

Population genetic consequences of a diapause:  
A species comparison of two calanoid copepods  
(*Eudiaptomus spec.*)

Dissertation  
zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität  
zu Kiel

vorgelegt von  
Martina Zeller

Max-Planck-Institut für Limnologie  
Kiel 2006

Für meinen Vater

## Vorbemerkung

Ergebnisse dieser Dissertation sind bereits wie folgt vorab veröffentlicht oder zur Publikation eingereicht worden:

Veröffentlicht:

Zeller, M., and T. H. Reusch. 2004. Identification and characterization of 10 microsatellite primers for the calanoid freshwater copepods *Eudiaptomus gracilis* and *E. graciloides* using enriched genomic libraries. *Mol. Ecol. Notes* 4:355-357.

Zeller, M., T. 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 different potential for dispersal. *Limnol. Oceanogr.* 51:117-124.

Eingereicht:

Bohonak, A. J., M. D. Holland, B. Santer, M. Zeller, C. M. Kearns, and N. G. Hairston Jr.. The population genetic consequences of diapause in *Eudiaptomus* copepods. In *Revision. Arch. Hydrobiol.*

Das folgende Manuskript befindet sich in der Vorbereitung für Evolution (brief communications):

Zeller, M., T. H. Reusch, and W. Lampert. Small effective population sizes in two planktonic freshwater copepod species (*Eudiaptomus*) with apparently large census sizes.

**Contents**

Summary	2
Zusammenfassung	4
Introduction	6
1. Genetic structure of populations	6
1.1. An overview	6
1.2. Isolation-by-distance: A model to describe the genetic structure of populations in accordance to their spatial distribution	7
1.3. Genetic drift and effective population size ( $N_e$ )	9
2. Microsatellites as molecular markers	10
3. Population genetics in freshwater zooplankton	14
4. The model system of the present study: <i>Eudiatomus spec.</i>	15
5. General goals and hypotheses	16
6. Thesis outline	17
Chapters:	
I Identification and characterization of 10 microsatellite primers for the calanoid freshwater copepods <i>Eudiatomus gracilis</i> and <i>Eudiatomus graciloides</i> using enriched genomic libraries	20
II A comparative population genetic study on calanoid freshwater copepods: Investigation of isolation-by-distance in two <i>Eudiatomus</i> species with a different potential for dispersal	25
III Small effective population sizes in two planktonic freshwater copepod species ( <i>Eudiatomus</i> ) with apparently large census sizes	42
IV The population genetic consequences of diapause in <i>Eudiatomus</i> copepods	59
References	80
Danksagung	90
Appendix 1	92
Appendix 2	93
Appendix 3	95
Curriculum Vitae	98
Erklärung	99

## Summary

The widespread calanoid copepods *Eudiaptomus graciloides* and *E. gracilis* co-occur in a number of European lakes and are morphologically similar, but only *E. graciloides* is capable to produce diapausing eggs. Therefore, these two sister-species are ideal to investigate the genetic impact of a life-cycle with a diapause in a between-species comparison.

In addition to my interest in the population genetic structure of *Eudiaptomus* spp. in general, I hypothesized that diapausing eggs, as a consequence of their high dispersal capacity, lead to higher gene flow and, hence, to less genetic differentiation between populations. This was tested in samples of 16 *E. gracilis* and 17 *E. graciloides* populations at three different spatial scales (investigation of isolation-by-distance in a range of 100, 1000, and >1000 km). Length polymorphism of 10 microsatellite markers from previously developed, enriched DNA libraries was used for the population genetic analyses. *Eudiaptomus* species generally revealed complex and highly scale dependent patterns of genetic differentiation with geographic distance, i.e. isolation-by-distance. In *E. graciloides*, a positive correlation between genetic and geographic distances was detected both at the smallest (100 km) and at the largest spatial scale (>1000 km), but was absent at the intermediate range (1000 km). Genetic differentiation in *E. gracilis* populations never depended on geographic distances neither at a small nor at an intermediate spatial scale. Isolation-by-distance at the largest scale (>1000 km) could not be investigated with the data of this study, because *E. gracilis* was not found in the most easterly lakes of the sampling area. Above findings suggested that in *Eudiaptomus* species short- and long-distance dispersal were the result of different mechanisms, with diapausing eggs only being advantageous for short-distance dispersal. The results at a 1000 km range rather suggested either similar moderate gene flow in both *Eudiaptomus* species over large geographic distances or persistence of historic patterns after postglacial recolonization.

Secondly, I hypothesized that the occurrence of a diapausing egg bank in *E. graciloides* results in higher effective population sizes ( $N_e$ ) compared to *E. gracilis*, due to the maintenance of genetic diversity. Effective population sizes were estimated from four temporal samples taken from a lake in northern Germany. Allele frequencies were obtained by assessing the polymorphism displayed by altogether 10 microsatellite loci.  $N_e$ s for both

*Eudiaptomus* species were much smaller than one would expect for a zooplankton species that is passively transported by movements of the water. This might be due to a substructuring of populations, a skewed reproduction success, unequal sex-ratios, non-random mating or fluctuating population sizes. Despite the fact that a diapausing egg bank existed in *E. graciloides* only, no significant differences between  $N_e$ s could be found between these species, suggesting that long-term diapausing eggs might not be as influential to the water column population in the lake under study than expected.

In order to obtain information on the genetic role of diapause in *Eudiaptomus* species over longer time scales, the variation of mitochondrial DNA sequences from populations from three lakes in northern Germany was studied in a joint project with A. J. Bohonak, M. D. Holland, B. Santer, C. M. Kearns, and N. G. Hairston (Jr.). Mitochondrial DNA variation was different in both *Eudiaptomus* species: Fewer alleles and a larger average divergence occurred in *E. graciloides*. Significant divergence among lakes in mitochondrial and nuclear markers (microsatellites) demonstrated that gene flow was low in *Eudiaptomus* species. Estimates of  $N_e\mu$  ( $\mu$  = mutation rate) backwards in time derived from a Bayesian approach were not significantly different for most time points, but tended to be larger in *E. gracilis* compared with *E. graciloides*. Above findings could be explained by phylogeographic analyses and Bayesian skyline plots: these analyses gave evidence for historic range expansions that occurred thousands of generations earlier in *E. gracilis* than in *E. graciloides*. Additionally, the populations of *E. graciloides* seem to have passed through a bottleneck. Thus, populations of *E. graciloides* are likely to have not yet reached equilibrium between gene flow and drift.

## Zusammenfassung

Die calanoiden Copepoden *Eudiaptomus graciloides* und *E. gracilis* sind in Seen Europas weit verbreitet, und sie koexistieren in einer Vielzahl von ihnen. Beide Arten sind nah miteinander verwandt und morphologisch sehr ähnlich. Dauereier treten nur im Lebenszyklus von *E. graciloides* auf. Daher stellen *E. graciloides* und *E. gracilis* ein ideales System dar, um in einem zwischenartlichen Vergleich den Einfluss einer Diapause als Bestandteil des Lebenszyklusses auf die genetische Populationsstruktur zu untersuchen.

Neben einem allgemeinen Interesse an der genetischen Populationsstruktur beider *Eudiaptomus*-Arten, wurde die Hypothese untersucht, dass Dauereier, bedingt durch ihr Verbreitungspotential, den Genfluss erhöhen und somit die genetische Differenzierung zwischen Populationen erniedrigen. Es wurden 16 *E. gracilis* und 17 *E. graciloides* Populationen innerhalb von drei räumlichen Größenskalen untersucht (Untersuchung der Isolation durch Distanz in einer Skala von 100, 1000 und >1000 km). Die genetische Analyse erfolgte mittels des Längenpolymorphismus von 10 Mikrosatellitenmarkern, die aus zuvor entwickelten angereicherten DNA Bibliotheken stammten. Es zeigten sich für beide *Eudiaptomus*-Arten komplexe Muster für die genetischen Distanzen in Abhängigkeit von der geografischen Entfernung, z.B. Isolation durch Distanz, die vom untersuchten Massstab abhingen. Für *E. graciloides* wurde eine positive Korrelation für die kleinste (100 km) und größte (>1000 km) untersuchte geografische Entfernung gefunden, nicht aber für eine intermediäre Entfernung (1000 km). Die genetische Differenzierung in den *E. gracilis* Populationen war weder auf der kleinsten noch auf einer mittleren Skala abhängig von der geografischen Entfernung. Isolation durch Distanz für Entfernungen >1000 km konnte mit diesem Datensatz nicht untersucht werden, weil *E. gracilis* in den östlichsten Teilen des Untersuchungsgebietes nicht gefunden wurde. Es wurde gefolgert, dass die Verbreitung von *Eudiaptomus* spec. über kurze und lange Distanzen durch jeweils verschiedene Prozesse erfolgt, und dass die Verbreitung durch Dauereier *E. graciloides* nur innerhalb kurzer Distanzen einen Vorteil verschafft. Die Ergebnisse für die 1000 km Distanz sprechen bei beiden Arten entweder für einen ähnlichen moderaten Genfluss über große Entfernungen oder für den Erhalt von historischen Mustern, die bei der postglazialen Wiederbesiedlung entstanden sind.

Die zweite Hypothese dieser Arbeit war, dass eine Dauereibank durch den Erhalt von genetischer Diversität zu höheren effektiven Populationsgrößen ( $N_e$ ) bei *E. graciloides* im Vergleich zu *E. gracilis* führt. Die effektiven Populationsgrößen wurden auf der Grundlage von vier zeitlich genommenen Proben aus einem See in Norddeutschland berechnet. Zum Erhalt der hierfür notwendigen Allelfrequenzen wurde der Polymorphismus von insgesamt 10 Mikrosatelliten Loci verwendet. Die effektiven Populationsgrößen waren für beide *Eudiptomus*-Arten kleiner, als es für Zooplanktonarten, die innerhalb eines Sees passiv verfrachtet werden, zu erwarten gewesen wäre. Hierfür könnten Faktoren wie Substrukturen innerhalb der Populationen, ungleich verteilter Reproduktionserfolg, ungleiche Geschlechterverhältnisse, nicht zufällige Verpaarungen oder fluktuierende Populationsgrößen eine mögliche Ursache sein. Es konnten keine signifikanten Unterschiede zwischen *E. graciloides* und *E. gracilis* in den  $N_e$ s gefunden werden. Daraus kann geschlossen werden, dass der Einfluss der Dauereibank von *E. graciloides* auf dessen pelagische Population nicht so groß ist, wie zuvor angenommen wurde.

Um Informationen über die genetische Rolle der Diapause in *Eudiptomus spec.* innerhalb eines größeren Zeitraums zu erhalten, wurde die Variation in mitochondriellen DNA-Sequenzen für Populationen aus drei norddeutschen Seen im Rahmen einer Zusammenarbeit mit A. J. Bohonak, M. D. Holland, B. Santer, C. M. Kearns und N. G. Hairston Jr. untersucht. Für *E. graciloides* wurden weniger Allele und eine größere durchschnittliche Divergenz gefunden als für *E. gracilis*. Signifikante Divergenzen zwischen den Seen, die sowohl mit mitochondriellen als auch mit nuklearen Markern (Mikrosatelliten) gefunden wurden, demonstrierten geringen Genfluss für beide *Eudiptomus*-Arten. Schätzungen einer Bayesischen Annäherung für  $N_e\mu$ , die auf der Zeitskala zurückgehen, unterschieden sich für die meisten Zeitpunkte nicht signifikant voneinander, waren aber tendenziell in *E. gracilis* größer als in *E. graciloides*. Diese Ergebnisse konnten durch phylogeografische Analysen und Bayesische Skyline-plots erklärt werden, die auf historische Expansionen für beide Arten hinwiesen: Die Expansionen von *E. graciloides* fanden wohl tausende von Generationen später statt als bei *E. gracilis*, und *E. graciloides* ist durch einen Populationsengpass gegangen. Daher ist anzunehmen, dass *E. graciloides* noch kein Gleichgewicht zwischen Genfluss und genetischer Drift erreicht hat.

## **Introduction**

The advances in molecular techniques and the increased accessibility of molecular markers over recent years have resulted in population genetic studies in numerous species. Gene flow, genetic drift, mutation and selection determine the genetic variability within and the genetic differentiation between populations, and therefore are most important parameters for population biology and conservation (Hedrick 1999). The utility of molecular genetic data is twofold: Firstly, such assessments allow the quantification of otherwise unaccessible demographic information. For example, the difficulties that arise when studying the migration of individuals or propagules can be insurmountable, in particular in species where individuals are small. Even if migration events can be observed properly, migration must not adequately reflect ongoing gene flow between populations, since the immigrating individual has to be able to survive, reproduce and contribute to the gene pool under the conditions of the new habitat (De Meester et al. 1996; Bohonak and Jenkins 2003). The second set of questions that greatly benefits from the use of genetic markers is related to population genetic processes such as inbreeding and genetic diversity in itself. Such processes are of relevance to ecology and conservation, because they determine population viability and individual fitness (Hedrick 1999).

### **1. Genetic structure of populations**

#### **1.1. An overview**

An ideal population can be defined as a group of interbreeding individuals, where each individual has an equal chance to be the parent of any progeny. Such an ideal population almost never exists, since there are various constraints upon the chances of any two individuals to exchange genetic information by recombination. Most species do not exist in one single large population, but rather in smaller subdivided units. Such substructuring can be caused by ecological, geographic or even behavioral dynamics. The rate of genetic exchange between subdivided populations will determine if breeding is more likely to occur within a subpopulation or rather in the whole population and therefore, determine the rate of genetic differentiation. If gene flow is high between subpopulations, it will have a homogenizing effect on gene frequencies, but if gene flow is reduced, genetic drift, selection and mutation can lead to genetic differences between these subpopulations. Thus,

gene flow is one component that can maintain genetic diversity of populations, especially if they are small, and thus retains its potential to respond to changing environmental conditions. Today usually the term “metapopulation” is used to describe any subdivided populations. However, in a more restricted view, a metapopulation was originally defined as a subdivided population, where subpopulations in single patches go extinct and are recolonized by individuals from the other subpopulations (Hanski 1998).

Sewall Wright developed a set of statistics in the 1930s to describe and quantify the genetic variation in and among populations. These so called F-statistics is briefly introduced in Box 1 together with the term heterozygosity, which is a common measure of genetic variation within a population.

## **1.2. Isolation-by-distance: A model to describe the genetic structure of populations in accordance to their spatial distribution**

To study genetic differentiation and gene flow between populations traditional attempts related estimates of regional  $F_{ST}$  to gene flow and drift with Wright’s equation  $F_{ST} = 1/(4Nm+1)$ , where  $N$  is the effective population size and  $m$  gives the portion of exchanged individuals per generation (Wright 1931). This method is a good first approximation, however, its utility depends upon a number of critical assumptions. For example, this method is based on the island-model where a gene flow-drift equilibrium has been reached between populations. This assumption is frequently violated, since natural populations often fluctuate in size and are founded and go extinct, which can cause that the equilibrium between the homogenizing effects of gene flow and the differentiating effects of genetic drift has not been reached (McCauley 1993; Whitlock and McCauley 1999). On the contrary, the isolation-by-distance hypothesis is based on the stepping stone model, and therefore is an appropriate alternative to study and understand the genetic structure of populations, even if the species under investigation has not reached a gene flow-drift equilibrium and if dispersal is limited (Slatkin 1993; Hutchison and Templeton 1999; Bohonak and Roderick 2001). The model of isolation-by-distance predicts a positive correlation of geographic and genetic distances (Wright 1943; Slatkin and Maddison 1990; Slatkin 1993), since gene flow is more likely to occur between populations located nearby than populations further away. Hence, to test for isolation-by-distance the genetic differentiation of all investigated population pairs is plotted as a function of geographic

distance and a Mantel test, which compares the matrices of both parameters, is performed to get information on how much of the genetic differentiation between populations can be explained by their geographic distance (Mantel 1967). A lack of correlation between genetic and geographic distance can be due to different reasons: (i) dispersal and gene flow, respectively, might be irregular and dependent on factors other than distance, (ii) the populations might not have reached a gene flow-drift equilibrium, and (iii) historical events might be reflected rather than ongoing gene flow (e.g. persistent founder effects).

*Box 1: Brief introduction into important population genetic parameters and F-statistics*

Heterozygosity is an important parameter to characterize within population diversity. The expected heterozygosity ( $H_E$ ) is the heterozygosity of a population that can be expected from the sampled allele frequencies under the assumption that the population is in Hardy-Weinberg equilibrium, and thus comes close to an ideal population (Hedrick 1999). The observed heterozygosity ( $H_O$ ) of a population denotes the measured heterozygosity. Both heterozygosity estimates are related to the inbreeding coefficient  $F_{IS}$ :

$$F_{IS} = H_E - H_O / H_E, \quad \begin{array}{l} H_O = \text{observed heterozygosity within a subpopulation} \\ H_E = \text{expected heterozygosity within a subpopulation} \end{array}$$

and thus measure the deviation from Hardy-Weinberg proportions within (sub)populations (Hedrick 1999). Inbreeding may be caused by the non-random mating of individuals within a subpopulation. Its mathematical theory was introduced by Wright (1921, 1931, 1969) and further improved by Crow and Kimura (1970). In inbreeding populations homozygosity will increase with every generation. The parameter  $F_{IT}$  is also a measure of the deviation from Hardy-Weinberg proportions, but of the total population rather than within subpopulations. Positive values of  $F_{IS}$  and  $F_{IT}$  indicate a deficiency of heterozygote individuals, and negative values indicate an excess of heterozygote individuals (Hedrick 1999).

$$F_{IT} = H_T - H_O / H_T, \quad H_T = \text{expected heterozygosity in the total population}$$

The differentiation between populations can be expressed in the  $F_{ST}$ -value. If allele frequencies in two subpopulations deviate, the average expected heterozygosity ( $H_E$ ) in those populations will always be less than that expected from the pooled allele frequencies ( $H_T$ ).

$$F_{ST} = H_T - H_E / H_T,$$

$F_{ST}$  takes a value between 0 (no genetic differentiation) and 1 (complete genetic differentiation). Accordingly, Weir and Cockerham (1984) developed a variance based method for the estimation of above F-statistics with  $\theta$  being an estimator for  $F_{ST}$ .

### 1.3. Genetic drift and effective population size ( $N_e$ )

Genetic drift changes allele frequencies of populations randomly due to the loss of alleles or chance fixation (Hedrick 1999). Drift has only little effect if a population is large, but with decreasing population size drift becomes more influential, which can, for example, result in the random loss of favourable alleles or the fixation of slightly deleterious mutations (Hedrick 1999; Frankham et al. 2002). How fast genetic drift removes genetic variation from the population is inversely proportional to the effective population size ( $N_e$ ). Thus, the effective population size is an important parameter in population biology. The concept of effective population size was introduced by Wright (1931): An ideal population of size  $N$  is considered, where all parents have the same expectation of being the parents of any progeny (Hedrick 1999). The effective population size can be either calculated as the inbreeding effective size or as the variance effective size. The inbreeding effective size is defined as the expected rate of increase in homozygosity and the variance effective size as the expected rate of increase of variance in allele frequency per generation (Crow and Denniston 1988). In other words, the effective population size gives the size of an idealized population that would produce the same amount of inbreeding and allelic frequency variance, respectively, as the studied population (Hedrick 1999). If populations are not changing over time, values from both effective sizes are quite similar, otherwise they can differ substantially (Kimura and Crow 1963). The effective population size can be much smaller than the total number of individuals found in an area (census population size,  $N_c$ ), (Frankham 1995). This discrepancy might be due to factors like the mating system, a skewed sex-ratio, variance in reproductive success or fluctuating population sizes (Caballero 1994). Thus, the calculation of the effective size of a population provides valuable information, that is often difficult or impossible to obtain from wild populations (Frankham 1995).

A number of methods have been developed to estimate effective population sizes, one of them being the “temporal approach”. In the temporal approach  $N_e$  is calculated from temporally spaced samples relating the changes in allele frequencies over time to genetic drift (e.g. Krimbas and Tsakas 1971; Pollack 1983; Waples 1989; Williamson and Slatkin 1999; Anderson et al. 2000). Factors other than genetic drift, that might influence allele frequencies, are neglected in all these approaches. Since migration is an important factor in most natural populations, Wang and Whitlock (2003) developed a novel method, that takes

migration between populations into account. These authors argue that neglecting migration will either lead to up- or downward biases in the estimate of  $N_e$  (Wang and Whitlock 2003). In the short term, migration events will change allele frequencies quickly. If these were interpreted as genetic drift, estimations of  $N_e$  will be too low. On the long run, by reaching an equilibrium state, migration will slow down changes in allele frequencies caused by genetic drift. In this case genetic drift would be underestimated and hence  $N_e$  overestimated.

## 2. Microsatellites as molecular markers

Microsatellites (msats) are short, tandemly repeated deoxyribonucleic acid (DNA) units of 2 to 6 base pairs, e.g.  $(CA/GT)_n$  or  $(CTAG/GATC)_n$ , that are abundant in eukaryotic genomes. These sequences are highly polymorphic in natural populations due to variability in the number of tandem repeats ( $n$ ). Microsatellites have discrete and co-dominant alleles, they are inherited in a Mendelian-way and are very likely to be selectively neutral (Goldstein and Schlötterer 1999). Their high mutation rates of  $10^{-4}$  to  $10^{-5}$  (Ellegren 1992; Schlötterer and Tautz 1992; Edwards et al. 1992) make them sensitive markers to detect the structure and differentiation at the intra- and interpopulational level. However, Nauta and Weissing (1996) showed within a theoretical framework that high mutation rates can also have a homogenizing effect, since the range of allele sizes in microsatellites is limited, by shifting different populations to the same distribution of alleles. As a consequence, the potential for genetic divergence can be limited, especially if populations are large and studied in a rather long term perspective (Nauta and Weissing 1996). The predominant mechanism of mutations of microsatellites is slipped strand mispairing of complementary bases during DNA replication (Levinson and Gutman 1987; Schlötterer and Tautz 1992), which increases or decreases the current number of repeats by one unit. The resulting length variation within microsatellites can be amplified by polymerase chain reaction (PCR) with primers in the conserved flanking regions. Three different types of microsatellites can be classified. The ‘perfect’ microsatellite consists of a stretch of DNA that contains the motif only (e.g. CACACACACACACACACA). The ‘imperfect’ or ‘interrupted’ microsatellite has one or more interruptions of the core motif (e.g. CACATTCACACACATTCACA), and in ‘compound’ microsatellites different motifs can be found adjacent to one another (e.g. CACACACAGAGAGAGAGAGA), (Ashley and

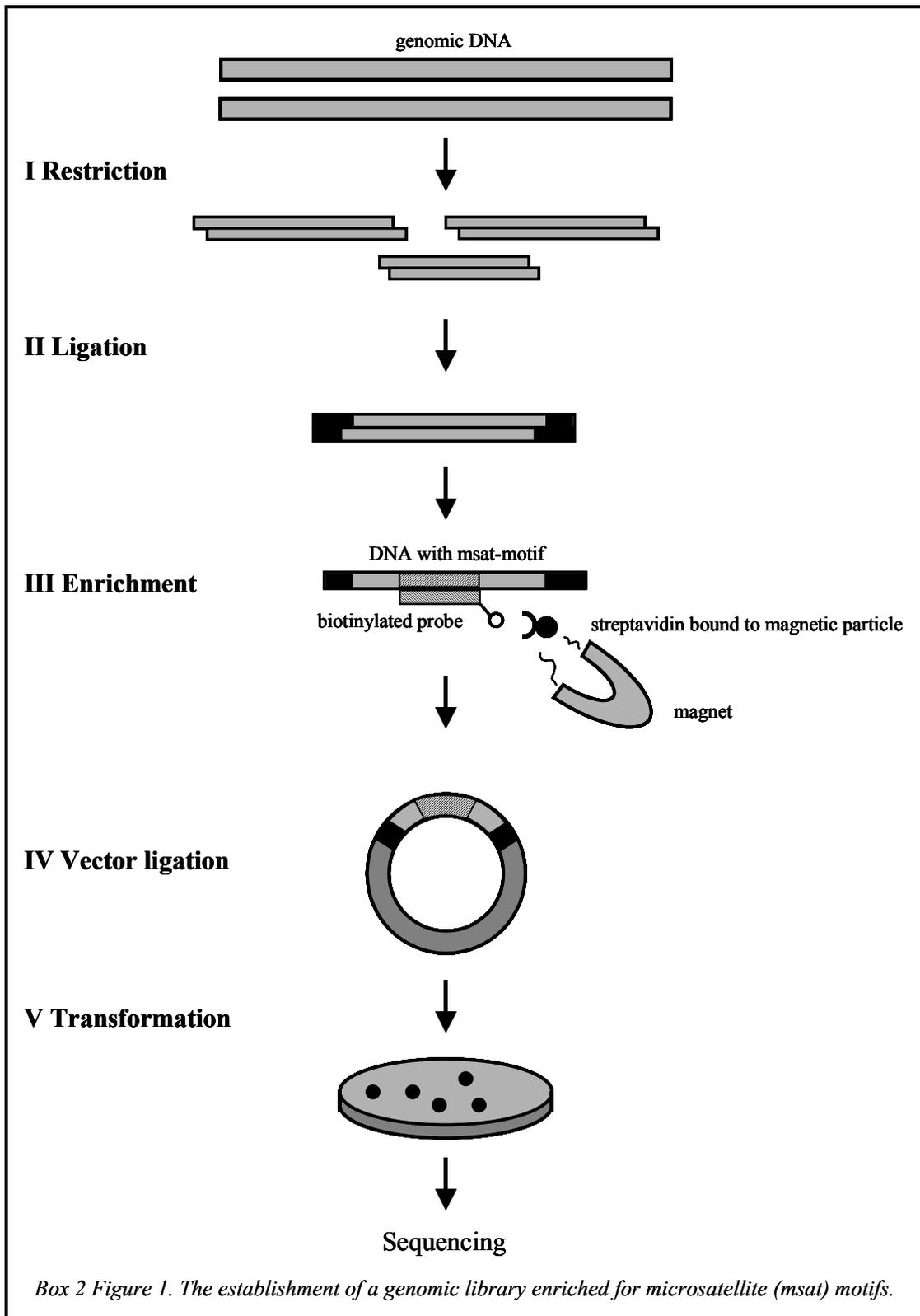
Dow 1994; Jarne and Lagoda 1996). Imperfect and compound motifs are likely to be less polymorphic compared to perfect microsatellites (reviewed in Jarne and Lagoda 1996). The major drawback of using microsatellite markers is that for most species being examined for the first time they need to be isolated *de novo* (Zane et al 2002). This establishment of a so called microsatellite library can be time consuming and expensive. The goal is to find microsatellite loci within the whole genome of the target species and to design appropriate primers in the flanking regions. However, once established, the amplification of microsatellite loci in PCRs is straight forward and can be assayed from minute quantities of DNA.

A short overview of the procedure to establish a genomic library that was used for the target species of this thesis is given in Box 2. Various alternative protocols exist, which can be found elsewhere (reviewed by Zane et al. 2002).

*Box 2: The establishment of a genomic library*

High purity genomic DNA is digested using restriction enzymes (B2 Fig. 1I). The desired length of DNA fragments is in the range of 200 – 1000 bp, because these fragments are long enough to minimize the risk that the microsatellite occurs too close to either end of the fragment, but still short enough to be completely sequenced. The use of different restriction enzymes increases the likelihood that every part of the genome is represented among the fragments with a desired size range of 200 – 1000 bp, which will form the basis of the genomic library. Accordingly, the appropriate size fraction of fragmented DNA can be size-selected by isolation from an agarose gel. The isolated DNA fragments are ligated to linkers, DNA fragments which consist of small self-complementary sequences (B2 Fig. 1II). Microsatellite loci have to be enriched, because they are too rare to be found in sufficient numbers in a raw library. For enrichment, DNA fragments that contain a microsatellite motif, are bound to biotinylated oligonucleotides (probes) in a hybridisation reaction (B2 Fig. 1III). The biotin has a high binding affinity to streptavidin, which is attached to magnetic particles (B2 Fig. 1III). A magnet is used to capture these particles during washing steps: the DNA fragments with a microsatellite motif are attached to the magnet while other fragments are washed away. Elution steps yield a fraction which is enriched with DNA fragments containing microsatellite motifs. The DNA fragments of the microsatellite enriched fraction are PCR amplified, using oligonucleotides of the F-sequence of the linker as primers, and subsequently ligated into a vector (B2 Fig. 1IV).

After ligation the vectors are transformed into bacteria cells. Each bacterium only incorporates one plasmid, which may be multiplied within the cell. Moreover, millions of cells build one colony on an agar plate (B2 Fig. 1V). The single colonies can be isolated and cultivated. This way, single sequences are obtained in multiple copies within each bacterial clone. The cultured bacteria colonies provide the sequencing material: the inserts of single bacteria clones are used as templates for PCR amplification and these PCR products are prepared for the sequencing. After the sequencing, primers are designed in the flanking regions of the microsatellite loci. It is essential that the microsatellites do not occur too close to either end of the sequenced DNA fragment, so that there is enough space for the design of appropriate primers. Prior to routine use, the primers have to be tested and amplification conditions have to be optimised. After the optimisation of primer pairs in single reactions, it is desirable to find primer pairs that work in the same PCR (multiplex PCR) to minimize expensive reaction material and pipetting time.



### 3. Population genetics in freshwater zooplankton

Freshwater zooplankton organisms provide highly interesting systems to study population genetic dynamics, e.g. the implications of gene flow and genetic differentiation on population structure. Firstly, populations of freshwater zooplankton live in discrete habitat patches, i. e. lakes and ponds, which are separated from other habitats by physical barriers. Therefore, distances at which gene flow occurs between patches can be well defined. In that respect, they conform very well to the original metapopulation concept of Hanski 1998 (see above). Secondly, a number of zooplankton organisms, e.g. rotifers, cladoceran, and copepods, produce resting stages, which have various effects on species ecology and evolution. Diapausing eggs, for example, are known to be transported via several vectors such as wind, rain, or waterfowl and therefore possess a high potential for passive dispersal (Stemberger 1995; Bilton et al. 2001; Havel and Shurin 2004). Thus, gene flow should be enhanced compared to a species that lacks diapause, and consequently genetic differentiation between populations should be reduced. Further, diapausing eggs can build resting propagule banks, which are comparable to seed banks in plants (Hairston et al. 1995). Hence, genotypes can be stored in lake sediments for decades to more than 200 years (Hairston et al. 1995; Cáceres 1998) and the hatching of long-term diapausing eggs should have great effects on the watercolumn population (Bilton et al. 2001). Genetic drift and directional evolution is assumed to be slowed down, presumably leading to the maintenance of genetic variation within the population and enhanced effective population sizes (Bilton et al. 2001).

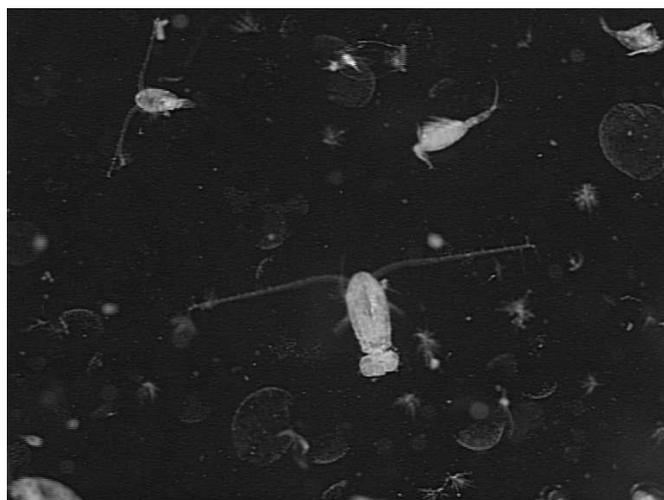


Figure 1. Egg-carrying *E. gracilis* female (centre) in a plankton sample from Lake Schöhsee, northern Germany.

Most population genetic studies in zooplankton organisms have concentrated on parthenogenetic rotifers and *Daphnia* so far, but investigations on obligately sexual freshwater species, especially copepods, are still rare (De Meester et al. 2002). Above investigations often revealed high genetic differentiation even among nearby populations, indicating restricted gene flow (Boileau et al. 1992; Lynch and Spitze 1994; De Meester 1996), and therefore contrasted with the broad opinion of zooplankton organisms being frequently dispersed over large distances. How this gene flow-dispersal paradoxon can be resolved is controversial. De Meester et al. (2002) suggested the “Monopolization hypothesis”, which assumes high dispersal rates in zooplankton organisms to be counteracted by “persistent founder effects” (Boileau et al. 1992), which are strengthened by the buffering effect of resting propagule banks and the consequences of rapid local adaptation. On the contrary, Bohonak and Jenkins (2003) concluded that passive dispersal is not generally frequent and widespread in freshwater invertebrates. Thus, they suspected that the monopolization hypothesis formulated by De Meester et al. (2002) could be applied for freshwater invertebrates in general (Bohonak and Jenkins 2003).

#### **4. The model system of the present study: *Eudiaptomus spec.***

The two calanoid copepods *Eudiaptomus graciloides* and *E. gracilis* provide an ideal system to study the genetic consequences of a life cycle with diapause. These sister species are morphological very similar and co-occur in a number of lakes in northern Europe (Kiefer 1978; Nauwerck 1980; Einsle 1993). However, they exhibit distinct life cycle strategies with diapausing eggs occurring in *E. graciloides* only (Fig. 2), (Santer et al. 2000). Thus, a genetic comparison of these two species should provide valuable information on the role of diapausing eggs as part of a life cycle on population structuring, and therefore genetic differentiation, gene flow, random drift and effective population sizes.

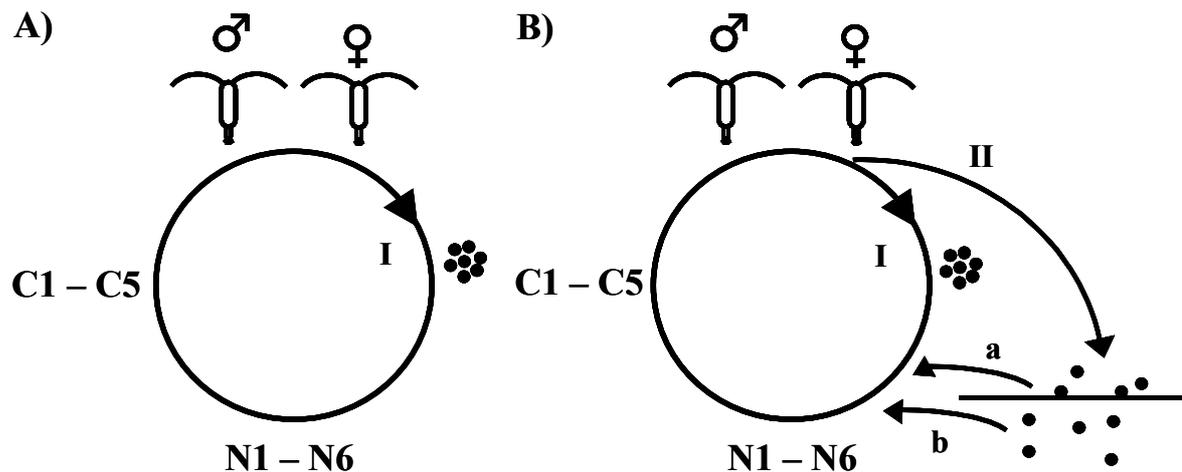


Figure 2. Simplified life cycles of *E. gracilis* (A) and *E. graciloides* (B). I: subitaneous (= immediately) hatching eggs, II: diapausing eggs that are produced in autumn/winter and sink to the lake sediment, a: diapausing eggs that hatch in the next growing season, b: long-term diapausing eggs that hatch in a subsequent growing season, N = naupliar stage, C = copepodid stage. Adult diapausing stages of *E. graciloides* were not depicted.

## 5. General goals and hypotheses

In the current thesis, the genetic implications of diapause were examined in a comparative study between *E. graciloides* and *E. gracilis*. The occurrence of diapausing eggs in *E. graciloides* only, led to the hypothesis that *E. graciloides* has an enhanced dispersal potential when compared to its sister species *E. gracilis*, which leads to higher gene flow and hence less genetic differentiation between populations. Further, it was hypothesized that *E. graciloides* has a larger effective population size compared to its congener *E. gracilis* due to the contribution of “old” genotypes from the resting propagule bank to the water column population in the former. In addition to being interested in the genetic structure of these two important members of zooplankton communities in European freshwaters, one goal of this work was to provide a new set of data in the recently discussed field of the dispersal gene flow paradoxon in freshwater organisms. The controversy of genetic differentiation versus dispersal capacity in freshwater zooplankton underlines the fact that our understanding of the general mechanisms that regulate dispersal, gene flow and therefore the population genetic structuring in zooplankton organisms is still poor, and results partly from a lack of empirical data to validate

competing hypotheses. Moreover, with this work the first population genetic analyses for planktonic freshwater copepods based on microsatellite polymorphism are presented.

## 6. Thesis outline

The current thesis contains four chapters, in which aspects of the population genetic structure of the calanoid freshwater copepods *E. graciloides* and *E. gracilis* were studied. These studies required the prior establishment of microsatellite libraries for both *Eudiaptomus* species, since no primer pairs were available to amplify microsatellite loci (Chapter 1). For Chapter II to IV brief introductions are given in this outline concerning the respective hypotheses and how they were tested. More detailed information is given in the single chapters. The fourth chapter resulted from a joint project with A. J. Bohonak, M. D. Holland, B. Santer, C. M. Kearns, and N. G. Hairston Jr. that was performed while the colleagues from Cornell University worked as guest researchers at the Max Planck Institute for Limnology. I contributed the data on microsatellite length polymorphism.

The thesis consists of the following chapters:

- I Identification and characterization of 10 microsatellite primers for the calanoid freshwater copepods *Eudiaptomus gracilis* and *Eudiaptomus graciloides* using enriched genomic libraries
- II A comparative population genetic study on calanoid freshwater copepods: Investigation of isolation-by-distance in two *Eudiaptomus* species with a different potential for dispersal
- III Small effective population sizes in two planktonic freshwater copepod species (*Eudiaptomus*) with apparently large census sizes
- IV The population genetic consequences of diapause in *Eudiaptomus* copepods

(I) A general introduction into microsatellite markers and the establishment of a genomic library was given above. Methodical details can be found in the chapter itself.

(II) In the second chapter I studied if populations of *E. graciloides* and *E. gracilis* were genetically isolated by distance and compared the results between species, in order to get information on the role of diapause (only occurring in *E. graciloides*) on population structuring. Specifically, I hypothesized for *E. graciloides* effective exchange of individuals over short distances and increasing genetic differentiation over long distances due to effective but decreasing dispersal via diapausing eggs. In contrast, I expected higher genetic differentiation, but no relationship with geographical distance for populations of *E. gracilis*, due to reduced gene flow and higher importance of genetic drift.

To detect a relationship between genetic and geographic distances the scale on which samples are taken should be crucial. On the one hand, if samples are only taken in a range where migration is high, a structuring of populations at greater geographical scales will remain undetected. On the other hand, substructures on small spatial scales may remain undetected if only large scales are studied. Since the range of dispersal and gene flow were not known in *Eudiaptomus* species *a priori* the population structure was investigated at three different spatial scales (100 km, 1000 km, 1340 km range), with copepod samples taken from lakes located on a longitudinal gradient spanning from northern Germany to Russia. A positive correlation of genetic and geographic distances in *E. graciloides* and high genetic differentiation of populations of *E. gracilis* at a comparable spatial scale would confirm our hypothesis of an enhanced dispersal potential of *E. graciloides*.

(III) One aim of this chapter was to relate effective population sizes ( $N_e$ s) to the ecological population sizes ( $N_c$ s).  $N_e$  is often much smaller compared to  $N_c$  in wildlife populations (Frankham 1995). However, one could assume large effective population sizes for *Eudiaptomus* species, because copepods were abundant in the water column of the study lake and passive transport within a relatively small water body should considerably mix-up plankton organisms. On the contrary,  $F_{IS}/F_{IT}$  values from previous studies (Chapter II) were relatively high for both species, indicating rather small  $N_e$ s. Secondly, it was hypothesized that *E. graciloides* has a greater  $N_e$  than *E. gracilis*, since genotypes are stored in the resting propagule bank in the former, which should lead to the maintenance of genetic variation within the water column population. To estimate  $N_e$ s, four temporal samples were taken from each *Eudiaptomus* species from a lake in northern Germany. Genetic data derived from microsatellite polymorphism were used to estimate  $N_e$ s with the

novel approach of Wang and Whitlock (2003), which considers migration as influencing factor in its calculations.

(IV) In chapter IV population genetic consequences of life-cycles with or without a diapause in *Eudiaptomus* species were investigated with mitochondrial DNA (mtDNA) sequences and microsatellite markers in a joint project (see above). MtDNA data can provide useful information on the population genetic structure over longer time scales because rates of evolution are slower than in microsatellite markers (Avice 1994). MtDNA variation in *E. graciloides* and *E. gracilis* was studied from three lakes in northern Germany by Nelson G. Hairston Jr., Cornell University, Ithaca, NY, and co-workers, and Andrew J. Bohonak from San Diego State University, California, that were also included in my study scheme. Genetic diversity within and pairwise genetic differences between population pairs were calculated for mitochondrial data as well as for parts of the length polymorphism data derived from my microsatellite investigations (Chapter II). It was assumed, that genetic diversity should be larger and genetic differentiation smaller in *E. graciloides* compared to *E. gracilis*, due to the occurrence of a resting propagule bank and diapausing eggs, respectively, in the former. A comparison between the results derived from two genetic markers that differ in their rates of evolution should give further important insights into the population genetic structure of both *Eudiaptomus* species.

Further, phylogeographic and Bayesian analyses were conducted for the data derived from mitochondrial DNA sequences, to get information on historical processes that shaped the observed population structures of *E. graciloides* and *E. gracilis*. Relative population sizes ( $N_e\mu$ ) together for populations from all three localities were estimated backwards in time and compared between both *Eudiaptomus* species under the assumption that, if mutation rates in both species are equally high, relative population sizes in *E. graciloides* should be greater compared to *E. gracilis*.

**Chapter I:****Identification and characterization of 10 microsatellite primers for the calanoid freshwater copepods *Eudiaptomus gracilis* and *Eudiaptomus graciloides* using enriched genomic libraries****Summary**

The calanoid copepods *Eudiaptomus gracilis* and *Eudiaptomus graciloides* are widespread in European lakes. We constructed enriched genomic libraries for both species in order to isolate and characterise microsatellite markers. The obtained seven polymorphic microsatellite loci for *E. gracilis* and three for *E. graciloides* are the first for any freshwater copepod. They display an expected heterozygosity ranging from 0.60 to 0.96 and 0.63 to 0.94 and observed heterozygosity ranging from 0.53 to 0.87 and 0.57 to 0.87, respectively. We were not able to cross-amplify the isolated loci across species, indicating long divergence among both congeneric species despite morphological similarity.

## Methodical report

Planktonic copepods are important members of the food chain in marine and freshwater habitats. The calanoid copepods *Eudiaptomus gracilis* and *Eudiaptomus graciloides* are common in European freshwater lakes and coexist in many lakes in northern Europe (Kiefer 1978; Nauwerck 1980). We constructed genomic libraries enriched for microsatellite motifs for both species. These identified microsatellites are the first for any freshwater copepod.

For each *Eudiaptomus* species, we extracted DNA from a pool of 350-500 copepods with the DNeasy Tissue Kit (Qiagen). 350-550 ng of genomic DNA were digested with the restriction enzymes Sau3a (Amersham Pharmacia Biotech), Csp6I and HpaII (MBI Fermentas) in single reactions. We ligated the digested DNA to linkers, which consisted of short self-complementary sequences (F: 5'-TTGCTTACGCGTGGACTC-3'; R: 5'-GAGTCCACGCGTAAGCAA-3' with a 5'-GATC-overhang (Sau3a-digested DNA), 5'-TA-overhang (Csp6I-digested DNA) or 5'-CG-overhang (HpaII-digested DNA)) (Edwards et al. 1996). Oligonucleotides of the F-sequence of the linker were used as primers in subsequent amplification steps. After the ligation we performed a PCR (see Table I.1, PCR 1 for exact reaction conditions) and purified PCR-products with the Qiaquick Purification Kit (Qiagen). We enriched these DNA fractions for microsatellite motifs with a biotin/streptavidin capture technique (Kijas et al. 1994) using magnetic particles (Dynal). We used 5'-biotinylated (CA)<sub>13</sub>- and (GA)<sub>13</sub>-oligonucleotides. They were 3'-dideoxy labeled to prevent amplification artifacts after the enrichment step (Koblizkova et al. 1998). The hybridization reaction (total volume 100 µl) was performed in 6x SSC at 60°C under gentle agitation for one hour. Washing steps (4 times 500 µl) were carried out for 5 minutes each with slow agitation in a heating block (50°C) with 2x SSC, 1x SSC, 0.5x SSC. All wash solutions contained 0.1 % SDS. The microsatellite enriched DNA fractions were eluted in 100 µl 0.2x SSC at 75°C and 80°C. We purified the eluted fractions (Qiagen) and amplified containing DNA fragments via PCR (Table I.1, PCR 2). Cloning was performed with Topo-TA-cloning-Kit (Invitrogen, according to manufacturer's instructions). We performed sequencing reactions (ABI PRISM BigDye Cycle Sequencing Kit) for DNA fragments of 353 clones (*E. gracilis*) and 426 clones (*E. graciloides*) and visualized sequences on an ABI-3100 capillary sequencer (Applied Biosystems). We were

able to design 33 and 18 primer pairs for *E. gracilis* and *E. graciloides*, respectively, using the Primer3-software (Rozen and Skaletsky 2000).

Table I.1. Contents and thermo profiles of polymerase chain reactions (PCRs) performed in a Whatmann-Biometra thermo cycler. PCR1: after digestion of genomic DNA, PCR2: after enrichment, PCR3: primer optimization; BSA: bovine serum albumin.

PCR	1	2	3
DNA ( $\mu$ l)	5.0	10.0	2.0
BSA (%)	0.1	0.1	0.1
Each dNTP ( $\mu$ M)	250.0	250.0	250.0
F-Primer ( $\mu$ M)	0.4	0.4	0.5
R-Primer ( $\mu$ M)	0.0	0.0	0.5
MgCl <sub>2</sub> (mM)	1.5	1.5	0.75/1.5/2.25
10x buffer <sup>1</sup> ( $\mu$ l)	5.0	10.0	2.0
Polymerase <sup>2</sup> (U)	2.5	5.0	1.0
Total volume ( $\mu$ l)	50.0	100.0	20.0
Initial Denaturation	94°C - 4min	94°C - 3min	94°C - 3min
Cycles	28	26	30
Denaturation	94°C - 30sec	94°C - 40sec	94°C - 50sec
Annealing	60°C - 1min	60°C - 1min	56.5°C - 30sec <sup>3</sup>
Elongation	72°C - 2min	72°C - 2min	72°C - 30sec
Final extension	72°C - 8min	72°C - 10min	72°C - 20min

<sup>1,2</sup> Promega.

<sup>1</sup> 1x buffer consisting of 10 mM Tris-HCl, 50 mM KCl, 0.1% TritonX-100.

<sup>3</sup> Annealing temperature was altered if necessary (see Table I.2).

Polymorphism of the identified microsatellite loci was tested in 20 individuals. We extracted DNA from single copepods using the DNeasy Tissue Kit (Qiagen). After optimization of the MgCl<sub>2</sub>-concentration (Table I.1, PCR 3), PCR products were separated and size scored with an ABI-3100 capillary sequencer (Applied Biosystems). Primer pairs that yielded polymorphic products were characterized in another 27 individuals per species (Table I.2). Deviation from Hardy-Weinberg equilibrium and linkage disequilibrium were tested with GENEPOP version 3.3 (Raymond and Rousset 1995) and FSTAT version 2.9.3 (Goudet 2001). For both species loci were in linkage equilibrium. In *E. gracilis* and *E. graciloides* three and one loci, respectively, were not in Hardy-Weinberg equilibrium, with observed heterozygosities lower than expected heterozygosities (Table I.2).

Table I.2. Polymorphic microsatellite loci for *Eudiptomus gracilis* (EGI) and *E. graciloides* (EGO). Repeat units and allele sizes are given for the cloned sequences. Expected heterozygosity ( $H_E$ ) and observed heterozygosity ( $H_O$ ) are calculated for 47 individuals from one population each. +, sequence interruptions between repeat motifs;  $T_a$ , annealing temperature; HEX, FAM, NED, 5' label. \*Significant deviation from Hardy-Weinberg equilibrium ( $p < 0.05$ ).

Locus name	Repeat Unit	Primer Sequence (5'- 3')	$T_a$ (°C)	MgCl <sub>2</sub> (mM)	Allele size (bp)	No. alleles	$H_E$	$H_O$	GenBank accession no.
EGI1	(CT) <sub>9+36</sub>	F: AATTTCTGGCGGTCATAGAGAA (hex) R: AGTAAAGGGCGATAGGGAGAAT	56.5	2.25	214	29	0.91	0.87	AY547395
EGI3	(CT) <sub>11+27+10</sub>	F: TCTGTTTATCAAATTGTCTCCAAGTC (hex) R: AAAAATAATTTTCGCACAGTGTGTTAC	56.5	1.5	199	39	0.96	0.83*	AY547396
EGI8	(CT) <sub>33+3+11</sub>	F: GGTTTGCATCCTCTCTGTCTCT (fam) R: TGCTGGCAGAATTTCAAAGATA	56.5	2.25	194	31	0.93	0.81	AY547397
EGI12	(CT) <sub>13</sub>	F: AAATTTCCAACCATTTCCCTCT (hex) R: CGCCTATCATAATGGGCTAAAC	56.5	2.25	203	15	0.84	0.55*	AY547398
EGI13	(CT) <sub>3+7+11</sub>	F: AGCACAAATAGCTCATGACTCG (fam) R: TCAAAAATCTCATGACACAGTGC	56.5	2.25	180	27	0.89	0.68*	AY547399
EGI17	(CTGT) <sub>5+3</sub>	F: GTATTGATTTGTCTTTGTGTTTTCCA (hex) R: AGAGTTGCACAGATAAACTCTTCGTA	56.5	2.25	234	6	0.60	0.53	AY547400
EGI35	(CT) <sub>15+3</sub>	F: ATTAGAGTGTAGAGCTGGCTGGA (fam) R: TTCTCAGGAGGCTAAAGGAGAATA	56.5	1.5	202	17	0.84	0.79	AY547401
EGO2	(GA) <sub>6+5+9+5+3+7</sub>	F: CATTCAATTTGTACCTGAAGGTGAG (hex) R: ATTAGCGACTCAGCCGTTTTAC	56.5	1.5	271	32	0.94	0.87	AY547394
EGO7	(GA) <sub>4+8</sub> (CA) <sub>2</sub> (GA) <sub>7+8</sub>	F: CCAGTACATAAGGATAAGTACAACCAG (fam) R: TCACCAAGATGACCTCTTAAAACA	56.5	2.25	218	31	0.94	0.75*	AY547392
EGO10	(CA) <sub>8+2</sub>	F: CCGAAATGGTGGGAAATACA (ned) R: CTGCCTCCCAGGTAGTTTTG	55.5	1.5	128	8	0.63	0.57	AY547393

We performed cross-amplifications with all 33 identified *E. gracilis* primer pairs and the three working *E. graciloides* primer pairs. PCRs were carried out for six to 10 copepods (Table I.1, PCR 3). Only few primer pairs yielded inconsistent PCR products in the other species. Therefore, no primer pair can be applied for cross-amplifications.

We experienced several difficulties to find working primer pairs for *E. graciloides* only. Sequence alignments revealed similarity in the flanking regions, suggesting that entire genomic regions with their respective microsatellites were duplicated. In support of this hypothesis four of our tested primer pairs amplified more than two alleles. Thus, we assume that the total of repetitive sequences and their flanking regions often occur multiply in the genome of *E. graciloides*. These insights into parts of the genome of *E. graciloides* and the unsuccessful cross-amplifications suggest that these congeneric species differ widely in their genomic structure.

### **Acknowledgements**

Many thanks to F. Alberto for help with the enrichment protocol, S. Carstensen and C. Schmuck for laboratory assistance, G. Nies for collaboration and N. G. Hairston Jr., C. M. Kearns and S. Bogdanowicz for their support when this project was born. TBHR was partly supported by DFG (Re 1108/4).

**Chapter II:****A comparative population genetic study on calanoid freshwater copepods: Investigation of isolation-by-distance in two *Eudiaptomus* species with a different potential for dispersal****Summary**

We examined patterns of genetic isolation-by-distance in two sister species of freshwater copepods, *Eudiaptomus graciloides* and *Eudiaptomus gracilis*, using the polymorphism of microsatellite markers. We assumed an enhanced dispersal potential of the species possessing diapausing eggs (*E. graciloides*) compared to *E. gracilis* with no diapause. On a longitudinal gradient spanning from northern Germany to Russia, we sampled 17 *E. graciloides* populations and 16 *E. gracilis* populations in order to investigate the scale dependence of isolation-by-distance from 0.4 - 1340 km distance. In the 100 - 1000 km range no isolation could be detected in either *Eudiaptomus* species, suggesting either similar moderate gene flow over large geographic distances or persistence of historic patterns after post glacial recolonization. Our hypothesis was partly supported at the small-scale level (100 km range) where differences in population structure among species existed. We found isolation-by-distance in *E. graciloides*, the species with diapausing eggs, whereas *E. gracilis* exhibited a pattern of persistent founder effects. Further, in *E. graciloides* we found significant isolation-by-distance in the 1340 km range with significantly reduced gene flow at distances >1000 km. In *E. gracilis*, isolation-by-distance at distances >1000 km could not be detected with our data because this species was not found in the Russian lakes. Our results suggest that short- and long-distance dispersal in *Eudiaptomus* are due to different processes, with diapausing eggs only being advantageous for short-distance dispersal. We also argue that different spatial scales must be sampled to understand the geographic partitioning of genetic variance at marker loci.

## Introduction

The study of patterns of genetic variation and differentiation at neutral marker loci has considerably advanced our knowledge on gene exchange in many species, especially those where direct measurement of the movement of individuals is impossible. In passively dispersing freshwater zooplankton, such studies often revealed high genetic differentiation even among nearby populations, indicating restricted gene flow (Boileau et al. 1992; Lynch and Spitze 1994; De Meester 1996). These results, achieved with genetic approaches, were at odds with the broad opinion that zooplankton organisms are frequently dispersed over large distances. At present, there is no general consensus to explain the discrepancy between genetic differentiation and dispersal capacity (Boileau et al. 1992; De Meester et al. 2002; Bohonak and Jenkins 2003). Boileau et al. (1992) suggested that these discrepancies were due to “persistent founder effects”, where zooplankton populations, which developed from a few individuals and grew rapidly after colonization, remained stable against the invasion of newly arriving individuals. De Meester et al. (2002) expanded this theory to the “Monopolization hypothesis” by including the buffering effect of resting propagule banks and the consequences of rapid local adaptation. Both these factors should increase the persistent founder effect. The above hypotheses were based on high dispersal capacity but reduced gene flow. On the contrary, Bohonak and Jenkins (2003) concluded that passive dispersal is not generally frequent and widespread in freshwater invertebrates. Thus, they suspected that the monopolization hypothesis formulated by De Meester et al. (2002) could be applied for freshwater invertebrates in general (Bohonak and Jenkins 2003). This controversy of genetic differentiation versus dispersal capacity underlines the fact that our understanding of the general mechanisms that regulate dispersal and gene flow -and therefore the population genetic structuring in zooplankton organisms- is still poor and partly is the result of a lack of empirical data to validate competing hypotheses. The motivation of the present work was to provide empirical data using high-resolution microsatellites for measuring genetic differentiation, and hence genetic structure across a range of spatial scales.

Studies on population genetics of zooplankton organisms have so far concentrated on parthenogenetic cladocerans and rotifers, but investigations on obligately sexual freshwater species, especially copepods, are still rare (De Meester et al. 2002). In this study we investigated the population genetic structure of two calanoid copepods *Eudiaptomus*

*graciloides* and *E. gracilis*. Both species are closely related but exhibit distinct life cycle patterns with diapausing eggs occurring only in *E. graciloides* (Santer et al. 2000). Diapausing eggs are known to be transported via several vectors such as wind, rain, or waterfowl and therefore possess a high potential for passive dispersal (Stemberger 1995; Bilton et al. 2001; Havel and Shurin 2004). Thus, we hypothesized an enhanced dispersal potential of *E. graciloides* when compared with its sister species *E. gracilis*. The between-species variation on the one hand and their relatedness and co-existence in many European freshwaters on the other hand (Kiefer 1978; Nauwerck 1980) provided an excellent system to investigate the population genetic structure of planktonic copepods and to compare these between species with different dispersal capacities. For *E. graciloides*, we hypothesized effective exchange of individuals over short distances and increasing genetic differentiation over long distances due to effective but decreasing dispersal via diapausing eggs. In contrast, we expected higher genetic differentiation, but no relationship with geographical distance for populations of *E. gracilis* due to reduced gene flow and higher importance of genetic drift.

## Methods

### *Study species*

The calanoid copepods *E. graciloides* and *E. gracilis* are small zooplankton organisms (1.0-1.5 and 1.0-2.0 mm body length, respectively; Einsle 1993) that are widely distributed in continental European freshwater ecosystems (Kiefer 1978; Einsle 1993). *E. graciloides* and *E. gracilis* co-occur in a number of lakes in northern Europe (Kiefer 1978; Nauwerck 1980). Both species are phylogenetically closely related, morphologically similar and they reproduce obligately sexual (Einsle 1993).

### *Sampling sites and sampling scheme*

Zooplankton samples were taken from lakes located on a longitudinal gradient spanning from northern Germany to Russia to look for scale dependence of genetic differentiation due to isolation-by-distance (Fig. II.1). The landscape in this area was formed after the last ice age about 11000 years ago, and lakes located here have the same age and geological history. We chose seven regions within this area on a logarithmic spatial scale and sampled populations of two lakes per region in order to investigate population genetic structure on a

large spatial scale. Distances between lakes varied from a few hundred meters up to a maximum of 1340 km. Additionally, in northern Germany (Region 1, Fig. II.1) we sampled a total of six populations of each *Eudiaptomus* species. These consisted of four populations from lakes where both species co-occurred and two populations of each species from lakes where the other *Eudiaptomus* species was missing. For our investigation of genetic differentiation on a small spatial scale (100 km range), we used the data of the above populations together with the data from populations of that lake in Region 2 where both species co-existed (Regions 1 and 2 = 100 km range; Fig. II.1). Names and positions of studied lakes are listed in Appendix 1. We took zooplankton samples with a plankton-net and stored them in 70% alcohol until further processing.

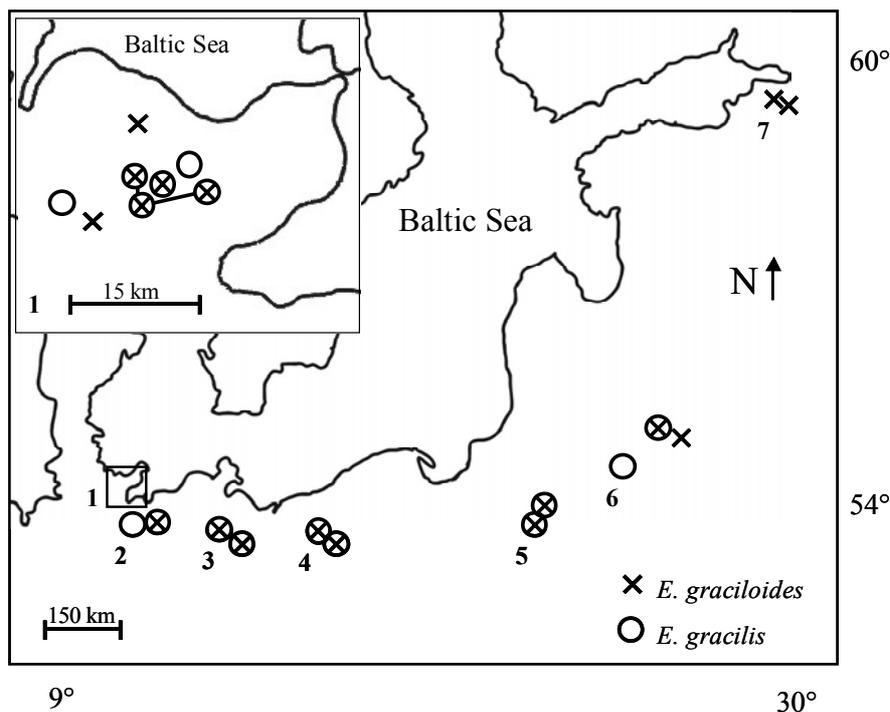


Figure II.1. Sampling sites of *Eudiaptomus* in the southern range of the Baltic Sea. Numbers indicate the sampling regions: Regions 1-3 are located in Germany, Regions 4 and 5 in Poland, Region 6 in Lithuania, and Region 7 in Russia. Regions 1-7 = 1340 km range, Regions 1-6 = 1000 km range, Regions 1 and 2 = 100 km range. The inset shows detail of Region 1. Solid lines between lakes indicate a connection by waterway.

#### *DNA extraction and microsatellite analysis*

For the population genetic studies on *Eudiaptomus* species we used the polymorphism displayed by 10 microsatellite markers that we had developed previously (Chapter I):

seven polymorphic loci for *E. gracilis* (EGI1, 3, 8, 12, 13, 17, and 35; Genbank accession numbers AY547395-547401) and three polymorphic loci for *E. graciloides* (EGO2, 7, and 10; Genbank accession numbers AY547392-547394). These markers represented a total of 127 alleles of *E. graciloides* and 314 alleles of *E. gracilis* in the studied populations.

For the genetic analysis, we isolated 45 females per lake and species from the zooplankton samples. We chose only females for our investigation, because it was difficult to distinguish between males of the two species. Prior microsatellite analysis of males and females originating from lakes where only a single *Eudiaptomus* species occurred did not show any sex-specific differences in the used markers.  $F_{ST}$ -values calculated between males ( $n = 46$ ) and females ( $n = 46$ ) of one population were not significantly different from zero in a total of two populations of each species that we investigated. In this study we extracted DNA from the isolated copepod females using the Invisorb DNA Tissue HTS 96-Kit/C from Invitex. We amplified microsatellite loci in polymerase chain reactions (PCRs). Loci of *E. gracilis* were amplified together in multiplex PCRs as follows: EGI1 + 13 + 17, EGI8 + 12, and EGI3 + 35 in a total reaction volume of 20  $\mu\text{L}$  using 2  $\mu\text{L}$  template DNA. Loci of *E. graciloides* were amplified in single reactions of 10  $\mu\text{L}$  with 1  $\mu\text{L}$  template DNA. Further PCR contents were the following: 0.1% bovine serum albumin (BSA), 250 pmole  $\mu\text{L}^{-1}$  each dinucleotide triphosphate (dNTP), 1  $\mu\text{L}$  forward primer (5 pmole  $\mu\text{L}^{-1}$  stock) and 1  $\mu\text{L}$  reverse primer (5 pmole  $\mu\text{L}^{-1}$  stock), for EGI3 1.5  $\mu\text{L}$ , and for EGI17 2  $\mu\text{L}$  of each primer (5 pmole  $\mu\text{L}^{-1}$  stock), 1x buffer (Promega), 1U Polymerase (Promega), 2.25 mmole  $\mu\text{L}^{-1}$   $\text{MgCl}_2$  (EGI1, 8, 12, 13, 17, EGO7), and 1.5 mmole  $\mu\text{L}^{-1}$   $\text{MgCl}_2$  (EGI3, 35, EGO 2, 10), respectively. An initial denaturation step at 94°C for 3 minutes was followed by 30 cycles of 94°C for 50s, 56.5°C (55.5°C for EGO10) for 30s, 72°C for 30s and a final extension step of 72°C for 20 minutes. Amplified microsatellite loci were electrophoretically separated and size scored on an ABI-3100 capillary sequencer (Applied Biosystems) using Rox 350 as internal standard. The data were analyzed with GENESCAN and GENOTYPER (Applied Biosystems). We excluded from further analysis individuals that resulted in no PCR-product for any primer. If only single primer pairs did not amplify, we repeated PCRs up to two times using only one failed primer pair per reaction, with a total reaction volume of 20  $\mu\text{L}$ . If necessary we increased cycling steps from 30 to 34. Where PCR-products were still missing, we interpreted these as homozygote null alleles.

*Data analysis*

## Null alleles:

Null alleles are alleles that do not amplify during PCR because of mutation events changing the DNA sequence in the primer binding region, which causes the primer no longer to bind to the DNA during the PCR (Callen et al. 1993). This results in either no PCR product, if the null allele is homozygote, or in false homozygote individuals, if the locus is heterozygote. Thus, an excess of homozygote individuals, as found for both *Eudiptomus* in our investigation (Appendix 2), could be caused by the presence of null alleles or by a real biological phenomenon. Available software for estimating null allele frequencies calculate them on the basis of heterozygosities, making it difficult to decide whether inbreeding or the presence of null alleles causes the excess of homozygote genotypes. Therefore, to assess the frequency of null alleles in our study we did not want to rely on calculations of null allele frequencies only, but rather on a combination of information about calculated null allele frequencies (Brookfield 2, Microchecker 2.2.1; Van Oosterhout et al. 2004),  $F_{IS}$ -values, and frequencies of nonamplified samples (Table II.1).

Further, we tested for correlations between the above parameters (means of information from microsatellite loci for each population) and their geographic occurrence to examine whether the frequency of null alleles was biased with distance. Populations in northern Germany were defined to be located at distance zero with all other populations oriented along an east-west transect.

## Hardy-Weinberg equilibrium and inbreeding:

We calculated the population genetic parameters observed heterozygosity ( $H_O$ ) and expected heterozygosity ( $H_E$ ), linkage disequilibrium and F-statistics with the software GENETIX v. 4.01 (Belkhir et al. 1998).  $F_{IS/IT}$ -values for *E. gracilis* were only calculated with information from three primer pairs, because it was likely that null alleles were abundant in the remaining four pairs (see results). We performed a Hardy-Weinberg equilibrium test in GENEPOP v. 3.3, which estimates exact  $p$  values using the Markov chain method (Raymond and Rousset 1995).

Isolation-by-distance and species comparison:

We performed an isolation-by-distance investigation for *Eudiaptomus* species, including geographic distances ranging from 0.4 to >1000 kilometres. The purpose was to address exchange of individuals on distinct spatial scales and to compare the population genetic structure between both species considering their different abilities for dispersal.

To test for isolation-by-distance we calculated pairwise  $F_{ST}$ -values with GENETIX v. 4.01 (Belkhir et al. 1998), which calculates the estimator  $\theta$  (Weir and Cockerham 1984). We plotted these estimates of pairwise population differentiation against the geographic distances separating each pair of populations. With the help of maps, geographic distances were measured as the shortest direct lines. Subsequently, we tested for a significant relationship between the matrix of pairwise  $F_{ST}$ -values ( $\theta$ ) and the matrix of geographic distances performing Mantel tests of matrix correlations, calculated as reduced major axis (RMA) regression with the programme IBD v. 1.52 (Bohonak 2002).

In order to investigate the population genetic structure on a large geographic scale, we performed Mantel tests including data from all 17 populations of *E. graciloides* and all 16 populations of *E. gracilis*, respectively. We used data of 12 populations where species co-occurred for species comparison. Matrices of geographic and genetic distances including all populations are available in Appendix 3. To compare the pairwise genetic differentiation between species, we performed a Mantel test between the matrix of pairwise  $F_{ST}$ -values for *E. graciloides* and the matrix of pairwise  $F_{ST}$ -values for *E. gracilis* (12 populations, programme IBD v. 1.52; Bohonak 2002). An additional pairwise *t*-test was performed with the programme Statistica v. 5 (StatSoft Inc. 1995).

For short geographic distances, we used a data subset that contained information about seven investigated populations of each species from lakes located in northern Germany (Regions 1 and 2 = 100 km range; from Region 2 we included only populations of one lake, where both species co-existed; Fig. II.1). Three of these lakes were connected by waterways providing three out of 21 pairwise comparisons on genetic differentiation. We neglected stream connections and measured geographic distances as shortest direct lines, since the number of data points were low and populations were partly dispersal-limited (see results). Mantel tests were performed, this time comparing the geographic and genetic

distance matrix for each species separately using the software IBD v. 1.52 (Bohonak 2002).

## Results

### *Null alleles*

Three of the seven primer pairs for *E. gracilis* provided low  $F_{IS}$ -values as well as low calculated frequencies of null alleles and low frequencies of nonamplified samples (loci EGI1, 8, 17; Table II.1). For the remaining four primer pairs the occurrence of null alleles was very likely. Therefore, we calculated  $F_{IS/IT}$ -values for population genetic analysis with information from the loci EGI 1, 8, and 17 only. For assessing isolation-by-distance we included the information of all primer pairs, because the occurrence of null alleles was not biased with distance (Brookfield 2,  $r = 0.05$ , not significant;  $F_{IS}$ -values,  $r = 0.15$ , not significant; frequencies of nonamplified samples,  $r = -0.02$ , not significant). Thus, calculations of pairwise  $F_{ST}$ -values should be unaffected by the occurrence of null alleles.

Table II.1. Null alleles in *Eudiptomus* species. Null allele 1: Frequency of nonamplified samples. Null allele 2: Frequency of calculated null alleles (Brookfield 2) with programme Microchecker 2.2.1 (Van Oosterhout et al. 2004). Std. dev., standard deviation; EGO: *E. graciloides*; EGI: *E. gracilis*. Inbreeding coefficients ( $F_{IS}$ ) and calculated frequencies of null alleles are shown as mean of all 17 *E. graciloides* populations and 16 *E. gracilis* populations, respectively. Primer pairs that were used for F-statistics in *E. gracilis* are printed in bold.

Primer	Null Allele 1	Null allele 2	Std. dev.	$F_{IS}$	Std. dev.
EGO10	0.005	0.064	0.079	0.103	0.035
EGO7	0.008	0.150	0.103	0.274	0.036
EGO2	0.043	0.184	0.145	0.260	0.028
<b>EGI1</b>	<b>0.003</b>	<b>0.061</b>	<b>0.065</b>	<b>0.120</b>	<b>0.033</b>
EGI3	0.033	0.207	0.087	0.267	0.028
<b>EGI8</b>	<b>0.003</b>	<b>0.057</b>	<b>0.050</b>	<b>0.121</b>	<b>0.021</b>
EGI12	0.043	0.362	0.140	0.577	0.036
EGI13	0.029	0.240	0.109	0.325	0.032
<b>EGI17</b>	<b>0.006</b>	<b>0.052</b>	<b>0.050</b>	<b>0.110</b>	<b>0.030</b>
EGI35	0.010	0.259	0.141	0.446	0.059

For *E. graciloides* we did not exclude information of any primer pairs for the calculations of  $F_{IS/IT}$ -values, but noted that null alleles might be abundant (Table II.1). Calculated frequencies of null alleles were not biased with geographic distance in *E. graciloides*

(Brookfield 2,  $r = 0.46$ , not significant).  $F_{IS}$ -values and nonamplified samples were positively correlated with increasing geographic distance to northern Germany ( $F_{IS}$ -values,  $r = 0.56$ ,  $p < 0.05$ ; frequencies of nonamplified samples,  $r = 0.49$ ,  $p < 0.05$ ; Fig. II.1). These correlations were driven by one of the Russian samples (Lake 7.1) which had high null allele frequencies (data not shown).  $F_{IS}$ -values and non-amplified samples were not correlated with distance when this population was excluded from the analysis ( $F_{IS}$ -values,  $r = 0.35$ , not significant; frequencies of nonamplified samples,  $r = 0.15$ , not significant). Thus, in *E. graciloides* null alleles were not correlated with distance; rather, more null alleles existed in one of the Russian populations (Lake 7.1) compared with the other investigated populations.

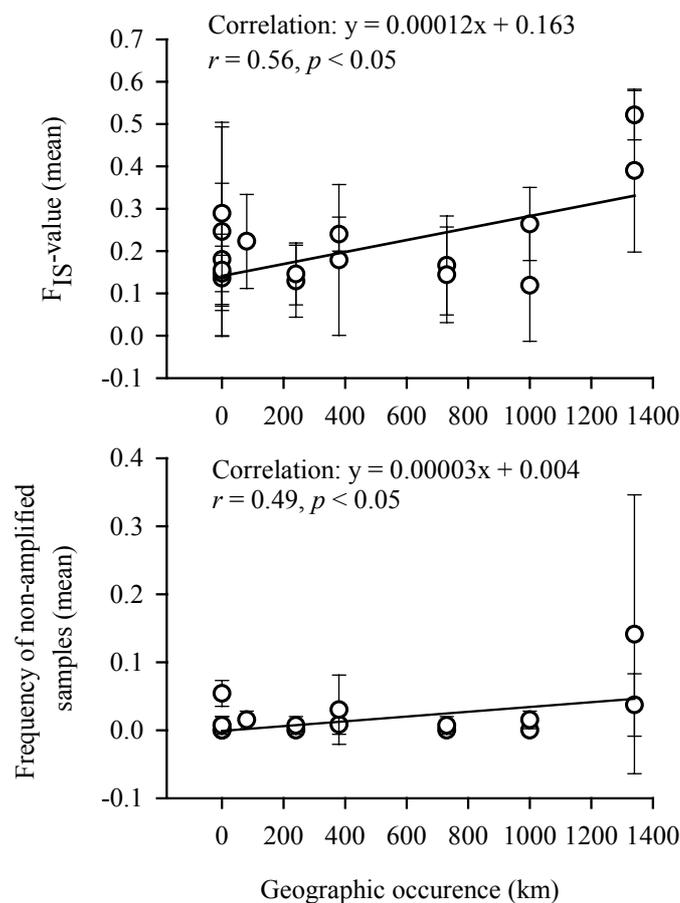


Figure II.2.  $F_{IS}$ -values and frequencies of nonamplified samples of *E. graciloides* populations as indicators for null-alleles along the sampled transect from northern Germany to Russia. Populations are shown in order of their geographic location from west (W) to east (E): distance 0 = northern Germany populations, distance 1340 = Russian populations. On the y-axis the means of  $F_{IS}$ -values and frequencies of nonamplified samples, respectively, of each microsatellite loci (EGO2, 7, 10) are shown. Error bars: standard deviation.

*Hardy-Weinberg equilibrium and inbreeding*

All loci were in linkage equilibrium for both species. Expected heterozygosities ( $H_E$ ) and observed heterozygosities ( $H_O$ ) for all loci and all populations, as well as the inbreeding coefficients ( $F_{IS}$ -values) and deviations from Hardy-Weinberg expectations, are given in Appendix 2. Calculations of F-statistics over all populations resulted for *E. gracilis* in  $F_{IS}$ -values of 0.12 (0.11 – 0.12),  $F_{IT}$ -values of 0.16 (0.13 – 0.23), and  $F_{ST}$ -values of 0.05 (0.01 – 0.13) based on information of primer EGI1, 8 and 17. For *E. graciloides* we found  $F_{IS}$ -values of 0.22 (0.10 – 0.27),  $F_{IT}$ -values of 0.26 (0.17 – 0.29), and  $F_{ST}$ -values of 0.05 (0.03 – 0.08).

*Isolation-by-distance and species comparison*

On a large spatial scale (1340 km), a high proportion of the pairwise genetic differentiation between all 17 *E. graciloides* populations could be explained by their geographic distances ( $r = 0.63$ ,  $p < 0.001$ ; Fig. II.3A). However, the scatter plot pointed to differences in the relationship of genetic and geographic distance. Genetic variation increased with geographic distance for populations located in a 100 km range, but no obvious relationship between genetic and geographic distance existed for populations separated by ~100 to 1000 km. We observed a further increase in genetic differentiation for populations separated by more than 1000 km (Fig. II.3A). In contrast, we could not find a significant correlation between genetic and geographic distances for the sister species *E. gracilis* (total data set, Fig. II.3B;  $r = 0.26$ ,  $p = 0.056$ ). Note that information for populations separated by more than 1000 km are missing in this data set, because *E. gracilis* was absent in samples from Russian lakes.

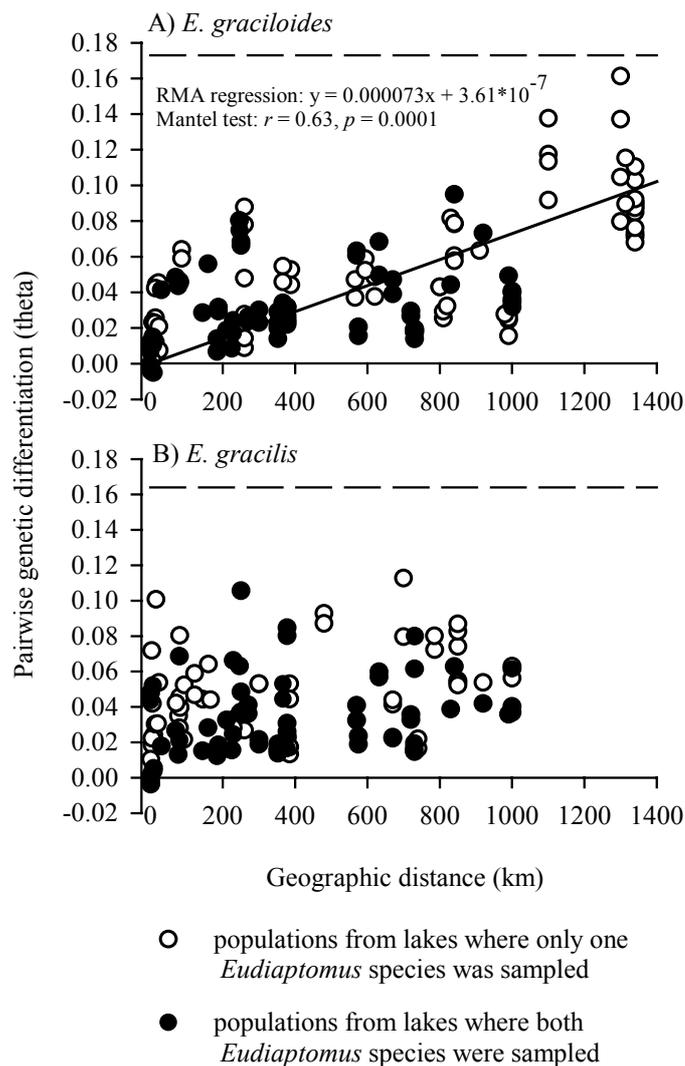


Figure II.3. Isolation-by-distance over the complete range (1000-1340 km). Shown are all pairwise comparisons among 17 populations of (A) *E. graciloides* and 16 populations of (B) *E. gracilis*. Solid line: RMA regression calculated with IBD v 1.52 for all populations (Bohonak 2002). Dashed lines: expected maximal value of genetic differentiation (Hedrick 1999).

To compare the population genetic structure between *E. graciloides* and *E. gracilis* at a geographical range of 0.4 - 1000 km, we reduced the data set to the 12 populations of both species originating from the same lakes. Neither populations of *E. graciloides* nor of *E. gracilis* were isolated by distance (*E. graciloides*,  $r = 0.32, p = 0.065$ ; *E. gracilis*,  $r = 0.25, p = 0.118$ ; Fig. II.3 filled circles). The pairwise  $F_{ST}$  -values for these co-occurring populations of both species were positively correlated ( $r = 0.45, p < 0.05$ ; Fig. II.4) and did not differ between species (paired  $t$ -test:  $t = -0.438, df = 130, p = 0.662$ ).

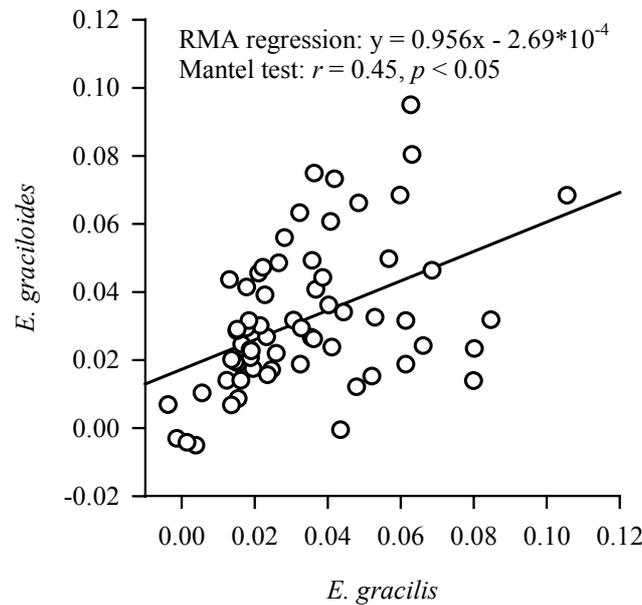


Figure II.4. Cross-correlation of pairwise genetic differentiation (theta) between 12 populations of *E. graciloides* and 12 *E. gracilis* populations originating from the same lakes (corresponding to filled circles in Fig. II.3). Solid line: RMA regression calculated with IBD v 1.52 (Bohonak 2002).

At a small spatial scale (<100 km distance), we found differences between both species. At this scale we observed higher variability for *E. gracilis* than for *E. graciloides* (Fig. II.3A, B). For example, at a 50 km distance, pairwise  $F_{ST}$ -values between *E. gracilis* populations ranged from 0 to 0.1 (Fig. II.3B), whereas pairwise  $F_{ST}$ -values between *E. graciloides* populations only attained 0 to 0.05 (Fig. II.3A). Figure II.5 shows the pairwise  $F_{ST}$ -values from seven populations of lakes located in northern Germany plotted against the geographic distances (100 km range). We found a significant relationship between genetic differentiation and geographic distance in *E. graciloides* ( $r = 0.81, p = 0.004$ ), but not in *E. gracilis* (Fig. II.5;  $r = -0.08$ , not significant), although the latter species had a larger range of variation. Three out of the 21 pairwise genetic comparisons resulted from population pairs of lake locations that were connected by streams (filled circles, Fig. II.5). Pairwise  $F_{ST}$ -values for one population pair from interconnected lakes (Lake 1.3 and 1.4) differed significantly (geographic distance = 8.5 km:  $F_{ST} = 0.01$  ( $p < 0.05$ ), and  $F_{ST} = 0.006$  ( $p < 0.05$ ) for *E. graciloides* and *E. gracilis*, respectively, see Appendix 3,  $p$ -values not shown). Since only few data points were affected and populations were already partly dispersal limited, we neglected lake connectivity as influencing factor in our study.

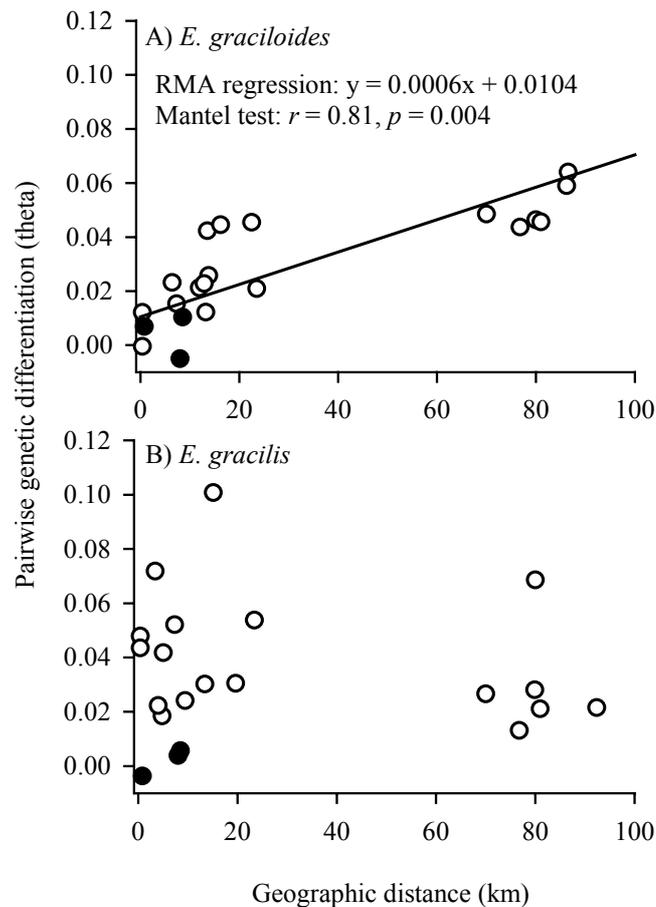


Figure II.5. Isolation-by-distance (100 km range). Pairwise comparisons of seven populations of (A) *E. graciloides* and (B) *E. gracilis*. Solid line: RMA regression calculated with IBD v 1.52 (Bohonak 2002). Filled circles: pairwise comparisons between populations from interconnected lakes.

## Discussion

Isolation-by-distance is expected between populations if gene flow is prevalent but dependent on geographical distance, because the exchange of individuals should be more frequent among nearby populations than among populations located further away (e.g., Wright 1946; Slatkin 1993; Hutchison and Templeton 1999). However, we predicted to observe no isolation-by-distance when gene flow is very low or if current patterns of gene flow are largely independent of any spatial dependence of the vector for propagules. Our results were, at least in part, consistent with above hypotheses. We could only detect isolation-by-distance for *E. graciloides* producing diapausing eggs, but not for *E. gracilis*. Genetic and geographic distances of populations of *E. graciloides* were positively

correlated on a large spatial scale (1340 km) and on a short spatial scale (< 100 km). Surprisingly, we could not find isolation-by-distance in the 1000 km range in either *Eudiptomus* species, but rather an irregular pattern of similar and genetically differentiated populations. This pattern might be due to past imprints of recolonization patterns after the last ice age, or, at least in the case of *E. graciloides*, to a different process of genetic exchange that was unaffected by distance. However, on a small spatial scale species-specific differences did occur. In *E. gracilis* genetic differentiation between populations was partially higher compared with *E. graciloides* and seemed to be randomly distributed with respect to geographic distance. Thus, we found our hypothesis to be true for a smaller geographic scale than originally expected. Due to our nested sampling design, we were able to detect these species-specific difference, and to find for *E. graciloides* that isolation-by-distance (100 km range) was hidden in a larger pattern of genetic differentiation that seemed to be unaffected by geographic distance (1000 km range). These results underline the importance of both the chosen sampling scheme for isolation-by-distance investigations in general and the necessary care that must be taken in the interpretation of studies addressing only a single spatial scale.

We can only speculate on the possible causes for the apparent dissimilarities in the patterns of genetic differentiation we observed at different spatial scales. The irregular patterns of population differentiation in both *Eudiptomus* between distances of 100 and 1000 km may indicate a similar moderate exchange of individuals over large distances for both species, for example through migratory waterfowl (Charalambidou and Santamaría 2002; Figuerola et al. 2003). Alternatively, this pattern does not reflect contemporary gene flow, but rather an average historic differentiation after a rapid recolonization after the last ice age. Within this matrix of moderate population differentiation due to founder effects, cells of isolation-by-distance developed in *E. graciloides*, whereas small- and large-scale patterns were congruent in *E. gracilis*. If we assume contemporary gene flow as shaping the observed patterns, it implies that the occurrence of diapausing eggs in *E. graciloides* does not enhance its dispersal ability in terms of long-distance dispersal compared to its sister species *E. gracilis*.

On small geographic scales (100 km range), however, populations of *E. gracilis* differed substantially in their genetic structure even at nearby locations. This pattern of high genetic

differentiation between nearby populations is known in a variety of freshwater ecosystems (Boileau et al. 1992; De Meester 1996; Nies and Reusch 2005). De Meester et al. (2002) explained such patterns with persistent founder effects (Boileau et al. 1992), which are enhanced by resting propagule banks and local adaptation. Persistent founder effects should be largely independent of geographical distance and should be stronger in *E. graciloides* compared to *E. gracilis* due to the existence of resting propagule banks (De Meester et al. 2002). Thus the observed species-specific differences at short distances were probably caused by differences in short-distance dispersal and current gene flow, respectively, rather than by population history alone. Possible vectors for short distance dispersal may include both wind and rain (Cáceres and Soluk 2002; Cohen and Shurin 2003) or waterfowl (Green et al. 2002). There is some evidence that at small distances zooplanktonic organisms are in fact highly effective dispersers; however, animals may be a less important dispersal vector than wind (Cáceres and Soluk 2002; Cohen and Shurin 2003).

For the total investigative range of 1340 km, we were not able to draw comparisons about genetic differentiation between the sister species, since we could only find individuals of *E. graciloides* in the Russian lakes. The increasing genetic differentiation of *E. graciloides* on distances above 1000 km indicated restricted gene flow that might be caused by restricted dispersal. However, dispersal events must not necessarily lead to successful gene flow. The dispersers are not always as well adapted to the environmental conditions of the new habitat as the residents; therefore, they are not consequently able to reproduce or survive (De Meester et al. 1996; Bohonak and Jenkins 2003). In our study population pairs with geographical distances greater than 1000 kilometers always included one of the two investigated *E. graciloides* populations sampled in the vicinity of St. Petersburg, Russia. These populations were not only located at the most easterly regions of this study, but also at a higher latitude (59-60 °N) than all other investigated populations. At high latitudes, the summer season is shorter and lakes are covered with ice for at least 6 months of the year (Alekseev, V. pers. comm.). Local adaptations are an expected phenomenon if populations are faced with very distinct environmental conditions. *E. graciloides*, for example, is only able to exhibit one new generation per year at this high latitude of 60°, whereas populations from lower latitudes are able to produce two or more generations per year (Nauwerck 1963; Ekman 1964; Bosselmann 1975). Furthermore, the timing of

reproduction and diapause should match with the seasonal changes in the environment to ensure survival (Hairston and Olds 1984; Santer 1998). Thus, gene flow between Russian and northern German populations might additionally be restricted by selection and local adaptation.

An alternative explanation for the identified pattern of genetic differentiation over the whole geographic range in *E. graciloides* might be source and sequence of postglacial recolonization. Different glacial refugia existed during the last ice age in Europe, from where species recolonized habitats after deglaciation (Hewitt 1996; Pamilo and Savolainen 1999). Lakes in northern Germany might be recolonized by *E. graciloides* spreading from other glacial refugia than those colonizing the Russian lakes. However, information from other molecular markers that allow a coalescence based reconstruction of colonization, such as SNPs (Morin et al. 2004) or mtDNA (Hewitt 1999), would be necessary to adequately answer questions about different recolonization routes in *Eudiaptomus*. We have to admit that the correlation of genetic differentiation and geographic distance in *E. graciloides* at the large spatial scale (1340 km) might have partly been driven by the frequent occurrence of null alleles in one of the two investigated Russian populations. Null alleles are suggestive of mutations in the nonrepetitive regions of microsatellites (Callen et al. 1993). However, one may argue that the enhanced occurrence of null alleles in Russian populations is just another manifestation of diverging alleles, and thus of increasing genetic divergence. In that respect, our main conclusions are rather strengthened.

Our data represent the first population genetic analysis for planktonic freshwater copepods based on microsatellite polymorphism. We gave first insights into the population genetic structuring of two important members of zooplankton communities in European freshwaters and provided a new set of data in the recently discussed field of the dispersal gene flow paradoxon in freshwater organisms. We showed the importance of the sampling scale for detecting whether populations were isolated by distance. We found that a species with diapausing eggs exhibited distance-correlated genetic exchange (isolation-by-distance) over short spatial scales. In the other species lacking resting eggs, patterns seemed rather to confirm persistent founder effects with no current gene flow.

## **Acknowledgements**

We are grateful to Siegfried Voss, Harald Deiwick, Victor Alekseev, Andrzej Mikulski, Adam Petrusek, Peter Kasprzak, and Kestutis Arbaciauskas for sampling assistance, and Silke Carstensen, Ilka Dankert, Nadine Ryk, and Tanja Sonntag for their laboratory assistance. We thank Nicole Aberle-Malzahn for helpful comments on an earlier version of the manuscript. TBHR was partly funded by DFG (Re 1108/4).

## Chapter III:

### Small effective population sizes in two planktonic freshwater copepod species (*Eudiaptomus*) with apparently large census sizes

#### Summary

In planktonic organisms census population sizes ( $N_c$ s) are huge, suggesting large effective population sizes ( $N_e$ s). Yet reliable estimates using temporal assessments of allele frequencies and high-resolution molecular markers are rare. Here, we report a seven-order of magnitude difference between  $N_e$  and  $N_c$  in two freshwater copepod species. Further, we hypothesized larger  $N_e$  in *E. graciloides* compared to *E. gracilis* due to the occurrence of a resting propagule bank in the former species. We took four temporal samples ( $n = 177$ ) between 2002 and 2004 for each species in a lake in northern Germany and genotyped copepods at polymorphic microsatellite loci.  $N_e$  and the fraction of migrants ( $m$ ) were estimated jointly using a pseudo-likelihood approach. We explored different scenarios of generation times combined with source populations from which migrants were received to investigate how these differences might influence estimates of  $N_e$  and  $m$ .  $N_e$ s were very small compared to  $N_c$ s in both *Eudiaptomus* species ( $N_e/N_c$ -ratio =  $10^{-7}$  to  $10^{-8}$ ) and migration rates were low. Contrary to expectations,  $N_e$ -estimates between both *Eudiaptomus* species were statistically not different (*E. graciloides*,  $N_e$  (mean of all scenarios) = 672.7, confidence intervals ranging from 276 to 1949; *E. gracilis*,  $N_e$  (mean of all scenarios) = 1027.4, confidence intervals ranging from 449 to 2495). Both, point estimates of  $N_e$  and confidence intervals were biased upwards using more populations as *a priori* defined source populations and more generations per year in the calculations, respectively. Small  $N_e$ s in *Eudiaptomus* might be caused by a substructuring of populations, a skewed reproduction success, unequal sex-ratios, non-random mating or fluctuating population sizes.  $N_e$ s of both *Eudiaptomus* species suggested that the resting propagule bank of *E. graciloides* had no detectable influence on  $N_e$ s compared to *E. gracilis*.

## Introduction

The effective population size,  $N_e$ , summarizes the relative magnitude of random processes (genetic drift) as opposed to deterministic processes (selection) on allele frequencies in a single variable (e.g., Wright 1931, 1938; Hedrick 1999). Estimations of  $N_e$  from life-history data are difficult to obtain in wild populations (Schwartz et al. 1998; Wang and Caballero 1999; Waples 2002). Often, estimations from genetic marker data provide the only alternative. From those data the so called variance effective size can be obtained by relating the measured variance in allele frequency of a “real” population to that of an ideal population of infinite population size (Hedrick 1999).

In various species there are pronounced discrepancies between the ecological size of a population (census size =  $N_c$ ) and its population genetic, effective population size (Frankham 1995). In planktonic organisms,  $N_e/N_c$ -ratios have rarely been studied so far. A few investigations have been made in order to estimate  $N_e$  from single genetic samples as  $N_\mu$  and  $N_m$  using the polymorphism of microsatellite loci or the variation of mtDNA sequences (Bucklin and Wiebe 1998; Pálsson 2000). However, even if we would expect much lower  $N_e$ s compared to  $N_c$ s, effective population sizes should nevertheless be considerable in freshwater planktonic species, since they are passively transported by movements of the water, what should prevent individuals from spatial isolation.

Several “temporal methods” have been developed to estimate  $N_e$  from changes in allele frequencies over time (e.g., Krimbas and Tsakas 1971; Waples 1989; Williamson and Slatkin 1999; Wang 2001). A common assumption of all these approaches is that allele frequencies only change through genetic drift, while the role of migration, mutation and selection are negligible. Temporal methods either follow a moment-based or a maximum-likelihood based approach (ML). The latter was shown to have higher accuracy and precision compared to moment-based approaches (e.g., Williamson and Slatkin 1999; Berthier 2002; Wang and Whitlock 2003), and Wang (2001) found that rare alleles are the main cause for the differences in performance. Since most natural populations are not completely isolated but open to gene flow, emigration or immigration of alleles by propagules or migrant individuals will change local allele frequencies over time. Recently, Wang and Whitlock (2003) developed a pseudo-likelihood approach to estimate  $N_e$  that accounts for both, migration and drift, rather than drift alone. Their approach allows to

estimate  $N_e$  and the fraction of migrants ( $m$ ) jointly from temporal samples (Wang and Whitlock 2003), and will therefore be employed in this study.

Here, we examined  $N_e$ s of the two planktonic freshwater copepods *Eudiaptomus graciloides* and *E. gracilis*. We estimated  $N_e$ s from four temporal samples taken from a lake in northern Germany and compared these to estimates of  $N_c$ , which we calculated on the basis of abundances data from earlier investigations (Hofmann 1979; Fußmann 1996). Thus, with our data we provide first  $N_e/N_c$ -ratios for any freshwater plankton organisms as well as the first  $N_e/N_c$ -ratios, where  $N_e$ s were estimated from temporal data, for any plankton organism. Further, we compared  $N_e$ s between *E. graciloides* and *E. gracilis*. While both species co-occur in a number of lakes in northern Europe and reveal strong morphological similarities (Kiefer 1978; Nauwerck 1980; Einsle 1993), they exhibit marked differences in their life cycle, namely the presence of a diapausing egg bank in one, but not the other species (Santer et al. 2000). Long-term diapausing eggs can remain viable over decades and establish resting propagule banks (Hairston et al. 1995; Cáceres 1998), which can slow down the rate of microevolution (Hairston and DeStasio 1988). Hence, ‘old’ genotypes are stored in the sediments and the recruitment of stages that have remained dormant over longer time periods should contribute to the maintenance of genetic diversity of today's populations (reviewed in Bilton et al. 2001). Here, we tested the hypothesis that *E. graciloides* has a larger effective population size compared to its congener *E. gracilis* due to the existence of a resting propagule bank in the former species. Additionally, we explored different scenarios of likely generation times in *Eudiaptomus* species combined with different source populations, from which migrants were received, to investigate how these differences might influence the estimates of  $N_e$  and  $m$ .

## Methods

### *Sampling*

Population samples of each *Eudiaptomus* species were taken in lake Schöhsee (Northern Germany, 54°09'N 10°26'E) at four sampling dates: March 2002, March/April 2003, August 2003 and March 2004, representing a total of approximately four to six generations (see below). Samples were taken in vertical hauls with a plankton-net and stored in 70% ethanol until further processing. Since samples were taken from populations apparently

much larger than the sample sizes (Waples 1989, plan 1), we assume that our sampling did not change the available pool of reproductive individuals (Wang and Whitlock 2003).

#### *DNA extraction and microsatellite analysis*

Female copepods (n=177/sampling date/species) were isolated with dissecting needles and DNA was extracted with the Invisorb DNA Tissue HTS 96-Kit/C from Invitex following the manufacturer's instructions. For estimating allele frequencies, we used the polymorphism displayed at three and seven microsatellite loci for *E. graciloides* and *E. gracilis*, respectively (EGO2, 7 and 10, Genbank accession numbers AY547392-547394; EGI1, 3, 8, 12, 13, 17 and 35, Genbank accession numbers AY547395-547401, chapter I). In the present data set, microsatellite loci represented a total of 119 alleles in *E. graciloides* and 292 alleles in *E. gracilis*. Microsatellite procedures were carried out as described in chapter II.

#### *Census population size*

We made an educated guess on census population sizes ( $N_c$ ) for both *Eudiaptomus* species from the data collected by Hofmann (1979). Hofmann (1979) sampled adult *E. graciloides* and *E. gracilis* in lake Schöhsee from April to December 1974 bi-weekly and calculated their abundances (Hofmann 1979). We averaged these abundances (Figure 1 in Hofmann 1979) for each species (*E. graciloides*,  $0.062 \times 10^6$  individuals/m<sup>2</sup>; *E. gracilis*,  $0.025 \times 10^6$  individuals/m<sup>2</sup>). Considering a lake area of 82.3 ha for lake Schöhsee (Rai 1982) estimates of census population sizes were  $5.1 \times 10^{10}$  and  $2.1 \times 10^{10}$  for *E. graciloides* and *E. gracilis*, respectively. Since the distribution of zooplankton in a water body might be rather patchy and abundances might vary extensively between years, we conservatively assume census sizes in the order of magnitude  $10^{10}$  instead of above estimates. The order of magnitude of  $10^{10}$  was confirmed by the study of Fußmann (1996). He reported abundances for *Eudiaptomus* species in lake Schöhsee in 1993/1994, which ranged from five to 60 copepods/L corresponding to  $5.4 \times 10^{10}$  to  $6.5 \times 10^{11}$  copepods in the entire Schöhsee (lake area: 82.3 ha, average water depth: 13 m; Rai 1982).

#### *Genetic diversity, Hardy-Weinberg equilibrium and F-statistics*

To obtain background information on within population diversity such as expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_O$ ) and allelic richness we measured these

parameters in both *Eudiaptomus* species. We calculated expected heterozygosity ( $H_E$ ) and observed heterozygosity ( $H_O$ ) with the programme GENETIX v. 4.01 (Belkhir et al. 1998) and allelic richness with the programme FSTAT (Goudet 1995). We refrained to test for differences in genetic diversity between both *Eudiaptomus* species, since different markers were used for each species. To test for linkage disequilibrium and to calculate F-statistic parameters we used GENETIX v. 4.01 (Belkhir et al. 1998). Hardy-Weinberg probability tests were run with the software GENEPOP v. 3.3 (Raymond and Rousset 1995).

#### *Effective population size ( $N_e$ ) and migration ( $m$ )*

In order to estimate  $N_e$  in *Eudiaptomus* species, we used the temporal approach developed by Wang and Whitlock (2003) implemented in MLNE (available at <http://www.zoo.cam.ac.uk/ioz/people/wang.htm>). This method calculates the variance effective population size,  $N_e$ , together with  $m$  using a pseudo-likelihood approach, and is based on the assumption that changes in allele frequencies over time are caused by drift and migration, rather than drift only. Therefore, using this method migration sources and gene flow, respectively, must be known *a priori*. Per definition temporal samples are taken from a “focal population”, which is receiving migrants from an infinitely large “source population” (Wang and Whitlock 2003). We calculated  $N_e$  for *E. graciloides* and *E. gracilis* inhabiting lake Schöhsee in northern Germany (focal populations). To define source populations we used information on allele frequencies, and hence pairwise genetic differentiation, from 14 and 15 populations of *E. graciloides* and *E. gracilis*, respectively, sampled in 2002/2003 from lakes situated on a longitudinal gradient between northern Germany and Lithuania (Chapter II). Genetic exchange between these populations and the Schöhsee populations differed in intensity. Thus, we rather defined two different source populations *a priori* in order to investigate how these differences might influence our estimates of  $N_e$  and  $m$ . In one source population we included only three nearby populations, and in a second one we included information from all available populations (for information on source populations see Table III.1).

Table III.1. Information on *a priori* defined source populations used in the calculations to estimate  $N_e$  and  $m$ . Source 1: copepods were sampled from lakes located a few hundred meters apart from the Schöhsee (focal) population, source 2: copepods were sampled from lakes located in a range of a few hundred meters to 1000 km to the Schöhsee (focal) population, nb pop. = number of populations sampled, n = individuals, std. dev. = standard deviation. <sup>1</sup> values from chapter II.

	<i>E. graciloides</i>		<i>E. gracilis</i>	
	source 1	source 2	source 1	source 2
nb pop (n)	3 (120)	14 (612)	3 (135)	15 (674)
mean pairwise $F_{ST}$ <sup>1</sup> (Std. dev.)	0.01 (0.008)	0.02 (0.017)	0.05 (0.004)	0.08 (0.018)

Subsequently, we combined the two differently defined source populations with different possible generation times of *Eudiaptomus* in our estimations. Since both *Eudiaptomus* species are likely to exhibit two to three generations per year (Bosselmann 1975; Hofmann 1979; Santer et al. 2000), we calculated  $N_e$ s for two, three and additionally four generations per year. Accordingly, the number of elapsed generations associated with our four sampling dates (March 2002/ March, April 2003/ August 2003/ March 2004) were 0/2/3/4, 0/3/4/6 and 0/3/5/6, respectively, and 0/4/6/8. Allele counts were analysed with the programme GENEPOP v. 3.3 (Raymond and Rousset 1995) and re-arranged for data input in MLNE. For the comparison between species we also calculated  $N_e$  and  $m$  for *E. gracilis* with the information from three primer pairs only, which represented a similar number of independent alleles (106; independent alleles = total number of alleles – number of loci, Wang 2001) as the three primer pairs used for *E. graciloides* (116), which is an important parameter in the pseudo-likelihood approach (Wang 2001).

Our previous studies indicated that null alleles might be abundant in some of our microsatellite loci (Chapter II). Temporal changes in allele frequencies should be little affected by null alleles as long as their occurrence is not markedly different between sampling dates. We used three parameters (the frequencies of nonamplified samples for individual loci,  $F_{IS}$ -values, and calculated null allele frequencies; programme Microchecker 2.2.1, Brookfield 2; Van Oosterhout et al. 2004) as indicators for the abundance of null alleles. To test whether significant differences between sampling dates existed, we performed one-way-ANOVAs for  $F_{IS}$ -values and calculated null allele frequencies. For

statistical comparison between sampling dates testing nonamplified samples we used a Kruskal-Wallis ANOVA, since the data contained a high number of zero-values.

## Results

### *Genetic diversity, Hardy-Weinberg equilibrium and F-statistics*

Genetic diversity was measured as expected heterozygosity ( $H_E$ ), observed heterozygosity ( $H_O$ ) and allelic richness (Table III.2). Observed heterozygosity was lower than expected heterozygosity in both *Eudiatomus* species, indicating an excess of homozygote individuals (Table III.2).  $F_{IS}$ -values ranged among sampling dates from 0.03 to 0.26 in *E. graciloides* and from 0.18 to 0.34 in *E. gracilis*. The majority of  $F_{IS}$ -values for single microsatellite loci deviated from Hardy-Weinberg equilibrium with positive values (Table III.2). Deviations from Hardy-Weinberg assumptions have no effect on the used pseudo-likelihood method for calculations of  $N_e$  and  $m$ , because this method is based on allele frequencies. All loci were in linkage equilibrium over all sampling dates for both species.

Pairwise  $F_{ST}$ -values among different sampling dates ranged from  $-0.0002$  to  $0.0046$  in *E. graciloides* and  $0.0002$  to  $0.0058$  in *E. gracilis* calculated for seven microsatellite loci (Table III.3). In *E. graciloides*, all pairwise comparisons that included the March 2004 sample were significantly different from zero (Table III.3). In *E. gracilis*, all pairwise comparisons were significantly different from zero except for the time interval March 2002 – March/April 2003 and August 2003 – March 2004 independent of the number of microsatellites used in the calculations (Table III.3).

Table III.2. Genetic diversity in *E. graciloides* and *E. gracilis*.  $H_E$ , expected heterozygosity;  $H_O$ , observed heterozygosity; std. dev. , standard deviation; EGO10, 7, and 2, primer pairs *E. graciloides*; EGI1, 3, 8, 12, 13, 17, and 35, primer pairs *E. gracilis*; M = March, Ap = April, A = August.  $F_{IS}$ - values that deviate significantly from Hardy-Weinberg expectations are printed in bold.

<i>E. graciloides:</i>	M2002	M/Ap2003	A2003	M2004
$H_E$ (Std. dev.)	0.83 (0.16)	0.85 (0.15)	0.83 (0.17)	0.82 (0.17)
$H_O$ (Std. dev.)	0.69 (0.04)	0.71 (0.06)	0.70 (0.04)	0.63 (0.09)
Allelic richness (Std. dev.)	29.2 (19.3)	31.2 (17.7)	28.3 (15.9)	30.3 (18.6)
$F_{IS}$ EGO 10	-0.027	0.047	-0.047	<b>0.136</b>
$F_{IS}$ EGO 7	<b>0.270</b>	<b>0.233</b>	<b>0.257</b>	<b>0.297</b>
$F_{IS}$ EGO 2	<b>0.227</b>	<b>0.191</b>	<b>0.225</b>	<b>0.245</b>
<i>E. gracilis:</i>				
$H_E$ (Std. dev.)	0.86 (0.09)	0.87 (0.08)	0.84 (0.13)	0.85 (0.12)
$H_O$ (Std. dev.)	0.62 (0.16)	0.66 (0.12)	0.61 (0.16)	0.65 (0.15)
Allelic richness (Std. dev.)	31.5 (17.6)	31.5 (16.3)	32.4 (18.9)	33.6 (20.2)
$F_{IS}$ EGI 1	<b>0.142</b>	<b>0.116</b>	<b>0.111</b>	0.097
$F_{IS}$ EGI 3	<b>0.182</b>	<b>0.235</b>	<b>0.254</b>	<b>0.234</b>
$F_{IS}$ EGI 8	<b>0.175</b>	<b>0.225</b>	<b>0.173</b>	<b>0.141</b>
$F_{IS}$ EGI 12	<b>0.517</b>	<b>0.434</b>	<b>0.403</b>	<b>0.362</b>
$F_{IS}$ EGI 13	<b>0.325</b>	<b>0.331</b>	<b>0.403</b>	<b>0.353</b>
$F_{IS}$ EGI 17	<b>0.246</b>	<b>0.231</b>	<b>0.221</b>	<b>-0.008</b>
$F_{IS}$ EGI 35	<b>0.439</b>	<b>0.137</b>	<b>0.409</b>	<b>0.462</b>

Table III.3: Pairwise genetic differentiation between sampling dates. M = March, Ap = April, A = August. In brackets: number of used microsatellite loci.  $F_{ST}$ -values that differed significantly from zero are printed in bold.

	<i>E. graciloides</i>	<i>E. gracilis</i> (7)	<i>E. gracilis</i> (3)
M2002 - M/A2003	-0.0002	0.0002	0.0005
M2002 - Aug2003	0.0004	<b>0.0015</b>	<b>0.0026</b>
M2002 - M2004	<b>0.0027</b>	<b>0.0013</b>	<b>0.0030</b>
M/A2003 - Aug2003	0.0006	<b>0.0058</b>	<b>0.0103</b>
M/A2003 - M2004	<b>0.0035</b>	<b>0.0046</b>	<b>0.0097</b>
Aug2003 - M2004	<b>0.0046</b>	0.0002	-0.0010

*Effective population size ( $N_e$ ) and migration ( $m$ )*

For *E. graciloides* we estimated  $N_e$ s in the hundreds (mean- $N_e$  from all different scenarios with standard deviation: 672.7 ( $\pm$  141.7)) with confidence intervals ranging from 276 to 1949 (Table III.4). For *E. gracilis* mean estimates of  $N_e$  calculated with information from all seven microsatellite loci were 1027.4 ( $\pm$  239.1) with confidence intervals between 449 and 2495 (Table III.4). Mean estimates of  $N_e$  for *E. gracilis* calculated with three primer pairs were 696.4 ( $\pm$  171.2) with confidence intervals ranging from 277 to 2084 (Table III.4). Mean-migration was 0.03 ( $\pm$  0.01) for *E. graciloides*, 0.02 ( $\pm$  0.01) and 0.01 ( $\pm$  0.004) for *E. gracilis* calculated with information from three and seven microsatellite loci, respectively (data for single scenarios see Table III.4). Estimates of  $N_e$  increased consistently with increasing generation times for all scenarios within both *Eudiptomus* species (Table III.4).  $N_e$  estimates and the associated confidence intervals were shifted upwards considering more populations in the *a priori* defined source population (Table III.4). In *E. gracilis* values of  $N_e$  calculated with seven microsatellite loci were higher compared to *E. graciloides*, but confidence intervals were overlapping (Table III.4).

Table III.4. Effective population sizes ( $N_e$ ) and migration ( $m$ ) for *E. graciloides* and *E. gracilis* estimated by MLNE according to a pseudo-likelihood approach (Wang and Whitlock 2003). GS = generation sequence assumed in the calculation; s1 = source population built from 3 nearby populations ( $n = 120$  and  $135$  for *E. graciloides* and *E. gracilis*, respectively); s2 = source population built from 14 populations in *E. graciloides* ( $n = 612$ ) and 15 populations in *E. gracilis* ( $n = 674$ ); CI = confidence interval; in brackets number of microsatellite loci used; values printed in bold belong to unsmooth log-likelihood curves (Fig. III.1, 2).

	GS	$N_e$ , s1	95 % CI ( $N_e$ , s1)	$m$ , s1	95 % CI ( $m$ , s1)	$N_e$ , s2	95 % CI ( $N_e$ , s2)	$m$ , s2	95 % CI ( $m$ , s2)
<i>E. graciloides</i> :	0/2/3/4	452.9	[276.3-917.0]	0.045	[0.020-0.083]	551.8	[317.4-1258.7]	0.025	[0.010-0.048]
	0/3/4/6	601.8	[364.1-1190.0]	0.032	[0.015-0.058]	745.5	[431.4-1609.0]	0.018	[0.008-0.033]
	0/3/5/6	621.9	[367.0-1268.0]	0.031	[0.014-0.057]	789.4	[445.5-1729.1]	0.016	[0.007-0.032]
	0/4/6/8	726.0	[433.6-1435.6]	0.025	[0.012-0.046]	892.5	[531.4-1949.1]	0.014	[0.006-0.026]
<i>E. gracilis</i> (7):	0/2/3/4	677.3	[449.4-1201.5]	0.022	[0.011-0.035]	837.3	[510.8-1697.2]	<b>0.017</b>	<b>[0.007-0.029]</b>
	0/3/4/6	968.6	[617.8-1613.5]	<b>0.015</b>	<b>[0.008-0.024]</b>	<b>1322.1</b>	<b>[765.6-2495.3]</b>	<b>0.010</b>	<b>[0.008-0.018]</b>
	0/3/5/6	867.3	[589.9-1412.6]	0.016	[0.010-0.026]	1114.8	[712.5-1987.6]	<b>0.012</b>	<b>[0.006-0.019]</b>
	0/4/6/8	1062.9	[718.6-1724.6]	<b>0.013</b>	<b>[0.008-0.020]</b>	<b>1368.8</b>	<b>[913.2-2459.6]</b>	<b>0.009</b>	<b>[0.007-0.015]</b>
<i>E. gracilis</i> (3):	0/2/3/4	455.5	[276.9-919.4]	0.032	[0.015-0.060]	548.6	[309.9-1260.6]	0.021	[0.008-0.042]
	0/3/4/6	728.3	[418.6-1624.3]	0.019	[0.008-0.037]	<b>871.2</b>	<b>[502.1-2465.4]</b>	<b>0.013</b>	<b>[0.004-0.024]</b>
	0/3/5/6	564.8	[351.6-1037.6]	0.025	[0.013-0.045]	672.9	[407.4-1361.9]	0.016	[0.008-0.030]
	0/4/6/8	767.3	[463.8-1486.0]	0.018	[0.009-0.032]	962.3	[558.7-2083.7]	0.011	[0.005-0.021]

In Figure III.1 and 2 the log-likelihood curves for  $N_e$  and  $m$ , respectively, are presented. These curves give background information on the calculated  $N_e$  and  $m$  values and confidence intervals in that respect, that estimates belonging to unsmooth curves should be interpreted with caution (Wang J. pers. comm.).

For *E. graciloides*, log-likelihood curves of  $N_e$  were smooth and narrow for all different generation time/source population scenarios (Fig. III.1E, F). In *E. gracilis*, log-likelihood curves for  $N_e$  were relatively smooth and narrow for all estimations, which allowed gene flow with three nearby populations (Fig. III.1A, C). Taking more populations as source population into account led to rugged lines in the second half of the log-likelihood curves for  $N_e$  calculations based on information from all seven microsatellite loci (Fig. III.1B). The same pattern occurred for the calculation with information from three microsatellite loci and a generation sequence of 0/3/4/6 (Fig. III.1D).

Log-likelihood curves for  $m$  were rather smooth for all *E. graciloides* calculations (Fig. III.2E, F). For *E. gracilis* the curves of  $m$  for all calculations with information from seven microsatellite loci and with 15 source populations were either rugged or exhibited a very steep increase in the first half of the curves (Fig. III.2B). The  $m$ -curves for generation times of 0/3/4/6 show a strong zig-zag for the calculation with all microsatellite loci and three source populations, and for the calculation with three microsatellite loci and 15 source populations (Fig. III.2A, D). Unsmooth log-likelihood curves are known from data sets that contain a lot of private rare alleles in both source and focal populations (Wang J. pers. comm.).

In *E. graciloides*, data sets of focal populations contained private rare alleles with a mean of 2 % ( $\pm 1$  % standard deviation) if 14 populations were taken into account as source population and 4 % ( $\pm 2$  %) if three populations were taken into account as source population. Data sets of the total source populations contained 13 % and 1 % rare alleles for 14 and three populations taken into account as source population, respectively. In *E. gracilis* 4 % ( $\pm 2$  %) and 6 % ( $\pm 3$ %) private rare alleles occurred in the focal populations and 18 % and 7 % in the total source population for 15 and three populations taken into account as source population when seven microsatellite loci were used. Focal populations in data sets derived from three loci contained 4 % ( $\pm 3$  %) and 5 % ( $\pm 3$  %) private rare

alleles and source populations 20 % and 8 % for 15 and three populations taken into account as source population.

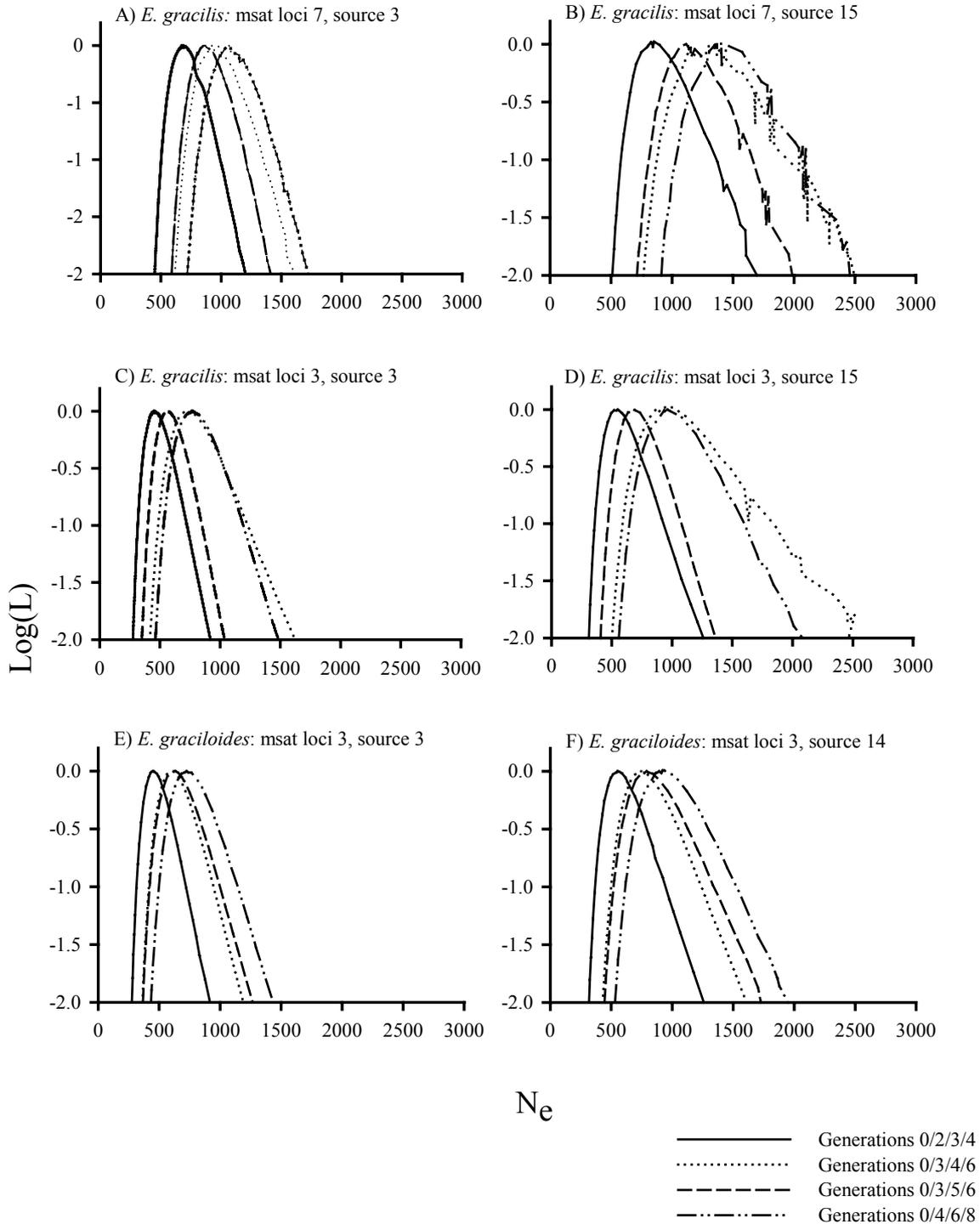


Figure III.1. Log-likelihood curves for effective population sizes ( $N_e$ ) in *E. graciloides* and *E. gracilis*. Msat, microsatellite; source, numbers of populations that were pooled as source population. The 95% confidence interval for the method implemented here can be calculated as the range of support associated with a drop of 2 in the log-likelihood.

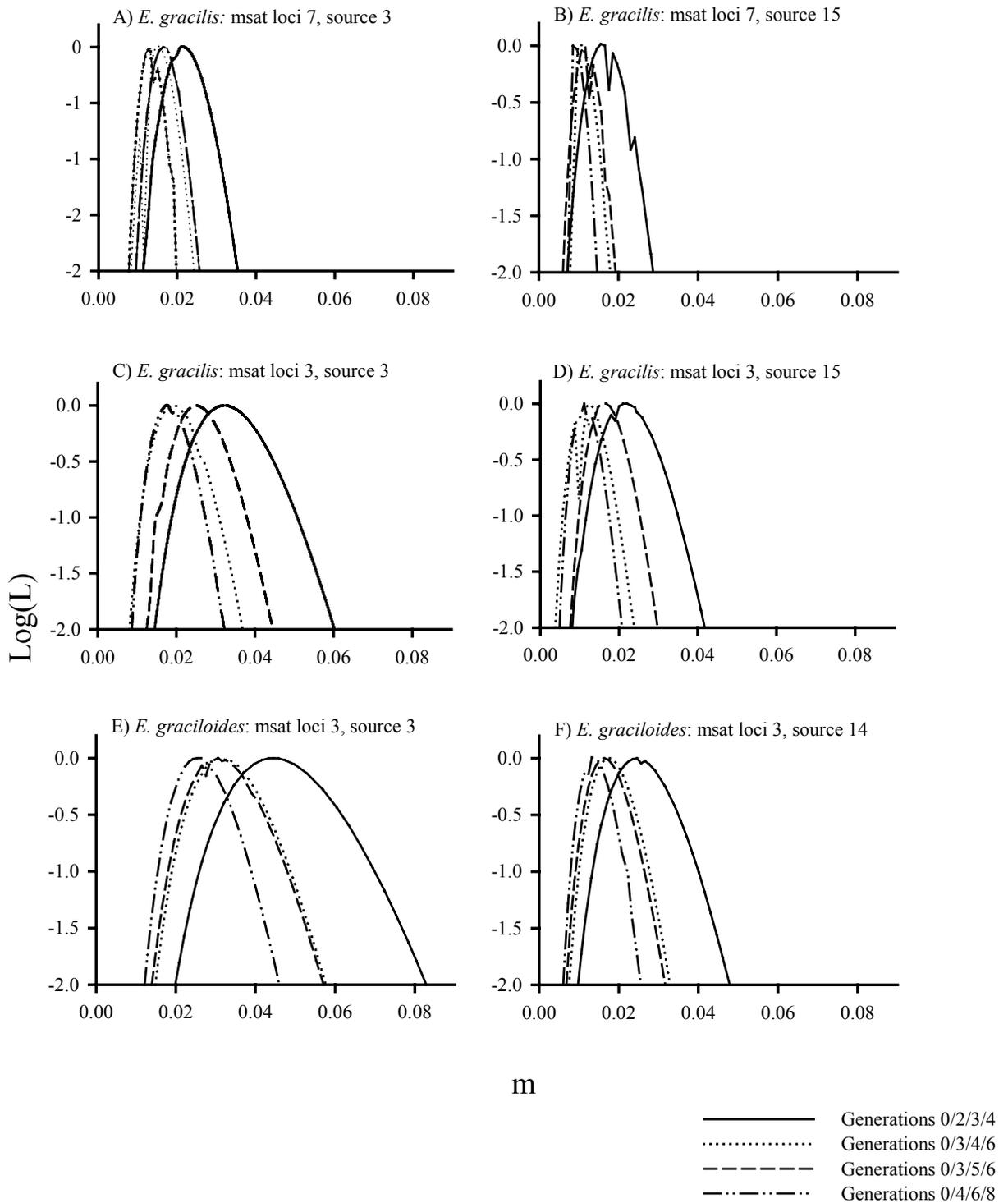


Figure III.2. Log-likelihood curves for migration ( $m$ ) in *E. graciloides* and *E. gracilis*. Msat, microsatellite; source, numbers of populations that were pooled as source population. The 95% confidence interval for the method implemented here can be calculated as the range of support associated with a drop of 2 in the log-likelihood.

Null alleles were equally distributed among sampling dates in both *Eudiaptomus* species, and should not have influenced the used pseudo-likelihood approach. In *E. graciloides*,  $F_{IS}$ -values (1-way-ANOVA,  $F = 0.24$ ,  $df = 3$ ,  $p = 0.87$ ), calculated null allele frequencies (1-way-ANOVA,  $F = 0.46$ ,  $df = 3$ ,  $p = 0.72$ ) and frequencies of nonamplified samples (Kruskal-Wallis-ANOVA,  $H(3, N=12) = 1.55$ ,  $p = 0.67$ ) were not significantly different between sampling dates. The same was true for *E. gracilis* ( $F_{IS}$ -values: 1-way-ANOVA,  $F = 0.27$ ,  $df = 3$ ,  $p = 0.84$ ; calculated null allele frequencies: 1-way-ANOVA,  $F = 0.13$ ,  $df = 3$ ,  $p = 0.94$ ; frequencies of nonamplified samples: Kruskal-Wallis-ANOVA,  $H(3, N=28) = 0.82$ ,  $p = 0.84$ ).

## Discussion

In this study we present first estimates of  $N_e$  in any planktonic organism using high-resolution microsatellite markers together with a novel temporal approach that takes migration into account. We found effective population sizes for two species of copepods that were several orders of magnitude smaller than their likely census sizes. Despite the knowledge of census sizes often being greater than effective population sizes from a number of species (Frankham 1995), such small  $N_e/N_c$ -ratios are surprising for freshwater zooplankton organisms, since one would assume considerable mix-up of individuals due to passive transport within a relatively small water body. Further, we found that  $N_e$ s were equally small in both *Eudiaptomus* species in lake Schöhsee. Thus, our initial hypothesis of differences in  $N_e$  due to the presence/absence of a resting egg bank was not supported. Earlier studies already pointed to small effective population sizes due to the occurrence of high  $F_{IS}$ -values and departures from Hardy-Weinberg equilibrium in a number of *E. graciloides* and *E. gracilis* populations (Chapter II). However, it was difficult to decide whether these findings were caused by a real biological phenomenon or, for example, by the presence of null alleles. It is likely that null alleles occurred in the present study, but the calculation of  $N_e$  and  $m$  should not have been affected, as null alleles were equally distributed between sampling dates (Jehle et al. 2001; Wang and Whitlock 2003).

Census sizes ( $N_c$ ) were not known for populations of *E. graciloides* and *E. gracilis*, but it is likely that they are at least in a range of  $10^{10}$  for each species in the entire lake. Thus,  $N_e/N_c$ -ratios in *E. graciloides* and *E. gracilis* were in a range of approximately  $10^{-7}$  to  $10^{-8}$ .

Using the variation in mitochondrial DNA sequences, Bucklin and Wiebe (1998) estimated effective population sizes for the marine copepods *Calanus finmarchicus* and *Nannocalanus minor*. Using the number of haplotypes, they found  $N_e$ s in a range of approximately  $10^8$ , and effective female population sizes (from nucleotide diversities) of  $\sim 10^5$ . These estimates were  $10^7$  to  $10^{10}$  smaller than their estimated census female population sizes ( $N_f$ ). When using sequence polymorphism data, it is difficult to decide whether such small  $N_e/N_f$ -ratios are caused by contemporary evolutionary processes or rather by population fluctuations in the distant past. Bucklin and Wiebe (1998) hypothesized that small effective population sizes in *C. finmarchicus* and *N. minor* were caused by range reductions due to glaciation during the past 20,000 years. In the present study, we used information from microsatellite loci, which should be better suited than mtDNA to estimate  $N_e$  due to faster evolution, and a temporal approach to calculate variance effective sizes.

Small effective population sizes despite much larger census sizes can be caused by the substructuring of populations, a skewed reproduction success, unequal sex-ratios, non-random mating and fluctuating populations sizes (e.g., Franklin 1980; Caballero 1994; Frankham 1995; Hedrick 2005). Earlier studies showed that abundances of both *Eudiaptomus* species fluctuated strongly within years in lake Schöhsee which might contribute to low  $N_e$ s (Hofmann 1979; Fußmann 1996; Santer et al. 2000). More life-history and demographic information for *E. graciloides* and *E. gracilis* would be needed to identify factors that are likely to reduce effective population sizes seven to eight order of magnitude below their census sizes. Our previous population genetic studies already pointed to small effective populations sizes in both *Eudiaptomus* species, since we found high  $F_{IS/IT}$ -values in 16 *E. graciloides* and 15 *E. gracilis* populations (*E. graciloides*,  $F_{IS} = 0.22$  (0.10 – 0.27),  $F_{IT} = 0.26$  (0.17 – 0.29); *E. gracilis*,  $F_{IS} = 0.12$  (0.11 – 0.12),  $F_{IT} = 0.16$  (0.13 – 0.23); Chapter II).  $F_{IS}$ -values in this study ( $n = 177$ ) were equally high for all four sampling dates, supporting our earlier findings.

Estimates of  $N_e$  and confidence intervals were both shifted upwards when we included more populations in the *a priori* defined source populations and using more generations per year in our calculations. In the short term, drift would be generally overestimated, and hence  $N_e$  underestimated, if migration was not taken into account for calculations of  $N_e$

(Wang and Whitlock 2003). In our study migration from only three surrounding populations seemed to have not reflected all migration events properly, leading to smaller  $N_e$ s compared to greater source populations taken into account. However, migration was low in both species and even lower in the case of the larger source population, which consisted of 612 individuals of *E. graciloides* and 674 individuals of *E. gracilis*. Thus, the impact of gene flow on changes in allele frequencies seemed higher compared to only three populations functioning as source population. On the other hand, more alleles seemed to be abundant in the source population that differed from alleles occurring in the focal population, and thus migration occurred to be lower compared to calculations were the smaller source populations were used. Nevertheless,  $N_e$ s were small for all different scenarios.

Despite a tendency of  $N_e$  estimates being higher in *E. gracilis* compared to *E. graciloides*, differences of  $N_e$ s between both *Eudiaptomus* species were statistically not significant. This suggested that genotypes stored in the resting propagule banks of lake Schöhsee did not affect the effective population size of the watercolumn population of *E. graciloides* as initially hypothesised.  $N_e$  estimates and confidence intervals were even closer, when comparable amounts of independent alleles were used in the calculations. On the contrary, the differences in  $N_e$  estimates were more pronounced if we considered prolonged generation times in *E. graciloides* compared to *E. gracilis* due to the hatching of long-term diapausing eggs from the sediments. In chapter IV a mean generation time of 0.48 per year for *E. graciloides* in lake Schöhsee will be estimated from the average of copepods originating from subitaneous eggs and animals that hatched from the egg bank, which approximately corresponded to the generation sequence 0, 2, 3, 4 in our calculations. *E. gracilis* is most likely to exhibit two to three generations per year (Hofmann 1979; Santer et al. 2000; Santer et al. unpublished). Therefore, for *E. gracilis* in the present investigation  $N_e$  probably equals an intermediate value of the estimates made for generation sequence 0, 2, 3, 4 and 0, 3, 5, 6 / 0, 3, 4, 6. In an independent study (Chapter IV) relative population sizes ( $N_e\mu$ ) were calculated, for *E. graciloides* and *E. gracilis* for three lakes in northern Germany with information from mitochondrial DNA sequences. A phylogeographic analyses and a Bayesian skyline-plot resulted in  $N_e\mu$  being rather greater in *E. gracilis* compared to *E. graciloides*, although the contrast was statistically not significant for most time points, including the most recent one (*E. gracilis*:  $N_e\mu \approx 0.5$ , *E. graciloides*:  $N_e\mu \approx$

0.05; see chapter IV). This difference was attributed to an increase of population size in *E. graciloides* ten thousands of years later compared with *E. gracilis* and probably a bottleneck in population size in the former (Chapter IV). If we assume a mutation rate of  $10^{-3}$  to  $10^{-4}$  for microsatellite loci in this study,  $N_e\mu$  for *E. graciloides* (generation sequence 0, 2, 3, 4 mean of source 1 and 2) was  $0.5 (\pm 0.07)$  and  $0.05 (\pm 0.01)$ , respectively. For *E. gracilis*  $N_e\mu$  varied from 0.7 to 1.3 and 0.07 to 0.13 assuming a mutation rate of  $10^{-3}$  and  $10^{-4}$  for seven microsatellite loci and from 0.5 to 0.9 and 0.05 to 0.09 for three microsatellite loci (range given for estimates from combinations of generation sequences 0, 2, 3, 4 / 0, 3, 4, 6 and 0, 3, 5, 6 with source 1 and 2). Thus, we did not find differences in  $N_e s$  between *E. gracilis* and *E. graciloides* of the order of magnitude found from the analysis of mtDNA variation (Chapter IV). However, a congruent pattern was found in both studies, despite the fact that the markers and methods used to estimate  $N_e\mu$  differed and that  $N_e\mu$  was calculated together for populations from three lakes, only including the Schöhsee, making it likely that a more general phenomenon was found in *Eudiaptomus*. Nevertheless, within our model system of *Eudiaptomus* species inhabiting lake Schöhsee we could not confirm the general assumption of effective population sizes being greater due to the existence of a resting propagule bank if compared to a closely related species where no such egg bank exists.

Our study is a further data point for an emerging consensus that in small planktonic species in general, effective population sizes are much smaller than expected. It will be interesting to see whether this pattern holds once more species have been rigorously studied using high-resolution genetic markers and temporal variation in allele frequencies.

### **Acknowledgements**

We thank Silke Carstensen, Ilka Dankert and Rahel Klein for their laboratory assistance. TBHR was partly funded by DFG (Re 1108/4).

**Chapter IV:****The population genetic consequences of diapause in *Eudiaptomus* copepods****Summary**

Prolonged diapause (extended dormancy) is thought to greatly influence evolution in freshwater invertebrates by lengthening generation time, promoting higher levels of dispersal among populations by wind or animal vectors, and increasing effective population size. However, empirical tests of these predictions are relatively rare. Comparative studies can be informative in this regard, if the comparisons involve sympatric, closely related species that differ only in the presence or absence of a dormant life history stage. We built upon the study of chapter II, which used this approach to study patterns of microsatellite variation in *Eudiaptomus* copepods from northern Europe. *E. graciloides* possesses diapausing eggs and adults, whereas the closely related species *E. gracilis* is ecologically and trophically similar but lacks diapause. To separate further historical processes from recent anthropogenic influences, we examined mitochondrial DNA sequence variation in these species from three lakes in northern Germany where they are sympatric. Biotic and abiotic variation among contemporary and historical populations was minimized by focusing on hydrologically linked lakes separated by less than 10 km. *E. gracilis* and *E. graciloides* possess very different patterns of mtDNA variation. Both species possess significant population structure on small spatial scales, but *E. graciloides* has far fewer alleles despite a larger average divergence. Phylogeographic analyses and Bayesian skyline plots revealed evidence for historic population expansions in both species, with the growth phase beginning thousands of generations earlier in *E. gracilis* than in *E. graciloides*. Levels of genetic diversity suggest that effective population size may be an order of magnitude larger in *E. gracilis* than in *E. graciloides*, even though *E. graciloides* possesses a diapausing egg bank. Although this would seem to be an ideal system for quantifying the genetic role of diapause, we found no support for the assumption that freshwater invertebrates with diapause categorically possess larger effective population sizes and higher rates of gene flow than closely related species that lack diapause.

## Introduction

Prolonged diapause (i.e., lasting longer than a single active generation) is generally assumed to have pronounced effects on patterns of genetic variation in freshwater invertebrates. “Banks” of diapausing eggs typically contain billions of propagules that can be viable for several decades (e.g., Hairston et al. 1995; Hairston 1996; Cousyn et al. 2001), increasing effective population size beyond individuals found in the water column, and reducing the relative importance of genetic drift. Extended diapause can also substantially increase generation times, slowing the rates of microevolutionary processes (Hairston and De Stasio 1988). For traits that experience fluctuating selection, extended diapause can in theory act to maintain genetic variation (Ellner and Hairston 1994; Hedrick 1995), which has been shown for a population of diaptomid copepods (Ellner et al. 1999). Diapause is also believed to promote higher levels of dispersal among populations, since many encysted diapausing eggs are resistant to harsh environmental conditions and possess attributes that appear to promote transport by vertebrates or wind (e.g., Brendonck et al. 1992; Korovchinsky and Boikova 1996; Brendonck and Riddoch 1999).

The estimation of movement among populations (dispersal and gene flow) occupies a particularly important role in population biology, community ecology and conservation. Theoretical debates about topics as varied as population persistence, local adaptation and community assembly rules are based in part on different underlying assumptions about the spatial and temporal frequency of movement (e.g., Belyea and Lancaster 1999; Harrison and Bruna 1999; Kawecki and Ebert 2004). There is no general consensus regarding the extent to which populations of freshwater invertebrates are connected by dispersal and gene flow, and it is not even clear that there should be a consensus. Most species of zooplankton and other freshwater invertebrates have explicitly or implicitly been assumed to disperse widely and often through various means (reviewed by Bilton et al. 2001; Bohonak and Jenkins 2003). However, experiments specifically focused on estimating dispersal have found that this generalization does not hold, since the range of dispersal rates runs from very low to very high for any particular spatial scale (e.g., Jenkins and Underwood 1998; Brendonck and Riddoch 1999; Cáceres and Soluk 2002; Cohen and Shurin 2003; Louette and De Meester 2005). In numerous studies of zooplankton, the highest gene flow estimates are not always found in species with attributes that are thought to confer high dispersal ability (Boileau et al. 1992). In the most striking cases, significant

genetic differences have been found at even the smallest spatial scales (e.g., Weider and Hebert 1987; Vanoverbeke and De Meester 1997; Gómez et al. 2002).

For neutral genetic markers, genetic differences among populations are primarily due to a balance between mutation, drift and gene flow. Interpreting empirical data is not straightforward, because it is almost always the case that observed patterns can be explained by more than one parameter combination. The variety of interpretations increases even further under nonequilibrium conditions, because the genetic signature of a founder event (or historically higher rates of gene flow) may last for thousands of generations or longer (Boileau et al. 1992; Bohonak and Roderick 2001; De Meester et al. 2002). As a result, freshwater biologists have interpreted levels of genetic divergence on a case-by-case basis by invoking specific combinations of founder effects, drift, dispersal, gene flow and the uncoupling of dispersal and gene flow due to local adaptation. For example, the “monopolization hypothesis” (De Meester et al. 2002) assumes that zooplankton dispersal rates are relatively high, and that diapausing egg banks exacerbate genetic founder effects by inflating population sizes and lengthening generation times. However, Bohonak and Jenkins (2003) speculated that the majority of freshwater invertebrates may not possess the unique combination of parameters (e.g., dispersal rate, generation time) required for this hypothesis.

Comparisons between organisms with different life histories provide one way to assess the assumed genetic consequences of prolonged diapause. The ideal contrast would be between sympatric, closely related species that differ only in the presence or absence of a dormant life history stage. In chapter II this approach was used to study patterns of microsatellite variation in *Eudiatomus* copepods from northern Europe at different spatial scales. *E. graciloides*, which possesses diapausing eggs and adults, shows patterns of isolation by distance (IBD) on small and very large spatial scales (0 – 100 km and up to 1300 km) but not intermediate scales (100 – 1000 km). Patterns of IBD (decreasing genetic similarity with increasing geographic distance) generally indicate that gene flow is too infrequent and/or spatially restricted to create panmixia over the spatial scale being studied (Slatkin 1993). The closely related species *E. gracilis* is ecologically similar, regionally sympatric, and often coexists with *E. graciloides* in the same lake (Kiefer 1978; Hofmann 1979; Nauwerck 1980; Adrian 1997). However, *E. gracilis* does not possess

diapause (Santer et al. 2000), and does not display IBD on small or intermediate scales (Chapter II), (Populations >1000 km were not studied for this species). High variance in the *E. gracilis* IBD plot on small scales could reflect recent, irregular colonization patterns and high levels of drift (Hutchison and Templeton 1999). However, persistent founder effects are actually expected to be weaker in *E. gracilis* than in *E. graciloides*, because diapausing egg banks in *E. graciloides* should greatly increase population size and slow the approach to genetic equilibrium. In chapter II was hypothesized that small-scale dispersal rates are higher in *E. graciloides* than in *E. gracilis*, eliminating founder effects in the former but not the latter. Alternative interpretations for all three spatial scales may involve specific dispersal vectors or historical processes.

To separate further the effects of population history, drift and gene flow in these two species, we examined mitochondrial DNA sequence variation in three lakes from northern Germany where they are sympatric. The use of DNA sequence variation offers some advantages over microsatellite markers, including a wider array of analyses to assess historical demographic changes. Two of the three lakes are connected by a river system that has been modified for boat traffic, facilitating the separation of very recent anthropologic changes in the landscape from those that took place during the post-Pleistocene. Our overall goals were to build upon the results of chapter II and assess the role of diapause in shaping genetic population structure. We specifically focused on whether prolonged dormancy in *E. graciloides* is associated with more persistent founder effects, increased rates of gene flow, increased effective population sizes, and less dramatic bottlenecks in population size when compared to *E. gracilis*.

## Methods

*Eudiaptomus gracilis* and *E. graciloides* were collected from the open water of three lakes in Schleswig-Holstein, Germany: the Grosser Plöner See, the Keller See and the Schöhsee. Collections were taken during May 1999 (approximately 5 individuals per species per lake) and June and July 2000 (ca. 15 individuals per species per lake). The Keller See is approximately 8.6 km east and upstream of the Grosser Plöner See, and the two lakes are separated by the small river Schwentine, and two intervening lakes (the Dieksee and the Behler See successively; see Fig. IV.1). The outflow from the Schöhsee enters the Behler

See through a small stream inhospitable to planktonic copepods that flows only when the water level in Schöhsee is very high. The Behler See then flows into the Grosser Plöner See, with a total hydrologic distance between the Schöhsee and the Grosser Plöner See of ca. 2.3 km. These three lakes were also sampled for microsatellite analyses (Chapter II), although this study used different individuals.

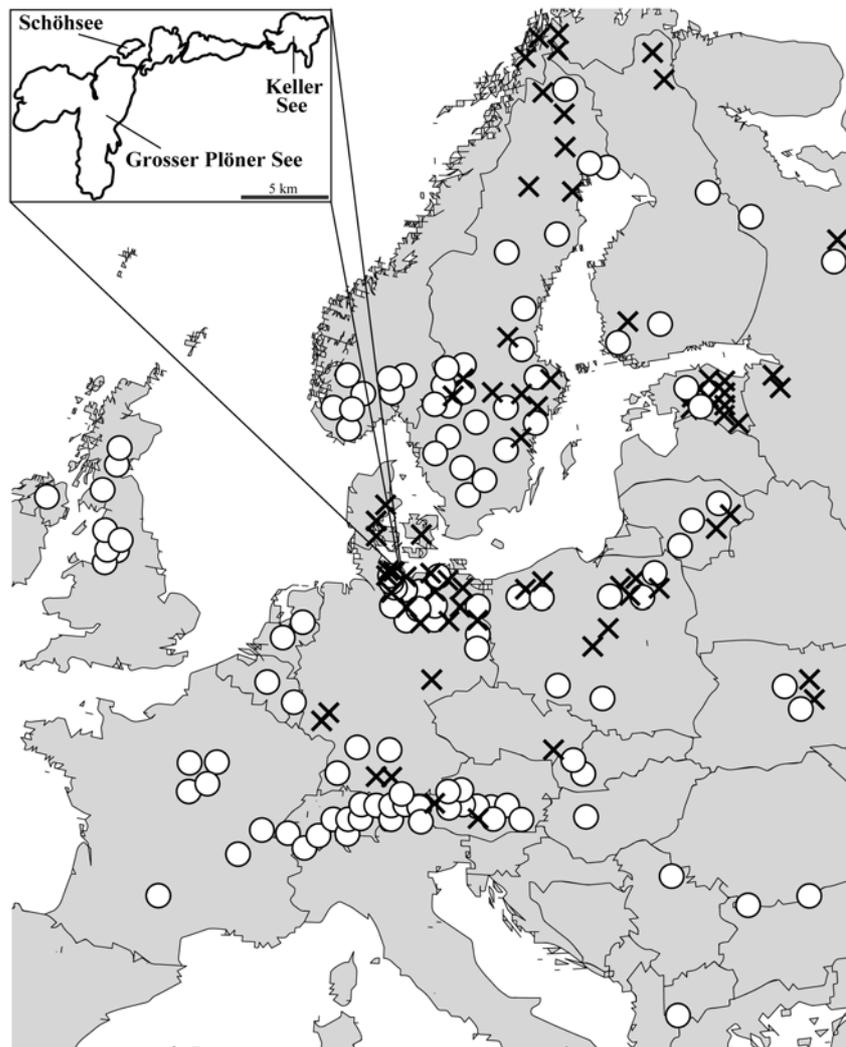


Figure IV.1. Occurrence of *E. gracilis* (circles) and *E. graciloides* (crosses) in Europe, as reported in previously published studies (references available from the author on request). One symbol may refer to multiple bodies of water in areas where they are closely spaced. Kiefer (1978) also stated that *E. gracilis* is found as far east as Siberia, and that the range of *E. graciloides* extends even further to China. The inset shows detail of the lakes under study.

Adult females were sorted, identified and frozen at  $-70^{\circ}\text{C}$  prior to DNA extraction with Chelex-100 beads (Biorad; Walsh et al. 1991). A portion of the mitochondrial gene cytochrome oxidase I (CO I) was amplified using the primer pair C1-J-1718 and C1-N-2191 (Simon et al. 1994). 50  $\mu\text{l}$  reactions contained 2  $\mu\text{l}$  of 1:10 diluted DNA template, 1.5 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  of each dNTP, 400 nM each primer, 1 unit each Taq polymerase and Platinum Taq antibody (GibcoBRL) and reaction buffer. Samples were amplified with a 2 minute hot start at  $94^{\circ}\text{C}$  followed by 35 cycles of  $94^{\circ}\text{C}$  (30 sec.),  $52^{\circ}\text{C}$  (30 sec.),  $72^{\circ}\text{C}$  (45 sec.) and a final 7 minute extension at  $72^{\circ}\text{C}$ . PCR products were verified on a 1.5% agarose gel with ethidium bromide staining and purified using Qiaquick kits (Qiagen). Following cycle sequencing with BigDye reaction mix (Perkin Elmer), sequences were visualized on an ABI 377 automated sequencer. In approximately half of the individuals, PCR products were sequenced in both directions to clarify ambiguous sites. After excluding beginning sites that could not be reliably scored in all individuals, sequence length was 437 bp. Because there were no insertions or deletions, sequences were aligned visually using Sequencher v. 4.0 (Gene Codes Corporation). Sequences have been submitted to NCBI Genbank (accession numbers DQ457129-DQ457160 and DQ459087-DQ459093). The analyses described below were conducted using all individuals (whether or not they possessed identical haplotypes).

Variation for each species (over all individuals, and within each lake) was summarized using the standard statistics  $K$  (number of alleles = unique haplotypes),  $S$  (number of segregating sites), and  $\pi$  (nucleotide diversity: the average pairwise divergence between all individuals). We estimated the population genetic parameter  $\theta = 4N_e\mu$  from  $K$ ,  $S$  and  $\pi$  (This parameter determines the amount of diversity in a population and the shape of a gene genealogy, where  $N_e$  is the effective population size and  $\mu$  is the mutation rate; Halliburton 2004.) Divergence among lakes was tested for statistical significance using exact tests (100,000 Markov chain steps), and quantified using Wright's (1931)  $F_{ST}$  and Excoffier et al.'s (1992)  $\Phi_{ST}$  (10,000 randomizations or permutations in each case). In contrast to  $F_{ST}$ ,  $\Phi_{ST}$  considers the genealogical or evolutionary distance among alleles, in addition to their frequencies in each population. Because the distances among alleles were small, we used the matrix of observed pairwise distances for these calculations, rather than a corrected distance measure. We also compared patterns of mtDNA variation from this study with the microsatellite data of chapter II. For each species, we calculated expected heterozygosity

$H_E$  in each lake for microsatellites, and we calculated differentiation among lakes as  $F_{ST}$ . Statistical significance for microsatellite  $F_{ST}$  in *E. gracilis* was estimated by bootstrapping over seven variable loci. Significance for *E. graciloides* could not be determined because only three loci were available. Calculations were conducted in Arlequin (Schneider et al. 2001), GENEPOP on the Web (Raymond and Rousset 1995) and FSTAT (Goudet 1995).

We conducted two “neutrality tests” for each species (overall, and within each lake) that compare the above summary statistics to expectations from neutral gene genealogies at equilibrium: Tajima’s (1989)  $D$  (which compares  $\theta_S$  to  $\theta_\pi$ ), and Fu’s (1997)  $F$  (which evaluates  $K$ , conditional on  $\pi$ ). Statistical significance was evaluated based on 10,000 simulated samples for  $D$  and  $F$  under the null hypothesis of neutral evolution in a stable population. For each species, we also compared the mismatch distribution shape to that expected under an expanding population using 10,000 bootstrapped replicates. The mismatch distribution is a histogram of pairwise genetic distances between all samples, and its shape can reflect recent demographic changes (Rogers and Harpending 1992). The null hypothesis for this test (as implemented in Arlequin; Schneider et al. 2001) is population expansion.

We estimated the genealogical relationships within each species using two methods. First, we used the parsimony-based network reconstruction method of Templeton and colleagues (e.g., Templeton et al. 1992; Templeton 1998) because it explicitly allows for multifurcations and the persistence of ancestral nodes. Thus, it is likely to be more appropriate for recently diverged alleles than standard phylogenetic tree-building techniques (reviewed by Posada and Crandall 2001). Mitochondrial haplotype networks were estimated for each species using TCS v. 1.13 (Clement et al. 2000) and nested cladistic relationships were inferred using the algorithms of Templeton et al. (1987, 1992). Non-random geographic associations among haplotypes were tested statistically using 10,000 permutations for each clade with the computer program GeoDis (Posada et al. 2000), and the results were interpreted using the most recent inference key (2005) of Templeton and collaborators. The dichotomous inference key examines the geographic distribution of each clade member, assessing whether patterns are consistent with processes such as restricted gene flow (i.e., isolation by distance), allopatric fragmentation or long-distance colonization. Templeton (1998, 2004) provides an overview of the entire process,

known as nested clade phylogenetic analysis (NCPA). Some critics have argued that the final phase of NCPA (the inference key) does not appropriately distinguish among alternative historical hypotheses (Knowles and Maddison 2002), although Templeton (2004) has addressed many of those criticisms and made some modifications to the inference key. For this study, we treated the inferences from NCPA as working hypotheses, and evaluated them based on the neutrality tests described above and qualitative patterns in the gene genealogies.

We also estimated the genealogy for each species with a Bayesian analysis conducted in BEAST (Drummond et al. 2002; Drummond and Rambaut 2003). To help interpret the results of the neutrality tests and NCPA, we used BEAST to estimate  $N_e\mu$  backwards in time based on coalescent events in the genealogy. The results are presented as a Bayesian “skyline plot” generated in TRACER v. 1.2.1 (Drummond et al. 2005). Assuming that mutation rates are equal in *E. gracilis* and *E. graciloides*, the skyline plot depicts relative estimates of population sizes throughout the species’ histories. To interpret the historical processes represented in these plots, we estimated generation times for each species using published and unpublished data.

## Results

*E. gracilis* and *E. graciloides* possessed remarkably different patterns of nucleotide-level variation even though all mtDNA substitutions within each species were silent (i.e., they do not change amino acid composition and are therefore selectively neutral). The maximum divergence was 1.8 % within *E. gracilis*, and 3.0 % within the diapausing species *E. graciloides*. Consequently, the average sequence divergence  $\pi$  was higher in *E. graciloides* (0.0087 versus 0.0075 in *E. gracilis*). However, *E. gracilis* had twice as many polymorphic sites (30 versus 14) and over four times as many unique haplotypes (32 versus 7; see Table IV.1). This led to large discrepancies between  $\theta$  estimated from K and S, and  $\theta$  estimated from  $\pi$  (Table IV.1). As a result, Fu’s F and Tajima’s D showed significant departures from neutral evolution in *E. gracilis* ( $p < 0.05$ ), whereas neither of these tests was statistically significant in *E. graciloides*. Furthermore, the null hypothesis of a sudden population expansion was accepted in *E. gracilis* and rejected in *E. graciloides* ( $p < 0.05$  for all three lakes and for the pooled data set). Consistent with the mtDNA data,

microsatellite variation for the three lakes was higher in *E. gracilis* ( $H_E = 0.85 - 0.87$  in each lake) than in *E. graciloides* ( $H_E = 0.78 - 0.83$ ).

Table IV.1. Summary of mtDNA variation within lakes and across all three lakes, in terms of sample size  $N$ , unique alleles  $K$ , estimates of  $\theta$ , and two neutrality tests. \*  $p \leq 0.05$ ; for  $F$ ,  $p < 0.05$  corresponds to a 2% cutoff in the randomized distribution, rather than 5% (Fu 1997).

	<i>E. gracilis</i>				<i>E. graciloides</i>			
	All individuals	Grosser Plöner See	Keller See	Schöhsee	All individuals	Grosser Plöner See	Keller See	Schöhsee
$N$	65	25	20	20	62	25	19	18
$K$	32	15	12	9	8	5	3	4
$\theta_S$	6.32	5.03	4.51	3.10	2.98	3.44	2.00	4.07
$\theta_\pi$	3.28	3.40	3.88	1.63	3.81	4.44	2.00	4.51
$\theta_K$	26.13	14.91	11.75	5.72	2.22	1.59	0.74	1.28
Tajima's $D$	-1.55*	-1.16	-0.52	-1.68*	0.81	-0.17	0.51	-0.34
Fu's $F$	-12.37*	-7.14*	-3.84	-4.22*	2.59	4.16	3.12	4.67

For mtDNA, exact tests of allele distributions and  $F$ -statistics both demonstrated that the two copepods have diverged genetically among the lakes. Relative levels of differentiation from microsatellites (estimated as  $F_{ST}$ ) and mtDNA ( $\Phi_{ST}$  and  $F_{ST}$ ) were generally consistent with each other, despite some interesting discrepancies (Fig. IV.2, Table IV.2). For mtDNA,  $\Phi_{ST}$  was higher than  $F_{ST}$  in *E. gracilis*, while the opposite was true in *E. graciloides*. Mitochondrial  $\Phi_{ST}$  and microsatellite  $F_{ST}$  were both higher in *E. gracilis* than in *E. graciloides*, but mtDNA  $F_{ST}$  was higher in *E. graciloides*. Lake by lake comparisons also showed differences between some estimates of divergence and hydrologic connectivity. For example, the Keller See and the Grosser Plöner See are separated by the greatest geographic distance, but possess hydrologic connections hospitable to copepods. As a result, these two lakes should have the lowest pairwise genetic divergence, which was true for  $F_{ST}$  in both species for both types of markers, and for mtDNA  $\Phi_{ST}$  in *E. gracilis*. Conversely, the Schöhsee and the Keller See should possess the lowest levels of

connectivity (since both are upstream of the Grosser Plöner See in different directions). The Schöhsee and the Keller See were the most divergent lakes in *E. gracilis* for microsatellites and for  $F_{ST}$  from mtDNA, but only for microsatellites in *E. graciloides* (Table IV.2).

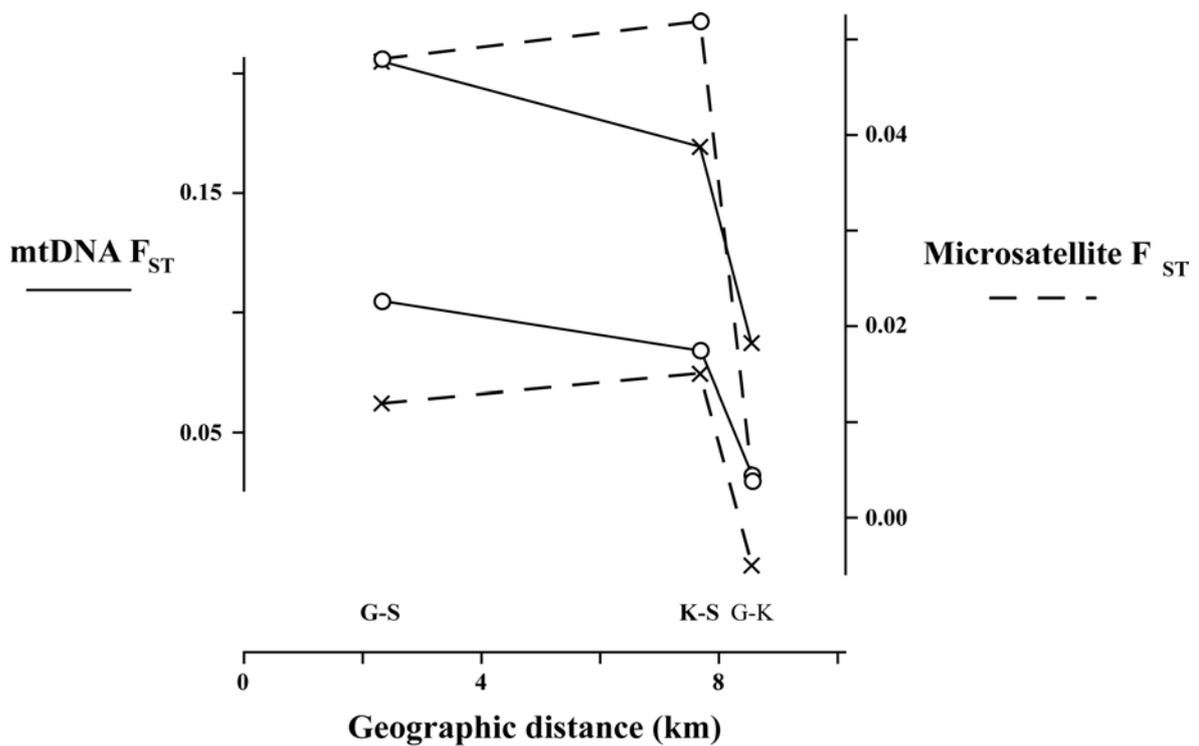


Figure IV.2. Estimates of  $F_{ST}$  from mtDNA (solid line) and microsatellites (dashed line) between pairs of lakes for *E. gracilis* (circles) and *E. graciloides* (crosses). Lake pairs are labeled along the geographic distance axis for the Grosser Plöner See, the Schöhsee, and the Keller See.

Table IV.2. Genetic differentiation among lakes for each *Eudiaptomus* species. G, Grosser Plöner See; K, Keller See; S, Schöhsee. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ , \*\*\*\*\*  $p < 0.00001$ , † significance cannot be determined, although 3/3 loci  $> 0$ .

Contrast:	<i>E. gracilis</i>			<i>E. graciloides</i>				
	across all lakes	G vs. K	G vs. S	K vs. S	across all lakes	G vs. K	G vs. S	K vs. S
<u>mtDNA</u>								
Exact test	*****	*	**	***	*****		****	****
$F_{ST}$	0.07**	0.03	0.11**	0.08**	0.16***	0.09	0.20***	0.17**
$\Phi_{ST}$	0.12***	0.07*	0.09**	0.22*****	0.04	0.07	-0.001	0.05
<u>microsatellites</u>								
$F_{ST}$	0.035*	0.004	0.048*	0.052**	0.009†	-0.005	0.012†	0.015†

Haplotype networks demonstrated the dramatic genetic differences between these species both within and among populations. *E. gracilis* possessed two common alleles and 29 with frequencies of 1 – 3 %. Only three alleles were found in more than one lake. *E. graciloides* also had three alleles present in more than one lake, but had only three common alleles and four rare alleles in total. Geographic patterns of population subdivision were obvious in both species (Fig. IV.3, 4), although the higher levels of genetic diversity in *E. gracilis* facilitate more inferences about evolutionary processes. In *E. gracilis*, significant ( $p < 0.05$ ) geographic associations were found in Clades 3-1, 3-3 and for the total cladogram (see Fig. IV.3). NCPA suggested contiguous range expansions for clades 3-1 and 3-3, and either a range expansion or restricted gene flow at the level of the total cladogram. For *E. graciloides*, NCPA found significant geographic associations ( $p < 0.0001$ ) and evidence for recent allopatric fragmentation in Clade 1-1, representing isolation of the Schöhsee from the other lakes. The older Clade 3-1 showed evidence for restricted gene flow with isolation-by-distance across all three lakes (Fig. IV.4), although the tests of geographic association were only marginally significant ( $p$ -values in six associated tests ranged from 0.07 to 0.09).

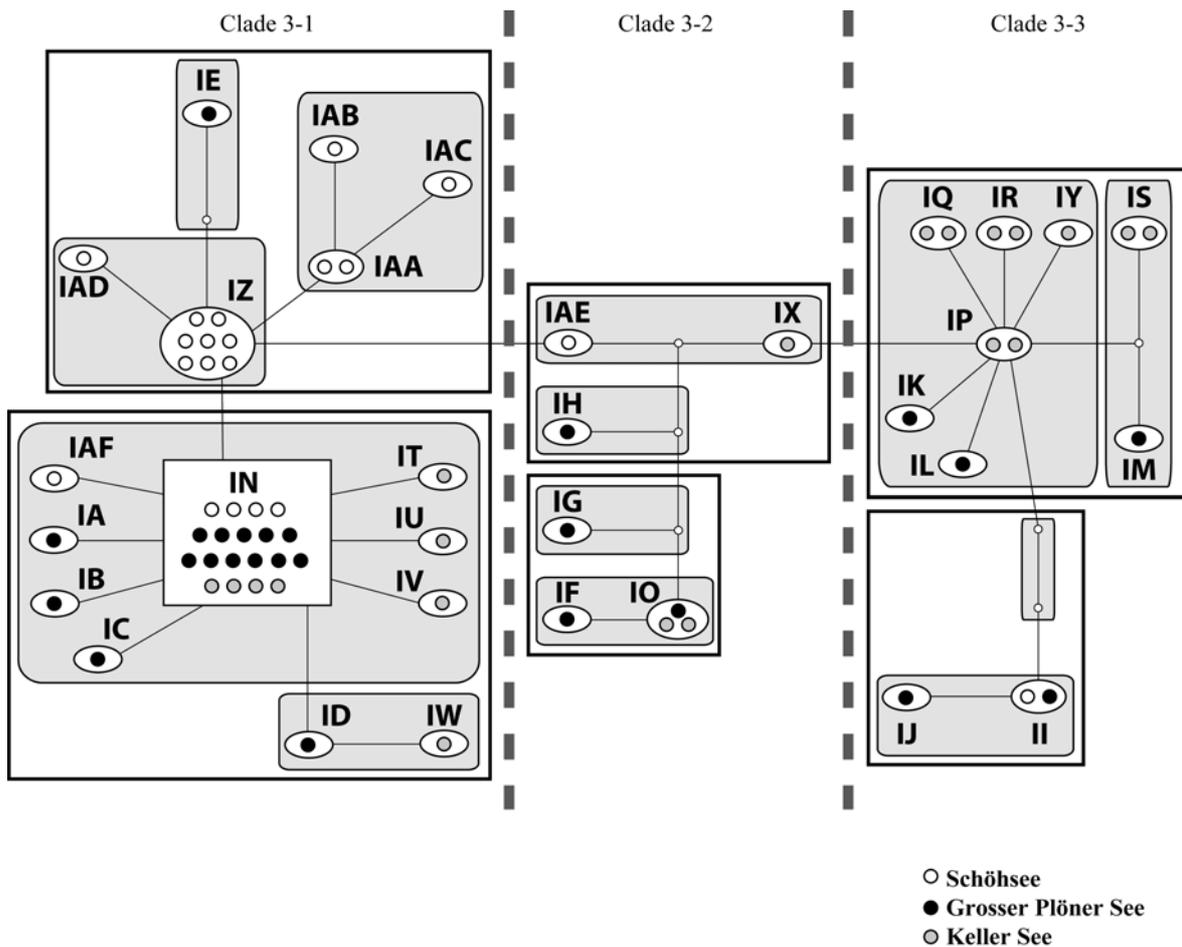


Figure IV.3. Haplotype network for *E. gracilis* (without diapause). Each haplotype is represented as an oval and labeled with a two letter code. Size is proportional to the number of individuals that possess the haplotype. Each line is one mutation, and empty circles represent hypothetical intermediate haplotypes that were not sampled. Boxes and dashed lines separate clades at three hierarchical levels. Haplotype IN (square) has the highest probability of being ancestral. Two ambiguous "loops" (i.e., incompletely resolved portions of the network) were resolved using the criteria of Crandall et al. (1994) and Templeton et al. (1995).

The Bayesian consensus genealogy for *E. graciloides* was completely congruent with the parsimony network in Figure IV.4. The Bayesian consensus for *E. gracilis* was nearly congruent with Figure IV.3, conflicting only because haplotypes IG and IH were monophyletic in the Bayesian tree (weakly supported with a posterior probability of 55 %). As a result, the Bayesian genealogies are not presented here. Topologies in the Bayesian genealogies were interpreted in a skyline plot that depicts mutation-scaled effective population size over time (Fig. IV.5). *E. gracilis* appears to have an effective population size an order of magnitude higher than *E. graciloides*, although the contrast is statistically

significant only for a brief point during their history. Both species have experienced range and/or population expansions, but the time scales are markedly different. The range expansions inferred from NCPA and the neutrality tests in *E. gracilis* predate (in generations) a more recent increase in effective population size for *E. graciloides*. However, *E. graciloides* has a longer generation time due to the diapausing egg bank in this species. To compare the absolute timing of these events, we estimated generation times for both species as follows.

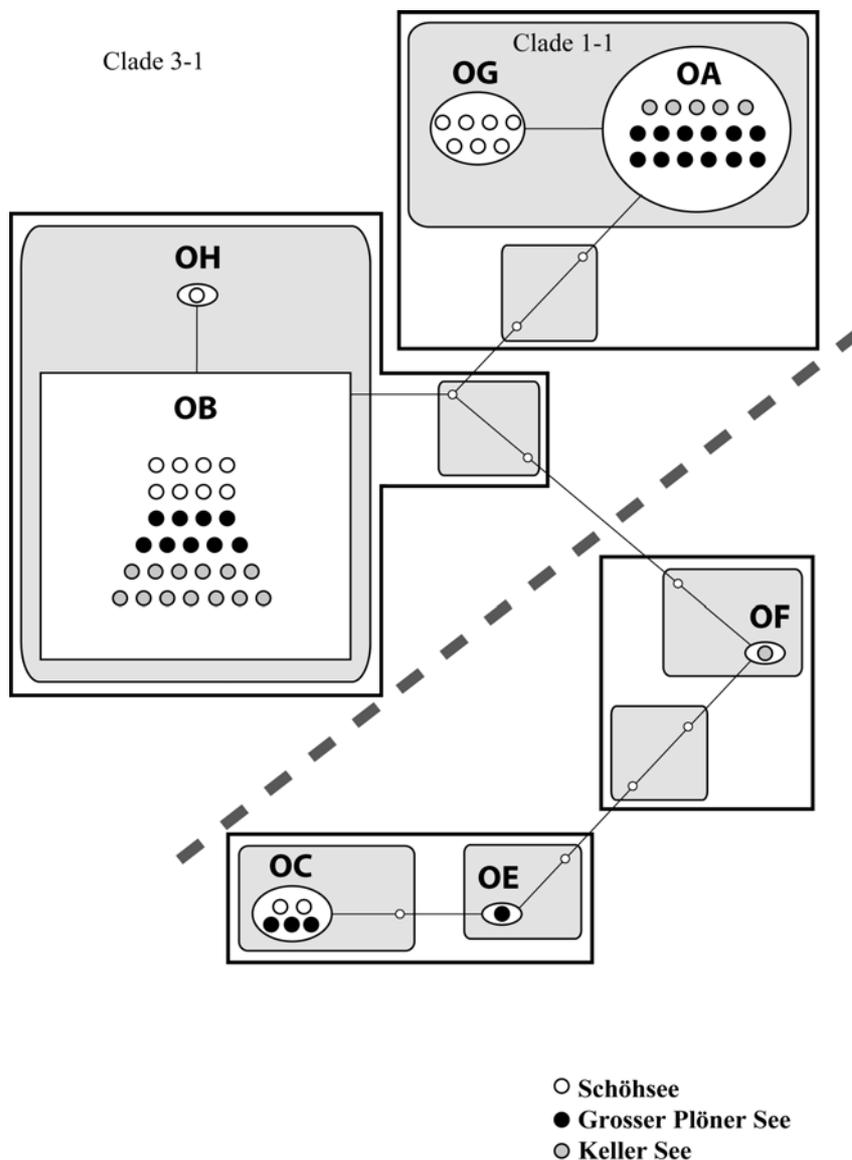


Figure IV.4. Haplotype network for *E. graciloides* (with diapause). Each haplotype is represented as an oval and labeled with a two letter code. Size is proportional to the number of individuals that possess the haplotype. Each line is one mutation, and empty circles represent hypothetical intermediate haplotypes that were not sampled. Boxes and dashed lines separate clades at hierarchical levels. Haplotype OB is hypothesized to be ancestral.

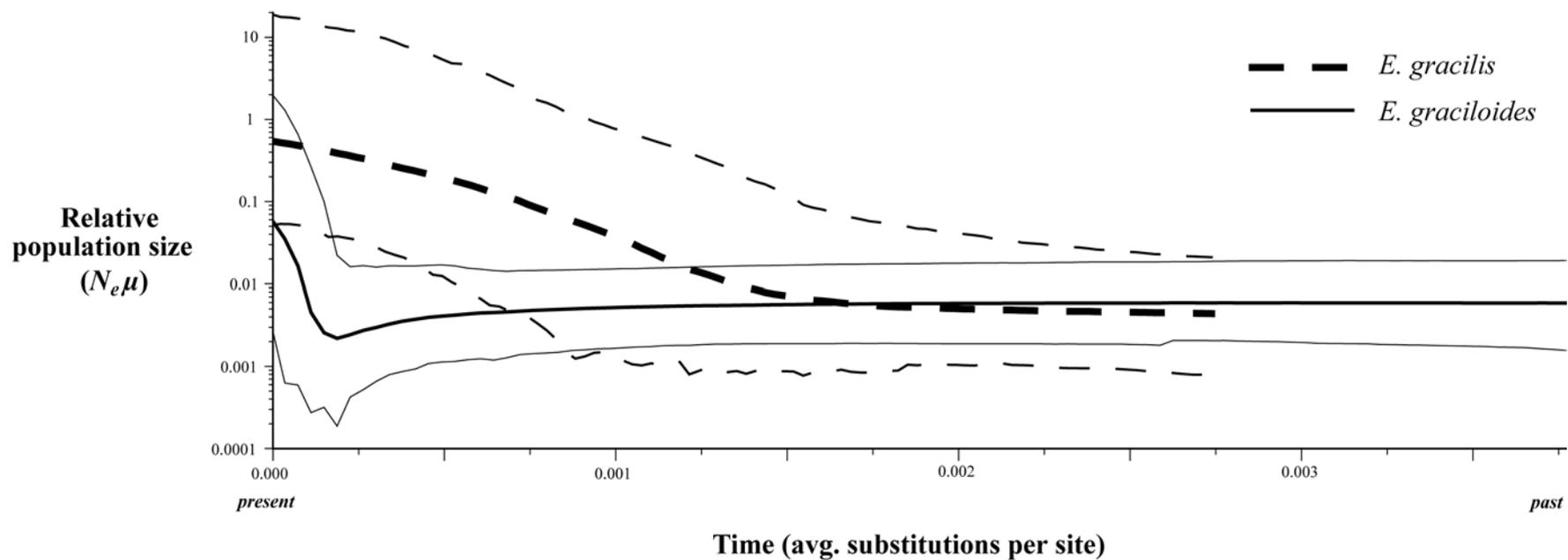


Figure IV.1. "Skyline plot" of  $N_e\mu$  backwards in time (average number of substitutions per site) for *E. gracilis* (without diapause, dashed lines) and *E. graciloides* (with diapause, solid lines). The median value and a 95% confidence interval are shown for each species.

Because *E. gracilis* has nonoverlapping generations and two or three generations per year (Hofmann 1979; Santer et al. 2000; B. Santer, unpublished), we assumed a mean generation time of 0.4 years. For organisms with dormant propagule banks, Nunney (2002) defined generation time as the average age of all parents, regardless of whether they hatch from subitaneous (immediately hatching) eggs or the diapausing egg bank. Thus, mean generation time for *E. graciloides* will be the average for animals derived from subitaneous eggs (weighted by the water-column population size), and animals hatching from the egg bank (weighted by the number of diapausing eggs in the lake sediments and the fraction of these that hatch each year). Like *E. gracilis*, *E. graciloides* also has a generation time of 0.4 years for eggs that immediately hatch in the water column (Hofmann 1979; Santer et al. 2000; B. Santer, unpublished). Based on lake volumes calculated by Santer et al. (2000) and densities reported by Hofmann (1979), the water column population sizes are estimated to be  $5.10 \times 10^{10}$ ,  $6.45 \times 10^{10}$  and  $2.15 \times 10^{12}$  in the Schöhsee, the Keller See and the Grosser Plöner See respectively. Hofmann (1979) showed for the period of his study that population sizes remain fairly constant throughout 2-3 water column generations, suggesting moderate to strong density regulation.

Because we do not have complete information on egg bank dynamics in *E. graciloides*, we rely here on data from another diaptomid copepod population. Diapausing eggs hatch from the well mixed zone in the top 2 cm of lake sediments (Kearns et al. 1996; Hairston and Kearns 2002). In the Schöhsee, the recent sedimentation rate is 0.8 mm per year (Hofmann 1983). Using this sedimentation rate, the mean age of eggs emerging from the top 2 cm is 12.5 years. Based on data from Santer et al. (2000), there are  $8.20 \times 10^9$  diapausing eggs in the top 2 cm of the Schöhsee. The annual emergence rate is unknown for *E. graciloides*, so we estimated the proportion hatching from diapause as somewhere between 1 % and 10 % based on data for *Onychodiaptomus sanguineus* (De Stasio 1989; Hairston et al. 1995; Hairston and Kearns 2002). Taking the 10 % value, the average generation time of the  $8.20 \times 10^8$  eggs emerging annually from diapause (age 12.5 years) and the  $1.28 \times 10^{11}$  subitaneous eggs produced annually in the water column (generation time 0.4 years) is 0.48 years. Comparable calculations for *E. graciloides* in the Keller See and the Grosser Plöner See produced estimates of 0.81 years and 0.41 years per generation, respectively. Averaging over all the three lakes, we estimated 0.57 years per generation. Thus, the *E. graciloides* egg bank increases generation time by 0.17 years (41 %) in comparison with *E.*

*gracilis*. The difference in generation times does not adequately explain the extreme time difference between the two species for start of the population growth phase. Depending on whether the upper or lower confidence intervals in Figure IV.5 are used for calculation, the growth phase in *E. gracilis* begins 5.5 - 13.7 times earlier. Because generation time in *E. graciloides* is only 41 % longer than in *E. gracilis*, the diapausing egg bank of *E. graciloides* cannot account entirely for this discrepancy.

We performed a limited sensitivity analysis on the generation time calculations by arbitrarily changing the fraction of diapausing eggs that emerge annually, since this is the only portion of the calculation that we inferred from another species. Increasing the hatching fraction to 25 % yields an average generation time of 0.59 years, still insufficient to account for differences in timing of the growth phase. Even if 100 % of the eggs in the upper 2 cm of lake sediment hatched annually, generation time is only increased by a factor of 3.3. Other parameter combinations that would increase generation time by at least 5.5 times are also improbable. For example, the timing of population growth in Figure IV.5 could be the same in both species if water column populations of *E. graciloides* are 88 % smaller than estimated, or if adults hatching from diapausing eggs are on average 374 years old.

In summary, *E. graciloides* has colonized this region more recently than *E. gracilis*. Regional population size in *E. gracilis* appears to have increased by two orders of magnitude, leading to an excess of rare, recent mutations. Growth of *E. graciloides* populations has been less dramatic (approximately one order of magnitude), and follows a decline in effective population size consistent with a bottleneck. It is not possible to date these events precisely without assuming (and calibrating) a molecular clock, and it is possible that generation time has recently decreased in both species due to cultural eutrophication. However, it is clear that the growth phase in each species predates any recent human-induced changes in lake ecology or connectivity, and probably occurred thousands of years ago in *E. graciloides*, and on the order of ten thousand years ago in *E. gracilis*.

## Discussion

Prolonged diapause has the potential to influence patterns of genetic variation within and among populations. However, its attributes may have contradictory effects on factors commonly considered important for the genetic structure of natural populations. Diapausing egg banks should increase population size and generation times, permitting founder effects associated with colonization to persist for thousands of generations or longer (Boileau et al. 1992; Gómez et al. 2002). Conversely, diapausing eggs are also assumed to confer high dispersal ability, because the encysted eggs of many species can be passively dispersed in wind or survive gut passage through dispersing vertebrates (e.g., Green et al. 2002). This should lead to greater genetic homogenization of populations, and the rapid dissolution of founder effects if gene flow is high enough. Depending on the relative importance of these factors, high genetic differentiation in species with prolonged diapause may reflect either founder effects or low gene flow despite high dispersal; similarly, the absence of genetic differentiation can be attributed either to a homogeneous founder event or to high gene flow. The analysis of neutral DNA sequence data can help distinguish among some alternative interpretations, because time is an inherent part of haplotype networks, with the most recent mutations found at the tips of the genealogy.

For microsatellites, patterns of isolation-by-distance (IBD) in *E. graciloides* were found on scales <100 km, but not in *E. gracilis* (Chapter II). The absence of IBD in this data for *E. gracilis* and high variance in population divergence at even the smallest spatial scales could be due either to patterns of gene flow that are irregular and not distance dependent, or to persistent founder effects and very low ongoing gene flow. Phylogeographic and demographic analyses of mtDNA sequence data allow for a more precise interpretation. Historical growth in population sizes are evident in both species (Fig. IV.5). In *E. gracilis*, this growth phase corresponds to range expansions seen in the oldest (most inclusive) clades. Thus, the mtDNA genealogy reflects the initial colonization of this region by this species (likely just post-Pleistocene). Standard neutrality tests detected a relative excess of unique mtDNA haplotypes in *E. gracilis* (Table IV.1), as would be expected for a range expansion combined with an associated growth in effective population size. Because population size increased so dramatically (Fig. IV.5), founder effects still persist in this species.

Increases in *E. graciloides* regional population size came at a much later time, possibly associated with a bottleneck in population size (Fig. IV.5). Founder effects in this species were apparently less pronounced than in *E. gracilis*, because changes in effective population size have been less dramatic (Fig. IV.5). This is probably the reason that the null hypothesis of neutral evolution is not rejected in *E. graciloides* using simple statistics such as Tajima's D and Fu's F. Patterns of microsatellite differentiation for *E. graciloides* suggest that gene flow is highest among lakes that are closer together (Chapter II). However, the mtDNA genealogy shows very low levels of gene flow currently between the Schöhsee and the other two lakes (Clade 1-1 in Fig. IV.4), as well as unsorted ancestral variation (haplotype OB). Thus, overland gene flow via animal or wind vectors must be very rare, even when populations are separated by < 3 km. We acknowledge the limitations that must accompany any mtDNA population genetic study, including the assumption of selective neutrality and the inherent stochasticity associated with gene trees (Knowles and Maddison 2002). However, our interpretations are strengthened by an absence of amino acid polymorphisms, and qualitative agreements between the mtDNA and microsatellite loci in terms of intraspecific variation and divergence among populations.

Significant divergence among lakes in mitochondrial and nuclear markers demonstrates that gene flow is low in these copepods, even on small spatial scales among connected lakes. Discrepancies between  $\Phi_{ST}$  and  $F_{ST}$  estimates also demonstrate that genetic differentiation has developed more rapidly in the absence of diapause. Both of these statistics quantify population differentiation, but  $\Phi_{ST}$  accounts for the evolutionary distances between alleles as well as their frequencies. *E. gracilis* has been isolated for enough generations that lake-specific lineages have evolved through mutation and drift (Fig. IV.3), so that  $\Phi_{ST} > F_{ST}$ . In *E. graciloides*, we interpret  $F_{ST} > \Phi_{ST}$  to mean that colonization events involved haphazard assemblages of alleles, and the only new lake-specific mutation to have reached appreciable frequencies is haplotype OG, restricted to the Schöhsee (Fig. IV.4). The microsatellite data provide additional support for this interpretation. Microsatellite mutation rates are several orders of magnitude higher than those for typical nucleotide sequences. Because *E. gracilis* has apparently been established in the lakes we studied for a longer period of time than *E. graciloides*, *E. gracilis* has had more time for unique microsatellite alleles to evolve, and population differentiation is higher. The poor overall match in *E. graciloides* between levels of divergence between

pairs of lakes and hydrologic connectivity provides further evidence that in this species, the lakes we studied are not yet in drift-gene flow equilibrium. Our estimates of generation time suggest that the diapause egg bank of *E. graciloides* has slowed the decay of genetic founder effects, although not as much as one might initially suspect. It is likely that these results extend to traits under natural selection, although the specific patterns would depend on the strength of selective differences among lakes.

The broad geographic distributions of *E. graciloides* and *E. gracilis* (Fig. IV.1) appear to be consistent with our interpretation of the genetic patterns. Although there has not been a systematic biogeographic study of these two species, published data from a variety of sources demonstrate that *E. gracilis* is broadly distributed throughout Europe, and the three lakes included in our study lie near the center of this range. In contrast, our study lakes occur at the western edge of the distribution of the *E. graciloides*, an observation consistent with this species having arrived in our study area more recently.

Nauwerck (1980) has previously noted that *E. gracilis* is more common throughout Sweden. However, in northern Sweden, this species is found only near the Baltic Sea shoreline, corresponding to an area that was at times inundated after the last ice age. In contrast, *E. graciloides* was nearly absent in southern Sweden, more abundant in northern Sweden, and found in mountain regions up to 1000 m. Nauwerck (1980) suggested that *E. gracilis* was the original inhabitant of continental Europe, and colonized southern Sweden and Britain at a time when lower sea levels provided land bridges (with freshwater bodies) that are not present today. *E. graciloides* was assumed to have colonized Scandinavia from the east following glacial retreat in this region. Nauwerck (1980) also hypothesized that *E. gracilis* might have inhabited what is now the Baltic Sea during its freshwater phases: the Baltic Ice Lake > 10,000 years before present, and the Ancylus Lake ca. 8500 years before present. This could have provided the source for *E. gracilis* to colonize lakes near the northern Sweden shoreline. These hypotheses seem to be generally supported by our analyses, and the large amount of mtDNA diversity in *E. gracilis* could be explained by a large regional population of *E. gracilis* in the Baltic Ice Lake and the Ancylus Lake.

Although the presence of diapausing propagule banks is expected to slow drift through increased effective population sizes, we did not find such an effect in *E. graciloides*.

Coalescent analyses suggest that effective population sizes may be an order of magnitude lower in *E. graciloides* than in *E. gracilis*, and there is no indication that *E. gracilis* goes through (contemporary) population bottlenecks of any consequence. The smaller effective population size in *E. graciloides* has led to lower levels of variation for both mtDNA and microsatellites in this species. Independent estimates of effective population size from a temporal analysis of microsatellite variation confirm that population sizes in *E. gracilis* are comparable to those in *E. graciloides*, if not larger (Chapter III).

In summary, we sought to assess the genetic effects of prolonged dormancy empirically with as few confounding factors as possible. The copepods *E. gracilis* and *E. graciloides* are ecologically and trophically similar, but diapause occurs only in *E. graciloides*. We focused on lakes where the two species are found sympatrically, and on a spatial scale of < 10 km to minimize differences among the habitats in terms of regional history and abiotic factors. We found dramatic differences among patterns of genetic differentiation between these two species that we had not predicted based on their life histories. These were at least in part due to differences in their colonization sequence in the region. Interestingly, we found no support for the hypothesis that species with prolonged diapause possess larger effective population sizes than species without diapause. The clear persistence of founder effects in *E. graciloides* precludes an accurate estimate of gene flow in this species. However, it is clear that even on small spatial scales, gene flow is not high enough to homogenize populations in either species.

### **Acknowledgements**

This research began as a challenge from Winfried Lampert to find diapausing eggs of *Eudiaptomus* in the Schöhsee: we did and we didn't, and that turned out to be interesting. We thank E. L. Gross, C. A. Cáceres and E. Blohm-Sievers for help in the lab and field, C. A. Aquadro, S. M. Bogdanowicz, C. A. Cáceres, R. G. Harrison, W. Lampert, T.B.H. Reusch and R.A. Thum for productive discussions. The Bohonak lab provided comments on an early version of the manuscript. Sequencing was performed at the Cornell University Evolutionary Genetics Core Facility. The research was funded by a cooperative international travel grant from the U.S. National Science Foundation (INT- 9603204) to NGH and the Deutscher Akademischer Austauschdienst to W. Lampert, the Max-Planck

Society, a Research Experiences for Undergraduates supplement to NSF grant DEB-9815365 to MDH and NGH, and a Cornell University undergraduate research grant to MDH.

---

## References

- Adrian, R. 1997. Calanoid-cyclopoid interactions: Evidence from an 11-year field study in a eutrophic lake. *Freshw. Biol.* **38**:315-325.
- Anderson, E. C., E. G. Williamson, and E. A. Thompson. 2000. Monte Carlo evaluation of the likelihood for  $N_e$  from temporally spaced samples. *Genetics* **156**:2109-2118.
- Ashley, M. V., and B. D. Dow. 1994. The use of microsatellite analysis in population biology: Background, methods and potential applications. pp. 185-201. *In* B. Schierwater, B. Streit, G. P. Wagner, and R. De Salle [ed.], *Molecular Ecology and Evolution: Approaches and Applications*. Birkhäuser Verlag, Basel.
- Avise, J. C. 1994. *Molecular markers, natural history and evolution*. Chapman and Hall, New York.
- Belkhir, K., P. Borsa, J. Goudet, L. Chikhi, and F. Bonhomme. 1998. Genetix, logiciel sous Windows pour la génétique des populations. Laboratoire Génome et Populations Cnrs Upr 9060: Université de Montpellier II, France.
- Belyea, L.R., and J. Lancaster. 1999. Assembly rules within a contingent ecology. *Oikos* **86**:402-416.
- Berthier, P., M. A. Beaumont, J. M. Cornuet, and G. Luikart. 2002. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: A genealogical approach. *Genetics* **160**:741-751.
- Bilton, D. T., J. Freeland, and B. Okamura. 2001. Dispersal in freshwater invertebrates. *Annu. Rev. Ecol. Syst.* **32**:159-181.
- Bohonak, A. J. 2002. IBD (isolation by distance): A program for analyses of isolation by distance. *J. Hered.* **93**:153-154.
- Bohonak, A. J., and G. K. Roderick. 2001. Dispersal of invertebrates among temporary ponds: Are genetic estimates accurate? *Isr. J. Zool.* **47**:367-386.
- Bohonak, A. J., and D. G. Jenkins. 2003. Ecological and evolutionary significance of dispersal by freshwater invertebrates. *Ecol. Lett.* **6**:783-796.
- Boileau, M. G., P. D. N. Hebert, and S. S. Schwartz. 1992. Non-equilibrium gene frequency divergence: Persistent founder effects in natural populations. *J. Evol. Biol.* **5**:25-39.
- Bosselmann, S. 1975. Population dynamics of *Eudiaptomus graciloides* in Lake Esrom. *Arch. Hydrobiol.* **75**:329-346.

- Brendonck, L., and B. J. Riddoch. 1999. Wind-borne short-range egg dispersal in anostracans (Crustacea: Branchiopoda). *Biol. J. Linn. Soc.* **67**:87-95.
- Brendonck, L., M. Hamer, and A. Thiery. 1992. Occurrence of tetrahedral eggs in the Streptocephalidae Daday (Branchiopoda: Anostraca) with descriptions of a new subgenus, *Parastreptocephalus*, and a new species, *Streptocephalus (Parastreptocephalus) zuluensis*. *J. Crustac. Biol.* **12**:282-297.
- Bucklin, A., and P. H. Wiebe. 1998. Low mitochondrial diversity and small effective population sizes of the copepods *Calanus finmarchicus* and *Nannocalanus minor*: Possible impact of climatic variation during recent glaciation. *J. Hered.* **89**:383-392.
- Caballero, A. 1994. Developments in the prediction of effective population size. *Heredity* **73**:657-679.
- Cáceres, C. E. 1998. Interspecific variation in the abundance, production and emergence of *Daphnia* diapausing eggs. *Ecology* **79**:1699-1710.
- Cáceres, C. E., and D. A. Soluk. 2002. Blowing in the wind: A field test of overland dispersal and colonization by aquatic invertebrates. *Oecologia* **131**:402-408.
- 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)<sub>n</sub> microsatellite markers. *Am. J. Hum. Genet.* **52**:922-927.
- Charalambidou, I., and L. Santamaría. 2002. Waterbirds as endozoochorous dispersers of aquatic organisms: A review of experimental evidence. *Acta. Oecol. Int.* **23**:165-176.
- Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: A computer program to estimate gene genealogies. *Mol. Ecol.* **9**:1657-1659.
- Cohen, G. M., and J. B. Shurin. 2003. Scale-dependence and mechanisms of dispersal in freshwater zooplankton. *Oikos* **103**:603-617.
- Cousyn, C., L. De Meester, J. K. Colbourne, L. Brendonck, D. Verschuren, and F. A. M. Volckaert. 2001. Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proc. Natl. Acad. Sci. U.S.A.* **98**:6256-6260.
- Crandall, K. A., A. R. Templeton, and C. F. Sing. 1994. Intraspecific phylogenetics: Problems and solutions. pp. 273-297. In R. W. Scotland, D. J. Siebert, and D. M. Williams [ed.], *Models in phylogeny reconstruction*. Clarendon Press, Oxford.

- Crow, J. F., and C. Denniston. 1988. Inbreeding and variance effective population numbers. *Evolution* **42**:482-495.
- Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Harper and Row, New York.
- De Meester, L. 1996. Local genetic differentiation and adaptation in freshwater zooplankton populations: Patterns and processes. *Ecoscience* **3**:385-399.
- De Meester, L., A. Gomez, B. Okamura, and K. Schwenk. 2002. The monopolization hypothesis and the dispersal-gene flow paradox in aquatic organisms. *Acta. Oecol. Int.* **23**:121-135.
- De Stasio, B. T. Jr. 1989. The seed bank of a freshwater crustacean: Copepodology for the plant ecologist. *Ecology* **70**:1377-1389.
- Drummond, A. J., and A. Rambaut. 2003. BEAST (Bayesian Evolutionary Analysis Sampling Trees). 1.2. <http://evolve.zoo.ox.ac.uk/beast/>
- Drummond, A. J., G. K. Nicholls, A. G. Rodrigo, and W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**:1307-1320.
- Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**:1185-1192.
- Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakraborty. 1992. Genetic variation at 5 trimeric and tetrameric tandem repeat loci in 4 human-population groups. *Genomics* **12**:241-253.
- Edwards, K. J., J. H. A. Barker, A. Daly, C. Jones, and A. Karp. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques* **20**:758-760.
- Einsle, U. 1993. Crustacea: Copepoda: Calanoida und Cyclopoida. Süßwasserfauna von Mitteleuropa, Vol. 8, Gustav Fischer Verlag, Stuttgart.
- Ekman, S. 1964. Die jährliche Populationsentwicklung des planktischen Kopepoden *Diaptomus graciloides* im subarktischen Nordschweden. *Zool. Bidrag. Uppsala* **36**:277-293.
- Ellegren, H., M. Johansson, K. Sandberg, and L. Andersson. 1992. Cloning of highly polymorphic microsatellites in the horse. *Anim. Genet.* **23**:133-142.

- Ellner, S. P. and N. G. Hairston Jr. 1994. Role of overlapping generations in maintaining genetic variation in a fluctuating environment. *Am. Nat.* **143**:403-417.
- Ellner, S. P., N. G. Hairston Jr., C. M. Kearns, and D. Babai. 1999. The roles of fluctuating selection and long-term diapause in microevolution of diapause timing in a freshwater copepod. *Evolution* **53**:111-122.
- 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.
- Figuerola, J., A. J. Green, and L. Santamaría. 2003. Passive internal transport of aquatic organisms by waterfowl in Donana, south-west Spain. *Global Ecol. Biogeogr.* **12**:427-436.
- Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: A review. *Genet. Res.* **66**:95-107.
- Frankham, R., J. D. Ballou, and D. A. Briscoe. 2002. *Introduction to conservation genetics*. Cambridge University Press, Cambridge.
- Franklin, I. R. 1980. Evolutionary change in small populations. pp. 11-34. *In* M. E. Soule, and B. A. Witcox [ed.], *Conservation Biology: An Evolutionary-Ecological Perspective*. Sinauer, Sunderland, MA.
- Fu, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915-25.
- Fußmann, G. 1996. Die Kontrolle der Rotatorien im Pelagial eines mesotrophen Sees durch Bottom-up- und Top-down-Prozesse: Freilandbeobachtungen und Enclosure-Experimente. Ph.D Thesis, Univ. Kiel, Germany.
- Goldstein, D. B., and C. Schlotterer. 1999. *Microsatellites Evolution and Applications*. Oxford University Press, Oxford.
- Gómez, A., G. J. Adcock, D. H. Lunt, and G. R. Carvalho. 2002. The interplay between colonization history and gene flow in passively dispersing zooplankton: Microsatellite analysis of rotifer resting egg banks. *J. Evol. Biol.* **15**:158-171.
- Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J. Hered.* **86**:485-486.
- Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995).

- Green, A. J., J. Figuerola, and M. I. Sánchez. 2002. Implications of waterbird ecology for the dispersal of aquatic organisms. *Acta. Oecol. Int.* **23**:177-189.
- Hairston, N. G. Jr. 1996. Zooplankton egg banks as biotic reservoirs in changing environments. *Limnol. Oceanogr.* **41**:1087-1092.
- Hairston, N. G. Jr., and E. J. Olds. 1984. Population differences in the timing of diapause: Adaptation in a spatially heterogeneous environment. *Oecologia* **61**:42-48.
- Hairston, N. G. Jr., and B. T. J. De Stasio. 1988. Rate of evolution slowed by a dormant propagule pool. *Nature* **336**:239-242.
- Hairston, N. G. Jr., and C. M. Kearns. 2002. Temporal dispersal: Ecological and evolutionary aspects of zooplankton egg banks and the role of sediment mixing. *Integ. and Comp. Biol.* **42**:481-491.
- Hairston, N. G. Jr., R. A. van Brunt, C. M. Kearns, and D. R. Engstrom. 1995. Age and survivorship of diapausing eggs in a sediment egg bank. *Ecology* **76**:1706-1711.
- Halliburton, R. 2004. Introduction to population genetics. Prentice Hall, Upper Saddle River, NJ.
- Hanski, I. 1998. Metapopulation dynamics. *Nature* **396**:41-49.
- Harrison, S., and E. Bruna. 1999. Habitat fragmentation and large-scale conservation: What do we know for sure? *Ecography* **22**:225-232.
- Havel, J. E., and J. B. Shurin. 2004. Mechanisms, effects, and scales of dispersal in freshwater zooplankton. *Limnol. Oceanogr.* **49**:1229-1238.
- Hedrick, P. W. 1995. Genetic polymorphism in a temporally varying environment: Effects of delayed germination or diapause. *Heredity* **75**:164-170.
- Hedrick, P. W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**:313-318.
- Hedrick, P. 2005. Large variance in reproductive success and the  $N_e/N$  ratio. *Evolution* **59**:1596-1599.
- Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* **58**:247-276.
- Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* **68**:87-112.
- Hofmann, W. 1979. Characteristics of syntopic populations of *Eudiaptomus gracilis* (SARS) und *E. graciloides* (LILLJEBORG) in three lakes of different trophic levels. *Arch. Hydrobiol.* **86**:1-12.

- Hofmann, W. 1983. Stratigraphy of Cladocera and Chironomidae in a core from a shallow north German lake. *Hydrobiologia* **103**:235-239.
- Hutchison, D. W., and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**:1898-1914.
- Jarne, P., and P. J. L. Lagoda. 1996. Microsatellites, from molecules to populations and back. *Tree* **11**:424-429.
- Jehle, R., J. W. Arntzen, T. Burke, A. P. Krupa, and W. Hödl. 2001. The annual number of breeding adults and the effective population size of syntopic newts (*Triturus cristatus*, *T. marmoratus*). *Mol. Ecol.* **10**:839-850.
- Jenkins, D. G., and M. O. Underwood. 1998. Zooplankton may not disperse readily in wind, rain, or waterfowl. *Hydrobiologia* **387-388**:15-21.
- Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. *Ecol. Lett.* **7**:1225-1241.
- Kearns, C. M., N. G. Hairston Jr., and D. H. Kesler. 1996. Particle transport by benthic invertebrates: Its role in egg bank dynamics. *Hydrobiologia* **332**:63-70.
- Kiefer, F. 1978. Freilebende Copepoda. Die Binnengewässer, Schweizbart'sche-Verlagsbuchhandlung, Stuttgart.
- Kijas, J. M. H., J. C. S. 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 particles. *Biotechniques* **16**:656-662.
- Kimura, M., and J. F. Crow. 1963. The measurement of effective population numbers. *Evolution* **17**:279-288.
- Knowles, L. L., and W. P. Maddison. 2002. Statistical phylogeography. *Mol. Ecol.* **11**:2623-2635.
- Koblizkova A., J. Dolezel, and J. Macas. 1998. Subtraction with 3' modified oligonucleotides eliminates amplification artifacts in DNA libraries enriched for microsatellites. *Biotechniques* **25**:32-38.
- Korovchinsky, N. M., and O.S. Boikova. 1996. The resting eggs of the Ctenopoda (Crustacea: Branchiopoda): A review. *Hydrobiologia* **320**:131-140.
- 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.

- Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**:203-221.
- Louette, G., and L. De Meester. 2005. High dispersal capacity of cladoceran zooplankton in newly founded communities. *Ecology* **86**:353-359.
- Lynch, M., and K. Spitze. 1994. Evolutionary genetics of *Daphnia*, pp. 109-128. In L.A. Real [ed.], *Ecological Genetics*. Princeton University Press, Princeton.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**:209-220.
- McCauley, D. E. 1993. Genetic consequences of extinction and recolonization in fragmented habitats. pp. 217-233. In P. M. Kareiva, J. G. Kingsolver, and R. B. Huey [ed.], *Biotic interactions and global change*. Sinauer, Sunderland, MA.
- Morin, P. A., G. Luikart, R. K. Wayne, and S. W. Group. 2004. Applications of single nucleotide polymorphisms (SNPs) in ecology, evolution, and conservation. *Trends Ecol. Evol.* **19**:208-216.
- Nauta, M. J., and F. J. Weissing. 1996. Constraints on allele size at microsatellite loci: Implications for genetic differentiation. *Genetics* **143**:1021-1032.
- Nauwerck, A. 1963. Die Beziehungen zwischen Zooplankton und Phytoplankton im See Erken. *Symb. Bot. Uppsal.* **17**:1-163.
- Nauwerck, A. 1980. Die Verbreitung der Familie *Diaptomidae* Sars in Nordschweden. *Arch. Hydrobiol.* **89**:247-264.
- Nies, G., and T. H. Reusch. 2005. Evolutionary divergence and possible incipient speciation in postglacial populations of a cosmopolitan aquatic plant. *J. Evol. Biol.* **18**:19-26.
- Nunney, L. 2002. The effective size of annual plant populations: The interaction of a seed bank with fluctuating population size in maintaining genetic variation. *Am. Nat.* **160**:195-204.
- Pálsson, S. 2000. Microsatellite variation in *Daphnia pulex* from both sides of the Baltic Sea. *Mol. Ecol.* **9**:1075-1088.
- Pamilo, P., and O. Savolainen. 1999. Post-glacial colonization, drift, local selection and conservation value of populations: A northern perspective. *Hereditas* **130**:229-238.
- Pollak, E. 1983. A new method for estimating the effective population size from allele frequency changes. *Genetics* **104**:531-548.

- Posada, D., and K. A. Crandall. 2001. Intraspecific gene genealogies: Trees grafting into networks. *Trends Ecol. Evol.* **16**:37-45.
- 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. *Mol. Ecol.* **9**:487-488.
- Rai, H. (1982): Primary production of various size fractions of natural phytoplankton communities in a north German lake. *Arch. Hydrobiol.* **95**:395-412.
- Raymond, M., and F. Rousset. 1995. Genepop (version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* **86**:248-249.
- Rogers, A. R., and H. C. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**:552-569.
- Rozen, S., and H.J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. pp. 365-386. In S. Krawetz, and S. Misener [ed.], *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowo, NJ. (Code available at [http://www-genome.wi.mit.edu/genome\\_software/other/primer3.html](http://www-genome.wi.mit.edu/genome_software/other/primer3.html).)
- Santer, B. 1998. Life cycle strategies of free-living copepods in freshwaters. *J. Mar. Syst.* **15**:327-336.
- Santer, B., E. Blohm-Sievers, C. E. Cáceres, and N. G. Hairston Jr.. 2000. Life-history variation in the coexisting freshwater copepods *Eudiaptomus gracilis* and *Eudiaptomus graciloides*. *Arch. Hydrobiol.* **149**:441-458.
- Schlötterer, C., and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**:211-215.
- Schneider, S., D. Roessli, and L. Excoffier. 2001. Arlequin: A software for population genetics data analysis, 2.0 (2001 update). <http://anthro.unige.ch/arlequin>
- Schwartz, M. K., D. A. Tallmon, and G. Luikart. 1998. Review of DNA-based census and effective population size estimators. *Anim. Cons.* **1**:293-299.
- 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. *Ann. Entomol. Soc. Am.* **87**:651-701.
- Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**:264-279.

- Slatkin, M., and W. P. Maddison. 1990. Detecting isolation by distance using phylogenies of genes. *Genetics* **126**:249-260.
- Stemberger, R. S. 1995. Pleistocene refuge areas and postglacial dispersal of copepods of the northeastern United States. *Can. J. Fish. Aquat. Sci.* **52**:2197-2210.
- Tajima, F. 1989. The effect of change in population size on DNA polymorphism. *Genetics* **123**:597-601.
- Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**:413-418.
- Templeton, A. R. 2004. Statistical phylogeography: Methods of evaluating and minimizing inference errors. *Mol. Ecol.* **13**:789-809.
- Templeton, A. R., E. Boerwinkle, and C. F. Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping: I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**:343-352.
- Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data: III. Cladogram estimation. *Genetics* **132**:619-633.
- Templeton, A. R., E. J. Routman, and C. A. Phillips. 1995. Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**:767-782.
- Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. Microchecker: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**:535-538.
- Vanoverbeke, J., and L. De Meester. 1997. Among-population genetic differentiation in the cyclical parthenogen *Daphnia magna* (Crustacea, Anomopoda) and its relation to geographic distance and clonal diversity. *Hydrobiologia* **360**:135-142.
- Walsh, P. S., D. A. Metzger, and R. G. Higuchi. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**:506-513.
- Wang, J. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genet. Res.* **78**:243-257.

- Wang, J., and A. Caballero. 1999. Developments in predicting the effective size of subdivided populations. *Heredity* **82**:212-226.
- Wang, J., and M. C. Whitlock. 2003. Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* **163**:429-446.
- Waples, R. S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**:379-391.
- Waples, R. S. 2002. Evaluating the effect of stage-specific survivorship on the  $N_e/N$  ratio. *Mol. Ecol.* **11**:1029-1037.
- Weider, L. J., and P. D. N. Hebert. 1987. Microgeographic genetic heterogeneity of melanic *Daphnia pulex* at a low-Arctic site. *Heredity* **58**:391-400.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population-structure. *Evolution* **38**:1358-1370.
- Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration:  $F_{ST} = 1/(4Nm + 1)$ . *Heredity* **82**:117-125.
- Williamson, E. G., and M. Slatkin. 1999. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics* **152**:755-761.
- Wright, S. 1921. Systems of mating. *Genetics* **6**:111-178.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**:97-159.
- Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* **87**:430-431.
- Wright, S. 1943. Isolation by distance. *Genetics* **28**:114-138.
- Wright, S. 1946. Isolation by distance under diverse systems of mating. *Genetics* **31**:39-59.
- Wright, S. 1969. *Evolution and genetics of populations*, Vol 2., Univ. Chicago Press, Chicago.
- Zane, L., L. Bargelloni, and T. Patarnello. 2002. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **11**:1-16.

## **Danksagung**

An erster Stelle möchte ich mich herzlich bei Thorsten B. H. Reusch und Winfried Lampert für die umfassende Betreuung bei der Anfertigung dieser Arbeit bedanken.

Ein weiterer Dank gilt Nelson G. Hairston Jr., der mir vor der Anfertigung dieser Arbeit einen Aufenthalt an der Cornell-Universität ermöglichte. In dieser Zeit hatte ich die Möglichkeit, molekularbiologische Arbeitsmethoden zu lernen und mich in die Prozesse zur Etablierung von Mikrosatellitenbibliotheken einzuarbeiten. Das war der erste Baustein für meine Doktorarbeit. Vielen Dank!

Kirsten Kessler hat mit mir einige Jahre nicht nur das Büro, sondern alle Höhen und Tiefen geteilt. Vielen Dank für alles, und dass diese Freundschaft über alle Länder- und Institutsgrenzen bestehen bleibt! (Und vielen Dank fürs Korrekturlesen...)

Bei Georg Nies bedanke ich mich für die gemeinsame Zeit, in der er mir als Kollege und besonders als Freund zur Seite stand.

Weiterhin bedanke ich mich bei Silke Carstensen, Ilka Dankert, Catherine Schmuck und Sybille Liedtke für den Einsatz im Labor, und insbesondere bei Dir Silke für eine schöne gemeinsame (Arbeits-)Zeit.

Barbara Santer hat mir stets bei Fragen rund um die Copepoden hilfreich zur Seite gestanden. Vielen Dank!

Nicole Aberle-Malzahn danke ich nicht nur für die in den letzten Jahren entstandene Freundschaft, sondern auch für die Funktion als inoffizielle "Mentorin". Vielen Dank, dass ich mich immer mit allen Fragen an Dich wenden konnte.

Die Mitarbeiter vom Rechenzentrum haben mich in allen Fragen rund um den Computer immer tatkräftig unterstützt. Danke schön! Ein besonderer Dank für die Hilfe und Geduld geht an dieser Stelle an Derk Wachsmuth.

Harald Deiwick und Siegfried Voß danke ich für die Hilfe bei den Probenahmen im Freiland. Die „Ausflüge“ mit Euch haben immer viel Spaß gemacht.

Ich danke unserer Tagesmutter Frauke Hachmann-Wachsmuth, für die Betreuung von Pelle. Außerdem gibt es noch viele Freunde, die mich auf diesem Weg begleitet haben: Mira Christen, Michael Häberli, Svenja Babel, Anne Kreutz, Sandra Gieske, Andrea Strichau-Plügg. Bei Euch allen, und denen die nicht namentlich erwähnt wurden, möchte ich mich auch bedanken.

Ein besonderer Dank geht nun an meine Familie. Von ganzem Herzen danke ich meinem Mann Marc, der mich mit ganzer Kraft bei dieser Arbeit und weit darüber hinaus unterstützte. Dem kleinen Pelle danke ich für jeden Tag! Du bringst so viel Freude in mein Leben, dass ich auch ohne Schlaf und mit viel Unordnung noch so einiges nebenbei schaffe. Meiner Mutter und Schwester danke ich für ihre Unterstützung und für das „Schaffen“ von noch ein bisschen mehr Arbeitszeit. Papa, danke für alles!

## Appendix 1

Names and positions of sampled lakes. No = numbering of lakes, N = North, E = East.

no	Lake name	Lake position	
1.1	Schöhsee	54° 09' N	10° 26' E
1.2	Großer Plöner See	54° 08' N	10° 24' E
1.3	Tramm	54° 10' N	10° 25' E
1.4	Kellersee	54° 10' N	10° 35' E
1.5	Schierensee	54° 11' N	10° 28' E
1.6	Holzsee	54° 09' N	10° 11' E
1.7	Stolpe	54° 07' N	10° 12' E
1.8	Selent	54° 18' N	10° 26' E
2.1	Schwerin	53° 39' N	11° 28' E
2.2	Medewege	53° 39' N	11° 25' E
3.1	Großer Vätersee	53° 00' N	13° 37' E
3.2	Hardenbecker Haussee	53° 15' N	13° 29' E
4.1	Zerdno	53° 36' N	16° 12' E
4.2	Drawsko	53° 35' N	16° 10' E
5.1	Mikolajki	53° 46' N	21° 35' E
5.2	Betdany	53° 45' N	21° 35' E
6.1	Stirniu	55° 15' N	25° 37' E
6.2	Dusios	54° 17' N	23° 04' E
6.3	Asveja	55° 00' N	25° 36' E
7.1	Chernoe	59° 34' N	30° 07' E
7.2	Filkino	59° 33' N	30° 08' E

## Appendix 2

Genetic diversity within *E. gracilis* and *E. graciloides* populations. Localities are abbreviated with numbers (see Fig. II.1). <sup>(1)</sup> = individuals that have been used to test for polymorphism of markers (Chapter I).  $H_E$  = expected heterozygosity,  $H_O$  = observed heterozygosity,  $n$  = number of individuals. Inbreeding coefficients ( $F_{IS}$ -values) for each used primer pair and population calculated after Weir and Cockerham (1984). \* =  $F_{IS}$ -values deviate significantly from Hardy-Weinberg-expectations.

Locality	<i>E. graciloides</i> :						<i>E. gracilis</i> :									
	n	H <sub>(E)</sub>	H <sub>(O)</sub>	F <sub>IS</sub> for microsatellite loci			n	H <sub>(E)</sub>	H <sub>(O)</sub>	F <sub>IS</sub> for microsatellite loci						
				EGO2	EGO7	EGO10				EG11	EG13	EG18	EG112	EG113	EG117	EG135
1.1 <sup>(1)</sup>	45	0.84	0.73	0.092	0.224*	0.092	45	0.85	0.72	0.054	0.149*	0.151*	0.354*	0.234*	0.145	0.081
1.2	31	0.79	0.58	0.217*	0.506*	0.014	45	0.87	0.60	0.100	0.200*	0.242*	0.760*	0.376*	0.168	0.453*
1.3	45	0.83	0.67	0.298*	0.269*	-0.028	45	0.88	0.68	0.127*	0.105*	0.176	0.602*	0.354*	-0.014	0.327*
1.5							45	0.83	0.54	0.237*	0.269*	0.253*	0.476*	0.414*	0.112*	0.697*
1.6							45	0.78	0.58	0.069	0.196*	0.133	0.489*	0.288*	-0.143	0.615*
1.7	45	0.81	0.70	0.186*	0.154*	0.101										
1.8	45	0.84	0.71	0.234*	0.167*	0.065										
1.4	45	0.78	0.57	0.518*	0.256*	0.092	45	0.87	0.70	0.019	0.206*	0.075	0.611*	0.346*	0.192	0.005
2.1	45	0.78	0.60	0.266*	0.307*	0.097	44	0.83	0.50	0.112	0.347*	0.243*	0.720*	0.501*	0.363*	0.593*
2.2							45	0.83	0.60	0.128	0.339*	0.186*	0.580*	0.346*	-0.019	0.436*
3.1	45	0.70	0.62	0.189	0.165*	0.032	45	0.76	0.64	-0.029	0.348*	0.018	0.187	0.056*	0.123	0.479*
3.2	45	0.80	0.69	0.100*	0.230*	0.108	45	0.82	0.62	-0.056	0.320*	-0.012	0.629*	0.270*	0.120	0.552*
4.1	41	0.75	0.57	0.236*	0.282*	0.203*	45	0.80	0.54	0.280*	0.210*	0.094*	0.554*	0.570*	0.078	0.429*
4.2	45	0.79	0.63	0.289*	0.274*	-0.026	45	0.81	0.54	0.446*	0.239*	0.023	0.559*	0.316*	0.135	0.709*
5.1	45	0.84	0.70	0.283*	0.166*	0.05	45	0.83	0.60	0.077*	0.271*	0.129*	0.533*	0.209*	0.280*	0.560*
5.2	45	0.85	0.72	0.151*	0.253*	0.027	45	0.85	0.60	0.109*	0.197*	0.118	0.691*	0.255*	0.186	0.549*
6.1	45	0.88	0.78	0.268*	0.015	0.074	45	0.80	0.65	0.020	0.311*	0.023*	0.668*	0.363*	0.022	0.047
6.2							45	0.78	0.48	0.246*	0.609*	0.075	0.856*	0.308*	-0.008	0.692*
6.3	45	0.86	0.63	0.298*	0.328*	0.166										
7.1	45	0.86	0.42	0.551*	0.557*	0.454*										
7.2	45	0.88	0.54	0.350*	0.599*	0.221*										

**Appendix 3**

Pairwise comparison of population differentiation between 17 populations of *E. graciloides* and 16 populations of *E. gracilis*, respectively. Above diagonal: pairwise  $F_{ST}$ -values ( $\theta$ ). Below diagonal: geographical distances (km). Localities are abbreviated with numbers (see Fig. II.1). In brackets: abbreviations of lakes that were connected with the studied lake by watersheds (100 km range).

*E. graciloides*

	1.1	1.2	1.3	1.4	1.7	1.8	2.1	3.1	3.2	4.1	4.2	5.1	5.2	6.1	6.3	7.1	7.2
1.1		0.012	-0.001	0.015	0.026	0.012	0.046	0.068	0.024	0.032	0.023	0.014	0.019	0.032	0.016	0.088	0.073
1.2 (1.3, 1.4)	0.4		0.007	-0.005	0.023	0.045	0.044	0.075	0.009	0.027	0.025	0.018	0.019	0.041	0.028	0.103	0.085
1.3 (1.2, 1.4)	0.4	0.8		0.010	0.021	0.023	0.046	0.066	0.017	0.032	0.022	0.017	0.020	0.036	0.025	0.090	0.071
1.4 (1.2, 1.3)	7.3	8.0	8.5		0.021	0.042	0.049	0.080	0.019	0.033	0.034	0.027	0.029	0.049	0.028	0.111	0.092
1.7	13.8	6.4	11.9	23.5		0.045	0.064	0.088	0.028	0.053	0.044	0.026	0.029	0.040	0.029	0.090	0.068
1.8	13.2	16.2	12.9	13.5	22.5		0.059	0.078	0.048	0.055	0.046	0.043	0.043	0.034	0.025	0.087	0.076
2.1	80.0	76.8	81.0	70.0	86.5	86.2		0.056	0.029	0.030	0.023	0.047	0.039	0.073	0.063	0.115	0.090
3.1	250.0	247.0	250.0	246.0	260.0	260.0	159.0		0.041	0.029	0.032	0.063	0.061	0.095	0.082	0.161	0.137
3.2	229.0	225.0	229.0	210.0	260.0	260.0	144.0	30.0		0.014	0.007	0.021	0.016	0.044	0.032	0.105	0.080
4.1	378.0	378.0	378.0	366.0	388.0	366.0	300.0	188.0	184.0		-0.003	0.029	0.023	0.068	0.049	0.138	0.118
4.2	378.0	378.0	378.0	366.0	388.0	366.0	300.0	188.0	184.0	0.5		0.020	0.014	0.050	0.038	0.113	0.092
5.1	730.0	730.0	730.0	720.0	810.0	800.0	670.0	570.0	575.0	352.0	352.0		-0.004	0.024	0.009	0.078	0.061
5.2	730.0	730.0	730.0	720.0	810.0	800.0	670.0	570.0	575.0	352.0	352.0	2.0		0.026	0.014	0.079	0.058
6.1	1000.0	1000.0	1000.0	990.0	1000.0	1000.0	920.0	840.0	830.0	632.0	632.0	270.0	270.0		0.007	0.047	0.037
6.3	990.0	990.0	990.0	980.0	990.0	990.0	910.0	830.0	820.0	620.0	620.0	260.0	260.0	24.0		0.059	0.052
7.1	1340.0	1340.0	1340.0	1340.0	1340.0	1340.0	1314.0	1300.0	1300.0	1100.0	1100.0	840.0	840.0	566.0	594.0		0.002
7.2	1340.0	1340.0	1340.0	1340.0	1340.0	1340.0	1314.0	1300.0	1300.0	1100.0	1100.0	840.0	840.0	566.0	594.0	5.0	

*E. gracilis*

	1.1	1.2	1.3	1.4	1.5	1.6	2.1	2.2	3.1	3.2	4.1	4.2	5.1	5.2	6.1	6.2
1.1		0.048	0.044	0.052	0.072	0.101	0.069	0.080	0.106	0.066	0.085	0.080	0.080	0.061	0.061	0.074
1.2 (1.3, 1.4)	0.4		-0.004	0.004	0.018	0.024	0.013	0.035	0.036	0.016	0.023	0.016	0.017	0.016	0.037	0.055
1.3 (1.2, 1.4)	0.4	0.8		0.006	0.022	0.030	0.021	0.040	0.048	0.025	0.031	0.026	0.020	0.015	0.040	0.054
1.4 (1.2, 1.3)	7.3	8.0	8.5		0.042	0.054	0.027	0.042	0.063	0.033	0.053	0.044	0.036	0.033	0.036	0.052
1.5	3.4	4.8	4.0	5.0		0.031	0.028	0.046	0.037	0.033	0.023	0.022	0.015	0.015	0.063	0.083
1.6	15.1	9.4	13.4	23.4	19.6		0.022	0.052	0.027	0.028	0.013	0.017	0.017	0.022	0.056	0.087
2.1	80.0	76.8	81.0	70.0	79.9	92.4		0.010	0.028	0.015	0.022	0.019	0.022	0.023	0.042	0.072
2.2	81.0	77.8	82.0	71.0	80.9	93.4	1.0		0.064	0.044	0.053	0.053	0.041	0.044	0.054	0.080
3.1	250.0	247.0	250.0	246.0	250.0	260.0	159.0	160.0		0.018	0.018	0.018	0.032	0.041	0.063	0.113
3.2	229.0	225.0	229.0	210.0	229.0	229.0	144.0	145.0	30.0		0.012	0.014	0.019	0.024	0.039	0.080
4.1	378.0	378.0	378.0	385.0	378.0	385.0	300.0	300.0	188.0	184.0		-0.001	0.015	0.019	0.060	0.093
4.2	378.0	378.0	378.0	385.0	378.0	385.0	300.0	300.0	188.0	184.0	0.5		0.014	0.016	0.057	0.087
5.1	730.0	730.0	730.0	720.0	730.0	740.0	670.0	670.0	570.0	575.0	352.0	352.0		0.001	0.041	0.059
5.2	730.0	730.0	730.0	720.0	730.0	740.0	670.0	670.0	570.0	575.0	352.0	352.0	2.0		0.036	0.047
6.1	1000.0	1000.0	1000.0	990.0	1000.0	1000.0	920.0	920.0	840.0	830.0	632.0	632.0	270.0	270.0		0.044
6.2	850.0	850.0	850.0	850.0	850.0	850.0	786.0	785.0	700.0	700.0	480.0	480.0	122.0	122.0	166.0	

## **Curriculum Vitae**

von Martina Zeller, geb. 27.05.1972 in Kiel, deutsche Staatsangehörigkeit

- 2001 – 2006           Doktorarbeit am Max-Planck-Institut für Limnologie in Plön
- 2000 (1 Semester)   Wissenschaftliche Mitarbeit im Labor von Nelson G. Hairston Jr.,  
Department of Ecology and Evolutionary Biology, Limnology  
Section, Cornell University, Ithaca, NY, USA
- 1999 – 2000           Diplomarbeit am Max-Planck-Institut für Limnologie in Plön  
Titel: Untersuchungen zur Induktion der Dauereiproduktion beim  
calanoiden Copepoden *Eudiatomus graciloides*
- 1994 – 1999           Studium der Biologie and der Christian-Albrechts-Universität zu  
Kiel mit Zoologie als Hauptfach sowie Biochemie und Biologische  
Meereskunde als Nebenfächer (10 Semester)
- 1991 – 1994           Ausbildung zur Bankkauffrau an der Sparkasse Eckernförde
- 1991                   Allgemeine Hochschulreife erlangt an der Jungmannschule  
Eckernförde

## **Erklärung**

Hiermit versichere ich, dass diese Abhandlung -abgesehen von der Beratung durch meine akademischen Lehrer- nach Form und Inhalt meine eigene Arbeit ist und dass ich keine anderen als die angegebenen Hilfsmittel und Quellen verwendet habe. Die mitochondriellen Rohdaten aus Kapitel IV sind teilweise in die Master Thesis von Matt D. Holland, Cornell University, Ithaca, NY, USA, mit eingegangen. Ansonsten hat diese Arbeit weder ganz noch zum Teil an anderer Stelle im Rahmen eines Prüfungsverfahrens vorgelegen.

Teile dieser Arbeit wurden bereits veröffentlicht oder als Manuskripte eingereicht. Kapitel I wurde in Molecular Ecology Notes publiziert mit Thorsten B. H. Reusch als Koautor. Kapitel II ist in Limnology and Oceanography erschienen mit Thorsten B. H. Reusch und Winfried Lampert als Koautoren. Kapitel IV ist eine Gemeinschaftsarbeit von A. J. Bohonak, M. D. Holland, B. Santer, M. Zeller, C. M. Kearns, and N. G. Hairston Jr. und befindet sich beim Archiv für Hydrobiologie in Revision. Kapitel III ist in der Vorbereitung, um bei Evolution (Brief Communications) eingereicht zu werden mit Thorsten B. H. Reusch und Winfried Lampert als Koautoren.

Kiel, den

Martina Zeller