (1)	3 (2)	-	20 (9)	3 (1)
AGC	1 (1)	2 (2)	-	1 (1)	2 (2)	-	3 (1)	-	-	1 (1)	-	3 (3)	5 (4)
AGG	-	2 (1)	-	1 (1)	-	1 (1)	-	-	-	2 (2)	-	1 (1)	3 (3)
ATC	-	-	-	-	-	3 (3)	-	-	-	1 (1)	-	-	5 (4)
CCG	-	-	1 (1)	-	-	-	-	-	-	1 (1)	-	-	1 (1)
AAAC	5 (5)	3 (3)	-	4 (3)	3 (3)	1 (1)	1 (1)	-	1 (1)	-	1 (1)	-	4 (3)
AAAG	6 (6)	1 (1)	5 (5)	12 (12)	3 (3)	9 (6)	-	-	1 (1)	2 (2)	-	-	-
AAAT	2 (2)	1 (1)	-	4 (3)	-	7 (4)	-	-	-	2 (2)	1 (1)	-	-
AACC	-	1 (1)	-	-	1 (1)	-	-	-	-	-	-	-	2 (2)
AACG	-	-	-	1 (1)	-	-	-	-	-	-	-	-	1 (1)
AAGC	-	-	-	3 (3)	-	-	-	-	-	-	-	-	2 (2)
AAGG	9 (7)	1 (1)	-	7 (3)	2 (2)	-	-	1 (1)	-	-	-	-	1 (1)
AAGT	6 (6)	1 (1)	1 (1)	1 (1)	-	-	-	-	-	-	-	-	1 (1)
AATC	6 (6)	-	-	-	1 (1)	-	-	-	-	-	-	-	2 (1)
AATG	-	-	-	2 (1)	-	-	-	-	-	1 (1)	-	-	-
ACAG	29 (20)	13 (11)	14 (12)	26 (14)	23 (13)	2 (2)	-	-	5 (1)	8 (2)	-	-	12 (5)
ACAT	8 (6)	2 (2)	1 (1)	10 (8)	3 (2)	13 (11)	10 (9)	2 (2)	5 (5)	-	3 (2)	-	4 (2)
ACCG	-	-	-	-	-	-	-	1 (1)	-	-	-	-	3 (1)
ACCT	-	-	-	-	-	-	1 (1)	-	-	-	-	-	2 (2)

ACGC	-	3 (3)	-	4 (3)	2 (1)	3 (2)	1 (1)	1 (1)	2 (2)	-	-	-	39 (29)
ACGG	-	1 (1)	-	-	-	-	-	-	-	-	-	-	1 (1)
ACTC	1 (1)	1 (1)	-	3 (2)	6 (6)	1 (1)	-	-	2 (2)	-	-	-	4 (3)
ACTG	1 (1)	-	-	-	-	-	-	-	-	-	-	-	-
AGAT	3 (3)	2 (2)	2 (2)	7 (5)	5 (5)	2 (2)	2 (2)	1 (1)	4 (2)	-	-	-	-
AGCC	-	-	-	-	1 (1)	-	-	-	-	-	-	-	-
AGCG	2 (2)	-	-	-	-	-	-	-	-	-	-	-	3 (2)
AGCT	-	-	-	1 (1)	-	-	-	-	-	-	-	-	-
AGGC	-	-	-	1 (1)	-	-	-	-	-	-	-	-	1 (1)
AGGG	-	-	-	-	-	-	-	-	-	1 (1)	-	-	2 (2)
ATCC	-	-	-	1 (1)	-	-	-	-	-	-	-	-	4 (2)
ATGC	-	1 (1)	-	-	-	-	-	-	-	-	-	-	1 (1)
CCCG	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
AAAAC	-	-	1 (1)	-	-	-	-	-	-	-	-	-	-
AAACC	-	-	-	-	1 (1)	-	-	-	-	-	-	-	-
AAAGC	-	-	-	-	-	-	1 (1)	-	-	-	-	-	-
AACAC	-	1 (1)	-	-	-	1 (1)	-	-	-	-	-	-	4 (2)
AACCT	3 (3)	4 (4)	5 (3)	-	2 (2)	1 (1)	-	-	-	1 (1)	-	-	-
AACGC	-	-	-	-	-	-	-	-	-	-	-	-	3 (3)
AAGAC	-	15 (8)	-	-	-	-	-	-	-	-	-	-	-
AAGAG	1 (1)	22 (7)	-	-	-	1 (1)	-	-	3 (3)	-	5 (1)	-	12 (12)
AAGAT	-	-	-	-	1 (1)	-	-	-	-	-	-	-	-
AAGGC	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
AAGGG	1 (1)	-	-	-	-	-	-	-	1 (1)	-	-	-	-
AAGGT	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
AAGTC	1 (1)	-	-	-	-	-	-	-	-	-	-	-	-
AATAC	-	1 (1)	-	-	1 (1)	-	-	-	-	-	-	-	-
AATAT	-	-	-	-	-	1 (1)	-	-	-	1 (1)	-	-	-
AATCT	-	-	-	-	1 (1)	-	-	1 (1)	1 (1)	-	-	-	-
AATAG	-	2 (1)	-	-	-	-	-	-	-	-	-	-	-
AATGC	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACACG	-	-	-	-	-	-	1 (1)	-	-	-	-	-	-
ACAGC	-	-	-	-	-	-	-	-	-	-	-	-	10 (7)
ACATC	-	-	-	-	-	-	-	-	-	-	-	-	3 (3)
ACATG	-	-	-	-	-	-	-	-	-	-	-	-	2 (2)
ACCTC	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)

ACACAG	1 (1)	-	-	-	-	-	-	-	-	-	-	-	-	7 (7)
ACACAT	-	-	2 (2)	5 (5)	1 (1)	-	1 (1)	-	1 (1)	-	-	-	-	1 (1)
ACACCC	-	-	-	1 (1)	-	-	-	-	-	-	-	-	-	1 (1)
ACACCT	-	-	-	-	-	-	1 (4)	-	-	-	-	-	-	-
ACACGC	2 (2)	2 (2)	-	3 (3)	7 (7)	6 (3)	3 (2)	-	-	-	-	-	-	16 (16)
ACACGT	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACACTC	-	-	-	-	1 (1)	-	-	-	-	-	-	-	-	1 (1)
ACACTG	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACAGAG	-	-	1 (1)	4 (4)	-	-	-	-	-	-	-	-	-	5 (5)
ACAGAT	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACAGCC	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACAGGT	-	-	-	1 (1)	-	-	-	-	-	-	-	-	-	-
ACATAG	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACATAT	-	-	-	-	-	1 (1)	1 (1)	-	1 (1)	-	-	-	-	1 (1)
ACATCT	-	-	-	-	-	-	1 (1)	-	-	-	-	-	-	-
ACATGG	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACCACT	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACCATC	-	-	-	-	-	-	-	-	-	-	-	-	-	2 (2)
ACCCGC	-	-	-	-	-	-	-	-	-	-	-	-	-	3 (3)
ACCCTC	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACCGCC	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACCTGG	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACGCAG	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACGCCC	-	-	-	1 (1)	-	-	-	-	-	-	-	-	-	-
ACGTGC	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACTCCC	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)
ACTGCC	-	2 (2)	-	-	-	-	-	-	-	-	-	-	-	-
AGAGAT	1 (1)	-	1 (1)	3 (3)	-	-	-	-	1 (1)	-	-	-	-	-
AGAGGC	-	-	-	-	-	-	-	-	-	-	-	-	-	2 (2)
AGAGGG	2 (1)	-	-	-	-	-	-	-	-	-	-	-	-	-
AGCAGG	-	-	-	-	-	-	-	-	-	-	-	-	-	3 (2)
AGGCAT	-	-	-	-	-	-	-	-	-	-	-	-	-	3 (1)
ATCGCC	-	-	-	-	-	-	-	-	-	-	-	-	-	1 (1)

Discussion

In exploratory studies based on microsatellite data the use of an efficient enrichment protocol and strategies for subsequent data analysis are essential to obtain many appropriate microsatellite loci within a short time and at low cost. This study showed that the reporter genome protocol is a universally applicable and very efficient enrichment protocol that has unique advantages.

Efficiency of the reporter genome protocol

The average success rate of the reporter genome protocol within the thirteen genomic libraries created from eight taxa was 85.5% (90.5% without the outlier *S. septemcarinata*). This is comparable to the success of other recent protocols (Glenn and Schable 2005; Korpelainen et al. 2007; Zane et al. 2002). In our laboratory, previous attempts to isolate microsatellites from *C. trilobitoides* using the screening of (i) a non-enriched library and (ii) an enriched libraries with radioactively labelled, synthetic oligomer probes (Rassmann et al. 1991), and (iii) the identification of microsatellites using the RAPD-based PIMA protocol (Lunt et al. 1999) were conducted. Whereas protocol (i) and (ii) failed to isolate sufficiently many appropriate microsatellites in an acceptable time, protocol (iii) failed at all.

The lower success rate from the enrichment using *S. septemcarinata* as target genome (25%) was still high enough to identify 63 microsatellites within the 146 clones. This was sufficient to establish an informative set of polymorphic markers for population genetic studies (Leese et al. in press).

This study also indicates that there is no significant difference in enrichment efficiency regarding the three different taxonomic target groups. Consequently, the reporter genome protocol is capable of detecting microsatellites from many different eukaryotic genomes without prior knowledge about repeat classes and their frequencies in the target genome.

To estimate an enrichment factor for a given repeat type, its frequency in the genome has to be known. For *C. trilobitoides*, screening a non-enriched library for $(AC)_n$ repeats allowed the calculation of the genomic density of $(AC)_n$ repeats to be approximately 343 bp/Mbp (see results). Cloning efficiency for $(AC)_n$ repeats was 26/3456 = 0.75%, which means that every 134th clone contains a $(AC)_n$ repeat on average. In the three enriched libraries for *C. trilobitoides* (1-3), the proportion of clones with at least one $(AC)_n$ repeat was 25%, 43%, and 60% for *M. musculus*, *H. sapiens* and *D. melanogaster* as reporter genome corresponding to enrichment factors of 33.3, 57.3 and 80.0 for this repeat type. When considering the repeat type density, which can include more than one microsatellite of a specific type per insert, enrichment factor for the three libraries are as high as 104.7, 212.8 and 180.2.

The two different approaches of defining an enrichment factor yielded considerably different results showing the difficulty in defining this measure as a consequence of hitchhiking repeats.

The density of 343 bp/Mbp *for C. trilobitoides* can be compared to densities found in other genomes analysed. The average genomic density of genomic perfect $(AC)_n$ repeats in arthropods estimated by Toth et al. (2000) is 825 bp/Mbp for perfect repeats >12bp. As density calculations are sensitive to the parameter settings and the algorithms of the search tool (Leclerc et al. 2007) we calculated average densities for the three fully sequenced arthropods *Drosophila melanogaster*, *Apis mellifera* and *Daphnia pulex* using Phobos with identical parameter settings. Genomic densities of $(AC)_n$ repeats were 492, 1115, and 2057 bp/Mbp for *A. mellifera*, *D. pulex*,

and *D. melanogaster* respectively. Consequently, *C. trilobitoides* seems to have a low density of $(AC)_n$ repeats.

Influence of the reporter genomes

The three reporter genomes used in this study differed considerably in their global microsatellite densities, their length characteristics and in the relative frequencies of different repeat classes (Tab. 3, Fig. 3).

As a first result, the global microsatellite density of the reporter genome does not seem to have a significant influence on protocol efficiency: Whereas *D. melanogaster* and *H. sapiens* have a microsatellite density of about half of that of *M. musculus* (Tab. 3), protocol performance was equally high for *M. musculus* and *H. sapiens* (libraries 1 and 3), but was significantly lower for *D. melanogaster* as reporter genome in the library pairs (1,2) (4,5) (6,7) (8,9). This indicates that the number of hybridization sites is sufficient in all three reporter genomes used in this study and that protocol performance is independent of the microsatellite density of reporter genomes in this density range (Tab. 2). In the present data set, protocol efficiency is best explained with genetic distances between target and reporter genome. The smaller the genetic distance between target and reporter species, the more hybridisations occur at loci other than microsatellites, and the lower is the enrichment. The best example is library 2, where a BLAST search revealed 19 target genome inserts of *C. trilobitoides* to be highly similar to conserved regions of the 18S and 28S rDNA regions of the reporter genome (*D. melanogaster*). This was avoided when using DNA of *M. musculus* or *H. sapiens* as reporter genomes indicating that sequence similarity between the two arthropod species *C. trilobitoides* and *D. melanogaster* was still high enough to allow hybridization of homologous loci, thus defying the purpose of the enrichment procedure.

The relative repeat class densities of di- and trinucleotide repeats in the reporter genomes have an influence on the relative repeat class densities in the enriched library. A permutation test confirms that the considerably higher relative dinucleotide repeat density of *M. musculus* as compared to *D. melanogaster* lead to a significantly higher enrichment of this repeat class in all four library pairs with the same target species and these two reporter genomes. The higher relative trinucleotide repeat density of *D. melanogaster* as compared to *M. musculus* lead to a significant trace only in the library pair (4,5), whereas in the other library pairs the data basis is too small to get significant results. Tetra- to hexanucleotide repeat densities were not compared, since their repeat densities were usually high only in very few insert, making a meaningful statistical comparison impossible.

The influence of the reporter genome can even be traced down to individual repeat types. It has been shown that the libraries produced with *M. musculus* as reporter genome contain significantly more $(AG)_n$ repeats, whereas the libraries produced with *D. melanogaster* contain significantly more $(AC)_n$ repeats.

It is difficult to infer exactly how strongly the repeat content of a reporter genome determines the libraries produced with it and whether it is possible to even find an approximate functional relationship. There are at least four main limiting factors here: (i) Hitchhiking of repeats which are only in the library due to their occurrence in the same insert as a repeat that has hybridized against the reporter genome, (ii) biochemical differences of the different repeat types, such as different binding strengths, (iii) the ability of some repeat types to form secondary structures, and (iv) different mean imperfections of repeat types. These limiting factors might make it impossible

to find a desired functional relationship. Ignoring these limiting factors, the probability that a certain repeat type in the target genome hybridisates against the reporter is proportional to the repeat density in each of the two genomes. Thus the expected density of this repeat type in the library is

$$D_{library}(\text{repeat type}) \sim D_{reporter}(\text{repeat type}) * D_{target}(\text{repeat type}).$$

It was tested whether this relationship holds by means of the ratio of the $(AC)_n$ to $(AG)_n$ repeat densities in all libraries. However, this relationship could not be confirmed, since *D. melanogaster* reports much more $(AC)_n$ repeats and *M. musculus* reports much more $(AG)_n$ repeats than expected from this relationship.

Information on the dominant repeat types in target genome

For the genomic libraries created in this study we showed that the repeat content of the reporter genome has a significant influence on the densities in the libraries, even though a direct proportionality relation had to be rejected. This makes it possible to retain at least some information about the dominant repeat types in the target genome. In the enriched libraries (1-3) of *C. trilobitoides* the AG/AC and AG/AT density ratios are significantly higher than in the corresponding reporter genomes. The same holds true for libraries (6-9), indicating that for *C. trilobitoides*, *G. antarcticus* and the bopyrid isopod, the $(AG)_n$ repeat type is likely to be a dominant repeat type. For the two libraries of *S. paradoxa*, however, the results are ambiguous.

From a comparative genomics point of view, the high $(AG)_n$ repeat content has been reported from only few species (e.g. Estoup et al. 1993; Thoren et al. 1995; Tóth 2000; Tóth et al. 2000; Xu et al. 1999). Generally, $(AC)_n$ and $(AT)_n$ dinucleotide repeat types constitute the major fraction of repeats in other crustaceans investigated so far (Katti et al. 2001; Kong and Gao 2005, Mayer, Leese and Tollrian unpublished data; Tassanakajon et al. 1998).

Reduced selection bias of the reporter genome protocol

The reporter genome protocol substitutes synthetic repeat probes with repeats naturally occurring in whole genomes as hybridization templates (Fig. 1). Using the repeat detection and analysis tools Phobos and sat-stat we have identified 476 different and independent di- to hexanucleotide repeat motifs in the reporter genome of *M. musculus*, 454 motifs in *H. sapiens*, and 467 motifs in *D. melanogaster* (Tab. 3). Combining the three reporter genomes, all 499 possible different and independent motifs are available as hybridization probes. Although many motifs are rare in the reporter genomes this still ensures a much less biased search than standard protocols which rarely screen more than 20 different types. Thus, without knowledge about relative densities of certain repeat types, which are commonly dissimilar among different taxa (Tóth et al. 2000), the reporter genome protocol allows the isolation of many different microsatellites that are frequent in the target genome without prior knowledge about repeat motifs (see Tab. 2). This significantly reduces the selection bias for the reporter genome protocol compared to all other protocols and explains it's comparatively high success in detecting as many as 133 different repeat types in the eight genomic libraries screened in this study.

Comments and Recommendations

Choice of reporter genomes

Our study suggests choosing reporter genomes distantly related to the target genomes to avoid hybridization at homologous non-repetitive regions. Using completely sequenced reporter genomes has the advantage that rare but possible contaminations by mobilized reporter genome fragments can be detected by BLAST searches. If a great diversity of different repeat types is desired, it is useful to select at least two reporter genomes that differ in their relative genomic repeat type densities, e.g. *Mus musculus* and *Drosophila melanogaster*.

Microsatellites for population genetic studies

For population genetic studies, highly perfect microsatellites (Fig. 2b, c) are preferred over imperfect, compound or complex microsatellites (Fig. 2d-f). The latter are more likely to have not just one but several sources of length variation and therefore show a higher degree of size homoplasy that can only be reliably assessed by sequencing (Adams et al. 1993; Estoup et al. 2002; Estoup et al. 1995). For population genetic and pedigree analyses, longer microsatellites (> 8 repeat units) are preferred over shorter microsatellites as their higher variability allows a better resolution. Very long microsatellites often suffer from *in vitro* artefacts (Shinde et al. 2003), and are frequently biased in mutation towards shorter repeats (Ellegren 2004; Nauta and Weissing 1996; Wierdl et al. 1997; Xu et al. 2000). This advises using microsatellite markers with < 40 repeat units for analysis, especially for dinucleotide repeats. Besides the structure of a microsatellite itself, a crucial point is the quality of its flanking regions. Frequently, additional structural processes, e.g. inversions, duplications or higher order repeat structures (see Fig. 2 g-f), can be detected in the flanking regions, which may complicate or bias analysis or even lead to the complete failure of PCR amplification. The process of identifying suitable markers from the large amount of microsatellites was greatly facilitated by a software workflow based on the Staden software package, a version of which will soon be available (Bestzeri et al. in prep.) including Phobos and other modules geared towards high throughput establishment of high quality microsatellite loci.

A common feature of enriched libraries is that many microsatllite loci isolated from them have disadvantageous properties and should be excluded for various reasons (Fig. 5). However, the number of microsatellites found in the enriched libraries in this study was so high that a sufficient number of appropriate, high quality loci were available for primer design (see Leese and Held (2008) and Leese et al. (2008) for details on primer design).

Conclusions

The reporter genome protocol is a highly efficient protocol to isolate microsatellites from unknown genomes. Its major advantages are (i) the high enrichment efficiency across a large taxonomic diversity of target genomes, (ii) the large number of different repeat types that can be detected without the need to specify synthetic reporter probes, and (iii) the possibility to retain some information about relative densities of repeat types in the target genome.

The hybridization step is expected to yield maximum numbers of microsatellites if the target and the reporter genomes share similarities in repeat regions only, but are sufficiently dissimilar by descent outside repeat regions. We are confident that the reporter genome protocol as

outlined in this study will make microsatellites more attractive for researchers in a wide variety of fields that previously avoided the development of this powerful marker system.

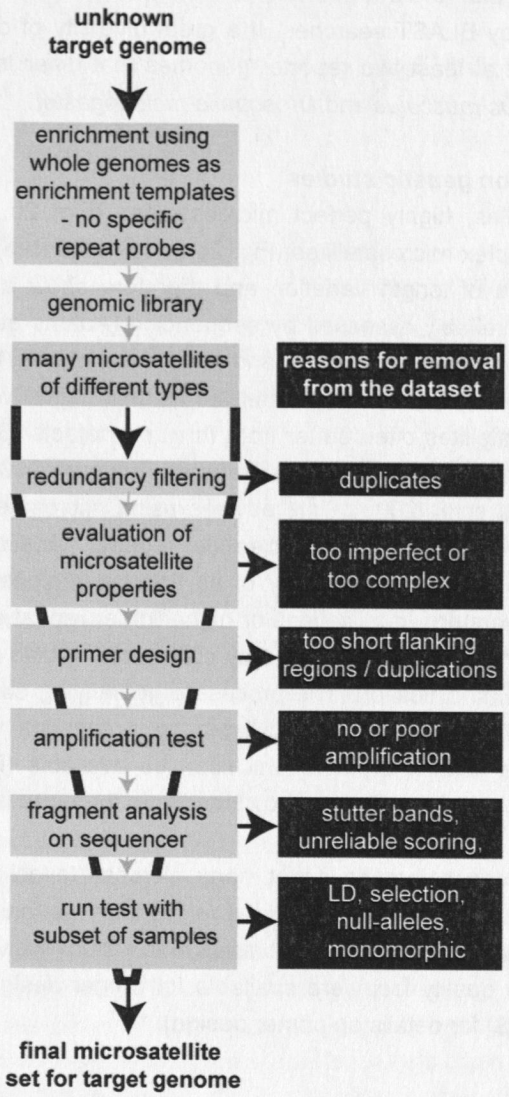


Figure 5: Workflow for isolating microsatellites and evaluating their utility (white boxes). If rigorous standards for quality of microsatellite markers are applied, the number of candidate loci rejected for different reasons (black boxes) often exceeds the number of loci that are eventually kept for the analysis. Only an efficient protocol that generates large numbers of candidate loci will guarantee that the properties of the selected microsatellites for the final data set suit the requirements of the study rather than being imposed by what is left. See Leese et al. (in press), Bestzeri et al. (in prep.) for details.

Acknowledgements: The authors would like to thank Arne Nolte (Université Laval, Canada) for helpful advice and technical support. Tilman Alpermann (AWI Bremerhaven) created the genomic library for *A. tamutum* and *P. multistriata* and gave helpful technical advice. Alec Lüdeking (Napels) kindly provided samples from the diatom *P. multi-striata* Shane Windsor (University of Auckland, New Zealand) made linguistic improvments corrected an earlier version.
Finally, thanks are due to Wolfgang Wägele (Museum Koenig, Bonn) for support and helpful discussions. This work was supported by the Deutsche Forschungsgemeinschaft grant HE 3391/3 to Christoph Held (AWI Bremerhaven) and NSF grant OPP-0132032 to H.W. Detrich (Northeastern University, Boston, USA).

References

Adams, R. P., T. Demeke, and H. A. Abulfaith. 1993. RAPD DNA fingerprints and terpenoids: Clue to past migrations of Juniperus in Arabia and east Africa. Theor Appl Genet **87:** 22-26.

Armour, J. A., R. Neumann, S. Gobert, and A. J. Jeffreys. 1994. Isolation of human simple repeat loci by hybridization selection. Hum Mol Genet **3:** 599-605

Balding, D. 1999. Forensic applications of microsatellite markers, p. 198-210. *In* D. B. Goldstein and C. Schlötterer [eds.], Microsatellites - Evolution and Applications. Oxford University Press.

Beszteri, B. and others in preparation. STAMP: Extensions for the Staden sequence analysis package for automating microsatellite marker design.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455-457.

Dib, C. and others 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature **380:** 152-154.

Ellegren, H. 2004. Microsatellites: Simple sequences with complex evolution. Nat Rev Genet **5:** 435-445.

Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. Mol Ecol **11:** 1591-1604.

Estoup, A., M. Solignac, M. Harry, and J. M. Cornuet. 1993. Characterization of $(GT)_n$ and $(CT)_n$ microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. Nucleic Acids Res **21:** 1427-1431.

Estoup, A., C. Tailliez, J. M. Cornuet, and M. Solignac. 1995. Size homoplasy and mutational processes of interrupted micorsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol Biol Evol **12:** 1074-1084.

Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res **8.**

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res **8:** 175-185.

Glenn, T. C., and N. A. Schable. 2005. Isolating microsatellite DNA loci. Meth Enzymol **395:** 202-222.

Held, C. 2003. Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda), p. 135-139. *In* A. H. L. Huiskes et al. [eds.], Antarctic biology in a global context. Backhuys Publishers, Leiden.

Jones, M. W. and others 2003. Development, characterisation, inheritance, and cross-species utility of American lobster (*Homarus americanus*) microsatellite and mtDNA PCR-RFLP markers. Genome **46:** 59-69.

Junier, T., and M. Pagni. 2000. Dotlet: diagonal plots in a web browser. Bioinformatics **16:** 178-179.

Kandpal, R. P., G. Kandpal, and S. M. Weissman. 1994. Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. Proc Natl Acad Sci U S A **91:** 88-92.

Karagyozov, L., I. D. Kalcheva, and V. M. Chapman. 1993. Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. Nucleic Acids Res **21:** 3911-3912.

Katti, M. V., P. K. Ranjekar, and V. S. Gupta. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol Biol Evol **18:** 1161-1167.

Kijas, J. M., J. C. Fowler, C. A. Garbett, and M. R. Thomas. 1994. Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic beads. BioTechniques **16:** 656-662.

Kong, J., and H. A. Gao. 2005. Analysis of tandem repeats in the genome of Chinese shrimp *Fenneropenaeus chinensis*. Chinese Science Bulletin **50:** 1462-1469.

Korpelainen, H., K. Kostamo, and V. Virtanen. 2007. Microsatellite marker identification using genome screening and restriction-ligation. BioTechniques **42:** 479-486.

Leese, F., and C. Held. in press. Identification and characterization of microsatellites from the Antarctic isopod *Ceratoserolis trilobitoides* - nuclear evidence for cryptic species. Conservation Genetics.

Leese, F., A. Kop, S. Agrawal, and C. Held. in press. Isolation and characterization of microsatellite markers from the marine serolid isopods *Serolis paradoxa* and *Septemserolis septemcarinata* (Crustacea: Peracarida). Mol Ecol Res **8:** 818-821.

Lunt, D. H., W. F. Hutchinson, and G. R. Carvalho. 1999. An efficient method for PCR-based isolation of microsatellite arrays (PIMA). Mol Ecol **8:** 891-893.

Nauta, M. J., and F. J. Weissing. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics **143:** 1021-1032.

Nolte, A. W., K. C. Stemshorn, and D. Tautz. 2005. Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Mol Ecol Notes **5:** 628-636.

Paetkau, D., W. Calvert, I. Stirling, and C. Strobeck. 1995. Microsatellite analysis of population-structure in canadian Polar Bears. Mol Ecol **4:** 347-354.

Rassmann, K., C. Schlötterer, and D. Tautz. 1991. Isolation of simple sequence loci for use in polymerase chain reaction based fingerprinting. Electrophoresis **12:** 113-118.

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. *In* S. Krawetz and S. Misener [eds.], Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press.

Schlötterer, C. 2004. The evolution of molecular markers - just a matter of fashion. Nat Rev Genet **5:** 63-69.

Selkoe, K., and R. J. Toonen. 2006. Microsatellites for Ecologists: A practical guide to using and evaluating microsatellite markers. Ecol Lett **9:** 615-629.

Shinde, D., Y. L. Lai, F. Z. Sun, and N. Arnheim. 2003. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)(n) and (A/T)(n) microsatellites. Nucl Acids Res **31:** 974-980.

Staden, R. 1996. The Staden Sequence Analysis Package. Mol Biotechnol **5:** 233-241.

Tassanakajon, A. and others 1998. Isolation and characterization of microsatellite markers in the black tiger prawn *Penaeus monodon*. Mol Mar Biol Biotech **7:** 55-61.

Thoren, P. A., R. J. Paxton, and A. Estoup. 1995. Unusually high frequency of $(CT)_n$ and $(GT)_n$ microsatellite loci in a yellowjacket wasp, *Vespula rufa* (L.) (Hymenoptera: Vespidae). Insect Mol Biol **4:** 141-148.

Tóth, G. 2000. SSRDB: A database of simple sequence repeats (SSRs) in eukaryotes.

Tóth, G., Z. Gaspari, and J. Jurka. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res **10:** 967.

Vos, P. and others 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res **23:** 4470.

Weber, J. L. 1990. Informativeness of human $(dC\text{-}dA)_n$ - $(dG\text{-}dT)_n$ polymorphisms. Genomics **7:** 524-530.

Weetman, D., A. Ruggiero, S. Mariani, P. W. Shaw, A. R. Lawler, and L. Hauser. 2007. Hierarchical population genetic structure in the commercially exploited shrimp *Crangon crangon* identified by AFLP analysis. Mar Biol (Berl) **151:** 565-575.

Weissenbach, J. and others 1992. A second-generation linkage map of the human genome. Nature **359:** 794-801.

Wierdl, M., M. Dominska, and T. D. Petes. 1997. Microsatellite instability in yeast: Dependence of the length of the microsatellite. Genetics **146:** 769-779.

Xu, X., M. Peng, Z. Fang, and X. P. Xu. 2000. The direction of microsatellite mutations is dependent upon allele length. Nature Genetics **24:** 396-399.

Xu, Z., A. K. Dhar, J. Wyrzykowski, and A. Alcivar-Warren. 1999. Identification of abundant and informative microsatellites from shrimp (*Penaeus monodon*) genome. Anim Genet **30:** 150-156.

Zane, L., L. Bargelloni, and T. Patarnello. 2002. Strategies for microsatellite isolation: a review. Mol Ecol **11:** 1-16.

Zeller, M., T. B. H. Reusch, and W. Lampert. 2006. A comparative population genetic study on calanoid freshwater copepods:Investigation of isolation-by-distance in two *Eudiaptomus* species with a differentpotential for dispersal. Limnol Oceanogr **51:** 117-124.

Chapter 4

STAMP: Extensions for the Staden sequence analysis package for high throughput interactive microsatellite marker design

Bánk Beszteri, Stefanie Gäbler, Christoph Held, Lars Kraemer, **Florian Leese,** Christoph Mayer, Stephan Frickenhaus

Manuscript, submission to BMC Bioinformatics

Abstract

Background: Microsatellites are DNA markers with high analytical power, which are widely used in population genetics, mapping, and forensic studies. Available software for high-throughput marker design has some shortcomings that motivated our work to develop a new tool for this purpose. Our tool is implemented as an extension of the *Staden Package*, which provides the environment for the design of flexible sequence analysis workflows that integrate an extendible set of sequence analysis programs, from base calling, vector and quality clipping, assembly to marker design. The possibility to assemble overlapping reads (also provided by the basic functionality of the Staden package) is important to avoid redundancies in designed markers, a feature missing from most other similar tools.

Results: Our extensions provide the following enhancements to facilitate microsatellite marker design in the Staden Package: The new modules (i) allow to integrate the state-of-the-art repeat detection software Phobos into workflows, (ii) enable separate repeat detection steps for assembling sequencing reads (masking repetitive regions) and for designing repeat-flanking primers, (iii) incorporate the widely used primer design program PRIMER3 into Staden workflows, enabling the interactive design and visualization of flanking primers for microsatellites, and (iv) provide the functionality to find optimal locus- and primer pair combinations for multiplex primer design. Furthermore, our extensions include a module for storing analysis results in an SQLite database, providing a transparent solution for data access from within as well as independent of the Staden Package.

Conclusions: The Staden Package is extended by our modules into a highly flexible, high throughput, interactive tool for conventional and multiplex microsatellite marker design. It gives the user detailed control over the workflow, enabling flexible combinations of manual and automated analysis steps. The software is available under the GNU GPL from www.awi.de/en/go/bioinformatics/software. Our automated marker design workflow has been tested and its high efficiency has been proven in three microsatellite development projects.

Background

Microsatellites (MSs), also called simple sequence repeats (SSRs) or short tandem repeates (STRs), are repetitive genomic sequence stretches with repeat unit lengths of 1-6 nucleotides. The high variability in the number of repeat units of many microsatellites is the basis of their usage as population genetic markers Their high analytical power (multi locus, multiallelic, co-dominant, nuclear, single-copy markers; Goldstein and Pollock 1997, Amos 1999, Schlötterer 2004) made microsatellites one of the most widely used molecular marker systems in the last two decades. Genotyping of MS markers is usually performed by amplification of the repetitive sequence with primers positioned in relatively conserved flanking regions. However, the design of suitable flanking primers, in particular for multiplex amplification, is still a major bottleneck in terms of handling time in MS marker design workflows.

Typical microsatellite marker design workflows should be able/have to deal with data originating from different sources, such as whole genome sequences, sequences from repeat-enriched genomic libraries or even EST (expressed sequence tag) data. The input data format can either be text files (e.g, from sequence databases) or trace files obtained directly from the sequencer. A software tool which allows for a fast microsatellite marker design should ideally combine a high throughput with a high flexibility, enabling the user to work with the whole spectrum of possible data sources. While the software should provide the possibility to manually interact with the analysis process at any point, it should be able to automatically perform a number of basic tasks when working with large genomic data sets or just individual sequence reads (e.g., base calling, vector and quality clipping and assembly in case of *de novo* sequencing as well as repeat detection and flanking primer design). Of great importance is the possibility to manually check and backtrack results of each analysis steps, and to introduce changes manually if required, e.g., shift primer candidates by a few nucleotides based on sequencing quality information. Last but not least, it should provide the functionality to aid the researcher in finding locus or primer combinations for multiplexing. Although a couple of software tools have been developed for automatic MS marker design, MSATCOMMANDER (Faircloth 2008), read2marker (Fukuoka et al. 2005), SSR primer (Robinson et al. 2004), msatfinder (Thurston and Field 2005), we found none that could satisfy all these criteria. Each of these tools had one or more of the following weaknesses: lack of redundancy checks, weak or inconsistent solutions for the repeat detection problem and lack of integration and interactivity. This motivated our own developments presented in this paper. Below we address these issues in more detail.

Only one of the above mentioned pipelines, the TROLL module for the Staden package (Matins et al. 2006), explicitly addresses the problem of marker redundancy: in most cases microsatellite markers are developed based on sequencing genomic shotgun libraries which generally contain redundant sequence fragments. This increases the risk of developing redundant molecular markers when uncritically treating sequenced clones as representatives of different genomic loci, as it has happened in several published studies (e.g., Locus Ca5 AF277577 & Locus Ca10 AF277582 in Dimsoski et al. (2000)). The TROLL module of Martins et al. (2006) addresses this issue by using the assembly functionality provided by the Staden package to detect and assemble overlapping reads into contigs.

Another basic weakness of currently available MS marker design pipelines is repeat detection. While detecting perfect repeats is a relatively straightforward task which is implemented satisfactorily in several tools, detecting imperfect repeats is a more complex

problem and several repeat detection tools perform poorly with them. However, detecting imperfect repeats is crucial in microsatellite design workflows at least for two distinct reasons. First, not just perfect but also imperfect repeats can seriously impede the assembly of sequence databases, especially in the case of sequences coming from repeat-enriched genomic libraries (sequences sharing similar types of repeats and no other sequence similarity can be joined into contigs). Second, imperfect MS markers are more common than perfect ones and often represent the major proportion of the markers reported in Molecular Ecology Notes. Although imperfect MSs may suffer from a higher degree of detectable homoplasy than perfect repeats (Selkoe and Toonen 2006), they are less prone to slippage mutations than perfect MSs (Sturzeneker et al. 1998). From the above cited MS marker design pipelines, only SSR primer (using the repeat detection tool Sputnik) and msatfinder (using Sputnik or an algorithm designed by the authors) can detect imperfect repeats, however only with a limited accuracy.

The novel repeat detection program Phobos (Christoph Mayer, www.rub.de/spezzoo/cm) implements an efficient and highly accurate algorithm to scan molecular sequences for perfect and imperfect tandem repeats without the need to specify a repeat pattern library. Its high accuracy stems from an exact, i.e. non-probabilistic, search algorithm which detects all tandem repeats that obey user defined repeat characteristics. An important feature is its ability to detect alternative (e.g. higher order) repeat pattern/units for the same repeat locus. Search settings and the output format of Phobos can be adjusted very flexible, making it an ideal multi purpose tandem repeat search tool.

Phobos has also been chosen for extending the Staden pipeline since it is significantly more accurate than other tandem repeat detection programs currently available. Typical shortcomings of repeat detection programs are, e.g., a possibly strong dependence of search results on whether a sequence is scanned in forward or backward direction (Sputnik and SciRoKo), the problem of identifying a well defined beginning and end of a tandem repeat (IMex), or an inherent inaccuracy of programs using a probabilistic, i.e. non-exact, search algorithm (TRF) which aim at finding just a specified portion (usually 95%) of repeats that obey predefined repeat characteristics. The accuracy of probabilistic search algorithms is is particularly low when searching for s shorter microsatellites. Furthermore, Phobos is significantly faster than TRF when searching for di- to hexanucleotide tandem repeats.

Another major issue motivating our work was the lack of flexibility and interactivity in most MS marker design tools. Most available MS marker design pipelines implement a pre-defined and mostly strictly linear workflow. However, an interactive workflow control can improve the usability at several steps of the design process. The TROLL module for the Staden package (Martins et al. 2006) demonstrated the power of an approach where a modular sequence analysis software is used as a backbone for a special purpose application, realized by adding some specialized extensions to the basic workflow. Their module integrated the repeat detection tool TROLL (finder.sourceforge.net) into the pregap module of the Staden package, making it possible to define workflows that combine base calling, vector and quality clipping as well as assembly of sequencing reads with repeat detection. Finally, the results from the pipeline were used in a stand-alone primer design program wrapping PRIMER3 (Rozen et al. 2000) for designing flanking primers. Unfortunately, the TROLL module has drawbacks, which impede the MS marker design process significantly: First, the tandem repeat search tool TROLL detects only perfect repeats. Second, the repeat detection step is only invoked once as part of the pre-processing pipeline for sequence assembly, i.e., no post-assembly repeat detection can be performed. A problem with

this approach is that the ideal search parameters for a repeat detection for masking out repetitive sequences before the assembly and for designing flanking primers of repeats after the assembly are often not the same. Whereas the user might be interested, for instance, in only designing primers for dinucleotide repeats, repeats with longer units can equally well impede the assembly process and should thus be masked out as well. Third, the primer design process was implemented as a stand-alone wrapper script for PRIMER3, an unsatisfying solution concerning integration. Thus, the user has no possibility to interactively validate the primers in the context of their contigs or sequencing reads. This step would be necessary to manually check, e.g., sequence coverage and quality in the primer binding region, or to perform multiple rounds of primer design with the aim of incrementally adding primer candidates from these rounds to a list of most promising primer candidates.

In view of the major drawbacks of existing MS marker design tools we decided to develop a set of extension modules for the Staden package (Staden 1996) for a flexible, convenient and efficient development of MS markers. Using the basic workflow of the Staden package as a backbone (data input from multiple sources, base calling, vector and quality clipping, assembly with redundancy checking, interactive contig editor, possibilities to manually and programmatically tag selected regions), we extended its functionality of the following features: (i) integration of the microsatellite search tool Phobos, (ii) independent detection of microsatellites for assembly and marker design with different search parameters, (iii) more flexible integration of repeat detection and primer design steps which can now be performed in an iterative process allowing the user to repeat steps without having to go through the complete workflow again, (iv) design of MS markers optimized for multiplex PCR (Fig. 1).

In order to improve upon the probably most important weakness of the Staden package for such applications, namely the difficult programmatic handling of in-memory Gap4 databases, we also implemented an extension providing the possibility to store sequence features (repeats detected, primer candidates) in a more transparent manner in SQLite (http://www.sqlite.org/) databases (Fig. 1).

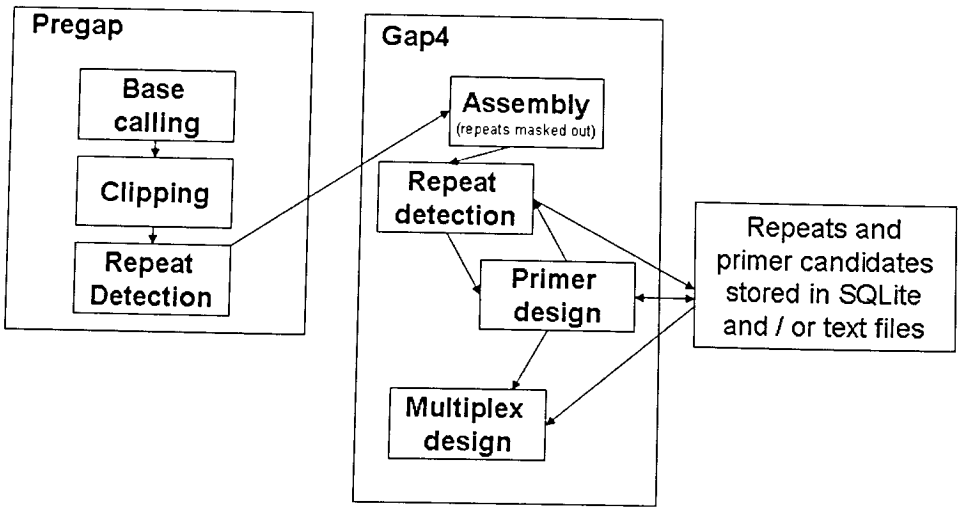


Figure 1: Diagram illustrating possible workflows for microsatellite marker design in the Staden package using the STAMP extensions modules.

Implementation

All extensions described below were implemented as tcl modules following the conventions used in the internal modules of the Staden package.

Phobos module for Pregap

Our first extension is a tcl module integrating the repeat detection tool Phobos into the Staden pre-processing tool pregap. It extends the existing functionality of the pre-processing pipeline, i.e. base calling, quality and vector clipping, filtering of contaminant clones etc. by a tandem repeat detection step. The user has full control over the Phobos search parameters and detected tandem repeats can be filtered according to user preferences in the module. Tandem repeats that pass the filter are added to the experiment files.

Phobos module for Gap4 (Fig. 2)

After the sequence pre-processing step using pregap, the experiment files (which include information on repetitive regions) can be used in an assembly process in the Gap4 module of the Staden package. When using the "Normal shotgun assembly" option, the tagged repeats can be masked out. Either all experiment files or a subset of them can be used as input for the assembly process. After assembly, the individual contigs can be viewed using the contig editor window of Gap4 with tagged regions being marked. If the project started from trace files, the individual traces can be displayed together with the contigs.

On the basis of assembled contigs, a tandem repeats search can be invoked using Phobos with search parameters independent of a previous search (Fig. 2). Selected tandem repeats can be saved in experiment files or an SQLite database, or can be used as targets for designing flanking primers in the following step.

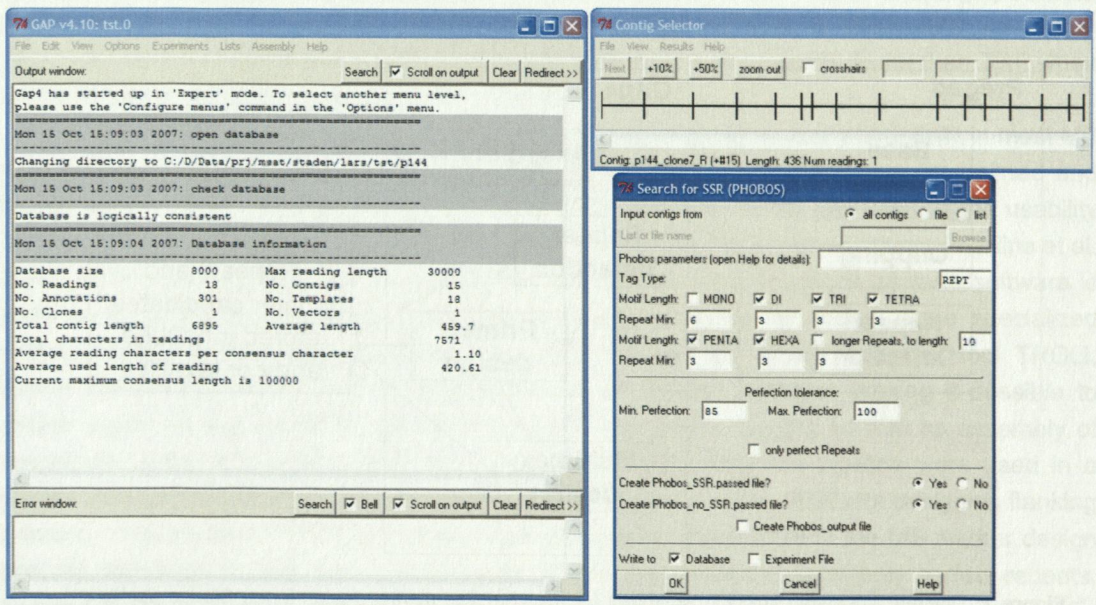


Figure 2: Gap4 dialog to configure microsatellite detection with Phobos.

Module for designing flanking primers with Gap4

This module integrates PRIMER3 (Rozen et al. 2000) into Gap4 and enables the creation of PCR primer pairs flanking any user specified tag in the Gap4 database. Even though our work focuses on designing primers for tagged tandem repeats, the module was developed so as to provide the functionality to design flanking primers for other kinds of features as well. In this module the user is offered a number of options to choose primer target regions and primer design parameters. The user can specify whether the primer pairs should flank just a single tandem repeat or several tandem repeat at once (Fig. 3).

After completion of a PRIMER3 run, a table listing the best five primer pairs for each contig is being displayed. By clicking on rows of this table, an instance of the contig editor is opened showing the contig corresponding to the primer pair that has been tagged (Fig. 4). Check boxes allow the user to manually select or deselect primer pairs for individual contigs (by default, the best primer pair based on PRIMER3 penalties is selected). Subsequently, all checked primer pairs can be written either into the Gap4 database, into experiment files or a SQLite database. Upon completion of manual checks, the list of selected primer pairs can be exported into a tab delimited text file, useful for e.g., ordering primers.

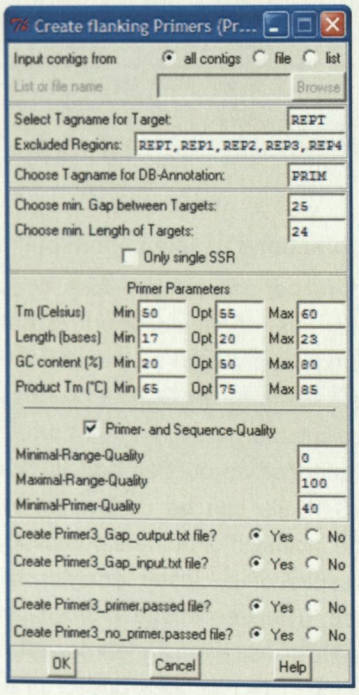


Figure 3: Dialog to start design of flanking primers for user defined tags present in a Gap4 database.

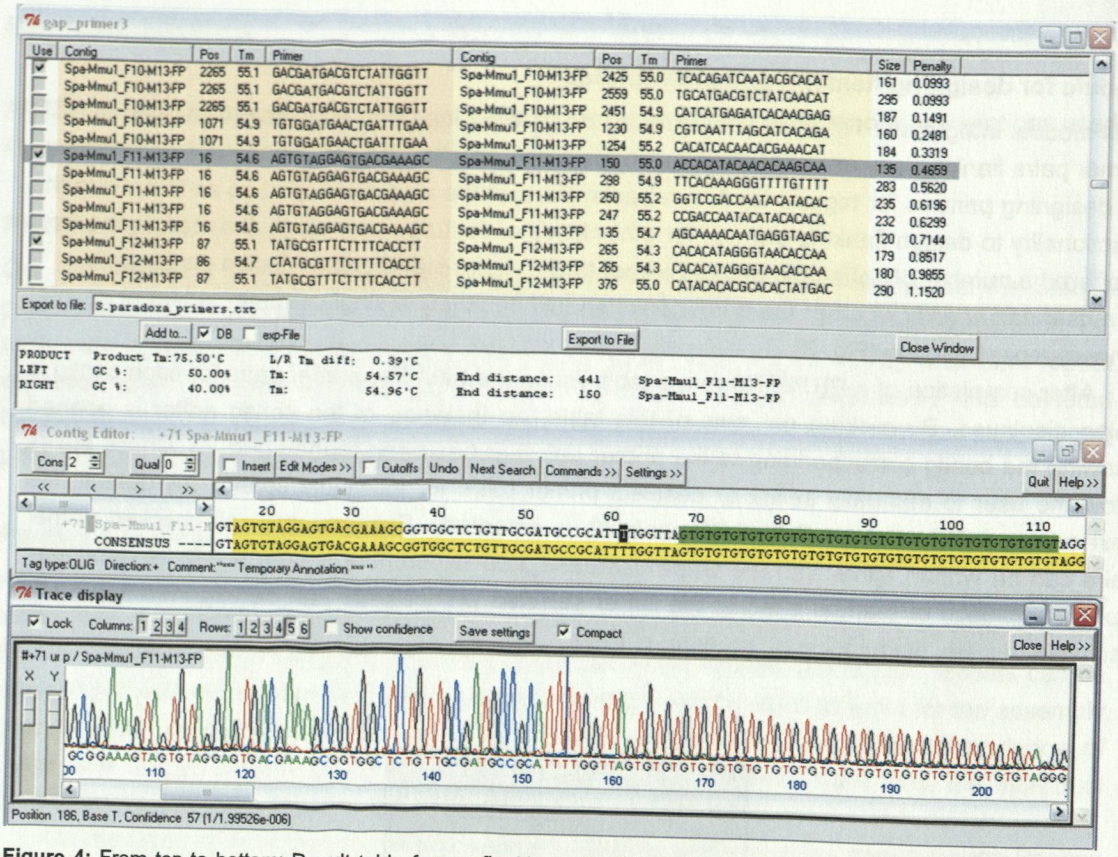
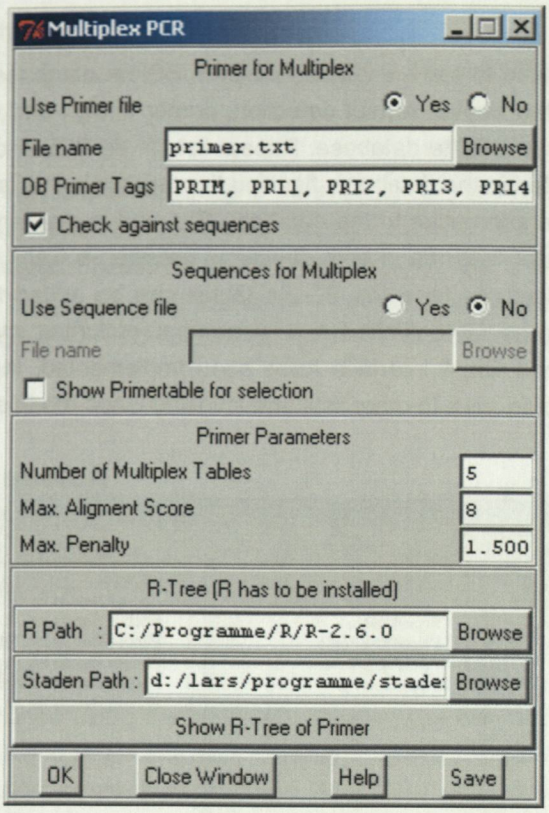


Figure 4: From top to bottom: Result table from a flanking primer design process; contig editor window with tagged tandem repeat (on the right) together with a forward primer candidate (on the left); trace display of the corresponding sequence segment.

Module for selecting primer combinations usable in multiplex PCR experiments for Gap4

This module allows the selection of primer pair combinations, which could be used, in a multiplex PCR experiment. The module uses a list of candidate primer pairs as input, which can read in from tab-delimited text files (including ones written by the above PRIMER3 module), primer tags in the Gap4 database or primer lists from an SQLite database. To create the primer pair combinations, a score matrix is calculated by aligning all primers with each other. In addition, BLAST searches using primer sequences as queries against sequences from either the Gap4 database or a user specified FASTA file can be performed, to check for cross-hybridization among different loci and prevent the inclusion of primers potentially binding to non-target loci. If R (www.r-project.org) is installed, results of a cluster analysis of primer pairs based on their alignment scores can be visualized. The tree image created using R shows the compatibility between different primer pairs: pairs in direct neighborhood are highly compatible, pairs on different branches are less compatible. This display is based on a hierarchical cluster analysis of primer alignment scores, calculated on both the forward and reverse complement strands. The created primer sets can be written as tags into the Gap4 database, experiment files or an SQLite database. It is also possible to export the lists as tab delimited text files (Fig. 5).



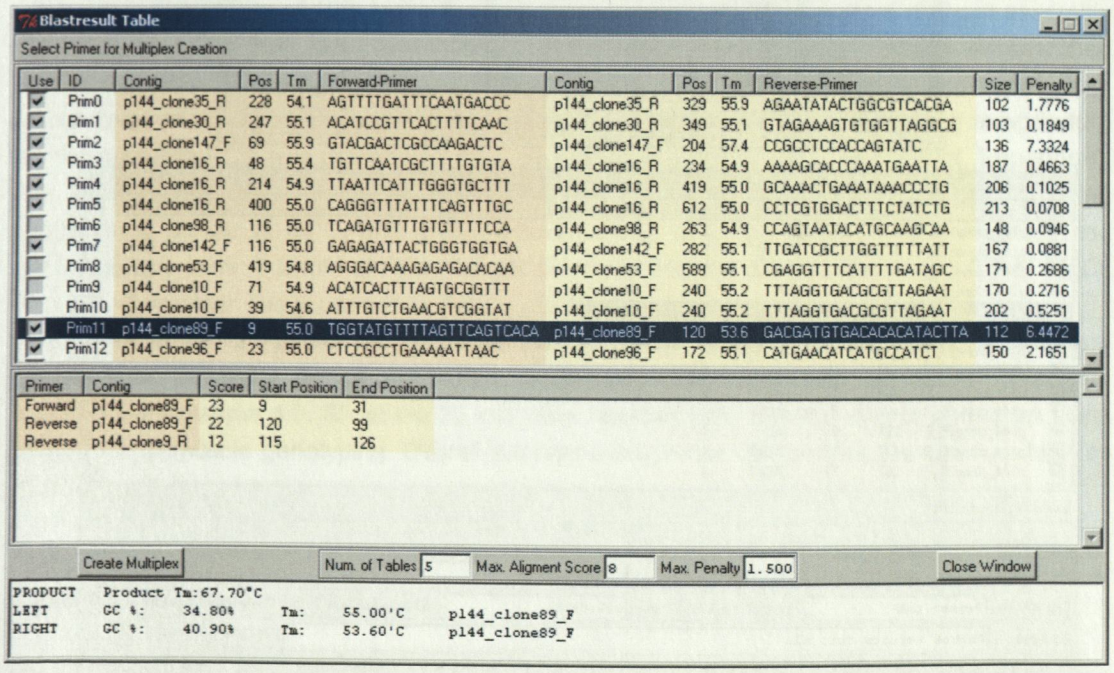


Figure 6: Top: Dialog to start multiplex design. bottom: Table showing similarities between primers and other contigs to detect possible cross-hybridizations.

SQLite module for Gap4

With this module it is possible to store analysis results in SQLite databases. All results created with the other modules listed above (repeat detection, primer design and multiplex design) can (optionally) be written into an SQLite database. Repeat lists in an SQLite database also show a table of primer pairs, included in the database, flanking the repeats in the list (Fig. 6). This adds a transparent feature storage alternative to the database files and in-memory databases of Gap4, with the possibility to transfer features in both directions between a Gap4 database and SQLite databases. Furthermore, features from the SQLite tables can be added to Staden experiment files and can be exported as tab-delimited text. A number of further methods to manipulate SQLite databases through a Gap4 interface have been implemented, including creating new, empty lists, exporting selected rows to other lists and invoking Gap4 modules with the list as input data.

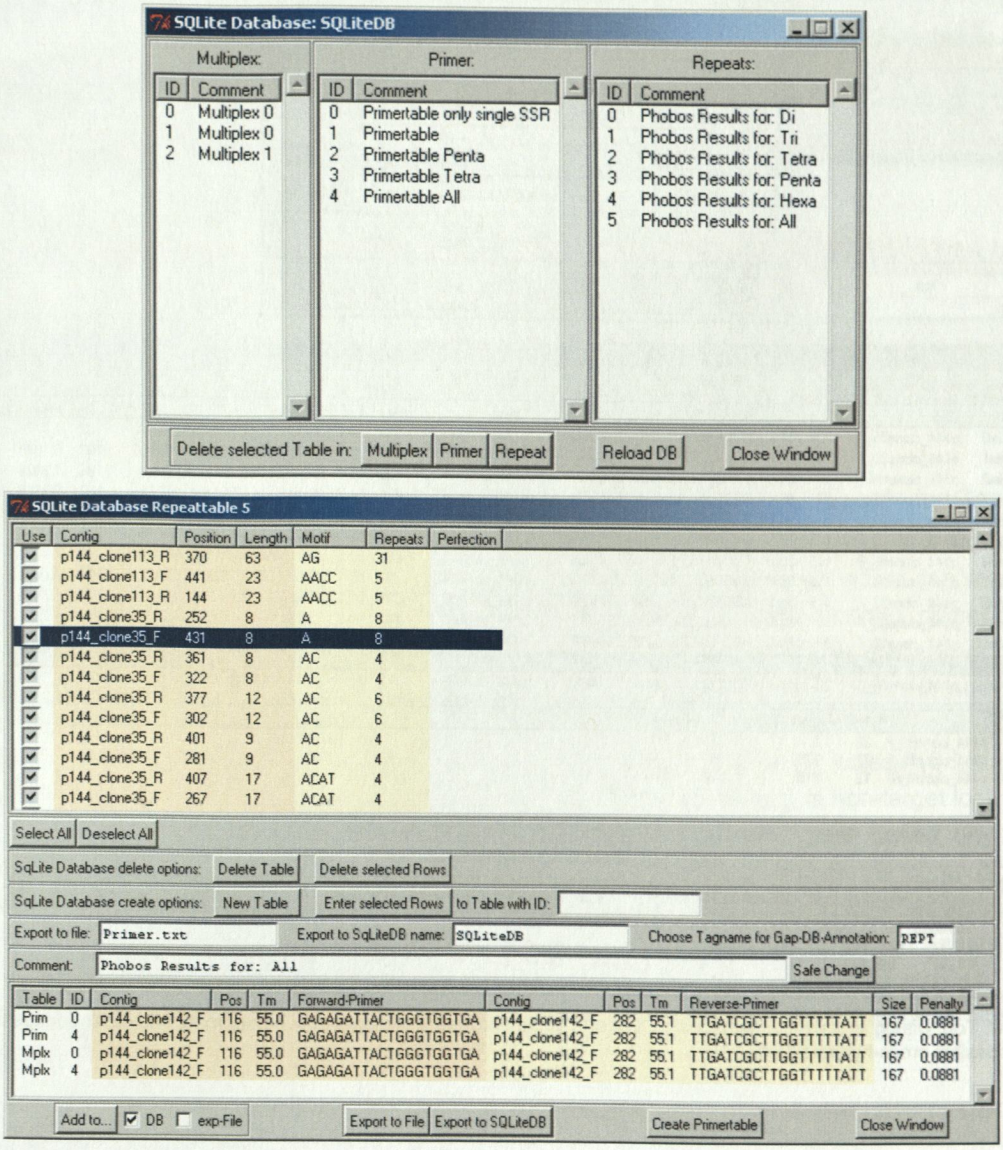


Figure 6: Top: Dialog to load data from SQLite, listing available analysis results. Bottom: Table of tandem repeats and their flanking primers loaded from an SQLite database.

Module for exporting the Gap4 database

This module allows exporting all contigs of a Gap4 database into a FASTA file.

Installation

The STAMP extensions can readily be added to an existing installation of the Staden package by extracting the STAMP archive into the Staden home folder. STAMP archives containing all necessary binaries (Phobos, PRIMER3, NCBI blast, SQLite) are available for the Windows, Linux, and Mac (except SQLite support) platforms. The archives also include documentation for each of the extension modules in PDF and HTML format as well as tutorials.

Results and Discussion

The pipeline was tested in three microsatellite marker design projects, described in detail in Leese et al. (2008), Gäbler et al. (in preparation) and Pöhlmann et al. (in preparation).

Lab test *Serolis paradoxa*

Microsatellites were enriched using the reporter genome protocol (Leese et al. accepted) with DNA from *Drosophila melanogaster* and *Mus musculus* as enrichment templates. A short-insert genomic library was created, of which inserts from 167 clones were sequenced directly. Allowing a mismatch percentage of 10% in the flanking regions using the GAP4 normal shotgun assembly algorithm, we found forty redundant inserts. For the remaining 127 unique inserts we searched for microsatellites with a percentage perfection of ≥95% using Phobos. For 22 appropriate loci, primer pairs were automatically developed using PRIMER3. The optimal annealing temperature for microsatellite PCR was determined on a gradient cycler (48 to 65°C, epgradient, Eppendorf). The PCR protocol for 20 µl reactions was 2 min 94°C, 34 cycles: 20 sec 94°C, 15 sec annealing temperature, 30 sec at 65°C, and a final extension step of 5 min at 65°C. PCR reagents and concentrations were: 0.2 mM dNTPs, 0.5 µM primer (unlabeled), 0.5 M Betaine, 2.5 mM MgCl, 0.03 U/µl Hotmaster *Taq* (Eppendorf), 2-40 ng DNA.

Twenty of the 22 primer combinations yielded distinct PCR products (91%). These were analyzed for variability on an ABI3130xl sequencer. Genotyping was performed using the software *Genemapper* 4.0. Six of the 20 loci were rejected from analysis due to artifactual allele patterns or unreliable genotyping. Overall success from primer design was therefore as high as 63.6%.

Conclusions

We developed the following set of extension modules to the Staden package to facilitate the process of microsatellite marker development:

- a Pregap module integrating the tandem repeat detection program Phobos for masking tandem repeats prior to the assembly,
- a Gap4 module integrating Phobos for detection of suitable microsatellites after assembly,

- Gap4 module integrating PRIMER3 for designing primers which flank tagged sequence regions,
- Gap4 module for multiplex primer design,
- Gap4 module for storing analysis results in SQLite databases.

All workflows can take full advantage of the rich functionality of the basic Staden package including the possibility to import data from a variety of sources, to visualize data as well as results, and to export results into several formats. The major advantages of this sequence analysis environment over other currently available tools that serve the same purpose are the consequent integration of the state-of-the-art repeat detection and flanking primer design tools into a single, flexibly customizable and interactive analysis environment.

Laboratory tests demonstrated that microsatellite marker design with a workflow using the new enhancement modules not only led to a high proportion of useful primer pairs but also to a strong increase in productivity during the design process.

Availability and requirements

Project name: STAMP

Project home page: http://www.awi.de/en/go/bioinformatics/software; http://aforge.awi.de/gf/project/stamp (source code repository)

Operating system(s): Platform independent

Programming language: tcl

Other requirements: none

License: GNU GPL

Any restrictions to use by non-academics: none

References

Amos, W. 1999. A comparative approach to the study of microsatellite evolution, p. 66-79. *In* D.B. Goldstein, and C. Schlötterer [eds.], Microsatellites - Evolution and Applications. Oxford University Press.

Dimsoski, P., G.P. Toth, and M.J. Bagley. 2000. Microsatellite characterization in central stoneroller *Campostoma anomalum* (Pisces: Cyprinidae). Molecular Ecology **9**: 2187-2189.

Faircloth, B. C. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources **8**: 92-94.

Fukuoka, H., T. Nunome, Y. Minamiyama, I. Kono, N. Namiki, and A. Kojima. 2005. read2Marker: a data processing tool for microsatellite marker development from a large data set. BioTechniques **39**: 472-476.

Goldstein, D. B., and D.D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. Journal of Heredity **88**: 335-342.

Leese, F., A. Kop, S. Agrawal, and C. Held. 2008. Isolation and characterization of microsatellite markers from the marine serolid isopods *Serolis paradoxa* and *Septemserolis septemcarinata* (Crustacea: Peracarida). Molecular Ecology Resources **8**: 818-821.

Leese, F., C. Mayer, and C. Held. In press. Enrichment of microsatellites from unknown genomes using knwon genomes as enrichment templates. Limnology and Oceanography: Methods.

Martins, W., D. de Sousa, K. Proite, P. Guimaraes, M. Moretzsohn, and D. Bertioli. 2006. New softwares for automated microsatellite marker development. Nucleic Acids Research **34**: e31.

Nolte, A. W., K.C. Stemshorn, and D. Tautz. 2005. Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Molecular Ecology Notes **5**: 628-636.

Robinson, A. J., C.G. Love, J. Batley, G. Barker, and D. Edwards. 2004. Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics **20**: 1475-1476.

Rozen, S., and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers, p. 365-386. *In* S. Krawetz, and S. Misener [eds.], Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press.

Schlötterer, C. 2004. The evolution of molecular markers - just a matter of fashion? Nature Reviews Genetics **5**: 63-69.

Selkoe, K., and R.J. Toonen. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters **9**: 615-629.

Staden, R. 1996. The Staden Sequence Analysis Package. Molecular Biotechnology **5**: 233-241.

Sturzeneker, R., L.A. Haddad, R.A. Bevilacqua, A.J. Simpson, and S.D. Pena. 1998. Polarity of mutations in tumor-associated microsatellite instability. Human Genetics **102**: 231-235.

Thurston, M. I., and D. Field. 2005. Msatfinder: detection and characterisation of microsatellites. Distributed by the authors at http://www.genomics.ceh.ac.uk/msatfinder/

Chapter 5

Identification and characterization of microsatellites from the Antarctic isopod *Ceratoserolis trilobitoides* – nuclear evidence for cryptic species

Florian Leese and Christoph Held

Conservation Genetics, 2008

Abstract

We report the successful isolation of ten polymorphic microsatellite markers for one species of the marine isopod species complex *Ceratoserolis trilobitoides* (Eights, 1833) from the Southern Ocean. The number of alleles per locus ranged from six to 30 for the 148 specimens analysed and heterozygosity ranged from 0.34 to 0.98. Seven microsatellites amplified successfully in a cryptic sister species, which is restricted to the Antarctic Peninsula. This novel marker set provides the opportunity to study and monitor population structure, demography and gene flow patterns for a benthic model taxon in a region that is now subject to rapid climate change.

Introduction

The species *Ceratoserolis trilobitoides* (Eights, 1833) is a large and locally abundant model invertebrate for ecological, physiological and biogeographic studies on the High Antarctic shelf (e.g. Luxmoore 1981; Clarke 1984; Wägele 1987; Brandt 1991; Clarke & Gore 1992). Recent molecular genetic studies have shown that *C. trilobitoides sensu* Wägele (1986) in fact consists of at least two cryptic species that differ in their distribution ranges, ecological preferences and morphology (Held 2003, Held unpubl. data). One of these species represents the type, *Ceratoserolis trilobitoides sensu stricto* (Eights, 1833, group 2 in Held 2003), which is restricted to the Antarctic Peninsula. The other, broadly distributed species, *Ceratoserolis* n. sp. 1 (group 1 in Held 2003), is new to science (species description Held in prep.).

This study reports the successful development of ten microsatellite markers for *Ceratoserolis* n. sp. 1. In addition, the utility of the microsatellites in the cryptic species *Ceratoserolis trilobitoides s. str.* is tested.

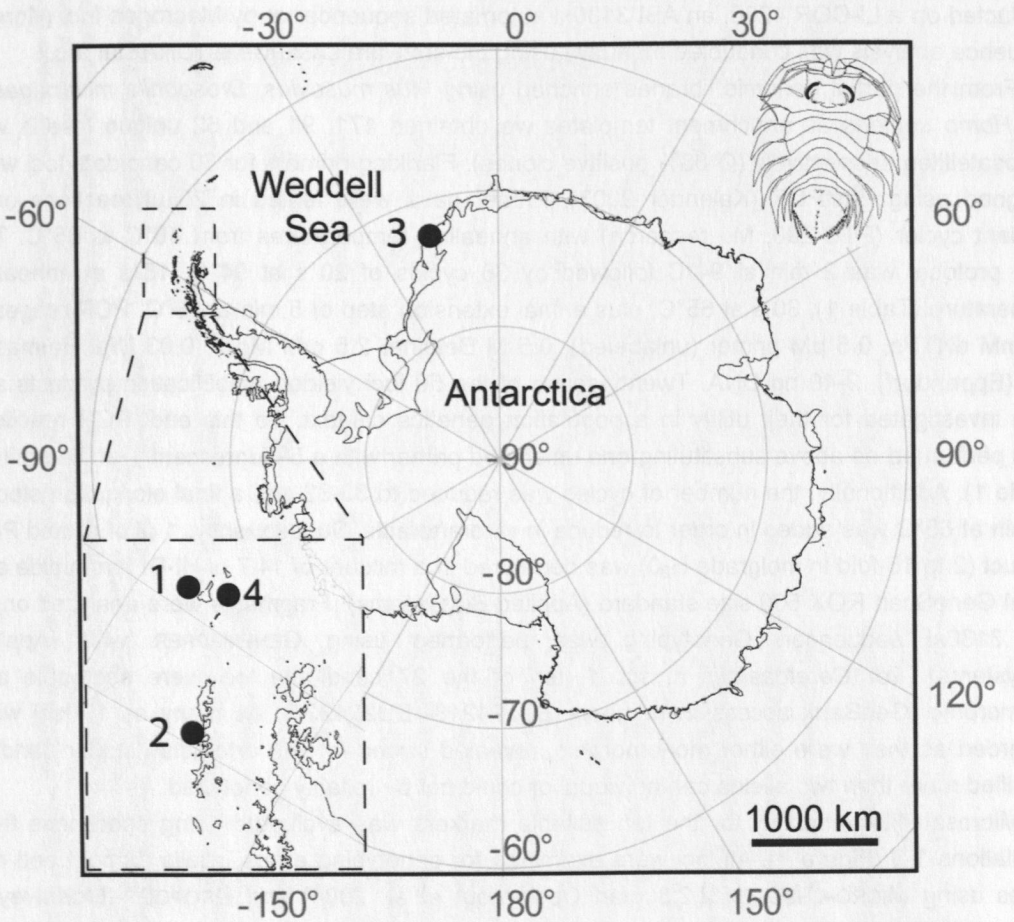


Figure 1: Sampling sites for specimens of *Ceratoserolis trilobitoides sensu* Wägele (1986): The cryptic species *Ceratoserolis* n. sp. 1 was sampled at stations 1-3, specimens of the *Ceratoserolis trilobitoides* s. str. at station 4. Drawing of *Ceratoserolis* n. sp. 1 from CH.

Material and Methods

Material was collected on expeditions ANTXIII/3 and ANTXIV/2 to the Antarctic with "RV Polarstern". Specimens of *Ceratoserolis* n. sp. 1 were collected with Agassiz trawls near Elephant Island (population 1, n=78), King George Island (population 2, n=47) and from the eastern Weddell Sea (population 3, n=33). Specimens of *C. trilobitoides s. str.* were collected near Elephant Island (population 4, n=8, Figure 1).

Genomic DNA of *Ceratoserolis* n. sp. 1 was isolated from ethanol preserved muscle tissue using standard spin column protocols (Qiagen DNeasy). A genomic library enriched for microsatellites was created using the reporter genome protocol (Nolte et al. 2005) as described in Held and Leese (2007). Hybridization chips (Hybond N+, GE Healthcare) with single stranded genomic DNA from *Mus musculus, Drosophila melanogaster* (strain Canton S) and *Homo sapiens* as reporter genomes were used as natural enrichment probes (reporter genomes). Enriched fragments were ligated into pCR2.1 TOPO vectors (Invitrogen) and transformed into *E. coli* (Promega JM 109, Invitrogen TOP10F'). Positive clones were grown overnight in LB media. Alternatively, plasmids were directly amplified using Eppendorf OriMaster Kit. Sequencing was conducted on a LI-COR 4200, an ABI 3130xl automated sequencer or by Macrogen Inc. (Korea). Sequence analysis was conducted manually using the software LASERGENE (DNAStar Inc.).

From the partial genomic libraries enriched using *Mus musculus*, *Drosophila melanogaster* and *Homo sapiens* as enrichment templates we obtained 171, 91 and 62 unique inserts with microsatellites, respectively (Ø 88% positive clones). Flanking primers for 50 candidate loci were designed using FastPCR (Kalender 2003). Primer pairs were tested in 20 µl reactions on a gradient cycler (PTC-200, MJ research) with annealing temperatures from 48°C to 65°C. The PCR protocol was 2 min at 94°C followed by 36 cycles of 20 s at 94°C, 15 s at annealing temperature (Table 1), 30 s at 65°C, plus a final extension step of 5 min at 65°C. PCR reagents: 0.2 mM dNTPs, 0.5 µM primer (unlabeled), 0.5 M Betaine, 2.5 mM MgCl, 0.03 U/µl Hotmaster *Taq* (Eppendorf), 2-40 ng DNA. Twenty-seven of the 50 loci yielded amplification products and were investigated for their utility in a population genetics context. To this end, PCR reactions were performed as above substituting one unlabeled primer with a 5'-fluorescently labeled primer (Table 1). Additionally, the number of cycles was reduced to 30-32 and a final elongation step of 45 min at 65°C was added in order to reduce *in vitro* artefacts. Subsequently, 1 µl of diluted PCR product (2 to 15-fold in molgrade H_20) was denatured in a mixture of 14.7 µl HI-DI formamide and 0.3 µl GeneScan ROX 500 size standard (Applied Biosystems). Fragments were analyzed on an ABI 3130xl sequencer. Genotyping was performed using GENEMAPPER v4.0 (Applied Biosystems). For *Ceratoserolis* n. sp. *1*, ten of the 27 candidate loci were scoreable and polymorphic (GenBank accession numbers EU294218- EU294227). As many as 17 loci were discarded as they were either monomorphic, revealed strong *in vitro* artefacts ('stutter bands'), amplified more than two alleles per individual or could not be reliably genotyped.

Microsatellite variability for the ten suitable markers was evaluated using specimens from populations 1-3 (Figure 1). All loci were examined for genotyping errors, allelic dropout and null alleles using MICRO-CHECKER 2.2.3 (van Oosterhout et al. 2004) and DROPOUT (McKelvey & Schwartz 2005). Tests for Hardy-Weinberg equilibrium (HWE) were performed using ARLEQUIN 3.11 (Excoffier et al. 2005) and for genotypic disequilibrium using GENEPOP 4.0.6 (Rousset 2007). The unbiased probability of identity for all 148 specimens was calculated using GIMLET 1.3.3 (Valiere 2002).

Table 1: Microsatellite loci developed for one cryptic species in the *Ceratoserolis trilobitoides* sensu Wägele (1986) species complex, *Ceratoserolis* n .sp. 1, using the reporter genome protocol: Locus name, primer sequences, 5'-fluorescent dyes, repetitive core sequence, number of alleles identified (N_A), product size range of the identified alleles, annealing temperature (T_a), observed (H_o) and expected (H_e) heterozygosity for the populations from Elephant Island (1), King George Island (2) and the eastern Weddell Sea (3), unbiased probability of identity (PI) and GenBank accession number.

Locus name	Primer sequence (5'-3')	Dye	Repeat motif	N_A	size range [bp]	T_a [°C]	H_o / H_e pop 1	H_o / H_e pop 2	H_o / H_e pop 3	PI (unbiased)	Accession number
C0M03	AGGCATTGCACTGGAGCCAC	FAM	$(CA)_{15}$	6	134-146	60	0.62	0.63	0.58	2.283×10^{-1}	EU294223
	CAGAGGAACCCAGAGGTGTGTG						0.62	0.65	0.53		
C0M05	TTGGTGCAAATGAGAGGTGA	TMR	$(AG)_{27}$	30	176-262	60	0.83*	0.87	0.84	4.649×10^{-3}	EU294227
	CGTCTTTTGCCTGTAGGTCG						0.94	0.95	0.86		
C0M07	GCTATGTCAAACAAAGCTGTTGC	FAM	$(AG)_{21}$	17	226-268	60	0.69**	0.77*	0.88	2.098×10^{-2}	EU294226
	AAAAGTCTCCAAGCTGGGTTC						0.88	0.88	0.83		
C0M11	TCAGTGCGCCTGGTGCCATG	FAM	$(AC)_{20}$	10	174-203	60	0.78	0.91	0.67	6.420×10^{-2}	EU294224
	AATGTCCAGGTTACATACGAAATGT						0.78	0.81	0.71		
C0M12	TGGGACAAAATAATCAAAGCACC	TMR	$(AG)_{16}C(AG)_8$	22	135-177	57	0.92	0.85	0.76*	7.755×10^{-3}	EU294225
	CGGTAGATGGATGAGTGGATGG						0.93	0.92	0.92		
C1D05	TCCAATGAAGTGTTCCGCAGG	FAM	$(AC)_{13}...(CA)_{11}$	11	185-197	57	0.77	0.62	0.73	9.821×10^{-2}	EU294222
	GACCCCACAACAGGACCCCAC						0.73	0.73	0.73		
C1Df14	CTTCTCCGACGCCGCCCAAGAC	FAM	$(AC)_{17}CTA(AC)_9$	10	184-202	63	0.54**	0.64	0.47**	8.711×10^{-2}	EU294221
	AAGGCGTTTTGTCGGCGCTGG						0.71	0.70	0.79		
C1H55	AGCCTCCTGCAGCATCGGAT	HEX	$(AG)_{27}$	23	153-208	63	0.92	0.98	0.76	7.538×10^{-3}	EU294218
	AGTTGCGCCGATGGGTTTGAGG						0.94	0.94	0.88		
C1H93	CCGCGAGGTCCAGATGACGCTCT	HEX	$(TCTG)_{17}$	8	248-260	4x56	0.49*	0.34**	0.63	1.536×10^{-1}	EU294219
	AGAGAGAGAGAGAGACAGAC					30x52	0.63	0.63	0.74		
C3D10	GGTTAGGGACATGACATCGTG	FAM	$(AC)_{11}$	15	127-157	57	0.86	0.74	0.77	2.658×10^{-2}	EU294220
	TGTCGGGTCACTCACGGACTG						0.85	0.84	0.79		

*, ** refer to markers that depart from HWE at $P < 0.05$ and $P < 0.01$, respectively.

Results and Discussion

The number of alleles per locus ranged from 6 to 30. The observed heterozygosity ranged from 0.34 to 0.98. No evidence for global genotypic disequilibrium was found across all pairs of loci across all samples (fisher's exact method). After sequential Bonferroni correction (Rice 1989) only loci C1H93 and C1Df14 deviated from HWE ($P<0.01$) with significantly reduced heterozygosities (Table 1). This inflated homozygosity results most likely from the the presence of null alleles as inferred from analyses using MICRO-CHECKER. The overall probability of identity for the data set investigated (n=148) is high, $PI=2.920 \times 10^{-15}$.

Amplification success of the ten microsatellites was tested in *C. trilobitoides* s. str. from Elephant Island (population 4, Figure 1). Seven microsatellites amplified successfully of which five markers showed private alleles (Table 2). Consequently, eight of the ten microsatellite markers developed in this study add support from the nuclear genome to the hypothesis that *Ceratoserolis* n. sp. 1 and *C. trilobitoides s. str.* represent two reproductively isolated species, as suggested on the basis of mitochondrial (Held 2003) and morphological data (Held in prep.).

Table 2: Cross-amplification success and product length ranges of microsatellite markers in the species *C. trilobitoides* s. str. from Elephant Island (n=8). The length range of the markers in *Ceratoserolis* n. sp. *1* are shown for comparison.

Locus name	amplification success	range [bp] *C. trilobitoides s. str.*	range [bp] *Ceratoserolis* n. sp. *1*
C0M03	-	-	134-146
C0M05	+	218-230	176-262
C0M07	+	301-305	226-268
C0M11	+	167-187	174-203
C0M12	+	121-141	135-177
C1D05	+	185-209	185-197
C1Df14	-	-	184-202
C1H55	+	170-201	153-208
C1H93	-	-	248-260
C3D10	+	129-161	127-157

This marker set provides sufficient resolution for detailed population genetic, phylogeographic and molecular ecological studies on these two benthic key taxa in the Southern Ocean. These findings are especially important to understand the microevolutionary causes of frequent, often cryptic, speciation in benthic taxa inhabiting the High Antarctic shelf (Held 2000, 2003; Held & Wägele 2005; Wilson et al. 2007). Data collected on population structure and demography of these benthic key species is additionally of major importance to study and monitor the consequences of global and regional warming on the Antarctic biota that inhabit one of the world's most affected regions to climate change (Vaughan et al. 2003; Meredith & King 2005).

Acknowledgements: We thank Wolfgang Wägele (Museum Koenig, Bonn, Germany) and Tilman Alpermann (AWI Bremerhaven, Germany) for helpful discussions and Andrea Eschbach (AWI Bremerhaven) for technical assistance. Anna Kop (York University, Toronto, Canda) proofread an earlier version of the manuscript. This work was supported by a DFG grant HE-3391/3 to Christoph Held.

References

Brandt A (1991) Colonization of the Antarctic shelf by the Isopoda (Crustacea, Malacostraca). *Berichte zur Polarforschung*, 98, 1-240.

Clarke A (1984) Lipid composition of two species of *Serolis* (Crustacea, Isopoda) from Antarctica. *British Antarctic Survey Bulletin*, 64, 37-53.

Clarke, A., Gore, D.J. (1992) Egg size and composition in *Ceratoserolis* (Crustacea: Isopoda) from the Weddell Sea. *Polar Biol*, 12, 129-134.

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

Held C (2000) Phylogeny and biogeography of serolid isopods (Crustacea, Isopoda, Serolidae) and the use of ribosomal expansion segments in molecular systematics. *Mol Phylogenet Evol*, 15, 165-178.

Held C (2003) Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda). In: *Antarctic biology in a global context* (eds. Huiskes AHL, Gieskes WWC, Rozema J, Schorno RML, van der Vies SM, Wolff WJ), pp. 135-139. Backhuys Publishers, Leiden.

Held C, Leese F (2007) The utility of fast evolving molecular markers for studying speciation in the Antarctic benthos. *Polar Biol*, 30, 513-521.

Held C, Wägele J-W (2005) Cryptic speciation in the giant Antarctic isopod *Glyptonotus antarcticus* (Isopoda: Valvifera: Chaetiliidae). *Sci Mar*, 69, 175-181.

Kalendar R (2003) FastPCR: a PCR primer and probe design and repeat sequence searching software with additional tools for the manipulation and analysis of DNA and protein. (www.biocenter.helsinki.fi/bi/programs/fastpcr.htm). Institute of Biotechnology, University of Helsinki.

Luxmoore RA (1981) The ecology of Antarctic serolid isopods. pp. 1-231. Council for National Academic Awards.

McKelvey KS, Schwartz MS (2005) DROPOUT: A program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Mol Ecol Notes*, 5, 716-718.

Meredith MP, King JC (2005) Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. *Geophysical Research Letters*, 32, L19604.

Nolte AW, Stemshorn KC, Tautz D (2005) Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. *Mol Ecol Notes*, 5, 628-636.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, 43, 223-225.

Rousset F (2007) GENEPOP '007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Mol Ecol Notes*, Online Early doi:10.1111/j.1471-8286.2007.01931.x.

Valiere N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Mol Ecol Notes*, 2, 377-379.

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

Vaughan DG, Marshall GJ, Connolley WM, Parkinson C, Mulvaney R, Hodgson DA, King JC, Pudsey CJ, Turner J (2003) Recent rapid regional climate warming on the Antarctic Peninsula. *Clim Change*, 60, 243-274.

Wägele J-W (1986) Polymorphism and distribution of *Ceratoserolis trilobitoides* (Eights, 1833) (Crustacea, Isopoda) in the Weddell Sea and synonymy with *C. cornuta* (Studer, 1879). *Polar Biol*, 6, 127-137.

Wägele J-W (1987) On the reproductive biology of *Ceratoserolis trilobitoides* (Crustacea: Isopoda): Latitudinal variation of fecundity and embryonic development. *Polar Biol*, **7**, 11-24.

Wilson, N.G., Hunter, R.L., Lockhart, S.J., Halanych, K.M. (2007) Multiple lineages and absence of panmixia in the "circumpolar" crinoid *Promachocrinus kerguelensis* from the Atlantic sector of Antarctica. *Mar Biol* 152, 895-904.

Chapter 6

Isolation and characterization of microsatellite markers from the marine isopods *Serolis paradoxa* and *Septemserolis septemcarinata* (Crustacea: Peracarida)

Florian Leese, Anna Kop, Shobhit Agrawal, Christoph Held

Molecular Ecology Resources, 2008

Abstract

This study reports the successful isolation of highly informative microsatellite marker sets for two marine serolid isopod species. For *Serolis paradoxa* (Fabricius, 1775), 13, and for *Septemserolis septemcarinata* (Miers, 1875), eight polymorphic microsatellite markers were isolated using the reporter genome enrichment protocol. The number of alleles per locus (N_A) and the observed heterozygosity (H_O) encompass a wide range of variation within *S. paradoxa* (N_A 3-31, H_O 6-89%) and *S. septemcarinata* (N_A 2-18, H_O 9-94%). The suitability of the newly isolated markers for population genetic studies is evaluated.

Introduction

Members of the marine isopod family Serolidae Dana, 1852 are predominately distributed on the continental shelves in the Southern hemisphere (Brandt 1988, Wägele 1994). *Serolis paradoxa* (Fabricius, 1775) is restricted to the marine sublittoral around the Falkland Islands and the Magellan Strait region where it is locally abundant (Gappa and Sueiro, 2007). In contrast, *Septemserolis septemcarinata* (Miers, 1875) is distributed in shallow waters of remote Antarctic islands (Brandt, 1991; see Figure 1). Both species brood their offspring and lack pelagic larval stages. The species' dispersal should thus be limited. However, the amount of gene flow between populations and the species' realized dispersal has never been estimated. This study reports two highly polymorphic microsatellite marker sets that allow to estimate population substructure and gene flow patterns.

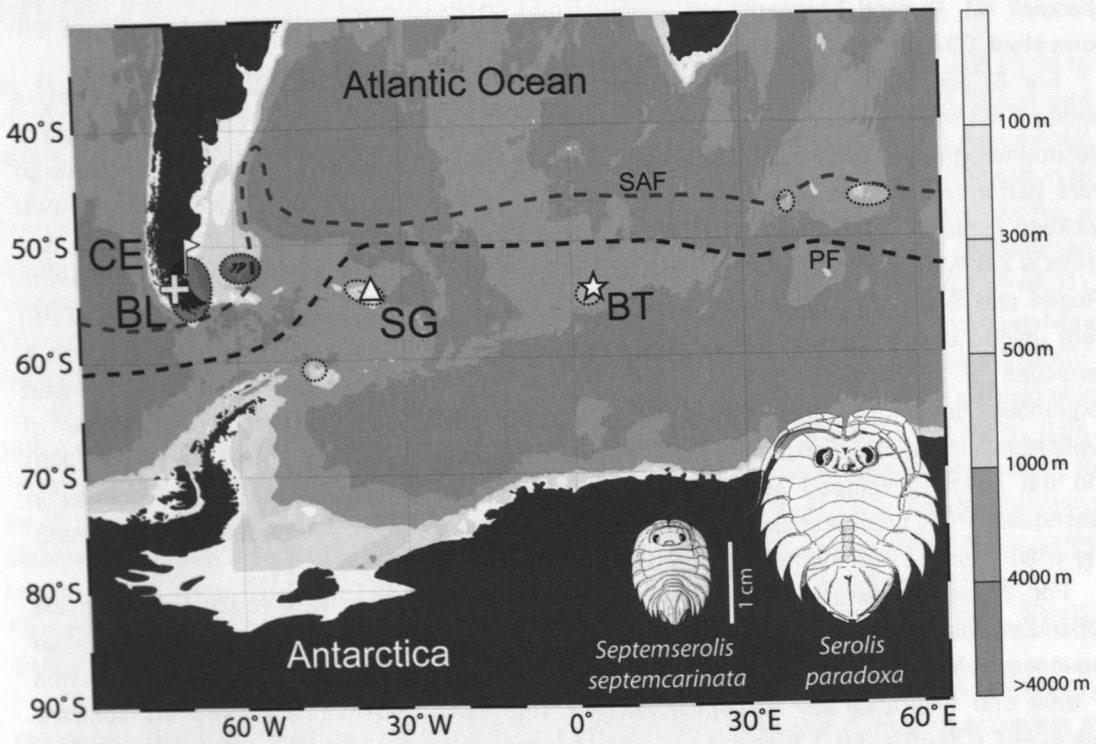


Figure 1: Distribution and sampling sites of *Serolis paradoxa* (dark grey circles, BL and CE) and *Septemserolis septemcarinata* (bright grey circles, SG and BT). Illustrations of *S.paradoxa* from Wägele (1994), of *S. septemcarinata* from Brandt (1991). SAF=Subantarctic Front, PF=Polar Front, according to Belkin and Gordon (1996) and Cortese and Gersonde (2007).

Material and methods

Specimens of *S. paradoxa* were sampled at Bahia Laredo near Punta Arenas, Chile (BL) and from the Atlantic opening of the Strait of Magellan (CE). Specimens of *S. septemcarinata* were sampled around South Georgia (SG) and Bouvetoya (BT, Figure 1). Genomic DNA was isolated from muscle tissue using the Qiagen DNeasy Mini Kit. A genomic library enriched for microsatellites was created for each species using the reporter genome protocol (Nolte et al., 2005) as described in Held and Leese (2007). Hybridization chips (Hybond N+, GE Healthcare) with DNA from *Mus musculus* and *Drosophila melanogaster* (Canton S) as reporter genomes were used for enrichment. As a modification to Held and Leese (2007), 0.03 U/µl Hotmaster *Taq* (Eppendorf) were used in PCR. Also, nick repair and PCR were carried out in one reaction tube by incubating for 10 min at 65°C prior to PCR (94°C for 2 min followed by 25 cycles of 30 s at 94°C, 45 s at 52°C, 80 s at 65°C and 10 min final elongation at 65°C). For elution, hybridization chips were transferred into 500 µl TE buffer (pH 8.0, 80°C) for 5 min. DNA was precipitated using a standard isopropanol-sodiumacetate protocol.

The enriched fragments were PCR amplified in 25 µl reactions and purified using the Qiagen Qiaquick Kit. Purified fragments were cloned into pCR2.1-TOPO vector and transformed into competent TOP10F' *E.coli* (Invitrogen).

For *S. paradoxa*, plasmid preparation of 167 colonies and shotgun sequencing using a standard M13-forward primer was conducted by GATC-Biotech (Konstanz, Germany). Analysis of electropherograms, vector clipping, assembly of contigs, redundancy filtering and primer design were performed using a newly developed, automated software pipeline based on the STADEN package (Staden, 1996, Beszteri et al., in preparation) to which the microsatellite search tool PHOBOS 3.0 (Mayer, in preparation, www.rub.de/spezzoo/cm) and the primer design tool PRIMER3 (Rozen and Skaletsky, 2000) had been added. Out of the 167 inserts sequenced, 124 (74%) were unique and contained at least one microsatellite. Only inserts with microsatellites with a perfection of ≥95% were chosen using PHOBOS and considered for primer design. Insert sequences outside the microsatellites were additionally screened for the presence of duplications, inversions and higher-order repeat structures using DOTLET (Junier and Pagni, 2000) and the newly developed software SEROLIS (Schulz et al. in prep) to avoid problems in subsequent PCR amplification. Primer pairs for 22 adequate inserts were devised by PRIMER3. The entire process was made highly automated due to the software pipeline.

For *S. septemcarinata* plasmid preparation of 161 colonies was conducted using the Eppendorf Fast Plasmid Mini Kit. All 161 inserts were sequenced on an ABI 3130xl automated sequencer using M13-forward and reverse primers. A total of 103 inserts (64%) were unique with at least one microsatellite. Sequence analysis was conducted manually using the software LASERGENE (DNAStar Inc.). PHOBOS and DOTLET were used to select 16 adequate microsatellites from the 103 sequences. Primer pairs were designed using FASTPCR (Kalender, 2003).

For both species, the optimal annealing temperature for microsatellite PCR was determined on a gradient from 48°C to 65°C. The PCR protocol for 15 µl or 20 µl reactions was 2 min at 94°C followed by 34 cycles of 20 sec at 94°C, 15 sec at annealing temperature (Table 1a/b), 30 sec at 65°C, plus a final extension step of 5 min at 65°C. PCR reagents consisted of 0.2mM dNTPs, 0.5 µM primer (unlabeled), 0.5M Betaine, 2.5 mM MgCl, 0.03U/µl Hotmaster *Taq* (Eppendorf), 2-40

ng DNA.

For *S. paradoxa*, 21 of 22 and for *S. septemcarinata*, 14 of the 16 primer combinations yielded distinct PCR products. Microsatellite variability for *S. paradoxa* was evaluated using specimens from populations BL (n=35) and CE (n=32) and for *S. septemcarinata* using specimens from SG (n=23) and BT (n=52). PCR reactions were repeated substituting one unlabeled primer with a 5'-fluorescently labeled primer (Table 1a/b), reducing the number of cycles to 28-34 plus adding a final elongation step of 45 minutes at 65°C. The denatured PCR products were analyzed on an ABI 3130xl sequencer using ROX GS500 size standard (ABI). Genotyping implemented the software GENEMAPPER 4.0. The data sets were examined for genotyping errors, allelic dropout and null alleles using MICROCHECKER 2.2.3 (van Oosterhout et al., 2004). Tests for HWE and genotypic disequilibrium were performed using GENEPOP 4.0.6 (Rousset, 2007) and ARLEQUIN 3.11 (Excoffier et al., 2005). The unbiased probability of identity was calculated using GIMLET 1.3.3 (Valiere, 2002).

Results and Discussion

For *S. paradoxa*, 15 loci could be reliably genotyped of which 13 were polymorphic with three to 31 alleles per locus. The observed heterozygosities ranged from 0.059 to 0.89. Loci Spa04 and Spa41 (BL), Spa08, Spa10, Spa30 and Spa39 (BL and CE), and Spa34 (CE) were characterized by a significant global heterozygosity deficiency ($P<0.01$). Although this might be an indication of null alleles, it needs to be considered whether it could be a consequence of local inbreeding, a Wahlund effect, a sampling bias or of recent population expansion in the Strait of Magellan instead. A significant genotypic disequilibrium was reported for the loci Spa10/Spa41 ($P<0.01$). The probability of identity (excluding locus Spa10) was high, $PI=2.214x10^{-17}$. Two microsatellites were monomorphic for the populations investigated: Spa01 (EU127455) and Spa16 (EU127456).

For *S. septemcarinata* 13 loci could be reliably genotyped of which eight were polymorphic with two to 18 alleles per locus. The observed heterozygosities ranged from 0.09 to 0.94. Population SG displayed a highly significant homozygosity excess for locus Sse14 that is likely to be due to null alleles. None of the other loci deviated from HWE. Significant genotypic disequilibrium was reported for loci Sse14/Sse15 ($P=0.025$), which are located on the same insert, 131 bp apart. The probability of identity (excluding locus Sse14) was high, $PI=5.689x10^{-10}$. Five additional microsatellites were monomorphic for the populations investigated: Sse01, Sse02, Sse06, Sse12, Sse16 (EU056267, EU056268, EU056271, EU056275, EU056279).

The novel marker sets reported in this study are appropriate for studying microevolutionary processes, especially gene flow, to quantify the species' dispersal capabilities in the context of their unique biology and habitat characteristics. In addition, the markers can provide insight into the poorly understood reproductive strategies of these two benthic key species.

Table 1a: Microsatellite loci for *Serolis paradoxa*: Locus name, primer sequences, 5'-fluorescent dyes, repetitive sequence, number of alleles identified (N_A), product size range of the identified alleles, annealing temperature (T_a), observed (H_o) and expected (H_e) heterozygosity for the populations from Bahia Laredo (BL) and the Opening of the Strait of Magellan to the Atlantic Ocean (CE), unbiased probability of identity (PI) and GenBank accession numbers.
*, **refer to markers that depart from HWE at $P < 0.05$ and $P < 0.01$, respectively.

Species	Locus name	Primer sequence (5'-3')	Dye	Repeat motif	N_A	size range [bp]	T_a [°C]	H_o / H_e for populations	PI (unbiased)	Accession number
S. paradoxa	Spa04	F:gagcttacgaacaaaactgc R:cgtctccaaccttacttcag	HEX	$(CA)_9$	5	126-140	62	BL: 0.47 / 0.68** CE: 0.47 / 0.44	$1.594x10^{-1}$	EU127468
S. paradoxa	Spa07	F:tgtctgtctgttggtcgata R:aagcaaacaggcagtctaac	6FAM	$(TGTC)_5$	3	109-117	62	BL: 0.059 / 0.058 CE: monomorphic	$9.368 x10^{-1}$	EU127457
S. paradoxa	Spa08	F:aagataatccagaaggcgta R:gcagtgcttctttctctgtt	HEX	$(AGTG)_{15}$	16	264-332	55	BL: 0.32 / 0.90** CE: 0.13 / 0.92**	$1.700 x10^{-2}$	EU127458
S. paradoxa	Spa10	F:tgttttggtgatactgacga R:agtgtaggagtgacgaaagc	NED	$(AC)_{23}$	24	256-318	55	BL: 0.71 / 0.95** CE: 0.78 / 0.95**	$3.496 x10^{-3}$	EU127459
S. paradoxa	Spa12	F:caaatccaaaaaggaatctg R:ttccttctgttcgttcattt	HEX	$(AC)_{18}$	6	188-202	55	BL: 0.26 / 0.30 CE: 0.13 / 0.12	$2.885 x10^{-1}$	EU127460
S. paradoxa	Spa13	F:tcctcaaagaatttcacgtt R:gcatttttcttcaagtgtcc	6FAM	$(CA)_{25}$	6	153-181	60	BL: 0.50 / 0.71 CE: 0.42 / 0.54	$2.172 x10^{-1}$	EU180576
S. paradoxa	Spa30	F:aggtagccccactcatttac R:agtgttgttcaatgcacgta	HEX	$(AC)_{23}AA(AC)_3$	31	155-198	62	BL: 0.80 / 0.94* CE: 0.63 / 0.95**	$2.192 x10^{-3}$	EU127461
S. paradoxa	Spa34	F:ctcccaaaaagtagcacatc R:agaaagggatcagcgaata	NED	$(AC)_{23}$	20	145-191	60	BL: 0.88 / 0.92 CE: 0.66 / 0.74*	$1.373 x10^{-2}$	EU127462
S. paradoxa	Spa35	F:tatttgcctgtgcatgttta R:atgatctgagtgtgcgtgt	HEX	$(CA)_8$	10	227-255	62	BL: 0.68 / 0.67 CE: 0.63 / 0.73	$6.910 x10^{-2}$	EU127463
S. paradoxa	Spa39	F:tgtctcgaacgagaaactct R:gtgtgcaagtgtatcgatgt	NED	$(ACAG)_{20}$	19	172-256	62	BL: 0.63 / 0.92** CE: 0.69 / 0.89**	$7.899 x10^{-3}$	EU127464
S. paradoxa	Spa41	F:agtgtaggagtgacgaaagc R:accacatacaacacaagcaa	6FAM	$(GT)_{22}$	28	120-280	62	BL: 0.73 / 0.95** CE: 0.87 0.95	$3.292 x10^{-3}$	EU127465
S. paradoxa	Spa42	F:tatgcgtttcttttcacctt R:cacacatagggtaacaccaa	NED	$(GT)_{22}$ AGG$(GT)_6$/ $(TG)_4CG(TG)_5$	20	160-208	55	BL: 0.89 / 0.91 CE: 0.78 / 0.89	$1.205 x10^{-2}$	EU127466
S. paradoxa	Spa43	F:gagggaaggaaagaatgaat R:gtttaggtcctcctctggtc	HEX	$(GAAT)_3/(AGA)_3$/ $(TGAA)_4$	4	174-182	59	BL: 0.40 / 0.39 CE: 0.38 / 0.48	$2.136 x10^{-1}$	EU127467

*, **refer to markers that depart from HWE at $P < 0.05$ and $P < 0.01$, respectively.

Table 1b: Microsatellite loci for *Septemserolis septemcarinata*: Locus name, primer sequences, 5'-fluorescent dyes, repetitive sequence, number of alleles identified (N_A), product size range of the identified alleles, annealing temperature (T_a), observed (H_o) and expected (H_e) heterozygosity for the populations from South Georgia (SG) and Bouvetoya (BT), unbiased probability of identity (PI) and GenBank accession numbers.

Species	Locus name	Primer sequence (5'-3')	Dye	Repeat motif	N_A	size range [bp]	T_a [°C]	H_o / H_e for populations	PI (unbiased)	Accession number
S. septemcarinata	Sse04	F:tatttgtgtcggcgtgtg R: tccacgtgcaagtaggcggt	6FAM	$(AT)_7$	12	230-258	65	SG: 0.48 / 0.56 BT: 0.73 / 0.80	4.106×10^{-2}	EU056269
S. septemcarinata	Sse05	F:agcacaagcgcttagagggtccag R:agtacgtctagagctagcaagtgtg	6FAM	$(CT)_9$	2	215-217	63	SG: 0.09 / 0.09 BT: 0.25 / 0.25	6.497×10^{-1}	EU056270
S. septemcarinata	Sse07	F:acgcgtgattcactggcagagttc R: agattcggccaagcggctgttc	HEX	$(ATT)_6$	4	210-223	65	SG: 0.39 / 0.37 BT: 0.42 / 0.43	3.557×10^{-1}	EU056272
S. septemcarinata	Sse08	F: tcgaaagtcgaattgcgtgtg R: agaaaccgcccagagtgg	HEX	$(AG)_{42}$	16	215-254	65	SG: 0.70 / 0.68 BT: 0.83 / 0.86	2.873×10^{-2}	EU056273
S. septemcarinata	Sse10	F:gccccaacacaatatggaggctgtg R:agaaggccgtgacatcggttaggg	HEX	$(GT)_4AT(GT)_{21}$	18	155-203	60	SG: 0.70 / 0.88 BT: 0.94 / 0.93	5.388×10^{-3}	EU056273
S. septemcarinata	Sse13	F:tcttgacagggtggagcgcaaacc R:ggcagcgagcctagtgcctcgattc	NED	$(AG)_{26}AT(AG)_4$	11	188-212	65	SG: 0.87 / 0.85 BT: 0.85 / 0.83	2.774×10^{-2}	EU056276
S. septemcarinata	Sse14	F:ggtctaagggtagatgactcgaccg R:ggcgattctactggtgccgccatca	NED	$(AC)_8CT(AC)_{39}$	13	263-288	65	SG: 0.22 / 0.73** BT: 0.83 / 0.85	2.323×10^{-2}	EU056277
S. septemcarinata	Sse15	F:tggcggcaccagtagagtcgccatg R:acggtgacgcagtggggcttcgag	NED	$(AG)_{11}$	15	126-182	60	SG: 0.78 / 0.80 BT: 0.83 / 0.86	1.396×10^{-2}	EU056277

*, **refer to markers that depart from HWE at $P < 0.05$ and $P < 0.01$, respectively.

Acknowledgements: We thank Erika Mutschke and Carlos Rios (Universidad de Magallanes, Punta Arenas) for providing us with material from the 2nd Cruce estrecho, 2003. This work was supported by a DFG grant HE-3391/3 to CH, NSF grant OPP-0132032 to H.W. Detrich, and a DAAD scholarship to FL and AK. This is publication number 20 from the ICEFISH Cruise of 2004.

References

BELKIN, I. M. & GORDON, A. L. (1996) Southern Ocean fronts from Greenich meridian to Tasmania. *Journal of Geophysical Research,* 101, 3675-3696.

BESZTERI, B., FRICKENHAUS, S., GAEBLER, S., HELD, C., KRAEMER, L., LEESE, F. & MAYER, C. (in preparation) Extensions for the Staden sequence analysis package for automating microsatellite marker design.

BRANDT, A. (1988) Antarctic Serolidae and Cirolanidae (Crustacea: Isopoda): New genera, new species and redescriptions. Theses Zoologicae. IN FRICKE, R. (Ed.) 10 ed., Koeltz Scientific Books.

BRANDT, A. (1991) Colonization of the Antarctic shelf by the Isopoda (Crustacea, Malacostraca). *Berichte zur Polarforschung,* 98, 1-240.

CORTESE, G. & GERSONDE, R. (2007) Morphometric variability in the diatom *Fragilariopsis kerguelensis*: Implications for Southern Ocean paleoceanography. *Earth and Planetary Science Letters,* 257, 536-544.

EXCOFFIER, L., LAVAL, G. & SCHNEIDER, S. (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online,* 1, 47-50.

GAPPA, L. & SUEIRO, M. (2007) The subtidal macrobenthic assemblages of Bahia San Sebastian (Tierra del Fuego, Argentina). *Polar Biology,* 30, 679-687.

HELD, C. & LEESE, F. (2007) The utility of fast evolving molecular markers for studying speciation in the Antarctic benthos. *Polar Biology,* 30, 513-521.

JUNIER, T. & PAGNI, M. (2000) Dotlet: diagonal plots in a web browser. *Bioinformatics,* 16, 178-179.

KALENDAR, R. (2003) FastPCR: a PCR primer and probe design and repeat sequence searching software with additional tools for the manipulation and analysis of DNA and protein. (www.biocenter.helsinki.fi/bi/programs/fastpcr.htm)

MAYER, C. (in preparation) Phobos: Detection of perfect and imperfect tandem repeats.

NOLTE, A. W., STEMSHORN, K. C. & TAUTZ, D. (2005) Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. *Molecular Ecology Notes,* 5, 628-636.

ROUSSET, F. (2007) GENEPOP '007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Notes,* online first.

ROZEN, S. & SKALETSKY, H. J. (2000) Primer3 on the WWW for general users and for biologist programmers. IN KRAWETZ, S. & MISENER, S. (Eds.) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Totowa, NJ, Humana Press.

STADEN, R. (1996) The Staden Sequence Analysis Package. *Molecular Biotechnology,* 5, 233-241.

VALIERE, N. (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes,* 2, 377-379.

VAN OOSTERHOUT, C., HUTCHINSON, W. F., WILLS, D. P. M. & SHIPLEY, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology,* 4 535-538.

WÄGELE, J.-W. (1994) Notes on Antarctic and South American Serolidae (Crustacea, Isopoda) with remarks on the phylogenetic biogeography and a description of new genera. *Zoologische Jahrbücher Abteilung für Systematik Ökologie und Geographie der Tiere,* 121, 3-69.

Chapter 7

Population differentiation and historical demography of *Ceratoserolis* n. sp. 1 from the Antarctic shelf: Evidence for survival in independent refugia during last glacial maximum

Florian Leese and Christoph Held

Manuscript

Abstract

Benthic habitats on the Antarctic shelf are largely ice-free at present but were covered with grounded ice that reached several times as far as the shelf break during the glacial maxima. The strikingly high biodiversity on the Antarctic shelf has often been taken as indirect evidence that not all benthic life may have been eradicated during the glacial maxima but there is a lack of hard evidence supporting this view.

In this study, we investigate ten fast evolving microsatellite loci and a fragment of the mitochondrial COI gene in the broadly distributed benthic isopod *Ceratoserolis* n. sp. 1 from the Antarctic Peninsula (AP) and the Eastern Weddell Sea (EWS). The patterns of intraspecific genetic variation in the two marker systems are concordant with a deep genetic divergence between both regions on either side of the Weddell Sea indicating a lack of genetic exchange since the late Pleistocene. There is also a signature of rapid and recent population expansion in AP and EWS, which coincides with a recovery of populations after the last glacial maximum (LGM). The molecular data demonstrate that glaciations had a pronounced impact on genetic diversity but also suggest that continued survival in separate refuge areas in the EWS and around the AP was possible.

Abbreviations

AMOVA	Analysis of Molecular Variance
AP	Antarctic Peninsula
CI	Confidence Interval
COI	Cytochrome oxidase subunit I gene
EWS	Eastern Weddell Sea
FCA	Factorial Component Analysis
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HT	Haplotype
IAM	Infinite Allele Model
IBD	Isolation by distance
Ln Pr	Estimated logarithmic probability of the data
N_A	Number of alleles
N_e	Effective population size
P1-13	Population 1-13
SMM	Stepwise Mutation Model
TPM	Two Phase Model

Introduction

The fauna of the Southern Ocean is highly diverse and endemic (Clarke and Crame 1989, Dayton 1990, Arntz et al. 1994, Knox 1994, Crame 1997, Clarke and Johnston 2003, Clarke et al. 2004, Gutt et al. 2004). Brooding benthic macroinvertebrates of the taxon groups echinoderms, molluscs, isopods and amphipods are particular diverse (Brandt 1988, De Broyer and Jazdzewski 1993, Arntz et al. 1994, Wägele 1994, Poulin and Feral 1996, Brandt et al. 1999, Clarke and Johnston 2003, Gutt et al. 2004, Linse et al. 2006). Currently, approximately 4000 species are known to coexist on the Antarctic shelf (Clarke and Johnston 2003), however, the true number of species living on the shelf must be considered a magnitude higher than that (Gutt et al. 2004). This estimate might be further augmented by the discovery of species complexes that consists of cryptic species (Allcock 2002, Page and Linse 2002, Held 2003, Held and Wägele 2005, Linse et al. 2007, Wilson et al. 2007, Leese and Held 2008).

Antarctic marine macroinvertebrate biodiversity is concentrated predominantly on the narrow Antarctic shelf and to a lesser extend to the continental slope and the deep sea (but see Brandt et al. 2007). However, during the recurrent glacial periods on Milankovitch timescales (Imbrie et al. 1993) a massive ice shield that exceeded the extent of today's ice sheet by far covered most areas of the Antarctic continent and the shelf. Based on geophysiological evidence and mathematical models it must be expected that grounded ice reached as far as to the outer shelf break (Elverhoi 1981, Gingele et al. 1997, Anderson et al. 2002, Huybrechts 2002, Evans et al. 2004, Ingolfsson 2004). For the Antarctic benthic communities this must have imposed a major disturbance as the grounded ice sheets expelled the benthic fauna from the shelf to a large degree. In addition to this extensive ice disturbance, primary production was generally reduced at glacial maxima due to thick and permanent multiannual sea ice (see Thatje et al. 2008). While most studies on evolution of the Antarctic communities mainly focussed on phylogenetic aspects to explain on evolutionary timescales how species evolved in an increasingly cold and hostile habitat (e.g. Eastman 1991, Lecointre et al. 1997, Bargelloni et al. 2000, Held 2000, Near et al. 2003) we also have to ask the question how benthic species, in particular the less mobile brooding species, survived periods of high disturbance on shorter, ecological time scales. Either benthic communities went extinct during these glacial periods or they had to survive in 'refuge areas' such as the deep sea, shelf areas of South America, the continental slope or in ice-free refugia on the shelf. The eurybathy of many Antarctic taxa is regarded as an indication for 'vertical evasive actions' (Brey et al. 1996). While the existence of refuge areas is often taken for granted (Clarke and Crame 1992, Crame 1997), biological data providing evidence on where those could have been located and whether populations of a species were really separated in different refuge areas are scarce (e.g. Linse et al. 2007). Very few studies investigated the intraspecific genetic variability in the context of historical processes (Janko et al. 2007). Often, the availability of sufficient sample sizes and highly informative markers such as microsatellites also limited the scope of the studies (see Held and Leese 2007).

This study investigates the partitioning of genetic polymorphisms in one more broadly distributed species of the *Ceratoserolis trilobitoides* (Eights, 1833) species complex (Wägele 1986, Held 2003, Leese and Held 2008, see Fig. 1). Only one species of this complex, *Ceratoserolis* n. sp. 1 (group 1 in Held 2003), has a broad distribution and occurs around the South Shetland Islands at the tip of the Antarctic Peninsula (AP) and the eastern Weddell Sea (EWS). If glaciations affected the present day population structure, i.e. the subdivision between

fferent populations and their present-day genetic variability, we expect to find genomic gnatures in particular in species with limited dispersal capabilities (Poulin et al. 2002, Thatje et l. 2005). *Ceratoserolis* n. sp. 1, like all members of the serolid isopods known so far, is strictly enthic, broods their offspring and lives half buried in the soft-bottom sediments. For *eratoserolis* the impact of glaciations must have been particularly severe due to its limited ability evade the ice. This, however, makes *Ceratoserolis* n. sp. 1 an appropriate model organism to tudy the effects of large-scale ice disturbances during Southern Ocean history on the genome of characteristic key species. For this purpose we use a novel set of ten nuclear microsatellite arkers (Held and Leese 2008) and a fragment of the cytochrome c oxidase subunit I gene COI).

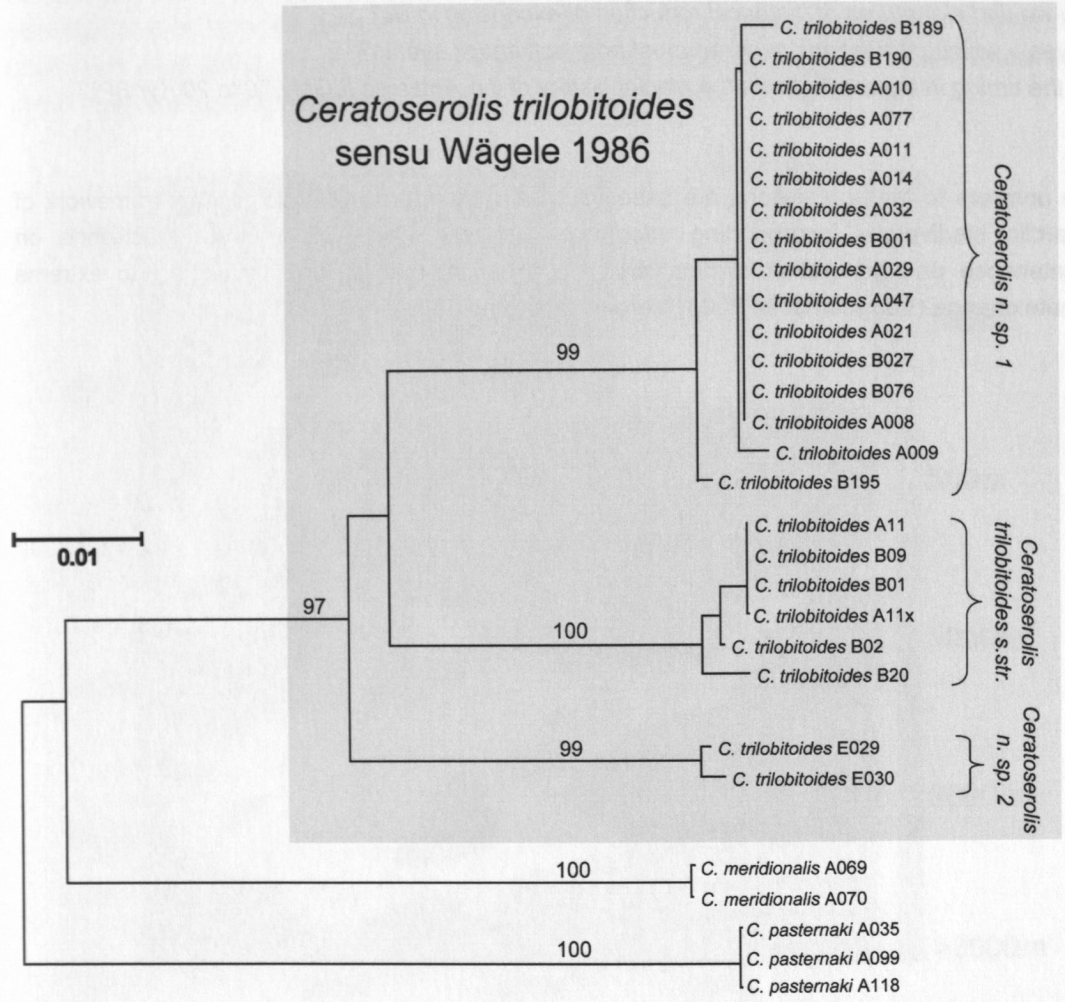


Figure 1: Neighbor-Joining tree based on the 3'-terminus of the mitochondrial large subunit rRNA gene (16S) of members of the genus *Ceratoserolis* Brandt 1988. The numbers correspond to the bootstrap support (1000 replicates) of the corresponding Maximum Parsimony tree. All OTUs in the grey box belong to the *Ceratoserolis trilobitoides sensu* Wägele, 1986 group, which consist of at least three distinct groups: the type, *Ceratoserolis trilobitoides sensu strico* (group 2 in Held, 2003), a widely distributed group *Ceratoserolis* n. sp. 1 (group 1 in Held, 2003, Leese and Held 2008), and a group found only around the South Sandwich Islands, *Ceratoserolis* n. sp. 2 (Held and Leese unpublished data).

By focussing on assessing present-day and past population structure with this marker set we address the following questions to make inferences about the evolutionary past of *Ceratoserolis* n. sp. 1 on the Antarctic shelf.

Present genetic structure

- *are populations of Ceratoserolis* n. sp. 1 *from the EWS and the AP genetically differentiated?*
- *what is the present-day level of gene flow?*

Present genetic diversity

- *do populations differ in their genetic diversity?*
- *what is the effective populations size (N_e) for the populations from different regions?*

Historical genetic structure:

- *do we find signatures of historical reduction or expansion in N_e?*
- *if yes – which populations suffered most from bottleneck events?*
- *is the timing in agreement with the glacial history of the Antarctic (LGM ~70 to 20 Kyr BP)?*

The answers to these questions are essential to add data to a microevolutionary framework of Antarctic biodiversity incorporating information on the effects of recurrent glaciations on maintenance and generation of biodiversity in a habitat that is currently subject to extreme climate change (Vaughan et al. 2003, Meredith and King 2005).

Material and Methods

Taxon sampling

Specimens were collected by bottom trawling during the expeditions ANT-XIII/3, ANT-XIV/2, ANT-XIX/3 and ANT-XXI/2 to the EWS and the tip of the AP with the German research vessel "POLARSTERN". Specimens were collected at 19 stations (Tab. 1). The animals were sorted directly on deck, immediately preserved in pre-chilled 96% ethanol and kept at 4°C for the first months to ensure adequate tissue preservation for maximum yield of high quality DNA from the cells. Due to limited numbers of specimens available at some sampling locations, specimens from neighbouring sites were pooled to increase sample size for some regions (Tab. 1). To avoid biasing results by pooling specimens of distinct gene pools (Wahlund effect), we tested that neither significant violations of Hardy-Weinberg equilibrium (HWE) nor significant linkage disequilibrium (LD) were observed (spatial pooling). According to the same principle we also pooled specimens from the same site sampled a different times (temporal pooling). In total, 13 independent populations, P1 to P13, with sample sizes of minimum n=14 (P11) and maximum n=78 (P4) were analysed in this study (Tab. 1).

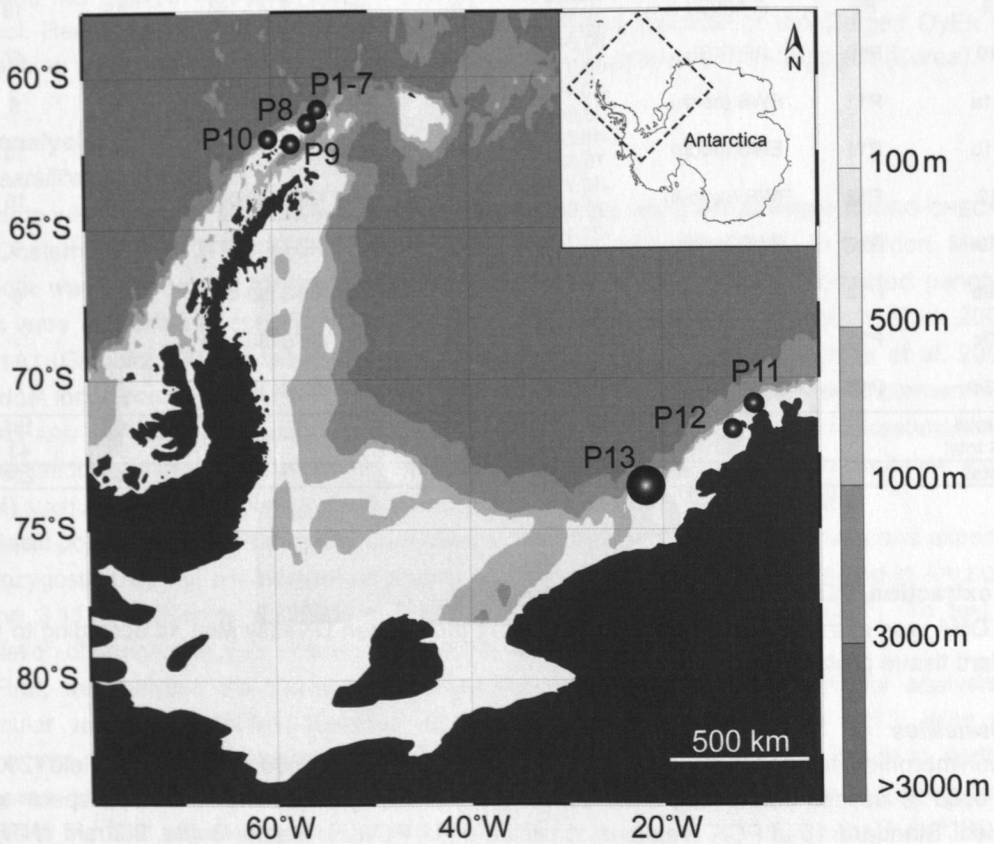


Figure 1: Overview over the sampling sites around the tip of the Antarctic Peninsula (populations P1-P10) and from the shelf from the eastern Weddell Sea (P11-13). P5, P7, P11 and P13 consist of specimens sampled at different stations at the same region (see Tab. 1).

Table 1: Sampling sites (AP=Antarctic Peninsula, EI=Elephant Island, KGI=King George Island, EWS=Eastern Weddell Sea) with precise geographical position (decimal longitude / latitude), depth and cruise information for the populations of *Ceratoserolis* n. sp. 1 investigated in this study. Msat n=400, mtDNA n=224.

Population	Label	Region	Lon/Lat [°W / °S]	depth [m]	Cruise / Station	N_{MSAT}	N_{mtDNA}
1	P1	AP (EI)	56.423 61.232	403	ANTXIV-2 (42-022)	39	3
2	P2	AP (EI)	56.180 61.140	380	ANTXIV-2 (42-003)	31	28
3	P3	AP (EI)	56.125 61.102	371	ANTXIV-2 (42-023)	18	-
4	P4	AP (EI)	55.970 60.982	415	ANTXIV-2 (42-027)	78	53
5a	P5	AP (EI)	55.659 60.824	454	ANTXIX-3 (61-064)	6	-
5b	P5	AP (EI)	55.622 60.837	344	ANTXIV-2 (42-032)	26	16
6	P6	AP (EI)	54.813 61.023	503	ANTXIV-2 (42-040)	30	11
7a	P7	AP (EI)	54.560 61.160	278	ANT-XIX-3 (61-048)	11	-
7b	P7	AP (EI)	54.633 61.233	333	ANTXIV-2 (42-014)	8	-
8	P8	AP (EI)	56.252 61.502	572	ANTXIV-2 (42-019)	19	5
9	P9	AP (KGI)	57.983 62.183	476	ANTXIV-2 (42-164)	19	18
10	P10	AP (KGI)	60.345 62.015	374	ANTXIV-2 (56-184)	47	47
11a	P11	EWS (north)	-10.472 70.845	281	ANT-XXI-2 (65-336)	5	5
11b	P11	EWS (north)	-10.596 70.835	269	ANT-XXI-2 (65-121)	9	10
12	P12	EWS (central)	-12.738 71.685	227	ANT XIII-3 (39-005)	33	15
13a	P13	EWS (south)	-21.168 73.302	459	ANT XIII-3 (39-012)	3	3
13b	P13	EWS (south)	-21.165 73.300	668	ANT XIII-3 (39-017)	12	6
13c	P13	EWS (south)	-21.166 73.366	338	ANT XIII-3 (39-011)	1	-
13d	P13	EWS (south)	-22.508 73.700	446	ANT XIII-3 (39-015)	5	4
AP total						332	181
EWS total						68	43
Total						**400**	**224**

DNA extraction, PCR, sequencing and genotyping.

Total DNA was extracted from muscle tissue using the Qiagen DNeasy Mini Kit according to the standard tissue protocol.

Microsatellites

Ten polymorphic microsatellite loci for *Ceratoserolis* n. sp. 1 developed by Leese and Held (2008) were used to assess partitioning of intraspecific genetic polymorphisms for all 400 specimens sampled. Standard 15 µl PCR reactions consisted of 1x PCR HotMaster Buffer, 0.2 mM dNTPs, 0.5-0.75 µM of each primer (one labeled, one unlabeled), 0.03 U/µl HotMaster *Taq* (Eppendorf, 5-prime), 0.5 M Betaine (Sigma Aldrich) and 20 ng of genomic DNA (equilibrated). Cycling conditions on an epgradient thermocycler (Eppendorf) were 2 min at 94°C followed by 28 to 32 cycles with 20 s at 94°C, 15 s at annealing temperature, 30 s at 65°C. A final extension step of 45 min at 65°C was performed to reduce stutter artefacts from incomplete adenylation of products.

PCR products were controlled on a 2% TBE agarose gel, diluted 1-15 fold with molgrade water. Of the diluted PCR product 1 µl was denatured in a mixture of 14.7 µl HI-DI formamide with 0.3 µl GeneScan ROX 500 size standard (both Applied Biosystems). Fragment analyses were performed on an ABI 3130xl automated sequencer. Allele length scoring was performed using the software GENEMAPPER 4.0 (Applied Biosystems). To minimize errors due to *in vitro* artefacts, PCR and fragment analysis was performed up to four times for a subset of samples. To minimize genotyping errors, genotyping was objectified by two independent persons.

Cytochrome c oxidase subunit I (COI)

A 563 bp fragment of the mitochondrial COI was amplified for a subset of 224 specimens (Tab. 1). Primers HCO2198 and LCO1490 (Folmer et al. 1994) were used for amplification. Reactions were carried out in 25 µl volumes with 1x HotMaster reaction buffer, 0.2 mM dNTPs, 0.5 µM of each primer, 0.025 U/µl Hotmaster *Taq*. Reaction conditions were initial denaturation for 2 min at 94°C followed by 36 cycles of 20 s at 94°, 15 s at 46°C and 80 s at 65°C plus a final elongation step of 5 min at 65°C. PCR products were purified using Qiagen Qiaquick PCR purification kit or the ExoSap-IT enzyme mix (GE Healthcare). Cycle-sequencing using 1 µM either of the amplification primers was conducted in 10 µl reaction volumes using 1 µl of the purified template DNA and the BigDye Termiator Kit 3.1 (Applied Biosystems) according to the manufacturer's protocol. Reactions were purified according to the 'modified protocol' of the Qiagen DyEx Kit. Sequencing was conducted on an ABI 3130xl sequencer or outsourced to Macrogen (Korea).

Data analysis

Microsatellites

Raw data were checked and corrected for genotyping errors using the software MICRO-CHECKER (van Oosterhout et al. 2004) and DROPOUT (McKelvey and Schwartz 2005). In addition, MICRO-CHECKER was used to test for the presence of null alleles in populations. Corrected genotype tables were converted to specific software formats using the software MSTOOLKIT (Park 2001), CONVERT (Glaubitz 2004), FORMATOMATIC (Manoukis 2007) and CREATE (Coombs et al. 2007). Individual locus contribution to a strict SMM was calculated using F_{CTR} tests as implemented in ANIMALFARM version 1.0 (Landry et al. 2007). This is necessary to avoid overestimation of differentiation estimates that assume for mutations according to a strict stepwise mutation model (SMM) such as Slatkin's or Rousset's R_{ST} estimates (Slatkin 1995, Rousset 1996).

Basic population genetic diversity parameters (Nei 1987) such as the observed and expected heterozygosity (H_O, H_E) and number of alleles (N_A) were calculated as implemented in ARLEQUIN version 3.11 (Excoffier et al. 2005) and GENEPOP version 4.0.6 (Rousset 2007). To test for population differentiation, four different approaches were followed:

First, we analysed the hierarchical distribution of genetic polymorphisms by analysis of molecular variances (AMOVA, Excoffier et al. 1992, Cockerham and Weir 1973, Weir and Cockerham 1984) as implemented in ARLEQUIN. This is a straightforward approach to partition variances at different hierarchical levels (among individuals, subpopulations, subpopulation groups or the total population). We defined as subpopulation groups (P1-P10) vs (P11-P13), i.e. subpopulations from the AP versus the populations from the EWS and assessed the relative contribution of molecular variance at the level of (i) subpopulations (Φ_{ST}) and (ii) population groups (Φ_{CT}). The significance of the fixation indices was assessed using a non-parametric permutation test with 10,000 random permutations (Excoffier et al. 1992).

In a second step, we performed a Factorial Correspondence Analysis (FCA) as implemented in GENETIX, version 4.0.5 (Belkhir et al. 1996). FCA is a multidimensional statistical method summarizing and visualizing large datasets into informative essential subsets that represent trends in the original data without making assumptions on underlaying population structure (e.g. idealized Fisher-Wright population) or mutational dynamics of the molecular markers applied (i.e. different mutation models).

Third, we calculated pairwise differentiation estimates between populations based on Weir and Cockerham's (1984) coancestry coefficient *Theta* (referred to as F_{ST} in this paper) and Rousset's (1996) SMM-based differentiation estimate *Rho* (referred to as R_{ST} in this paper) using the software GENEPOP. Both differentation measures are the most frequently applied estimates to infer population structure based on microsatellite data and rely on an analysis of variance (ANOVA) framework analysing allele frequency (F_{ST}) and allele size (R_{ST}) distributions in subpopulations compared to the total population. F_{ST} analyses rely on the assumptions of Wright's n-Island model, most importantly the assumption that migration is much greater than mutation ($m >> \mu$) and mutation dynamics is in accordance with an Infinite Allele Model (IAM, Kimura and Crow 1964). For R_{ST} this is no prerequisite and mutation rate can be high (even $\mu >> m$), as long as mutation occurs in a stepwise fashion, i.e. it is conform with the Stepwise Mutation Model (SMM, Otha and Kimura 1973), which is, however, typical for many microsatellites (Valdes et al. 1993).

F_{ST} estimates were standardized according to the method outlined by Hedricks (2005) in an ANOVA framework as outlined by Meirmans (2006) using the program RECODEDATA (Meirmans 2006) and subsequently GENEPOP to calculate the maximum value possible for F_{ST}. In the following, the F_{ST} estimate calculated was divided by $F_{ST\ max}$ to yield the standardized value F'_{ST} that has not only a theoretical range of 0 to 1. For the fourth approach we applied a Bayesian cluster analysis to detect population structure and assign individuals based on their genotypes probabilistically to one or several of the populations detected using the program STRUCTURE version 2.2.2 (Pritchart et al. 2000, Falush et al. 2003). The advantage of this approach is that STRUCTURE does not rely on calculating differentiation values on predefined populations but estimates the most likely number of populations (K) without prior information on population origin, assuming HWE and no or weak LD within populations. By calculating the log likelihood for the data given a certain K over a predefined range ($K_{i\ 1\rightarrow n}$ ln $Pr(D|K_i)$) the most likely number of distinct populations under different underlying models can be computed. In addition, STRUCTURE can use the information on population origin to calculate the log likelihood for this particular value for K_i for comparison. For the *Ceratoserolis* n. sp. 1 data set, the most likely number of populations was inferred with and without prior information on population origin, with and without assuming admixture. The number of MCMC steps needed to reach convergence during the burn-in was determined by using the software AMTY (Wilgenbusch et al. 2004). Based on the results of AWTY we used number of 5,000 MCMC burn-in steps and subsequently sampled 100,000 MCMC steps for $K_{i\ 1\rightarrow 13}$. Ten independent replicates were calculated, which were used to compute a consensus population membership coefficient matrix with the program CLUMPP, version 1.1.1. (Jakobsson and Rosenberg 2007). Both, the individual membership coefficient matrix and averaged population membership matrix (Q-matrices) were visualized using the program DISTRUCT, version 1.1 (Rosenberg 2004). As very strong genetic structure between population groups can camouflage present but weaker differentiations among populations within

regions (e.g. Craig et al. 2006) we ran STRUCTURE under the same conditions for populations from the two different geographic regions, i.e. (AP: P1 – P10) and (EWS: P11 – P13), separately.

The data set was tested for isolation by distance correlation using the program IBD (Bohonak 2002) This program applies a Mantel test to detect correlations between matrices of pairwise F_{ST} values (normal and logarithmic) and geographical (Euclidean) distances (normal and logarithmic) between populations. IBD analyses were also performed for the data from each region (AP and EWS) separately.

Population demography

We estimated the present effective population size (N_e) using the linkage disequilibrium method proposed by Hill (1981), modified by Waples (2006) to account for bias correction when only small sample sizes are available (Waples 2006). This method is implemented in the LDNE, version 1.3 (Waples and Do 2007). Calculations of N_e and the confidence intervals were estimated including alleles with a frequency of $c \geq 0.05$ and $c \geq 0.01$, respectively.

Tests for historical population bottlenecks were performed using the program BOTTLENECK (Piry et al. 1999, Cornuet and Luikart, 1996). Tests are based on the finding that populations that have experienced recent reductions in their effective population size (N_e) show a reduction in allelic richness and heterozygosity. Following a bottleneck, the number of alleles drops off more quickly than heterozygosity. Vice versa, in expanding populations, the number of alleles increases faster than heterozygosity until equilibrium is reached. From the correlation of both parameters from microsatellite data, it is possible to make inferences on historical demography of a population. The expected heterozygosity is computed from the observed number of alleles (N_A) under the assumption of a constant-size population (Luikart et al. 1998). Analyses were performed assuming a strict IAM, a strict SMM, and a so-called two-phase model (TPM, Di Rienzo, 1994) allowing for varying proportions of both models (IAM and SMM). In our case, we performed analyses with SMM proportions of 70 – 90%.

Data analysis COI:

Assembly of forward and reverse strands and editing was performed using the software CODONCODE ALIGNER version 1.6 (CodonCode Corporation). Sequence alignment was performed using the CLUSTALW program as implemented in BIOEDIT 7.0.9 (Hall 1999). Sequence data were collapsed to a haplotype frequency file using COLLAPSE version 1.2 (Posada 2004, http://darwin.uvigo.es/software/collapse.html). File conversion was done using FABOX version 1.32 (Villesen 2007). Sequence variation was analysed using MEGA 4.0 (Tamura et al. 2007). Statistical parsimony networks were calculated using TCS 1.21 (Clement et al. 2000). PAUP version 4b10 (Swafford 2000) was used to calculate neighbor joining distance trees and bootstrap support (1,000 replicates) of clades. Population genetic parameters were inferred using ARLEQUIN 3.11. Haplotype diversity (H) and pairwise differences (π) were calculated according to Nei (1987).

Detecting demographic expansions from mtDNA

Traces of population expansions were examined by using three different approaches. First, we calculated Tajima's *D* (Tajima 1989) which is based on the magnitude of difference between the number of segregating sites (*S*) and the average pairwise nucleotide differences (π). Population expansions after bottleneck or founder events inflate *S* relative to π leading to negative values for

D. The second statistics applied, Fu's F_S statistic (Fu 1997), is even more sensitive in detecting populations expansions from smaller samples (Ramos-Onsins 2002). Fu's F_S statistic (Fu 1997) is based on the probability of finding a number of alleles greater than or equal to the observed number in a sample drawn from a stationary population. Significance of Fu's F_S and Tajima's *D* were estimated by coalescent simulation as implemented in ARLEQUIN 3.11 using 10,000 permutations. In the third test, pairwise mismatch distributions among individuals were plotted and tested for goodness-of-fit to a model of sudden expansion using parametric bootstrapping with 50,000 replicates (Schneider and Excoffier 1999). The parameter *tau* inferred by mismatch analyses was used to infer the timing of population expansion events (Rogers and Harpending 1992). Because no calibrated clock for the COI of *Ceratoserolis* exists (but see Held 2001 for transversions only) a range substitution rates calibrated on other Crustacea (Knowlton et al. 1993: 2.2 – 2.6% Myr^{-1}, Schubart et al. 1998: 1.66% - 2.3% Myr^{-1}, Knowlton and Weigt 1998: 1.4% Myr^{-1}) was applied to estimate the timing of population expansion.

Nested Clade Analysis (NCA)

Present day population structure is the result of limited gene flow, historical processes (habitat fragmentation, range expansion, colonization), or a combination thereof (Avise 2000). To statistically infer the processes that most likely led to the present-day partitioning of sequence data, Nested Clade Analysis, NCA (Templeton et al. 1998), provides a powerful means to disentangle historical and contemporary processes and tests several scenarios for the data observed. NCA was performed with GEODIS version 2.5 (Posada et al. 2000, 2006) based on the topology of a 95% connection limit parsimony network calculated by TCS. Alternative connections of the network were tested. Inferences on the underlying processes structuring populations variation we used the latest Inference Key for the results of the program GEODIS (Templeton and Posada 2005, available at: Darwin.uvigo.es/download/geodisKey_11Nov05.pdf).

Results

Microsatellites

Genetic diversity and marker characteristics

The 10 microsatellite loci had six to 42 alleles each for the n=400 specimens analysed (C0M03, C0M05 respectively). Allele lengths of microsatellites C0M03, C0M07, C0M12, C1Df14 and C1H93 were always a multiple of the size of the repeat unit. For microsatellite loci C0M05, C0M11, C1D05, C1H55, C3D10 some rare alleles differed by only ±1 bp from the frequent alleles thus being incompatible with a strict SMM, indicating that a second process besides slip-strand mispairing (SSM) causes variation at this locus. Significant overall evidence for null alleles was indicated for locus C1Df14 only (P<0.01) and supported by high F_{IS} values. Neither evidence for global LD between locus pairs nor deviations from HWE were observed after sequential Bonferroni corrections for multiple testing (Rice, 1989). Allelic diversity was significantly larger for populations from the AP with $N_{A\ mean}$ = 11.29 than for EWS with $N_{A\ mean}$ = 7.77 (T-test, P=0.004). However, sample size around the AP was around 5 times larger than for the EWS. Gene diversity in terms of expected heterozygosity corrects for unequal sample sizes (Nei 1987) and there is no conclusive evidence for significantly lower genetic diversity for populations from the EWS (H_E=0.774±0.012) than for populations from the AP (H_E=0.805±0.009) (T-test, P=0.64).
Genetic diversity estimates based on the highly variable microsatellites thus indicate only a minor difference in genetic diversity between AP and EWS populations.

Population differentiation

For calculations of AMOVA we defined an additional hierarchical level representing the regional groups AP (P1-P10) and EWS (P11-P13). Results of the AMOVA (Tab. 3) indicated that most of the variation was distributed within individuals (87.97%, Φ_{IT}=0.12). A significant amount of variation was also distributed among individuals within populations (5.44%, Φ_{IS}=0.058). Only a very minor and non-significant fraction of the variation was partitioned among populations within regions (-0.19%, Φ_{SC}=-0.002) but a large and significant amount between the two major geographical regions AP and EWS (6.77%, Φ_{CT}=0.068). The FCA plot clearly shows the split between the two population groups along axis 1 (Fig. 3). Axes 2 and 3 had considerably lower discriminating power and added no further discernable information concerning population structure.

Table 2: Total number of specimens scored for each population and locus (N_S), number of different alleles (N_A), inbreeding coefficient (F_{IS}), observed heterozygosity (H_O) and expected heterozygosity (H_E) for the 10 microsatellite loci and the 13 populations of *Ceratoserolis* n. sp. 1. Bonferroni adjusted significance level of 5% equals a P-value of P=0.0038. Significant deviations of HWE are printed in bold and indicated by */** (P<0.05, P<0.01).

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
C0M03													
N_S	78	60	36	154	64	60	38	38	36	92	28	66	42
N_A	5	3	4	4	5	4	4	4	4	4	2	4	4
F_{IS}	0.118	-0.161	-0.054	-0.012	0.054	-0.131	-0.013	0.214	-0.043	0.028	-0.28	-0.094	0.015
H_o	0.538	0.733	0.667	0.623	0.594	0.667	0.579	0.474	0.722	0.630	0.643	0.576	0.619
H_e	0.610	0.633	0.633	0.616	0.627	0.591	0.586	0.599	0.694	0.648	0.516	0.523	0.628
C0M05													
N_S	78	62	36	154	64	60	36	38	38	94	28	64	42
N_A	28	29	18	27	26	24	19	18	21	26	8	13	11
F_{IS}	0.034	0.068	0.012	0.114	-0.006	-0.039	0.007	-0.051	0.236	0.082	-0.090	0.025	-0.003
H_o	0.923	0.903	0.944	0.831	0.969	1.000	0.944	1.000	0.737	0.872	0.929	0.844	0.857
H_e	0.955	0.968	0.956	0.937	0.963	0.963	0.951	0.953	**0.959****	0.949	0.855	0.865	0.855
C0M07													
N_S	78	62	36	154	64	58	36	38	38	94	28	64	42
N_A	12	13	10	16	13	11	11	9	14	14	6	9	8
F_{IS}	0.211	0.216	0.187	0.220	0.140	0.154	0.120	0.351	0.177	0.134	0.031	-0.059	-0.168
H_o	0.692	0.710	0.667	0.688	0.750	0.724	0.778	0.526	0.737	0.766	0.786	0.875	0.952
H_e	0.875	0.902	0.816	**0.881****	0.871	0.854	0.881	0.804	0.890	0.884	0.810	0.827	0.819
C0M11													
N_S	76	62	36	156	64	60	38	38	38	94	28	66	42
N_A	8	8	8	9	8	6	6	8	6	9	5	5	5
F_{IS}	-0.067	0.104	0.090	0.004	0.122	0.016	-0.209	-0.032	-0.011	-0.135	0.041	0.063	0.017
H_o	0.816	0.742	0.778	0.782	0.688	0.800	0.947	0.842	0.789	0.915	0.714	0.667	0.762
H_e	0.765	0.827	0.852	0.785	0.781	0.813	0.788	0.817	0.781	0.808	0.743	0.710	0.775
C0M12													
N_S	72	62	36	154	54	60	34	38	38	94	28	66	42
N_A	17	17	13	20	19	14	13	15	14	17	14	14	13
F_{IS}	0.106	0.035	0.099	0.009	-0.027	0.117	0.044	0.220	0.272	0.074	0.031	0.182	0.046
H_o	0.833	0.903	0.833	0.922	0.962	0.800	0.882	0.737	0.684	0.851	0.857	0.758	0.857
H_e	0.930	0.935	0.922	0.930	0.939	0.905	0.922	**0.939***	**0.933***	0.919	0.884	0.923	0.898

Table 2: continued

C1D05													
N_S	78	62	36	156	64	60	38	38	38	94	28	66	38
N_A	8	6	6	8	7	5	5	7	5	8	5	7	6
F_{IS}	0.074	-0.111	-0.121	-0.058	0.079	-0.005	0.071	0.154	0.081	0.161	0.212	0.005	0.049
H_o	0.692	0.806	0.833	0.769	0.719	0.700	0.684	0.684	0.579	0.617	0.571	0.727	0.632
H_e	0.747	0.727	0.746	0.727	0.779	0.697	0.735	0.805	0.629	0.734	0.720	0.731	0.663
C1Df14													
N_S	70	56	30	142	60	56	36	36	32	88	28	64	42
N_A	6	6	4	8	7	7	6	4	5	7	5	7	6
F_{IS}	0.168	0.140	0.055	0.248	0.358	0.235	0.266	0.177	0.201	0.091	0.546	0.413	0.371
H_o	0.571	0.607	0.533	0.535	0.467	0.571	0.556	0.556	0.563	0.636	0.357	0.469	0.476
H_e	0.685	0.704	0.563	0.710	**0.722***	0.743	0.751	0.671	0.700	0.699	**0.770***	**0.794***	0.750
C1H55													
N_S	78	62	36	156	64	60	38	38	36	92	28	66	42
N_A	19	21	13	21	18	18	17	19	14	20	9	15	10
F_{IS}	0.077	-0.034	0.069	0.013	-0.042	-0.027	0.045	-0.003	0.054	-0.047	0.080	0.141	0.033
H_o	0.872	0.968	0.833	0.923	0.969	0.967	0.895	0.947	0.889	0.978	0.786	0.758	0.810
H_e	0.944	0.937	0.894	0.935	0.930	0.942	0.936	0.945	0.938	0.935	0.852	0.880	0.836
C1H93													
N_S	58	56	32	154	64	58	36	28	38	88	28	64	42
N_A	6	5	6	8	7	6	6	4	6	5	5	6	5
F_{IS}	0.445	0.384	0.375	0.218	0.179	0.309	0.154	0.414	-0.219	0.465	0.008	0.160	-0.218
H_o	0.379	0.393	0.438	0.494	0.594	0.448	0.611	0.357	0.789	0.341	0.714	0.625	0.952
H_e	**0.678***	0.633	0.692	0.630	0.721	0.645	0.720	0.601	0.651	**0.633****	0.720	0.742	0.786
C3D10													
N_S	78	62	36	154	64	50	38	38	38	94	28	62	42
N_A	16	15	10	15	15	14	12	8	14	12	8	9	9
F_{IS}	-0.016	0.060	0.048	-0.014	-0.032	-0.060	-0.133	0.007	0.030	0.116	0.127	0.016	0.248
H_o	0.897	0.806	0.778	0.857	0.906	0.933	1.000	0.842	0.842	0.745	0.643	0.774	0.619
H_e	0.883	0.857	0.816	0.845	0.878	0.882	0.886	0.848	0.868	0.842	0.733	0.787	0.818
$H_{o\ mean}$	0.722	0.757	0.730	0.743	0.762	0.761	0.788	0.697	0.733	0.735	0.707	0.754	0.738
$H_{e\ mean}$	0.807	0.812	0.789	0.800	0.821	0.803	0.815	0.798	0.804	0.805	0.760	0.779	0.783

The overall F_{ST} estimate averaged over all loci and populations was moderate, F_{ST}=0.021, but highly significant, indicating population subdivision. Pairwise F_{ST} between populations within either region (EWS and AP, respectively) were low and non-significant thus suggesting that there is no further geographical structure (Tab. 4). As in the AMOVA, significant pairwise population differentiation was only observed for comparisons between populations from the two different regions. Comparing populations from either side of the Weddell Sea there is evidence for strong differences; Weir and Cockerham's pairwise F_{ST} estimates of 0.057-0.093 (Tab. 4). The standardized estimates for significant pairwise F_{ST} were about five times higher with values of 0.274 to 0.440 (Tab. 5). This corroborates that without such a standardization approach the magnitude population subdivision based on multilocus F_{ST} is severely underestimated (Balloux and Lugon-Moulin 2002, Hedrick 2005, Meirmans 2006).

Tests for equal variances performed using ANIMALFARM prior to R_{ST} calculations indicated a significantly disproportionate variance for locus C0M05 among populations (P=0.01, 10,000 permutations). In fact, 60.4% of the total R_{ST} interpopulation distance is contributed by locus C0M05 only. Consequently, this locus was excluded and R_{ST} estimates were calculated based on the other nine microsatellites that contributed equally to the SMM (Tab. 5). Results of R_{ST} estimates revealed differentiation patterns similar to those derived from the F_{ST} estimates (Tab. 4, Tab. 5). For the absolute values of differentiation estimates we found that $R_{ST} > F_{ST}$ but $R_{ST} < F'_{ST}$ (Tab. 4, 5). Between-region pairwise R_{ST} estimates ranged from 0.067 – 0.273 (Tab. 5).

Testing genetic differentiation across the Weddell Sea (P1-P10 vs P11-P13) was highly significant (P<0.001). F_{ST} estimates were F_{ST}=0.068 for all loci, and F_{ST}=0.069 excluding locus C0M05. R_{ST} estimates were R_{ST}=0.333 for all loci, and only R_{ST}=0.132 excluding locus C0M05.

The overall higher values for R_{ST} estimates indicates that gene flow across the Weddell Sea is very low and has been for extended periods of time making mutations according to a SMM dominant over the effects of genetic drift alone.

Bayesian analyses performed using STRUCTURE reported the highest likelihood for the data for K=2 predefined clusters, ln $Pr_{max}(D|K)_{K=2}$) (Tab. 6). All individual genotypes were correctly assigned to the two major genotype groups (AP, EWS, Fig. 4). Two individuals in P5 had almost similar assignment probabilities for the two clusters indicating possible ancestral admixture. No additional substructure was observed from the Q-matrix calculated for K=3, which had the second highest log likelihood (Fig. 4). Also, no additional substructure was observed when Q-matrices were calculated for P1-10 and P11-13 separately, with K=1 being supported by highest log likelihood values (Fig. 5). When changing the model (no admixture, alleles independent), the results remained the same, however, admixture proportions were much less prominent. Again, the results support that there is a very strong split between populations from the AP and the EWS while no substructure is detectable among populations within the two regions.

Although isolation by distance analyses (IBD) revealed a significant and strong positive correlation between genetic and geographic distances for the whole data set (n=400), this correlation is spurious. When testing for IBD at regional scales, i.e. within the two geographic regions (AP and EWS), there was no evidence for IBD (Fig. 6). The significance of IBD between regions rests entirely on the two geographical extremes and is caused by the fact that we could not sample geographically intermediate populations.

Table 3: Hierarchical analysis of molecular variance (AMOVA) among *Ceratoserolis* n. sp. 1 populations within and between the two major regions investigated (AP and EWS) using 10 microsatellite markers. P refers to the probability of randomly getting a similar result based on 10,000 independent permutations.

Component of differentiation	df	variance [%]	Φ statistics	P
Among regions	1	6.77	Φ_{CT}=0.068	0.003
Among populations within regions	9	-0.19	Φ_{SC}=-0.002	0.969
Among individuals within populations	387	5.44	Φ_{IS}=0.058	0.000
Within individuals	400	87.97	Φ_{IT}=0.120	0.000

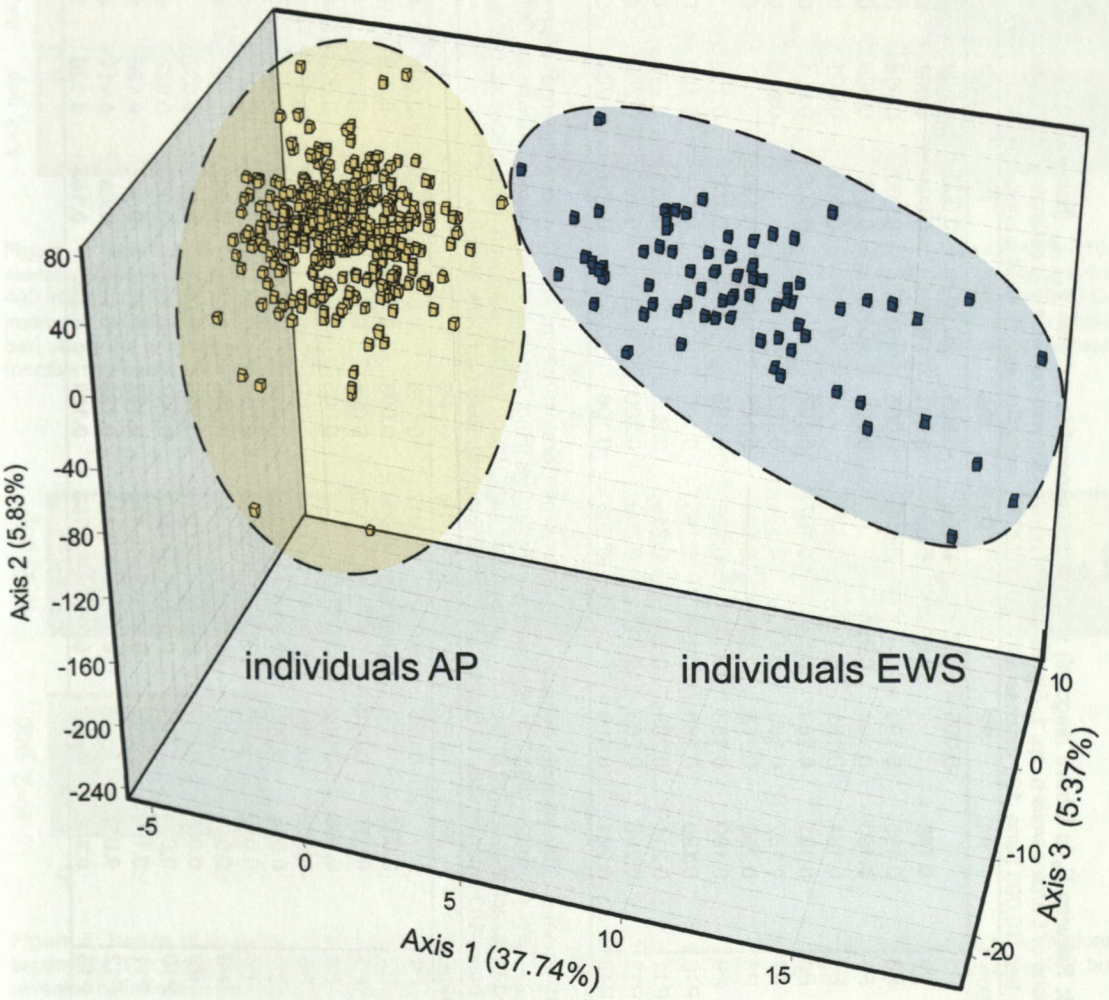


Figure 3: Three-dimensional Factorial Correspondence Analysis plot for n=400 specimens of *Ceratoserolis* n. sp. 1 genotyped at 10 microsatellite loci. The major proportion of the variation is distributed along axis 1, supporting a distinct split between populations from the Antarctic Peninsula (yellow) and the eastern Weddell Sea (blue).

Table 4: Pairwise F_{ST}'s (above diagonal) and maximum possible pairwise $F_{ST\ max}$ (below diagonal) calculated according to Meirmans (2006) between populations of *Ceratoserolis* n. sp. 1. Significance of above diagonal pairwise F_{ST} estimates (theta according to Weir and Cockerham, 1984) was performed by an G-test (10,000 burin, 100 batches with 10,000 MCMC sampling steps). Bold numbers: P<0.01.

Population	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
P1	-	-0.005	0.003	-0.003	-0.005	-0.002	-0.008	-0.002	-0.001	-0.003	**0.064**	**0.069**	**0.073**
P2	0.185	-	0.002	0.001	0.001	0.001	-0.004	0.002	-0.001	0.000	**0.057**	**0.060**	**0.064**
P3	0.195	0.193	-	0.006	0.005	0.006	0.000	-0.008	0.007	0.004	**0.085**	**0.084**	**0.093**
P4	0.193	0.191	0.200	-	0.001	-0.002	-0.003	0.000	0.001	-0.001	**0.070**	**0.072**	**0.074**
P5	0.184	0.181	0.190	0.191	-	-0.003	-0.009	-0.001	0.003	-0.002	**0.060**	**0.063**	**0.061**
P6	0.191	0.189	0.199	0.197	0.187	-	-0.008	0.001	0.002	-0.002	**0.059**	**0.059**	**0.062**
P7	0.187	0.185	0.195	0.194	0.182	0.191	-	-0.005	-0.002	-0.008	**0.058**	**0.057**	**0.060**
P8	0.190	0.188	0.199	0.195	0.185	0.194	0.189	-	0.005	-0.002	**0.065**	**0.068**	**0.075**
P9	0.190	0.188	0.198	0.196	0.185	0.194	0.189	0.192	-	-0.003	**0.069**	**0.069**	**0.070**
P10	0.190	0.188	0.197	0.196	0.186	0.194	0.190	0.192	0.192	-	**0.063**	**0.067**	**0.068**
P11	0.211	0.209	0.221	0.216	0.205	0.215	0.211	0.215	0.214	0.216	-	-0.004	0.004
P12	0.205	0.203	0.213	0.209	0.200	0.208	0.204	0.208	0.207	0.263	0.228	-	-0.005
P13	0.201	0.199	0.211	0.206	0.196	0.205	0.201	0.206	0.204	0.203	0.226	0.218	-

Table 5: Genetic differentiation between populations of *Ceratoserolis* n. sp. 1. Above diagonal represent standardized F'_{ST} estimated according to $F_{ST}/F_{ST\ max}$ (Tab. 4). Below diagonal: pairwise R_{ST} estimates (Rho, Rousset 1996, calculation without C0M05). Significance of differentation valules was assessed by exact G-tests (10,000 burin, 100 batches with 10,000 MCMC sampling steps). Bold numbers: P<0.01.Significance was assessed by a G-test (10,000 burin, 100 batches with 10,000 MCMC sampling steps). Bold numbers: P<0.01.

Population	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13
P1	-	-0.028	0.014	-0.017	-0.026	-0.013	-0.042	-0.008	-0.006	-0.015	**0.305**	**0.336**	**0.362**
P2	0.002	-	0.010	0.004	0.003	0.003	-0.021	0.011	-0.004	-0.002	**0.274**	**0.297**	**0.321**
P3	0.037	0.010	-	0.029	0.025	0.028	-0.001	-0.038	0.033	0.019	**0.385**	**0.392**	**0.440**
P4	0.008	-0.002	0.003	-	0.005	-0.011	-0.014	-0.002	-0.006	-0.006	**0.322**	**0.346**	**0.357**
P5	-0.002	-0.005	0.003	-0.006	-	-0.014	-0.049	-0.005	0.017	-0.011	**0.293**	**0.314**	**0.312**
P6	0.008	-0.002	-0.002	0.003	-0.011	-	-0.042	0.006	0.010	-0.011	**0.276**	**0.284**	**0.303**
P7	-0.016	-0.001	0.032	0.002	-0.014	-0.011	-	-0.027	-0.010	-0.040	**0.275**	**0.278**	**0.300**
P8	0.018	0.043	0.050	0.014	0.026	0.056	0.033	-	0.026	-0.008	**0.300**	**0.325**	**0.364**
P9	0.003	-0.017	0.012	-0.004	-0.005	-0.002	-0.004	0.037	-	-0.016	**0.324**	**0.335**	**0.344**
P10	0.001	-0.008	0.008	-0.003	-0.007	-0.003	-0.006	0.027	-0.006	-	**0.297**	**0.324**	**0.335**
P11	**0.100**	**0.176**	**0.272**	**0.175**	**0.187**	**0.216**	**0.149**	**0.130**	**0.156**	**0.163**	-	-0.017	0.019
P12	**0.067**	**0.102**	**0.176**	**0.118**	**0.129**	**0.129**	**0.087**	**0.111**	**0.083**	**0.100**	0.011	-	-0.024
P13	**0.106**	**0.172**	**0.273**	**0.169**	**0.183**	**0.224**	**0.150**	**0.126**	**0.151**	**0.162**	0.007	0.043	-

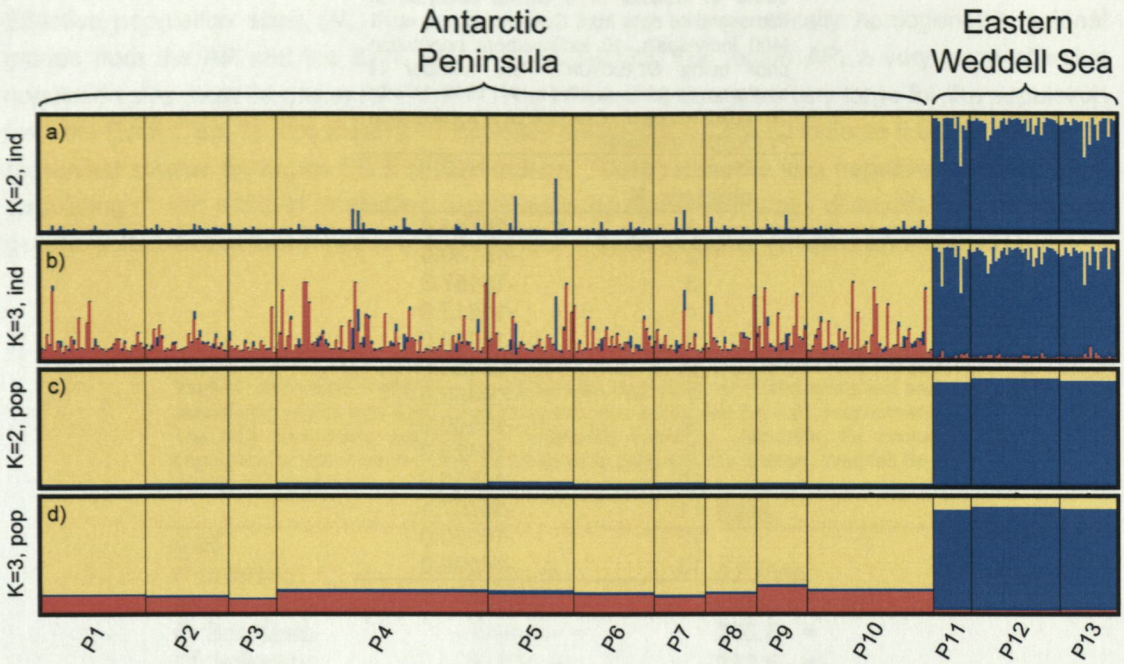


Figure 4: Results of Bayesian cluster analyses of STRUCTURE for all n=400 individuals. Antarctic Peninsula = P1-P10, eastern Weddell Sea = P11-P13. The graphs display the consensus membership coefficients matrices (Q-matrices) for 400 individuals from 13 regions (P1-P13) of *Ceratoserolis* n. sp. 1 using 10 microsatellites. In a) and b) the individual Q-matrix for K=2 and K=3 clusters, respectively, is visualized. The genotype of each individual is represented by a single bar, where the proportion of the colour refers to the probability of assignment to a certain cluster. In c) and d) the average membership coefficient for each of the 13 populations to a certain cluster is visualized for K=2 and K=3.

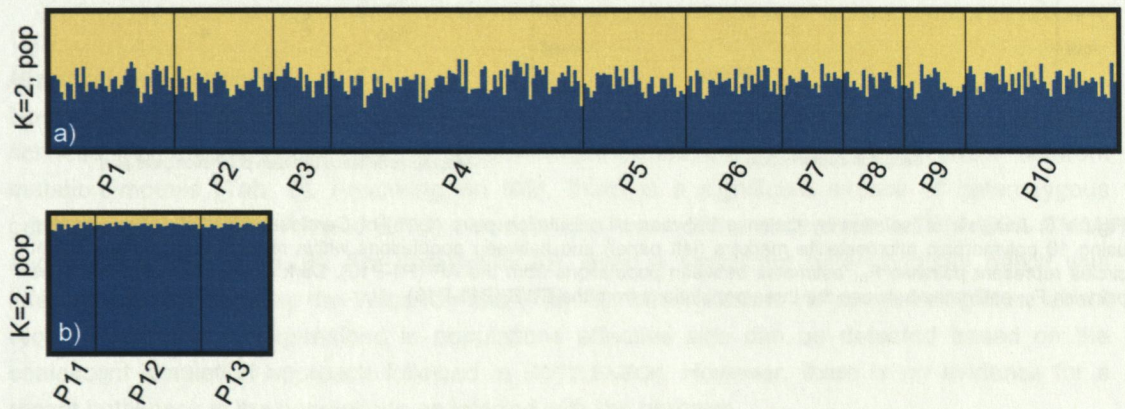


Figure 5: Results of Bayesian cluster analyses of STRUCTURE for individuals within the two major geographical regions separately. The log likelihood for the data given K=1 was maximum. The second highest probability was for K=2 but revealed no indication for intraregional structure.

Table 6: Results of a cluster analysis of microsatellite data from *Ceratoserolis* n. sp. 1 (400 individuals, 10 loci) without population prior using STRUCTURE. The number of clusters was computed from K=1 to K=13 with 10 replicates each. Only the best replicate (ln Pr (D|K)) is displayed.

Number of clusters, K	Ln Pr(D|K)
1	-16784.4
2	-16130.0
3	-16157.3
4	-16212.0
5	-16458.0
6	-16435.0
7	-16458.0
8	-16589.9
9	-16423.9
10	-16656.6
11	-16707.4
12	-16970.0
13	-16899.0

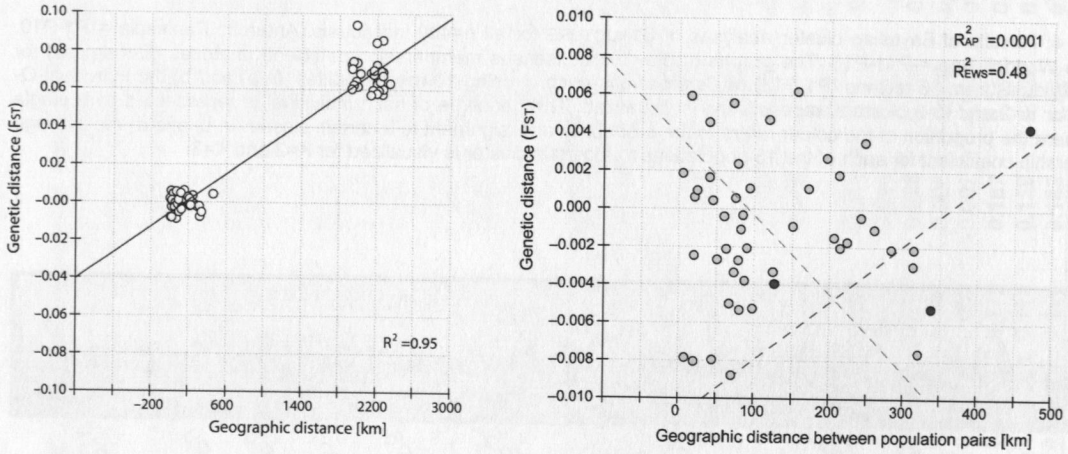


Figure 6: Analysis of isolation by distance between all population pairs (1-13) of *Ceratoserolis* n. sp. 1 investigated using 10 polymorphic microsatellite markers (left panel) and between populations within regions (right panel). Bright circles represent pairwise F_{ST} estimates between populations from the AP (P1-P10). Dark circles represent the three pairwise F_{ST} estimates between the three populations from the EWS (P11-P13).

Present and past effective population sizes

Effective population sizes (N_e) were calculated for the two genetically homogenous regional groups from the AP and the EWS [(P1-P10)(P11-P13)]. For region AP, a very large effective population size must be assumed (Tab. 7). The inferred N_e was also very large for the population from the EWS (Tab. 7). The mean and the lower values for the 5% CI indicate that N_e tends to be somewhat smaller for region EWS in comparison. The N_e estimate was negative for EWS when calculating r^2 with $c \geq 0.01$, indicating that there is no evidence for any disequilibrium caused by genetic drift due to a finite number of parents, see manual of LDNe (Waples and Do 2007).

Table 7: Estimates of effective population size (N_e) using 10 microsatellite loci and considering alleles with a minimum frequency of $c \geq 0.05$ and $c \geq 0.01$, respectively. The 95% confidence intervals (CI parametric bootstrap, Jackknife) for pooled population groups from the Antarctic Peninsula (AP) and the Eastern Weddell Sea (EWS) are listed.

	AP	EWS
0.05		
r^2 (overall)	0.00324	0.01561
N_e	5816.5	-1934.8 (= ∞)*
CI Bootstrap	1084.3 - ∞	290.8 - ∞
CI Jackknife	918.2 - ∞	222.5 - ∞
0.01		
r^2 (overall)	0.00318	0.01587
N_e	30431.5	3432.3
CI Bootstrap	2642.8 - ∞	351.2 - ∞
CI Jackknife	2255.4 - ∞	309.9 - ∞

*a negative N_e is results from the estimation procedure if in a sample there is no evidence for any disequilibrium caused by genetic drift due to a finite number of parents, see manual of LDNe (Waples and Do 2007).

Historical demography

No consistent trends were derived from the correlatives between genetic diversity (H_o) and allelic richness (N_A) based on coalescent simulation performed using BOTTLENECK under different mutation models (Tab. 8). Assuming an IAM, there is a significant excess of heterozygous genotypes (P=0.001 for AP and EWS), however, this was not the case under a TPM (80% SMM) and a strict SMM (Tab. 8). For population AP, even a significant heterozygosity deficit was inferred under a SMM by the Wilcoxon test (Tab. 8). In conclusion, no consistent picture about recent reductions or expansions in populations effective size can be detected based on the coalescent simulation approach followed in BOTTLENECK. However, there is no evidence for a recent bottleneck in the populations as inferred with the program.

Table 8: Statistical tests for significant heterozygosity (*H*) excess or deficiency in the two regional groups of *Ceratoserolis* n. sp. 1 (AP, EWS) applying 10 microsatellite loci and assuming three different mutation models (IAM, TPM, SMM) using the program BOTTLENECK. P-values of the Sign Test and Standardized Differences Test and two-tailed probability for heterozygosity excess or deficiency are based on a 1000 permutations. Significant P-values are printed in bold. The TPM was adjusted to allow for 80% mutations according to a SMM and 20% to an IAM model. The exponents d and e refer to the effect causing the significant P-value (d=significant heterozygosity deficit, e=significant heterozygosity excess).

		Sign Test				Wilcoxon Test
Population	**Model**	Expected no of loci with H excess	Observed no. of loci with H excess	observed no of loci with H deficiency	P	P (two tailed for H excess or deficiency)
AP	IAM	5.91	10	0	0.005	**0.001**^{e}
	TPM	5.86	6	4	0.597	1.000
	SMM	6.00	3	7	0.054	**0.019**^{d}
EWS	IAM	5.91	10	0	0.005	**0.001**^{e}
	TPM	5.92	6	4	0.612	0.322
	SMM	5.84	3	7	0.066	0.105

Mitochondrial DNA

Genetic diversity

The average base pair frequencies for the 563 bp fragment of the COI for the 224 specimens analysed were A=37.5%, C=22.5%, G=16.1% and T=23.9%. Sixty-six polymorphic sites were detected in the alignment, accounting for 65 unique haplotypes (HT1-65, Tab. 8, GenBank accession numbers EU597358-597422). Among the 66 polymorphic sites there were nine replacement substitutions. Among the 66 variable sites, 34 were parsimony-informative, 32 were singleton variable sites, of which 20 were derived from one of the two major haplotype groups (HT1, HT2, HT53) contributing to a star-like shape of the networks. Remarkably, only four of all 66 variable sites were shared between regions (AP, EWS) yet no haplotype was shared between them. This suggests a pattern that is also seen in more detailed analyses below that the division among regions has a long history. Both population clusters are connected by a long branch from which three intermediate clades with few haplotypes split (Fig. 7). The average number of nucleotide differences among populations from the AP (P1-P10) was 2.37±1.30 and 1.06±0.72 for the populations from the EWS (P11-P13). The maximum uncorrected haplotype distance within regions was 1.8% between HT5 and HT6, which occur in population 4 (Elephant Island) and 10 (King George Island). This value is in the expected range of intraspecific COI distances within other serolid isopods but is exceeded by an order of magnitude when comparing *Ceratoserolis* n. sp. 1 to specimens assigned to *C. trilobitoides sensu stricto* even when they occur in sympatry (Held, unpubl.; this study Fig. 1). The identification of a bimodal distribution of pairwise genetic distances, their persistence in sympatry around KGI and the congruence between groups delimited on the basis of molecular (COI and microsatellites) and morphological

characters support the notion that we are comparing species as meaningful, evolutionary units rather than arbitrarily defined entities.

Haplotype diversity ($\hat{H}$) within populations ranged from 0.26 to 1.0, nucleotide diversity (π) from 0.0007-0.0095. Populations furthest south (P12 and P13) showed the lowest haplotype and nucleotide diversities (Tab. 10) Population P11 was equal in the magnitude of $\hat{H}$ and π to populations from the AP. The differences in the median values between the two groups were not large enough to exclude the possibility that the differences are due to random sampling variability.

In the EWS populations, divergence is shallow with a mean number of 1.06 pairwise differences among haplotypes. The maximum intraregional p-distance is low (0.9%) and found between haplotypes 58/59 and HT60 (0.9%), which occur in population 11 (northern most population from the EWS).

The average number of nucleotide differences between the two groups is 11.03 corresponding to 1.96% uncorrected sequence divergence.

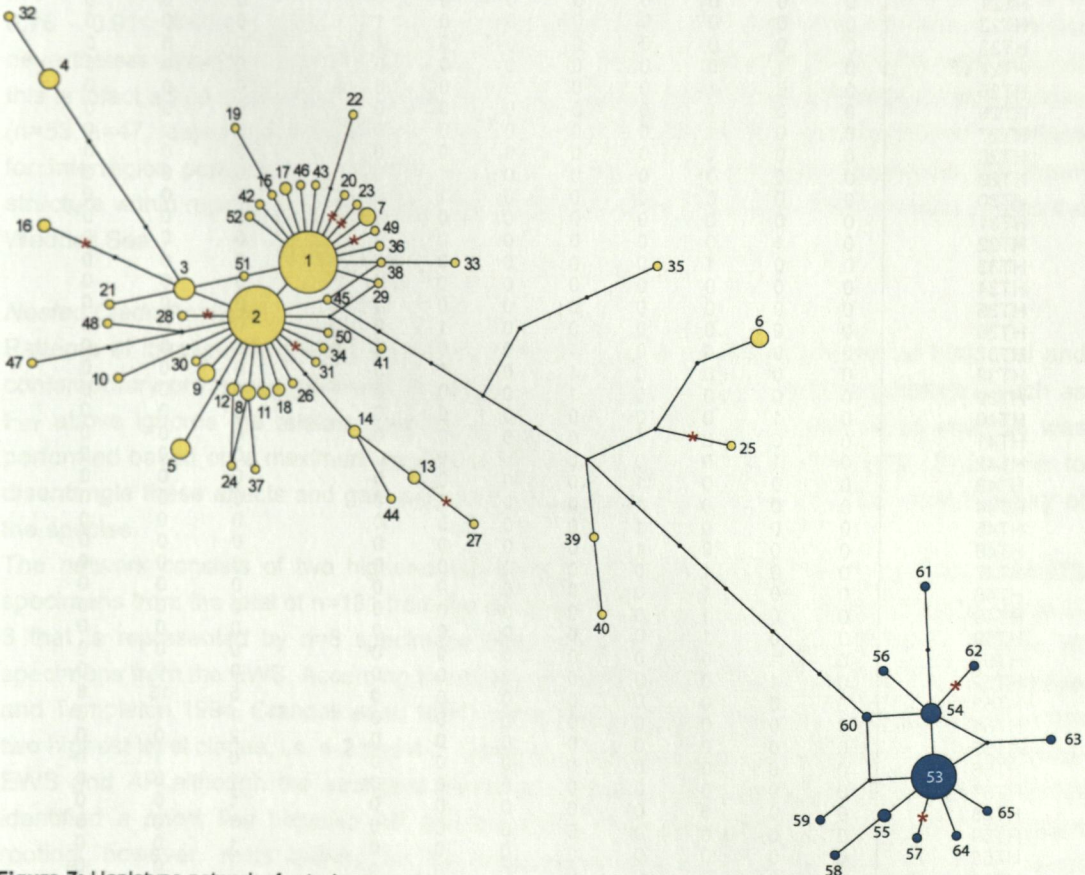


Figure 7: Haplotype network of cytochrome oxidase 1 sequences of 224 specimens of *Ceratoserolis* n. sp. 1 (circles, 1-65) as deferred by statistical parsimony (TCS, 95% connection limit). Yellow haplotypes are derived from populations around the Antarctic Peninsula, blue haplotypes from the eastern Weddell Sea. Every black circle represents an unsampled, hypothetical haplotype. Red asterisks indicate replacement substitutions. The diameter of circles corresponds to the number of specimens that share the particular haplotype (see Tab. 9 for precise numbers).

Table 9: Distribution of the 224 COI sequences on 65 distinct haplotypes for the populations for *Ceratoserolis* n. sp. 1 from the Antarctic Peninsula and from the eastern Weddell Sea

Haplotype	Antarctic Peninsula								Eastern Weddell Sea		
	1	2	4	5	6	8	9	10	11	12	13
HT1	1	7	17	2	5	2	6	11	0	0	0
HT2	1	7	7	5	3	1	7	17	0	0	0
HT3	0	2	3	1	0	0	0	0	0	0	0
HT4	0	0	2	1	0	0	1	1	0	0	0
HT5	0	0	4	0	0	0	0	1	0	0	0
HT6	1	0	0	1	0	1	0	1	0	0	0
HT7	0	1	2	0	0	0	0	1	0	0	0
HT8	0	1	2	0	0	0	0	0	0	0	0
HT9	0	1	1	0	0	0	0	2	0	0	0
HT10	0	0	0	0	0	0	0	2	0	0	0
HT11	0	1	0	0	0	0	0	1	0	0	0
HT12	0	0	2	0	0	0	0	0	0	0	0
HT13	0	0	2	0	0	0	0	0	0	0	0
HT14	0	0	0	0	0	0	0	2	0	0	0
HT15	0	0	0	0	0	0	1	0	0	0	0
HT16	0	0	0	1	0	1	0	0	0	0	0
HT17	0	0	2	0	0	0	0	0	0	0	0
HT18	0	2	0	0	0	0	0	0	0	0	0
HT19	0	0	0	0	0	0	0	1	0	0	0
HT20	0	0	0	0	0	0	0	1	0	0	0
HT21	0	0	0	0	0	0	0	1	0	0	0
HT22	0	0	0	0	0	0	0	1	0	0	0
HT23	0	0	0	0	0	0	0	1	0	0	0
HT24	0	1	0	0	0	0	0	0	0	0	0
HT25	0	0	1	0	0	0	0	0	0	0	0
HT26	0	0	1	0	0	0	0	0	0	0	0
HT27	0	0	1	0	0	0	0	0	0	0	0
HT28	0	0	1	0	0	0	0	0	0	0	0
HT29	0	0	0	0	0	0	1	0	0	0	0
HT30	0	0	0	0	0	0	1	0	0	0	0
HT31	0	0	0	0	0	0	0	1	0	0	0
HT32	0	1	0	0	0	0	0	0	0	0	0
HT33	0	0	1	0	0	0	0	0	0	0	0
HT34	0	0	0	0	0	0	0	1	0	0	0
HT35	0	0	0	0	0	0	0	1	0	0	0
HT36	0	0	0	0	0	0	1	0	0	0	0
HT37	0	0	0	0	1	0	0	0	0	0	0
HT38	0	0	0	0	1	0	0	0	0	0	0
HT39	0	0	0	0	1	0	0	0	0	0	0
HT40	0	1	0	0	0	0	0	0	0	0	0
HT41	0	0	0	1	0	0	0	0	0	0	0
HT42	0	1	0	0	0	0	0	0	0	0	0
HT43	0	0	0	1	0	0	0	0	0	0	0
HT44	0	0	0	1	0	0	0	0	0	0	0
HT45	0	0	0	1	0	0	0	0	0	0	0
HT46	0	0	0	1	0	0	0	0	0	0	0
HT47	0	0	1	0	0	0	0	0	0	0	0
HT48	0	0	1	0	0	0	0	0	0	0	0
HT49	0	0	1	0	0	0	0	0	0	0	0
HT50	0	0	1	0	0	0	0	0	0	0	0
HT51	0	1	0	0	0	0	0	0	0	0	0
HT52	0	1	0	0	0	0	0	0	0	0	0
HT53	0	0	0	0	0	0	0	0	5	13	8
HT54	0	0	0	0	0	0	0	0	1	1	3
HT55	0	0	0	0	0	0	0	0	2	0	0
HT56	0	0	0	0	0	0	0	0	0	1	0
HT57	0	0	0	0	0	0	0	0	1	0	0
HT58	0	0	0	0	0	0	0	0	1	0	0
HT59	0	0	0	0	0	0	0	0	1	0	0
HT60	0	0	0	0	0	0	0	0	1	0	0
HT61	0	0	0	0	0	0	0	0	1	0	0
HT62	0	0	0	0	0	0	0	0	1	0	0
HT63	0	0	0	0	0	0	0	0	1	0	0
HT64	0	0	0	0	0	0	0	0	0	0	1
HT65	0	0	0	0	0	0	0	0	0	0	1

Genetic differentiation

Results of the hierarchical AMOVA based on either a matrix of pairwise nucleotide distance between haplotypes or the unweighted haplotype frequencies, consistently revealed a high and significant proportion of the variance being distributed within populations and between regions (Tab. 11). No significant amount of variation was partitioned among populations within regions ($p>0,1$).

The absence of any shared haplotypes between the two geographical regions results in very high pairwise F_{ST} differentiation estimates (Tab. 12). Pairwise F_{ST} estimates based on the haplotype frequencies and on the distance matrix were high and significant for population comparisons between regions. All populations from the AP have significantly different nucleotide and haplotype frequencies from any population from the EWS. Only P1 and P11 are an exception because the null hypothesis of equal haplotype frequencies cannot be rejected ($F_{ST}=0.069$), although nucleotide differences are highly significant between the two populations (Tab. 12). Rather than being a feature, this is the consequence of sample size of n=3 for P1. Conversely, the populations are much less divergent within the same regions with F_{ST} values ranging from (F_{ST} 0.76 - 0.911, $P<0.01$) (Tab. 12). Only populations P4 and P10 from the AP have a low but nevertheless significant differentiation ($F_{ST}=0.029$, $P<0.05$). In this case, it must be regarded that this is infact a true differentiation signal, as P4 and P10 are populations with largest sample sizes (n=53, n=47, respectively). All distance-based calculations yielded high and significant F_{ST} values for interregion population comparisons only. In summary, there is no widespread significant structure within regions in the mitochondrial COI gene but a very strong differentiation across the Weddell Sea.

Nested Clade Analysis

Patterns of intraspecific genetic diversity are the result of a complex interplay of historical and contemporary processes (reviewed in Neigel 1997). Because single summary statistics such as F_{ST} above ignores the relative contribution of either processes, a nested clade analysis was performed based on a maximum parsimony network Templeton et al. (1995) (Fig. 6) in order to disentangle these effects and gain a clearer picture about driving forces in the recent history of the species.

The network consists of two higher-order clades, 4-1 and 4-2. Clade 4-1 consists of n=173 specimens from the total of n=181 from the AP only. Clade 4-2 encompasses the interior clade 3-3 that is represented by n=8 specimens from AP only, and tip clade 4-4, which houses all specimens from the EWS. According to coalescent theory and common usage of NCA (Castelloe and Templeton 1994, Crandall et al. 1994) one would place the root of the network between the two highest level clades, i.e. 4-2 and 4-1. Doing so creates a clade 4-2 that unites clades from the EWS and AP although the strongest barrier to dispersal, the Weddell Sea itself that can be identified *a priori*, lies between AP and the EWS. The correctness of this implicit mid-point rooting, however, rests heavily on the assumption that sampling is complete or at least geographically unbiased (Dahlgren et al. 2000). Given that the sample size from AP was almost five times larger than from the EWS and large parts of potentially suitable shelf habitat in East Antarctica and the western Weddell Sea were unavailable, rooting the network with an outgroup is preferred because it is more robust against geographic sampling bias and missing haplotypes.

Rooting the network with congeners places the root somewhere along the long branch connecting clade 3-4 with the rest of the network, thus creating a binary split between haplotypes occuring exclusively around the AP or the EWS, respectively.
The clade with the *Ceratoserolis* outgroup species splits from the branch connecting clade 3-3 and 3-4. This information was used to define 4-2 as interior and 4-1 as tip-clade in the network. Ambiguous connections of the parsimony network (Fig. 7) had no influence on the results of the NCA. The mtDNA clade variation at different geographical locations revealed a significant level of phylogeographic structure for the total cladogram of *Ceratoserolis* n. sp. 1 haplotypes (Tab. 13, Tab. 14) as suggested by significant chi-square analyses. According to the most recent inference key for NCA the topology of clades observed structure is likely to be the result of allopatric fragmentation. The large number of steps connecting [3-4, 3-3 and 4-1] and [4-2] provide additional support for this scenario. However, due to the lack of samples from geographically intermediate positions (i.e. Filchner Ice Shelf) the two regions it cannot be excluded that this pattern is due isolation by distance from this data set alone.

Table 10: Haplotype diversity ($\hat{H}$) and nucleotide diversity (π) according to Nei (1987) for a 563 bp fragment of the mitochondrial cytochrome c oxidase subunit I for 224 specimens from 11 populations of *Ceratoserolis trilobitoides*. Populations 1-10 come from the tip of the Antarctic Peninsula. Populations 11-13 from the eastern Weddell Sea.

Population	Number of sequences	Number of Haplotypes	$\hat{H}$	π
P1	3	3	1.0±0.27	0.0095±0.0078
P2	28	14	0.88±0.04	0.0038±0.0024
P4	53	20	0.88±0.04	0.0044±0.0027
P5	16	11	0.91±0.06	0.0056±0.0034
P6	11	5	0.76±0.11	0.0036±0.0024
P8	5	4	0.90±0.16	0.0078±0.0054
P9	18	7	0.76±0.07	0.0025±0.0018
P10	47	18	0.82±0.04	0.0039±0.0025
P11	15	10	0.90±0.07	0.0036±0.0024
P12	15	3	0.26±0.14	0.0007±0.0008
P13	13	4	0.60±0.13	0.0012±0.0011
AP (P1-P10)	181	52	0.849±0.019	0.0042±0.0025
EWS (P11-P13)	43	13	0.628±0.082	0.0019±0.0014

Table 11: Hierarchical analysis of molecular variance among *Ceratoserolis* n. sp. 1 populations within and between two regions of the Southern Ocean. Table includes calculation for pairwise differences matrix and haplotype frequency matrix. P refers to the probability of randomly getting the same result based on 10,000 independent permutations.

Component of differentiation	df	variance [%]	Φ statistics	P
pairwise differences matrix				
Among regions	1	81.42	Φ_{CT}=0.814	0.005
Among populations within regions	9	0.02	Φ_{SC}=0.001	0.357
within populations	213	18.55	Φ_{ST}=0.814	0.000
Haplotype frequencies				
Among regions	1	24.15	Φ_{CT}=0.241	0.006
Among populations within regions	9	0.85	Φ_{SC}=0.011	0.109
within populations	213	75.00	Φ_{ST}=0.250	0.000

Table 12: Pairwise F_{ST} estimates and exact test of population differentiation among the eleven populations of *C. trilobitoides* tested for mtDNA COI differences. The lower matrix was calculated using pairwise nucleotide differences, the upper matrix was calculated using haplotype frequencies. Exact test probabilities of non-differentiation was calculated using 1,000,000 MCMC steps and initial 20,000 dememorization steps (burn-in).

Population	P1	P2	P4	P5	P6	P8	P9	P10	P11	P12	P13
P1	-	-0.113	-0.084	-0.133	-0.131	-0.288	-0.127	-0.118	0.069	**0.577***	**0.278***
P2	0.099	-	-	-0.011	-0.010	-0.048	-0.007	-0.004	**0.111****	**0.385****	**0.239****
P4	0.088	-0.008	-	0.024	-0.006	-0.048	0.020	**0.029***	**0.116****	**0.360****	**0.235****
P5	-0.061	0.004	0.011	-	0.023	-0.049	-0.003	-0.009	**0.098****	**0.412****	**0.239****
P6	0.039	-0.025	-0.020	0.012	-	-0.081	-0.029	0.001	**0.167****	**0.514****	**0.320****
P8	-0.291	0.042	0.045	-0.077	0.011	-	-0.044	-0.029	**0.103***	**0.538****	**0.284****
P9	0.179	-0.012	-0.014	0.017	-0.023	0.080	-	-0.017	**0.172****	**0.476****	**0.311****
P10	0.078	-0.005	-0.002	0.000	-0.010	0.044	-0.008	-	**0.147****	**0.394****	**0.267****
P11	**0.760****	**0.815****	**0.793****	**0.774****	**0.820****	**0.768****	**0.848****	**0.810****	-	0.185	0.037
P12	**0.894****	**0.860****	**0.821****	**0.836****	**0.900****	**0.876****	**0.911****	**0.842****	0.035	-	0.053
P13	**0.858****	**0.846****	**0.810****	**0.815****	**0.878****	**0.845****	**0.896****	**0.831****	0.005	-0.023	-

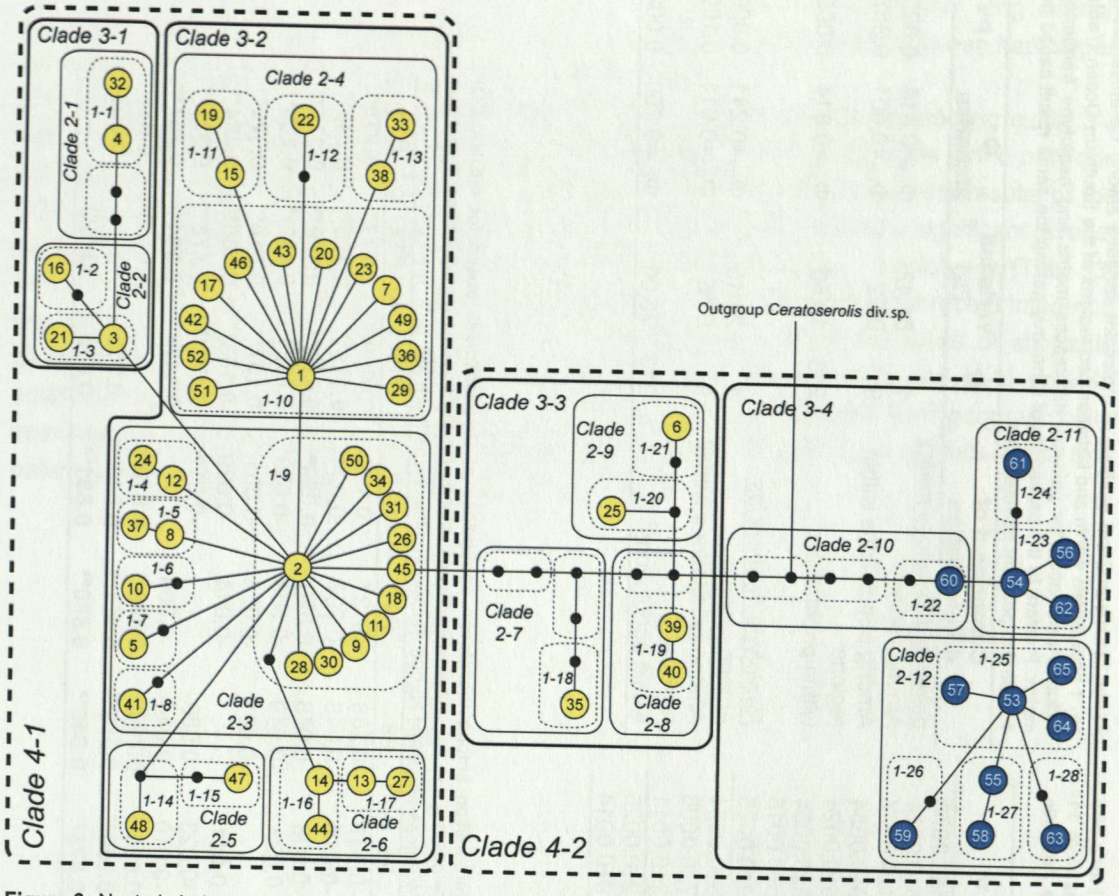


Figure 8: Nested clade network for cytochrome oxidase subunit I haplotypes of n=224 *Ceratoserolis* n. sp. 1 specimens observed in 11 populations based on the topology of the strict parsimony network from Fig. 7. Haplotypes from the Antarctic Peninsula are yellow (1-52), from the Eastern Weddell Sea blue (53-65). Haplotypes in the statistical parsimony network are connected by one mutation. Missing haplotypes are indicated by black dots.

Table 13: Nested clade contingency analysis of geographical association for COI data from n=224 *Ceratoserolis* n. sp. 1 specimens based on 10,000 permutations. Clades that do not contain either genetic or geographical variation are excluded. Significant P-values are printed in bold.

Clade	Permutational chi-square statistic	Probability
1-1	6.000	0.674
1-3	0.875	1.000
1-4	3.000	0.331
1-5	4.000	0.498
1-9	48.77	0.764
1-10	78.194	0.507
1-16	3.000	0.331
1-19	2.000	1.000
1-23	5.600	0.284
1-25	7.919	0.171
2-2	6.107	0.163
2-3	36.010	0.380
2-4	13.556	0.680
2-6	6.000	0.100
2-9	5.000	1.000
2-11	1.905	1.000
2-12	12.811	**0.004**
3-1	2.361	1.000
3-2	12.634	0.860
3-3	11.200	0.929
3-4	2.426	0.824
4-1	6.415	0.468
4-2	51.000	**0.000**
Total Cladogram	183.204	**0.000**

Table 14: Significant results of the nested clade analysis. P-values refer to the probability of observing getting a chi-square value of greater or equal to the observed one. The inference chain and the biological interpretation according to the inference key are listed.
* refers to ambiguous results

Clades being tested	x^2	P-value	Inference chain
2-12	12.8	0.0035	1N; 2N; 11N; 17N: inconclusive results
4-2	51.0	<0.0001	1Y; 19Y; 20Y/N: allopatric fragmentation / IBD*
total cladogram	183.2	<0.0001	1N; 2Y; 3N; 4Y; 9Y; 10 Y/N: allopatric fragmentation / IBD*

Historical demography

Historical demography was evaluated separately for the two population groups (AP, EWS). Results for Tajima's *D* and Fu's F_s statistics were significant and negative for both groups indicating an excess of low frequency mutations among haplotypes in both regions (Tab. 14), which is typical for a recent population expansion. The star-like shape with many singleton mutations derived from the major haplotypes (HT1, HT2, HT53) is also indicative of an expanding population. Results of the mismatch distribution were in agreement with expectations under a sudden expansion model (Rogers and Harpending 1992). Greater negative values for both statistics were observed for the samples for the AP group (*D*=-2.23 and F_S=-23.87, respectively) indicating that the genomic signatures of a population expansion are more pronounced in this region. The mismatch analysis, revealed a strong unimodal distribuion of pairwise genetic distances, which is typical for a non-equilibrium population under a sudden expansion. Therefore, the expectation of an expanding population cannot be rejected (SSD P-value >0.05, Tab. 14).

Estimates of the time t since the population expansion for each of the two groups were calculated according to Rogers and Harpending (1992) using the formula $T=2\mu\tau$, where τ was inferred from the mismatch distribution and a range of values for the mutation rate μ were applied. Values of τ calculated from the mismatch distributions were $\tau = 0.93$ (95% CI: 0.29-4.15) for the EWS and $\tau = 1.44$ (95% CI: 0.18-2.80) for the AP (Tab. 15). Using reported mutation rates of 1.4% per Myr and 2.3% per Myr (average 1.9% Myr) (Knowlton and Weigt 1998, Schubart et al. 1998, see Held 2001) the time t since expansion for population AP was calculated as t=43.5 Kyr (95% CI: 11.2 Kyr – 245.7 Kyr). For population EWS the time was calculated as t=67.3 Kyr (95% CI: 6.9 Kyr – 165.8 Kyr).

The postulated expansion of population sizes in recent times in both regions is further supported by the fact that all nine replacement substitutions occurred in both regions and along terminal branches in the network. Silent subsitutions, i.e. substitutions on the DNA level that lead to the incorporation of an unchanged amino acid in the gene product, are protected from effects of natural selection and are thus expected to be nearly neutral. On the other hand, the conserved nature of COI on an amino acid level shows that the protein is most likely under stabilising selection. On theoretical grounds, expanding populations are expected to experience relaxed selection and thus feature elevated rates of change in traits under selection. Finding all amino acid substitutions restricted to terminal as opposed to internal branches is in line with the prediction of a period of relaxed selection and, by extension, population expansion in both regions in recent times because a comparison of external vs internal branches roughly means a comparison of recent vs ancient events.

The signature of recent population expansion is stronger around AP than in the EWS, which is in congruence with a more pronounced or a more recent population expansion. A population expansion in recent times is in agreement with time estimates based on molecular clock estimates and the placement of mutations in characters under selection on the network.

Table 14: Tajima's *D*, Fu's F_S, corresponding P-values, mismatch analysis parameter estimates.

Sample	Tajima's *D*		Fu's F_S		Mismatch analysis							
	D	P	F_S	P	τ (mean)	τ (CI)	θ_0	θ_0 (CI)	θ_1	θ_1 (CI)	SSD	P (SSD)
AP	-2.23	0.00	-26.87	0.00	0.93	0.29-4.15	1.24	0.00-1.62	∞	7.34-∞	0.00	0.52
EWS	-1.98	0.01	-10.15	0.00	1.44	0.18-2.80	0.01	0.00-0.04	2.87	1.93-∞	0.00	0.86

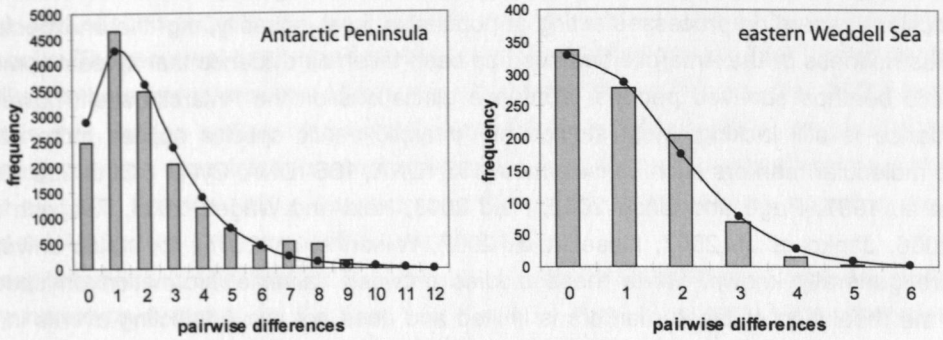


Figure 9: Mismatch distribution plots showing observed pairwise differences (bars), and the expected mismatch distribution under the sudden expansion model (solid line) of the cytochrome oxidase c subunit I haplotypes in *Ceratoserolis* n. sp. 1 from the Antarctic Peninsula (n=181, left panel) and the eastern Weddell Sea (n=43, right panel).

Table 15: Estimates of the time since population expansion in the two population groups of *Ceratoserolis* n. sp. 1 (AP and EWS) based on different molecular clocks from Crustacea (Knowlton et al. 1993, Knowlton and Weigt 1998, Schubart et al. 1998, see also Held 2003). Expansion times T are estimated for the average τ (T=2μ τ) and the upper ($_{max}$) and the lower ($_{min}$) estimates 95% confidence interval estimates of τ. Mutation rate μ is given per site per year.

	τ	T_{min} ($\mu=2.6*10^{-8}$ y^{-1})	T_{avr} ($\mu=2.0*10^{-8}$ y^{-1})	T_{max} ($\mu=1.4*10^{-8}$ y^{-1})
AP $_{min}$	0.29	9,906 y	12,877 y	18,397 y
AP $_{avr}$	0.93	31,766 y	41,296 y	58,994 y
AP $_{max}$	4.15	141,754 y	184,280 y	263,257 y
EWS $_{min}$	0.18	6,148 y	7,992 y	11,418 y
EWS $_{avr}$	1.44	49,187 y	63,943 y	91,347 y
EWS $_{max}$	2.80	95,641 y	124,333 y	177,619 y

Discussion

In this study, we directly estimated population structure, diversity and population demography for 400 specimens of *Ceratoserolis* n. sp. 1 using 10 hypervariable microsatellite loci and a fragment of the mtDNA gene across several geographic scales. Although the investigation of population subdivision, speciation and the underlying microevolutionary processes in general have been identified as major topics in Antarctic research (e.g. Allcock et al. 1997, Gaffney 2000, Page and Linse 2002, Rogers et al. 2007, Wilson et al. 2007) this is the first and most detailed study on population genetic processes in the Antarctic benthos using nuclear microsatellite data.

Until recently, the requirements of fast evolving markers has greatly forced researchers to infer conclusions regarding processes acting at population level indirectly, e.g. the unexpectedly high species richness of the Antarctic benthos has been taken as evidence that at least some of the Antarctic benthos survived periods of intense glaciations on the Antarctic shelf, however, direct evidence is still lacking. Most studies on Antarctic bentic species applied more easily obtainable molecular markers such as isozymes, 16S rDNA, 18S rDNA, Cyt b, COI or 28S rDNA (Allcock et al. 1997, Page and Linse 2002, Held 2003, Held and Wägele 2005, Raupach and Wägele 2006, Janko et al. 2007, Linse et al. 2007, Wilson et al. 2007), for which universal primers are generally known. While these studies provided valuable information on species structure, the resolution of these markers is limited and does not allow detecting events in the recent population history. Such markers with high variability below species level, most importantly, co-dominant and bi-parentally inherited microsatellite markers have only been applied to few commercially important species and Antarctic vertebrates. Recent improvements in marker isolation techniques and screening strategies (Zane et al. 2002, Glenn and Schable 2005, Nolte et al. 2005, Held and Leese 2007, Leese et al. 2008, Beszteri et al. in prep.) have brought the fast-evolving microsatellites encoded in the nuclear genome as an independent, complementary class of informative markers within reach for studies on non-model organisms.

Marker characteristics

Due to their different mutation rates mitochondrial and microsatellite data resolve different time windows but they complement each other well. For population genetic and phylogenetic inferences on processes, research of two different markers systems are important in particular to detect effects of incomplete lineage sorting and bias of shared ancestral polymorphisms (Maddison and Knowles 2006).

All molecular markers lose some information over time as it gets overprinted several times, making identity by descent and identity by state indistinguishable. This homoplasy expresses itself in microsatellites in the form of overlappping size-ranges for the alleles under conditions of large genetic differentiation (Nauta and Weissing 1996). In addition, extremely long alleles may fail to amplify in vitro. This and the four times larger effective population size of the nuclear microsatellites explain an overlap in allele length found in the microsatellite data between populations on either side of the Weddell Sea that do not share a single haplotype of the more slowly evolving COI gene. Nevertheless, despite the overlap in microsatellite allele sizes, differentiation between AP and EWS is highly significant in both marker systems.

Population structure of *Ceratoserolis* n. sp. 1

The overarching impression conveyed by the two marker systems is a deep divergence between regions on either side of the Weddell Sea (AP and EWS) and surprisingly little differentiation between regions. Microsatellites and mitochondrial data differ with respect to their evolutionary dynamics. On the one hand, microsatellites have a much higher mutation rate (Levinson and Gutman 1987, Weber and Wong 1993) on the other hand, genes encoded on the mitochondrial genome have a four times smaller effective population size (Birkey et al. 1983) and are thus more susceptible to the effects of random genetic drift (Wright 1943, Ballard and Whitlock 2004 for review). This effect is expected to be particularly acute in species undergoing frequent population bottlenecks. The recurrent large-scale glaciations of the habitat occupied by *Ceratoserolis* n. sp. 1 are likely to be a major force that affects genetic polymorphisms (Thatje et al. 2005). A practical upshot is that lineage sorting of ancestral polymorphisms and increase in genetic distances between groups are expected to happen faster under conditions encountered during a bottleneck (Nei et al. 1975, Hedrick 1999). The reciprocal monophyly of haplotypes occurring exclusively around the AP and the EWS, respectively, support this.

Despite the strong differentiation between the AP and EWS haplotypes, they are more similar to any haplotype assigned to *Ceratoserolis* n. sp. 1 than they are to the closest relative, *Ceratoserolis trilobitoides sensu stricto* (group 2 in Held 2003).

The time estimates derived from the mitochondrial data imply that effective gene flow is and has been interrupted or extremely low for a time far exceeding the last glaciations on the Antarctic shelf. In more recent times, however, there is evidence for significant and rapid population expansion in both regions in the mitochondrial data. These time estimates agree well with dates for the LGM estimated primarily by calibrated radiocarbon dating (Anderson et al. 2002, Ingolfsson et al. 2004 for review), especially if the elevated observed mutation rates following population bottlenecks are taken into account. The position of the replacement substitutions exclusively along terminal branches which tend to stand for recent events is consistent with elevated observed mutation rates under conditions of population expansion (effects of drift negligible). The excess of low-frequency haplotypes connected to few very frequent, internal haplotypes is another finding supporting that populations are recovering from severe reduction in size around the LGM.

Due to the fact that populations were sampled in two major geographic regions only and no intermediate samples are incorporated in the study design, e.g. from the Western Antarctic Peninsula or the Filchner shelf, it cannot be excluded that results are due to a simple isolation by distance scenario. Three points, however, suggest that allopatric fragmentation is a much more plausible scenario: (1) specimens that were found on the Western part of the Antarctic Peninsula do not belong to the *Ceratoserolis* n. sp. 1 group, but to *Ceratoserolis trilobitoides sensu stricto* (Held, unpubl. data), (2) there is morphological evidence that members of the genus *Ceratoserolis* sampled from the Filchner shelf by J.-Wolfgang Wägele, are morphologyically distinct from *Ceratoserolis* n. sp. 1. What remains unclear is whether they are part of *C. trilobitoides s. str.* or yet another species (Held, in prep.). No DNA-based analysis were possible due to the preservation in formaldehyde, however, (3) the lack of any detectable trend of IBD

within regions, even over several hundreds of kilometres, makes it unlikely that IBD without additional major barriers to dispersal is the cause of the observed partitioning of genetic polymorphisms.

The rare haplotypes branching off of the long branch connecting the AP and EWS terminal subnetworks (clade 3-3) are not part of the recent, expanding populations. Instead, they connect to the oldest part of the network as evidenced by the outgroup rooting. They can be interpreted as descendants from haplotypes that were present around the Antarctic Peninsula before the LGM. The low frequency of these ancient haplotypes (clade 3-3) today cannot be taken as evidence that the populations were less diverse before the LGM. The population bottleneck that we assume the LGM has imposed on the population has purged haplotypes from the population randomly so that the clade 3-3 are merely the haplotypes that survived the genetic drift associated with the bottleneck. The fact that pre-LGM haplotypes (clade 3-3) are today found only around the AP but not in the EWS may be interpreted as a consequence that the AP is structurally more diverse and has the mildest climatic conditions (Palma et al. 2007).

Mobility of *Ceratoserolis* n. sp. 1 – a comparison

In view of the strictly benthic life strategy of *Ceratoserolis* n. sp. 1, the lack of any detectable population substructure within the two geographical regions is unexpected (Wägele 1986, Held 2003). In particular for populations from the EWS (P11 – P13), which are separated by up to 500 kilometres along the High Antarctic shelf. The high resolving power of the markers used makes it unlikely that the absence of a proof of differentiation resulted from too low resolution (see Held and Leese 2007). It must rather be asked how these specimens are capable of maintaining gene flow. One possibility is that the strong currents occasionally carry specimens over distances they may not be capable of dispersing actively on the shelf. As maturation takes several years in these species (Luxmoore 1982, Wägele 1987) the distance an individual may cover during its life span may be unexpectedly large. It must be stated, however, that even migration of very few specimens per generation may suffice to counteract differentiation in two very large populations, e.g. with very large effective population sizes, as are likely for *Ceratoserolis* n. sp. 1 (simulation results Leese, unpublished, Robin Waples pers. comm.).

Implications for speciation events on the Antarctic shelf

Recurrent glaciations are regarded as a key element in structuring population diversity. Their impact on the zoogeographic distribution of species and their respective genetic structure is well researched for the Northern Hemisphere (Dynesius and Jansson 2000, Hewitt 2000). There, disjunct refuge areas during the glacial maxima have given rise to several genetically distinct lineages which co-exist in what is a continuous habitat today, making recolonization pathways traceable in extant genomes (Bernatchez et al. 1999, Avise 2000, Hewitt 2000, Liebers et al. 2004). For the Southern Ocean, the scarcity of population genetic studies on benthic Southern Ocean organisms greatly limits our understanding of the influences of recurrent glaciations on species structure. The first study using two fast evolving genes to compare genomic patterns of population demography was conducted by Janko et al. (2007), who found striking differences in population demography between pelagic and benthic nothotenioid species, which they causally linked to Pleistocene glaciations. However, also from a palaeogeologists view, the debate about

the extension of grounded shelf ice during the Pleistocene ice ages and the timing is controversial (see Denton et al. 1991, Ingolfsson 2002, Anderson et al. 2002, Huybrechts 2002, Evans et al. 2004, Heroy and Anderson 2005). Most studies deal with the extents of the LGM, as earlier ice ages are much more difficult to estimate precisely, because the LGM has left the most pronounced imprint on the geosphere. Several studies document and simulations suggest that the peripheral domes of the Antarctic Ice Sheet were 500-1000 m thicker than at present and provide evidence that grounded ice extended to the continental shelf break around most of Antarctica (Denton et al. 1991, Huybrechts 2002). However, other studies state that ice extent was less severe (see Ingolfsson 2004 and references therein). Several recent studies investigating the extent of grounded shelf ice at LGM provide strong evidence that at several sites the grounding ice did not exceed as far as to the continental edge (see Anderson 2002 for review). These regions include parts of the outer Ross Ice shelf, King George V Land, Prydz Bay and the shelf off Queen Maud Land (Andersen 2002, Fig. 10). The western Antarctic Penisula was almost covered by grounded ice to the continental shelf (Heroy and Anderson 2005). Little data is available from the South Shetland Islands and Elephant Island. However, based on the data available, Anderson (pers. comm.) suspects that to the north of the South Shetland Islands and the platform off Elephant Islands were ice-free during glacial maxima. These regions could have acted as potential refuge areas. It must be stated, however, that the absence of grounded ice is just an essential, however not sufficient condition (Thatje et al. 2008). Huge multiannual ice layers covered the open sea, limiting primary production and thus bentho-pelagic energy coupling. These effects are increased further to the south while the tip of the Antarctic Peninsula at approximately 60°S was less affected by the seasonality. For the shelf region off Queen Maude Land, there is evidence from ice core analyses that even during glacial maxima a stable and large *Polynya* was present and might have promoted a primary production in the surface waters and energy coupling to the benthos. The data of this study is in accordance with such a scenario.

In summary, the data in this study provide, for the first time, empirical evidence for a common benthic macroinvertebrate species that despite the large-scale severe glaciations survival on the shelf was possible over evolutionary times.

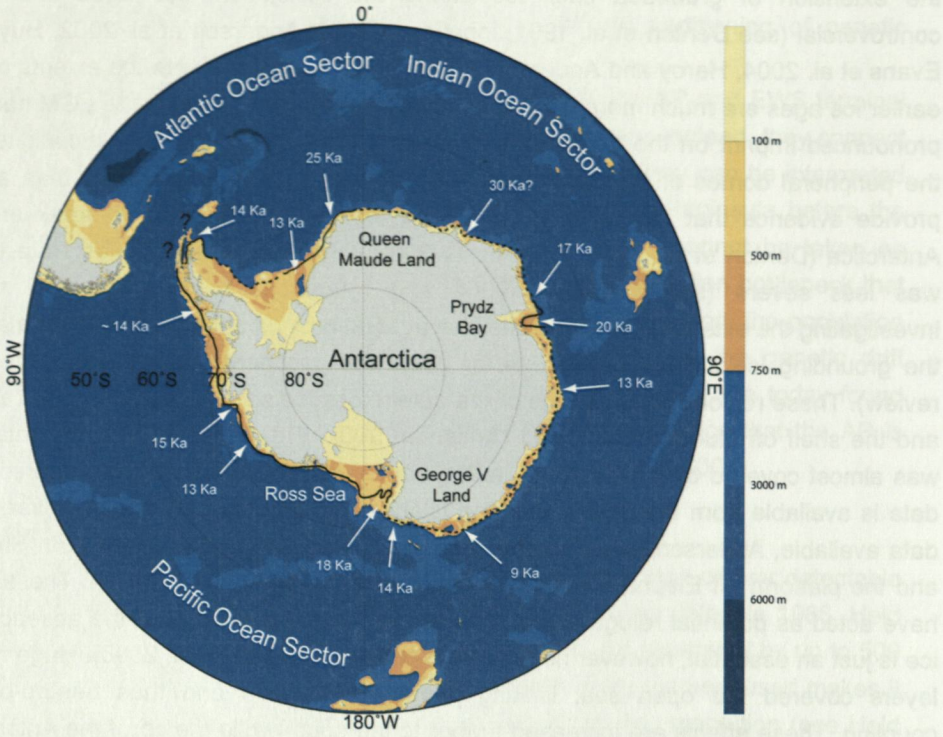


Figure 10: Maximum extent of grounding shelf ice around the Antarctic continent during the last glacial maximum (black line) according to Anderson et al. (2002), Evans et al. (2004), Heroy and Anderson (2005). Dashed parts of the line refer to estimates of the extent of grounded ice. White numbers are estimates of the minimum age (in 1,000 years) of ice-sheet retreat from the Antarctic shelf during deglaciation based on radiocarbon dating (Anderson et al. 2002). Bathymetry data: Ocean Data View.

Conclusions

We provide for the first time data from fast evolving microsatellite markers and a mtDNA marker to link population structure and demography of an Antarctic model species with a highly immobile life strategy to the recurrent glacial events. Our data support strong divergence of *Ceratoserolis* n. sp. 1 from either side of the Weddell Sea, but prove populations within regions to be genetically relatively homogenous. Both populations reveal very large present effective population sizes. The dating of the split between the two population groups predates the LGM and supports that populations of *Ceratoserolis* n. sp. 1 survived the LGM in independent refugia on the Antarctic shelf. The slower evolving mtDNA marker shows significant signatures of a recent population expansion after the LGM. The data presented strongly support the glacial refugia theory, indicating that Antarctic species richness for benthic taxa may largely be explained by lineage sorting of populations in small, independent refugia on the shelf during glacial maxima. Results are in accordance with geophysical data that suggest one putative ice-free refuge area for the shelf of the eastern Weddell Sea or off Queen Maude Land and another putative refuge area to the north of the South Shetlands and Elephant Island (J. Anderson pers. comm.).

References

Anderson, J. B., S.S. Shipp, A.L. Lowe, J.S. Weller, and A.B. Mosola. 2002. The Antarctic ice sheet during the last glacial maximum and its subsequent retreat history: a review. Quaternary Science Reviews **21**: 49-70.

Arntz, W. E., T. Brey, and V.A. Gallardo. 1994. Antarctic zoobenthos. Oceanography and Marine Biology: An Annual Review **32**: 241-304.

Avise, J. C. 2000. Phylogeography. The history and formation of species. Harvard University Press.

Ballard, J. W., and M.C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology **13**: 729-744.

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

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

Bargelloni, L., S. Marcato, L. Zane, and T. Patarnello. 2000. Mitochondrial phylogeny of notothenioids: A molecular approach to antarctic fish evolution and biogeography. Systematic Biology **49**: 114-129.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 1996. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations.

Bernatchez, L., A. Chouinard, and G.Q. Lu. 1999. Integrating molecular genetics and ecology in studies of adaptive radiation: whitefish, *Coregonus sp.* as a case study. Biological Journal of the Linnean Society **68**: 173-194.

Beszteri, B., S. Frickenhaus, S. Gaebler, C. Held, L. Kraemer, F. Leese, and C. Mayer. In prep. STAMP: Extensions for the Staden sequence analysis package for automating microsatellite marker design.

Birky, C. W., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts and some results. Genetics **103**: 513-527.

Bohonak, A. J. 2002. IBD (isolation by distance): a program for analyses of isolation by distance). Journal of Heredity **93**: 153-154.

Brandt, A. 1988. Antarctic Serolidae and Cirolanidae (Crustacea: Isopoda): New genera, new species and redescriptions, p. 143. *In* R. Fricke [ed.], Theses Zoologicae. Koeltz Scientific Books.

Brandt, A. 1991. Colonization of the Antarctic shelf by the Isopoda (Crustacea, Malacostraca). Berichte zur Polarforschung **98**: 1-240.

Brandt, A., A.J. Gooday, S.N. Brandao, S. Brix, W. Brokeland, T. Cedhagen, M. Choudhury, N. Cornelius, B. Danis, I. De Mesel, R.J. Diaz, D.C. Gillan, B. Ebbe, J.A. Howe, D. Janussen, S. Kaiser, K. Linse, M. Malyutina, J. Pawlowski, M. Raupach, and A. Vanreusel. 2007. First insights into the biodiversity and biogeography of the Southern Ocean deep sea. Nature **447**: 307-311.

Brandt, A., K. Linse, and U. Mühlenhardt-Siegel. 1999. Biogeography of Crustacea and Mollusca of the Subantarctic and Antarctic regions. Scientia Marina **63**: 383-389.

Brey, T., C. Dahm, M. Gorny, M. Klages, M. Stiller, and W.E. Arntz. 1996. Do Antarctic benthic invertebrates show an extended level of eurybathy? Antarctic Science **8**: 3-6.

Castelloe, J., and A.R. Templeton. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. Molecular Phylogenetics and Evolution **3**: 102-113.

Clarke, A., and A.J. Crame. 1989. The origin of the Southern Ocean marine fauna, p. 253-268. *In* A.J. Crame [ed.], Origin and evolution of the Antarctic biota. The Geological Society.

Clarke, A., and J.A. Crame. 1992. The Southern Ocean benthic fauna and climate change: a historical perspective. Philosophical Transactions of the Royal Society of London B Biological Sciences **338**: 299-309.

Clarke, A., and N.M. Johnston. 2003. Antarctic marine benthic diversity. Oceanography and Marine Biology: an Annual Review **41**: 47-114.

Clarke, A., R.B. Aronson, J.A. Crame, J.M. Gil, and D.B. Blake. 2004. Evolution and diversity of the benthic fauna of the Southern Ocean continental shelf. Antarctic Science **16**: 559-568.

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

Cockerham, C. C., and B.S. Weir. 1973. Descent measures for two loci with some applications. Theoretical Population Biology **4**: 300-330.

Coombs, J. A., B.H. Letcher, and K.H. Nislow. 2007. CREATE: a software to create input files from diploid genotypic data for 52 genetic software programs. Molecular Ecology Notes online early.

Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144**: 2001-2014.

Crame, J. A. 1997. An evolutionary framework for the polar regions. Journal of Biogeography **24**: 1-9.

Crandall, K. A., A.R. Templeton, and C.F. Sing. 1994. Intraspecific phylogenetics: Problems and solutions, p. 273-297. *In* R.W. Scotland, D.J. Sieber, and D.M. Williams [eds.], Models in Phylogeny Reconstruction. Clarendon Press.

Dahlgren, T. G., J.R. Weinberg, and K.M. Halanych. 2000. Phylogeography of the ocean quahog (*Arctica islandica*): influences of paleoclimate on genetic diversity and species range. Marine Biology **137**: 487-495.

Dayton, P. K. 1990. Polar benthos, p. 631-685. *In* W.O. Smith [eds.], Oceanography Part B: Chemistry, biology and geology. Academic Press.

De Broyer, C., and K. Jazdzewski. 1993. Biodiversity of the Southern Ocean: Towards a new synthesis for the Amphipoda (Crustacea). Bollettino del Museo Civico di Storia Naturale di Verona **20**: 547-568.

Denton, G. H., M.L. Prentice, and L.H. Burckle. 1991. Cainozoic history of the Antarctic ice sheet, p. 365-433. *In* R.J. Tingey [eds.], Geology of Antarctica. Oxford University Press.

Di Rienzo, A., A.C. Peterson, J.C. Garza, A.M. Valdes, M. Slatkin, and N.B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences of the USA **91**: 3166-3170.

Dynesius, M., and R. Jansson. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. Proceedings of the National Academy of Sciences of the USA **97**: 9115-9120.

Eastman, J. T. 1991. Evolution and diversification of Antarctic notothenioid fishes. American Zoologist **31**: 93-109.

Elverhoi, A. 1981. Evidence for a late Wisconsin glaciation of the Weddell Sea. Nature **293**: 641-642.

Evans, J., J.A. Dowdeswell, and C. O Cofaigh. 2004. Late Quaternary submarine bedforms and ice-sheet flow in Gerlache Strait and on the adjacent continental shelf, Antarctic Peninsula. Journal of Quaternary Science **19**: 397-407.

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

Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**: 479-491.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**: 1567-1587.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology **3**: 294-299.

Fu, Y. -. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**: 915-923.

Gaffney, P. M. 2000. Molecular tools for understanding population structure in Antarctic species. Antarctic Science **12**: 288-296.

Gingele, F., G. Kuhn, B. Maus, M. Melles, and T. Schöne. 1997. Holocene ice retreat from the Lazarev Sea shelf, East Antarctica. Continental Shelf Research **17**: 137-163.

Glaubitz, J. C. 2004. CONVERT: A user-friendly program to reformat diploid genotypc data for commonly used population genetic software packages. Molecular Ecology Notes **4**: 309-310.

Glenn, T. C., and N.A. Schable. 2005. Isolating microsatellite DNA loci. Methods in Enzymology **395**: 202-222.

Gutt, J., B.I. Sirenko, I.S. Smirnov, and W.E. Arntz. 2004. How many macrozoobenthic species might inhabit the Antarctic shelf? Antarctic Science **16**: 11-16.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41**: 95.

Hedrick, P. 1999. Highly variable loci and their interpretation in evolution and conservation. Evolution **53**: 313-318.

Hedrick, P. 2005. A standardized genetic differentiation measure. Evolution **59**: 1633-1638.

Held, C. 2001. No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). Polar Biology **24**: 497-501.

Held, C. 2003. Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda), p. 135-139. *In* A.H. Huiskes, W.W. Gieskes, J. Rozema, R.M. Schorno, S.M. van der Vies, and W.J. Wolff [eds.], Antarctic biology in a global context. Backhuys Publishers.

Held, C., and F. Leese. 2007. The utility of fast evolving molecular markers for studying speciation in the Antarctic benthos. Polar Biology **30**: 513-521.

Heroy, D. C., and J.B. Anderson. 2005. Ice-sheet extent of the Antarctic Peninsula region during the Last Glacial Maximum (LGM)—Insights from glacial geomorphology. GSA Bulletin **117**: 1497-1512.

Hewitt, G. M. 2000. The genetic legacy of the quaternary ice ages. Nature **405**: 907-913.

Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. Genetical Research **38**: 209-216.

Huybrechts, P. 2002. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. Quaternary Science Reviews **21**: 203-231.

Imbrie, J., A. Berger, E.A. Boyle, S.C. Clemens, A. Duffy, W.R. Howard, G. Kukla, J. Kutzbach, D.G. Martinson, A. McIntyre, A.C. Mix, B. Molfino, J.J. Morley, L.C. Peterson, N.G. Pisias, W.L. Prell, M.E. Raymo, N.J. Shackleton, and J.R. Toggweiler. 1993. On the structure and origin of major glaciation cycles. 2. The 100,000-year cycle. Paleoceanography **8**: 699-735.

Ingolfsson, O. 2004. Quaternary glacial and climate history of Antarctica, p. 3-43. *In* J. Ehlers, and P.L. Gibbard [eds.], Quaternary glaciations - extent and chronology, Part III. Elsevier.

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: a cluster matching and permutationprogram for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23**: 1801-1806.

Janko, K., G. Lecointre, A. Devries, A. Couloux, C. Cruaud, and C. Marshall. 2007. Did glacial advances during the Pleistocene influence differently the demographic histories of benthic and pelagic Antarctic shelf fishes?--Inferences from intraspecific mitochondrial and nuclear DNA sequence diversity. BMC Evolutionary Biology **7**: 220.

Kimura, M., and J.F. Crow. 1964. The number of alleles that can be maintained in a finite population. Genetics **49**: 725-738.

Knowlton, N., and L.A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. Proceedings of the Royal Society of London, Series B: Biological Sciences **265**: 2257-2263.

Knowlton, N., L.A. Weigt, L.A. Solórzano, D.K. Mills, and E. Bermingham. 1993. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. Science **260**: 1629-1632.

Knox, G. A. 1994. The biology of the Southern Ocean. Cambridge University Press.

Landry, P. A., M.T. Koskinen, and C.R. Primmer. 2002. Deriving evolutionary relationships among populations using microsatellites and (deltamu)(2): all loci are equal, but some are more equal than others. Genetics **161**: 1339-1347.

Lecointre, G., C. Bonillo, C. Ozouf-Costaz, and J.-.C. Hureau. 1997. Molecular evidence for the origins of Antarctic fishes: Paraphyly of the Bovichtidae and no indication for the monophyly of the Notothenioidei (Teleostei). Polar Biology **18**: 193-208.

Leese, F., and C. Held. 2008. Identification and characterization of microsatellites from the Antarctic isopod *Ceratoserolis trilobitoides*: nuclear evidence for cryptic species. Conservation Genetics *online early*.

Leese, F., C. Mayer, and C. Held. Enrichment of microsatellites from unknwn genomes using known genomes as enrichment templates. Limnology and Oceanography: Methods in press.

Levinson, G., and G.A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Molecular Biology and Evolution **4**: 203-221.

Liebers, D., P. de Knijff, and A.J. Helbig. 2004. The herring gull complex is not a ring species. Proceedings of the Royal Society of London B Biological Sciences **271**: 893-901.

Linse, K. 2006. Biiodiversity and biogeography of Antarctic and sub-Antarctic Mollusca. Deep-Sea Research Part II **53**: 985-1008.

Linse, K., T. Cope, A.-.N. Lörz, and C. Sands. 2007. Is the Scotia Sea a centre of Antarctic marine diversification? Some evidence of cryptic speciation in the circum-Antarctic bivalve *Lissarca notorcadensis* (Arcoidea: Philobryidae). Polar Biology **30**: 1059-1068.

Luikart, G., F.W. Allendorf, J.M. Cornuet, and W.B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity **89**: 238-247.

Luxmoore, R. A. 1982. Moulting and growth in serolid isopods. Journal of Experimental Marine Biology and Ecology **56**: 63-85.

Maddison, W. P., and L.L. Knowles. 2006. Inferring phylogene despite incomplete lineage sorting. Systematic Biology **55**: 21-30.

Manoukis, N. C. 2007. FORMATOMATIC: a program for converting diploid allelic data between common formats for population genetic analysis. Molecular Ecology Notes **7**: 592.

McKelvey, K. S., and M.S. Schwartz. 2005. DROPOUT: A program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. Molecular Ecology Notes **5**: 716-718.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution **60**: 2399-2402.

Meredith, M. P., and J.C. King. 2005. Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. Geophysical Research Letters **32**: L19604.

Nauta, M. J., and F.J. Weissing. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. Genetics **143**: 1021-1032.

Near, T. J., J.J. Pesavento, and C.H.C. Cheng. 2003. Mitochondrial DNA, morphology, and the phylogenetic relationships of Antarctic icefishes (Notothenioidei: Channichthyidae). Molecular Phylogenetics and Evolution **28**: 87-98.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution **29**: 1-10.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press.

Neigel, J. E. 1997. Population genetics and demography of marine species p. 274-292. *In* R.F. Ormond, J.D. Gage, and M.V. Angel [eds.], Marine biodiversity: Patterns and processes. Cambridge University Press.

Nolte, A. W., K.C. Stemshorn, and D. Tautz. 2005. Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Molecular Ecology Notes **5**: 628-636.

Otha, T., and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetical Research **22**.

Page, T. J., and K. Linse. 2002. More evidence of speciation and dispersal across the Antarctic Polar Front through molecular systematics of Southern Ocean *Limatula* (Bivalvia: Limidae). Polar Biology **25**: 818-826.

Palma, A. T., E. Poulin, M.G. Silva, R.B. San Martin, C.A. Munoz, and A.D. Diaz. 2007. Antarctic shallow suhtidal echinoderms: is the ecological success of broadcasters related to ice disturbance? Polar Biology **30**: 343-350.

Pearse, D. E., and K. Crandall. 2004. Beyond Fst: Analysis of population genetic data for conservation. Conservation Genetics **5**: 585-602.

Park, S.D,E. 2001 Trypanotolerance in west African cattle and the population genetic effects of selection. Ph.D. Thesis University of Dublin.

Piry, S., G. Luikart, and J.-.M. Cornuet. 1999. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. Journal of Heredity **90**: 502-503.

Posada, D., K.A. Crandall, and A.R. Templeton. 2000. GeoDis: A program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Molecular Ecology Notes **9**: 489-488.

Posada, D., K.A. Crandall, and A.R. Templeton. 2006. Nested clade analysis statistics. Molecular Ecology Notes **6**: 590-593.

Poulin, E., A.T. Palma, and J.P. Feral. 2002. Evolutionary versus ecological success in Antarctic benthic invertebrates. Trends in Ecology and Evolution **17**: 218-222.

Poulin, E., and J.P. Feral. 1996. Why are there so many species of brooding Antarctic echinoids? Evolution **50**: 820-830.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**: 945-959.

Ramos-Onsins, S. E., and J. Rozas. 2002. Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution **19**: 2092-2100.

Raupach, M. J., and J.W. Wägele. 2006. Distinguishing cryptic species in Antarctic Asellota (Crustacea : Isopoda) - a preliminary study of mitochondrial DNA in *Acanthaspidia drygalskii*. Antarctic Science **18**: 191-198.

Rogers, A. R., and H.C. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution **9**: 552-569.

Rogers, A. D. 2007. Evolution and biodiversity of Antarctic organisms: a molecular perspective. Philosophical Transactions of the Royal Academy of London B Biological Sciences **362**: 2191-2214.

Rosenberg, N. A. 2004. Distruct: a program for the graphical display of population structure. Molecular Ecology Notes **4**: 137-138.

Rousset, F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics **142**: 1357-1362.

Rousset, F. 2007. GENEPOP '007: a complete reimplementation of the GENEPOP software for Windows and Linux. Molecular Ecology Notes **online first**.

Schneider, S., and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics **152**: 1079-1089.

Schubart, C. D., R. Diesel, and S.B. Hedges. 1998. Rapid evolution to terrestrial life in Jamaican crabs. Nature **393**: 363-365.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics **139**: 457-462.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585-595.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24**: 1596-1599.

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Molecular Ecology **7**: 381-397.

Templeton, A. R., E. Routman, and C.A. Philips. 1995. Separating population structure from population history: a cladistic analysis of geographical distribution of mitochondrial DNA haplotypes of the tiger salamander, *Ambystoma tigrinum*. Genetics **140**: 767-782.

Thatje, S., C.-.D. Hillenbrand, A. Mackensen, and R. Larter. 2008. Life hung by a thread: Endurance of Antarctic fauna in glacial periods. Ecology **89**: 682-692.

Thatje, S., C.D. Hillenbrand, and R. Larter. 2005. On the origin of Antarctic marine benthic community structure. Trends in Ecology and Evolution **20**: 534-540.

Valdes, A. M., M. Slatkin, and N.B. Freimer. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics **133**: 737-749.

van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**: 535-538.

Vaughan, D. G., G.J. Marshall, W.M. Connolley, C. Parkinson, R. Mulvaney, D.A. Hodgson, J.C. King, C.J. Pudsey, and J. Turner. 2003. Recent rapid regional climate warming on the Antarctic Peninsula. Climatic Change **60**: 243-274.

Villesen, P. 2007. FaBox: an online toolbox for fasta sequences. Molecular Ecology Notes **7**: 965-968.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics **7**: 167-184.

Waples, R. S., and C. Do. 2007. LDNe: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Notes **online first**.

Wägele, J. 1986. Polymorphism and distribution of *Ceratoserolis trilobitoides* (Eights, 1833) (Crustacea, Isopoda) in the Weddell Sea and synonymy with *C. cornuta* (Studer, 1879). Polar Biology **6**: 127-137.

Wägele, J.-W. 1987. On the reproductive biology of *Ceratoserolis trilobitoides* (Crustacea: Isopoda): Latitudinal variation of fecundity and embryonic development. Polar Biology **7**: 11-24.

Wägele, J.-W. 1994. Notes on Antarctic and South American Serolidae (Crustacea, Isopoda) with remarks on the phylogenetic biogeography and a description of new genera. Zoologische Jahrbücher Abteilung für Systematik Ökologie und Geographie der Tiere **121**: 3-69.

Weber, J. L., and C. Wong. 1993. Mutation of human short tandem repeats. Human Molecular Genetics **2**: 1123-1128.

Weir, B. S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**: 1358-1370.

Wilgenbusch, J. C., D.L. Warren, and D.L. Swofford. 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. http://ceb.csit.fsu.edu/awty .

Wilson, N. G., R.L. Hunter, S.J. Lockhart, and K.M. Halanych. 2007. Multiple lineages and absence of panmixia in the "circumpolar" crinoid *Promachocrinus kerguelensis* from the Atlantic sector of Antarctica. Marine Biology **152**: 895-904.

Wright, S. 1943. Isolation by distance. Genetics **28** 114-138.

Zane, L., L. Bargelloni, and T. Patarnello. 2002. Strategies for microsatellite isolation: a review. Molecular Ecology **11**: 1-16.

Chapter 8

An exception to the rule? Long-distance dispersal of the brooding benthic isopod *Septemserolis septemcarinata* from remote Southern Ocean islands

Florian Leese, Shobhit Agrawal, Christoph Held

Manuscript, submission to Journal of Biogeography

Abstract

Background: The Antarctic benthos differs from all known marine benthic communities in that a great number of species brood and are highly immobile during all life stages. Recent molecular genetic studies revealed an unexpectedly high number of cryptic species on the High Antarctic shelf and also the Antarctic islands, supporting that limited capability of long-distance dispersal is the rule rather than the exception. In this context, zoogeographic reports of the brooding shallow water isopod *Septemserolis septemcarinata* from most remote Antarctic and Subantarctic Islands must be regarded as doubtful, since shallow water habitats are separated by several hundreds to several thousands of kilometres of uninhabitable deep sea. This study applied one mitochondrial DNA marker and seven nuclear microsatellite markers to analyse the structure of *S. septemcarinata* populations from the highly remote Antarctic and Subantarctic islands South Georgia, Bouvet and Marion Island.

Results: Summary statistics and Bayesian inferences of population structure reveal that genetic variation is strongly partitioned geographically with very distinct gene pools for each of the three islands. However, haplotype and allele sharing patterns and the lack of a typical bimodal distribution of intraspecific variation do not support that populations from the different islands belong to different, yet unrecognized cryptic species. Unexpectedly, genetic diversity is increased in the populations from the two very young and small volcanic islands Bouvet and Marion to the east compared to the population around rather old and large South Georgia in the west. In addition, the population from South Georgia also reveals signs of a recent reduction in effective population size, while in particular the population from Bouvet resembles a large and stable population.

Conclusions: The maintenance of a partly shared gene pool relies on effective gene flow and consequently dispersal of specimens between the highly remote islands. Accordingly, our results imply that *S. septemcarinata* is much more mobile than expected from its strict benthic life cycle, its restriction to shallow water habitats and its reproduction mode (brooding). Based on own observations we suppose passive eastward drift on large floating macro algae mats or other organic substrata with the Antarctic Circumpolar Current (ACC) as a plausible mechanism for long-distance dispersal. The asymmetry of the ACC provides one explanation for inflated genetic diversity in the small islands Bouvet and Marion further to the east: We conclude that Bouvet and Marion Islands receive allelic diversity by passive drift with the ACC of specimens from source populations do not reach South Georgia. In addition, less severe effects of ice disturbance and interspecific niche-competition further explain the patterns observed.

Background

The Southern Ocean is different from most other marine environments on this planet in that it is geographically and thermally isolated due to the strong eastward flowing Antarctic Circumpolar Current (ACC). The permanently very cold waters (-1.86°C to 2.0°C) are characterized by strong vertical mixing and extreme seasonality (Dayton 1990, Knox 1994). Recurrent advancing sea and shelf ice greatly affect pelagic and benthic communities on ecological and evolutionary time scales (Gutt et al. 2001, Poulin et al. 2002, Thatje et al. 2005, Barnes and Convey 2007). These extreme abiotic conditions have led to striking adaptations by the organism inhabiting this ecosystem, e.g. the development of antifreeze proteins and the loss of red blood cells in icefish (Di Prisco 1997, 1998, 2000, Battaglia et al. 1997). They have also structured the benthic communities on the whole: there are many more brooding species present in the Antarctic benthos than on other benthic communities (Dell 1972, Clarke 1979, Picken 1980, Arntz et al. 1994, 1997, Poulin and Feral 1996, Poulin et al. 2002, Thatje et al. 2005). Even though, species with pelagic larvae can become very abundant (Pearse et al. 1991, Poulin et al. 2002) in shallow water habitats, highly disturbed by iceberg scouring (Gutt et al. 2001, Barnes and Conlan 2007). This is, however, restricted to few species and mainly shallow water habitats, predominantly due to the fact that they are better colonizers on an ecological time scale (Poulin et al. 2002).

The high incidence of brooding species in the Antarctic benthos leads to the expectation that genetic variation within brooding species is highly structured geographically, in particular in species with limited mobility in the adult stage (Palumbi 1994, Bohonak 1999). This has indirectly been supported by the recent findings that many species that were previously believed to have a broad, even circumpolar distribution in the Antarctic in fact consists of a series of cryptic species (Allcock et al. 1997, Page and Linse 2002, Held 2003, Held and Wägele 2005, Raupach and Wägele 2006, Linse et al. 2007). The serolid isopods are a particularly well-investigated taxon in this regard: they are strictly benthic, live often half-buried in the sediment and thus have only a limited mobility (Luxmoore 1982, 1984 Wägele 1987, Brandt 1991, Held 2000). Several cryptic species with mainly allopatric distribution and only small geographical overlap have already been discovered (Held 2003, Held and Wägele 2005, Leese and Held 2008) but these must be considered to represent just the first examples of a much greater, yet unidentified species diversity (Gutt et al. 2004, Held in prep.).

In this paper we investigate the genetic structure of different populations of *Septemserolis septemcarinata* (Miers, 1875), a brooding benthic serolid isopod that is widely distributed in shallow waters around Antarctic and Subantarctic islands (Brandt 1991, Wägele 1994). We focus on populations from three remote islands: South Georgia, Bouvet (Bouvetøya) and Marion Island (Fig. 1). All three islands are separated by several thousand kilometres of deep sea, which we expect to be a strong barrier for species with poor dispersal capabilities (Bohonak 1999). Despite their unifying feature of being remote, the islands strongly differ in their age, their size and their location within the ACC. South Georgia is a rather large and old Antarctic island (approximately 3750 km^2 landmass and >100 Myr) and located in the north-eastern Scotia Sea (54°12'S, 36°36'W). Bouvet Island is a very small and young volcanic island (approximately 50 km^2, approximately 1 Myr, Linse (2006) and citations therein) on the Mid-Atlantic Ridge (54°24'S, 3°21'E) just south of the Antarctic Polar Front (PF in Fig. 1) Marion Island is a small and geologically very young island of the Prince Edwards archipelago (approximately 290 km^2 and 0.25 Myr, Branch et al. 1991) south of Africa (see Fig. 1).

This study focuses on microevolutionary key aspects of genetic structuring between populations (1) and on analyses of intrapopulation variability (2) and demography (3).

1. In view of the poor dispersal of serolid isopods and strong genetic structure of related species on the High Antarctic shelf (Held 2003, Leese and Held 2008, Leese and Held in prep.) we expect a very strong geographical structuring of genetic variability for *S. septemcarinata* between the different islands and even predict that the separate islands may be inhabited by cryptic species. We thus test the null-hypothesis
   H_0: Populations of S. septemcarinata from highly isolate Antarctic and Subantarctic islands belong to one species and maintain a shared gene pool.

2. The prevailing differences in geological age, isolation and size of the three islands are expected to have imposed differences in the magnitude of genetic diversity. In particular we expect that genetic diversity is highest in populations from the rather old and large island South Georgia, lower around the relatively small and young island Bouvet and lowest around the very young and small Marion Island due to increased effects of genetic drift and founder events. We test the null hypothesis
   H_0: Genetic diversity is independent of the islands size, age and geological history.

3. We expect that the old age and the great availability of shallow water habitats around South Georgia has allowed the continuous maintenance of large populations at mutation-drift equilibrium, thus without traces of recent bottlenecks. On the contrary, the very young and much more remote islands Bouvet and Marion are expected to house much smaller populations that are much more affected by genetic drift and thus reveal genomic traces of recent reductions in effective populations size. We test the null hypothesis
   H_0: Present population size and past population oscillations are independent of the respective island characteristics.

Methods

Taxon sampling

Specimens were collected by bottom trawling during the ICEFISH 2004 expedition from shallow waters around South Georgia and Bouvet Island. Specimens from Marion Island were sampled on the MIOS-IV expedition by Christoph Held using a 1.5 m Agassiz' trawl and a metal stone dredge (Tab. 1). Animals were immediately sorted on deck and preserved in pre-chilled 96% ethanol. Specimens were kept at 4°C for the first months to guarantee adequate tissue preservation and maximum yield for subsequent DNA extraction.

Microsatellite analyses were performed for n=56 specimens from Bouvet Island and n=34 specimens from South Georgia and n=5 from Marion Island. Mitochondrial diversity was assessed for a subset of n=21 specimens from Bouvet, n=33 specimens from South Georgia and all n=5 specimens from Marion Island (Tab. 1). See Fig. 1 for location of islands.

Table 1: Sampling site details and Cruise information for the populations of *Septemserolis septemcarinata*.

Island (age, size)	Lat/Lon	Depth [m]	Cruise	N_{MSAT}	N_{mtDNA}
South Georgia (>100 Myr, 1700 km^2)	54°01'S 36°49'W	<200 m	ICEFISH 2004	34	33
Bouvet Island (1-2 Myr, 50 km^2)	54°27'S 3°17'E	<200 m	ICEFISH 2004	56	21
Marion Island (0.25 Myr, 290 km^2)	46°53'S 37°44'S	<200 m	MIOS-IV	5	5

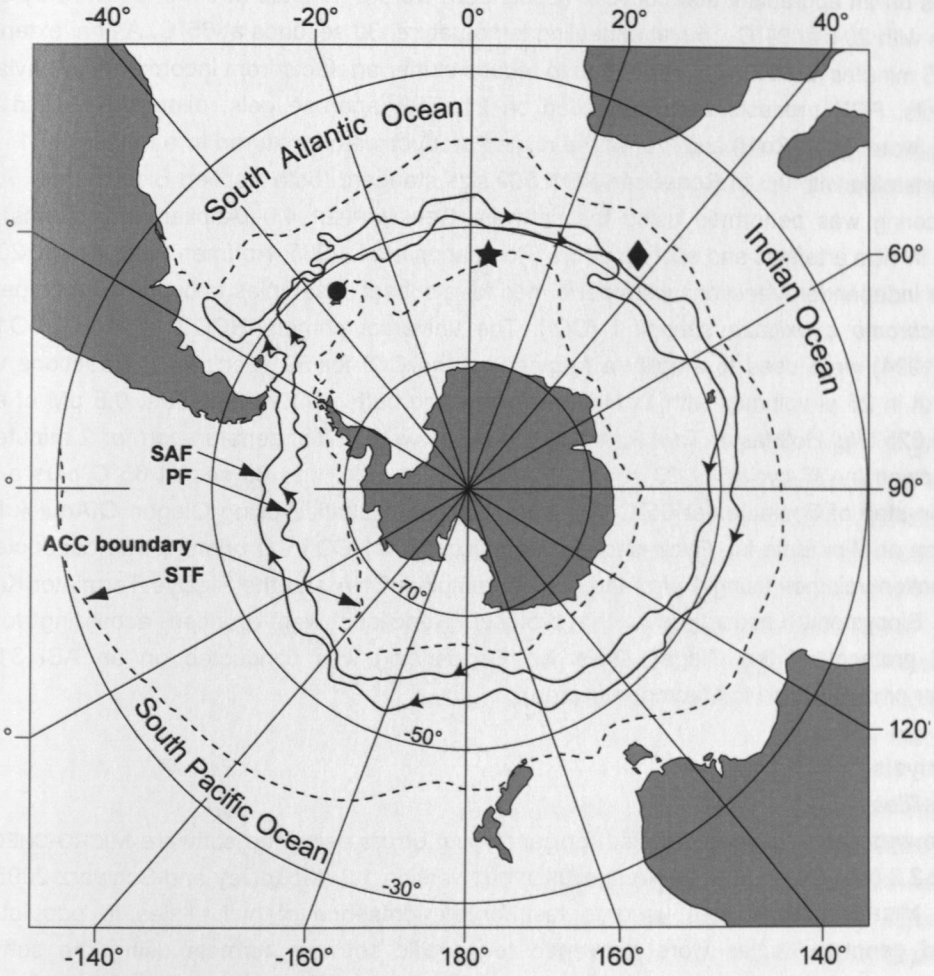


Figure 1: Sampling sites of *Septemserolis septemcarinata* in the Southern Ocean (South Georgia, circle; Bouvet Island, star) and the southern Indian Ocean (Marion Island, rhombus). Polar front systems according to Orsi et al. (1995) and Stewart (2007). STF=Subtropical Front, ACC=Antarctic Circumpolar Current, PF=Polar Front, SAF=Subantarctic Front. Arrowheads indicate the direction of the current.

DNA extraction, PCR, sequencing/genotyping

Total DNA from specimens from Bouvet and South Georgia was extracted from muscle tissue using the Qiagen DNeasy Mini Kit according to the standard tissue protocol using only 50 µl of elution buffer to increase concentration. DNA of specimens from Marion Island was extracted from one leg using the DNeasy Micro Kit (Qiagen).

Microsatellites: Eight polymorphic microsatellite loci for *S. septemcarinata* developed by Leese et al. (2008) were applied to assess intraspecific genetic polymorphisms for all specimens from Bouvet and South Georgia. Standard 15 µl reactions consisted of 1x PCR HotMaster Buffer, 0.2 mM dNTPs, 0.5-0.75 µM of each primer (one labeled, one unlabeled), 0.03 U/µl HotMaster *Taq* (Eppendorf, 5-prime), 0.5 M Betaine (Sigma Aldrich) and 5-20 ng of genomic DNA. Cycling conditions on an epgradient thermocycler (Eppendorf) were 2 minutes at 94°C followed by 30 to 34 cycles with 20 s at 94°C, 15 s at annealing temperature, 30 seconds at 65°C. A final extension step of 45 minutes at 65°C was performed to reduce stutter artefacts from incomplete adenylation of products. PCR products were controlled on 2% TBE agarose gels, diluted 1-15 fold with molgrade water (Carl Roth) and 1 µl of the diluted product was denatured in a mixture of 14.7 µl HI-DI formamide with 0.3 µl GeneScan ROX 500 size standard (both Applied Biosystems). Allele length scoring was performed using the software GENEMAPPER 4.0 (Applied Biosystems). To minimize in vitro artefacts and scoring errors (Pompanon et al. 2005, Hoffmann and Amos, 2005), up to four independent reactions were performed for a subset of samples to confirm genotypes.

Cytochrome c oxidase subunit I (COI): The universal primers HCO2198 and LCO1490 (Folmer 1994) were used to amplify a fragment of the COI for 59 specimens. Reactions were carried out in 25 µl volumes with 1x HotMaster reaction buffer, 0.2 mM dNTPs, 0.5 µM of each primer, 0.025 U/µl HotMaster *Taq*. Reaction conditions were: initial denaturation for 2 minutes at 94°C followed by 36 cycles of 20 sec at 94°, 15 sec at 46°C and 80 sec at 65°C plus a final elongation step of 5 minutes at 65°C. PCR products were purified using Qiagen QIAquick PCR purification or Mini Elute kit. Cycle-sequencing using 1 µM HCO/LCO primers was conducted in 10 µl reaction volumes using 1 µl of the purified template DNA and the BigDye Termiator Kit 3.1 (Applied Biosystems) according to the protocol. Reactions were purified according to the 'modified protocol' of the Qiagen DyEx Kit. Sequencing was conducted on an ABI 3130xl sequencer or outsourced to Macrogen (Korea).

Data analysis

Microsatellites

Raw data was checked and corrected for genotyping errors using the software MICRO-CHECKER version 2.2.3 (van Oosterhout 2004) and DROPOUT version 1.3 (McKelvey and Schwarz 2005). In addition, MICRO-CHECKER was used to test for the presence of null alleles in populations. Corrected genotype tables were converted to specific software formats using the software MSTOOLKIT version 3.1 (Park 2001), CONVERT version 1.3.1 (Glaubitz 2004) and FABOX version 1.3 (Villesen 2007). Bonferroni corrected significane was assessed using a permutation test with 10,000 permutations. Tests for Hardy Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were performed using GENEPOP version 4.0.6 (Rousset 2007).

Population differentiation was assed by calculating a coancestry coefficient similar to Weir and Cockerham's F_{ST} estimate θ (Weir and Cockerham 1984). Calculations were performed using the software GENEPOP version 4.0.6 (Rousset 2007). Significance of differentiation was assessed by an exact test (G-test according to Rousset 2007, parameters: 10,000 burnin, 50

batches with 10,000 MCMC steps each). In general, F_{ST} measures the relatedness of pairs of alleles within a subpopulation (S) relative to the total population (T), therefore, the estimators have the suffix ST. F_{ST} estimates (e.g. Weir and Cockerham 1984, Weir 2002) are based on the assumptions of the infinite allele model (IAM, Kimura and Crow 1964). This mutation model assumes that each mutation gives rise to a new, unique allele with no relation to the ancestral allelic state. The derived F_{ST} estimates the probability of identity in state among alleles within subpopulations compared to the probability of identity in the total populations. These estimates can handle multiple alleles and are in general calculated according to an Analysis of Variance (ANOVA) framework:

$$F_{ST} = \frac{Q_T - Q_S}{1 - Q_T} \qquad (1)$$

Where Q stands for the probability of identity among alleles within subpopulations (Q_S) or the total population (Q_T). GENEPOP calculates the estimate for F_{ST} for multiple loci according to the formula

$$F_{ST} = \frac{\sum_i \left[n_c s_P^2 \right]_i}{\sum_i \left[n_c s_P^2 + n_c s_I^2 + n_c s_G^2 \right]_i T} \qquad (2)$$

where n_c is a function of the sample size of the genotyped specimens for each locus to give more weight to more intensively genotyped loci. The index letter for the estimators of the variance components σ, G, I, and P, refer to there the variance components for Gametes, Individuals, and Populations. This estimation method by Rousset (2007) is regarded to provide a less biased estimate of F_{ST} than the traditional methods suggested by Weir and Cockerham (1984). F_{ST} estimates were standardized according to the method outlined by Hedrick (2005) in an ANOVA framework as outlined by Meirmans (2006) using the program RECODEDATA (Meirmans 2006) and subsequently GENEPOP to calculate the maximum value possible for F_{ST}. In the following, the F_{ST} estimate calculated was divided by F_{ST} max to yield the standardized value F'_{ST} that has not only a theoretical range of 0 to 1.

In addition to these summary statistics based estimates we applied a Bayesian multilocus-based clustering method implemented in the program STRUCTURE (Pritchard et al. 2000, Falush et al. 2003), of which version 2.2.2 was used to simultaneously infer population structure and assign individuals to populations. The advantage of the Bayesian clustering algorithm of STRUCTURE is that no classification of populations has to be done *a priori*. This algorithm of STRUCTURE assumes that there are *K* populations. Assuming Hardy-Weinberg equilibrium and no or only weak linkage disequilibrium within subpopulations, STRUCTURE assigns individuals in the data set probabilistically to populations, or jointly to two or more populations if genotypic data of individuals indicates that they are admixed. STRUCTURE calculates the likelihoods of the data (D) for the number for a given number of population (K_i), i.e. ln Pr(D|K) using a Markov Chain Monte Carlo algorithm. For the *S. septemcarinata* data set, the most likely number of populations was inferred without prior information on geographic origin of individuals using STRUCTURE. The number of MCMC steps needed to reach convergence was first estimated by comparing run

lengths between 10,000 and 2,000,000 steps. Convergence was reached in general with 10,000 steps. Therefore, for the parameter sets 5 independent runs with a burn-in of 10,000 and subsequent 50,000 MCMC steps were performed with and without assuming population admixture, with and without allele frequencies correlated The number of clusters (K) to infer was set from K=1 to K=4 to allow detecting potential cryptic subpopulations. Alpha was inferred from the data for each population separately in the admixture mode. All 5 independent runs were analysed in CLUMPP, version 1.1.1 (Jakobsson and Rosenberg 2007) in order to compute a consensus matrices from all 5 matrices. Both, the individual membership coefficient matrix and averaged population membership matrix (Q-matrices) were visualized using DISTRUCT, version 1.1 (Rosenberg, 2004).

To assess estimates of the present effective population size (N_e), we applied the linkage disequilibrium method proposed by Hill (1981), modified by Waples (2006) to account for bias correction. This method is implemented in the program LDNE, version 1.3 (Waples and Do 2007). Calculations of N_e and the confidence intervals were estimated considering alleles with a frequency of c ≥0.05 and c ≥0.02, respectively.

Tests for historical population bottlenecks were performed using the program BOTTLENECK (Piry et al. 1999). Tests are based on the finding that populations that have experienced recent reductions in their effective population size (N_e) show a reduction in both, allelic richness and heterozygosity (Nei et al. 1975, Maruyama and Fuerst 1985). Following a bottleneck, the number of alleles are faster reduced than heterozygosity (Nei et al. 1975). Vice versa, in expanding populations, the number of alleles increases faster than heterozygosity until equilibrium is reached. From the correlation of both parameters from microsatellite data, it is possible to make inferences on historical demography of a population. We tested for patterns of heterozygosity excess or deficiency using three different microsatellite mutation models: the Infinite Allele Model (IAM, Kimura and Crow 1964), the Stepwise Mutation Model (SMM, Otha and Kimura 1973) and the a model that combines features of both, the Two Phase Model (TPM, DiRienzo et al. 1994) with 90% of the mutations according to a SSM according to the recommendations of Piry et al. 1999. Statistical significance of the sign test and a Wilcoxon rank test were assessed by 1,000 permutations (Cornuet and Luikart 1996, Luikart et al. 1998).

Data analysis COI

Assembly of forward and reverse strands and editing was performed using the software SEQMAN (DNAstar, Lasergene). Sequence alignment was performed using the CLUSTALW program as implemented in BIOEDIT 7.09 (Hall 1999). Sequence variation was analysed using MEGA 4.0 (Tamura et al. 2007). A statistical parsimony network based on COI haplotype data was calculated using TCS 1.21 (Clement et al. 2000). Population differentiation was estimated by conventional F-statistics based on haplotype frequencies using ARLEQUIN 3.11 (Excoffier et al. 2005). Significance of differentation estimates was assessed by an exact test (Raymond and Rousset 1995) as implemented in ARLEQUIN. Parameter settings were: burnin 10,000, MCMC steps 100,000. Several demographic parameters were also inferred from the COI data using ARLEQUIN 3.11. First, we calculated Tajima's D (Tajima 1989) that is based on the magnitude of differences between the number of differences between the number of segregating sites (S) and the average pairwise nucleotide differences (π). Population expansions after bottleneck or founder events inflate S relative to π. Consequently, this statistic tends to be negative when there is an excess of recent mutations (or rare alleles). Fu's F_S statistic (Fu 1997) is based on the

probability of finding a number of alleles greater or equal to the observed number in a sample drawn from a stationary population and is even more sensitive in detecting population expansions from smaller samples (Ramos-Onsins 2002). Significance of Fu's F_S and Tajima's D are estimated using a coalescent simulation approach as implemented in ARLEQUIN 3.11 using 10,000 permutations. In the third test, pairwise mismatch distributions among individuals were plotted and tested for goodness-of-fit to a model of sudden expansion using parametric bootstrapping with 50,000 replicates (Schneider and Excoffier 1999).

Results

(Remark: Tables 2-9 and Figure 2 are at the end of the manuscript in the appendix)

Both markers were highly informative with respect to the scientific questions addressed in this study (Hypotheses 1-3, introduction).

Microsatellites

Microsatellites were highly polymorphic in all three both populations. Allele sizes and variability estimates for the eight microsatellites are given in Tab. 2. Frequency distributions are displayed in figure 2. The number of alleles per microsatellite locus for all specimens screened ranged from N_A=2 (Sse05) to N_A=25 (Sse10). Observed heterozygosity (H_o) ranged from 0.059 (Sse05, South Georgia population) to 0.911 (Sse10, Bouvet population). No significant deviations from HWE were found for the population from Bouvet and Marion Island (Tab. 2). A significant heterozygosity deficiency was reported for locus Sse14 for the population from South Georgia. Analyses with MICRO-CHECKER indicate that this may be best explained due to the presence of null alleles (non-amplifying alleles) at this locus, which artificially inflate homozygosity. When removing locus Sse14, global tests for deviations from HWE fail to detect a significant heterozygosity deficiency (Tab. 2). Global linkage disequilibrium (LD) was observed for locus pair Sse14 and Sse15 and remained significant ($P<0.01$) after sequential Bonferroni correction for multiple testing (Rice, 1989). As locus Sse14 also shows a significantly inflated homozygosity and could not be genotyped for the n=5 specimens from Marion island, this locus was excluded from population genetic parameter inference in this study. The microsatellite data provide unique resolution to test hypotheses 1-3:

Genetic structure

Genetic polymorphisms show a strong geographical association, with strong differences in allele types and allele frequencies between the different islands (Fig. 2). Population differentiation estimates based on Weir and Cockerham's F_{ST} estimate (*theta,* Weir and Cockerham 1984, as implemented according to Rousset 2007) reveal significant population differentiation between all pairs of population (Tab. 3). Non-standardized F_{ST} estimates are similarly high between population pairs South Georgia/Bouvet and South Georgia/Marion but lower for Bouvet/Marion, indicating that gene flow between Bouvet and Marion island is higher. Calculations of standardized F_{ST} (= F'_{ST}) according to Meirmans (2006), which present the degree of allele sharing on a scale of 0 to 1 (Wright 1978), show that non-standardized F_{ST} estimates tend to

underestimate differentiation. The magnitudes of pairwise F'_{ST} reveal that gene flow between populations is very limited but, however, present.

Results of the Bayesian cluster analyses of STRUCTURE report the highest probability of the data when using information on population origin as a prior: Ln Pr $(D|K)_{K=3}$ = -2322.5. All individuals are assigned correctly to their population and reveal only very minor genotypic admixture proportions. When searching for the K with the highest likelihood without the information on population origin of individuals, results depend on the model assumtions. Under an admixture model, STRUTURE reports the highest likelihood for K=3: Ln Pr $(D|K)_{K=3}$ = -2350.8 (standard deviation ±11.27). The three clusters inferred correspond to the three islands sampled (Fig. 3). However, in particular populations from Bouvet and Marion Island also contain specimens with shared genotype proportions that indicate genetic admixture. However, one individual from Bouvet shows a genotype very similar to the dominant genotype of South Georgia population, and also one individual with almost equal proportions of a typical Bouvet and a typical Marion Island genotype. If no admixture is assumed in the model, structure reports the highest probability for K=2. Ln Pr $(D|K)_{K=2}$ = -2369.38 (standard deviation ±1.62). The individuals from South Georgia had high proportions for cluster 1 and individuals from Bouvet and Marion for cluster two. Interestingly, two of the n=5 specimens from Marion Island head high admixture proportions q_n of cluster 1 (South Georgia): q_1=0.32 and 0.45, respectively. No individual with a higher admixture proportion than 0.1 was found in the Bouvet Island population.

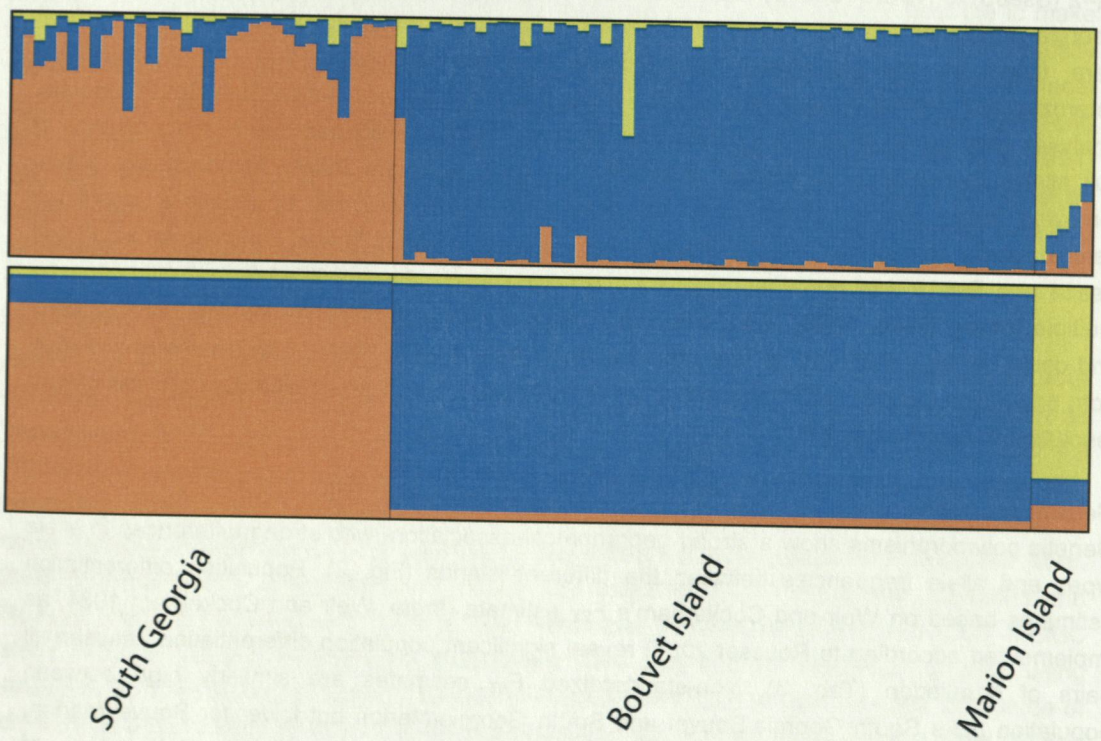


Figure 3: Results of cluster analyses performed with STRUCTURE with the highest log likelihood without prior information on population origin. The graphs display the consensus membership coefficients matrices (Q-matrices) for 95 individuals from three populations of *Septemserolis septemcarinata* (South Georgia = orange, Bouvet = blue, Marion Island = yellow) using seven microsatellite markers. The genotype of each individual is represented by a single bar, where the proportion of the colour refers to the probability of assignment to one of the three populations.

Genetic diversity

Intrapopulation gene diversity in terms of expected heterozygosity (H_E, Nei 1987) was higher for populations from the small and young islands Bouvet and Marion (H_E=0.712 and 0.717, respectively) compared to the population from South Georgia (0.630). As this measure is not sensitive to sample size and only the loci without evidence for null alleles were included, this must be regarded a distinguishing feature of the three different populations (South Georgia vs. Marion and Bouvet).

Demography

Estimates of N_e revealed the presence of large populations around South Georgia and Bouvet (Tab. 4). For Marion Island, sample size was too small to obtain reliable results. The upper confidence interval was infinite for both inference methods. However, mean values for N_e were consistently larger for Bouvet than for South Georgia (Tab. 4). Results of BOTTLENECK provide no consistent evidence for a significant heterozygosity excess or deficiency for the population from Bouvet or Marion Island (Tab. 5). However, there was a significant heterozygosity deficiency reported from both, the Sign and the Wilcoxon test for the South Georgia populations when assuming an underlying TPM and a SMM ($P<0.05$, Tab. 5).

Mitochondrial DNA

Base pair frequencies of the 551 bp fragment of the partial COI gene, were A=22.9%, C=18.9%, G=21.9%, and T=36.3%, respectively. Eight positions were variable, four being parsimony-informative and four singleton sites. All substitutions were silent mutations. The n=33 specimens from South Georgia contribute to one single haplotype only (HT1). Haplotype 2 (HT2) is shared among specimens from Bouvet (n=7) and Marion Island (n=4). Haplotype 3 (HT3) is only reported for specimens from Bouvet. Haplotypes 4-6 (HT4-6) are reported from single specimens from Bouvet, and haplotype 7 (HT7) from one specimen from Marion Island (Tab. 6). Haplotype sequences were deposited to GenBank [EU597351-EU597357].

Genetic structure

The average pairwise distances between groups were low (Tab. 7). The maximum uncorrected pairwise genetic distance was between HT1 and HT6 and comprises 1.1%. This is within a range typically observed for other species of isopods (Held 2003, Held and Wägele 2005, Raupach and Wägele 2006, Leese et al. submitted). Haplotypes from specimens sampled around Bouvet and Marion Island reveal partly overlapping haplotypes. No bimodal distribution of genetic distances was observed, i.e. no barcoding gap (Hebert et al. 2003) was detected. As a consequence based on the COI data of the different populations we must regarded populations to belong to one species rather than to several cryptic species.

Genetic diversity

Nucleotide diversity and pairwise differences (π) were (0 / 0) for the populations of South Georgia (no variation), comparatively high for Marion Island (0.0007 / 0.400) and highest for the population from the smallest island Bouvet (0.0023 / 1.248). This is unexpected in terms of the expectations due to age and size of the islands.

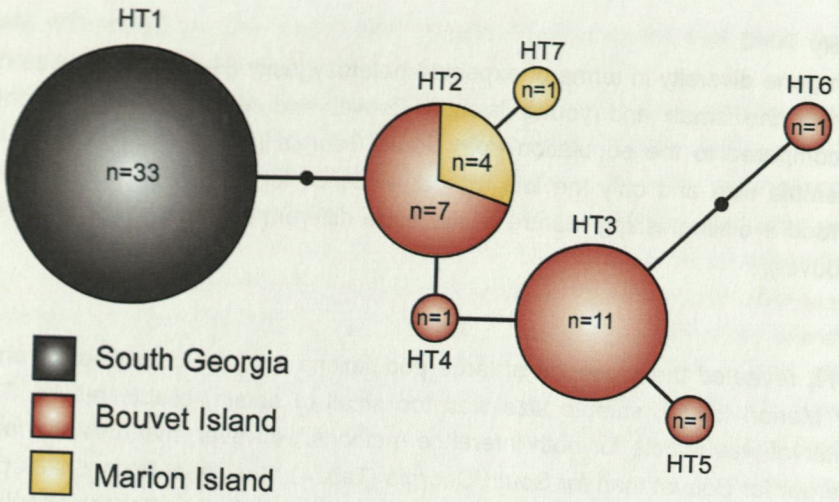


Figure 4: Statistical parsimony network of n=59 cytochrome oxidase I sequences in seven haplotypes (HT1-HT7) from *Septemserolis septemcarinata* (TCS, 95% parsimony connection limit). The diameter of circles corresponds to the number of specimens that share a particular haplotype. Black dots represent hypothetical, unsampled haplotypes.

Present and historical demography

To test the hypothesis of recent population decline or growth from a low-diversity founder population, we only applied the three demographic tests to the population from Bouvet, as all specimens from South Georgia are genetically identical for the COI, thus representing a typical low-diversity founder population, and from Marion Island, only 5 samples were available.

Neither Tajima's *D*, nor Fu's F_S support the hypothesis of an expanding population for populations from Bouvet and Marion Island (Tab. 9). The mismatch distribution for the data observed differs strongly from the distribution of a simulated population under a sudden expansion model (Fig. 5). However, the null hypothesis of population expansion could not be rejected based on the mismatch distribution, although the probability is nearly significant (P=0.073).

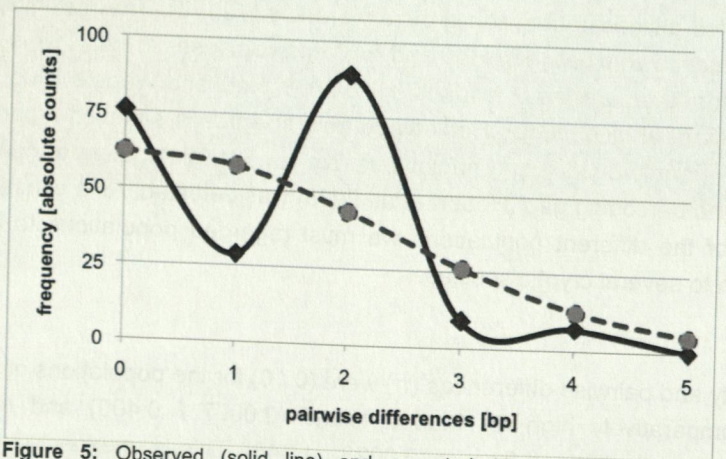


Figure 5: Observed (solid line) and expected (dotted line) mismatch distribution under the assumption of a sudden population expansion based on the mitochondrial cytochrome oxidase I sequence data from n=21 specimens of *Septemserolis septemcarinata* from Bouvet Island.

Taken evidence from both marker systems together, all three null hypotheses (1: genetic homogeneity among populations, 2: no differences in genetic diversity, 3: no differences in the genomic signatures of bottlenecks) can be rejected.

Discussion

Mobility of the species' genome

Data on genetic differentiation based on eight microsatellites and the mitochondrial partial COI gene unambiguously indicate that genetic variation within *S. septemcarinata* is significantly partitioned geographically. The magnitudes of the subdivision prove that populations from the three different islands maintain distinct gene pools with limited gene flow between the populations. However, the partial overlap in microsatellite alleles and COI haplotypes (for Bouvet and Marion Island), the lack of a typical bimodal distribution of genetic variation, i.e. the absence of a 'barcoding gap' (Hebert et al. 2004), and finally the overall low magnitude of genetic differences in the COI gene, provide strong evidence that populations belong to one species that is capable of exchanging alleles over long distances, at low frequencies though. Additional support for these findings stem from geological evidence: phylogenetic analyses, based on molecular clock estimates that utilize the opening of the Drake Passage (Barker and Burrell 1977, Livermore et al. 2005) as a calibration point (Held 2001), calculate the age of the taxon *S. septemcarinata* to approximately 6 Myr. Thus, the taxon itself is older than the young volcanic islands Bouvet (1-2 Myr) and Marion Island (0.25 Myr). As the islands have never been in close contact to the Antarctic Islands or other continents, the present-day population must have arrived on the islands by migration via the sea late in its history.

The finding of long-distance dispersal (LDD) in *S. septemcarinata* is unexpected given the enormous isolation of the three island populations, in particular of Bouvet Island, and the recent finding that brooding isopods from the Antarctic shelf (Held 2003, Held and Wägele 2005, Leese and Held in prep.), the shelf of Patagonia (Leese et al. under review) and from temperate subtidal waters (Wares et al. 2007) often consist of complexes of cryptic species with generally much smaller distribution ranges. Studies on other brooding Antarctic animals, e.g. molluscs (Page and Linse 2002, Linse et al. 2007), echinoderms (Lee et al. 2004), octopods (Allcock et al. 1997) and even some Antarctic species with pelagic larvae (Wilson et al. 2007) revealed the presence of cryptic species and thus provide additional support that limited dispersal is the rule rather than the exception.

In conclusion it must be accepted that specimens of *S. septemcarinata* are much more mobile than one would predict based on the strictly benthic life cycle. In the absence of active means for dispersal, i.e., no free-swimming life stages, other means of genome transport across large distances must be present in this taxon. Highsmith (1985) proposed that passive floating on the water surface and rafting on algal mats may provide a means for brooding species to greatly increase their capability to disperse. Other than most serolid isopods, which normally live benthic and not on substrata that is likely to become detached, *S. septemcarinata* is an exception as it has enlarged pereiopods 5-7, which it uses to climb on e.g. large Antarctic rossellid sponges (pers. observ. CH and FL during ICEFISH 2004 expedition). While the possibility of passive drift on sponge fragments or even entire sponges with detached anchor ice (Dayton et al. 1969) may provide a means for dispersal, this is likely to be of minor importance for LDD (McClintock et al.

2005), in particular in the Subantarctic islands. Drift on large algal mats, in particular those of *Macrocystis* species, are very likely to provide a much more reasonable means for long-term dispersal (Helmuth et al. 1994, Thiel and Gutow 2004, Hobday 2000a,b). After detachment these algae can remain alive and even still grow in the water for several month (Hobday 2000b) and therefore provide a unique means for LDD over hundreds to several thousands of kilometres – unidirectional – with the ACC. With an average speed of approximately 40 cm s^{-1} in the ACC (Whitworth et al. 1982, Hofmann 1985), transportation over 1000 km would take about 30 days. However, net drift speed depends highly on other factors such as large-scale eddies which can easily decrease the net transportation speed. The linear distance between South Georgia and Bouvet is about 2500 km. The same holds true for the distance between Bouvet and Marion Island. Transportation with the ACC for 2500 km would at least last 70 days, which can certainly be realized by drifting macro algae.

Small, remote AND diverse?

When comparing gene diversity estimates (H_E) based on microsatellite data it must always be questioned whether the estimates are valid or suffer from artificial biases (Bonin et al. 2004, Pompanon et al. 2005). For the present study it must be assured in particular that inflated homozygosity around South Georgia is not the result of artificial biases of the analysis method or the marker characteristics. To avoid errors of the analysis method (amplification problems, scoring errors), microsatellite amplifications and fragment analysis were performed up to four times independently for a subset of specimens. Allele scoring was performed independently by SA and FL in order to maximize inter-rater reliability of the data. To test for potential biases related to the microsatellite markers used the allele tables were analysed for biases applying DROPOUT and MICRO-CHECKER in the data set. The potentially most severe bias might be due to non-amplifying alleles, so called null alleles, which can lead to artificial heterozygosity deficiencies (Callen et al. 1993, Paetkau and Strobeck 1995). In general, point mutations in the primer-binding site are the cause of null alleles. Allelic dropout can occur due to preferential amplification of a particular allele (e.g. shorter ones) and may also inflate homozygosity (Gagneux et al. 1997). However, as there was significant evidence for null alleles only for locus Sse14, which was excluded for parameter inference, and no evidence for allelic dropout, these artificial biases are expected to have only a minor impact on the data. In addition, the consistent depauperate genetic diversity as inferred from microsatellites and the COI supports that differences in genetic diversity are a consistent population feature and not artificially caused by a bias.

In the absence of selection three major microevolutionary forces shape genetic diversity at neutral loci in an idealized population: random genetic drift, migration and mutations. While the latter two increase the genetic diversity, random genetic drift, i.e. the random loss of alleles due to finite population size, decreases diversity. The effects of genetic drift are antiproportional to N_e (Nei et al. 1975, Maruyama and Fuerst 1985). As a consequence, the partitioning of intrapopulation genetic diversity among the islands is unexpected, while the geographical subdivision of genetic variation can for the most part be explained with the isolation of the islands with little gene flow in between. Based on the island characteristics (age and size), the largest populations and highest intrapopulation diversity were expected around South Georgia. Interestingly, our results indicate that it is the smallest island, Bouvet Island, that has the highest genetic diversity and vice versa the oldest and largest island, South Georgia, that has the lowest

genetic diversity. While the high genetic diversity around Bouvet and Marion Island can in principle be explained by migration of specimens from islands other than South Georgia, most likely from the South Sandwich Islands or the South Orkney Islands, historical demographic processes might also pose a major factor affecting genetic diversity of populations from the islands in different ways.

Population demography

The finding of a single mitochondrial haplotype in the populations from South Georgia only (nucleotide diversity = 0) and significantly lower heterozygosities than expected under mutation-drift equilibrium lead to the conclusion that the population around South Georgia has recently experienced a severe reduction in effective population size from which it is currently recovering. A recent (re)colonization by a small founder population is another possible scenario that would explain the genomic signatures inferred.

While at first glance it is counterintuitive that South Georgia has been affected by reduction in N_e more than both other, much smaller islands, two processes with high explanatory power must be taken into account: (1) at both, evolutionary and ecological time scales, major glaciations with grounded ice eradicating shallow water habitats had a much more severe impact on the shallow water shelf fauna around South Georgia than at Bouvet or Marion Island (Clapperton 1989, but see also Ronqvist et al. 1999), (2) since South Georgia is much older and was in close contact with the Antarctic Peninsula it houses a very rich and old benthic fauna (Winkler 1994 and references therein) including several larger serolid isopods, which occur in great abundances in the same habitat, e.g. *Serolella pagenstecheri*. Consequently, niche competition must be discussed as another possible mechanism limiting present population size around South Georgia. On both the other islands, Bouvet and Marion, *S. septemcarinata* is the only serolid isopod and might benefit from less niche competition.

Conclusions

Our results support that specimens of the brooding shallow water isopod *S. septemcarinata* from the three highly remote Southern Ocean islands, South Georgia, Bouvet and Marion Island, belong to one species. Genetic variation is significantly partitioned geographically indicating very little gene flow between islands. In the absence of pelagic dispersal stages of this brooding isopod, long-distance dispersal on floating macro algal mats or other substrata with the Antarctic Circumpolar Current seems most likely.

The population from the largest and oldest island South Georgia is genetically less diverse than populations from the young and small islands Bouvet and Marion and reveals genomic signatures of a recent bottleneck. This evidence supports that the population from South Georgia might be more affected by processes reducing effective population size (and therefore diversity). We argue that more severe glaciations and increased niche competition due to the presence of other serolid species around South Georgia compared to the smaller volcanic islands Bovuet and Marion are two key aspects that can explain the patterns observed. The great allelic diversity around the young and remote islands Bouvet and Marion strongly suggests that both populations are recipients of gene flow from other source populations to the west in the Antarctic Circumpolar Current, which do not disperse to South Georgia.

Acknowledgements: FL and CH are grateful to Wolfgang Wägele (Museum Koenig, Bonn, Germany) for helpful discussions and support. We thank Andrea Eschbach and Annegret Müller for technical assistance. This work was supported by a DFG grant HE-3391/3 to CH and NSF grant OPP-0132032 to H.W. Detrich III. This is publication number 23 from the ICEFISH Cruise of 2004.

Appendix

Table 2: Total number of specimens scored for each population and locus (N_S), number of different alleles (N_A), inbreeding coefficient (F_{IS}), observed heterozygosity (H_O), expected heterozygosity, and probability (unbiased estimate of type-I error) for Hardy-Weinberg departure proportions (P_{HW}). The mean H_O and H_E for all microsatellites and the data set and their respective overall P_{HW} are listed at the end of the table. Due to very limited amount of DNA, locus Sse14 was not amplified for population from Marion Island.

	South Georgia	Bouvet	Marion Island
Sse04			
N_S	34	56	5
N_A	10	13	2
F_{IS}	0.113	0.077	-0.333
H_o	0.588	0.750	0.600
H_e	0.662	0.812	0.467
P_{HW}	0.105	0.353	1.000
Sse05			
N_S	34	53	5
N_A	2	2	3
F_{IS}	-0.015	0.000	-0.067
H_o	0.059	0.245	0.400
H_e	0.058	0.245	0.378
P_{HW}	1.000	1.000	1.000
Sse07			
N_S	34	56	5
N_A	2	4	2
F_{IS}	-0.060	0.036	-0.600
H_o	0.471	0.429	0.800
H_e	0.444	0.444	0.533
P_{HW}	1.000	0.613	0.428
Sse08			
N_S	30	54	5
N_A	7	13	6
F_{IS}	-0.072	0.093	0.314
H_o	0.767	0.815	0.600
H_e	0.716	0.857	0.844
P_{HW}	0.114	0.990	0.162
Sse10			
N_S	34	56	5
N_A	15	18	7
F_{IS}	0.186	0.018	0.135
H_o	0.706	0.911	0.911
H_e	0.864	0.927	0.800
P_{HW}	**0.045**	0.553	0.491
Sse13			

N_S	34	55	5
N_A	13	11	7
F_{IS}	-0.046	-0.007	-0.081
H_o	0.882	0.836	1.000
H_e	0.844	0.831	0.933
P_{HW}	0.969	0.719	1.000
Sse14			
N_S	23	56	0
N_A	8	13	-
F_{IS}	0.734	0.064	-
H_o	0.192	0.804	-
H_e	0.713	0.858	-
P_{HW}	**0.000**	0.272	-
Sse15			
N_S	34	56	5
N_A	13	15	8
F_{IS}	0.1429	0.055	-0.056
H_o	0.706	0.821	1.000
H_e	0.822	0.869	0.956
P_{HW}	0.057	0.202	1.000
Total			
Mean H_E	0.640	0.730	-
Mean H_O	0.546	0.701	-
P_{HW}	**0.000**	0.064	-
Excl. Sse14			
Mean H_E	0.630	0.712	0.717
Mean H_O	0.597	0.687	0.742
P_{HW}	0.103	0.870	0.943

Table 3: Upper diagonal: Pairwise F_{ST} estimates and maximum possible F_{ST} estimates between island populations of *Septemserolis septemcarinata*. Lower diagonal: Standardized pairwise F'_{ST}. All pairwise differenatiation estimates are significant at P<0.001.

	South Georgia	Bouvet Island	Marion Island
South Georgia	-	0.139 (0.325)	0.141 (0.339)
Bouvet Island	0.427	-	0.088 (0.285)
Marion Island	0.420	0.311	-

Table 4: Estimates of effective population size using seven microsatellite loci and considering alleles with a minimum frequency of c≥0.05 and c≥0.02. The 95% confidence intervals (CI) based on parametric bootstrap and Jackknife testing are listed. *Sample size for the Marion Island population was too low (n=5) to perform reliable tests.

	South Georgia	Bouvet	Marion Island*
c ≥0.05			
r^2 (overall)	0.03710	0.02043	0.33331
N_e	117.7	302.9	-12.7
CI Bootstrap	29.1 - ∞	78.8 - ∞	8.7 - ∞
CI Jackknife	21.9 - ∞	84.7 - ∞	-27.5 - ∞
c ≥0.02			
r^2 (overall)	0.03528	0.01946	0.33331
N_e	190.4	2109.9	-12.7
CI Bootstrap	53.7 - ∞	173.0 - ∞	8.7 - ∞
CI Jackknife	42.9 - ∞	163.0 - ∞	-27.6 - ∞

Table 5: Statistical tests for significant heterozygosity (H) excess or deficiency in populations of *Septemserolis septemcarinata* from Bouvet and South Georgia assuming three different mutation models (IAM, TPM, SMM) using BOTTLENECK. P-values of the Sign Test and for one-tailed probability of heterozygosity deficit (P_D) or excess (P_E) of the Wilcoxon Test are based on 1000 independent permutations. Significant P-values are printed in bold. The TPM was adjusted to allow for 90% mutations according to a SMM and 10% to an IAM model.

		Sign Test				Wilcoxon Test
Population	**Model**	Expected no of loci with H excess	Observed no. of loci with H excess	Observed no of loci with H deficiency	P	P (one tailed for H deficiency)
South Georgia	IAM	3.83	4	3	0.604	P_D=0.594
	TPM	3.85	1	6	**0.034**	P_D=**0.027**
	SMM	3.88	1	6	**0.032**	P_D=**0.020**
Bouvet Island	IAM	4.03	7	0	**0.020**	P_D=1.000
	TPM	3.97	2	5	0.131	P_D=0.188
	SMM	3.91	2	5	0.141	P_D=**0.020**
Marion Island	IAM	3.46	5	1	0.199	P_E=**0.031**
	TPM	3.55	5	1	0.212	P_E=**0.039**
	SMM	3.23	5	1	0.147	P_E=0.055

Table 6: Frequency and geographic origin of the mitochondrial cytochrome c oxidase subunit I haplotypes for the populations of *Septemserolis septemcarinata* from South Georgia, Bouvet and Marion Island.

Haplotype	South Georgia	Bouvet Island	Marion Island	GenBank accession number
HT1	33	0	0	EU597351
HT2	0	7	4	EU597352
HT3	0	11	0	EU597353
HT4	0	1	0	EU597354
HT5	0	1	0	EU597355
HT6	0	1	0	EU597356
HT7	0	0	1	EU597357

Table 7: Average pairwise genetic distances between (lower diagonal) and within populations of *Septemserolis septemcarinata* (mean diagonal) for a 551 bp fragment of the mitochondrial cytochrome c oxidase subunit I gene.

	South Georgia (n=33)	Bouvet (n=21)	Marion Island (n=5)
South Georgia	*0.000*		
Bouvet	0.006	*0.002*	
Marion	0.003	0.004	*0.001*

Table 8: Pairwise F_{ST} estimates for populations of *Septemserolis septemcarinata* based on mitochondrial cytochrome c oxidase subunit I gene data (lower diagonal) and their respective P values (upper diagonal).

	South Georgia	Bouvet Island	Marion Island
South Georgia	-	**0.000**	**0.000**
Bouvet Island	0.854	-	0.078
Marion Island	0.980	0.400	-

Table 9: Cytochrome oxidase I nucleotide diversity, pairwise differences, goodness-of-fit test (raggedness index, Harpending 1992) of a 'sudden-expansion' population model and Tajima's and Fu's tests for demographic expansion of the populations of *Septemserolis septemcarinata* sampled around Bouvet (n=21 specimens). Samples from South Georgia revealed no polymorphism. Number of specimens from Marion Island (n=5) were inadequate for mismatch analyses.

Population	Nucleotide div./ pairwise diff.	Goodness-of-fit test		Fu's F_S test		Tajima's *D* test	
		Rag. index	P	F_s	P	Tajima's D	P
South Georgia	0	-	-	-	-	-	-
Bouvet	0.0023/1.248	0.278	0.076	-0.485	0.376	-0.308	0.418
Marion Island	0.0007/0.400	-	-	0.09	0.299	-0.817	0.306

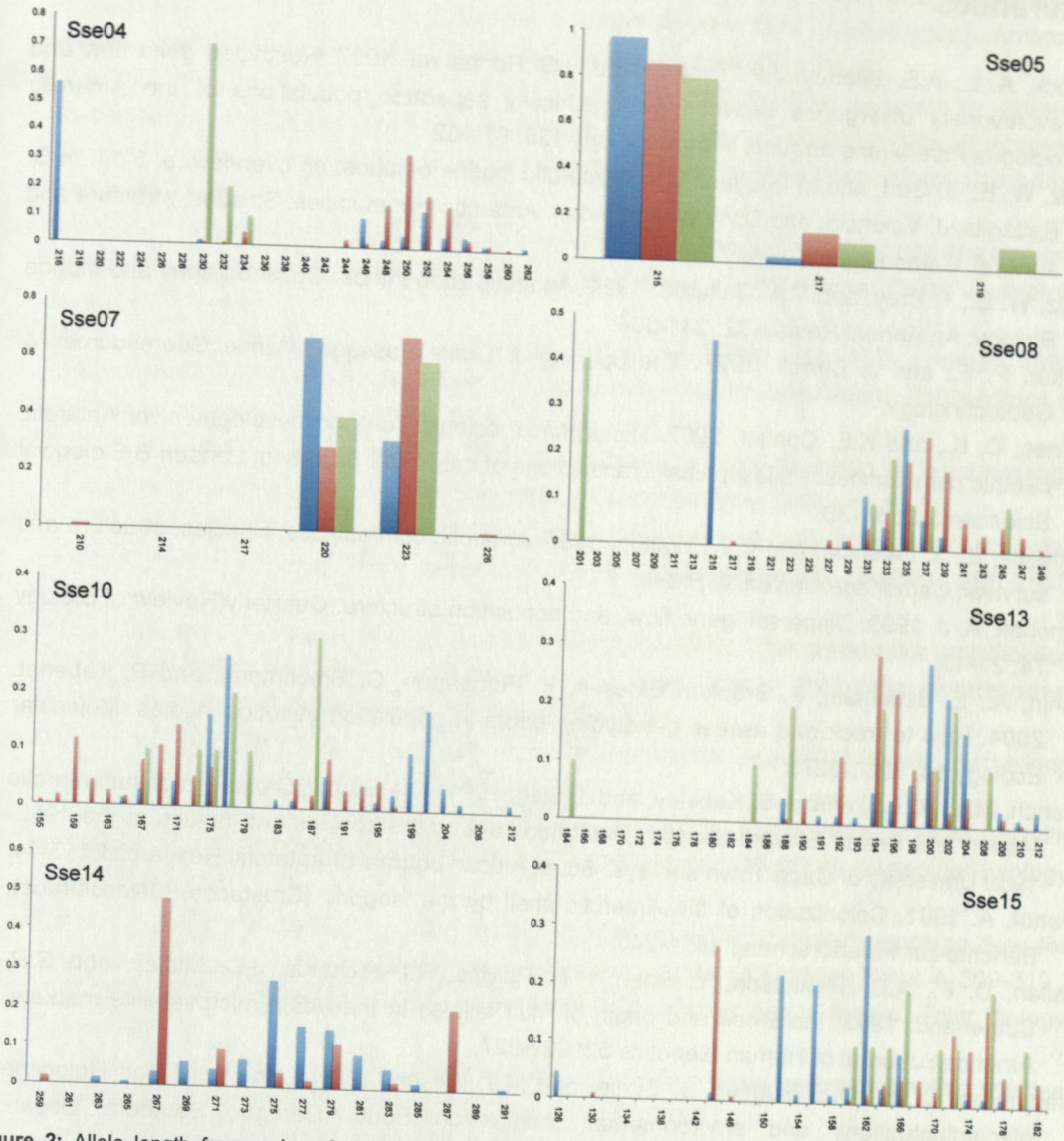


Figure 2: Allele length frequencies of eight microsatellite loci [GenBank: EU597351-EU597357] genotyped for the populations of *Septemserolis septemcarinata* from South Georgia (blue), Bouvet Island (red) and Marion Island (green, no data for Sse14).

References

Allcock, A. L., A.S. Brierley, J.P. Thorpe, and P.G. Rodhouse. 1997. Restricted gene flow and evolutionary divergence between geographically separated populations of the Antarctic octopus *Paraledone turqueti*. Marine Biology **129**: 97-102.

Arntz, W. E., J. Gutt, and M. Klages. 1997. Antarctic marine benthos: an overview, p. 3-14. *In* B. Battaglia, J. Valencia, and D.W. Walton [eds.], Antarctic communities: Species, structure and survival. Cambridge University Press.

Arntz, W. E., T. Brey, and V.A. Gallardo. 1994. Antarctic zoobenthos. Oceanography and Marine Biology: An Annual Review **32**: 241-304.

Barker, P. F., and J. Burrell. 1977. The opening of Drake Passage. Marine Georesources & Geotechnology.

Barnes, D. K., and K.E. Conlan. 2007. Disturbance, colonization and development of Antarctic benthic communities. Philosophical Transactions of the Royal Society of London B Biologcial Sciences **362**: 11-38.

Battaglia, B., J. Valencia, and D.W. Walton. 1997. Antarctic communities: Species, structure and survival. Cambridge University Press.

Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. Quarterly Review of Biology **74**: 21-43.

Bonin, A., E. Bellemain, P. Bronken Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. Molecular Ecology **13**: 3261-3273.

Branch, M. L., C.L. Griffiths, B. Kensley, and J. Sieg. 1991. The benthic Crustacea of subantarctic Marion and Prince Edward Islands: Illustrated keys to the species and results of the 1982-1989 University of Cape Town surveys. South African Journal of Antarctic Research **21**: 3-44.

Brandt, A. 1991. Colonization of the Antarctic shelf by the Isopoda (Crustacea, Malacostraca). Berichte zur Polarforschung **98**: 1-240.

Callen, D. F., A.D. Thompson, Y. Shen, H.A. Phillips, R.I. Richards, J.C. Mulley, and G.R. Sutherland. 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. American Journal of Human Genetics **52**: 922-927.

Clapperton, C. M., D.E. Sugden, J. Birnie, and M.J. Wilson. 1989. Late-glacial and Holocene glacierfluctuations and environmental change on South Georgia, Southern Ocean. Quaternary Research **31**: 210-228.

Clarke, A. 1979. On living in cold water: K-strategies in Antarctic benthos. Marine Biology **55**: 111-119.

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

Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144**: 2001-2014.

Dayton, P. K. 1990. Polar benthos, p. 631-685. *In* W.O. Smith [eds.], Oceanography Part B: Chemistry, biology and geology. Academic Press.

Dell, R. K. 1972. Antarctic benthos, p. 1-216. *In* F.S. Russell, and M. Yonge [eds.], Advances in marine biology. Academic Press.

di Prisco, G. 1997. Physiological and biochemical adaptations in fish to a cold marine environment, p. 251-260. *In* B. Battaglia, J. Valencia, and D.W. Walton [eds.], Antarctic communities: Species, structure and survival. Cambridge University Press.

di Prisco, G., M. Tamburini, and R. D'Avino. 1998. Oxygen transport systems in extreme environments: multiplicity and structure-function relationship in haemoglobins of Antarctic fish, p. 143-165. *In* H.O. Pörtner, and R.C. Playle [eds.], Cold ocean physiology. Cambridge University Press.

Di Prisco, G., V. Carratore, E. Cocca, A. Riccio, and M. Tamburrini. 2000. Molecular structure and functional adaptations of hemoglobins from Antarctic marine organisms. Italian Journal Of Zoology **67**: 37-46.

Di Rienzo, A., A.C. Peterson, J.C. Garza, A.M. Valdes, M. Slatkin, and N.B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences of the USA **91**: 3166-3170.

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

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**: 1567-1587.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology **3**: 294-299.

Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**: 915-923.

Gagneux, P., C. Boesch, and D.S. Woodruff. 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. Molecular Ecology **6**: 861-868.

Glaubitz, J. C. 2004. CONVERT: A user-friendly program to reformat diploid genotypc data for commonly used population genetic software packages. Molecular Ecology Notes **4**: 309-310.

Gutt, J. 2001. On the direct impact of ice on macrobenthic communities, a review. Polar Biology **24**: 553-564.

Gutt, J., B.I. Sirenko, I.S. Smirnov, and W.E. Arntz. 2004. How many macrozoobenthic species might inhabit the Antarctic shelf? Antarctic Science **16**: 11-16.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series **41**: 95.

Hebert, P. D., M.Y. Stoeckle, T.S. Zemlak, and C.M. Francis. 2004. Identification of Birds through DNA Barcodes. PLoS Biology **2**: e312.

Hedrick, P. 2005. A standardized genetic differentiation measure. Evolution **59**: 1633-1638.

Held, C. 2000. Phylogeny and biogeography of serolid isopods (Crustacea, Isopoda, Serolidae) and the use of ribosomal expansion segments in molecular systematics. Molecular Phylogenetics and Evolution **15**: 165-178.

Held, C. 2001. No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). Polar Biology **24**: 497-501.

Held, C. 2003. Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda), p. 135-139. *In* A.H. Huiskes,

W.W. Gieskes, J. Rozema, R.M. Schorno, S.M. van der Vies, and W.J. Wolff [eds.], Antarctic biology in a global context. Backhuys Publishers.

Held, C., and J.-. Wägele. 2005. Cryptic speciation in the giant Antarctic isopod *Glyptonotus antarcticus* (Isopoda: Valvifera: Chaetiliidae). Scientia Marina **69**: 175-181.

Helmuth, B., R.R. Veit, and R. Holberton. 1994. Long-distance dispersal of a subantarctic brooding bivalve (Gaimardia trapesina) by kelp-rafting. Marine Biology **120**: 421-426.

Highsmith, R. C. 1985. Floating and algal rafting as potential dispersal mechanisms in brooding invertebrates. Marine Ecology Progress Series **25**: 169-179.

Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. Genetical Research **38**: 209-216.

Hobday, A. J. 2000a. Age of drifting *Macrocystis pyrifera* (L.) C. Agardh rafts in the Southern California Bight. J Exp Mar Bio Ecol **253**: 97-114.

Hobday, A. J. 2000b. Persistence and transport of fauna on drifting kelp (Macrocystis pyrifera (L.) C. Agardh) rafts in the Southern California Bight. Journal of Experimental Marine Biology and Ecology **253**: 75-96.

Hoffman, J. I., and W. Amos. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology **14**: 599-612.

Hofmann, E. E. 1985. The large-scale horizontal structure of the Antarctic Circumpolar Current from FGGE drifters. Journal of Geophysical Research - Oceans **90**: 7087-7097.

Jakobsson, M., and N.A. Rosenberg. 2007. CLUMPP: a cluster matching and permutationprogram for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23**: 1801-1806.

Kimura, M., and J.F. Crow. 1964. The number of alleles that can be maintained in a finite population. Genetics **49**: 725-738.

Knox, G. A. 1994. The biology of the Southern Ocean. Cambridge University Press.

Lee, Y. H., M. Song, S. Lee, R. Leon, S.O. Godoy, and I. Canete. 2004. Molecular phylogeny and divergence time of the Antarctic sea urchin (*Sterechinus neumayeri*) in relation to the South American sea urchins. Antarctic Science **16**: 29-36.

Leese, F., and C. Held. 2008. Identification and characterization of microsatellites from the Antarctic isopod *Ceratoserolis trilobitoides* : nuclear evidence for cryptic species. Conservation Genetics **online early**.

Linse, K. 2006. New records of shelled marine molluscs at Bouvet Island and preliminary assessment of their biogeographic affinities, p. 120-127. *In* Polar Biology. .

Linse, K., T. Cope, A.-.N. Lörz, and C. Sands. 2007. Is the Scotia Sea a centre of Antarctic marine diversification? Some evidence of cryptic speciation in the circum-Antarctic bivalve *Lissarca notorcadensis* (Arcoidea: Philobryidae). Polar Biology **30**: 1059-1068.

Livermore, R., A. Nankivell, G. Eagles, and P. Morris. 2005. Paleogene opening of Drake Passage. Earth and Planetary Science Letters **236**: 459-470.

Luikart, G., F.W. Allendorf, J.M. Cornuet, and W.B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity **89**: 238-247.

Luxmoore, R. A. 1982. Moulting and growth in serolid isopods. Journal of Experimental Biology and Ecology **56**: 63-85.

Luxmoore, R. A. 1984. A comparison of the respiration rate of some Antarctic isopods with species from lower latitude. British Antarctic Survey Bulletin **62**: 53-65.

Maruyama, T., and P.A. Fuerst. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics **111**: 675-689.

McKelvey, K. S., and M.S. Schwartz. 2005. DROPOUT: A program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. Molecular Ecology Notes **5**: 716-718.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution **60**: 2399-2402.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution **29**: 1-10.

Otha, T., and M. Kimura. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res **22**.

Paetkau, D., and C. Strobeck. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. Molecular Ecology **4**: 519-520.

Page, T. J., and K. Linse. 2002. More evidence of speciation and dispersal across the Antarctic Polar Front through molecular systematics of Southern Ocean *Limatula* (Bivalvia : Limidae). Polar Biology **25**: 818-826.

Palumbi, S. R. 1994. Genetic divergence, reproductive isolation and marine speciation. Annual Review of Ecology and Systematics **25**: 547-572.

Park, S. D. E. 2001. Trypanotolerance in west African cattle and the population genetic effects of selection. University of Dublin.

Pearse, J. S., J.B. McClintock, and I. Bosch. 1991. Reproduction of Antarctic benthic marine invertebrates: Tempos, modes, and timing. American Zoologist **31**: 65-80.

Picken, G. B. 1980. Reproductive adaptations of Antarctic benthic invertebrates. Biological Journal of the Linnean Society **14**: 67-75.

Piry, S., G. Luikart, and J.-.M. Cornuet. 1999. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. Journal of Heredity **90**: 502-503.

Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: causes, consequences and solutions. Nature Reviews Genetics **6**: 847-859.

Poulin, E., A.T. Palma, and J.P. Feral. 2002. Evolutionary versus ecological success in Antarctic benthic invertebrates. Trends in Ecology & Evolution **17**: 218-222.

Poulin, E., and J.P. Feral. 1996. Why are there so many species of brooding Antarctic echinoids? Evolution **50**: 820-830.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**: 945-959.

Ramos-Onsins, S. E., and J. Rozas. 2002. Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution **19**: 2092-2100.

Raupach, M. J., and J.W. Wägele. 2006. Distinguishing cryptic species in Antarctic Asellota (Crustacea : Isopoda) - a preliminary study of mitochondrial DNA in Acanthaspidia drygalskii. Antarctic Science **18**: 191-198.

Raymond, M., and F. Rousset. 1995. GENEPOP Version 1.2: Population genetics software for exact tests and eucumenism. Journal of Heredity **86**: 248-249.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution **43**: 223-225.

Rosenberg, N. A. 2004. Distruct: a program for the graphical display of population structure. Molecular Ecology Notes **4**: 137-138.

Rosqvist, G. C., M. Rietti-Shati, and A. Shemesh. 1999. Late glacial to middle Holocene climatic record of lacustrine biogenic silica oxygen isotopes from a Southern Ocean island. Geology **27**: 967-970.

Rousset, F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics **142**: 1357-1362.

Rousset, F. 2007. GENEPOP '007: a complete reimplementation of the GENEPOP software for Windows and Linux. Molecular Ecology Notes **online first**.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585-595.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24**: 1596-1599.

Thatje, S., C.D. Hillenbrand, and R. Larter. 2005. On the origin of Antarctic marine benthic community structure. Trends in Ecology and Evolution **20**: 534-540.

Thiel, M., and L. Gutow. 2005. The ecology of rafting in the marine environment - I - The floating substrata, p. 181-263. *In* Oceanography And Marine Biology: An Annual Review, Vol 42. .

van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes **4**: 535-538.

Villesen, P. 2007. FaBox: an online toolbox for fasta sequences. Molecular Ecology Notes **7**: 965-968.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics **7**: 167-184.

Waples, R. S., and C. Do. 2007. LDNe: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Notes **online first**.

Wares, J. P., S. Daley, R. Wetzer, and R.J. Toonen. 2007. An evaluation of cryptic lineages of *Ideotea balthica* (Isopoda: Idoteidae): Morphology and microsatellites. An evaluation of cryptic lineages of Ideotea balthica (Isopoda: Idoteidae): Morphology and microsatellites .

Wägele, J. 1987. On the reproductive biology of *Ceratoserolis trilobitoides* (Crustacea: Isopoda): Latitudinal variation of fecundity and embryonic development. Polar Biology **7**: 11-24.

Weir, B. S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**: 1358-1370.

Weir, B. S., and W.G. Hill. 2002. Estimating F-statistics. Annual Review of Genetics **36**: 721-750.

Whitworth, T., W.D. Nowlin, and S.J. Worley. 1982. The net transport of the Antarctic Circumpolar Current through Drake Passage. Journal of Physiological Oceanography **12**: 960-971.

Wilson, N. G., R.L. Hunter, S.J. Lockhart, and K.M. Halanych. 2007. Multiple lineages and absence of panmixia in the "circumpolar" crinoid *Promachocrinus kerguelensis* from the Atlantic sector of Antarctica. Marine Biology **152**: 895-904.

Winkler, H. 1994. Characterization of the Scotia Arc isopod fauna (Crustacea, Malacostraca) from a biogeographical point of view: A multivariate approach. Berichte zur Polarforschung : 1-196.

Wright, S. 1978. Evolution and the genetics of populations. University of Chicago Press.

Chapter 9

Extensive genetic differentiation among populations of the brooding isopod *Serolis paradoxa* (Crustacea, Peracarida) from the Falkland Islands and the Strait of Magellan

Florian Leese, Anna Kop, Johann-Wolfgang Wägele, Christoph Held

BMC Frontiers in Zoology, under review

Abstract

Background

The Falkland Islands and Patagonia are traditionally assigned to the Magellan Biogeographic Province. Most marine species in Falkland waters are also reported from Patagonia, however, it remains unclear whether and how relatively immobile marine benthic, shallow-water species maintain gene flow. Past fluctuations in sea level during glacial maxima are regarded as a possible mechanism that might have recurrently allowed genetic exchange between both regions. The genetic exchange between Falkland Islands and Patagonia has never been estimated.

Results

This study analyses the genetic structure of populations of the marine shallow water isopod *Serolis paradoxa* (Fabricius, 1775) from the Falkland Islands and the Strait of Magellan and assesses genetic differentiation among populations based on seven nuclear microsatellites and a fragment of the mitochondrial 16S rRNA gene. Both marker systems report very strong differentiation among populations from the Strait of Magellan and the Falkland Islands with on average 2.0% uncorrected p-distance for the 16S and a standardized differentiation coefficient for the microsatellite data of $F'_{ST} > 0.86$. The magnitude of inferred differences leads to the conclusion that effective gene flow is absent between regions for this species today and has been for an amount of time far in excess of the age of the last glacial maximum (LGM). Genomic signatures of recent population expansion suggest that the LGM greatly shaped present-day population structure on a regional scale in the central Strait of Magellan. We argue that specimens from the Strait of Magellan and the Falkland Islands very likely represent two distinct species that separated in the late Pliocene or early Pleistocene (2.7-1.4 myr BP).

Conclusions

The results of this study advise caution concerning the untested assignation of the Falkland Islands to the Magellan zoogeographic province. The distance to the Falkland Islands may be insurmountable in particular for shallow water species as demonstrated for *S. paradoxa*, yet almost no pairs of potential sister species have been scrutinized for cryptic species.

Introduction

Present-day distribution of a species is the result of a complex interplay between (1) extrinsic factors such as isolation of landmasses, climatic conditions and availability of niches and (2) intrinsic factors such as dispersal ability and physiological tolerance. Extrinsic factors typically influence the distribution of many species in the same way because they act on an ecosystem scale. Over time, this leads to a characteristic assemblage of species with similar distribution patterns within larger geographical areas, so called biogeographic provinces, and distinct gaps between them. On the southern hemisphere, the Magellan Biogeographic Province has obvious close ties to the whole of South America to which it is connected today, but some of its species inventory stems from times before the Gondwana breakup [1, 2 and references therein]. The Falkland Islands are connected to the South American shelf and located approximately 500 km to the east of Patagonia (Fig. 1) and are commonly assigned to the Magellan Biogeographic Province [3 for review] based on current knowledge of their species inventory. However, originally the Falkland Islands drifted to their current position on a microplate that formed part of the African continental plate. Their current position was reached approximately 130 Myr BP [4-6]. Close biogeographic ties between the Falkland Islands and Patagonia are widely accepted and even more plausible in the marine realm because here biotic exchange does not depend on the existence of land bridges [7-10]. Furthermore, major ocean current systems facilitate dispersal of specimens. There are few reports of species endemic to the Falkland Islands [2, 11-13]. On the whole, marine evidence supports that the Falkland Islands form a part of the Magellan Biogeographic Province and that migration of species between the continental South America and the Falkland Island is occurring repeatedly. However, recent molecular studies have shown that unrecognized cryptic species may be more common than previously assumed [10, 14-23]. They indicate that morphological and ecological similarity may be an unreliable token on which to base taxonomic identifications and, by extension, the definition of biogeographic provinces derived from them.

In this study we investigate spatial partitioning of intraspecific molecular polymorphism in nominal *Serolis paradoxa* (Fabricius, 1775), a marine benthic isopod, using two independent genetic marker systems. *S. paradoxa* was originally described from the Falkland Islands but is also frequently reported from the Strait of Magellan, the Patagonian coastal shelf (Atlantic and Pacific side), and also from regions further to the equator [11, 24]. For the current taxonomic status and synonyms of *S. paradoxa* see [25]. In the Magellan region, *S. paradoxa* can be locally very abundant [26, Held pers. obs.]. Like nearly all isopods, *S. paradoxa* lacks free-swimming distribution stages and broods its offspring in a ventral brood pouch, the marsupium, and is thus expected to be limited in its dispersal capabilities. Direct measurements of dispersal and migration over large geographical distances are impractical in most marine systems and do not provide a means of directly assessing effective gene flow. Indirect genetic estimates use tools that analyse the genetic structure and its underlying microevolutionary forces and thus represent a more reliable and easier method [27].

However, when estimating present-day population structure, historical extrinsic factors must be considered that might have exerted a structuring force on present-day distribution of a species. One extrinsic factor that is known to have had a major impact on genetic structure and distribution of species are glaciation events [28]. Their influence on the marine fauna is two-fold: large-scale glaciations may directly render entire coastal habitats unavailable during glacial

maxima [29-31] and lead to a decrease in sea level of up to 130 m [30]. The latter may have disrupted inshore habitats on either side of an emerging barrier (e.g. appearance of the Panama land bridge [7], or connected shallow water habitats that are disjunct during periods of high sea level.

In this study we investigate the genetic structure of *S. paradoxa* from the Falkland Islands and the Patagonian shelf. The present-day situation suggests that the deeper waters on the South American shelf may present an insurmountable barrier to *S. paradoxa* inhabiting shallow waters around Patagonia and the Falkland Islands. Historically, the lower sea level during glacial maxima may have brought both regions closer together and facilitated migration between habitats thus counteracting independent lineage sorting in the two regions.

By investigating the coherence between gene pools and construction of an approximate timeline we test whether the disruptive or unifying forces predominated in the case of *S. paradoxa* and if the influence of the last glaciations, exerted a major influence on the evolutionary history of the species. Furthermore, we test whether the relatively young age of the marine habitats in the central Strait of Magellan (maximum 14 Kyr BP, [32-38]) left signatures in the populations' genome that differ from those of populations from shallow water habitats around the Falkland Islands, which were little affected by glaciations [39]. In particular, we test for patterns of recent population expansions. We hypothesize that populations in the central Strait of Magellan are genetically less diverse than populations from the coast or the Falkland Islands due to recent range expansion into the Magellan Strait subsequent to the retreat of the glaciers after the LGM.

Material and Methods

Taxon sampling

Specimens from the Falkland Islands (FI) were collected by bottom trawling during the ICEFISH 2004 expedition, specimens from the Strait of Magellan near Punta Arenas (PA) were collected by CH SCUBA diving and specimens from the Atlantic opening of the Strait of Magellan (AO) were provided from the 2nd Cruce estrecho, 2003, by Carlos Rios and Erika Mutschke, Universidad de Magallanes, Punta Arenas (see Fig. 1 and Tab. 1). Animals were immediately preserved in 96% ethanol. Microsatellite analyses were performed for 35 specimens from PA, 33 from AO and 23 from FI. A subset of 19 specimens from PA and 22 samples from FI were analysed for variation of the 3'-terminus of the mitochondrial 16S rRNA gene.

Table 1: Sampling sites (PA=Strait of Magellan near Punta Arenas, AO=Atlantic opening of the Strait of Magellan, FI=Falkland Islands) and numbers of specimens studied using nuclear microsatellite markers (N_{Msat}) and mitochondrial markers (N_{mtDNA}) of the populations of *S. paradoxa* investigated in this study.

Region	depth [m]	Cruise	N_{MSAT}	N_{mtDNA}
Strait of Magellan near Punta Arenas	8 m	SCUBA diving by CH	35	19
Opening of the Strait of Magellan to the Atlantic Ocean	<20 m	2nd Cruce estrecho 2003	33	0
West Falkland Islands	15 m	ICEFISH 2004	23	22

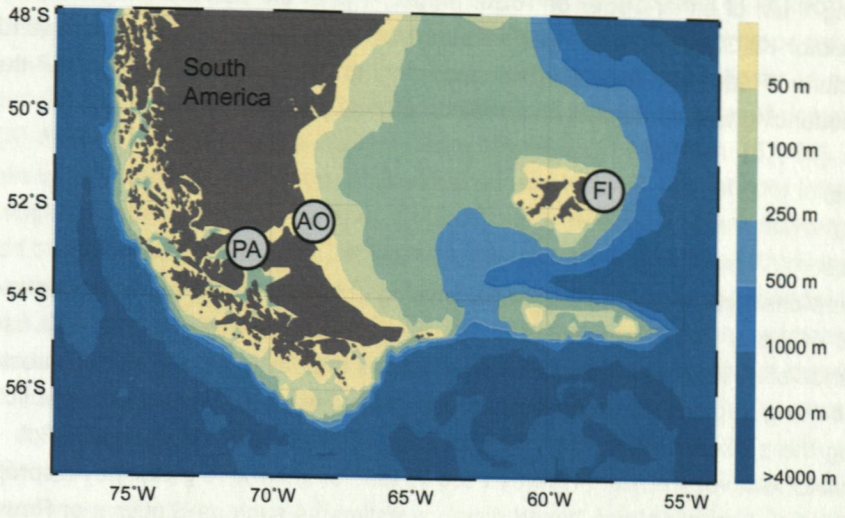


Figure 1: Sampling sites of *Serolis paradoxa* in the Strait of Magellan near Punta Arenas (PA), the opening to the Atlantic Ocean (AO) and the Falkland Islands (FI).

DNA extraction, PCR, sequencing/genotyping

Total DNA was extracted from muscle tissue using the Qiagen DNeasy Mini Kit according to the standard tissue protocol. Only 100 µl of elution buffer were used to increase DNA concentration.

Microsatellites

Microsatellite markers Spa04, Spa12, Spa34, Spa35, Spa39, Spa42 and Spa43 [23] were applied to assess intraspecific genetic polymorphisms for all specimens from the three sampling sites. Standard 15 µl reactions consisted of 1x PCR HotMaster Buffer, 0.2 mM dNTPs, 0.5-0.75 µM of each primer (one labelled, one unlabelled), 0.03 U/µl HotMaster *Taq* (Eppendorf, 5-prime), 0.5 M Betaine (Sigma Aldrich) and 5-20 ng of genomic DNA. Cycling conditions on an epgradient thermocycler (Eppendorf) were 2 min at 94°C followed by 30 to 34 cycles with 20 s at 94°C, 15 s at annealing temperature, 30 s at 65°C. A final extension step of 45 minutes at 65°C was performed to reduce *in vitro* artefacts due to incomplete adenylation of products [see 23 for details]. PCR products were visualized on 2% TBE agarose gels, diluted 1-15 fold with molgrade water (Carl Roth) and 1 µl of the diluted product was denatured in a mixture of 14.7 µl HI-DI formamide with 0.3 µl GeneScan ROX 500 size standard (both Applied Biosystems). Allele length scoring was performed using the software GENEMAPPER 4.0 (Applied Biosystems). To minimize genotyping errors [40, 41], up to four independent reactions were performed on a subset of samples to estimate allele calling errors.

16S rDNA

The universal primers 16Sar and 16Sbr [42] were used for amplification. Reactions were carried out in 25 µl volumes with 1x HotMaster reaction buffer, 0.2 mM dNTPs, 0.5 µM of each primer, 0.025 U/µl HotMaster *Taq* (Eppendorf, 5-Prime). Reaction conditions were initial denaturation for 2 min at 94°C followed by 36 cycles of 20 s at 94°, 15 s at 46°C and 80 s at 65°C plus a final elongation step of 5 min at 65°C. PCR products were purified using Qiagen QIAquick and Eppendorf Perfectprep Gel cleanup kits. Cycle-sequencing was performed in 10 µl reaction

volumes using 1 µM of either 16Sar or 16Sbr primer, 1 µl of the purified template DNA and the BigDye Termiator Kit 3.1 chemistry (Applied Biosystems) according to the recommendations of the manufacturer. Reactions were purified according to the 'modified protocol' of the Qiagen DyEx Kit. Sequencing was conducted on an ABI 3130xl sequencer.

Data analysis

Microsatellites

Raw data were checked and corrected for genotyping errors using the software MICRO-CHECKER version 2.2.3 [43] and DROPOUT version 1.3 [44]. In addition, MICRO-CHECKER was used to test for the presence of null alleles in populations, i.e. alleles that fail to amplify due to substitutions in the primer binding regions. Corrected genotype tables were converted to specific software formats using the software MSTOOLKIT version 3.1 [45] and CONVERT version 1.3.1 [46]. The program ANIMALFARM version 1.0 [47] was used to test for loci with significantly disproportionate variances that may bias allele-size based distance estimates such as Slatkin's or Rousset's R_{ST} estimates [48, 49]. Tests for Hardy Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were performed using GENEPOP version 4.0.6 [50]. Parameter settings: 10,000 dememorization steps, 50 batches, 20,000 MCMC sampling steps. HWE tests aim at testing whether there is a statistically significant deviation of genotype frequencies from those expected according to Mendelian inheritance. Linkage disequilibrium occurs when two genomic loci are not inherited independently, e.g. due to physical linkage or other processes at population level hindering independent recombination of loci.

To assess partitioning of genetic variability witin individuals, subpopulations and regions, we performed hierarchical analyses of molecular variance (AMOVA) using ARLEQUIN version 3.11 [51]. Therefore, populations PA and AO were assigned to one group, while FI constituted the other group. In addition, single and multilocus inbreeding coefficients (F_{IS}) and pairwise population coancestry coefficients (F_{ST}, similar to Weir and Cockerham's Theta) were estimated as in [52] using GENEPOP. We also calculated pairwise allele-size based differentiation estimates, R_{ST}, according to [49] using GENEPOP. Significance was assessed by exact G tests as implemented in GENEPOP. Since the interpretation of the F_{ST} values from multiallelic data is problematic because their maximum values depend on the amount of within-population variation and can, even in the absence of any shared allele, often not reach the theoretical maximum of 1 [53-55]. We therefore applied a standardization approach suggested by Hedrick (2005) for calculcations of G_{ST} [55] and derived by Meirmans (2006) for Analysis of Variance frameworks (ANOVA) [56].

The main principle of this standardization approach is to consider the maximum possible value for F_{ST} as follows: $F_{ST(max)}$ was calculated using GENEPOP applying the sampling bias correction suggested by Meirmans (2006) using the Software RECODEDATA [56] to reformat the data set. F'_{ST} was subsequently individually calculated by dividing F_{ST} by this inferred maximum value.

The standardized F_{ST} measure calculated range from 0 (populations equifrequent for all alleles) to 1 (populations fixed for different alleles) and therefore makes interpretation of the degree of subdivision much easier and results comparable among studies.

In addition to these ANOVA based coancestry estimates we performed Bayesian individual assignment tests on the data using the program STRUCTURE, version 2.2.2 [57] to investigate

population subdivision. The advantage of the Bayesian clustering algorithm of STRUCTURE is that no classification of populations has to be done *a priori*. Assuming HWE and no or only weak LD within subpopulations, STRUCTURE assigns individual genotypes probabilistically to populations and calculates the log probability of the genotype dataset for a given number of populations (K), i.e. ln Pr (D|K) for K=1 to K=n, using a Markov Chain Monte Carlo algorithm [57, 58]. For the *S. paradoxa* data set, the most likely number of populations was inferred without prior information on geographic origin of individuals. The number of MCMC steps needed to reach convergence was first estimated by comparing run lengths between 10,000 and 2,000,000 steps. Convergence was generally reached with <5,000 steps. Therefore, for the parameter sets 10 independent runs with a burn-in of 5,000 and subsequent 100,000 MCMC steps were performed with and without assuming recent admixture in the prior model, and considering alleles as correlated and uncorrelated. The number of clusters (K) to infer was defined from K=1 to K=4 to allow detection of potential cryptic subpopulations. Alpha was inferred from the data for each population separately. Results from 10 independent runs were analysed in CLUMPP, version 1.1.1 [59] to compute a consensus membership coefficient Q-matrix from all 10 independent Q-matrices. Both, the individual Q-matrix and averaged population Q-matrix were visualized using DISTRUCT, version 1.1 [60].

To assess estimates of the present effective population size (N_e), we applied the linkage disequilibrium method proposed by Hill (1981) [61], modified by Waples (2006) [62] to account for a bias correction when sample size is much smaller than effective population size. This method is implemented in the program LDNE, version 1.3 [63]. Calculations of N_e and the confidence intervals were estimated considering alleles with a frequency of c ≥0.05 and c ≥0.02 and ≥0.01, respectively.

Tests for historical population bottlenecks were performed using the program BOTTLENECK [64]. Tests implemented in this program are based on the hypothesis that populations that have experienced recent reductions in their effective population size (N_e) show a reduction in both allelic richness and heterozygosity. In populations decreasing in size, the number of alleles (N_A) drops faster than heterozygosity [65] and therefore the observed heterozygosity is larger than the expected heterozygosity (H_O>H_E). Conversely, in expanding populations often the number of alleles increases faster than heterozygosity until equilibrium is reached. From the correlation of both parameters, allelic diversity and heterozygosity, it is possible to make inferences on historical demography of a population. For each locus and population BOTTLENECK computes distribution of H_E expected from the observed N_A, given the sample size (n) under the assumption of mutation-drift equilibrium. This distribution is obtained through simulating the coalescent process of n genes under the three possible mutation models, i.e. a) the Infinite Allele Model (IAM), b) the Two-Phase Model (TPM), c) the Stepwise-Mutation Model (SMM). As recommended by Cornuet and Luikart (1996) we tested several proportions of the SMM for the TPM (70-90%) [66]. Statistical significance of the parameters were inferred applying a Sign-test and a Wilcoxon-rank-test [64, 66, 67].

16S rDNA

Assembly of forward and reverse strands and editing was performed using the software SEQMAN (Dnastar, Lasergene). Sequence alignment was performed using the CLUSTALW program with its default parameter settings as implemented in BIOEDIT V.7.09 [68]. The alignment required only little manual correction, which was based on secondary structure information [69]. Sequence

variation was analyzed using MEGA 4.0 [70]. Gene diversity and nucleotide diversity according to Nei (1987) [71] and Theta based on the number of segregating sites, Theta (S), were calculated with ARLEQUIN version 3.11. Assuming neutrality, evidence of a population expansion was tested applying Tajima's *D* [72] and Fu's F_S statistic [73] as implemented in ARLEQUIN 3.11 applying a coalescent simulation approach generating 10,000 selectively neutral samples for assessment of significance of results. A test for sudden population expansion based on the distribution of pairwise difference (mismatch distribution) was calculated using ARLEQUIN and significance was assessed by 50,000 pseudo replicates.

A neighbor joining tree [74] with bootstrap support (1000 replicates) was calculated using PAUP* version 4b10 [75] and edited using DENDROSCOPE version 1.0.5 [76]. Sequences of the serolid isopods *Cuspidoserolis luethjei* and *Cuspidoserolis johnstoni* (GenBank accession numbers AJ269802, AJ269803; see [15]) were used as outgroups based on their close relationship to *S. paradoxa* [15]. A statistical parsimony network of the sequence data was created using TCS version 1.21 [77].

Results

(Tables 2, 3 and 6 in Appendix)

Microsatellites

Seven microsatellite loci were analysed for three populations. All loci were highly polymorphic for all three populations (Tab. 2). Based on allele frequencies, the geographically intermediate population AO is clearly more similar to PA than FI (Tab. 3). The allele distribution of all microsatellite loci reveal strongly differing frequency spectra with several private and almost fixed different allele patterns between regions (e.g. locus Spa04, Spa12, Spa35, Spa43). Allele length spectra differ between populations but overlap (Tab. 3). The number of alleles per microsatellite locus ranged from 6 to 23 (Tab. 3). The observed heterozygosity ranged from 0.0 (locus 39, all specimens homozygous for populations FI) to 0.886. Significant deviations from HWE were detected for loci Spa04 and Spa39. Analyses with MICRO-CHECKER indicate that null alleles are likely to be the cause for inflated homozygosity. No significant global linkage disequilibrium was observed after sequential Bonferroni correction [78].

Results of ANIMALFARM confirmed that none of the loci contributed disproportionally to distance-based differentiation estimates after Bonferroni or Sidak adjusting of the significance level.

Results of the AMOVA indicate that most variation is distributed among individuals. Φ_{IS} and multilocus F_{IS} were significantly positive for all populations (Tab. 2, Tab. 4) indicating within-population structure. However, a large proportion of variation is distributed among major geographical regions [(PA + AO) vs. FI] and only a minor but nevertheless significant proportion between populations (Tab. 4). Strong differences between the populations from the two regions can be observed when comparing allele frequency patterns at all loci. In particular at loci Spa04, Spa12 and Spa43, populations from the two major regions [(PA, AO vs. FI)] are nearly fixed for different alleles (Tab. 3). PA and AO reveal very similar allele frequency patterns for most loci.

These differences are expressed by the high and significant pairwise F_{ST} and R_{ST} estimates between population PA and FI and populations AO and FI (Tab. 5). Differentiation estimates were even higher between populations AO and FI (Tab. 5). In contrast, F_{ST} estimates among populations PA and AO are low albeit significant (P=0.0005). R_{ST} do not support significance differentiation between PA and AO (Tab. 7).

Table 4: Hierarchical analyses of molecular variance (AMOVA) among *Serolis paradoxa* populations within and between two regions (2 populations from the Strait of Magellan, one population from the Falkland Islands) using 7 microsatellite markers.

Component of differentiation	df	variation [%]	Φ statistics	P
Among groups	1	30.99	Φ_{CT}=0.310	0.329
Among populations within regions	1	1.64	Φ_{SC}=-0.024	0
Among individuals within populations	88	8.02	Φ_{IS}=0.119	0
Within individuals	91	59.35	Φ_{IT}=0.407	0

Table 5: Genetic differentiation among populations of *Serolis paradoxa* from three stations as assessed by F-statistics (F_{ST}, lower diagonal) and R-statistics (R_{ST}, upper diagonal), based on seven polymorphic microsatellite loci. * refers to P<0.001 (exact G test). F_{ST} values in parentheses are differentiation estimates standardized according to Meirmans (2006).

	PA	AO	FI
PA	-	-0.006	0.217*
AO	0.023* *(0.065)*	-	0.258*
FI	0.322**(0.863)*	0.376**(0.901)*	-

R_{ST} estimates were lower than F_{ST} estimates in this study. In general, R_{ST} is hypothesized to be larger if an appreciable amount of differentiation between populations is not only caused by drift but by independent mutations in the different populations according to a stepwise mutation model (SMM). Consequently, R_{ST} distance measures are considered a 'memory' of past mutations [48]. F_{ST} is superior to R_{ST} when populations have diverged mostly by means of random genetic drift and migration m (i.e. mutation rate << migration).

Applying Hedrick's standardization approach for pairwise F_{ST} calculates in this study [56], differentiation between PA and FI is 0.86, between AO and FI 0.91, and among the Magellan Strait populations PA and AO 0.063. These values demonstrate that both regions are almost fixed for different alleles at the seven loci investigated. When removing locus 39, which is biased due to the presence of null-alleles in population FI, the standardized values do not change, however, the non-standardized F_{ST} values are almost twice as high.

Inferring the most likely number of populations without making assumptions concerning their delimitation, STRUCTURE identifies only two very distinct clusters, which correspond to the two major regions (AP+AO vs. FI; Fig. 6). When using the no-admixture model, all individuals are correctly assigned to the two regions with admixture proportions of 1.000 and 0.000, respectively. There was no additional substructure within populations (PA, AO, FI), i.e. when analysing the populations separately ln Pr (D|K) was highest for K=1. No significant differentiation between population PA and AO was detected (compare F_{ST} estimates, Tab. 5).

Estimating effective population sizes using linkage disequilibrium as outlined by Waples (2006) we find evidence for very large effective population sizes in both regions as indicated by negative N_e estimates and an upper limit of its confidence interval reaching infinity.

We tested for recent demographic contractions or expansions by looking for deviations from mutation-drift equilibrium under different mutation models using BOTTLENECK. We found a significant heterozygosity deficiency under particular mutation models: For population AO there is a significant heterozygosity deficit under both SMM and TPM models (Tab. 8), which provides strong evidence for recent population expansion. For FI the evidence for recent population expansion is somewhat weaker: a significant heterozygosity deficit is detected only using the SMM and the Wilcoxon test (P=0.0195, Tab. 8). Thus results of BOTTLENECK do not provide evidence of a similarly drastic decline and subsequent recovery in population size for FI. For population AO the under a TPM and a strict SMM, the significant excess of heterozygosity may indicate that this population is expanding presently. Although the evolutionary dynamics of microsatellites are not fully understood [79, 80] it is commonly accepted that the IAM model is not an appropriate descriptor of the dynamics of microsatellite markers and hence that its application often leads to unrealistic conclusions.

16S rDNA

We sequenced a 490 bp fragment of the 16S rRNA gene for a subset of 19 specimens from population PA and 22 specimens from population FI (Tab. 1) to test whether the strong pattern of differentation inferred using fast evolving microsatellites is also traceable with a slower evolving gene. The amplified fragment was AT-rich as typical for this gene (A 39.0%, C 13.5%, G 14.1%, T 33.4% [81]. Of the sixteen polymorphic positions, ten were parsimony informative and six represented singletons. Substitutions were located only in loop regions of the rRNA gene fragment (folding model: *Drosophila melanogaster* 16S rRNA, [69]. Ten haplotypes were characterized (HT1-HT10, Tab. 7). The statistical parsimony network constructed is characterized by two shallow subnetworks (≤ 3 segregating sites) representing PA and FI, which are connected by a long internal branch (8 segregating sites, Fig. 2). None of the haplotypes was shared among specimens from different regions. Phylogenetic analyses revealed that specimens from both regions form two reciprocally monophyletic clades each supported by high bootstrap values (Fig. 3). The average uncorrected genetic pairwise distance between both groups was 2.0% for transitions (1.6%) and transversions (0.4%). Variation within groups was an order of magnitude lower (0.2% among FI and 0.1% among PA and AO; Fig. 4). These values are amongst the lower values observed between reproductively isolated species of serolid isopods and other crustaceans [15, 17-19, 22]. The existence of distinct gene pools and the barcoding gap is further corroborated by the bimodal distribution of pairwise distance values (Fig. 4).

Table 7: Distribution of the n=41 16S rDNA sequences on the two sampling locations and GenBank accession number.

Haplotype	PA	FI	Accession no.
HT1	0	1	EU419766
HT2	0	9	EU419767
HT3	0	9	EU419768
HT4	0	1	EU419769
HT5	0	1	EU419770
HT6	0	1	EU419771
HT7	16	0	EU419772
HT8	1	0	EU419773
HT9	1	0	EU419774
HT10	1	0	EU419775

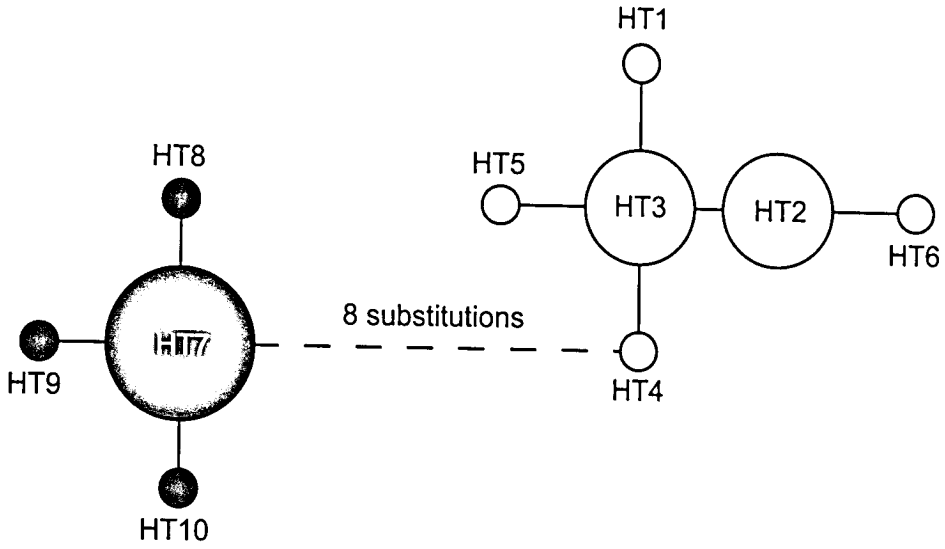


Figure 2: Statistical parsimony network of 16S rDNA haplotypes of *Serolis paradoxa* from the Strait of Magellan (grey) and the Falkland Islands (white). Branches in subnetwork represent one substitution except for the branch connecting HT4 and HT7, which differ by eight mutations.

Table 8: Genetic diversity and neutrality indices for the 16S rDNA data sets.

parameter	Magellan Strait (PA)	Falkland Islands (FI)
nucleotide diversity	0.0006 ± 0.001	0.002 ± 0.001
gene diversity	0.298 ± 0.133	0.688 ± 0.066
Theta (S)	0.858 ± 0.550	1.372 ± 0.733
Tajima's *D*	-1.719 (**P=0.022**)	-1.058 (P=0.165)
Fu's *Fs*	-2.677 (**P=0.001**)	-2.489 (**P=0.017**)

Nucleotide diversity and gene diversity were on average one order of magnitude lower for PA compared to FI (Tab. 8). Estimates of Theta (S) were also higher for FI than for PA (Tab. 8). Tajima's *D* and Fu's *Fs* were negative and significantly different from zero for PA, but only for Fu's *Fs* in FI (Tab. 3). *D* and *Fs* are measures of the relation of segregating sites and number of haplotypes. Recent population expansions are frequently associated with negative values of *D* and *Fs* because under these circumstances mutation generates more and closely related haplotypes than are eliminated by genetic drift. It should be considered that according to Fu [73] a significance level of 5% corresponds to P=0.02 thus Fu's *Fs* for population FI is only marginally significant. The mismatch analyses cannot the reject the assumptions of sudden population expansion for both, PA and FI (Tab. 9, Fig. 5) but the significance of the distribution of the mismatch analysis is only approaching a significance level of 5% for FI (raggedness P=0.0577, Tab. 9) whereas the situation is clearer for PA. In summary, there is a stronger signature for population expansion in PA than in FI based on the 16S rDNA data.

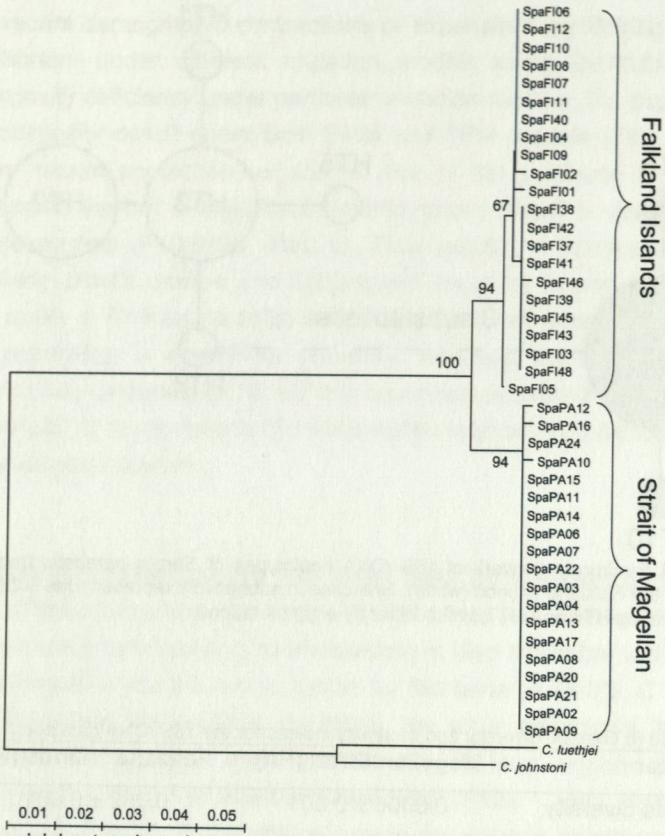


Figure 3: Neighbor joining tree based on uncorrected p-distances of 3'-terminus of the mitochondrial 16S rRNA gene with sequences of n=19 specimens from the Strait of Magellan (SpaPA) and n=22 specimens from the Falkland Islands (SpaFI). Sequences of *Cuspidoserolis luethjei*, AJ269802 and *C. johnstoni*, AJ269803 [15] were used as outgroup. Numbers on branches represent bootstrap support (1000 replicates).

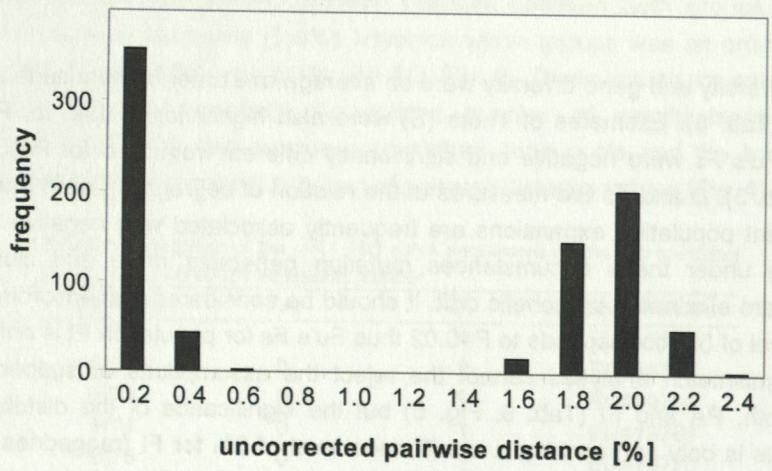


Figure 4: Frequency distribution of uncorrected pairwise genetic distances of the mitochondrial 16S rRNA gene among 41 specimens of *Serolis paradoxa* from the Strait of Magellan (SpaPA, n=19) and the Falkland Islands (SpaFI, n=22). Between population distances were an order of magnitude larger (1.6 – 2.4%) than within population distances (0.0 – 0.6%).

Table 9: Mismatch analysis based on the 16S rDNA data sets. Confidence intervals (CI give the 5% and 95% values, respectively for the parameters estimated by 50,000 bootstrap replicates.

Parameter	Magellan Strait (PA)	Falkland Islands (FI)
Tau	3.000	1.049
CI (Tau)	0.359-3.000	0.486-1.850
Theta 0	0.000	0.000
CI (Theta 0)	0.000-0.002	0.000-0.062
Theta 1	0.445	inf.
CI (Theta 1)	0.000-inf.	4.100-inf.
SSD	0.007 (P=0.455)	0.024 (P=0.120)
Raggedness	0.618 (P=0.618)	0.176 (P=0.058)

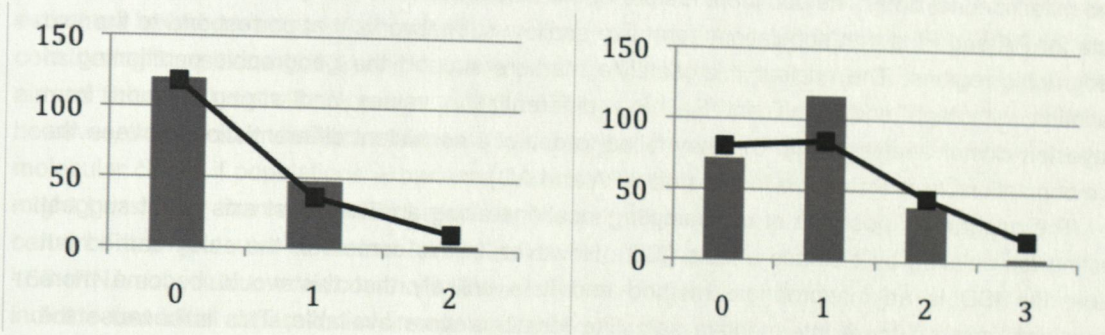


Figure 5: Mismatch analysis based on the 16S rRNA gene for population PA (left) and population FI (right). Columns represent the number of mismatches (x-axis) between haplotypes observed and the line the number of mismatches expected under a model of sudden population expansion.

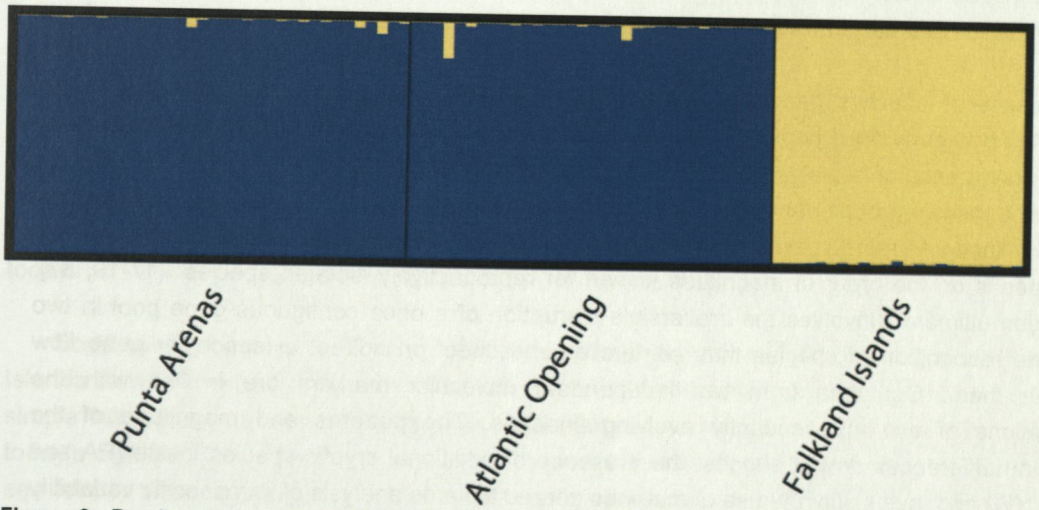


Figure 6: Results of cluster analyses performed with STRUCTURE (admixture model, allele frequencies correlated) with the highest log likelihood probability. The graphs display the consensus membership coefficients matrices (Q-matrices) for 91 individuals from three populations of *Serolis paradoxa* using seven microsatellite loci. The genotype of each individual is represented by a single bar, where the proportion of the colour refers to the probability of assignment to a certain cluster.

Discussion

The genetic variability within the nominal species *Serolis paradoxa* turned out to be extremely structured. The genetic differences exceed one order of magnitude in each marker system and reveal geographic structure. The differences in mutation rates and coalescent dynamics of the two marker systems help describe present-day population structure and reconstruct historical demographic processes.

Two genetically distinct lineages

There is a deep divergence between populations from both regions supported by microsatellite and mitochondrial data. The dominant feature of the intraspecific variability of mitochondrial DNA data for PA and FI is that populations form two shallow subnetworks that correspond to the two geographic regions. The nuclear microsatellite markers support the geographic partitioning of variation with high and significant F_{ST}, R_{ST} differentiation values and strong support from Bayesian cluster analyses (Fig. 6). They failed to detect a consistent differentiation between the two populations from Patagonia in this study (PA and AO).

The geographic positions of our sampling locations along an East-West axis might suggest testing for isolation by distance effects (IBD). However, in this context of the study carried out here, the IBD is an inappropriate method and it is unlikely that this would become more meaningful even if more intermediate sampling locations were available. This is because the central Strait of Magellan became available for (re)colonization only very recently (approximately 9-14 Kyr BP [35, 36]. This rapid range expansion is typically accompanied by loss of alleles and an excess of homozygosity [82] which violates a mutation-drift equilibrium assumed by the IBD model. Investigating distance effects on the distribution of intraspecific variance inside the Magellan Strait offers a means to trace the recolonization of this young habitat and would be appropriate for IBD but this requires more fine-scaled sampling and is outside the scope of this paper.

Absence of effective gene flow between the Falkland Islands and Patagonia is strongly suggested by nearly fixed population specific differences in fast evolving microsatellites and the perfect congruence of haplotype identity and geography for the 16S rDNA data. The long branch connecting the two groups of haplotypes (Fig. 2) and their reciprocal monophyly (Fig. 3) indicates complete lineage sorting in both groups. The magnitude of genetic differentiation between 16S genotypes is on the order of magnitude known for reproductively isolated species [17-19, 83]. Speciation ultimately involves the irreversible disruption of a once contiguous gene pool in two [84]. The recognition of species thus centers around direct or indirect evidence for gene flow between them. Our data from two independent molecular markers are in line with the expectations of two independently evolving lineages. The patterns and magnitude of the remaining differences do not suggest the presence of additional cryptic species inside (PA and AO) vs. (FI) and thus justify the use of methods geared towards analysis of intraspecific variability within the newly recognized species (see below).

The congruence between both marker systems supports that the 16S rRNA gene tree reflects the species tree rather than being a result of shared ancestral polymorphisms [85] or other processes affecting mitochondrial genes (see [86] for review).

Evolutionary history of nominal *S. paradoxa*

Genetic diversity estimates for FI (Tab. 8) indicate that population FI is relatively close to assumptions of a large and genetically diverse stationary population although there is also a weak signal for recent population expansion. In comparison, populations from Patagonia are less diverse for the 16S rDNA with one dominant haplotype only (HT7) and reveal strong evidence for recent population expansion in the Strait of Magellan. In summary, our data are in agreement with the following scenario: Populations of an ancestral species were separated geographically and evolved in allopatry (Falkland Islands vs. Patagonia). Applying a rate for the accumulation of substitutions in 16S rDNA estimated by Held (2001) [87] for the serolid isopod *Ceratoserolis trilobitoides* (Eights, 1833) with a rate of transversions of 0.14% per Myr and about 0.37% per Myr for transitions and transversions, the time of divergence can be estimated based on the distance estimates based on transversions (0.4%) as $T_{0\ Tv}$ = (0.4%/0.14% per Myr)*0.5 = 1.4 Myr. When considering transitions (1.6%) and transversions (0.4%) the divergence time is estimated to be almost twice as long with $T_{0\ Ti+Tv}$ = (2.0%/0.37% per Myr) *0.5 = 2.7 Myr. It must be stated, however, that genetic distances between two lineages increases much faster than predicted by molecular clocks if populations experience population bottlenecks [53] Thus, the coalescent time might be shorter than calculated based on the molecular clocks. In addition, it is not entirely certain if the molecular clock can be applied to *S. paradoxa*. The time estimates are based on 16S rRNA substitution rates commonly used for other Crustacea [83]. Our data, however, clearly indicate that initial separation of lineages predates the last glaciations and took place in the late Pliocene or the early Pleistocene (2.7-1.4 Myr BP). We must therefore reject the hypothesis that low sea levels during glacial periods led to significantly elevated levels of gene flow between populations of *S. paradoxa* due to greater proximity of shallow-water habitats as demonstrated by the strong genomic signatures of differentiation. A similar argument applies to potential migration between Patagonia and the Falkland Islands via passive rafting on drifting substrates. Although there are major directional ocean currents that frequently transport substrates suitable for transportation of even rather immobile species [88, 89] this apparently played no role in the evolutionary history of *S. paradoxa*. This species exclusively inhabits soft-bottom shallow waters and is frequently half-buried in the sediment (Held pers. observ.). Its capability to colonize new island habitats and maintain genetic continuity across barriers to dispersal and over evolutionary times is therefore small. Further sampling effort should focus on sampling specimens from the West Falkland Islands. It cannot be excluded that members of both lineages live in sympatry today.

Southern hemisphere glaciations affected both regions in a different way: The Falkland Islands were little affected by glacial advances, *S. paradoxa* had the opportunity to survive by simple following the rising and falling sea level changes. In Patagonia, however, major parts of today's distribution of nominal *S. paradoxa* became unavailable due to ice coverage and/or low sea levels. Western Patagonia was covered by a contiguous ice shield similar to the Antarctic Peninsula today and the central Strait of Magellan was inundated only after the LGM, approximately 14-9 Kyr BP (Fig. 7) [35]. *S. paradoxa* was thus forced to immigrate into the Strait of Magellan. The character of a founder population that recolonized this region is still preserved in the depauperate 16S rDNA which is characterized by one dominant haplotype and few low frequency haplotypes that are derived from this dominant haplotype, contributing to a star-like topology of the network which is typical for population in mutation-drift disequilibrium.

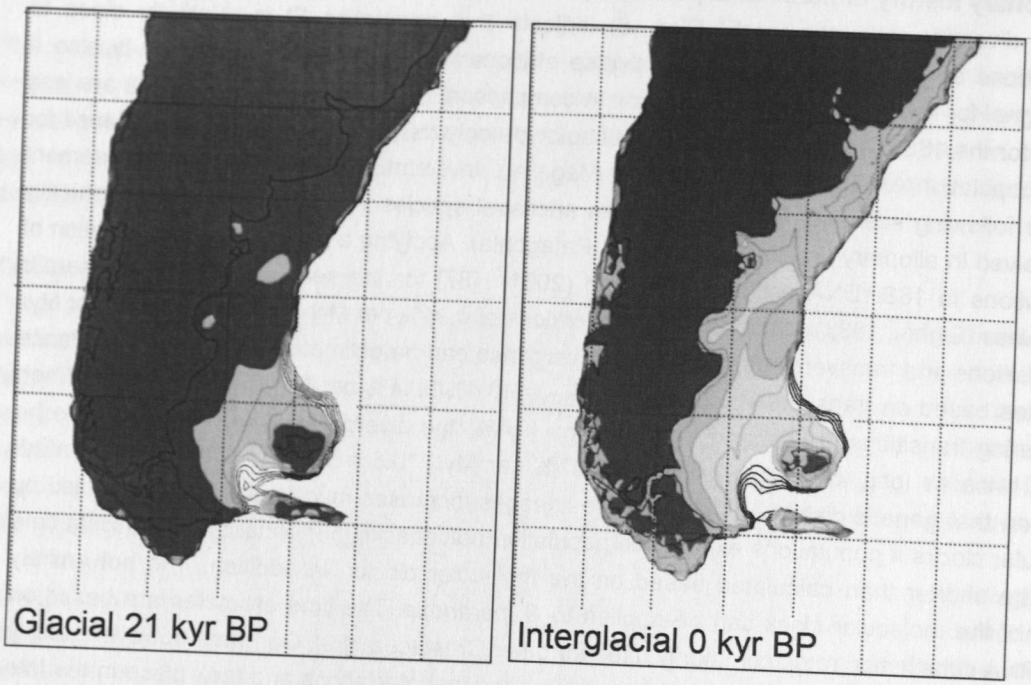


Figure 7: Terrestrial (black) and marine habitats (grey, 50 m contours) available at last glacial maximum (approximately 21 kyr BP, left panel) and at the present interglacial (right panel). During last glacial maximum the sea level was considerably lower thus bringing shallow water habitats of the Patagonian Coast and the Falkland Islands in close proximity. Adapted from [37].

The population around the Falkland Islands was not affected in the same way by glacial conditions (non-significant Tajima's *D*, mismatch analysis with much less support for unimodal distribution) and has a much higher genetic diversity, however, we cannot reject evidence of recent population expansion based on Fu's *Fs* and the mismatch analysis. This could however be the result of a sampling bias or other reasons for population expansion such as climatic ameliorizations.

Reliability and systematic bias in demography and diversity estimates

At first sight it is surprising that the fast evolving microsatellites display a higher expected heterozygosity for populations AO and PA compared to FI because the former are expected to be subject of a recent population expansion that is expected to bias diversity measures towards lower values. Because microsatellites were originally developed for PA [90], null alleles due to cross-species amplification can be expected to be the source of reduced H_o in FI, which is evolutionarily furthest from PA. In contrast, 16S rDNA is not similarly affected and shows the expected pattern of genetically depauperate populations in regions of recent population expansion (only PA sampled).

Reliability and systematic bias in differentiation estimates

The equilibrium F_{ST} estimate for totally isolated populations based on microsatellites can reach the maximum value of $F_{ST}=1$ only theoretically. Due to the high mutation rate of microsatellites [79, 91, 92] and often a restricted allelic spectrum [93- but see 95], the intrapopulation variability

is generally very high in particular after after a long time of independent evolution of large populations. However, F_{ST} and R_{ST} estimates are based on comparing total variance to the variance within subpopulations. Therefore, if the latter is relatively large compared to the total variance, F_{ST} will take on numerically small values, even though no allele may be shared between populations. Hedrick (2005) and Meirmans (2006) provided standardizations to F_{ST} estimates, which make F_{ST} estimates from different microsatellites comparable with a realized range of 0 to 1 even for highly variable markers. Applying Meirman's standardization approach for pairwise F_{ST} calculates in this study, differentiation between PA and FI is 0.86, between AO and FI 0.91 and among the Magellan Strait populations PA and AO 0.063 and thus about three times larger than without this correction. These values underpin that both populations are almost fixed for different alleles at the seven loci investigated. When removing locus 39, which is biased for population FI, the standardized values do not change, however, the not-standardized F_{ST} values are almost twice as high. The results of this study point out the importance of the recently introduced standardization approach [55, 56] in order to allow easier comparison and interpretation of the data.

Taxonomic and conservation status of the newly delimited species

The genetic data strongly suggest that nominal *Serolis paradoxa* (Fabricius, 1775) consists of two reproductively isolated species one of which occurs in Patagonia while the other is presumably confined to shallow waters around the Falkland Islands. As the type was originally described by Fabricius as *Oniscus paradoxum* Fabricius, 1775 from the Falkland Island the species from Patagonia is in need of formal description and a scientific name.

The occurrence of cryptic species has important implications for the conservation of biodiversity in general [96]. If a cryptic species is not recognized, unique and endangered local faunas cannot be efficiently protected. However, the estimates of effective population size for both species contained inside nominal *Serolis paradoxa* implies that both are highly abundant and neither needs to be considered endangered.

Conclusions

In summary, our data proves low differentiation among populations within the Strait of Magellan but the absence of effective gene flow among populations from the Strait of Magellan and the Falkland Islands. In fact, specimens from both regions belong to two cryptic lineages that probably diverged in the late Pliocene or early Pleistocene and may already represent reproductively isolated species. The 16S rDNA data supports that while the Falkland Island reveals a genetically rich population, the Strait of Magellan population is genetically depauperate and may represent the offspring of founder populations that recolonized the Strait of Magellan after the last glacial maximum approximately 9-14 Kyr after deglaciation of the habitat and sea level rise.

While the fauna of the Falkland Islands has often been accepted to reveal most marine species generally also present in Patagonia our results indicate that shallow water species with low mobility may in fact turn out to be strongly differentiated populations of one species or even reproductively isolated species.

Author's contributions

CH and FL designed the study and sampled the specimens. FL and AK carried out the molecular genetic analyses. FL performed the computational analyses and interpretation of population genetic data and drafted the manuscript. CH coordinated the project and made substantial contributions to the interpretation of the data and the drafting of the manuscript. JWW participated in the coordination of the study and contributed to drafting and revising the manuscript. AK helped drafting the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

We thank Erika Mutschke and Carlos Rios (Universidad de Magallanes, Punta Arenas) for providing us with material from the 2nd Cruce estrecho, 2003. This work was supported by a DFG grant HE-3391/3 to CH, NSF grant OPP-0132032 to H.W. Detrich III, and a DAAD scholarship to FL and AK. This is publication number 22 from the ICEFISH Cruise of 2004.

Appendix

Table 2: Total number of specimens scored for each locus (N_S), number of different alleles (N_A), inbreeding coefficient (F_{IS}), observed heterozygosity (H_O) and expected heterozygosity for the seven microsatellites and three populations of *Serolis paradoxa*.

	PA	AO	FI	Mean N_A/locus
Spa04				
N_S	32	33	23	
N_A	5	5	4	4.667
F_{IS}	0.316**	-0.075	0.539*	
H_O	0.769	0.485	0.130	
H_E	0.682	0.452	0.279	
Spa12				
N_S	35	33	23	
N_A	6	4	3	4.333
F_{IS}	0.172	-0.028	-0.012	
H_O	0.257	0.121	0.087	
H_E	0.310	0.118	0.086	
Spa34				
N_S	35	33	23	
N_A	20	12	13	15.000
F_{IS}	0.040	0.102	0.206	
H_O	0.886	0.667	0.696	
H_E	0.922	0.741	0.872	
Spa35				
N_S	34	30	17	
N_A	7	9	5	7.000
F_{IS}	-0.015	0.138	0.286	
H_O	0.676	0.633	0.353	
H_E	0.667	0.733	0.490	
Spa39				
N_S	35	27	13	
N_A	17	16	6	13.000
F_{IS}	0.322**	0.299**	1.000**	
H_O	0.629	0.630	0.000	
H_E	0.923	0.893	0.788	
Spa42				
N_S	35	33	23	
N_A	14	14	11	13.000
F_{IS}	0.034	0.119	0.069	
H_O	0.886	0.788	0.826	
H_E	0.907	0.892	0.886	
Spa43				
N_S	35	33	22	
N_A	2	4	5	3.667
F_{IS}	-0.033	0.230	-0.085	
H_O	0.40	0.364	0.636	
H_E	0.388	0.470	0.588	
Mean N_A per location	10.14	9.14	6.71	8.667
Mean H_O	0.600	0.527	0.390	
Mean H_E	0.686	0.614	0.570	
F_{IS}	0.110**	0.094**	0.181**	

Table 3: Allele frequencies at seven microsatellite loci for three *Serolis paradoxa* populations (PA=Punta Arenas, AO=Atlantic opening of the Strait of Magellan, FI=Falkland Islands). Private alleles are shown in bold.

Locus	Allele	Location		
		PA	AO	FI
Spa04				
	126	0.219	-	-
	128	-	-	0.063
	130	0.203	0.157	0.854
	132	0.063	0.057	0.042
	134	0.484	0.714	-
	136	-	0.057	0.042
	138	-	0.014	-
	140	0.031	-	-
Spa12				
	186	-	-	0.021
	188	-	0.015	0.958
	190	0.014	0.030	0.021
	192	0.071	-	-
	194	0.829	0.939	-
	196	0.043	0.015	-
	198	0.029	-	-
	202	0.014	-	-
Spa34				
	145	0.043	0.076	-
	147	-	-	0.042
	149	0.043	-	-
	151	0.014	-	-
	153	0.014	-	0.021
	155	-	-	0.021
	157	0.029	0.061	0.063
	159	0.057	0.030	0.292
	161	0.071	0.030	0.104
	163	0.043	0.015	0.042
	165	0.086	0.030	0.042
	167	0.014	0.015	0.083
	169	0.043	-	-
	171	0.157	0.061	0.042
	173	0.186	0.485	-
	175	0.057	0.121	-
	177	0.029	0.061	0.063
	179	0.014	-	0.146
	181	0.029	-	-
	183	-	0.015	0.021
	185	0.043	-	-
	187	0.014	-	0.021
	189	0.014	-	-
Spa35				
	227	-	0.017	-
	231	-	0.033	-
	235	0.485	0.467	-
	237	0.074	0.117	-
	239	0.015	0.050	0.118
	241	0.309	0.183	-
	243	0.074	0.100	0.029
	245	0.029	-	0.706
	247	-	0.017	0.088
	249	0.015	-	-
	251	-	-	0.059

	255	-	0.017	-
Spa39				
	172	0.071	-	-
	180	0.014	0.018	-
	184	-	0.018	-
	188	0.157	0.232	-
	192	0.043	0.161	0.077
	196	0.114	0.054	-
	200	0.071	0.089	-
	202	-	-	0.231
	204	0.071	0.071	0.154
	206	-	0.054	0.385
	208	0.071	-	-
	210	0.014	-	-
	212	-	0.018	0.077
	216	0.129	0.036	-
	220	0.043	0.107	-
	224	0.057	0.054	-
	228	0.086	0.018	-
	232	0.014	-	-
	234	-	-	0.077
	236	0.014	0.036	-
	240	0.014	0.018	-
	244	0.014	-	-
	256	-	0.018	-
Spa42				
	150	-	-	0.044
	154	-	-	0.022
	158	-	-	0.065
	160	-	-	0.109
	162	0.014	0.014	0.065
	164	-	-	0.087
	166	0.057	0.071	0.152
	168	0.086	0.086	0.217
	169	-	-	0.022
	170	0.143	0.143	0.174
	172	0.157	0.143	0.044
	174	0.114	0.214	-
	176	0.114	0.114	-
	178	0.129	0.043	-
	180	0.057	0.071	-
	182	0.014	0.029	-
	184	0.043	-	-
	186	-	0.014	-
	188	0.029	0.014	-
	190	-	0.014	-
	192	0.014	0.029	-
	206	0.029	-	-
Spa43				
	176	-	0.014	-
	177	0.257	0.300	0.022
	178	0.743	0.671	0.239
	182	-	0.014	0.109
	183	-	-	0.609
	184	-	-	0.022

Table 6: Statistical tests for significant heterozygosity (H) excess or deficiency in three populations of *Serolis paradoxa* (PA, AO, FI) assuming three different mutation models (IAM, TPM, SMM). P-values of the Sign Test and Standardized Differences Test and one-tailed probability for heterozygosity deficiency is based on a 1000 permutations. Significant P-values are printed in bold. The TPM was adjusted to allow for 80% mutations according to a SMM and 20% to an IAM model.

		Sign Test				Wilcoxon Test
Population	Model	Expected no of loci with H excess	Observed no. of loci with H excess	observed no of loci with H deficiency	P	P (one tailed for H deficiency)
PA	IAM	4.00	6	1	0.1212	0.9453
	TPM	4.04	4	3	0.6306	0.7109
	SMM	4.03	3	4	0.3360	0.2891
AO	IAM	4.17	3	4	0.2979	0.2343
	TPM	4.10	1	6	**0.0226**	**0.0117**
	SMM	4.09	0	7	**0.0021**	**0.0039**
FI	IAM	4.06	4	3	0.6244	0.3438
	TPM	4.14	2	5	0.1049	0.1484
	SMM	4.16	2	5	0.1024	**0.0195**

References

1. Knox GA, Lowry JK: A comparison between the benthos of the Southern Ocean and the North Polar Ocean with special reference of the Amphipoda and the Polychaeta. In *Polar Oceans.* Edited by Dunbar MJ. Calgary: Arctic Institute of North America; 1977:423-462.
2. San Martin AM: Biodiversity, zoogeography and ecology of polychaetes from the Magellan region and adjacent areas. *Berichte zur Polarforschung* 2005, 505:1-112.
3. McDowall RM: Falkland Islands biogeography: converging trajectories in the Southern Ocean. *Journal of Biogeography* 2005, 32:49-62.
4. Adie RJ: The position of the Falkland Islands in a reconstruction of Gondwanaland. *Geological Magazine* 89:401-410.
5. Mitchell C, Taylor GK, Cox KG, Shaw J: Are the Falkland Islands a rotated microplate?. *Nature* 1986, 319:131-134.
6. Marshall JEA: The Falkland Islands: A key element in Gondwana palaeogeography. *Tectonics* 1994, 13:499-514.
7. Knowlton N: Sibling species in the sea. *Annual Review of Ecology and Systematics* 1993, 24:189-216.
8. Palumbi SR: Genetic divergence, reproductive isolation and marine speciation. *Annual Review of Ecology and Systematics* 1994, 25:547-572.
9. Bohonak AJ: Dispersal, gene flow, and population structure. *Quarterly Review of Biology* 1999, 74:21-43.
10. Knowlton N: Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia* 2000, 420:73.
11. Brandt A: Colonization of the Antarctic shelf by the Isopoda (Crustacea, Malacostraca). *Berichte zur Polarforschung* 1991, 98:1-240.
12. Brandt A, Linse K, Mühlenhardt-Siegel U: Biogeography of Crustacea and Mollusca of the Subantarctic and Antarctic regions. *Scientia Marina* 1999, 63:383-389.
13. Linse K: New records of shelled marine molluscs at Bouvet Island and preliminary assessment of their biogeographic affinities. *Polar Biology* 2006, 29:120-127.
14. Allcock AL, Brierley AS, Thorpe JP, Rodhouse PG: Restricted gene flow and evolutionary divergence between geographically separated populations of the Antarctic octopus *Paraledone turqueti*. *Marine Biology* 1997, 129:97-102.
15. Held C: Phylogeny and biogeography of serolid isopods (Crustacea, Isopoda, Serolidae) and the use of ribosomal expansion segments in molecular systematics. *Molecular Phylogenetics and Evolution* 2000, 15:165-178.
16. Page TJ, Linse K: More evidence of speciation and dispersal across the Antarctic Polar Front through molecular systematics of Southern Ocean *Limatula* (Bivalvia: Limidae). *Polar Biology* 2002, 25:818-826.
17. Held C: Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda). In *Antarctic biology in a global context.* Edited by Huiskes AH, Gieskes WW, Rozema J, Schorno RM, van der Vies SM, Wolff WJ. Leiden: Backhuys Publishers; 2003:135-139.
18. Held C, Wägele J-W: Cryptic speciation in the giant Antarctic isopod *Glyptonotus antarcticus* (Isopoda: Valvifera: Chaetiliidae). *Scientia Marina* 2005, 69:175-181.

19. Raupach MJ, Wägele JW: Distinguishing cryptic species in Antarctic Asellota (Crustacea : Isopoda) - a preliminary study of mitochondrial DNA in Acanthaspidia drygalskii. *Antarctic Science* 2006, 18:191-198.
20. Linse K, Cope T, Lörz A-N, Sands C: Is the Scotia Sea a centre of Antarctic marine diversification? Some evidence of cryptic speciation in the circum-Antarctic bivalve *Lissarca notorcadensis* (Arcoidea: Philobryidae). *Polar Biology* 2007, 30:1059-1068.
21. Wares JP, Daley S, Wetzer R, Toonen RJ: An evaluation of cryptic lineages of *Idotea balthica* (Isopoda: Idoteidae): Morphology and microsatellites. *Journal of Crustacean Biology* 2007, 27:643-648.
22. Wilson NG, Hunter RL, Lockhart SJ, Halanych KM: Multiple lineages and absence of panmixia in the "circumpolar" crinoid *Promachocrinus kerguelensis* from the Atlantic sector of Antarctica. *Marine Biology* 2007, 152:895-904.
23. Leese F, Kop A, Agrawal S, Held C: Isolation and characterization of microsatellite markers from the marine serolid isopods *Serolis paradoxa* and *Septemserolis septemcarinata* (Crustacea: Peracarida). *Molecular Ecology Resources* 2008, 8:818-821.
24. Lancellotti DA, Vasquez JA: Zoogeografía de macroinvertebrados bentónicos de la costa de Chile: contribución para la conservación marina. *Revista Chilena de Historia Natural* 2002, 73:99-129.
25. Wägele J: Notes on Antarctic and South American Serolidae (Crustacea, Isopoda) with remarks on the phylogenetic biogeography and a description of new genera. *Zoologische Jahrbücher Abteilung für Systematik Ökologie und Geographie der Tiere* 1994, 121:3-69.
26. Gappa L, Sueiro M: The subtidal macrobenthic assemblages of Bahia San Sebastian (Tierra del Fuego, Argentina). *Polar Biology* 2007, 30:679-687.
27. Neigel JE: Population genetics and demography of marine species. In *Marine biodiversity: Patterns and processes.* Edited by Ormond RF, Gage JD, Angel MV. Cambridge: Cambridge University Press; 1997:274-292.
28. Hewitt GM: The genetic legacy of the quaternary ice ages. *Nature* 2000, 405 :907.
29. Anderson JB, Shipp SS, Lowe AL, Weller JS, Mosola AB: The Antarctic ice sheet during the Last glacial maximum and its subsequent retreat history: a review. *Quaternary Science Reviews* 2002, 21:49-70.
30. Huybrechts P: Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quaternary Science Reviews* 2002, 21:203-231.
31. Ingolfsson O: Quaternary glacial and climate history of Antarctica. In *Quaternary Glaciations - Extent and Chronology, Part III.* Edited by Ehlers J, Gibbard PL. Elsevier; 2004:3-43.
32. Polanski J: The maximum glaciation in the Argentine Cordillera. *Geological Society of America, Inc. Special Paper* 1965, 84:453-472.
33. Flint RF, Fidalgo F: Glacial drift in the eastern Argentine Andes between latitude 41° 10'S and latitude 43° 10'S. *Geological Society of America Bulletin* 1969, 80:1043-1052.
34. Hulton N, Sugden D, Payne A, Clapperton C: Glacier modeling and the climate of Patagonia during the last glacial maximum. *Quaternary Research* 1994, 42:1-19.
35. Clapperton CM, Sugden DE, Kaufman DS, McCulloch RD: The last glaciation in central Magellan Strait, southernost Chile. *Quaternary Research* 1995, 44:133-148.

36. McCulloch RD, Bentley MJ, Purves RS, Hulton NRJ, Sugden DE, Clapperton CM: Climatic inferences from glacial and palaeoecological evidence at the last glacial termination, southern South America. *Journal of Quaternary Sciences* 2000, 15:409-417.
37. Rostami K, Peltier WR, Mangini A: Quaternary marine terraces, sea-level changes and uplift history of Patagonia, Argentina: comparisons with predictions of the ICE-4G (VM2) model of the global process of glacial isostatic adjustment. *Quaternary Science Reviews* 2000, 19:1495-1525.
38. Hulton NRJ, Purves RS, McCulloch RD, Sugden DE, Bentley MJ: The Last Glacial Maximum and deglaciation in southern South America. *Quaternary Science Reviews* 2002, 21:233-241.
39. Clapperton CM: Quaternary glaciations in the Southern Ocean and the Antarctic Peninsula area. *Quaterary Science Reviews* 1990, 9:229-252.
40. Pompanon F, Bonin A, Bellemain E, Taberlet P: Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 2005, 6:847-859.
41. Hoffman JI, Amos W: Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 2005, 14:599-612.
42. Palumbi SR, Martin A, Romano S, Mcmillan WO, Stice L, Grabowski G: *The simple fool's guide to PCR. A collection of PCR protocols, version 2* 1991,
43. van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P: MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 2004, 4:535-538.
44. McKelvey KS, Schwartz MS: DROPOUT: A program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework.. *Molecular Ecology Notes* 2005, 5:716-718.
45. Park SDE: *Trypanotolerance in west African cattle and the population genetic effects of selection.* University of Dublin; 2001.
46. Glaubitz JC: CONVERT: A user-friendly program to reformat diploid genotypc data for commonly used population genetic software packages. *Molecular Ecology Notes* 2004, 4:309-310.
47. Landry PA, Koskinen MT, Primmer CR: Deriving evolutionary relationships among populations using microsatellites and (deltamu)(2): all loci are equal, but some are more equal than others. *Genetics* 2002, 161:1339-47.
48. Slatkin M: A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 1995, 139: 457-462.
49. Rousset F: Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* 1996, 142:1357-1362.
50. Rousset F: GENEPOP '007: a complete reimplementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Notes* 2007, online first:
51. Excoffier L, Laval G, Schneider S: Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 2005, 1:47-50.
52. Weir BS, Cockerham CC: Estimating F-statistics for the analysis of population structure. *Evolution* 1984, 38:1358-1370.
53. Hedrick P: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 1999, 53:313-318.
54. Whitlock MC, McCauley DE: Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1). *Heredity* 1999, 82 (Pt 2):117-25.

55. Hedrick P: A standardized genetic differentiation measure. *Evolution* 2005, 59:1633-1638.
56. Meirmans PG: Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 2006, 60:2399-402.
57. Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000, 155:945-959.
58. Falush D, Stephens M, Pritchard JK: Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003, 164:1567-1587.
59. Jakobsson M, Rosenberg NA: CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 2007, 23:1801-1806.
60. Rosenberg NA: Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* 2004, 4:137-138.
61. Hill WG: Estimation of effective population size from data on linkage disequilibrium. *Genetical Research* 1981, 38:209-216.
62. Waples RS: A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics* 2006, 7:167-184.
63. Waples RS, Do C: LDNe: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Notes* 2007, online first:
64. Piry S, Luikart G, Cornuet J-M: BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* 1999, 90:502-503.
65. Nei M, Maruyama T, Chakraborty R: The bottleneck effect and genetic variability in populations. *Evolution* 1975, 29:1-10.
66. Cornuet JM, Luikart G: Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 1996, 144:2001-14.
67. Luikart G, Allendorf FW, Cornuet JM, Sherwin WB: Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* 1998, 89:238-247.
68. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999, 41:95.
69. Gutell RR: Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Research* 1993, 21:3051-3054.
70. Tamura K, Dudley J, Nei M, Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.. *Molecular Biology and Evolution* 2007, 24:1596-9.
71. Nei M: *Molecular evolutionary genetics*. New York: Columbia University Press; 1987.
72. Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989, 123:585-595.
73. Fu Y-X: Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 1997, 147:915-923.
74. Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 1987, 4:406-426.
75. Swofford DL: *PAUP*. Phylogenetic analysis using parsimony (* and other methods)* 1998,
76. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R: Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 2007, 8:460.
77. Clement M, Posada D, Crandall KA: TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 2000, 9:1657-1659.

78. Rice WR: Analyzing tables of statistical tests. *Evolution* 1989, 43:223-225.
79. Ellegren H: Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics* 2004, 5:435-445.
80. Buschiazzo E, Gemmell NJ: The rise, fall and renaissance of microsatellites in eukaryotic genomes.. *Bioessays* 2006, 28:1040-50.
81. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P: Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene-Sequences and A Compilation of Conserved Polymerase Chain- Reaction Primers. *Annals of the Entomological Society of America* 1994, 87:651.
82. Ibrahim KM, Nichols RA, Hewitt GM: Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 1996, 77:282-291.
83. Schubart CD, Diesel R, Hedges SB: Rapid evolution to terrestrial life in Jamaican crabs. *Nature* 1998, 393:363-365.
84. Coyne JA, and Orr HA: *Speciation.* Sinauer Associates, Inc.; 2004.
85. Moore WS: Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 1995, 49:718-726.
86. Ballard JW, Whitlock MC: The incomplete natural history of mitochondria. *Molecular Ecology* 2004, 13:729-744.
87. Held C: No evidence for slow-down of molecular substitution rates at subzero temperatures in Antarctic serolid isopods (Crustacea, Isopoda, Serolidae). *Polar Biology* 2001, 24:497-501.
88. Helmuth B, Veit RR, Holberton R: Long-distance dispersal of a subantarctic brooding bivalve (*Gaimardia trapesina*) by kelp-rafting. *Marine Biology* 1994, 120:421-426.
89. Thiel M, Gutow L: The ecology of rafting in the marine environment - I - The floating substrata. In *Oceanography And Marine Biology: An Annual Review, Vol 42.* 2005:181-263.
90. Leese F, Held C: Identification and characterization of microsatellites from the Antarctic isopod *Ceratoserolis trilobitoides*: nuclear evidence for cryptic species. *Conservation Genetics* 2008, online early:
91. Weber JL, Wong C: Mutation of human short tandem repeats. *Human Molecular Genetics* 1993, 2:1123-1128.
92. Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF: The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster. Molecular Biology and Evolution* 1998, 15:1751-1760.
93. Nauta MJ, Weissing FJ: Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* 1996, 143:1021-1032.
94. Goldstein DB, Pollock DD: Launching microsatellites: a review of mutation processes and methods of phylogenetic interference.. *Journal of Heredity* 1997, 88:335-42.
95. Primmer CR, Ellegren H, Saino N, Moller AP: Directional evolution in germline microsatellite mutations. *Nature Genetics* 1996, 13:391-393.
96. Bickford D, Lohman DJ, Sohdi NS, Ng PK, Meier R, Winker K, Ingram KK, Das I: Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 2007, 22:148-155.

Chapter 10

Concluding Discussion

The results of this thesis provide significant advances and insights in two fields of research. First, a laboratory and software-based workflow is introduced to facilitate the isolation of sufficiently many and appropriate microsatellite markers from unknown genomes. Second, data obtained using these fast evolving microsatellite markers and a mitochondrial DNA marker are analysed for three serolid isopods from the Antarctic, the Subantarctic and the Magellan Region and discussed in the context of the species' recent evolutionary history. Some of the key aspects of our results are highlighted in this concluding chapter.

Isolation of microsatellites from unknown genomes

The great information content of microsatellite markers for population genetic studies is indisputable (Jarne and Lagoda 1996, Goldstein and Pollock 1997, Sunnucks 2000, Oliveira et al. 2006, Selkoe and Toonen 2006). Their usefulness for more exploratory studies, i.e. for studies investigating genetic structure and diversity in non-model organisms, is still restricted by an often unpredictably difficult isolation procedure. The laboratory strategies outlined in this thesis, using the novel Reporter Genome Protocol (Nolte et al. 2005) and the software pipeline based on the the *STADEN* package (Staden 1996) with the microsatellite search tool PHOBOS, were shown to allow the isolation and setup of a highly informative set of microsatellite markers within a short period of time. Other protocols, namely the radioactive screening of enriched and non-enriched libraries and the PIMA protocol did not succeed in isolating a sufficient number of appropriate microsatellites. The Reporter Genome Protocol was also proven to be successful for taxa other than Antarctic isopods. Altogether, the fast and economical workflow is expected to greatly facilitate marker isolation and design in subsequent studies on Antarctic and non-Antarctic taxa likewise.

Utility of the different molecular markers

While microsatellites provide a very detailed insight to study the recent evolutionary history, their resolution decreases with increasing coalescent time as a result of their high mutation rates (Fig. 1). The mitochondrial DNA allowed investigating these processes. One other marker system often used to address microevolutionary questions are AFLPs (Amplified Fragment Length Polymorphisms, Vos et al. 1995). Although the AFLP method is based on restriction fragment patterns of the total genome ("whole genome approach"), it has two major drawbacks: it is a dominant marker system and the results highly depend on the quality of DNA (Bensch and Akesson 2005). This limits its potential to address questions when working with Antarctic samples since samples from different cruises are generally of different age and preserved in different ways. Thus consistent differences are likely to be caused by preservation biases rather than by genomic similarity. Therefore, microsatellite loci, which can still be amplified from degraded material (but see Pompanon et al. 2005 for other biases), are the marker system of choice for studying Antarctic population genetics. To resolve events in the late Pleistocene between populations that have been isolated for more than 20 Kyr BP, the application of a nuclear, co-dominant allozymes might provide further resolution as mutation rates are one to several orders of magnitude lower.

Concerning the statistical analysis of the data, the results of this thesis show the importance of using standardized differentiation measures since standardized and non-standardized F_{ST} values may differ by more than one order of magnitude simply due to marker characteristics (Hedrick 2005, Meirmans 2006, Fig. 1). Without standardization, one may incorrectly infer low levels of

population differentiation, e.g. indicated by $F_{ST} < 0.05$ (Wright 1978), although in reality no alleles are shared between the populations investigated.

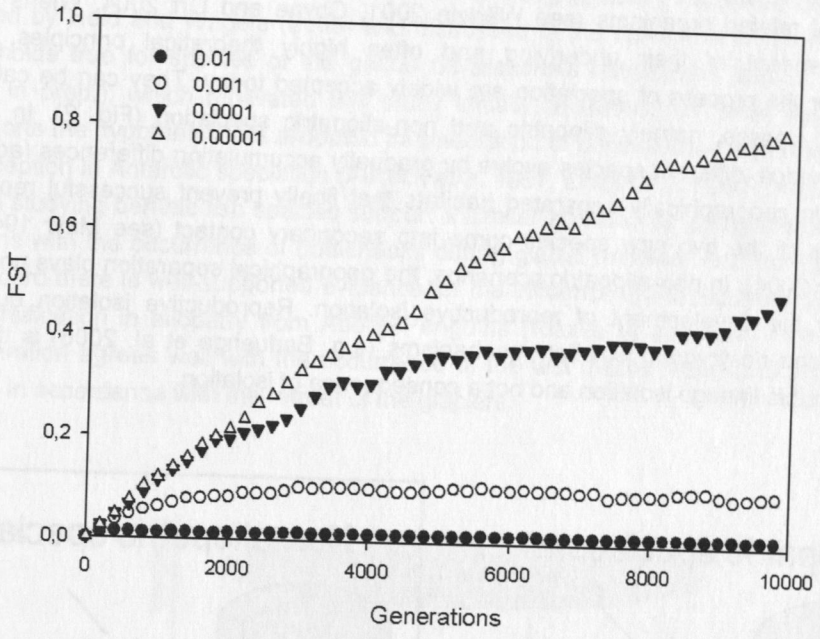


Figure 1: Equilibrium values of F_{ST} (Weir and Cockerham 1984) and period of time needed to approach equilibrium for a simulated data set (EASYPOP version 2.01, Balloux 2001) of two large populations (effective size N_e=2000) exchanging no migrants (m=0). Data points represent the average of 5 independent simulation runs of a microsatellite data set of n=10 loci with a limited number of possible allelic states (N_A=30) with different mutation rates (μ=1x10^{-2} to 1x10^{-5}) applying a two-phase mutation model (0.9 SMM, 0.1 IAM). The data set with the average mutation rate of 10^{-5} reaches equilibrium latest (>10,000 generations) and has highest differentiation values. The locus with mutation rate of 10^{-2} reaches equilibrium fast (400 generations) and maximum differentiation value for F_{ST} is low (0.013).

In this thesis we applied highly variable microsatellites and a mitochondrial DNA marker in three population genetic studies on three brooding serolid isopods from different regions. A common feature of the studied species is the strong and significant geographical partitioning of genetic variation. However, the degree of this partitioning was found to deviate significantly from expectations: a strong differentiation between two population groups of *Ceratoserolis* n. sp. 1 from the tip of the Antarctic Peninsula and the eastern Weddell Sea was detected. However, within the two geographical regions no genetic subdivision of populations was observed even over distances of more than 500 km on the shelf. Even more unexpected came the result that the degree of subdivision between populations of *Septemserolis septemcarinata* is less prominent than expected based on their highly disjunct habitats. The most plausible explanation for this genetic similarity is a very low but ongoing gene flow, which could for instance be maintained by passive drift of specimens on macroalgal mats with the Antarctic Circumpolar Current. For populations of the third species, *Serolis paradoxa,* populations from the Strait of Magellan and the Falkland Islands turned out to maintain distinct gene pools and must be regarded as different species. In view of the strong ocean currents between the regions and the comparatively shallow water depths on the South American shelf, this result is unexpected.

Speciation on the Antarctic shelf

Several species concepts exist and they describe, depending on their perspective, different assemblages of related organisms (see Wägele 2001, Coyne and Orr 2004, Wiens 2004 for review). Independent of their underlying and often highly theoretical principles, different mechanisms for the process of speciation are widely accepted today. They can be categorized into two major classes, namely allopatric and non-allopatric speciation (Fig. 2). In allopatric speciation scenarios, different species evolve by gradually accumulating differences (adaptive or non-adaptive) in geographically separated habitats that finally prevent successful reproduction when members of the two new species come into secondary contact (see Mayr 1942, 1963, Coyne and Orr 2004). In non-allopatric scenarios, the geographical separation plays no or only a lesser role for the development of reproductive isolation. Reproductive isolation due to fast evolving pre- and postzygotic isolation mechanisms (e.g. Barluenga et al. 2006) is the driving force of irreversible lineage isolation and not a consequence of isolation.

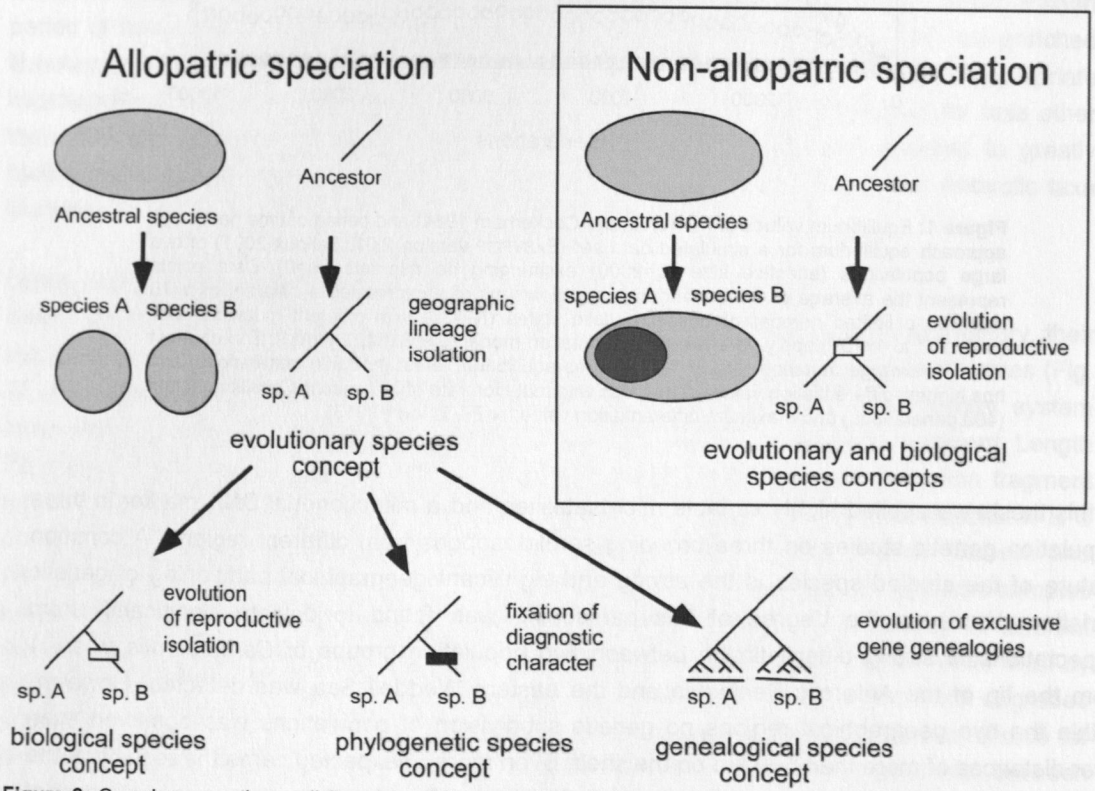


Figure 2: Overview over the explicit and implicit assumption of different species concepts under allopatric and non-allopatric speciation scenarios. Independently of their detailed assumptions the two different speciation scenarios differ in their assumptions on the ultimate cause of speciation: geographic lineage isolation for the allopatric and the evolution of reproductive isolation for the non-allopatric speciation model (Wiens 2004).

Models of allopatric speciation for the Antarctic benthos gained support with growing information on the influence of glacial advances on lineage sorting in the Northern Hemisphere during the Pleistocene epoch (2 Myr BP to 10 Kyr BP), during which peridocially advancing,

large-scale glaciers shaped the genetic structure of many species (Roberts 1998, Hewitt 2000) (Fig. 3). First indications that glaciations could have imposed a major mechanism leading to an allopatric speciations in the Antarctic emerged in recent studies: For the benthic isopod *Glyptonotus antarcticus* Eights, 1852 the existence of geographically separated cryptic species was revealed by Held and Wägele (2005) and discussed in the context of allopatric speciation. The same holds true for species of the genus *Ceratoserolis trilobitoides* (Eights, 1833) (Held 2003, Held in prep.), which motivated this study initially. Research on other benthic species further supports the hypothesis that allopatric fragmentation of populations may be the rule rather than the exception in Antarctic speciation (Allcock et a. 1997, Linse et al. 2007). Results of Janko et al. (2007) studying benthic fish species support a similar scenario for geographical separation of populations with the occurrence of bottlenecks during glacial maxima. Also from the Antarctic terrestrial record there is well-supported evidence for the influence of the recurrent glaciations on lineage diversification in allopatry from Adelie Penguins (Ritchie et al. 2004). In this example, lineage separation agrees well with the occurrence of the last glacial maximum, and population expansion is in accordance with the retreat of the glaciers.

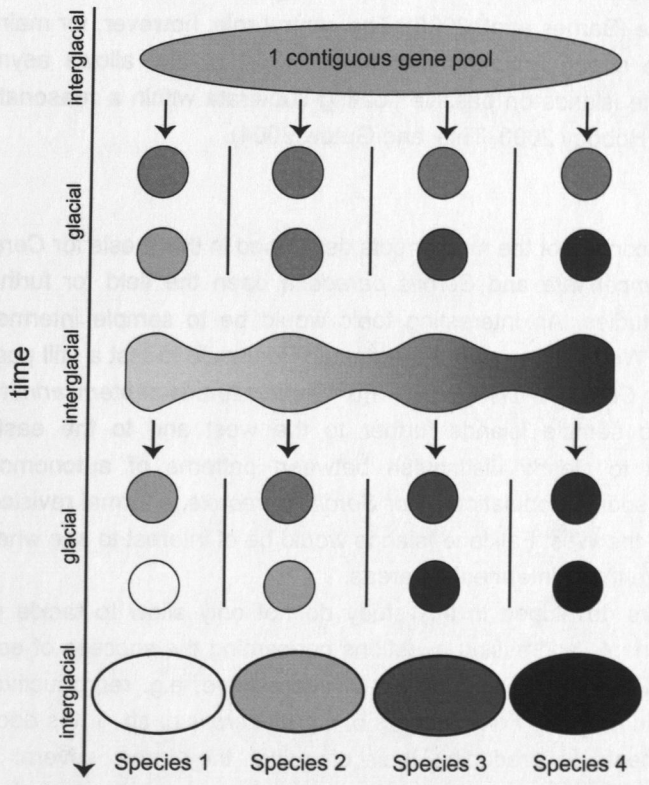


Figure 3: Schematic overview over population genetic differentiation processes in the context of the glacial refugia theory. A once contiguous gene pool of a broadly distributed species (top) is being distrupted during glacial maxima. Four small populations with their random genetic equipment are subject to random genetic drift (reciprocal to population size) during glacial maxima. This process may (after one, 2 as above, or n glacial maxima) lead to allopatric speciation.

Population genetic data from this thesis on the brooding benthic isopod *Ceratoserolis* n. sp. 1 from the High Antarctic shelf now provide for the first time highly resolved information on microevolutionary processes. The data support an allopatric speciation with lineage sorting in different refuge areas on the Antarctic shelf or the upper slope during glacial maxima (see Fig. 3 for schematic overview over processes). In addition, they provide evidence that two such glacial refuge areas have been located around the tip of the Antarctic Peninsula and on the partly ice-free shelf of the eastern Weddell Sea. These data are in best agreement with geophysical information (e.g. Andersen et al. 2002, Andersen pers. comm.).

The occurrence of a cryptic sister species of *Ceratoserolis* n. sp. 1 living in sympatry around the Antarctic Peninsula could be the result of earlier allopatric separation processes with overlapping habitats today (Held 2003).

Gene flow among highly disjunct Southern Ocean islands

The unexpected finding of shared gene pools among populations of *Septemserolis septemcarinata* from the highly disjunct islands supports that even for immobile brooding species a gene flow over long distances can be maintained at a low but significant level. In the context of increasing human impact in the Antarctic it can be expected that for several species, gene flow can artificially increase (Barnes et al. 2005). The central role, however, for maintaining gene flow must be seen in the strong Antarctic Circumpolar Current that allows asymmetric migration between highly remote islands on passive floating substrata within a reasonable period of time (Helmuth et al. 1994, Hobday 2000, Thiel and Gutow 2004).

Future research

The high information content of the marker sets developed in this thesis for *Ceratoserolis* n. sp. 1, *Septemserolis septemcarinata* and *Serolis paradoxa* open the field for further, more detailed population genetic studies. An interesting topic would be to sample intermediate shelf areas between the eastern Weddell Sea and the Antarctic Peninsula to test a still possible isolation-by-distance scenario for *Ceratoserolis* n. sp. 1. For *Septemserolis septemcarinata* a major point of interest would be to sample islands further to the west and to the east in the Antarctic Circumpolar Current to clearly distinguish between patterns of autonomous evolution and migration from other source populations. For *Serolis paradoxa*, a formal revision of the species is needed. Sampling of the West Falkland Islands would be of interest to see whether specimens of both lineages colonize these intermediate areas.

The new markers developed in this study do not only allow to tackle population genetic questions but also enable addressing questions concerning the success of ecological strategies of benthic species (Poulin and Feral 1996). Of interest are, e.g. reproductive strategies of the three brooding serolid isopods. For members of *Ceratoserolis* n. sp. 1 it is documented that they make large investments in producing their offspring: they grow several years before first reproduction, the eggs hatch and remain inside a brood pouch on the ventral side of the females (Luxmoore 1982, Wägele 1987). After reaching maturity, females reproduce at most every second year, and the fecundity is very low. Under these circumstances, females can be expected to be subject to intense selection to mate only with the best males, whereas males are less restricted and can mate with larger numbers of females. Due to the inaccessibility of the habitat and their extremely long life cycles, there are no direct observations of reproductive strategies of serolid isopods. The markers allow genotyping eggs or juvenile specimens in the marsupium to

test whether they are offspring of a single male or whether the eggs of one female were fertilized by more than one male. First results indicate that single clutches are fertilized by at least two males (Leese, Held, Frickenhaus, Spanier unpubl. results). It can be expected that Antarctic brooding taxa with long developmental times with high investments in terms of energy, apply different reproductive strategies compared to their relatives from warmer regions with shorter life cycles.

References

Allcock, A. L., A.S. Brierley, J.P. Thorpe, and P.G. Rodhouse. 1997. Restricted gene flow and evolutionary divergence between geographically separated populations of the Antarctic octopus *Paraledone turqueti*. Marine Biology **129**: 97-102.

Anderson, J. B., S.S. Shipp, A.L. Lowe, J.S. Weller, and A.B. Mosola. 2002. The Antarctic Ice Sheet during the last glacial maximum and its subsequent retreat history: a review. Quaternary Science Reviews **21**: 49-70.

Barnes, D. K. 2002. Biodiversity: invastions by marine life on plastic debris. Nature **416**: 808-809.

Bensch, S., and M. Akesson. 2005. Ten years of AFLP in ecology and evolution: why so few animals? Molecular Ecology **14**: 2899-2914.

Coyne, J. A., and H.A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland MA.

Futuyma, D. 1998. Evolutionary Biology. Sinauer Associates Inc., Sunderland MA.

Goldstein, D. B., and D.D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. Journal of Heredity **88**: 335-342.

Hedrick, P. 2005. A standardized genetic differentiation measure. Evolution **59**: 1633-1638.

Held, C. 2003. Molecular evidence for cryptic speciation within the widespread Antarctic crustacean *Ceratoserolis trilobitoides* (Crustacea, Isopoda), p. 135-139. *In* A.H. Huiskes, W.W. Gieskes, J. Rozema, R.M. Schorno, S.M. van der Vies, and W.J. Wolff [eds.], Antarctic biology in a global context. Backhuys Publishers, Leiden.

Held, C., and J.-W. Wägele. 2005. Cryptic speciation in the giant Antarctic isopod *Glyptonotus antarcticus* (Isopoda: Valvifera: Chaetiliidae). Scientia Marina **69**: 175-181.

Helmuth, B., R.R. Veit, and R. Holberton. 1994. Long-distance dispersal of a subantarctic brooding bivalve (Gaimardia trapesina) by kelp-rafting. Marine Biology **120**: 421-426.

Hobday, A. J. 2000. Persistence and transport of fauna on drifting kelp (*Macrocystis pyrifera* (L.) C. Agardh) rafts in the Southern California Bight. Journal of Experimental Marine Biology and Ecology **253**: 75-96.

Imbrie, J., A. Berger, E.A. Boyle, S.C. Clemens, A. Duffy, W.R. Howard, G. Kukla, J. Kutzbach, D.G. Martinson, A. McIntyre, A.C. Mix, B. Molfino, J.J. Morley, L.C. Peterson, N.G. Pisias, W.L. Prell, M.E. Raymo, N.J. Shackleton, and J.R. Toggweiler. 1993. On the structure and origin of major glaciation cycles. 2. The 100,000-year cycle. Paleoceanography **8**: 699-735.

Janko, K., G. Lecointre, A. Devries, A. Couloux, C. Cruaud, and C. Marshall. 2007. Did glacial advances during the Pleistocene influence differently the demographic histories of benthic and pelagic Antarctic shelf fishes? Inferences from intraspecific mitochondrial and nuclear DNA sequence diversity. BMC Evolutionary Biology **7**: 220.

Jarne, P., and P.J.L. Lagoda. 1996. Microsatellites, from molecules to populations and back. Trends in Ecology and Evoution **11**: 424-429.

Linse, K., T. Cope, A.-.N. Lörz, and C. Sands. 2007. Is the Scotia Sea a centre of Antarctic marine diversification? Some evidence of cryptic speciation in the circum-Antarctic bivalve *Lissarca notorcadensis* (Arcoidea: Philobryidae). Polar Biology **30**: 1059-1068.

Luxmoore, R. A. 1982. Moulting and growth in serolid isopods. Journal of Experimental Marine Biology Ecology **56**: 63-85.

Mayr, E. 1942. Systematics and the origin of species. Columbia University Press, New York.

Mayr, E. 1963. Animal species and evolution. Harvard University Press, Cambridge MA.

Meirmans, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution **60**: 2399-2402.

Nolte, A. W., K.C. Stemshorn, and D. Tautz. 2005. Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Molecular Ecology Notes **5**: 628-636.

Oliveira, E. J., J.G. Padua, M.I. Zucchi, R. Vencovsky, and M.L.C. Vieira. 2006. Origin, evolution and genome distribution of microsatellites. Genetics and Molecular Biology **29**: 294-307.

Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: causes, consequences and solutions. Nature Reviews Genetics **6**: 847-859.

Poulin, E., and J.P. Feral. 1996. Why are there so many species of brooding Antarctic echinoids? Evolution **50**: 820-830.

Ritchie, P. A., C.D. Millar, G.C. Gibb, C. Baroni, and D.M. Lambert. 2004. Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two Adelie penguin lineages in Antarctica. Molecular Biology and Evolution **21**: 240-248.

Roberts, N. 1998. The Holocene: an environmental history. Blackwell Publishing.

Selkoe, K., and R.J. Toonen. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters **9**: 615-629.

Staden, R. 1996. The Staden sequence analysis package. Molecular Biotechnology **5**: 233-241.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology and Evolution **15**: 199.

Thiel, M., and L. Gutow. 2005. The ecology of rafting in the marine environment - I - The floating substrata. Oceanography And Marine Biology: An Annual Review **42**: 181-263.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research **23**: 4470.

Wägele, J.-W. 1987. On the reproductive biology of *Ceratoserolis trilobitoides* (Crustacea: Isopoda): Latitudinal variation of fecundity and embryonic development. Polar Biology **7**: 11-24.

Wägele, J.-W. 2001. Grundlagen der phylogenetischen Systematik. Dr. Friedrich Pfeil, München.

Wiens, J. J. 2004. What is speciation and how should we study it? American Zoologist **163**: 914-923.

Wright, S. 1978. Evolution and the genetics of populations. University of Chicago Press, Chicago.

Summary

Recent molecular phylogenetic and phylogeographic studies revealed that the traditional concept of circumpolar distribution of most Antarctic benthic invertebrate species might be the exception rather than the rule. Yet, the underlying microevolutionary processes that ultimately lead to the disruption of a once contiguous gene pool in Antarctic macroinvertebrates remain poorly understood. For the great numbers of brooding invertebrates that inhabit the relative homogenous High Antarctic shelf, recurrent glaciations and lineage sorting in independent ice-free refugia resulting in allopatric speciation have been discussed for about two decades as a major mechanism for partitioning population genetic variation. For the fauna of the highly remote Antarctic and Subantarctic islands, the lack of free-swimming distributional stages in brooding species is seen as the major process driving genetic divergence and allopatric speciation. So far, however, no study has utilized highly informative, independent, nuclear microsatellite markers to study the genetic structure of brooding invertebrates in the context of recurrent glaciations or tested the species' realized dispersal capacity between remote islands.

This thesis focuses on two central aspects in order to contribute information to microevolutionary framework for the Southern Ocean benthos. First, I outline a straightforward strategy to obtain microsatellite markers much easier than with previous laboratory and software-based attempts and second, I investigate the genetic structure, diversity and demography of brooding serolid isopods (Crustacea: Peracarida) from (i) the High Antarctic shelf, (ii) three highly remote Antarctic and Subantarctic islands, and (iii) for comparison, from the Magellan Region using the novel microsatellites and mitochondrial DNA markers.

Three traditional radioactive and non-radioactive microsatellite isolation protocols did not yield any or any appropriate microsatellite. As a consequence, a novel approach to isolate and establish microsatellite markers from largely unknown genomes is outlined in the first part of the thesis. The strategy to obtain microsatellite markers with little effort in a relatively short time outlined is based on a novel microsatellite isolation method, the Reporter Genome Protocol (RGP) that utilizes whole genomic DNA of completely sequenced organisms as enrichment templates rather than using synthetic probes with only a limited number of repeat types that need to be specified *a priori*. The RGP facilitates the process of marker development. Besides its practical advantages in terms of time and costs, the RGP has the unique advantage that it isolates a much greater number of different microsatellites than other protocols. The importance for detecting and deselecting microsatellites with too complex structures based on dotplots are outlined and the success of the RGP is demonstrated over a broad range of crustacean and non-crustacean taxa from the Antarctic and lower latitudes. As a second major advantage in the process of isolating microsatellites, a newly developed, platform-independent open-source software pipeline is introduced that greatly facilitates the process of finding appropriate microsatellites in a flexible work-flow using a newly developed microsatellite search tool. The software pipeline is based on the STADEN PACKAGE software.

Three highly variable sets of microsatellites were developed, based on the genomic libraries obtained from the RGP and analysed in (i) *Ceratoserolis* n. sp. 1 from the High Antarctic shelf, (ii) *Septemserolis septemcarinata* from different Antarctic and Subantarctic Islands, and (iii) *Serolis paradoxa* from the Strait of Magellan and the Falkland Islands. The number of alleles per locus encompasses a wide range of variation and enables the detection of fine-scale differences in populations, even down to assignments of individual specimens to populations.

For *Ceratoserolis* n. sp. 1, populations from the tip of the Antarctic Peninsula and from the Eastern Weddell Sea were investigated using ten microsatellite markers and a fragment of the cytochrome oxidase subunit I gene (COI). Applying summary statistic approaches and Bayesian analyses of population differentiation, a strong geographical structure of genetic polymorphisms was found between regions on a larger geographical scale (Antarctic Peninsula, eastern Weddell Sea), indicating the lack of effective recent gene flow between them. On the other hand, no consistent genetic substructure was detected among populations within these regions indicating relatively high levels of genetic connectedness even over distances of in excess of 500 km. The divergence between Antarctic Peninsula and eastern Weddell Sea genotype clusters most likely developed prior to the Last Glacial Maximum (LGM) and provides strong direct support that members of *Ceratoserolis* n. sp. 1 have survived the LGM in two independent refugia, one at the tip of the Antarctic Peninsula and the other possibly on the shelf of the eastern Weddell Sea. Geophysical data indicate that these regions might have been ice-free during LGM. The COI data revealed significant signatures of a recent population expansion that can be estimated to have occurred most likely at the end of the LGM, indicating that the ice ages nevertheless imposed a strong reduction in effective population size on *Ceratoserolis* n. sp. 1. In conclusion, the results support the assumption that the benthic species were able to survive glacial maxima on the shelf or the slope of the Antarctic continent. Rapid lineage sorting owed to elevated levels of random genetic drift in small allopatric populations is regarded as the most parsimonious microevolutionary mechanism leading to two distinct gene pools in the different regions. It can be speculated that the two reciprocally monophyletic lineages of populations from both regions, presently separated by the Ronne-Filchner shelf, might further diverge and become two sister species in future.

Populations of *Septemserolis septemcarinata* from the remote islands South Georgia, Bouvet and Marion were found to be genetically strongly differentiated. However, the genetic distances were much lower than expected for distinct, reproductively isolated species, indicating that *S. septemcarinata* is capable of dispersing over enormous distances (>2500 km). Based on own observations and reports in literature from other isopods, it is most likely that rafting on macroalgae is responsible for the species capability of long-distance dispersal in the absence of free-swimming larval stages. The pattern of population diversity with lowest values found in populations around the old and large islands South Georgia to the West and much higher differentiation for the populations from Bouvet and Marion Island support the conclusion that asymmetric gene flow is maintained by passive drift of specimens with the eastward-flowing Antarctic Circumpolar Current (ACC). The high mobility of *S. septemcarinata* was unexpected and advises caution regarding the mobility of brooding species. In view of this long-distance dispersal, even to Marion Island that lies north of the Polar Front today, supports that small and remote islands such as Bouvet might in fact represent most important stepping-stones for dispersal to and from the Antarctic.

The only serolid isopod investigated from non-Antarctic regions, *Serolis paradoxa*, was sampled from the Strait of Magellan and from the East Falkland Islands. Both regions are situated on the South American continental shelf and are regarded as one zoogeographic province. For marine benthic taxa the strong current systems and the relatively shallow continental shelf should promote faunal exchange. In addition, the recurrent glacial maxima are expected to have facilitated dispersal between the two habitats as a result of lower sea levels at glacial maxima. For *S. pardaoxa*, however, populations from both regions differed substantially in microsatellite

data and a fragment of the 16S rRNA gene. The magnitude of differences supports the assumption that both populations belong to different reproductively isolated species rather than populations of one species. This was unexpected, as the distance between both habitats is much smaller than e.g. between the Antarctic Peninsula and the eastern Weddell Sea.

In conclusion, the results of this study provide evidence based on several independent nuclear microsatellites and a mitochondrial DNA marker that different processes in recent evolutionary history shaped the genetic structure of serolid isopod species on the High Antarctic shelf, the remote Southern Ocean islands and in the Magellan Region. They claim a major influence of the glacial advances on population structure and provide direct evidence that speciation towed to geographical separation is common in the Antarctic. However, results also show that brooding, low-mobile species can be capable to disperse over much greater distances than expected based on their lifestyle alone. The results of this dissertation hopefully encourage future studies on other benthic Antarctic taxa applying both marker systems of this study: a mitochondrial marker that is more susceptible to random genetic drift and several independent, fast-evolving microsatellite loci. Data from both marker systems allow to establish a comprehensive, evidence-based microevolutionary framework of speciation in the marine Antarctic benthos.

Zusammenfassung

Rezente molekulare Studien zur Phylogenie und Phylogeographie benthischer Invertebraten aus dem Südpolarmeer belegen, dass das traditionelle Konzept der zirkumpolaren Verbreitung von Arten eher die Ausnahme als die Regel darstellt. Für zahlreiche Arten mit postulierter zirkumpolarer Verbreitung zeigten diese Studien, dass es sich stets nicht um eine weitverbreitete, sondern um zwei oder mehrere kryptische Arten handelt, die ein weitaus kleinräumigeres, allopatrisches Verbreitungsmuster aufweisen. Die zu Grunde liegenden mikroevolutionären Prozesse, die ursächlich zur Trennung der ehemals einheitlichen Genpools geführt haben, sind bislang nur sehr unzureichend bekannt. Für die große Zahl Brutpflege betreibender Invertebraten auf dem relativ homogenen Schelf der Hochantarktis werden überwiegend die periodischen Vereisungen weiter Bereiche des Schelfs während glazialer Maxima mit „lineage sorting" innerhalb eisfreier Refugien auf dem Schelf als Ursache für allopatrische Artbildungsszenarien von zahlreichen Autoren diskutiert. Bislang existieren keine wissenschaftlichen Studien, welche mit hochauflösenden nukleären Mikrosatelliten die genetische Struktur von Populationen in der Hochantarktis mit Szenarien der glazialen Refugien Theorie verglichen haben. Des Weiteren gibt es zu dem tatsächlich realisierten Ausbreitungspotenzial von Brutpflege betreibenden benthischen Invertebraten der extrem isolierten Inseln im Südpolarmeer keine empirischen Daten. Genau diese Fragen werden in der vorliegenden Dissertation am Beispiel der Isopoden Familie Serolidae Dana 1852 (Crustacea: Peracarida) untersucht.

Zwei Themen stehen im Zentrum der Dissertation. Erstens wird eine Labor- und Softwareroutine entwickelt, die es ermöglicht, den schwierigen Prozess der Mikrosatellitenisolation aus weitgehend unbekannten Genomen zu vereinfachen. Zweitens wird die genetische Struktur, Diversität und Populationsdemographie von Brutpflege betreibenden Seroliden vom (i) Schelf der Hochantarktis, (ii) drei extrem isolierten Inseln der Antarktis und Subantarktis und (iii) zum Vergleich aus der Magellan Region mittels der neuen Mikrosatelliten und mitochondrialen DNA Marker analysiert.

Zur Etablierung von informativen Mikrosatelliten wurden zuerst drei traditionelle radioaktive und nicht-radioaktive Mikrosatelliten-Isolationsprotokolle herangezogen, die zwar das Vorhandensein von Mikrosatelliten im Genom einer Art bestätigten, deren Erfolg aber in keinem Verhältnis zum geleisteten Aufwand stand. Ein neues, innovatives und bis dahin kaum ausgearbeitetes und getestetes Verfahren zur Isolation und Etablierung hochinformativer Mikrosatelliten-Marker aus bislang unbekannten Genomen hat sich schließlich als sehr erfolgreich herausgestellt. Die Strategie, Mikrosatelliten mit vertretbarem Kosten- und Zeitaufwand zu finden, basiert auf einer neuartigen Methode, dem Reporter Genom Protokoll (RGP), welches DNA taxonomisch entfernt verwandter Organismen als Anreicherungssonden nutzt und keine spezifischen Repeat-Sonden benötigt. Es wird am Beispiel von sieben Arten gezeigt, dass mit dem RGP in kurzer Zeit eine große Anzahl unterschiedlicher Mikrosatelliten mit geringen Kosten gefunden werden können. In der Arbeit werden außerdem Strategien zur Erkennung ungeeigneter Mikrosatelliten diskutiert. Beim Auffinden von hochvariablen Mikrosatelliten kommt dabei die neue und in Kooperation entwickelte Software-Pipeline STAMP zum Einsatz.

Mit dem RGP werden drei hochvariable Mikrosatelliten Sets für (i) *Ceratoserolis* n. sp. 1 vom Schelf der Hochantarktis, (ii) für *Septemserolis septemcarinata* von unterschiedlichen antarktischen und subantarktischen Inseln und (iii) für *Serolis paradoxa* aus der Magellan Region

entwickelt. Die Anzahl der Allele je Locus umfasst ein großes Spektrum und ermöglicht dadurch die Erkennung geringster Unterschiede zwischen Populationen.

Für *Ceratoserolis* n. sp. 1 wurden Populationen von der Spitze der Antarktischen Halbinsel (AP) und vom Schelf des östlichen Weddellmeers (EWS) auf ihre Variabilität in zehn Mikrosatelliten und einem Fragment des Cytochromoxidase I Gens (COI) analysiert. Varianz-basierte und Bayes'sche Analysen der Populationsstruktur zeigen eine deutliche geographische Struktur in der Verteilung genetischer Polymorphismen. Eine um die AP herum verbreitete Gruppe von Genotypen ist klar abgesetzt von einer anderen, ausschließlich im EWS verbreiteten. Die Höhe der genetischen Unterschiede belegt das Fehlen eines effektiven rezenten Genflusses zwischen den beiden Regionen. Innerhalb der zwei Regionen findet sich keine weitere konsistente Strukturierung genetischer Variabilität, selbst nicht bei Distanzen von über 500 km. Innerhalb der Regionen ist Genfluss zwischen Populationen also relativ uneingeschränkt möglich. Die Höhe der Unterschiede zwischen den Regionen (AP und EWS) belegt, dass die Aufspaltung in zwei Gruppen mit ziemlicher Sicherheit bereits vor der letzten Eiszeit stattfand und daher beide Gruppen in voneinander unabhängigen Refugien die letzte Eiszeit überdauerten. Die Daten dieser Arbeit unterstützen die Annahme, dass ein Refugium höchstwahrscheinlich an der Spitze der AP und ein anderes auf dem Schelf des EWS lokalisiert waren. Diese Ansicht wird von neuen geophysikalischen Daten unterstützt, denen zufolge in beiden Regionen der Einfluss von Grundeis geringer war als bislang vermutet.

Populationen von *Septemserolis septemcarinata* von den extrem entlegenen Inseln Südgeorgien, Bouvet und Marion weisen genetisch deutlich unterschiedliche Merkmale auf. Andererseits zeigen die Daten zu *S. septemcarinata* in dieser Arbeit, dass in Ausnahmefällen auch gering mobile Arten in der Lage sind, Genfluss über evolutionäre Zeiträume und große geographische Distanzen hinweg aufrecht zu erhalten. Es erscheint uns am wahrscheinlichsten, dass die Ausbreitung der strikt benthischen Organismen mittels passiv driftender Makroalgen erfolgt. Die auf den ersten Blick überraschende Verteilung der populationsgenetischen Diversität, mit geringsten Werten um die geologisch alte und große Insel Südgeorgien und viel höheren Werten der genetischen Diversität für die Populationen von Bouvet und Marion, unterstützen den Schluss, dass Genfluss über passive Drift von Organismen mit dem ostwärts fließenden Zirkumpolarstrom asymmetrisch aufrecht erhalten wird. Die hohe Mobilität von *S. septemcarinata* war unerwartet und rät zur Vorsicht bei generellen Aussagen über die tatsächliche Mobilität Brutpflege betreibender Arten.

Populationen von *S. paradoxa* aus der Magellanstraße und von den Falklandinseln wiesen enorme Unterschiede in den Mikrosatelliten und 16S rDNA Daten auf. Beide Regionen liegen auf dem Schelf von Südamerika und werden traditionell als eine zoogeographische Provinz bezeichnet. Die geringe Tiefe zwischen beiden Regionen und die starken Strömungen lassen eigentlich einen regelmäßigen Faunenaustausch erwarten. Die Unterschiede legen die Schlussfolgerung nahe, dass es sich bei den untersuchten Populationen um unterschiedliche Arten handelt. Dieses Ergebnis ist unerwartet in Anbetracht der räumlich erheblich geringeren Isolierung im Vergleich zu *Ceratoserolis* n. sp. 1 und *S. septemcarinata*.

Zusammenfassend belegen die Ergebnisse dieser Dissertation erstmals an Hand eines auflösender nukleärer Mikrosatelliten und jeweils einem mitochondrialen DNA Marker, dass unterschiedliche Prozesse in der jüngeren Vergangenheit die genetischen Struktur innerhalb von Arten der Familie Serolidae auf dem Schelf der Hochantarktis, den entlegenen Inseln des Südpolarmeers und in der Magellanregion geformt haben. Die Ergebnisse unterstützen erstmals

mit empirischen Daten die Hypothese, dass Artbildung auf dem heutzutage relativ homogenen antarktischen Schelf durch wiederkehrende großflächige Vereisungen unter Allopatrie möglich war. Sie verdeutlichen zudem, dass die Mobilität Brutpflege betreibender Arten in Ausnahmefällen hoch sein kann. Auf der Grundlage des Auflösungsvermögens der angewandten Marker ermutigt diese Studie zu weiterer Forschung an anderen benthischen Taxa der Antarktis. Die Kombination mehrerer unabhängiger, hochvariabler Mikrosatelliten-Loci mit einem mitochondrialen Gen, welches erheblich sensitiver in Bezug auf genetische Drift ist, erwies sich als besonders geeignet, um unterschiedliche Zeitskalen in der jüngeren Vergangenheit der Arten auflösen zu können. Daten dieser Markersysteme können in Zukunft erheblich dazu beitragen eine empirisch fundierte Theorie über die wichtigsten mikroevolutionären Prozesse für die Artbildung im Südpolarmeer zu liefern.

Danksagung

Prof. Dr. Johann-Wolfgang Wägele (ZFMK, Bonn) danke ich für die Bereitschaft, die Arbeit zu betreuen. Seine Hilfsbereitschaft, raschen Urteile und wertvollen Diskussionsbeiträge waren von großer Hilfe.

Prof. Dr. Dominik Begerow (Ruhr-Universität Bochum) danke ich für die Bereitschaft as Zweitgutachter dieser Dissertation zu fungieren.

Dr. Christoph Held (AWI Bremerhaven) hat dieses Projekt ins Leben gerufen und mir die Doktorandenstelle für das Projekt anvertraut. Für dieses Vertrauen, auch während der langen "Durststrecken" bei der Entwicklung der Marker, bedanke ich mich. Vor allem aber möchte ich für die Betreuung, die vielen Gestaltungsfreiräume innerhalb des Projektes und die stetige Bereitschaft, sein enzyklopädisches molekulargenetisches und phylogenetisches Wissen weiterzugeben, herzlichst danken! Seine zahlreichen Denkanstöße und Anregungen haben wesentlich zu meinem Verständnis von Evolutionsprozessen generell und in der Antarktis speziell beigetragen.

Ohne die Hilfe von Andrea Eschbach, Shobhit Agrawal, Annegret Müller, Tilman Alpermann (AWI Bremerhaven) und Anna Kop (University of Toronto, Canada) wäre die Bearbeitung der Proben in dem vorliegenden Umfang und bei so guter Stimmung nicht möglich gewesen.

Besonderer Dank gilt Dr. Christoph Mayer (Ruhr-Universität Bochum) für wertvolle Diskussions-, Software- und Schriftbeiträge im Rahmen dieser kumulativen Doktorarbeit. Die standardisierte statistische Analyse der Mikrosatellitendaten ist das Resultat der von ihm in den Jahren der Doktorarbeit geschriebenen Programme *Sat-Stat* und *Phobos*.

Prof. Dr. W. Detrich III (Northeastern University, Boston) danke ich für die Möglichkeit, im Rahmen der US-amerikanischen Expedition ICEFISH 2004 mit dem Forschungseisbrecher *„Nathaniel B. Palmer"* wertvolle Proben für diese Arbeit in der Antarktis zu sammeln.

Dr. Erika Mutschke und Dr. Carlos Rios (Magellan Universität Punta Arenas, Chile) haben wertvolles Probenmaterial zur Verfügung gestellt.

PD Dr. Julian Gutt und Dr. Uwe John (AWI Bremerhaven) haben in regelmäßigen Abständen den Fortschritt der Arbeit in ihrer Funktion als AWI-internes Doktorandenkomitee begleitet. Dafür möchte ich herzlich danken.

Ohne die wertvollen Tipps von Dr. Arne Nolte (Universität Laval, Kanada) wäre die Isolation von Mikrosatelliten nicht so erfolgreich gewesen.

Am Alfred-Wegener-Institut hat Dr. Jan Schulz das Programm „SEROLIS" zur komfortablen Darstellung von Sequenzähnlichkeiten geschrieben. Hierfür sei an dieser Stelle gedankt.

Dr. Bank Beszteri und Prof. Dr. Stephan Frickenhaus danke ich für die Nutzung der UNIX/LINUX Cluster und ihre stetige Unterstützung bei Fragen zur Bioinformatik.

Dr. Jan Michels hat wesentliche Teile der Dissertation gelesen und wertvolle inhaltliche und sprachliche Kommentare zur Verbesserung der Arbeit geliefert.

Der Bochumer Arbeitsgruppe vom Lehrstuhl für Spezielle Zoologie sei herzlich für ihre Hilfe und Unterstützung gedankt!

Der DFG danke ich für finanzielle Förderung im Rahmen des Projekts He-3391/3.

Für ein Stipendium im Rahmen des RISE Programms möchte ich mich beim DAAD bedanken.

Bei meiner kleinen Familie möchte ich mich für die stetige Unterstützung vor allem in schwierigeren Zeiten danken. Und natürlich bei allen, die ich jetzt hier vergessen habe und die dennoch einen wichtigen Beitrag zum Erfolg dieser Arbeit geleistet haben

Curriculum Vitae Florian Leese (Dipl.-Biol., M.Ed.)

Date and place of birth

18th July 1978 Lauterbach (Hesse, Germany)

Material Status

Unmarried, one child

Father: Klaus Walter Leese (special education teacher), Mother: Regina Leese (nee Brakensiek; welfare worker).

University Education

Dec 2007 - ongoing

Scientific assistant at the Department of Animal Ecology, Evolution and Biodiversity, Prof. Dr. R. Tollrian, Ruhr-University Bochum

Nov 2006

M.Ed. in Biology and Sports Science ("very good")
Ruhr-University Bochum

Sep 2005 – Dec 2007

Continuing of the Ph.D. thesis at the Alfred Wegener Institute for Polar and Marine Sciences, Working Group "Marine Animal Ecology"
Advisors: Prof. Dr. J.-W. Wägele (ZFMK Bonn), Dr. C. Held (AWI Bremerhaven)
Thesis: "The recent evolutionary history of Antarctic and Subantarctic isopods: Development and analysis of fast evolving molecular markers" (DFG HE-3391/3 to Christoph Held)

Jul 2004 – Aug 2005

Ph.D. thesis at the Department of Animal Morphology and Systematics, Ruhr-University Bochum (Head: Prof. Dr. J.-W. Wägele)

May – Jul 2004

Participation at the 'International Collaboration to Sample Fishes indigenous to Subantarctic Habitats' (ICEFISH 2004) to the Antarctic and Subantarctic with Prof. Dr. H.W. Detrich III (Northeastern University, Boston, USA)

Mar 2004

M.Sc. in Biology ("with distinction")
Max Planck Institute for Limnology & University of Bochum, 2004
Advisors: Prof. Dr. R. Wagner, Prof. Dr. J.-W. Wägele, Prof. Dr. H. Häupler
Thesis: Molecular genetic, chemotaxonomic, and autecological investigations of central European Sericostomatidae (Trichoptera)

Mar 2003 – Feb 2004

Diploma thesis at the Max Planck Institute for Limnology, River Station Schlitz (Germany)

Sep 2001 – Mar 2003

Studies in Biology and Sports Science (since 2001) at the Ruhr-University Bochum (Germany)

Jul – Sep 2000

Internship at the Federal State University of Mato Grosso, Cuiaba (Brazil). Fieldwork as part of the SHIFT project (Max-Planck Institute for Limnology, Plön, Prof. Dr. W. Junk)

Oct 1997 – Sep 2001

Studies in Biology and Sports Science (since 2001) at the University of Marburg

School Education

Aug 1984 – Jun 1997

Waldorfschool Loheland, Fulda (Hesse)

Publications (exclusive those in the Ph.D. Thesis)

LEESE, F. and N. BRENKE, (2005) *Chauliodoniscus coronatus* sp. nov., a new deep-sea species from the Angola Basin (Asellota, Janiroidea, Haploniscidae). Organisms Diversity & Evolution 5: 189-201.

LEESE, F. and R. WAGNER, (2005) The "*Sericostoma*-Problem" - molecular, chemotaxonomic, and autecological approaches (Trichoptera: Sericostomatidae). Lauterbornia 54: 161-163.

WAGNER, R., F. LEESE and A. R. PANESAR, (2004) Aquatic dance flies from a small Himalayan mountain stream (Diptera: Empididae: Hemerodromiinae, Trichopezinae and Clinocerinae). Bonner Zoologische Beiträge 52: 3-32.