Klaus Richter

Genetic structure in European populations of the earthworm *Lumbricus terrestris*

kassel university press

Die vorliegende Arbeit wurde vom Fachbereich Ökologische Agrarwissenschaften der Universität Kassel als Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) angenommen.

Erster Gutachter: Prof. Dr. Bernard Ludwig, Universität Kassel
Zweiter Gutachter: Prof. Dr. Reiner Finkeldey, Universität Göttingen
Dritter Gutachter: PD. Dr. Martin Potthoff, Universität Göttingen

Tag der mündlichen Prüfung 02. März 2009

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar

Zugl.: Kassel, Univ., Diss. 2009
ISBN print: 978-3-89958-872-9
ISBN online: 978-3-89958-873-6
URN: http://nbn-resolving.de/urn:nbn:de:0002-8737

Printed in Germany

Contents

1. Introduction

1.1. Terrestrial ecosystems and ecosystem engineers

Terrestrial ecosystems consist of two strongly interacting compartments: the aboveground subsystem and the soil subsystem. The aboveground subsystem is, at least on a first glimpse, more eye-catching, but in most cases the soil subsystem exhibits a far greater biodiversity particularly in terms of microorganisms and invertebrate animals. The main task of soil organisms is the maintenance of a recycling system for organic material produced aboveground, either by direct mineralisation of these resources or by regulation of the soil subsystem and its food webs (Wardle 2002).

For a complete understanding of processes in soils and their tight interactions with the aboveground subsystem, it is required to explore the biodiversity of soil organisms on different taxonomical scales, the composition of soil communities and the structure within species and single populations. Because of the hidden life of soil organisms, which often inhibits the direct observation of populations, molecular markers can be helpful tools to indirectly investigate origin, dispersal and dynamics of populations and the differences between multiple populations of single species. For example, the genetic structure, gene flow, dispersal ability and isolation of

forest populations of the collembolan species *Orchesella cincta* (L.) (Collembola, Insecta) was inferred with two molecular techniques (van der Wurff et al. 2003).

Earthworms are a very important component in many soils. The importance of this group is clearly expressed by the term 'ecosystem engineers', introduced for organisms strongly altering the physical structure of habitats and influencing other organisms (Jones et al. 1994). Ecosystem engineers in soils are earthworms, ants and termites. These taxa interact with soil microorganisms, other soil animals and plants (Lavelle et al. 1995).

The present study was conducted to investigate the genetic diversity, population structure and postglacial migration of the European earthworm species *Lumbricus terrestris* L. (Lumbricidae; Annelida; Oligochaeta) by using AFLP (Amplified Fragment Length Polymorphism) as molecular marker.

1.2. Biology of *Lumbricus terrestris*

European earthworm species are commonly assigned to three ecological categories mainly based on the feeding behaviour, the burrowing patterns and the preferred soil horizon (Bouché 1971, 1977; Edwards and Bohlen 1996). The 'epigeic' species are surface dwelling, litter feeding, fast reproducing and often pigmented earthworms, while the 'endogeic' species inhabit the mineral soil horizon, ingest a mixture of mineral soil and organic material, reproduce more slowly and are less pigmented. *L. terrestris* is a typical member of the third group: the 'anecic' earthworms. *L. terrestris* lives in deep vertical burrows, reaching from the mineral soil horizon to the soil surface. These burrows are (semi-)permanent, i. e. inhabited by an

individual for a longer period of time, in contrast to the burrows of endogeic earthworms, which are shortly used. *L. terrestris* is a comparatively large earthworm species and has a pigmented body, which is darkest at the anterior parts on the dorsal side. The preferred food are leaves and other plant materials collected on the soil surface and pulled into the entrances of the burrows. During exploration for food and during mating, which also takes place on the soil surface mostly by night, individuals of *L. terrestris* attempt to stay attached to their burrows with their posterior parts (Edwards and Bohlen 1996; Nuutinen and Butt 1997).

L. terrestris is simultaneously hermaphroditic and reproduces biparentally by a reciprocal exchange of sperm between two individuals during copulation. After copulation, the clitellum, a glandular saddle-shaped part of the epidermis, produces cocoons containing a nutritional fluid in which eggs and sperm are released. In these cocoons, deposited in the soil, fertilisation and embryonic development takes place. *L. terrestris* shows a complex mating behaviour including visits of potential partners prior to copulation, a specific copulation position and piercing the partner's skin with the setae of some segments, probably to influence the uptake of sperm (Nuutinen and Butt 1997; Koene et al. 2005). This behaviour makes a self-fertilisation in *L. terrestris* despite simultaneous presence of male and female reproductionary organs impossible. Some other earthworm species are parthenogenetic (Edwards and Bohlen 1996).

L. terrestris and virtually all other European earthworm species are semi-continuous breeders with numerous reproduction events during their life, not taking place on specific points in time but during an extended breeding season with appropriate conditions for reproduction (Olive and Clark 1978). The production of cocoons is

highest in spring, in the early months of summer and in autumn, and it is reduced during winter (Gerard 1967).

It is difficult to estimate the length of life span and of life cycle components, like hatching of cocoons and maturation, in natural populations of earthworms. Incubation times of cocoons and growth to maturity strongly depend on environmental conditions like temperature and moisture. Under best conditions and in laboratory cultures *L. terrestris* needs approximately half a year to hatch and develop a clitellum (Lowe and Butt 2005). Development times in natural populations are longer and often lead to a hatching of cocoons in the year following copulation, indicated by a peak of hatching in spring. The first copulation of these individuals can occur in the autumn of the same year but will more likely take place in the following year. This estimate of time to reach sexual maturity and the average life span of *L. terrestris* of only 1.25 years (Satchell 1967), makes overlapping generations for the majority of earthworms in natural populations unlikely.

1.3. Biogeography of *Lumbricus terrestris*

The geographical distribution of European earthworms was severely altered at least two times in the past. The first event was the last glaciation in Europe, which lead to an extinction of earthworms and subsequent recolonization in great parts of Northern Europe. The second one was caused by human long range movements leading to transportation of European earthworms to other continents.

Dispersal of earthworms can either be an active or passive mechanism. In case of *L. terrestris*, a comparison of active dispersal speed and the current distribution of this species suggests a mainly passive

dispersal of this species. Whereas a passive transport of worms and cocoons by rivers and larger animals is theoretically possible, there is strong evidence that *L. terrestris* is dependent on disturbance of habitats to establish a new population in previously earthworm-free habitats (Enckell et al. 1986; Tiunov et al. 2006). This indicates that today's distribution of *L. terrestris* was facilitated by human transport of earthworms and cocoons and human impact on habitat soils.

The invasion of North America by European earthworms since the arrival of European settlers is an interesting example which is investigated with great effort. In this case, the earthworms were intentionally or unintentionally introduced to parts of Northern America which were free of earthworms since the last glaciation. Since then, European earthworms have been changing ecosystems in North America which had evolved without earthworms after the last glaciation, for example hardwood forest (Hendrix 2006).

In cases of introduced earthworms an accurate characterisation of the invasive populations regarding origin, expansion potential and population structure is of great interest. Molecular methods are a very promising tool to investigate these topics. Sequencing of mitochondrial (mt)DNA helped to identify the dispersal mechanism and the locations where multiple introductions of invasive *Dendrobaena octaedra* (Savigny) in boreal forests in Alberta, Canada most likely had occurred (Cameron et al. 2008). *Aporrectodea tuberculata* (Eisen) probably had also been introduced several times to the USA, indicated by a high number of polymorphisms at several enzyme loci analysed in individuals from 30 sampling locations in New York (Stille et al. 1980). By using RAPDs (Randomly Amplified Polymorphic DNA) it was possible to calculate similarity

indices between populations of *Aporrectodea* (Savigny) species in Australian agricultural soils, and it was proposed to use this technique to identify origins of these species (Dyer et al. 1998). However, these studies only investigated the introduced earthworm populations. Studies comparing both, European and non-European populations, are rare. A phylogenetic study based on mtDNA sequences in *Octolasion tyrtaeum* (Savigny) showed almost no genetic differences between a low number of individuals sampled in Germany and Canada (Heethoff et al. 2004). A comparison of *Octolasion cyaneum* (Savigny) from several countries in Europe and in Australia revealed a rather high clonal diversity in the Australian population measured by enzyme electrophoresis, indicating that this species had been introduced several times (Terhivuo and Saura 2006). The presumably most extensive study on relationships between European and introduced earthworms revealed a higher similarity between *D. octaedra* from Northern Europe and Canada than between these locations and Greenland, which probably was a refuge for this freeze-tolerant earthworm species during the last glaciation (Hansen et al. 2006).

The present study is predominantly based on European sampling locations and deals with postglacial expansion of *L. terrestris* in Europe. However, the topic of invasive expansion of *L. terrestris* is addressed by including some individuals from Canada to the analysis. This is probably the first attempt to compare European and invasive populations of a biparental reproducing earthworm species.

1.4. AFLPs

Methods used in genetic diversity research and population genetics either focus on a small number of loci, like enzyme electrophoresis,

microsatellite analysis or sequencing of certain parts of organellic or genomic DNA, or use multiple loci, like RAPD, RFLP (Restriction Fragment Length Polymorphism) and AFLP. While the former methods' loci are highly informative, the latter methods achieve their informative and statistical power by applying a high number of loci across the whole genome.

Because of the anonymous character of AFLPs, the origin of an AFLP fragment in terms of belonging to nuclear or organellic DNA or exact position on chromosomes can only be determined by laborious sequencing of single AFLP fragments. Therefore, only the high numbers of AFLP fragments enhances the probability of having a representative sampling across the whole genome. Nevertheless, some studies indicate that most AFLP markers originate from centromeric regions of chromosomes (Alonso-Blanco et al. 1998; Saliba-Colombani et al. 2000).

AFLPs are generated by restriction of DNA and subsequent PCR reactions (Vos et al. 1995). Restriction is generally performed with two restriction endonucleases, a 'rare cutter' and a 'frequent cutter', digesting DNA at a 6-base pair (bp) recognition site and a 4-bp recognition site, respectively. Prior to PCR, adapter molecules are ligated to the DNA fragments. The adapter molecules with known nucleotide sequence and the remaining parts of the restriction enzyme recognition sites then serve as binding sites for PCR primers. These primers have been extended on their 3'-ends with additional nucleotides. Generally two steps of PCR are applied using primer pairs with one additional nucleotide in the first reaction (preselective PCR) and three additional nucleotides in the second reaction (selective PCR). The PCR steps decrease the absolute number of fragments because the primers amplify only fragments contain-

ing nucleotides complementary to the additional nucleotides in the primers and increase the amount of DNA of the finally amplified fragments for a proper visualisation of the fragments. Visualisation is done after electrophoresis of the amplified fragments using polyacrylamide gels or by capillary electrophoresis leading to patterns (fingerprints) consisting either of bands or peaks (Meudt and Clarke 2007).

Each band or peak represents a fragment of DNA flanked by a recognition site for the 'rare' and the 'frequent cutter' and containing the nucleotides necessary for binding of the additional nucleotides of the primers (Vos et al. 1995). Absence of a fragment can be caused by mutations, insertions or deletions in the restriction sites, by absence of the nucleotides complementary to the additional nucleotides in the PCR primers or by insertions and deletions between two restriction sites leading to a different length of a fragment and therefore a different position in the AFLP fingerprint.

There are two ways to interpret AFLP fingerprints (Bonin et al. 2007). The first way is to use the obtained patterns of bands or peaks directly for analysis. Because there are no assumption about underlying alleles or inheritance of the bands/peaks, the AFLP fingerprints can be seen as a phenotypic marker and in this context AFLPs are sometimes referred as phenotypic fingerprints. Analysis methods following this interpretation usually rely on the calculation of band sharing indices between individuals. In their review on statistical analysis of polymorphic data Bonin and colleagues called these methods 'band-based methods' (Bonin et al. 2007). These methods are contrasted by methods based on the estimation of the underlying allele frequencies of the AFLP fingerprints. For this purpose, the stretch of DNA that produces a

band/peak is interpreted as an allele, the PRESENCE-allele, whereas every DNA at the same location in the genome which does not produce a band in the AFLP procedure is regarded as the corresponding null-allele, called ABSENCE-allele. The PRESENCE-allele is dominant, and in a diploid organism the combinations PRESENCE-allele/PRESENCE-allele and PRESENCE-allele/ABSENCE-allele produce the PRESENCE-phenotype (a band or peak), whereas the combination ABSENCE-allele/ABSENCE-allele produces the ABSENCE-phenotype. The estimation of allele frequencies can be done with different approaches, which are applied in this study. In difference to the individual-oriented band-based approaches these methods concentrate on the characteristics of a whole population. These methods are called 'frequency-based methods' (Bonin et al. 2007).

Even though AFLPs can be used for every organism from microorganisms to plants to animals and humans, the proportion of studies using this method in animals, especially in invertebrates, is low (Bensch and Akesson 2005). Expanding the taxonomic range of the AFLP method was another goal of the present study.

1.5. Objectives

Objectives of this study were:

- Adaption of the AFLP-technique to the species *L. terrestris*
- Estimation of genetic diversity at several sampling locations in Europe
- Estimation of genetic differentiation between pairs of locations and overall differentiation

- Comparison between genetic and geographic distances between locations
- Finding geographic regions in Europe where populations are connected via genetic exchange and genetic boundaries dividing these regions

2. Material and Methods

2.1. Sampling locations and geographic data

Lumbricus terrestris was sampled at single locations in Finland, Sweden, Austria and Bosnia-Herzegovina, at two sampling locations in France and at eight sampling locations in Germany. Most sampling locations were grasslands, like pastures, lawns or park-like areas, and one sampling location in Germany was located in a forest (table 2.1). Individuals from Canada were bought in shops selling fishing equipment. These earthworms had most likely been sampled at Canadian golf courses to be sold as fishing baits. For these individuals no exact geographic data were available. However, most of the available information suggested Ontario as the sampling region.

The geodetic coordinates of the sampling locations (except Canada) were projected to X- and Y-coordinates of a Cartesian coordinate system using the Lambert conformal conic projection. The central meridian was placed at 10° E, the origin latitude at 40° N and the standard parallels at 46° N and 58° N, respectively. To obtain only positive X- and Y-coordinates a false easting of 1000 km was performed. The Datum was World Geodetic System 84, the ellipsoid was WGS84. The projection was done with the software GEOTRANS Version 2. 4. 1 (National Geospatial Intelligence Agency of the United States Department of Defense 2007).

Table 2.1.: Abbreviations and description of sampling locations, sampling methods and number of individuals (n) used in AFLP procedures. Countries are encoded in International Olympic Committee country codes, and different locations in one country are additionally encoded with three letters taken from the site descriptions (bold letters). Longitude and latitude are in decimal notation.

Abbreviation	Country	Description	Lat.	Long.	sampled in	sampling method	n
FIN	Finland	lawn, Jokioinen, Campus MTT Agrifood Research	60.81 N	23.48 E	2004	handsorting	21
SWE	Sweden	lawn, Uppsala, Campus SLU Ultuna	59.80 N	17.65 E	2004	electrical	21
GER-Tim	Germany	pasture, near ***Tim***mendorfer Strand, Baltic Sea	53.96 N	10.76 E	2004	handsorting	32
GER-Sol	Germany	pasture, Relliehausen, ***Sol***ling	51.78 N	9.69 E	2004	electrical	19
GER-EgF	Germany	***f***orest, ***Egg***e Range	51.65 N	9.03 E	2006	electrical + chemical	27
GER-EgP	Germany	***p***asture, ***Egg***e Range	51.65 N	9.03 E	2006	electrical	27
GER-Pad	Germany	lawn, airport near ***Pad***erborn	51.54 N	8.58 E	2004	electrical	21
GER-Kas	Germany	pasture, Domain Frankenhausen, experimental farm University ***Kas***sel	51.41 N	9.44 E	2004	electrical	36
GER-Lei	Germany	pasture, near ***Lei***pzig	51.35 N	12.60 E	2004	electrical	21
GER-Bay	Germany	lawn, near ***Bay***reuth	49.94 N	11.72 E	2004	electrical	26
FRA-Bru	France	meadow, near ***Bru***z	48.40 N	1.45 W	2006	chemical	29
FRA-Béd	France	meadow, near ***Béd***ee	48.12 N	2.10 W	2007	chemical	26
AUT	Austria	lawn, near Vienna, experimental garden, University of Natural Resources and Applied Life Sciences Vienna	48.21 N	16.52 E	2004	electrical	30
BIH	Bosnia-Herzegovina	garden, Sarajevo	43.84 N	18.36 E	2006	handsorting	31
CAN	Canada	fishing bait, see text for details	not available	not available	not available	not available	27

2.2. Sampling of *Lumbricus terrestris*

Lumbricus terrestris was sampled by chemical extraction, electrical extraction or by collecting animals from the soil surface by night (table 2.1). For chemical extraction 10 L of formaldehyde solution (0.1 %) were applied to a 0.25 m^2 surface of soil delimited by a metal ring with a height of 10 cm (Raw 1959). Electrical extraction was done with an 'electrical octet machine', WORM-EX III, supplied by the 'Gesellschaft für angewandte Ökologie mbH', Germany (Thielemann 1986, 1989; Schmidt 2001). The octet machine consisted of eight rod-shaped electrodes encircling an area of 0.125 m^2 of soil and a unit to control intensity and duration of voltages applied to opposed electrodes. Prior to electrical extraction, litter and vegetation were carefully removed. The applied voltage and the duration of electrical extraction were adjusted to the particular on-site conditions. Several chemical or electrical extractions per sampling site were performed until a sufficient number of earthworms was sampled.

All sampled earthworms were washed with water and placed in buckets with wet filter paper for transportation to the lab. Earthworms were either placed in a refrigerator, if the date of DNA extraction was within a few days, or in buckets with soil at a temperature of 15 °C in the dark, if the earthworms had to be kept for a longer time. In cases of long-time storage, earthworms were fed with oat flakes every week. Prior to DNA extraction, identification of earthworms was done based on external characteristics (Schaefer 1992; Bouché 1972; Christian and Zicsi 1999). If there was any doubt about an earthworm belonging to the species *L. terrestris*,

for example in case of earthworms being in a subadult stage, the earthworm was not used for DNA extraction and AFLP analyses.

2.3. Isolation of DNA

Earthworms were placed on wet filter paper over night to allow emptying of guts. They were washed in distilled water, killed in 99 % ethanol and were kept for at least two hours in 99 % ethanol to harden tissues. Afterwards, earthworms were washed consecutively in distilled water and 70 % ethanol. Tissue sampling was performed by cutting a piece of tissue between prostomium and clitellum on the dorsal site of the earthworm with a scalpel, making sure that the cut was performed without perforating the gut (figure 2.1).

Tissue samples were washed in freshly prepared 70 % ethanol and placed in collection tubes for DNA isolation. Remains of earthworms were stored at −20 °C. In cases where DNA isolation had to be repeated, the corresponding stored sample was washed consecutively in distilled water and 70 % ethanol and then cutting of tissue was performed like described for fresh samples.

DNA isolation was done with a commercially available DNA isolation kit (DNeasy 96 Tissue Kit, Qiagen). DNA was stored at −20 °C. In total, the DNA of 394 individuals was isolated. DNA isolation of 98 individuals was done twice to obtain additional samples for repeatability tests of AFLP peaks.

Quantity and quality of DNA was assessed by electrophoresis of four µL sample on 1.5 % agarose gels stainend with ethidium bromide.

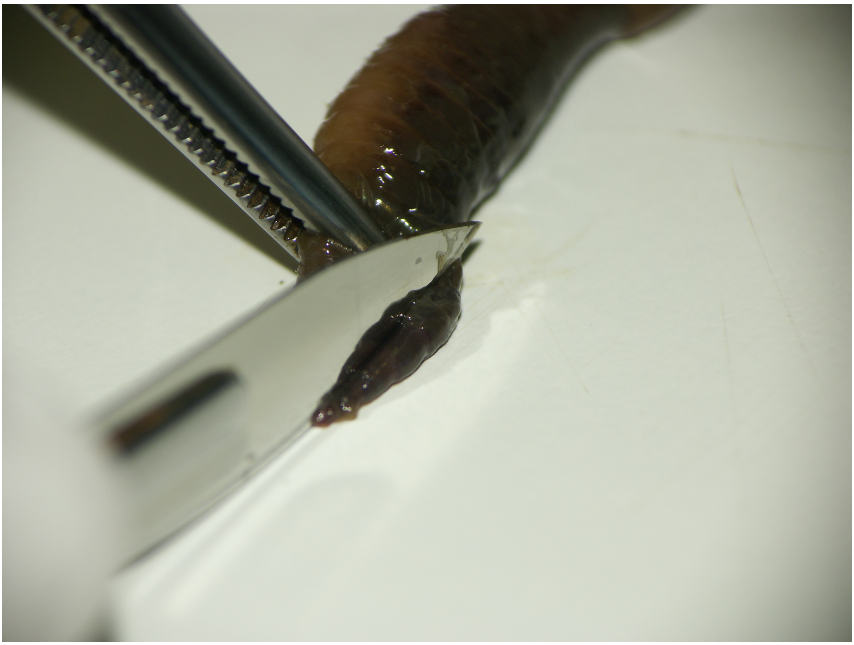

Figure 2.1.: Tissue sampling of an adult *Lumbricus terrestris*

2.4. AFLP

2.4.1. Production of AFLP fingerprints

Restriction of DNA and ligation of AFLP adapters were done simultaneously in one single reaction. Restriction was performed with the restriction endonucleases *Mse*I and *Eco*RI. To four µL of isolated DNA six µL of an adapter mix and two µL of an enzyme mix, containing both restriction enzymes and T4 DNA ligase, were added (table 2.2). The structure of the *Mse*I-adapter (metabion GmbH) was:

```
5'-GACGATGAGTCCTGAG
       TACTCAGGACTCAT-5'
```

and the structure of the *Eco*RI-adapter (metabion GmbH) was:

```
5'-CTCGTAGACTGCGTACC
      CATCTGACGCATGGTTAA-5'.
```

Incubation for restriction and ligation was done over night at room temperature. Afterwards water *ad* 50 µL was added to every sample.

Table 2.2.: Composition of adapter mix and enzyme mix for the restriction/ligation step of the AFLP procedure. Amounts for one sample. Water: Rotisolv® HPLC Gradient Grade, Carl Roth, Germany.

adapter mix		
	concentration	(µL)
T4 DNA ligase buffer	10 x	1
NaCl	0.5 mol/L	1
BSA	1 mg/mL	0.5
*Mse*I-adapter	5 pmol/µL	0.6
*Eco*RI-adapter	5 pmol/µL	0.6
water		2.3
enzyme mix		
	concentration	(µL)
T4 DNA ligase buffer	10 x	0.2
NaCl	0.5 mol/L	0.2
BSA	1 mg/mL	0.1
*Mse*I	10U/µL	0.08
*Eco*RI	10U/µL	0.4
T4 DNA ligase	1U/µL	0.19
water		0.83

Preselective amplification was performed using *Taq* polymerase (Qiagen), the preselective primers E01 and M03 and four µL of DNA from the restriction/ligation reaction (table 2.3). The sequence of the primer M03 was:

5'-GATGAGTCCTGAG TAA G-3'

and the sequence of primer E01 was:

5'-GACTGCGTACC AATTC A-3'.

The primers consisted of a core sequence, an enzyme specific sequence annealing with the adapters and the remains of the restriction sites and one selective nucleotide, i. e. G in M03 and A in E01 (Vos et al. 1995).

For the selective amplification the primers M72 and E39 were used. The sequence for the *Mse*I specific primer M72 was:

5'-GATGAGTCCTGAG TAA GGC-3'

and the sequence for the *Eco*RI specific primer was:

5'-GACTGCGTACC AATTC AGA-3'.

The general structure of the primers was similar to the preselective primers, but with three selective nucleotides, i. e. GGC in M72 and AGA in E39. Primer E39 was labelled with the fluorescent dye HEX^{TM} (Applied Biosystems) for automatic DNA fragment analysis (see below). Selective amplification was performed with four µL of the product of the preselective amplification diluted with water (1:20) and eleven µL of selective amplification mix (table 2.4). A touchdown PCR protocol was used (table 2.4).

Table 2.3.: Composition of preselective amplification mix per sample and PCR protocol. Water: Rotisolv® HPLC Gradient Grade, Carl Roth, Germany.

	concentration	(µL)
PCR buffer	10 x	2.5
dNTPs	10 mmol/L	0.375
Primer M03	5 pmol/µL	0.25
Primer E01	5 pmol/µL	0.25
Taq polymerase	5 U/µL	17.525
water		0.1
temperature (°C)	duration	
72	2 min	
94	10 s	
56	30 s	repeat 19 times
72	2 min	
60	3 min	
16	end	

The results of restriction/ligation and PCR reactions were checked by electrophoresis. Usually five µL of a sample were run on 1.5 % agarose gels with an appropriate size marker. Visualisation of DNA was performed with ethidium bromide and UV-light.

To prepare samples for capillary electrophoresis, the samples were diluted with water (1:5) and two µL of this dilution were mixed with HiDi Formamide and the ROX™-labelled size standard Genescan™ 500 ROX™ (Applied Biosystems) by two times vortexing followed by a short centrifugation. The size standard consisted of 16 labeled DNA fragments of a size ranging from 35 to 500 base pairs. Denaturation of DNA was done at 92 °C for two minutes followed by

Table 2.4.: Composition of selective amplification mix per sample and touchdown PCR protocol. Water: Rotisolv® HPLC Gradient Grade, Carl Roth, Germany.

	concentration	(µL)
PCR buffer	10 x	1.5
dNTPs	10 mmol/L	0.25
Primer M03	5 pmol/µL	0.25
Primer E01	5 pmol/µL	0.25
Taq polymerase	5 U/µL	0.068
water		8.77
temperature (°C)	duration	
94	2 min	
94	10 s	repeat with
65-57	30 s	annealing temp.
72	2 min	−1 °C per step
94	10 s	
56	30 s	repeat 19 times
72	2 min	
60	3 min	
16	end	

placing the samples on ice. Electrophoresis and analysis was done on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems), a fluorescence-based capillary electrophoresis analysis system. The software GENESCAN version 3.7 (Applied Biosystems) was used to process the electrophoresis data to electropherograms and for size calling of fragments. For size calling the 250 bp fragment of the size standard was not used.

2.4.2. Scoring of AFLP fingerprints

Scoring of AFLP fingerprints was based on repeatability of peaks in the fingerprints of 98 individuals used twice in DNA isolation and AFLP procedures. The AFLP fingerprints of these individuals were labelled automatically with the Software GENOTYPER version 3.7 (Applied Biosystems) using the function 'label peaks by size' to obtain a list of all peaks with their corresponding sizes in base pairs. This list was transferred to the software PEAKMATCHER which calculates peak categories with maximum repeatability across all samples (DeHaan et al. 2002). A peak category is defined as having a center at a given peak size and covering a given range of peak size. PEAKMATCHER was used several times with different settings regarding ranges of peak size and minimum peak repeatability. The obtained lists of categories were compared in terms of absolute number of categories and numbers of reproducible peaks per category. The settings finally used were: category ranges of $\pm$ 0.2, 0.3 and 0.4 base pairs, category increment of 0.1 base pairs and a minimum repeatability of 95 %. The categories obtained with these settings were checked visually in GENOTYPER 3.7. Categories including weak peaks, obviously different peaks and peaks very close to the borders of the categories were removed.

The remaining 125 categories, ranging from 38.4 ± 0.4 bp to 498.4 ± 0.2 bp, were used in GENTOYPER version 3.7 to automatically score all 394 samples (figure 2.2; for a complete list of categories, see appendix C), leading to a 0/1-matrix indicating absence and presence of peaks. Out of the 98 individuals which were used twice in AFLP fingerprinting only one sample per individual, i. e.

Figure 2.2.: Example section of AFLPs of three individuals from BIH (BOS04), GER-Kas (FHG19) and GER-Tim (SLH04) used twice in DNA isolation and fingerprinting for reproducible tests. The numbers on the x-axis are the lengths of fragments in bp. The ranges of six categories are shaded in grey. The category at 144.1±0.3 was not used for the final scoring because peaks were often located at the edges of the category. Screenshot from GENOTYPER 3.7

the sample with the higher intensity across all peaks, was used for scoring.

2.4.3. Analysis of AFLP data

Interpretation of AFLPs

AFLP fingerprints were interpreted in two ways. First, the obtained patterns of peaks were used directly for analysis as a phenotypic marker. For this purpose the band sharing indices Tanimoto (Jaccard 1908) and Simple matching (Sokal and Michener 1958) between individuals were calculated. Second, the allele frequencies underlying the AFLP fingerprints were estimated using different approaches (see below). Most of the downstream analyses of AFLPs were based on the band sharing indices (band-based methods) or the estimations of allele frequencies (frequency-based methods).

Calculation of band sharing indices The Tanimoto and Simple matching distances of all pairs of individuals were calculated. Mean (arithmetic) values of Tanimoto and Simple matching distances within and between sampling locations were also calculated. The Tanimoto distance between two individuals was calculated as

$$1 - \frac{a}{a + b + c} \tag{2.1}$$

and the Simple matching distance between two individuals was calculated as

$$1 - \frac{N - b - c}{N} \tag{2.2}$$

where a was the number of bands represented in both individuals, b and c were the numbers of bands represented in the one or the

other individual and N was the total number of band positions in the data set.

Estimation of allele frequencies at AFLPs Three different methods of allele frequency estimation were used:

The square-root method estimates the frequency of the null-allele at one sampling location by calculating the square-root of the Absence-phenotype's frequency (Bernstein 1930). The frequency of the Presence-allele was then calculated by one minus the frequency of the Absence-allele. The square-root method assumes Hardy-Weinberg proportions at the sampling locations.

The second method of estimating allele frequencies follows a Bayesian approach with a non-uniform prior distribution of allele frequencies (Zhivotovsky 1999). This method of allele frequency estimation was performed by using the software AFLP-SURV (Vekemans 2002; Vekemans et al. 2002). In the following this method is simply termed 'Bayesian method'. This method assumes Hardy-Weinberg proportions at the sampling locations, too.

The third approach uses also a Bayesian framework but is independent from assumptions concerning Hardy-Weinberg proportions; instead, it assumes similarity in F_{IS} and F_{ST} across loci, and frequency variation among sampling locations to follow a beta distribution (Holsinger 1999; Holsinger et al. 2002; Holsinger and Wallace 2004). The software HICKORY v1.1 was used for this estimation of allele frequencies (Holsinger and Lewis 2007). In the following the method is called 'HICKORY method'.

All calculations of genetic diversity or genetic structure demanding the estimation of allele frequencies were carried out by one or

more of these methods. See the description of the methods for details.

Genetic Diversity

Eight measures of genetic diversity at sampling locations were used. These measures belong to four different groups of methods. The first group uses band-based, the following groups use allele frequency-based approaches.

Band sharing indices The first group was represented by mean pairwise band sharing indices between individuals at sampling locations. Tanimoto and Simple matching were used as band sharing indices.

Nei's genetic diversity The second group consisted of Nei's genetic diversity (Nei 1973) calculated with allele frequencies estimated with the square-root method and the Bayesian method, respectively. Nei's genetic diversity gives the percentage of heterozygotes (expected heterozygosity) at a sampling location under Hardy-Weinberg proportions. Expected heterozygosity for one locus (h_e) was calculated following Nei's original approach as

$$h_e = 1 - \sum_{i=1}^{m} x_i^2 \tag{2.3}$$

with x_i denoting the estimated frequency of the ith allele at one locus and m denoting the number of alleles (m=2 in AFLPs). The multilocus expected heterozygosity H_e equals the arithmetic mean of all single locus heterozygosities. Under the assumption of Hardy-

Weinberg proportions, the theoretical maximum of H_e for a dominant marker with two alleles is 0.5 (25 % of individuals with Absence-phenotype).

Bayesian estimation of diversity — HICKORY The third group of genetic diversity measures consisted of Bayesian estimations of heterozygosity without assumptions about Hardy-Weinberg proportions based on the HICKORY method and named *hs* (Holsinger 1999; Holsinger et al. 2002; Holsinger and Wallace 2004). This method assumes that F_{IS} and F_{ST} are similar across all loci and that allele frequencies follow a beta distribution among populations. The software HICKORY v1.1 (Holsinger and Lewis 2007) was used to calculate *hs*. The software provides different models for estimating genetic diversity and Deviance Information Criterion (DIC) statistics to aid in choosing between these models (Spiegelhalter et al. 2002; Holsinger and Lewis 2007). Following the recommendations of interpreting the DIC statistics in the HICKORY manual (Holsinger and Lewis 2007) the full model and the f=0 model were chosen to calculate *hs*. The full model estimated f, an Bayesian analogue for F_{IS}, while the f=0 model assumed that f was zero, i. e. that there was no inbreeding in the population.

Genetic diversity after Hill (1973) The diversity measure of the fourth group was calculated as

$$v_2 = \left(\frac{1}{L} \times \sum_{l=1}^{L} \frac{1}{v_2(l)} \right)^{-1} \tag{2.4}$$

with L denoting a set of loci and $v_2(l)$ the single locus diversity at the lth locus (Hill 1973; Gregorius 1978). Single locus diversities were calculated as

$$v_2(p) = \left(\sum_{i=1}^{n} p_i^2 \right)^{-1} = v_2(l) \tag{2.5}$$

where p_i is the relative frequency of the ith allele and n the number of alleles (Gregorius 1978; Routledge 1979). This measure of genetic diversity can be interpreted as the inverse of the probability of drawing the same allele twice from individuals at one sampling location. The diversity over all loci equals the harmonic mean of all single locus diversities $v_2(p)$. This measure ranges theoretically from 1 to 2 and reaches its maximum with 25 % of the individuals showing the ABSENCE-phenotype. It was calculated using allele frequencies estimated with both the square-root method and the Bayesian method.

The spearman rank correlation coefficient was calculated to investigate the relationship between pairs of diversity measures (Spearman 1904).

Genetic Differentiation

Correlation between genetic and geographic distance The association between genetic and geographic distances was investigated using two scenarios. The first one predicted smaller genetic distances within sampling locations than between sampling locations, the second one predicted a linear relationship between genetic and geographic distances. The first scenario was investigated with Mantel tests, the second one with a different variant of Mantel tests and

with distograms (Mantel 1967). Both methods were based on inter-individual genetic distances, i. e. Tanimoto and Simple matching distances.

In order to test if genetic distances within sampling locations were smaller than between sampling locations, the correlation coefficient r between two matrices was calculated via Mantel tests (Mantel 1967). The matrices were an genetic distance matrix combined of inter-individual distances (Tanimoto and Simple matching) within and between sampling locations and a matrix indicating if a given value of the first matrix belongs to either the former or latter group. The null hypothesis that there were no differences in genetic distances within and between sampling locations was tested by a permutation test with 8000 permutations. The Mantel test and the permutations were performed using the software MANTELSTRUCT (Miller 1999).

In order to test if there was a linear relationship between genetic and geographic distances, Mantel tests and distograms were applied for two geographic scales. On the larger scale all 14 European sampling locations and on the smaller scale only sampling locations in Germany were used to compare genetic and geographic distances. Pairwise geographical distances between individuals were calculated using euclidean distances. The correlation coefficients r(G,S) between the genetic distance (Tanimoto and Simple matching) matrices (G) and the geographic distance matrix (S) were calculated with Mantel tests. The null hypothesis, i. e. no relationships between geographic and genetic distances, was tested with permutation tests. Therefore genetic data were permuted 8000 times over the locations, followed by a recalculation of r(G,S) for each permutation. Observed values of r(G,S) were then compared with an 95 % inter-

val of all permutated values of r(G,S) and if the observed value lay outside of this interval, the null hypothesis was rejected and the correlation coefficient r(G,S) was regarded as significant.

For the distograms all possible pairs of individuals were placed in 15 distance classes with a width of 150 km each for the larger geographic scale and in six distance classes with a width of 75 km for the smaller scale. Mean genetic distances in distance classes were calculated. To test the null hypothesis of no correlation between distance class and mean genetic distance, the 8000 permutations from the Mantel test (see above) were used to recalculate mean genetic distances for each distance class. The observed genetic distances and a 95 % interval of the permutated mean genetic distances of the distance classes were then depicted in the distograms. Observed values lying outside of the 95 % interval of permutations led to a rejection of the null hypothesis and the mean genetic distance of the particular distance class was considered as being significant. All calculations of r(G,S) and for the construction of distograms were done with the software DISTCLAS (provided by E. M. Gillet, Department of Forest Genetics and Forest Tree Breeding, University of Goettingen).

Genetic distances between sampling locations In order to measure the genetic differences between pairs of sampling locations, mean Tanimoto and Simple matching distances, D_S, and d_0 were calculated.

Nei's standard genetic distance D_S for a single locus was calculated as

$$D_S = -ln\left(\frac{\sum_{i=1}^{m} x_i y_i}{\sqrt{\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i^2}}\right) \tag{2.6}$$

where x and y are the frequencies of the ith of m alleles in the two populations. D_S for multiple loci was calculated by using the arithmetic means of $\sum_{i=1}^{m} x_i y_i$, $\sum_{i=1}^{m} x_i^2$ and $\sum_{i=1}^{m} y_i^2$ (Nei 1972). Allele frequencies were calculated by the square root method and the Bayesian method. Theoretically D_S ranges from 0 (equal frequencies of all alleles in the two populations) to ∞ (no common alleles).

Gregorius' d_0 (Gregorius 1984) for a single locus was calculated as

$$d_0 = \frac{1}{2} \times \sum_{i=1}^{m} |x_i - y_i| \tag{2.7}$$

where x and y are the frequencies of the ith of m alleles in the populations compared. The multilocus d_0 is calculated as the arithmetic mean of single locus d_0s. Allele frequencies were calculated by the square root method and the Bayesian method. The values of d_0 range from 0 (equal frequencies of all alleles in the two populations) to 1 (no common alleles).

The pairwise distance matrices were visualised via the four clustering methods Single linkage, Complete linkage, UPGMA (Unweighted Pair Group Method with Arithmetic mean) and WPGMA (Weighted Pair Group Method with Arithmetic mean). All clustering method were performed with the software STATISTICA 7.1 (StatSoft, Tulsa, USA) The cophenetic correlation coefficients of all combinations of distance metrics and clustering methods were calculated to test the goodness of fit for all dendrograms with the respective underlying distance matrix. For this the module 'COPH' implemented in NTSYSpc version 2.2 (Rohlf 2007) was used to calculate cophenetic matrices for every dendrogram. In the next step

the module 'MXCOMP' was used to compare these cophenetic matrices with the corresponding distance matrix.

Genetic Boundaries Genetic boundaries, defined as areas with sharp genetic change (Crida and Manel 2007), were assessed with two methods: Monmonier's maximum difference algorithm implemented in the software BARRIER version 2.2 and the Wombling algorithm implemented in the R package WOMBSOFT (Womble 1951; Monmonier 1973; Manni et al. 2004; Guérard and Manni 2004; Crida and Manel 2007).

The Monmonier approach starts by drawing a map including all sampling locations at their respective X and Y coordinates. A Voronoï tessellation is drawn on this map, which is, easily spoken, a pattern of borderlines enclosing every sampling location. Afterwards a Delaunay triangulation is produced, which appears as a set of triangles between adjacent sampling locations on the map. The edges of the triangulation are linked to a corresponding matrix consisting of pairwise genetic distances between sampling locations.

Four distance measures were used: Tanimoto, Simple matching, d_0 and D_S. The allele frequency-based distances d_0 and D_S were calculated with frequencies estimated with the Bayesian approach.

The algorithm then searches the edge of the triangulation with the highest distance and starts to draw a boundary. The boundary is continued on the adjacent edge with the next highest distance and so on. A boundary is stopped when it meets itself, another boundary or the borders of the whole map. The boundaries are always crossing edges of the triangulation and move along the Voronoï tessellation. In this way boundaries are always located equidistantly between the two sampling locations connected by the particular

edge. The location of a boundary is therefore always defined relatively ('between sampling location A an sampling location B') but not with exact geographic coordinates.

To set limits to the Voronoï tessellation, i. e. the whole map, six so called virtual points were placed in regions where no sampling had taken place surrounding the sampled area. By doing so, outer edges of the Delaunay triangulation connecting sampling locations over a great distance had to be removed. Namely, the edges connecting Bosnia-Herzegovina and Finland, Bosnia-Herzegovina and France and the edge connecting France and Sweden were excluded from the analysis.

The Wombling algorithm computed a function (systemic function) that was expressed as a map. On this map potential genetic barriers could be identified as regions with high values of the systemic function. Afterwards the potential barriers were confirmed using a binomial test. Wombling algorithm and binomial test where performed with the software WOMBSOFT (Crida and Manel 2007) with a resolution of the map of 100 to 100 units, a bandwidth of h=130 km, a distance from the border of m=100 km and for the binomial test a p-value of 0.001 and a percentile of the systemic function of 0.3. The coordinates of individuals were slightly (<1 mm) changed because the software did not work with individuals located exactly at the same coordinates.

Overall differentiation Treating groups of individuals at sampling locations as subpopulations of a total population, the differentiation among these subpopulations, i. e. the relative amount of genetic diversity in subpopulations compared to the total genetic diversity,

was estimated by calculating F_{ST}, a Bayesian estimator of F_{ST} and δ.

Single locus F_{ST} was calculated as

$$F_{ST} = \frac{H_T - H_S}{H_T} \tag{2.8}$$

with H_T giving the genetic diversity (expected heterozygosity) of the total population and H_S giving the genetic diversity in all subpopulations weighted by their relative sample size:

$$H_T = \left(1 - \sum_{i=1}^{n} p_i^2\right) \tag{2.9}$$

with p_i as the relative frequency of the $i-th$ of n alleles in the total population,

$$H_S = \sum_{j=1}^{m} c_j \times \left(1 - \sum_{i=1}^{n} p_i^2(j)\right) \tag{2.10}$$

with $p_i(j)$ as the relative frequency of the $i-th$ of n alleles in the $j-th$ of m subpopulations and c_j as the relative sample size of subpopulation j. F_{ST} for multiple loci was calculated with arithmetic means of single locus H_T and H_S. Allele frequencies were estimated by using the square-root method.

Additionally a bayesian estimator of F_{ST}, called $\theta^{(II)}$, was calculated using two models (full model and f=0 model) of the software HICKORY v1.1 (Holsinger and Lewis 2007).

F_{ST} and $\theta^{(II)}$ theoretically range from 0 to 1. F_{ST} or $\theta^{(II)}$=0 indicate no genetic differentiation between subpopulations, while F_{ST} or $\theta^{(II)}$=1 indicate no diversity within any subpopulation and

a fixation on different alleles in at least two subpopulations but not necessarily a complete differentiation between sampling locations.

δ was calculated as

$$\delta = \sum_{k=1}^{K} c_k D_k \tag{2.11}$$

with c_k denoting the investigated number of individuals of sampling location k, and D_k denoting the genetic distance between sample location k and its complement, i.e. all other sampling locations pooled. D_k was calculated as

$$D_k = \frac{1}{2} \times \sum_{i=1}^{m} |p_i(k) - \overline{p}_i(k)| \tag{2.12}$$

with $p_i(k)$ as the relative frequency of the ith of m alleles within the kth sampling location and $\overline{p}_i(k)$ as the relative frequency of that allele in the remaining pooled sampling locations (Gregorius and Roberds 1986). For $\delta = 0$ all subpopulations were identical in terms of allele frequencies.

Additionally, D_k was used as measure of how good subpopulation k represented the whole population with low values indicating a good representation and high values indicating a high differentiation and low representativeness. D_k and δ were calculated using both the square-root and the Bayesian method's estimates of allele frequencies.

Testing arrangements of regions with AMOVA The AMOVA procedure was used to calculate variance components and estimate differentiation on different hierarchical population levels (Excoffier et al. 1992). Variance components and differentiation were inves-

tigated within populations, among populations belonging to one region and among regions. The corresponding differentiation measures were called F_{ST}, F_{SC} and F_{CT}, respectively (originally described as Φ_{ST}, Φ_{SC} and Φ_{CT} in (Excoffier et al. 1992)).

For AMOVA the AFLP fingerprints of individuals were treated as haplotypes, i. e. fingerprinting phenotypes. The genetic distance matrix with all pairs of haplotypes was constructed using Euclidean squared distances.

Statistical significance of differentiation measures and the corresponding variance components were tested with permutation tests using 10,000 permutations with random allocations of individuals to populations regardless of regions (for F_{ST}), individuals within groups (for F_{SC}) and whole populations across groups (for F_{CT}). All calculations were performed with the software Arlequin 3.11 (Excoffier et al. 2005).

Different hypothetical sets of regions containing the sampling locations were tested. Two sets of regions (set 1 and set 2) were artificial, assuming no groupings (set 1) or a grouping which was based on the countries were *L. terrestris* had been sampled (set 2). The other sets were based on the results of the Monmonier algorithm (sets 3, 4, 5, 6a, 6b), the Wombling approach (set 7) and the genetic distance UPGMA distograms (set 6b, see results of these methods for details). This way, the regional resolution of the patterns of genetic boundaries and similarities between sampling locations could be tested. The sets were:

- Set 1. This set grouped all populations in one single region. F_{SC} and F_{CT} were not calculated.

- Set 2. This set contained regions representing the countries where *L. terrestris* had been sampled.
  - region 1: France
  - region 2: Germany
  - region 3: Sweden
  - region 4: Finland
  - region 5: Austria
  - region 6: Bosnia
- Set 3. This set contained regions which were based on the results of the Monmonier algorithm based on a Tanimoto distance matrix.
  - region 1: FRA-Bru, FRA-Béd
  - region 2: GER-Pad
  - region 3: GER-EgP, EgF, Sol, Kas
  - region 4: GER-Lei
  - region 5: GER-Bay
  - region 6: GER-Tim, SWE, FIN, AUT, BIH
- Set 4. This set contained regions which were based on the results of the Monmonier algorithm based on a Simple matching distance matrix.
  - region 1: FRA-Bru, FRA-Béd
  - region 2: GER-Pad
  - region 3: GER-EgP, EgF, Sol, Kas
  - region 4: GER-Tim

  - region 5: GER-Bay
  - region 6: GER-Lei, SWE, FIN, AUT, BIH

- Set 5. This set contained regions which were based on the results of the Monmonier algorithm based on a D_S distance matrix.
  - region 1: FRA-Béd
  - region 2: FRA-Bru
  - region 3: GER-Pad, EgP, EgF, Sol, Kas
  - region 4: GER-Tim
  - region 5: SWE, FIN
  - region 6: GER-Bay, Lei, AUT, BIH

- Set 6a. This set contained regions which were based on the results of the Monmonier algorithm based on a d_0 distance matrix.
  - region 1: FRA-Bru, FRA-Béd
  - region 2: GER-Lei
  - region 3: GER-Pad, EgP, EgF, Sol, Kas
  - region 4: GER-Tim
  - region 5: SWE, FIN
  - region 6: GER-Bay, AUT, BIH

- Set 6b. This set contained regions which were based on the results of the Monmonier algorithm based on a d_0 distance matrix like in set 6a but puts GER-Bay, AUT, BIH, FIN, and SWE together in one region based on the d_0 distance UPGMA

dendrogram in which a cluster consisting of BIH, SWE and FIN was found.

  - region 1: FRA-Bru, FRA-Béd
  - region 2: GER-Lei
  - region 3: GER-Pad, EgP, EgF, Sol, Kas
  - region 4: GER-Tim
  - region 5: GER-Bay, AUT, BIH, FIN, SWE

- Set 7. This set contained regions which were based on the results of the Wombling approach.
  - region 1: FRA-Bru, FRA-Béd
  - region 2: GER-Pad, EgF, EgP, Sol, Kas, Lei, Tim, SWE, FIN
  - region 3: GER-Bay
  - region 4: AUT, BIH

3. Results

3.1. General AFLP results

All of the investigated 125 AFLP loci were polymorphic across the 394 individuals of *Lumbricus terrestris*. At the 15 sampling locations the number of polymorphic loci ranged from 47 (37.6 %) at SWE to 89 (71.2 %) at GER-Kas. The majority of monomorphic loci were monomorphic for the ABSENCE-phenotype. At the sampling location in Sweden and at one sampling location in France (FRA-Bru) monomorphism exceeded polymorphism (table 3.1). Every individual showed an unique AFLP banding pattern.

3.2. Genetic Diversity

The range of genetic diversity and the ranking of sampling locations depended on the measure used to assess genetic diversity (table 3.2).

The mean Tanimoto distances between individuals belonging to the same sampling location ranged from 0.422 (FRA-Bru) to 0.663 (GER-Bay), whereas mean Simple Matching distances were much lower and ranged from 0.102 (FRA-Bru) to 0.197 (GER-Bay).

Nei's genetic diversity H_e ranged from 0.081 (SWE) to 0.146 (GER-Bay) when based on square root estimations of allele frequencies and from 0.097 (SWE) to 0.155 (GER-Bay) when allele

Table 3.1.: Proportion of polymorphic and monomorphic loci at sampling locations

Sampling location	polymorphic	monomorphic	
		Absence-phenotype	Presence-phenotype
FIN	0.520	0.440	0.040
SWE	0.376	0.568	0.056
GER-Tim	0.656	0.328	0.016
GER-Sol	0.592	0.392	0.016
GER-EgF	0.600	0.376	0.024
GER-EgP	0.632	0.344	0.024
GER-Pad	0.696	0.264	0.040
GER-Kas	0.712	0.288	0.000
GER-Lei	0.592	0.392	0.016
GER-Bay	0.672	0.320	0.008
FRA-Bru	0.416	0.544	0.040
FRA-Béd	0.528	0.416	0.056
AUT	0.680	0.312	0.008
BIH	0.568	0.392	0.040
CAN	0.624	0.336	0.040

Table 3.2.: Four groups of genetic diversity measures with standard deviations. 1: mean interindividual distances calculated with band sharing indices; 2: expected heterozygosity (H_e); 3: Bayesian estimation of heterozygosity (hs) following two different HICKORY models; 4: diversity measure v_2. Allele frequencies for groups 2 and 4 were estimated with the square root method and the Bayesian method.

	1		**2**		**3**		**4**	
location	Tanimoto	Simple Matching	H_e square root	H_e Bayesian	hs (full model)	hs (f=0 model)	v_2 square root	v_2 Bayesian
FIN	0.551 ± 0.07	0.145 ± 0.03	0.112 ± 0.16	0.127 ± 0.16	0.124 ± 0.01	0.123 ± 0.00	1.126 ± 0.28	1.146 ± 0.28
SWE	0.490 ± 0.08	0.108 ± 0.03	0.081 ± 0.14	0.097 ± 0.14	0.102 ± 0.01	0.102 ± 0.00	1.088 ± 0.23	1.107 ± 0.23
GER-Tim	0.611 ± 0.15	0.182 ± 0.06	0.144 ± 0.17	0.152 ± 0.17	0.142 ± 0.01	0.139 ± 0.00	1.168 ± 0.31	1.179 ± 0.31
GER-Sol	0.633 ± 0.14	0.180 ± 0.04	0.138 ± 0.17	0.150 ± 0.16	0.148 ± 0.01	0.145 ± 0.00	1.160 ± 0.30	1.176 ± 0.30
GER-EgF	0.584 ± 0.08	0.163 ± 0.03	0.126 ± 0.15	0.137 ± 0.15	0.132 ± 0.01	0.131 ± 0.00	1.144 ± 0.27	1.159 ± 0.27
GER-EgP	0.604 ± 0.08	0.171 ± 0.03	0.129 ± 0.15	0.141 ± 0.15	0.135 ± 0.01	0.133 ± 0.00	1.148 ± 0.27	1.164 ± 0.27
GER-Pad	0.645 ± 0.07	0.194 ± 0.03	0.133 ± 0.15	0.152 ± 0.14	0.143 ± 0.01	0.140 ± 0.00	1.153 ± 0.26	1.179 ± 0.25
GER-Kas	0.596 ± 0.10	0.165 ± 0.04	0.133 ± 0.16	0.138 ± 0.16	0.136 ± 0.01	0.134 ± 0.00	1.154 ± 0.28	1.160 ± 0.28
GER-Lei	0.577 ± 0.10	0.160 ± 0.04	0.127 ± 0.16	0.138 ± 0.16	0.133 ± 0.01	0.132 ± 0.00	1.146 ± 0.28	1.160 ± 0.28
GER-Bay	0.663 ± 0.09	0.197 ± 0.03	0.146 ± 0.15	0.155 ± 0.15	0.148 ± 0.01	0.145 ± 0.00	1.171 ± 0.27	1.184 ± 0.28
FRA-Bru	0.422 ± 0.07	0.102 ± 0.02	0.095 ± 0.16	0.104 ± 0.15	0.112 ± 0.00	0.112 ± 0.00	1.105 ± 0.28	1.116 ± 0.28
FRA-Béd	0.568 ± 0.07	0.142 ± 0.04	0.097 ± 0.14	0.115 ± 0.14	0.114 ± 0.01	0.113 ± 0.00	1.108 ± 0.23	1.130 ± 0.23
AUT	0.611 ± 0.10	0.165 ± 0.03	0.129 ± 0.15	0.136 ± 0.15	0.132 ± 0.01	0.130 ± 0.00	1.148 ± 0.27	1.158 ± 0.27
BIH	0.598 ± 0.07	0.157 ± 0.03	0.110 ± 0.14	0.124 ± 0.14	0.122 ± 0.01	0.120 ± 0.00	1.123 ± 0.23	1.141 ± 0.23
CAN	0.617 ± 0.08	0.163 ± 0.03	0.110 ± 0.13	0.126 ± 0.13	0.123 ± 0.01	0.120 ± 0.00	1.124 ± 0.21	1.144 ± 0.22

frequency estimation was performed with the Bayesian method. A high Spearman coefficient of 0.96 indicated that different approaches of allele frequency estimation had no great effect on the calculation of H_e (table 3.3).

The Bayesian estimations of heterozygosity (hs) calculated with HICKORY were similar to results for H_e, however, showing a more narrow range. They covered a range from 0.102 (SWE) to 0.145 and 0.148 with the f=0 model and the full model, respectively (GER-Sol and GER-Bay). The values obtained with the two models were only slightly different in some cases and these differences had no effect in ranking of sampling locations (Spearman coefficient of 1).

The third allele frequency-based measure, v_2, ranged from 1.088 (SWE) to 1.171 (GER-Bay) for square root estimates and from 1.107 (SWE) to 1.184 (GER-Bay) for Bayesian estimation of allele frequencies. Differences between v_2 values obtained with the two different allele estimation methods were small (Spearman coefficient of 0.96).

Comparisons between the four groups of diversity measures showed that Simple matching distances were more similar to allele frequency-based methods than Tanimoto distances (mean Spearman coefficients of 0.93 and 0.75, respectively). There were only small differences between hs and the other two allele frequency-based methods H_e and v_2 (mean Spearman coefficient of 0.97).

All allele frequency-based methods ranked the sampling locations in Sweden and France as the sampling locations with the lowest genetic diversity, followed by a group with medium genetic diversity consisting of Finland, Bosnia and Canada and finally a group of all German sampling locations and Austria as a group with the highest genetic diversity. Band-based methods placed one sampling location

Table 3.3.: Pairwise Spearman index of genetic diversity measures.

	Tanimoto	Simple Matching	H_e square root	H_e Bayesian	hs (full model)	hs (f=0 model)	v_2 square root	v_2 Bayesian
Tanimoto	—							
Simple Matching	0.8929	—						
H_e square root	0.7464	0.9321	—					
H_e Bayesian	0.7357	0.9321	0.9571	—				
hs(full model)	0.7571	0.9286	0.9714	0.9750	—			
hs(f=0 model)	0.7571	0.9286	0.9714	0.9750	1.0000	—		
v_2 square root	0.7464	0.9321	1.0000	0.9571	0.9714	0.9714	—	
v_2 Bayesian	0.7357	0.9321	0.9571	1.0000	0.9750	0.9750	0.9571	—

in France (FRA-Bru) and the sampling location in Sweden in a group with lowest genetic diversity and all other sampling locations in one group covering medium and highest genetic diversity.

3.3. Genetic differentiation

3.3.1. Comparison between interindividual distances within and between sampling locations

Interindividual genetic distances were lower within than between sampling locations (tables A.1 and A.2 in appendix). The Mantel tests, based on Tanimoto and Simple matching distances, led to correlation coefficients of r=0.289 and r=0.23, respectively. The null hypothesis that there were no differences between interindividual genetic distances within and between sampling locations was rejected for both correlation coefficients ($P<0.001$), based on 8000 permutations. Even earthworms from Canada, though possibly not originating from one single sampling location, possessed a lower interindividual genetic distance among each other than to earthworms from other locations.

3.3.2. Comparison between genetic and geographic distances — Mantel test

On the larger geographical scale (Europe) the Tanimoto distance matrix of individuals was positively correlated with the geographic distance Matrix (r(G,S)=0.132, upper limit of 8000 permutations: 0.053), in contrast to the genetic distance matrix based on Simple matching, which showed no correlation with the matrix of geo-

graphic distances (r(G,S)=0.014, upper limit of 8000 permutations: 0.061).

On the smaller scale, where only sampling locations in Germany were considered, both genetic distances (Tanimoto and Simple matching) were positively correlated with the geographic distance matrix (Tanimoto: r(G,S)=0.315, upper limit of 8000 permutations: 0.075; Simple Matching: r(G,S)=0.329, upper limit of 8000 permutations: 0.071).

3.3.3. Comparison between genetic and geographic distances — Distograms

Genetic distance (Tanimoto and Simple matching) between German *L. terrestris* increased with increasing geographic distance. In a distance less than 75 km the observed mean genetic distances were significantly lower than the genetic distances from the permutation test and became significantly higher at a geographic distance greater than 150 km (Tanimoto distance) and 225 km (Simple matching distance, figure 3.1), leading to the rejection of the hypothesis, i. e. no effect of geographic distance on genetic distance, for these distance classes. The larger geographic scale (Europe) did not show such a simple pattern. Significantly low and high genetic distances compared to the mean distances from the permutation tests occurred without a clear relation to distance class (figure 3.2).

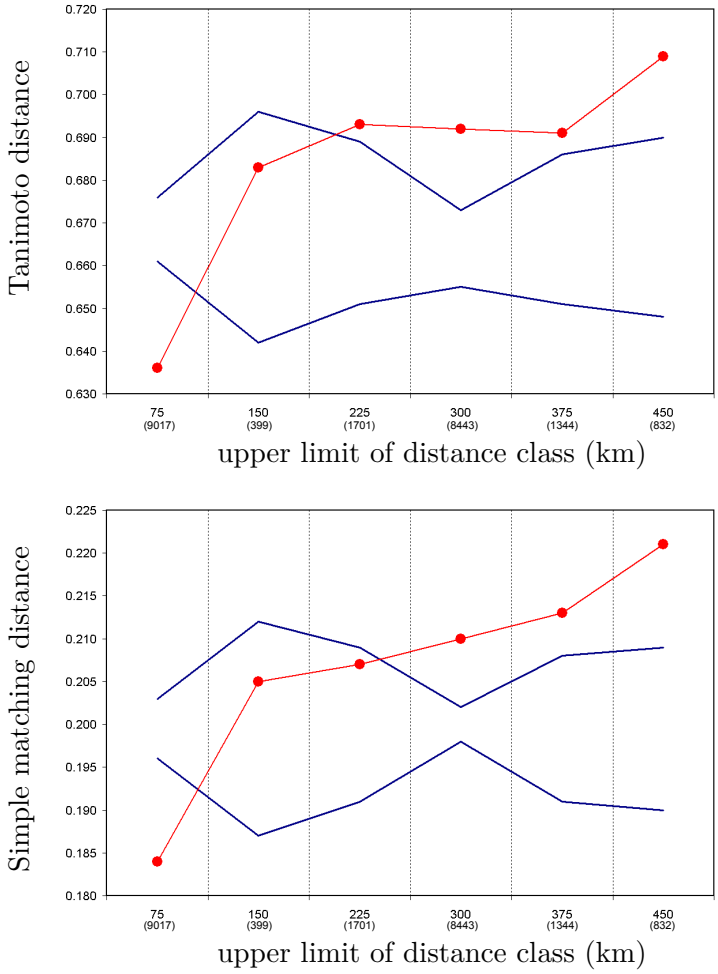

Figure 3.1.: Distogram of mean interindividual Tanimoto (above) and Simple matching (below) distances of *L. terrestris* from eight German locations. Red dots: observed mean distances, blue lines: 95 % interval of 8000 Permutations. Width of distance classes: 75 km. Numbers in parentheses give the numbers of pairs in the distance classes.

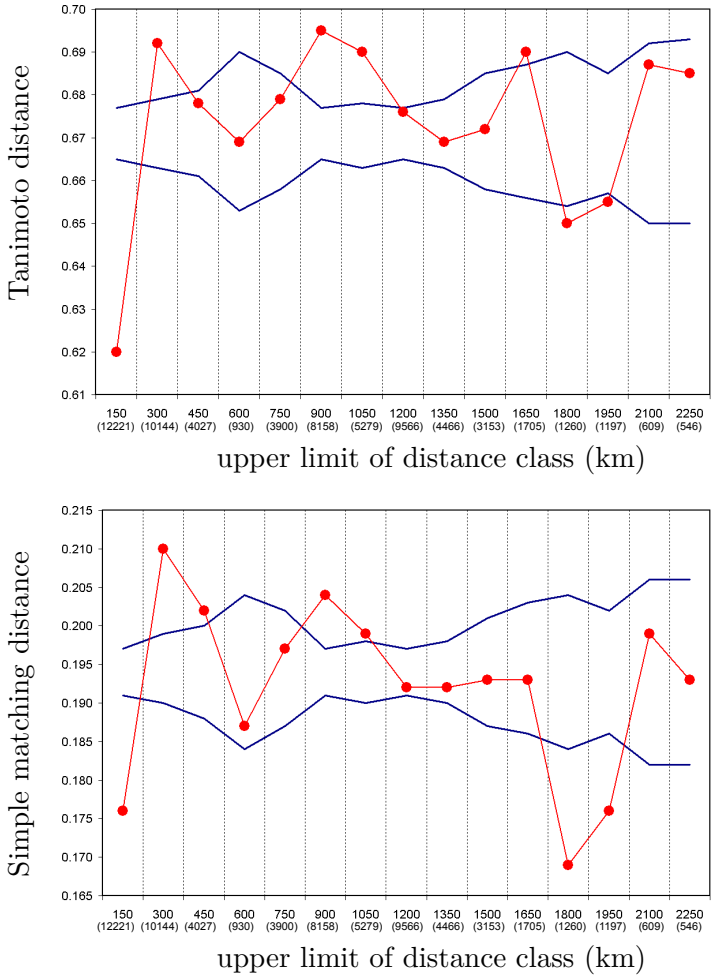

Figure 3.2.: Distogram of mean interindividual Tanimoto (above) and Simple matching (below) distances of *L. terrestris* from 14 locations in Europe. Red dots: observed mean distances, blue lines: 95 % interval of 8000 Permutations. Width of distance classes: 150 km. Numbers in parentheses give the numbers of pairs in the distance classes.

3.3.4. Pairwise distances between sampling locations

Simple matching distances between pairs of sampling locations were much lower than values from the other band-based distance method, the Tanimoto distance (mean distance: 0.196 and 0.676, tables A.2 and A.1 in appendix). Even Tanimoto distances had a wider range than Simple matching distances (from 0.584 to 0.725 and 0.148 to 0.225, respectively), both band-based methods showed only small differences between pairwise distances. Values for D_S ranged from 0.005 to 0.05 and differed only slightly between the two allele frequency estimation methods. The values for d_0 ranged from 0.04 to 0.13 for the square root estimation of allele frequencies and from 0.04 to 0.12 for the Bayesian estimation (tables A.3, A.4, A.5 and A.6 in appendix).

Out of four different clustering methods Single linkage had the lowest cophenetic correlations across all distance methods (mean cophenetic correlation coefficient r=0.73) followed by Complete linkage (mean r=0.77, table 3.4). However, both methods had a good correlation if cluster analysis was based on a D_S distance matrix, with a slightly higher correlation if allele frequencies had been estimated with the Bayesian method. UPGMA and WPGMA had good cophenetic correlations with allele frequency-based distance metrics but not with band-based methods (mean r=0.87 versus 0.78), with the exception of UPGMA, which showed a good correlation using the Tanimoto distance (r=0.80).

Because of the highest cophenetic correlation across all genetic distance measures, only the results for UPGMA are presented here and discussed below. However, for comparison purposes, the dendograms for Single linkage, Complete linkage and WPGMA are pre-

sented in the appendix. Likewise, the dendograms based on square root estimates of allele frequencies are presented in the appendix, too (figures B.1 to B.20 in appendix).

The UPGMA dendrograms of the 14 European sampling locations (Canada excluded) based on the band-based distance measures Tanimoto and Simple matching consisted mainly of two clusters. One cluster was comprised of the four German sampling locations EgP, EgF, Sol and Kas. The other cluster consisted of sampling locations in Austria, Bosnia, Finland, Sweden and GER-Lei. The sampling locations in France were also clustered (figures 3.3 and 3.4). The UPGMA dendogram based on D_S positioned Austria, GER-Bay and Bosnia-Herzegovina in one cluster and GER-EgP, GER-EgF, GER-Sol, GER-Pad and GER-Kas in another cluster. In contrast to band-based UPGMA dendograms, the two sampling locations from France did not form a cluster (figure 3.5). In the UPGMA dendrogram based on d_0 Bosnia-Herzegovina was placed in one cluster with Finland and Sweden and the two sampling locations in France were clustered (figure 3.6).

If the Canadian *L. terrestris* were considered in the cluster analysis, they were either placed together with the sampling location in Austria or France. In case of band-based distance measures the overall topology of the dendrograms was not strongly effected by the addition of Canadian earthworms, whereas in case of allele frequency-based measures the topology changed more markedly. Especially the position of FRA-Béd was strongly affected (figures 3.3 to figure 3.6).

Table 3.4.: Cophenetic correlation coefficients r of different methods of cluster analysis. Bold numbers indicate a good ($r \geq 0.8$) or very good ($r \geq 0.9$) fit of correlation.

	Single Linkage	Complete Linkage	UPGMA	WPGMA
Simple Matching	0.70	0.68	0.78	0.77
Tanimoto	0.76	0.68	**0.80**	0.79
D_S square root	**0.84**	**0.85**	**0.90**	**0.89**
D_S bayesian	**0.88**	**0.87**	**0.92**	**0.92**
d_0 square root	0.57	0.78	**0.81**	**0.80**
d_0 bayesian	0.64	0.78	**0.86**	**0.86**

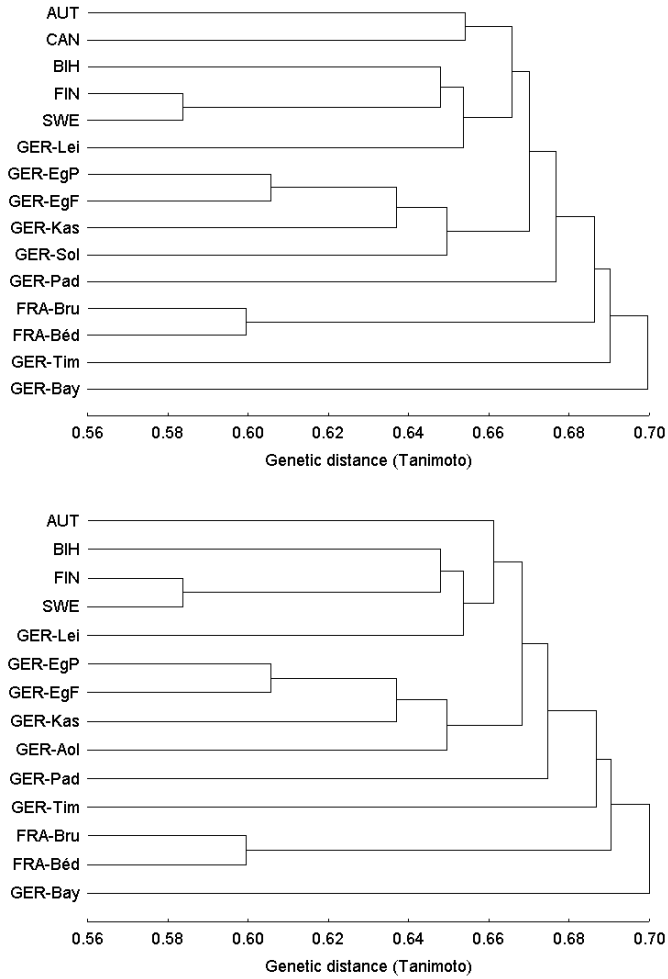

Figure 3.3.: UPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Tanimoto distance.

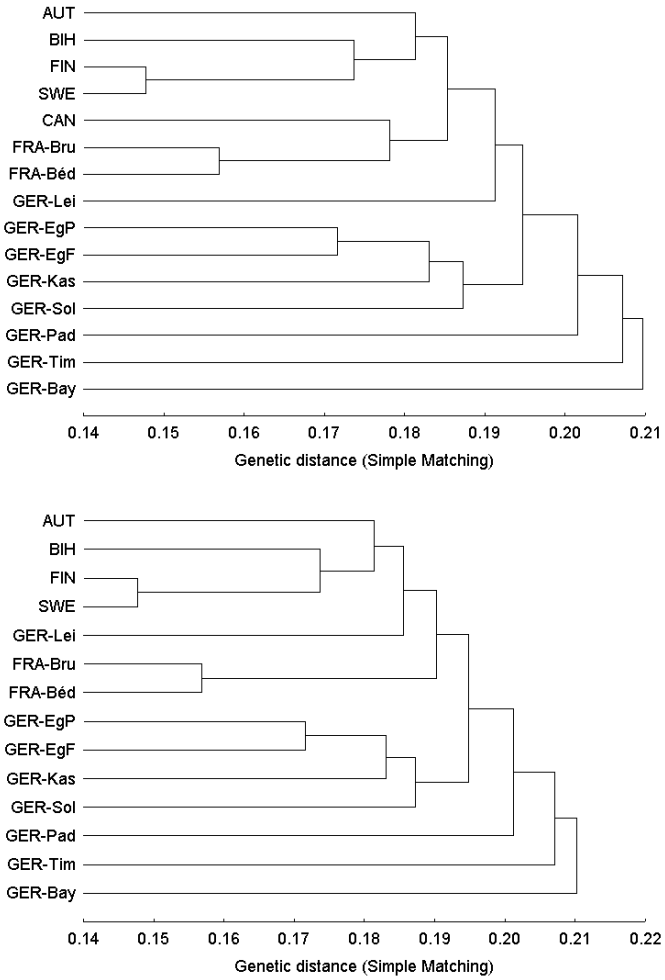

Figure 3.4.: UPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Simple Matching distance.

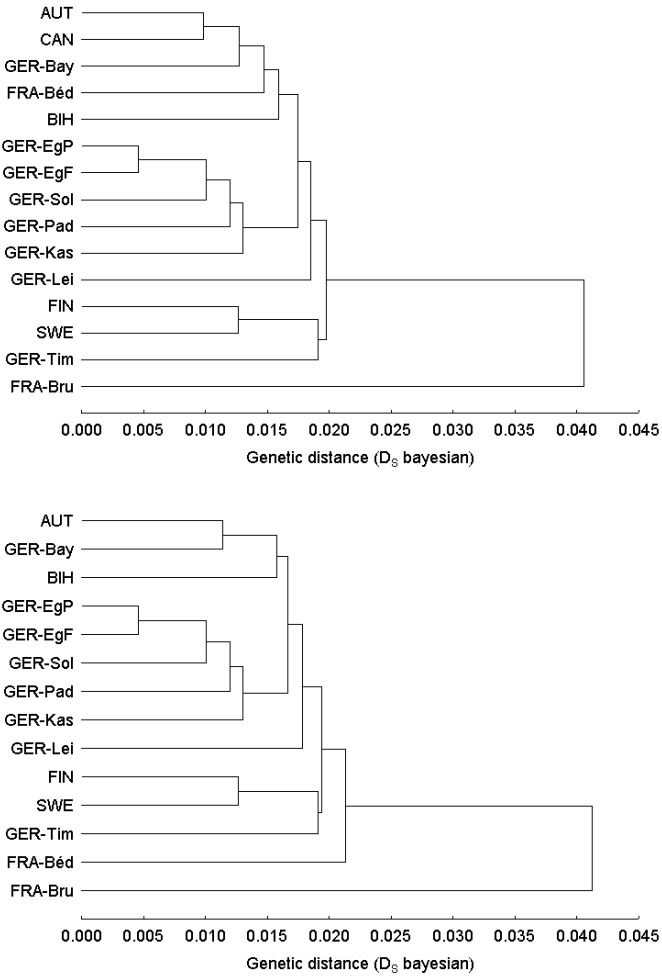

Figure 3.5.: UPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with the Bayesian method.

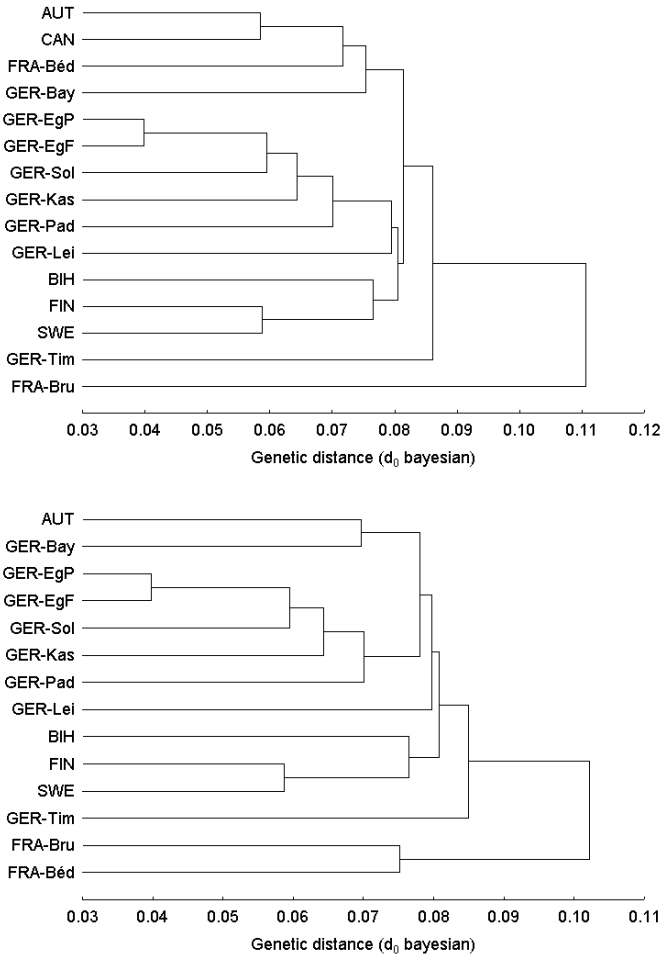

Figure 3.6.: UPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with the Bayesian method.

3.3.5. Genetic boundaries

Genetic boundaries calculated with the Monmonier algorithm based on a Tanimoto distance matrix of 14 European sampling locations divided the sampled area into six regions. France was separated from the rest of Europe. Austria and Bosnia-Herzegovina were placed in one region with a potential connection to Sweden and Finland through Eastern Europe (figure 3.7). The German locations GER-Pad, GER-Bay and GER-Lei were isolated and GER-Tim was placed in the same region as Sweden and Finland. On a smaller geographical scale a additional boundary isolated GER-Sol from the remaining German locations (figure 3.8). The map of genetic boundaries obtained with Simple matching distances showed similarities in the south-western part to the map based on Tanimoto distances. However, the Monmonier algorithm in this case isolated GER-Tim from the rest of the sampling locations and formed a greater eastern region ranging from Finland and Sweden to Bosnia-Herzegovina including the eastern most German sampling location GER-Lei (figure 3.7).

The genetic boundaries obtained by using the Monmonier algorithm based the on allele frequency-based distances D_S and d_0 disconnected Finland and Sweden completely from Austria and Bosnia. France was also separated from the rest of Europe and, in case of D_S, both French locations were separated from each other. In contrast to the band-based methods, GER-Bay was recognized as a separate region in Europe (figure 3.9). On the smaller geographical scale GER-Kas was isolated from the remaining German locations. For both, D_S and d_0, the boundaries at the smaller geographical scale were the same, but with a slightly different order in which the

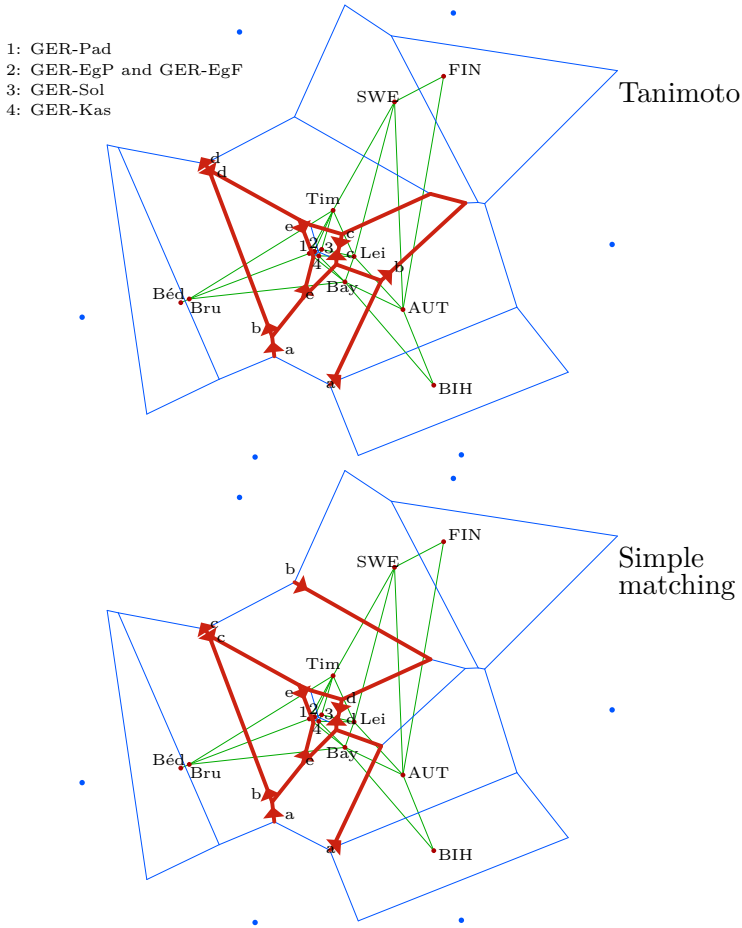

Figure 3.7.: Monmonier map of 14 *L. terrestris* sampling locations (red points). Distance measures: Tanimoto (above) and Simple matching (below). Red lines: Genetic barriers. Small letters: start and end points of barriers. Blue lines: Voronoï tessellation, green lines: Delaunay triangulation, blue points: virtual points. Note: GER-EgP and GER-EgF appear as one point because of close proximity of sampling locations but were analysed as two independent locations.

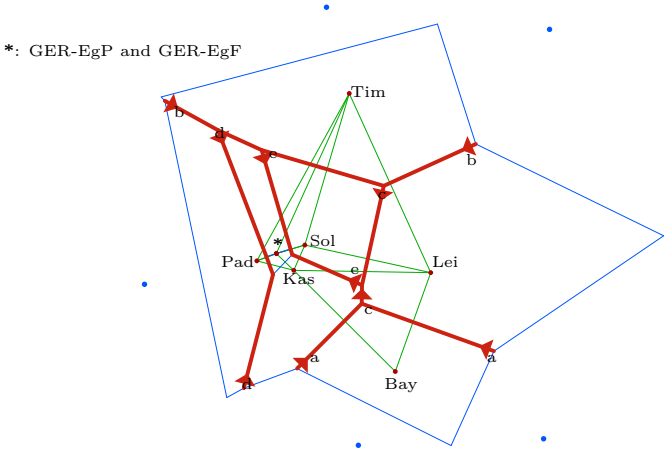

Figure 3.8.: Monmonier map of eight *L. terrestris* sampling locations in Germany (red points). Distance measures: Tanimoto and Simple matching, respectively. Red lines: Genetic barriers. Small letters: start and end points of barriers. Blue lines: Voronoï tessellation, green lines: Delaunay triangulation, blue points: virtual points. Note: GER-EgP and GER-EgF appear as one point because of close proximity of sampling locations but were analysed as two independent locations.

boundaries were drawn by the algorithm due to some differences in the distance matrices (figure 3.10).

The systemic maps obtained with the Wombling approach indicated potential genetic boundaries (figure 3.11). Regarding Europe, a big V-shaped boundary was confirmed by the binomial test. The boundary isolated the sampling locations in France and the sampling locations in Bosnia and Austria from all other sampling locations. On the smaller geographical scale the Wombling method divided Germany into a northern part, including GER-Tim, a southeastern part with GER-Bay and GER-Lei, and a western part including the sampling locations at the Egge-range, GER-Pad and GER-Kas (figure 3.12).

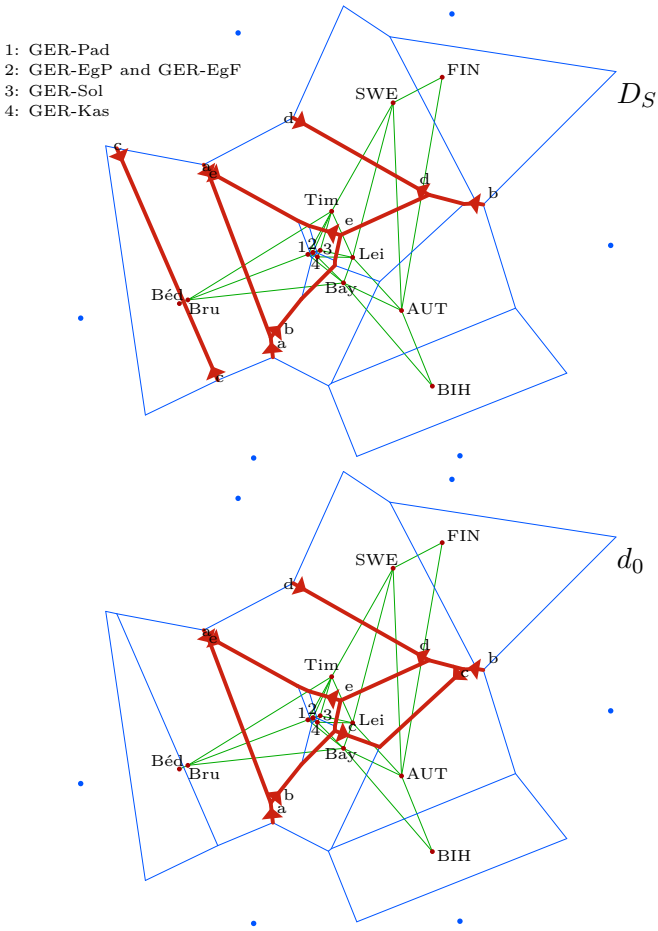

Figure 3.9.: Monmonier map of 14 *L. terrestris* sampling locations (red points). Distance measures: D_S (above) and d_0 (below). Red lines: Genetic barriers. Small letters: start and end points of barriers. Blue lines: Voronoï tessellation, green lines: Delaunay triangulation, blue points: virtual points. Note: GER-EgP and GER-EgF appear as one point because of close proximity of sampling locations but were analysed as two independent locations.

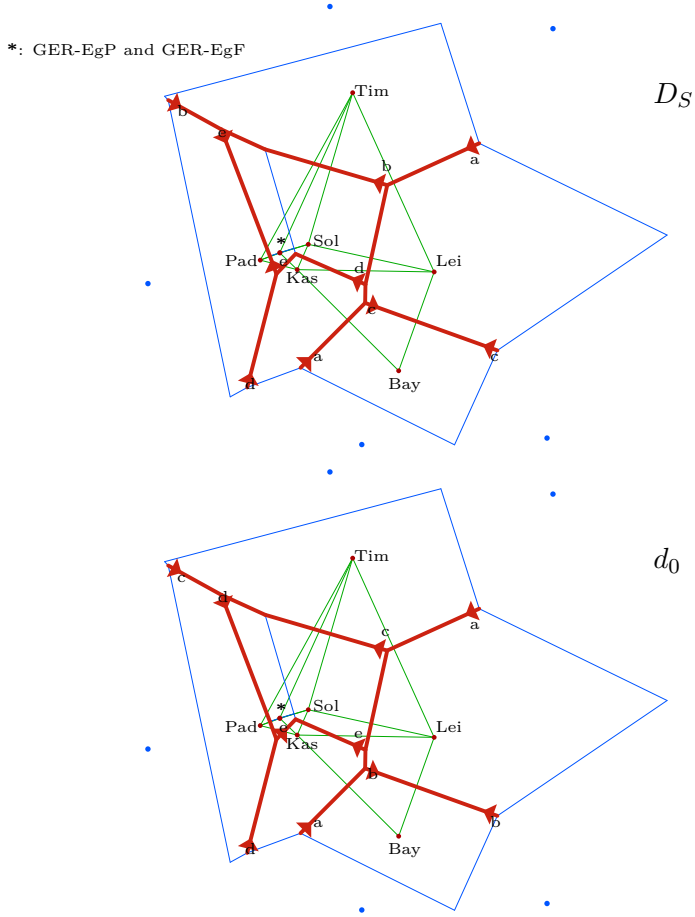

Figure 3.10.: Monmonier map of eight *L. terrestris* sampling locations in Germany (red points). Distance measures: D_S (above) and d_0 (below). Red lines: Genetic barriers. Small letters: start and end points of barriers. Blue lines: Voronoï tessellation, green lines: Delaunay triangulation, blue points: virtual points. Note: GER-EgP and GER-EgF appear as one point because of close proximity of sampling locations but were analysed as two independent locations.

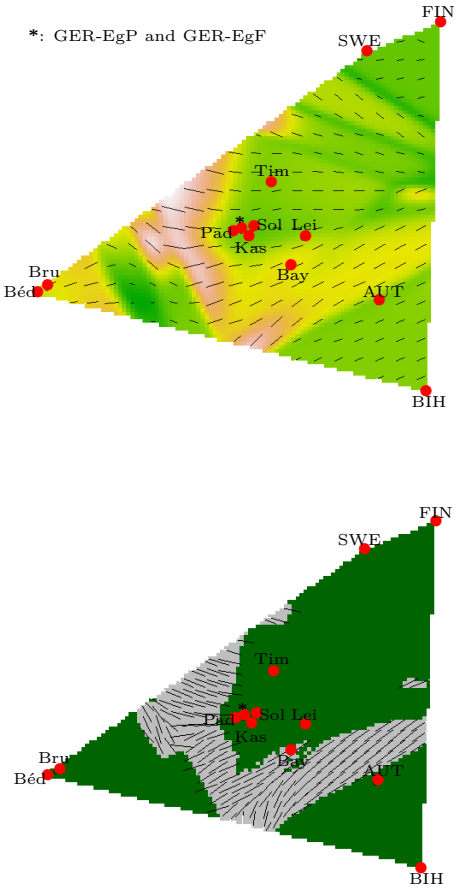

Figure 3.11.: Systemic function map (left) and genetic boundaries (right) of 14 European *L. terrestris* sampling locations (red points) obtained with the Wombling method. Yellow, pink and white: potential boundaries, grey: boundaries confirmed by binomial test. Note: GER-EgP and GER-EgF appear as one point because of close proximity of sampling locations but were analysed as two independent locations.

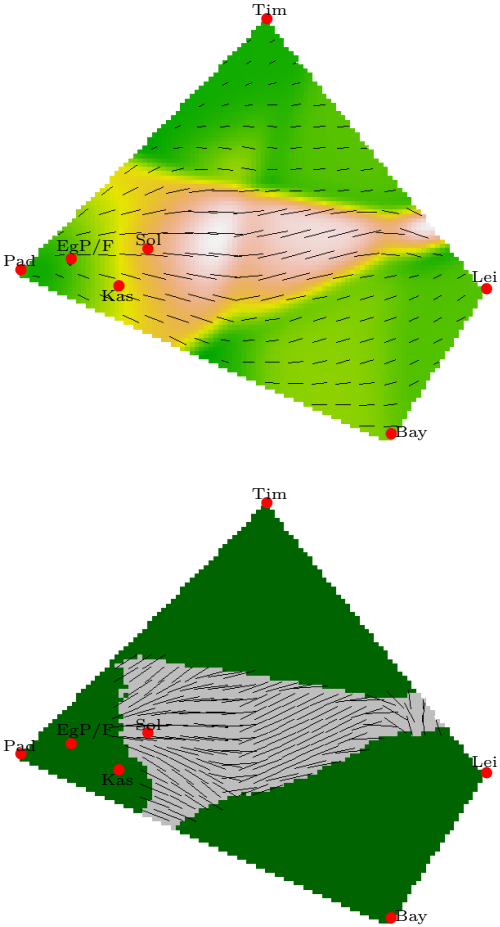

Figure 3.12.: Systemic function map (left) and genetic boundaries (right) of eight German *L. terrestris* sampling locations (red points) obtained with the Wombling method. Yellow, pink and white: potential boundaries, grey: boundaries confirmed by binomial test. Note: GER-EgP and GER-EgF appear as one point because of close proximity of sampling locations but were analysed as two independent locations

3.3.6. Quantification of overall differentiation

The common differentiation measure F_{ST} and its Bayesian estimator $\theta^{(II)}$ showed similar values of differentiation (F_{ST}: 0.124, $\theta^{(II)}$: 0.117 ± 0.011 (full model), 0.110 ± 0.006 (f=0 model)). The values for δ were lower, with 0.070 ± 0.044 for square-root and 0.065 ± 0.040 for Bayesian estimation of allele frequencies, respectively.

The highest values of D_k were observed in the French subpopulation FRA-Bru with 0.109 ± 0.134 for square root allele frequency estimates and 0.105 ± 0.131 for Bayesian allele frequency estimates, indicating that this subpopulation in France was least representative for the whole population. The subpopulation from the pasture in the Egge-range in Germany (GER-EgP) had the lowest D_k values of 0.053 ± 0.050 when allele frequency estimations with the square root method were applied and 0.049 ± 0.047 when allele frequency estimations were done with the Bayesian method and was therefore the most representative subpopulation. The Canadian earthworms had also comparatively low D_k values (0.059 ± 0.063 for square root estimates of allele frequencies, 0.054 ± 0.055 for Bayesian estimates of allele frequencies). In general, D_k values based on Bayesian estimation of allele frequencies were slightly lower than values obtained by the square root method of allele frequency estimation. The method of allele frequency estimation also altered the order of subpopulations in some cases (figure 3.13).

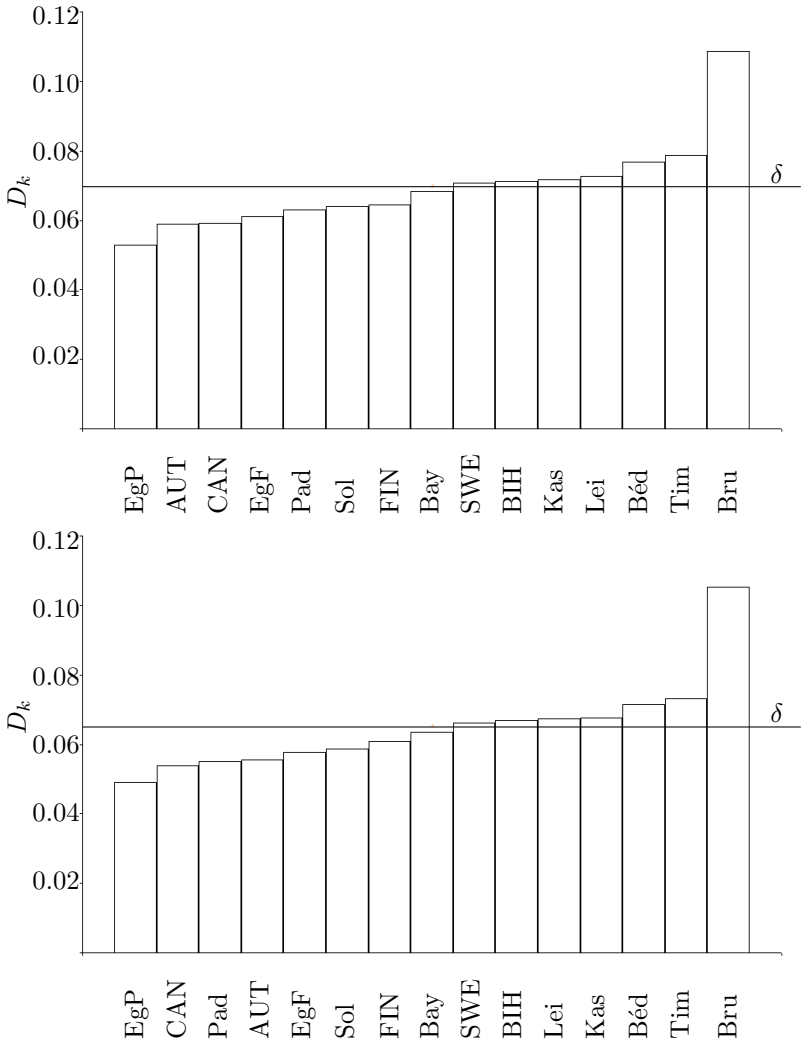

Figure 3.13.: Genetic distances between sampling locations and their complement (D_k) and overall differentiation (δ), based on square root (above) and Bayesian estimates (below) of allele frequencies.

3.3.7. Arrangements of regions tested with AMOVA

The patterns of genetic boundaries and similarities between sampling locations were used to define hypothetical sets of regions to be tested with the AMOVA approach. The greatest part of variance was found within populations in all of the tested sets of regions, followed by variance components among populations and among regions. Among regions variances ranged from 6.17 to 9.49 %. All variance components and F-statistics were significant. The overall F_{ST}, if no regions were assumed (set 1), was 0.189 and therefore slightly higher as the common F_{ST} measure based on allele frequencies. The highest among regions-variance and F_{CT} was found in set 5 which was based on the results of the Monmonier algorithm based on D_S (with Bayesian method of allele frequency estimation). The lowest regional resolution was obtained by the set assuming countries as regions (set 2). This set also showed the lowest significance for F_{CT}. The set based on the Wombling approach (set 7) had also a low regional resolution, possibly due to the comparatively low numbers of regions this set consisted of. The arrangement of FIN, SWE, BIH, AUT and GER-Bay in one region, based on the consideration of the FIN-SWE-BIH-cluster in the corresponding UPGMA dendogram (figure 3.6), led to a lower among regions-variance and F_{CT} in set 6b compared to set 6a, the set without the conection between FIN, SWE and BIH (table 3.5).

Table 3.5.: AMOVA results for different sets of regions.

	set 1(one region)			
Variance component	Variance	% total	p	F-statistics
Among regions				
Among populations/regions	2.329	18.94	***	
Within populations	9.957	81.06	***	$F_{ST} = 0.189$
	set 2 (countries)			
Among regions	0.773	6.17	*	$F_{CT} = 0.062$
Among populations/regions	1.801	14.36	***	$F_{SC} = 0.153$
Within populations	9.967	79.47	***	$F_{ST} = 0.205$
	set 3 (Tanimoto)			
Among regions	0.900	7.21	***	$F_{CT} = 0.072$
Among populations/regions	1.608	12.89	***	$F_{SC} = 0.$
Within populations	9.967	79.90	***	$F_{ST} = 0.201$
	set 4 (Simple matching)			
Among regions	0.978	7.84	***	$F_{CT} = 0.078$
Among populations/regions	1.527	12.24	***	$F_{SC} = 0.133$
Within populations	9.967	79.91	***	$F_{ST} = 0.201$
	set 5 (D_S)			
Among regions	1.187	9.49	***	$F_{CT} = 0.095$
Among populations/regions	1.360	10.87	***	$F_{SC} = 0.120$
Within populations	9.967	79.64	***	$F_{ST} = 0.206$
	set 6a (d_0)			
Among regions	1.140	9.13	***	$F_{CT} = 0.091$
Among populations/regions	1.379	11.04	***	$F_{SC} = 0.122$
Within populations	9.967	79.82	***	$F_{ST} = 0.202$
	set 6b (d_0+UPGMA dendrogram)			
Among regions	1.052	8.39	***	$F_{CT} = 0.084$
Among populations/regions	1.514	12.08	***	$F_{SC} = 0.132$
Within populations	9.967	79.53	***	$F_{ST} = 0.205$
	set 7 (Wombling)			
Among regions	0.843	6.68	***	$F_{CT} = 0.067$
Among populations/regions	1.811	14.35	***	$F_{SC} = 0.154$
Within populations	9.967	78.97	**	$F_{ST} = 0.210$

4. Discussion

4.1. AFLP in *Lumbricus terrestris*

After more than ten years, the AFLP method is now a standard procedure in population genetics, evolutionary ecology, phylogeography and related fields. However, most studies using AFLPs investigated plants and microorganisms and only a small number of studies used AFLPs in animals, mainly vertebrates (Bensch and Akesson 2005). The reasons for this bias towards plants, microorganisms and vertebrates are unknown. AFLPs can be applied universally and, if a certain quality of DNA is provided, should lead to results in every kind of organism after an appropriate combination of restriction enzymes and primers has been determined (Mueller and Wolfenbarger 1999). The AFLP technique is by far less laborious than the development of primers for sequencing or large numbers of microsatellite markers. After establishing a lab routine, AFLPs are quite fast and easy to perform. The present study was successful in applying the AFLP procedure in the earthworm *Lumbricus terrestris*. As far as it is known, it is the first attempt to use this method in an earthworm species.

Like every genetic marker AFLPs face advantages and disadvantages. Two characteristics of AFLPs are the main sources of problems: The first one is the anonymous nature of AFLPs. That means,

every DNA in a sample, regardless if it is the target DNA or a contamination, can be processed and will presumably lead to peaks in the electropherogram (Dyer and Leonard 2000). For this reason it is essential to prepare samples during DNA isolation and subsequent lab work in a way that contaminations are eliminated or at least reduced. Analysing all or several samples twice or more times can help to asses the amount of contaminations. This, of course, makes lab work more expensive and time consuming. In the presented study the earthworms were washed several times and the tissue sampling was performed keeping the gut intact. The gut is the most likely source of contaminations in earthworms, because it contains living microorganisms and undigested organic material. Letting the earthworms empty their guts before tissue sampling should also reduce the risk of contaminations in cases were oversighted injuries of the gut occured. The replicated sampling and AFLP profiling of a proportion of all samples in the present study did not indicate any variances in AFLP fingerprints caused by contaminations.

The second drawback of AFLPs is the dominant character of this method. It is not possible to distinguish all possible genotypes and therefore allele frequencies can not be measured directly. If certain informations about the populations, like validation of Hardy-Weinberg proportions, are missing, the estimation of allele frequencies exhibits some degree of uncertainty. New approaches try to estimate allele frequencies by circumvent such assumptions, like the approaches implemented in the HICKORY software which assesses allele frequencies without assumptions about Hardy-Weinberg proportions.

One crucial step in the analysis of AFLP fingerprints is the scoring of peaks or bands. There seems to be no common approach to

execute this step. It is possible to score AFLP fingerprints manually or by using a software including a feature for automatic scoring. Automatic scoring is faster than manual scoring, but often the scoring software is based on rigid definitions of 'containers' which collect peaks in a specified range of base pairs and merge them to one locus. Peaks near the minimum or maximum of these ranges are often ambiguously scored. Manual scoring is more flexible in this cases. However, manual scoring is much more time consuming, more error-prone and often the criteria on which the manual scoring was based are unknown. Therefore, manual scoring of the same samples can lead to different results if done by different researchers (Bonin et al. 2004). The scoring method applied in this study was based on a objective criterion, the repeatability of peaks. The software applied here automatically rejects 'containers' with a high number of ambiguously scored peaks. The resulting list of AFLP markers is totally independent on the person doing the analysis. However, some manual corrections (exclusion of markers with questionable peaks) had to be performed to obtain the 125 polymorphic AFLP markers finally used.

4.2. Genetic diversity

The four methods used to calculate expected heterozygosity, i. e. H_e based on two allele frequency estimations (square root method and Bayesian method) and hs calculated with two HICKORY models (full and f=0 model), showed similar values. Bayesian estimation of allele frequencies led to slightly higher estimates of H_e, which was previously reported in other studies (Krauss 2000), but did not effect the order of sampling locations in terms of heterozygos-

ity. The order of magnitude for heterozygosities in *L. terrestris* is in accordance with AFLP-derived estimates of heterozygosities in other invertebrates (Knowles and Richards 2005; Timmermans et al. 2005).

Lumbricus terrestris is an obligatory outcrossing species and it is likely that natural populations are in Hardy-Weinberg equilibrium. Observed heterozygosities, measured by enzyme electrophoresis in Italian *L. terrestris* were in good agreement with expected heterozygosities assuming Hardy-Weinberg proportions (Robotti 1982). Also, on the Faroe Islands ten populations of *L. terrestris* did not show significant deviations from Hardy-Weinberg proportions (Enckell et al. 1986). It is unknown if the assumption of Hardy-Weinberg proportions, required for the calculation of H_e, can be expanded to the *L. terrestris* populations of the present study. However, the model choice criteria in HICKORY proposed to use the f=0 model which assumes no inbreeding in the investigated populations. The full model from HICKORY also reported a low value for the Bayesian analogue of F_{IS}, f=0.11, but this estimation should be handled with care (Holsinger and Lewis 2007). Therefore, the assumption of Hardy-Weinberg proportions in the investigated populations of *L. terrestris* seems to be valid or at least deviations from Hardy-Weinberg proportions seem to be of minor importance when estimating heterozygosities at AFLPs.

The differences in interindividual Tanimoto and Simple matching distances were caused by the different treatment of the Absence-phenotype in these two methods. The Simple matching method counts the appearance of the Absence-phenotype in two individuals as a similarity between them, whereas the Tanimoto method completely ignores this condition. Because of the high percentage of

ABSENCE-phenotypes in the data set, Simple matching values are obviously biased to higher similarities (lower distances) compared to Tanimoto values. This may also be the reason for a higher Spearman index between Simple matching distances and allele frequency-based genetic diversity estimates, because these methods have in common that they are based on the frequency of the ABSENCE-phenotype.

Genetic diversity of earthworms was measured before in studies dealing with different species of Lumbricidae and using different methods. Mostly enzyme electrophoresis at a restricted number of loci, but also RAPDs and sequencing of mtDNA were applied. There was no clear relationship between genetic diversity and reproduction strategy (cross-fertilization vs. parthenogenetic) or ecological life form of Lumbricidae (endogeic, epigeic and anecic). Genetic diversity of the parthenogenetic and endogeic *Aporrectodea trapezoides* (Duges) living in Australian agricultural soils was reported as low, when measured with RAPDs (Dyer et al. 1998), just as the genetic (clonal) diversity of the parthenogenetic and endogeic *Octolasion tyrtaeum* in Northern Europe and the USA (Jeanike and Selander 1985; Terhivuo and Saura 2006). Other parthenogenetic Lumbricidae with a low genetic diversity at seven enzyme loci are *Octolasion cyaneum*, which was investigated in Central Europe, Northern Europe and Australia, and *Eiseniella tetraedra* (Savigny) (Terhivuo and Saura 2003, 2006).

In the parthenogenetic reproducing and epigeic earthworm *Dendrobaena octaedra* sampled in Canadian forests genetic diversity is high when 609 base pairs of the mitochondrial genome are sequenced (Cameron et al. 2008). Likewise, analysis of isozymes in European and North American populations of this species also led to high clonal diversity measures (Terhivuo and Saura 2006; Hansen et al.

2006). *Aporrectodea rosea* (Savigny) showed a high clonal diversity in Northern Europe when three enzyme loci were observed (Terhivuo and Saura 1997, 2006).

Among Lumbricidae reproducing by cross-fertilization, *Aporrectodea caliginosa* (Savigny) of agricultural soils in Australia exhibits low genetic diversity at RAPDs (Dyer et al. 1998) and *Lumbricus rubellus* (Hoffmeister) populations from Italy have a low genetic diversity at six enzyme loci (Robotti 1984). In contrast, genetic diversity of *Aporrectodea tuberculata* in the USA was reported as high when measured by enzyme electrophoresis (Stille et al. 1980).

The presumably first investigation of genetic diversity of *L. terrestris* was performed on individuals from Northern Italy in 1981/82 (Robotti 1982). Electrophoresis at seven enzyme loci led to a mean observed heterozygosity of 0.304. On the Faroe Islands heterozygosity measured at four enzyme loci of ten *L. terrestris* populations ranged from 0.233 to 0.476 (mean: 0.367) (Enckell et al. 1986). These results are higher than the heterozygosity values derived from AFLPs in the presented study but these differences are possibly caused by the maximum heterozygosity of 0.5 in AFLPs, the different types of markers applied and the different numbers of loci investigated, and therefore the better coverage of the whole genome by AFLPs. In the city of Münster (Germany) the high number of 17 different mtDNA haplotypes in 199 individuals of *L. terrestris* sampled at 14 locations were found. However, nucleotide diversity within sampling sites range from 0.010 to 0.049 (Field et al. 2007). Using RAPDs Kautenburger reported mean band-based Nei/Li distances (Nei and Li 1979) in five German *Lumbricus terrestris* populations ranging from 0.205 to 0.244 and from 0.265 to 0.395 at four sampling locations (Kautenburger 2006*a*,*b*; Nei/Li distances are re-

ported as similarities in the original studies). Genetic distances expressed as Nei/Li distances are similarly computed as Tanimoto distances. In general, Tanimoto distances are slightly higher then Nei/Li distances, but not as much as between the studies of Kautenburger and the presented study. Therefore, the sampling locations investigated in the present study are either genetically more diverse in terms of phenotypic fingerprints or AFLPs generally lead to higher diversity estimates than RAPDs.

The differences in numbers of sampling locations, general design of sampling regarding geographical resolution, the numbers of investigated loci and the varieties of molecular methods lead to difficulties in comparing the results of these studies, even when only a single species is considered. An elaborate comparison of methods, applied on a consistent set of sampling location over different geographical scales should be addressed in future surveys of single and multiple species of Lumbricidae to obtain comparable datasets of different earthworm species.

In the present study no clear geographical pattern of genetic diversity in *L. terrestris* was found. The lower genetic diversity of the subpopulations in France compared to the other sampling locations was possibly caused by a different origin of the French subpopulations with a low genetic diversity in the source refugial population. The low genetic diversity in Sweden possibly reflects the isolated geographic position of the Swedish subpopulation surrounded by water and harsh environmental conditions in the North of Sweden causing a bottleneck during postglacial recolonization and a reduced gene flow between Sweden and other locations. Comparatively high genetic diversities in German subpopulations were probably caused by a multiple colonization of Germany due to its central position

in Europe similar to patterns found in different vertebrate animals and in plants (Taberlet et al. 1998; Hewitt 2001; Petit et al. 2003).

4.3. Genetic differentiation

Genetic distances (Tanimoto and Simple matching) between individuals were significantly smaller within sampling locations than between sampling locations. Therefore, individuals from a given sampling location belong to a 'real' subpopulation with a higher gene flow within the subpopulation than to other subpopulations. Surprisingly, even the Canadian earthworms exhibit higher similarity among each other than to other sampling locations. It is safe to assume, that the Canadian earthworms have not been sampled at one single location but at different locations possibly far away from each other. The high similarity of these individuals can possibly be explained by a limited number of introductions from European *L. terrestris* to Canada. Nevertheless, conclusions based on this limited number of earthworms with a uncertain sampling should be made with caution.

There was a positive correlation between genetic and geographic distances of the German sampling locations. Regarding all European sampling locations, the results of the Mantel tests were ambiguous. In case of Tanimoto distances a low positive correlation between geographic and genetic distance was found, whereas in case of Simple matching distance no correlation was detected. These results were accordant to the distograms, which show a clear increasing pattern of mean genetic distances across the distance classes for the German populations and a considerable more complex pattern across distance classes containing all sampling locations. This

leads to the conclusion, that a simple relationship between geographic and genetic distances reflecting a distance-dependent gene flow pattern is only existing within a limited area (a geographic region). On a larger geographical scale these simple linear gene flow patterns are superimposed by different origins of far distant subpopulations on one hand and high genetic similarities due to long distance dispersals of subpopulations on the other hand. In contrast to the presented results, a correlation between geographic distance and genetic distance (Nei/Li distance) in *L. terrestris* sampled at five German locations 4.5 km to 310 km apart using 49 polymorphic RAPD loci was not found (Kautenburger 2006*a*).

The Monmonier algorithm and the Wombling approach were used to investigate genetic barriers between sampling locations. Both methods have drawbacks regarding the sampling locations at the outermost positions of the map, i. e. Bosnia-Herzegovina, France, Sweden and Finland. Possible connections or barriers between these distant sampling locations were therefore crosschecked by considering clusters in the UPGMA dendograms. Regions derived from the Monmonier results in general resemble the structure of clusters in the UPGMA dendrograms. The use of the Monmonier algorithm and the Wombling approach showed substructuring of the European *L. terrestris* population. To a certain amount the results depended on the nature of distance measure, but there were high similarities even between band-based and allele frequency-based approaches.

The Monmonier results and the dendrograms of the band-based measures, Tanimoto and Simple matching, suggested a structure of European *L. terrestris* populations consisting of a French region, a Balkan-Austria-Northern Europe region, and a core German region. The outer German locations (Lei, Tim, Bay, Pad) showed

ambiguous results and were often either isolated or assigned to the same region as the sampling locations in Sweden, Finland, Austria or Bosnia-Herzegovina depending on the kind of distance measure applied.

The separation of the Bavarian population is interesting and possibly constitutes the northern part of another region extending farther south towards Italy. In the soil invertebrate *Orchesella cincta* an Italian region isolated from the rest of Europe was identified using AFLPs and sequencing of mtDNA, supporting the hypothesis of the general existence of such a region for soil animals (Timmermans et al. 2005)

The very noticeable connection between the sampling locations in Bosnia-Herzegovina, Austria, Finland and Sweden when Tanimoto and Simple matching distances were applied was not found in case of allele-frequency based approaches. The genetic boundaries did not suggest this connection between south-eastern and northern Europe, but the UPGMA dendrogram for d_0 also contained a cluster of populations from Bosnia-Herzegovina, Finland and Sweden. In case of D_S a connection between Bosnia-Herzegovina, Austria, Finland and Sweden was not evident, neither in the Monmonier map nor in the UPGMA dendrogram. The same applies to the Wombling approach, not showing the connection between Northern Europe, Bosnia-Herzegovina and Austria, too, but which is also limited in the analysis of populations at the edges of the investigated area. The Wombling approach in general resulted in a more simplified map of genetic barriers between sampling locations in Europe. Analogue to the Monmonier algorithm, France was isolated from the rest of Europe. The strong structuring of German populations was not found by the Wombling approach with the exception of the

Bavarian sampling location (GER-Bay), almost isolated from every other sampling location.

The geographic structures derived from the Monmonier and the Wombling approaches were used as sets of hierarchical population structures to be tested in AMOVAs. The variance within populations was the greatest variance component in all of these sets, accounting for approximately 79 % of total variance. The among regions variance was small, but significant. The set of region with the highest regional resolution, i. e. the set with the highest among region variance and F_{CT} value, was the one based on the Monmonier algorithm and the D_S distance matrix. This set of regions (set 5) was also the most uncommon one, being the one separating the two French populations from each other. The corresponding UPGMA dendrogram of D_S also placed FRA-Bru basal to all other sampling locations. This condition was not found in any other dendrogram or map from the Monmonier and Wombling approaches.

The regional resolution in terms of genetic variance components was lower for band-based methods than for allele-frequency based methods, as seen in the AMOVA results (lower variance components and F_{CT} values). All sets of regions had lower variance components among regions compared to a study of population structure in the springtail *Orchesella cincta*, comparable because of being a study of a soil animal with a similar sampling design, where variance components of 25 % among regions and 59 % within populations were reported (Timmermans et al. 2005). This study perhaps measured a higher regional resolution because it included populations from an isolated Italian region, which was unfortunately not sampled in the present study.

4.4. Postglacial recolonization of Europe by *Lumbricus terrestris*

Today the 60° E longitude is the eastern limit of *L. terrestris* in Russia (Tiunov et al. 2006). However, most of *L. terrestris* found eastwards of the 40° E longitude were found in habitats strongly affected by men, indicating an eastwards orientated dispersal of *L. terrestris* mediated by human transport. In Russia (East of 30° E) *L. terrestris* was only occasionally found north of the 60° N latitude. In France the distribution of *L. terrestris* is limited in the South at approximately 47° N latitude. In the south-western part of France *L. terrestris* is substituted by *Lumbricus friendi* (Bouché 1972), but in the very North of Spain and North of Portugal *L. terrestris* is also present (Cosín et al. 1992; Rodríguez et al. 1997). In the Balkan region the Southern boarder of *L. terrestris*' distribution is approximately at the same latitude like in France.

The species *L. terrestris* presumably originated during the Weichsel glaciation (Bouché 1969, 1972), beginning 150,000 years ago and ending with the Last Glacial Maximum (LGM) 22,000 to 14,000 years ago. After the LGM, organisms in Europe had to overcome some less extreme changes in temperature, climate and vegetation. During the LGM a large ice sheet covered Scandinavia, Finland, northern Germany and Poland. Great Britain, except its southern-western parts, Iceland, the Alps and the Pyrenees were also covered by ice. Soils in northern and central Europe uncovered by ice were in permafrost condition reaching southwards presumably to a latitude of todays central France. In addition to low temperatures very dry conditions prevailed in Europe.

Ice-free regions of northern and central Europe were covered by the steppe-tundra vegetation type (Adams and Faure 1998). In more southern parts of Europe, i. e. southern France, the Iberian peninsula, Italy and the Balkan region, temperate dry steppe existed. Therefore, the investigated sampling locations in Europe were presumably either covered by ice (SWE, FIN, GER-Tim, GER-Sol), covered by steppe-like tundra and exhibited permafrost soils (FRA, AUT, GER, except GER-Tim and GER-Sol) or were covered by temperate dry steppe (BIH) during the LGM. Out of the investigated sampling location only the location in Bosnia-Herzegovina (BIH) could have been located in a potential refuge area for earthworms during the LGM.

Phylogeography and likely postglacial routes of colonization have been investigated in several animal and plant species. Repeated cycles of extinction and recolonisation due to climatic alterations were experienced differently by single taxa, leading to a more or less unique recolonisation pattern for every species (Taberlet et al. 1998). However, three general patterns for terrestrial animals could be derived, namely the patterns firstly described in the grasshopper *Chorthippus parallelus* (Zetterstedt), hedgehogs (*Erinaceus* L.) and the bear *Ursus arctos* L. (Hewitt 2000). Almost all species used the Iberian peninsula, Italy and Balkans as refuge areas. It is supposed that *L. terrestris* endured the LGM in refugial areas in the Balkan, in Italy and southern France, whereas the Iberian peninsula was no refuge (Bouché 1969, 1972).

The Balkan region was a potent recolonization source, inhabited by populations able to recolonise great parts of northern Europe, central Europe and even southwestern Europe in many cases. Accordingly, the results for *L. terrestris* suggest a connection between

the Balkan and Northern Europe. In contrast, Italy seems to has been a refuge with only restricted expansion potential. In only some cases (hedgehog, oaks) recolonisation routes starting in Italy and reaching beyond the alps were found (Hewitt 2000). In the present study *L. terrestris* was not investigated in the potential refuge Italy, but some Monmonier results indicate that Italy was isolated from the rest of Europe. The allocation of the Bavarian sampling location to this hypothetical Italian region despite its opposed position on the other site of the Alps may indicate a more recent leaking of Italian worms across the alps caused by human activities. The isolated position of the French subpopulations in the presented results suggests that potential refuge populations in Southern France were only able to recolonize the area of today's France.

In regions where descendants of different refuge areas met, so called suture zones or hybrid zones were formed. Again, every single species may or may not form suture zones, but there are general patterns. Two of these suture zones are related to geographic barriers and are located at the Alps and the Pyrenees, respectively. Two common other suture zones are not obviously correlated with geographic constraints of dispersal, one lying in Scandinavia probably caused by two meeting colonization routes coming from the North and South of Scandinavia and the other one is located in a more or less broad region between France and Germany, often created by meeting of recolonisation routes from the Balkan and Iberian refuge (Taberlet et al. 1998).

Suture zones result in sharp changes in genetic information and can therefore be identified with the Monmonier and Wombling methods. Two of these suture zones seem to be present in *L. terrestris*: the Alps are a potential geographical barrier for *L. terrestris* and

the results of the present study point to an isolated Italian region. Future investigations including Italian earthworms are needed to confirm this hypothesis. The isolation of the French populations and the strong structure among German sampling locations seem to represent another suture zone. The Pyrenees are nonrelevant as a suture zone for *L. terrestris*, simply because *L. terrestris* had no refuge on the Iberian peninsula. As expected, a suture zone in Scandinavia was also not found. Due to the absence of *L. terrestris* in Northern Scandinavia the colonization route from northern Scandinavia pointing southwards and creating a suture zone with the northwards directed colonization route simply did not exist.

Postglacial expansion was rapid for many species. All species investigated so far were able to disperse either by active movement or by pollen and seed dispersal. Earthworms are known to be slow dispersers, especially *L. terrestris* or other anecic species with a sedentary-like behaviour, inhabiting their vertical burrow system for longer times. The presumably fastest above ground movement of *L. terrestris* reported so far was up to 19 m in one night (Mather and Christensen 1988). Populations of *L. terrestris* spread at a speed of 10 m per year (Edwards and Bohlen 1996). Active dispersal of *L. terrestris* is to slow to explain today's range of *L. terrestris*. Even if a twofold dispersal speed of populations of 20 m per year is assumed, *L. terrestris* could only travel 200 km in the last approximately 10,000 years after the LGM, which is only one tenth of the distance between the southern-most and the northern-most sampling locations of this study (BIH and FIN). This leads to the suggestion that *L. terrestris* is strongly dependent on any mode of passive dispersal.

In areas at the boarders of the distribution of *L. terrestris*, which were colonised by earthworms rather recently, the dependence of *L. terrestris* on human activity to establish a population becomes visible. On the Faroe Islands, where man arrived between 1200 and 900 years ago, *L. terrestris* populations exist only in a few sites altered by human activity (Enckell et al. 1986). The north-easternmost populations of *L. terrestris* in Russia are also only occurring in habitats with a high anthropogenic impact (Tiunov et al. 2006).

Europe was colonised by humans from the Near East several times. The analysis of human mtDNA revealed colonisation waves 47,000, 28,000 and 16,500 years ago. The most recent colonisation wave was started by humans capable of agriculture approximately 9000 years ago (Lell and Wallace 2000). Considering the importance of human migration as vector for earthworm dispersal and agriculture as a prerequisite of creating suitable conditions for *L. terrestris* to establish a new population, this so called Neolithic Revolution can be seen as a starting point for the long range dispersal of wide-spread earthworm species like *L. terrestris* in Europe.

Two theories of spreading of agriculture into Europe exist. The demic-diffusion model depicts the migration of farmers from the Near East into Europe and bringing agricultural techniques and knowledge with them. The contrasting cultural-diffusion model states that only the knowledge of agriculture was delivered from the Near East farmers to European inhabitants without a substantial migration of people (Lell and Wallace 2000). For a dispersal mode of *L. terrestris* based on the movement of agriculture into Europe the demic-diffusion model seems more likely. The cultural-diffusion model should only be able to explain todays distribution of *L. terrestris*, if the transfer of agricultural knowledge was accompanied by

trading of plants or agricultural tools carrying earthworms or their cocoons. Different genetic methods and markers (gene frequencies, mtDNA, Y-chromosomal DNA) in Europeans lead to the conclusion that one-fifth to one half of the genetic information of Europeans today was brought by Neolithic farmers into Europe indicating that the demic-diffusion model is at least partly true (Lell and Wallace 2000).

Spread of farming across Europe was rather fast, lasting about 2500 years from Greece to the British Isles. One of the main directions of genetic information in humans points from southeastern to northwestern Europe (Richards 2003). If this depicts the movement of humans, agriculture and therefore a possible dispersal mechanism for *L. terrestris*, it explains the genetic structure found in the *L. terrestris* samples from Bosnia-Herzegovina, Austria, Sweden and Finland. The isolation of France is probably explained by another route of agricultural expansion.

Today's occurrence of *L. terrestris* in non-agricultural sites like grasslands or forest can be explained by a subsequent spread of *L. terrestris* from the initial populations in agricultural sites to other habitats. If these migrations have been completely independent of human activities or if they depended on utilization of non-agricultural sites by men is not known, but should be addressed by future research comparing different sites closely located. In the presented study this was addressed by investigating the grassland site and the forest site at the Egge-range (EgP and EgF). The results for genetic diversity and genetic distances for these locations revealed a high similarity between the two sites. This indicates that subpopulations are more affected by a common decent and not by differences in utilization of sites.

4.5. Conclusion

The present study successfully applied the AFLP method and revealed a genetic structure among several sampling locations or populations of *L. terrestris* across Europe. Possible postglacial recolonisation routes of *L. terrestris* resemble patterns of recolonisation in other animals but also exhibit unique features. *L. terrestris* was most likely dispersed by Neolithic farmers expanding into Europe from the Near East. The pattern of genetic barriers and similarities in *L. terrestris* can be explained by the expansion of agriculture into Europe. The genetic diversities measured with different methods in the subpopulations could partly interpreted in the light of different origins of subpopulations and by isolation of some locations, for example Sweden.

To thoroughly investigate the population genetic structure of soil organisms, molecular markers, like AFLP, are appropriate tools. In the future, extensive sampling across large geographical ranges including primary and introduced subpopulations of soil organisms should lead to deeper insights into population dynamics, origins, dispersal and migration of these organisms. A comparison of locations with different utilization by humans should lead to a better understanding how earthworms were able to start populations in forest or grassland after an initial establishment in agricultural sites. Finally, this should also lead to a better understanding of the role of these populations in soil functions and their connections to aboveground systems.

More attention is needed in the appropriate analysis and interpretation of results based on genetic markers. Different markers and alternative methods for assessing genetic diversity and differentia-

tion have to be compared and evaluated. It was shown in the present study that a variety of analysis methods can lead to more or less pronounced differences in results and interpretations. The use of mtDNA should lead to a better understanding of postglacial migration of *L. terrestris*, whereas microsatellites could complement the AFLP data to asses genetic diversity of populations. For a better understanding of genetic differentiation and postglacial recolonization patterns at a large scale the number of sampling locations and investigated individuals should be extended.

5. Summary

Genetic diversity and genetic differentiation of the anecic earthworm *Lumbricus terrestris* were investigated at 14 sampling locations in Europe and in individuals descended from introduction of earthworms to Canada by European settlers. A total of 394 individuals was investigated.

The AFLP method (Amplified Fragment Length Polymorphism) was successfully applied to this species, the first time this method was used in a bi-parentally reproducing earthworm species. Based on reproducibility of AFLP peaks, 125 polymorphic AFLP markers were identified and used to measure genetic diversity and genetic differentiation with a variety of methods.

There was no clear overall pattern of genetic diversity, but low values in France and Sweden could be explained by specific origins or geographical characteristics of sampling locations.

The Monmonier algorithm, the Wombling method and cluster analysis based on different genetic distance measures identified genetic boundaries in Europe, i. e. regions with a sharp change in genetic information. Using these boundaries it was possible to define several regions in Europe with a high probability of being colonised from the same refuge populations after the Last Glacial Maximum, approximately 14,000 years ago. The AMOVA procedure (Analysis of Molecular Variance) was used to investigate the proportion

of genetic Variance among these regions and revealed a small but significant amount of variation differentiating between the regions.

There seems to be a connection between *L. terrestris* populations from the Balkan to Northern Europe, i. e. Sweden and Finland, whereas populations from France and possibly Italy are more isolated from the rest of Europe. There is strong evidence that *L. terrestris* depended on human agricultural activity to spread across Europe and establish new populations. The pattern of spread of agriculture during the neolithic revolution, approximately 9000 years ago, can help to explain today's patterns found in the genetic structure of *L. terrestris* populations in Europe.

Molecular methods, like AFLP, are helpful tools to investigate origin, dispersal, genetic diversity and dynamics of populations of soil organisms which are often difficult to monitor.

Bibliography

Adams, J., and H. Faure, 1998. Quaternary Environments Network (QEN) website, http://www.esd.ornl.gov/projects/qen/adams1.html.

Alonso-Blanco, C., A. J. M. Peeters, M. Koornneef, C. Lister, C. Dean, N. van den Bosch, J. Pot, and M. T. R. Kuiper. 1998. Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. The Plant Journal **14**:259–271.

Bensch, S., and M. Akesson. 2005. Ten years of AFLP in ecology and evolution: why so few animals? Molecular Ecology **14**:2899–2914.

Bernstein, F. 1930. Über die Erblichkeit der Blutgruppen. Zeitschrift für Abstammung und Vererbungslehre **54**:400–426.

Bonin, A., E. Bellemain, P. B. Eidesen, F. Pompanon, C. Brochmann, and P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. Molecular Ecology **13**:3261–3273.

Bonin, A., D. Ehrich, and S. Manel. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. Molecular Ecology **16**:3737–3758.

Bouché, M. B. 1969. Comparison critique de methodes d'evaluation des populations de lumbricides. Pedobiologia **9**:26–34.

Bouché, M. B., 1971. Relations entre les structures spatiales et fonctionelles des écosystemes, illustrées par le rôle pédobiologique des vers de terre. Pages 187–209 *in* P. Pesson, editor. La Vie dans les Sols, Aspects Nouveaux, Études Experimentales. Gauthier-Villars, Paris.

Bouché, M. B. 1972. Lombriciens de France — Ecologie et Systématique. Institut National de la Recherche Agronomique, Paris.

Bouché, M. B., 1977. Stratégies lombriciennes. Pages 122–132 *in* U. Lohm and T. Persson, editors. Soil Organisms as Components of Ecosystems, volume 25. Ecological Bulletins, Stockholm.

Cameron, E. K., E. M. Bayne, and D. W. Coltman. 2008. Genetic structure of invasive earthworms *Dendrobaena octaedra* in the boreal forest of Alberta: insights into introduction mechanisms. Molecular Ecology **17**:1189–1197.

Christian, E., and A. Zicsi. 1999. Ein synoptischer Bestimmungsschlüssel der Regenwürmer Österreichs (Oligochaeta: Lumbricidae). Die Bodenkultur **50**:121–131.

Cosín, D. J. D., D. Trigo, and R. Mascato. 1992. Earthworms of the Iberian peninsula. Species list and some biogeographical considerations. Soil Biology and Biochemistry **24**:1351–1356.

Crida, A., and S. Manel. 2007. WOMBSOFT: An R package that implements the Wombling method to identify genetic boundary. Molecular Ecology Notes **7**:588–591.

DeHaan, L. R., R. Antonides, K. Belina, and N. J. Ehlke. 2002. Peakmatcher: Software for semi-automated fluorescence-based AFLP. Crop Science **42**:1361–1364.

Dyer, A. R., J. C. S. Fowler, and G. H. Baker. 1998. Detecting genetic variation in exotic earthworms, *Aporrectodea* spp. (Lumbricidae), in Australian soils using RAPD markers. Soil Biology and Biochemistry **30**:159–165.

Dyer, A. T., and K. J. Leonard. 2000. Contamination, Error, and Nonspecific Molecular Tools. Phytopathology **90**:565–567.

Edwards, C. A., and P. J. Bohlen. 1996. Biology and Ecology of Earthworms. 3 edition. Chapman & Hall, London, UK.

Enckell, P. H., S. A. Bengtson, P. Douwes, M. N. B. Stille, and B. Wiman. 1986. The dispersal pattern of an anthropochorous species: Genetic variation in populations of *Lumbricus terrestris* L. (Lumbricidae) in the Faroe Islands. Hereditas **104**:253–261.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of Molecular Variance Inferred From Metric Distances Along DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. Genetics **131**:479–491.

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

Field, S. G., M. Lange, H. Schulenburg, T. P. Velavan, and N. K. Michiels. 2007. Genetic diversity and parasite defense in a fragmented urban metapopulation of earthworms. Animal Conservation **10**:162–175.

Gerard, B. M. 1967. Factors affecting earthworms in pastures. Journal of Animal Ecology **36**:235–252.

Gregorius, H.-R. 1978. The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. Mathematical Biosciences **41**:253–271.

Gregorius, H.-R. 1984. A unique genetic distance. Biometrical Journal **26**:13–18.

Gregorius, H.-R., and J. H. Roberds. 1986. Measurement of genetical differentiation among subpopulations. Theoretical and Applied Genetics **71**:826–834.

Guérard, E., and F. Manni, 2004. Barrier version 2.2 Manual of the user version 1.0.

Hansen, P. L., M. Holmstrup, M. Bayley, and V. Simonsen. 2006. Low genetic variation for *Dendrobaena octaedra* from Greenland compared to populations from Europe and North America: Refuge or selection. Pedobiologia **50**:225–234.

Heethoff, M., K. Etzold, and S. Scheu. 2004. Mitochondrial COII sequences indicate that the parthenogenetic earthworm *Octolasion tyrtaeum* (Savigny 1826) constitutes of two lineages differing in body size and genotype. Pedobiologia **48**:9–13.

Hendrix, P. F. 2006. Biological invasions belowground — earthworms as invasive species. Biological Invasions **8**:1201–1204.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature **405**:907–913.

Hewitt, G. 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Molecular Ecology **10**:537–549.

Hill, M. O. 1973. Diversity and evenness: A unifying notation and its consequences. Ecology **54**:32.

Holsinger, K. 1999. Analysis of genetic diversity in geographically structured populations: A Bayesian perspective. Hereditas **130**:245–255.

Holsinger, K., P. O. Lewis, and D. H. Dey. 2002. A Bayesian approach to inferring population structure from dominant markers. Molecular Ecology **11**:1157–1164.

Holsinger, K., and L. E. Wallace. 2004. Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophaea* (Orchidaceae). Molecular Ecology **13**:887–894.

Holsinger, K. E., and P. O. Lewis, 2007. Hickory: A package for analysis of population genetic data v1.1. Department of Ecology & Evolutionary Biology, U-3043, University of Connecticut.

Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaudoise des Sciences Naturelles **44**:223–270.

Jeanike, J., and R. K. Selander. 1985. On the coexistance of ecologically similar clones of parthenogentic earthworms. OIKOS **44**:512–514.

Jones, C. G., J. H. Lawton, and M.Shachak. 1994. Organisms as ecosystem engineers. Oikos **69**:373–386.

Kautenburger, R. 2006*a*. Genetic structure among earthworms (*Lumbricus terrestris L.*) from different sampling sites in western Germany based on random amplified polymorphic DNA. Pedobiologia **50**:257–266.

Kautenburger, R. 2006*b*. Impact of different agricultural practices on the genetic structure of *Lumbricus terrestris*, *Arion lusitanicus* and *Microtus arvalis*. Animal Biodiversity and Conservation **29**:19–32.

Knowles, L. L., and C. L. Richards. 2005. Importance of genetic drift during Pleistocene divergence as revealed by analyses of genomic variation. Molecular Ecology **14**:4023–4032.

Koene, J. M., T. Pförtner, and N. K. Michiels. 2005. Piercing the partner's skin influences sperm uptake in the earthworm *Lumbricus terrestris*. Behavioral Ecology and Sociobiology **59**:243–249.

Krauss, S. L. 2000. Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. Molecular Ecology **9**:1241–1245.

Lavelle, P., C. Lattaud, D. Trigo, and I. Barois. 1995. Mutualism and biodiversity in soils. Plant and Soil **170**:23–33.

Lell, J. T., and D. C. Wallace. 2000. The peopling of Europe from maternal and paternal perspectives. American Journal of Human Genetics **67**:1376–1381.

Lowe, C. N., and K. R. Butt. 2005. Culture techniques for soil dwelling earthworms: A review. Pedobiologia **49**:401–413.

Manni, F., E. Guérard, and E. Heyer. 2004. Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. Human Biology **76**:173–190.

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

Mather, J. G., and O. Christensen. 1988. Surfache movements of earthworms in agricultural land. Pedobiologia **32**:399–405.

Meudt, H. M., and A. C. Clarke. 2007. Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. Trends in Plant Science **12**:106–117.

Miller, M. P. 1999. MANTEL-STRUCT: A Program for the Detection of Population Structure via Mantel Tests. Journal of Heredity **90**:258–259.

Monmonier, M. 1973. Maximum-difference barriers: An alternative numerical regionalization method. Geographical Analysis **3**:245–261.

Mueller, U. G., and L. L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting. Trends in Evolution and Ecology **14**:389–394.

National Geospatial Intelligence Agency of the United States Department of Defense, 2007. GEOTRANS Version 2.4.1 Software.

Nei, M. 1972. Genetic distance between populations. The American Naturalist **106**:283–292.

Nei, M. 1973. Analysis of Gene Diversity in Subdivided populations. Proceedings of the National Academy of Science of the USA **70**:3321–3323.

Nei, M., and W. H. Li. 1979. Mathematical models for studying genetic variation in terms of restriction endonucleases. Proccedings of the National Academy of Sciences of the USA **76**:5269–5273.

Nuutinen, V., and K. R. Butt. 1997. The mating behaviour of the earthworm *Lumbricus terrestris* (Oligochaeta: Lumbricidae). Journal of Zoology **242**:738–798.

Olive, P. J. W., and R. B. Clark, 1978. Physiology of reproduction. Pages 271–368 *in* P. J. Mill, editor. Physiology of Annelids. Academic Press, London.

Petit, R. J., I. Aguinagalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos, S. Fineschi, D. Grivet, M. Lascoux, A. Mohanty, G. Müller-Starck, B. Demesure-Musch, A. Palmé, J. P. Martín, S. Rendell, and G. G. Vendramin. 2003. Glacial refugia: Hotspots but not melting pots of genetic diversity. Science **300**:1563–1565.

Raw, F. 1959. Estimating earthworm populations by using formalin. Nature **184**:1661–1662.

Richards, M. 2003. The Neolitihic Invasion of Europe. Annual Review of Anthropology **32**:135–162.

Robotti, C. A. 1982. Biochemical Polymorphism Of Earthworms. II. Enzymes of *Lumbricus terrestris* (Annelida, Oligochaeta). Pedobiologia **24**:41–43.

Robotti, C. A. 1984. Biochemical polymorphism of earthworms. 5. Enzymes of *Lumbricus rubellus* (Annelida, Oligochaeta). Pedobiologia **27**:49–50.

Rodríguez, T., D. Trigo, and D. D. Cosín. 1997. Biogeographical zonation of the western Iberian peninsula on the basis of the distribution of earthworm species. Journal of Biogeography **24**:893–901.

Rohlf, F. J. 2007. NTSYSpc: Numerical Taxonomy System, ver. 2.20. Exeter Publishing, Ltd., Setauket, NY.

Routledge, R. D. 1979. Diversity indices: which ones are admissible? Journal of Theoretical Biology **76**:503–515.

Saliba-Colombani, V., M. Causse, L. Gervais, and J. Philouze. 2000. Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. Genome **43**:29–40.

Satchell, J. E., 1967. Lumbricidae. Pages 259–322 *in* A. Burgess and F. Raw, editors. Soil Biology. Academic Press, London.

Schaefer, M. 1992. Brohmer — Fauna von Deutschland. Quelle & Meyer, Wiebelsheim, Germany.

Schmidt, O. 2001. Appraisal of the electrical octet method for estimating earthworm populations in arable land. Annals of Applied Biology **138**:231–241.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. University of Kansas science bulletin **38**:1409–1438.

Spearman, C. 1904. The proof and measurement of association between two things. American Journal of Psychology **15**:72–101.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B **64**:483–689.

Stille, B., H. Ochman, and K. Selander. 1980. Genetic structure of populations of the earthworm *Aporrectodea tuberculata*. OIKOS **34**:195–201.

Taberlet, P., L. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology **7**:453–464.

Terhivuo, J., and A. Saura. 1997. Island biogeography of North European parthenogenetic Lumbricidae . 1. Clone pool affinities and morphometric differentiation of Aland populations. Ecography **20**:185–196.

Terhivuo, J., and A. Saura. 2003. Low clonal diversity and morphometrics in the parthenogenetic earthworm Octolasion cyaneum (Sav.). Pedobiologia **47**:434–439.

Terhivuo, J., and A. Saura. 2006. Dispersal and clonal diversity of North-European parthenogenetic earthworms. Biological Invasions **8**:1205–1218.

Thielemann, U. 1986. Elektrischer Regenwurmfang mit der Oktett-Methode. Pedobiologia **29**:296–302.

Thielemann, U., 1989. Untersuchungen zur Lumbricidenfauna mit neu entwickelten Methoden in erosionsgefährdeten Gebieten des Kraichgaus. Ph.D. thesis, University of Heidelberg.

Timmermans, M. J. T. N., J. Ellers, J. Mariën, C. Verhoef, B. Ferwerda, and N. M. van Straalen. 2005. Genetic structure in *Orchesella cincta* (Collembola): strong subdivision of European populations inferred from mtDNA and AFLP markers. Molecular Ecology **14**:2017–2024.

Tiunov, A. V., C. M. Hale, A. R. Holdsworth, and T. S. Vsevolodova-Perel. 2006. Invasion patterns of Lumbricidae into previously earthworm-free areas of northeastern Europe and the western Great Lakes region of North America. Biological Invasions **8**:1223–1234.

van der Wurff, A. W. G., J. A. Isaaks, G. Ernsting, and N. M. van Straalen. 2003. Population substructures in the soil invertebrate *Orchesella cincta*, as revealed by microsatellite and TE-AFLP markers. Molecular Ecology **12**:1349–1359.

Vekemans, X., 2002. AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Vekemans, X., T. Beauwens, M. Lemaire, and I. R. Roldan-Ruiz. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of relationship between degree of homoplasy and fragment size. Molecular Ecology **11**:139–151.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Fijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A New Technique For DNA Fingerprinting. Nucleic Acids Research **23**:4407–4414.

Wardle, D. A. 2002. Communities and Ecosystems: linking the aboveground and the belowground components. Princeton University Press, Princeton.

Womble, W. 1951. Differential systematics. Science **28**:315–322.

Zhivotovsky, L. A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology **8**:907–913.

Appendix A.

Genetic distances

Table A.1.: Mean interindividual Tanimoto distances between and within sampling locations.

	FIN	SWE	GER-Tim	GER-Sol	GER-EgF	GER-EgP	GER-Pad	GER-Kas	GER-Lei	GER-Bay	FRA-Bru	FRA-Béd	AUT	BIH	CAN
FIN	0.5505														
SWE	0.5838	0.4900													
GER-Tim	0.6555	0.6558	0.6109												
GER-Sol	0.6830	0.6772	0.7127	0.6334											
GER-EgF	0.6472	0.6490	0.6773	0.6435	0.5840										
GER-EgP	0.6361	0.6354	0.6805	0.6499	0.6057	0.6043									
GER-Pad	0.6694	0.6662	0.6997	0.6827	0.6629	0.6604	0.6453								
GER-Kas	0.6777	0.6631	0.6992	0.6548	0.6305	0.6435	0.6804	0.5956							
GER-Lei	0.6405	0.6569	0.6822	0.6812	0.6438	0.6458	0.6731	0.6942	0.5772						
GER-Bay	0.6890	0.6718	0.7085	0.7068	0.7009	0.6951	0.7180	0.7245	0.6987	0.6633					
FRA-Bru	0.6871	0.6571	0.6929	0.7089	0.6916	0.6892	0.6984	0.7060	0.6932	0.7014	0.4215				
FRA-Béd	0.6854	0.6569	0.7122	0.7071	0.6888	0.6931	0.6970	0.7085	0.6832	0.6905	0.5996	0.5681			
AUT	0.6582	0.6515	0.7034	0.6924	0.6667	0.6624	0.6862	0.6877	0.6659	0.6871	0.6882	0.6680	0.6114		
BIH	0.6533	0.6424	0.7017	0.6987	0.6681	0.6643	0.6916	0.6902	0.6638	0.7066	0.6952	0.6838	0.6695	0.5984	
CAN	0.6689	0.6543	0.7009	0.6918	0.6666	0.6706	0.6945	0.6892	0.6858	0.6945	0.6580	0.6552	0.6542	0.6725	0.6173

Table A.2.: Mean interindividual Simple matching distances between and within sampling locations

	FIN	SWE	GER–Tim	GER-Sol	GER-EgF	GER-EgP	GER-Pad	GER-Kas	GER-Lei	GER-Bay	FRA-Bru	FRA-Béd	AUT	BIH	CAN
FIN	0.1454														
SWE	0.1478	0.1083													
GER-Tim	0.1930	0.1804	0.1818												
GER-Sol	0.1997	0.1836	0.2203	0.1897											
GER-EgF	0.1863	0.1748	0.2057	0.1855	0.1629										
GER-EgP	0.1818	0.1698	0.2070	0.1878	0.1716	0.1712									
GER-Pad	0.1987	0.1843	0.2189	0.2052	0.1987	0.1973	0.1938								
GER-Kas	0.1972	0.1780	0.2133	0.1885	0.1804	0.1857	0.2047	0.1650							
GER-Lei	0.1831	0.1776	0.2072	0.2017	0.1872	0.1883	0.2028	0.2070	0.1602						
GER-Bay	0.2054	0.1846	0.2208	0.2152	0.2143	0.2111	0.2249	0.2232	0.2122	0.1965					
FRA-Bru	0.1993	0.1739	0.2084	0.2104	0.2058	0.2053	0.2107	0.2087	0.2047	0.2112	0.1016				
FRA-Béd	0.1928	0.1682	0.2103	0.2031	0.1976	0.1987	0.2033	0.2032	0.1946	0.1988	0.1569	0.1419			
AUT	0.1862	0.1708	0.2119	0.2022	0.1927	0.1906	0.2042	0.1998	0.1922	0.2021	0.1999	0.1833	0.1655		
BIH	0.1821	0.1652	0.2091	0.2020	0.1913	0.1897	0.2040	0.1983	0.1891	0.2089	0.1987	0.1873	0.1871	0.1568	
CAN	0.1875	0.1690	0.2077	0.1985	0.1896	0.1911	0.2046	0.1970	0.1972	0.2023	0.1819	0.1744	0.1794	0.1847	0.1626

Table A.3.: Pairwise genetic distances D_S (square-root method) between sampling locations

	FIN	SWE	GER-Tim	GER-Sol	GER-EgF	GER-EgP	GER-Pad	GER-Kas	GER-Lei	GER-Bay	FRA-Bru	FRA-Béd	AUT	BIH	CAN
FIN	—														
SWE	0.0147	—													
GER-Tim	0.0199	0.0249	—												
GER-Sol	0.0229	0.0269	0.0217	—											
GER-EgF	0.0182	0.0236	0.0204	0.0130	—										
GER-EgP	0.0143	0.0210	0.0197	0.0116	0.0049	—									
GER-Pad	0.0194	0.0216	0.0226	0.0182	0.0151	0.0132	—								
GER-Kas	0.0269	0.0290	0.0258	0.0135	0.0135	0.0150	0.0196	—							
GER-Lei	0.0194	0.0299	0.0249	0.0217	0.0185	0.0155	0.0233	0.0309	—						
GER-Bay	0.0212	0.0242	0.0193	0.0175	0.0204	0.0163	0.0207	0.0248	0.0202	—					
FRA-Bru	0.0463	0.0466	0.0473	0.0465	0.0471	0.0432	0.0467	0.0499	0.0471	0.0423	—				
FRA-Béd	0.0242	0.0273	0.0312	0.0252	0.0250	0.0223	0.0243	0.0304	0.0260	0.0205	0.0251	—			
AUT	0.0179	0.0213	0.0206	0.0168	0.0164	0.0134	0.0163	0.0207	0.0180	0.0130	0.0435	0.0192	—		
BIH	0.0185	0.0213	0.0280	0.0239	0.0192	0.0161	0.0189	0.0257	0.0189	0.0203	0.0421	0.0205	0.0164	—	
CAN	0.0175	0.0223	0.0243	0.0189	0.0162	0.0149	0.0207	0.0226	0.0233	0.0181	0.0337	0.0132	0.0122	0.0164	—

Table A.4.: Pairwise genetic distances D_S (Bayesian method) between sampling locations

	FIN	SWE	GER-Tim	GER-Sol	GER-EgF	GER-EgP	GER-Pad	GER-Kas	GER-Lei	GER-Bay	FRA-Bru	FRA-Béd	AUT	BIH	CAN
FIN	—														
SWE	0.0126	—													
GER-Tim	0.0170	0.0212	—												
GER-Sol	0.0201	0.0235	0.0201	—											
GER-EgF	0.0165	0.0208	0.0180	0.0108	—										
GER-EgP	0.0129	0.0177	0.0171	0.0094	0.0046	—									
GER-Pad	0.0170	0.0195	0.0176	0.0130	0.0125	0.0104	—								
GER-Kas	0.0239	0.0250	0.0234	0.0117	0.0117	0.0129	0.0158	—							
GER-Lei	0.0166	0.0255	0.0216	0.0187	0.0159	0.0132	0.0176	0.0270	—						
GER-Bay	0.0186	0.0205	0.0167	0.0147	0.0185	0.0146	0.0162	0.0228	0.0176	—					
FRA-Bru	0.0437	0.0430	0.0429	0.0431	0.0440	0.0405	0.0421	0.0469	0.0449	0.0402	—				
FRA-Béd	0.0227	0.0235	0.0262	0.0214	0.0221	0.0199	0.0199	0.0260	0.0229	0.0170	0.0236	—			
AUT	0.0159	0.0186	0.0191	0.0153	0.0149	0.0119	0.0130	0.0190	0.0161	0.0114	0.0406	0.0159	—		
BIH	0.0165	0.0186	0.0232	0.0195	0.0171	0.0142	0.0161	0.0224	0.0165	0.0174	0.0405	0.0182	0.0141	—	
CAN	0.0162	0.0194	0.0200	0.0151	0.0145	0.0129	0.0161	0.0191	0.0196	0.0142	0.0316	0.0114	0.0098	0.0140	—

Table A.5.: Pairwise genetic distances d_0 (square-root method) between sampling locations

	FIN	SWE	GER–Tim	GER-Sol	GER-EgF	GER-EgP	GER-Pad	GER-Kas	GER-Lei	GER-Bay	FRA-Bru	FRA-Béd	AUT	BIH	CAN
FIN	—														
SWE	0.0621	—													
GER-Tim	0.0848	0.0907	—												
GER-Sol	0.0890	0.0922	0.0888	—											
GER-EgF	0.0769	0.0878	0.0899	0.0644	—										
GER-EgP	0.0733	0.0833	0.0859	0.0634	0.0410	—									
GER-Pad	0.0807	0.0842	0.0918	0.0824	0.0775	0.0697	—								
GER-Kas	0.0933	0.0971	0.0911	0.0704	0.0692	0.0681	0.0805	—							
GER-Lei	0.0836	0.0944	0.0923	0.0882	0.0790	0.0779	0.0872	0.0969	—						
GER-Bay	0.0929	0.0915	0.0885	0.0825	0.0904	0.0822	0.0949	0.0971	0.0924	—					
FRA-Bru	0.1174	0.1116	0.1270	0.1248	0.1236	0.1192	0.1238	0.1221	0.1192	0.1160	—				
FRA-Béd	0.0939	0.0927	0.1068	0.1009	0.0982	0.0951	0.0920	0.1049	0.0975	0.0874	0.0772	—			
AUT	0.0830	0.0856	0.0860	0.0788	0.0780	0.0673	0.0808	0.0811	0.0788	0.0739	0.1154	0.0856	—		
BIH	0.0809	0.0805	0.1048	0.0931	0.0800	0.0719	0.0847	0.0899	0.0828	0.0890	0.1122	0.0877	0.0783	—	
CAN	0.0780	0.0820	0.0943	0.0820	0.0788	0.0742	0.0836	0.0874	0.0937	0.0813	0.0962	0.0690	0.0644	0.0751	—

Table A.6.: Pairwise genetic distances d_0 (Bayesian method) between sampling locations

	FIN	SWE	GER-Tim	GER-Sol	GER-EgF	GER-EgP	GER-Pad	GER-Kas	GER-Lei	GER-Bay	FRA-Bru	FRA-Béd	AUT	BIH	CAN
FIN	—														
SWE	0.0587	—													
GER-Tim	0.0792	0.0859	—												
GER-Sol	0.0852	0.0893	0.0853	—											
GER-EgF	0.0728	0.0823	0.0845	0.0605	—										
GER-EgP	0.0695	0.0779	0.0803	0.0585	0.0398	—									
GER-Pad	0.0784	0.0830	0.0833	0.0727	0.0717	0.0639	—								
GER-Kas	0.0861	0.0900	0.0864	0.0656	0.0645	0.0630	0.0718	—							
GER-Lei	0.0772	0.0879	0.0866	0.0830	0.0736	0.0724	0.0767	0.0914	—						
GER-Bay	0.0879	0.0871	0.0831	0.0755	0.0867	0.0779	0.0859	0.0929	0.0862	—					
FRA-Bru	0.1126	0.1068	0.1211	0.1196	0.1180	0.1144	0.1196	0.1183	0.1165	0.1144	—				
FRA-Béd	0.0901	0.0866	0.0996	0.0952	0.0922	0.0898	0.0882	0.0963	0.0913	0.0814	0.0752	—			
AUT	0.0773	0.0802	0.0822	0.0749	0.0738	0.0633	0.0718	0.0776	0.0748	0.0697	0.1109	0.0786	—		
BIH	0.0769	0.0761	0.0970	0.0863	0.0756	0.0682	0.0798	0.0827	0.0768	0.0836	0.1096	0.0822	0.0726	—	
CAN	0.0743	0.0781	0.0862	0.0746	0.0746	0.0698	0.0769	0.0798	0.0858	0.0750	0.0919	0.0648	0.0585	0.0715	—

Appendix B.

Dendrograms

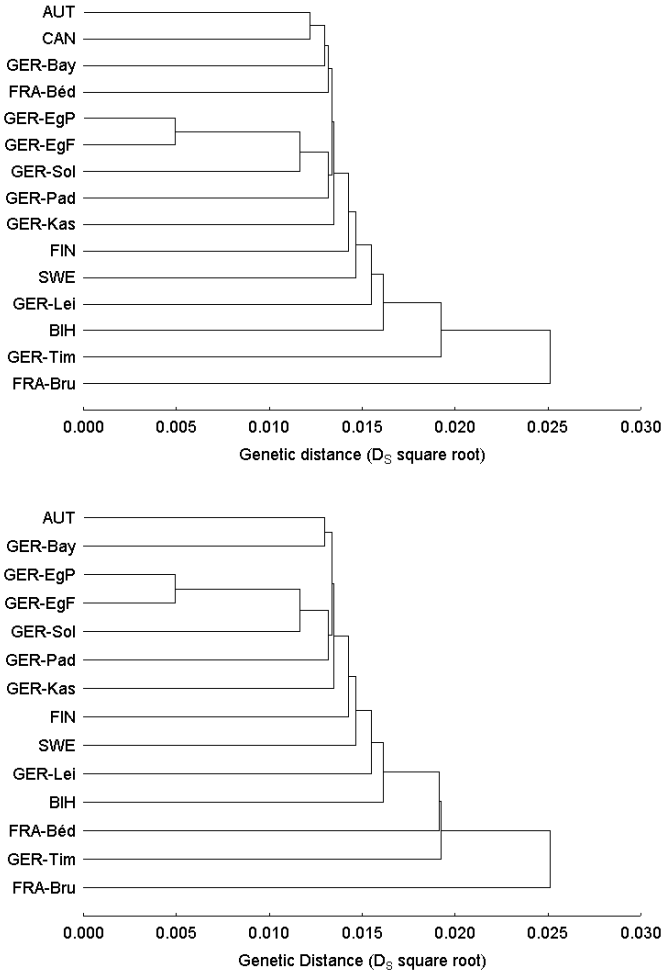

Figure B.1.: Single linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with square-root method.

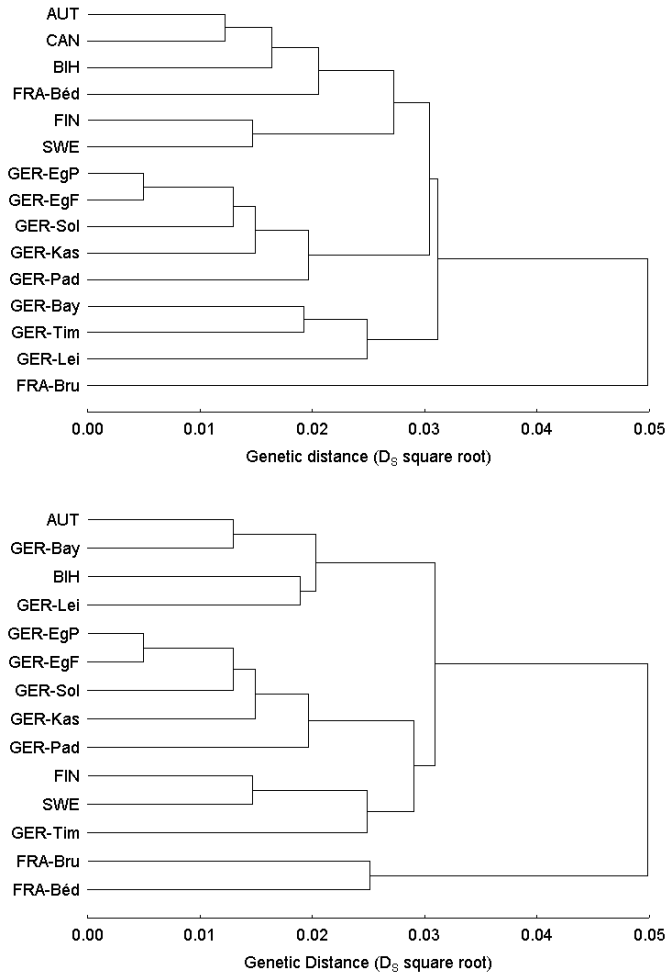

Figure B.2.: Complete linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with square-root method.

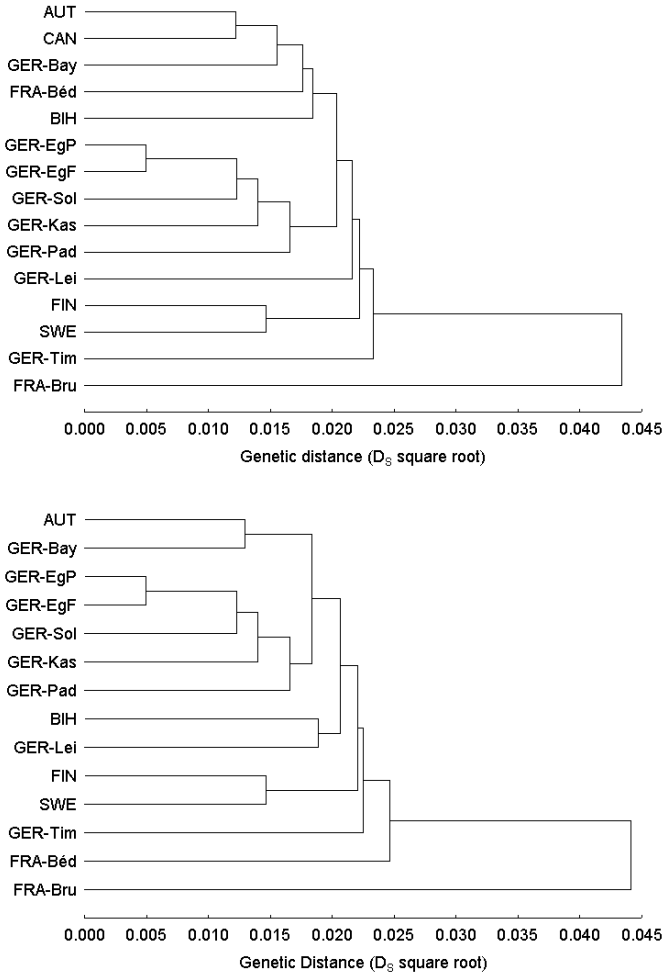

Figure B.3.: UPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with square-root method.

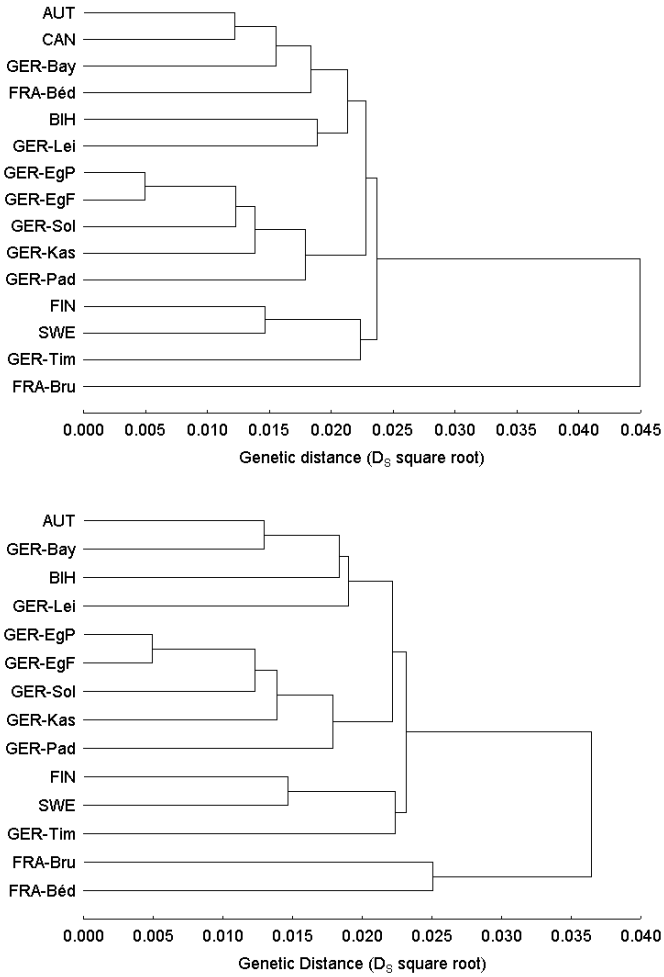

Figure B.4.: WPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with square-root method.

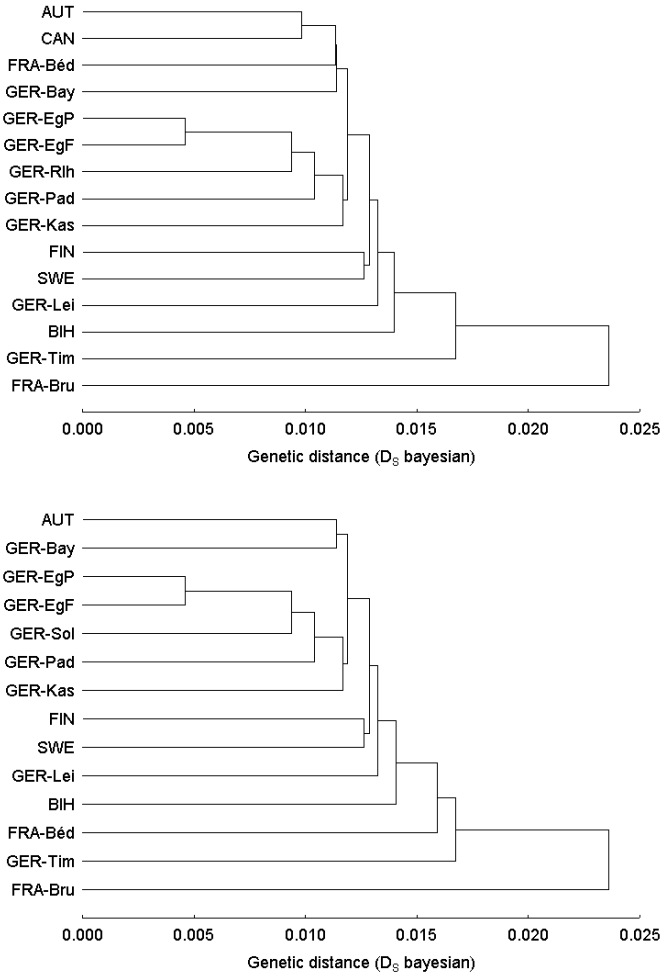

Figure B.5.: Single linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with Bayesian method.

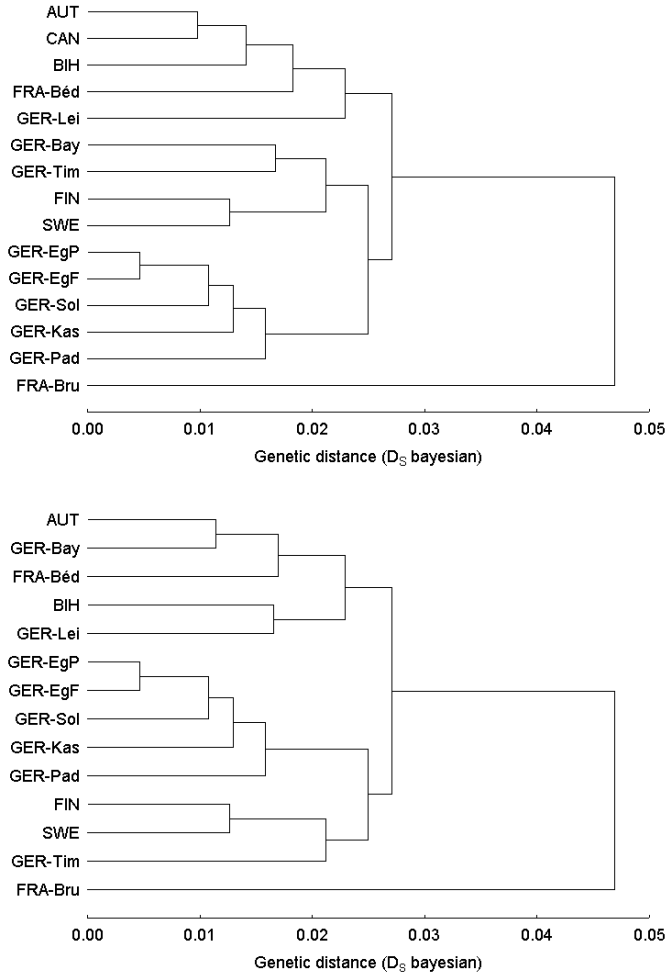

Figure B.6.: Complete linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with Bayesian method.

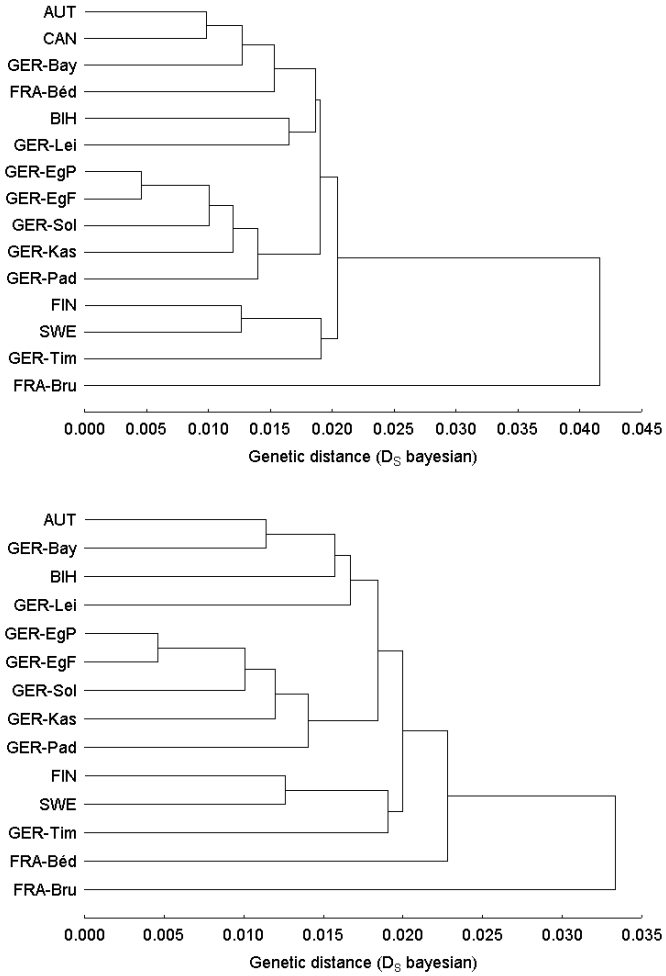

Figure B.7.: WPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: D_S, allele frequencies estimated with Bayesian method.

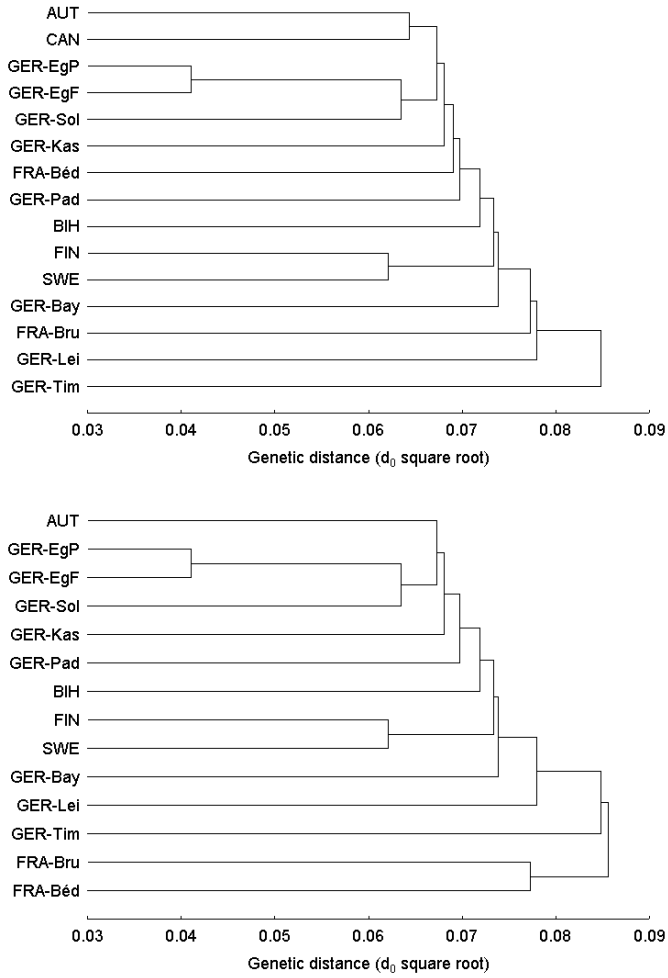

Figure B.8.: Single linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with square-root method.

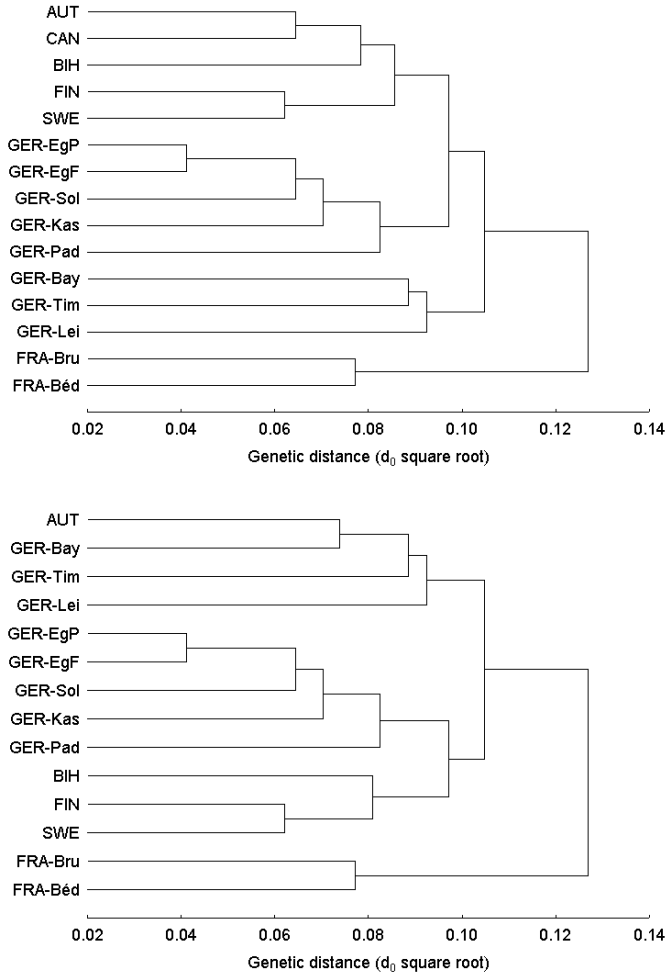

Figure B.9.: Complete linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with square-root method.

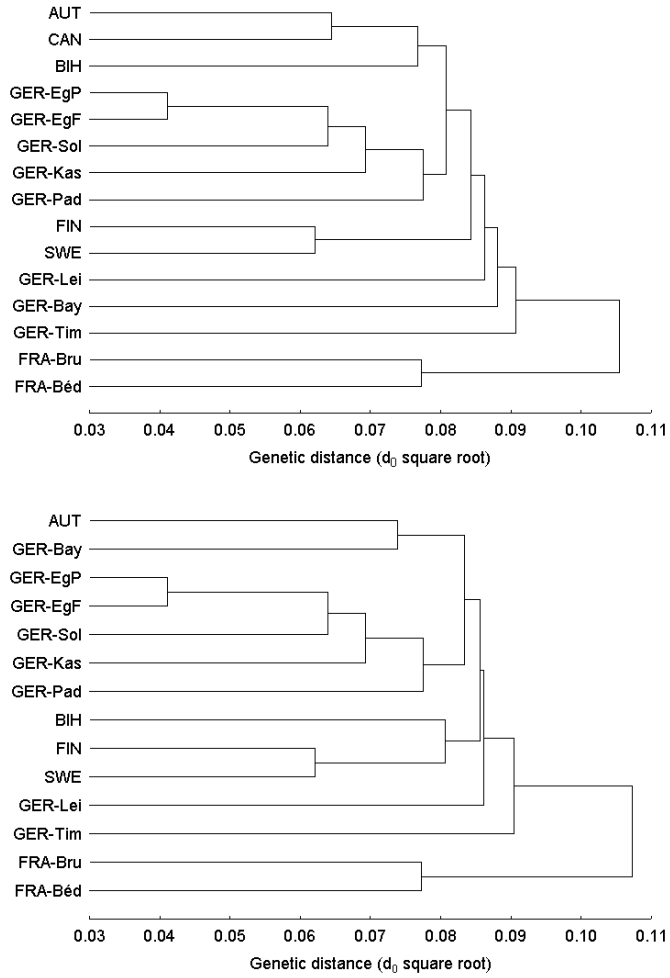

Figure B.10.: UPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with square-root method.

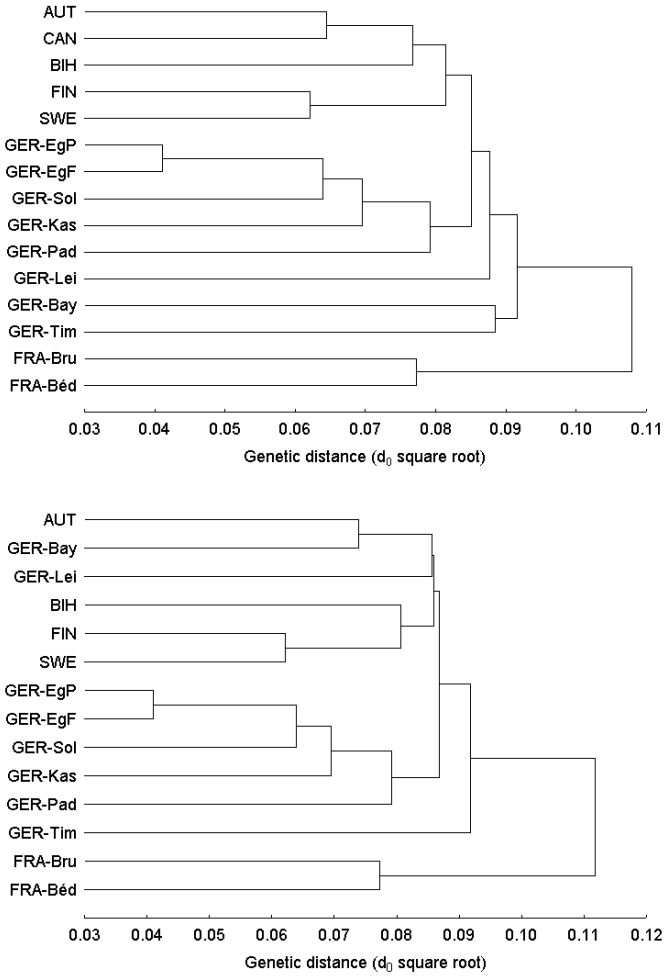

Figure B.11.: WPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with square-root method.

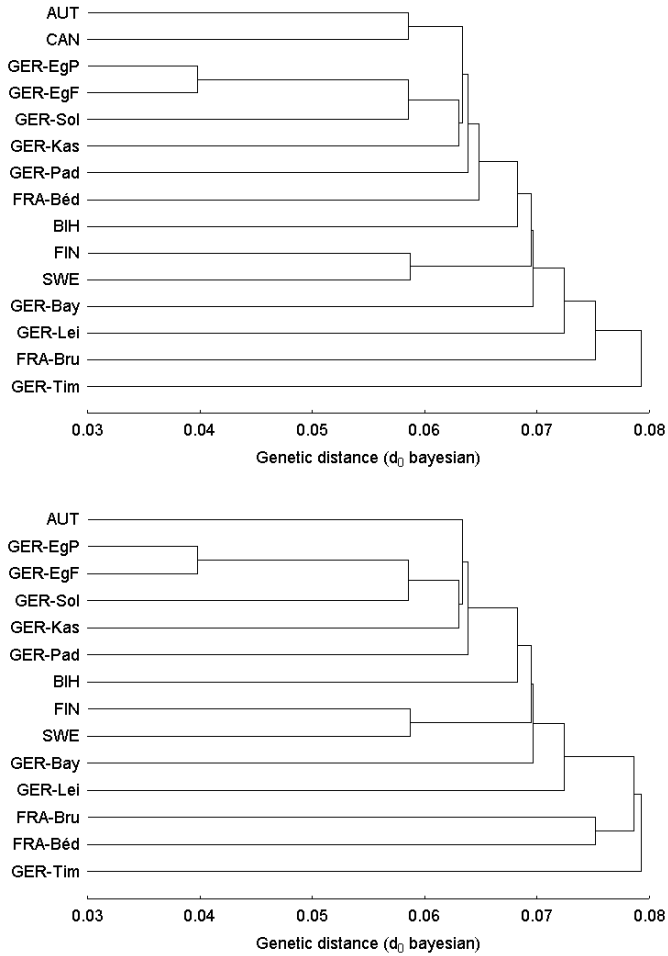

Figure B.12.: Single linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with Bayesian method.

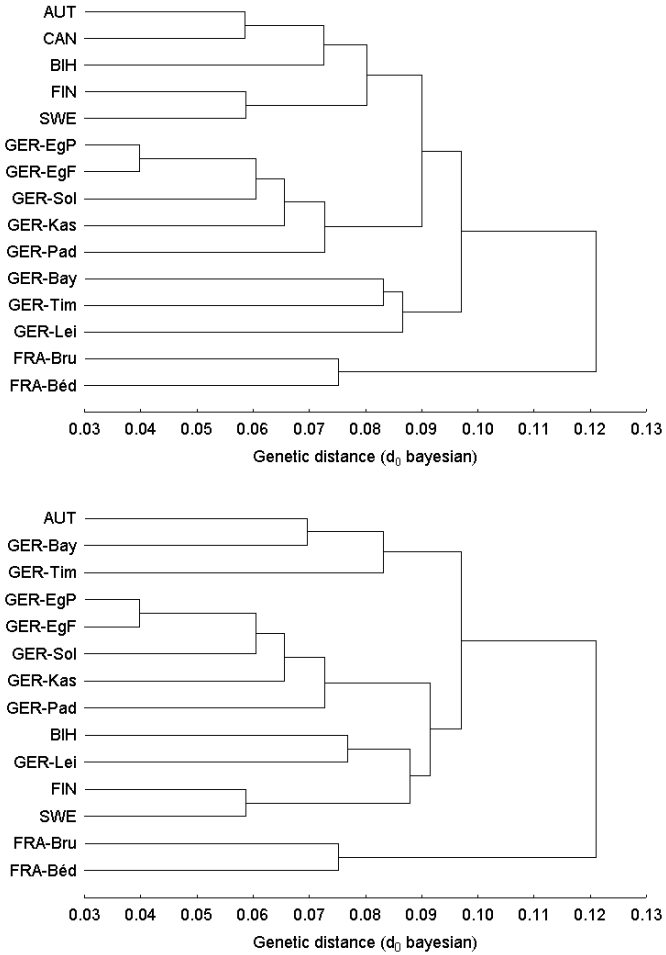

Figure B.13.: Complete linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with Bayesian method.

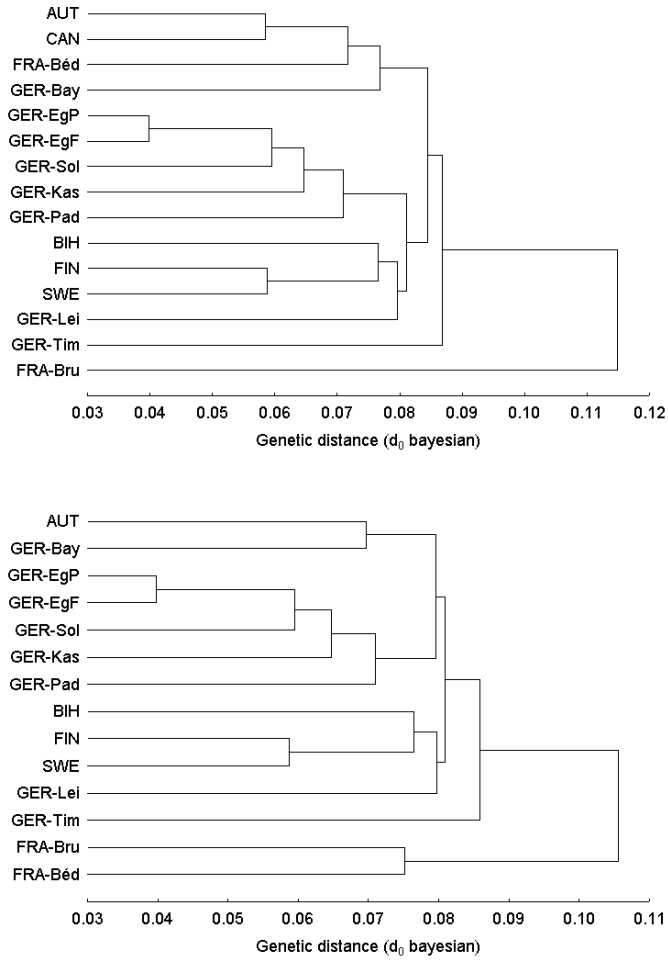

Figure B.14.: WPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: d_0, allele frequencies estimated with Bayesian method.

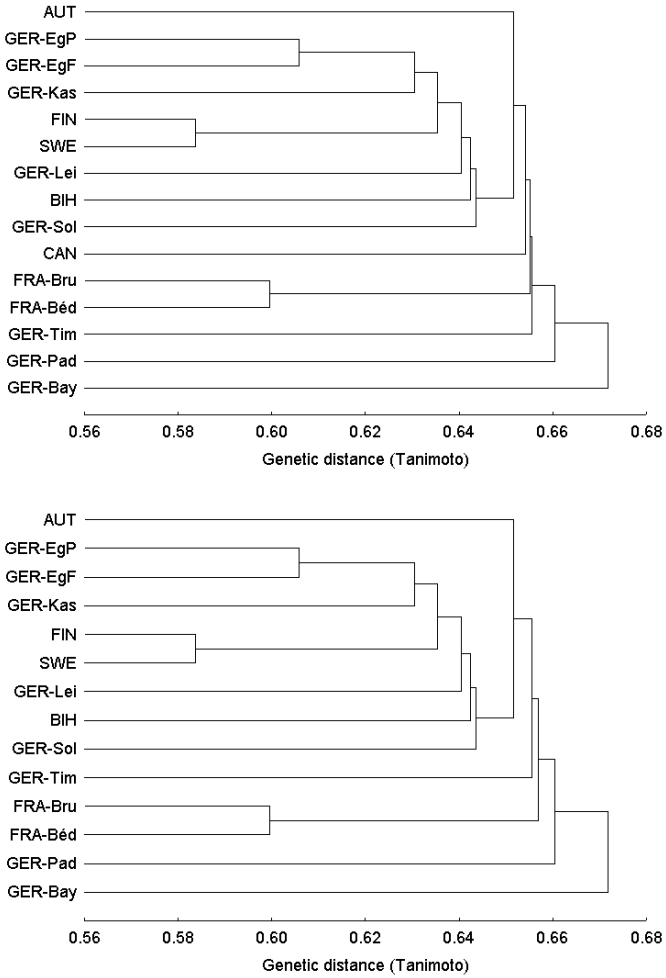

Figure B.15.: Single linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Tanimoto distance.

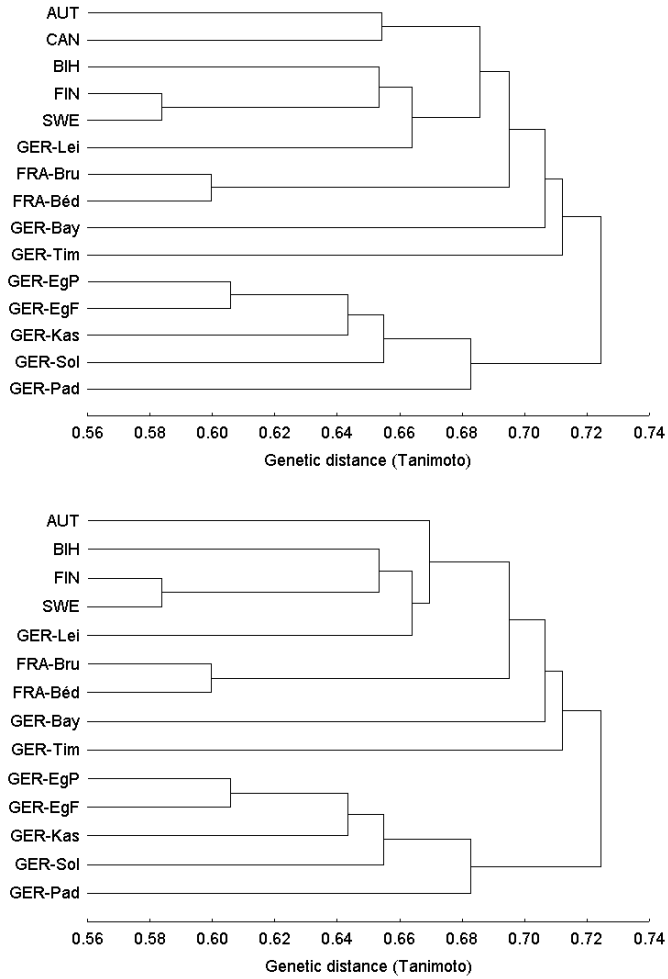

Figure B.16.: Complete linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Tanimoto distance.

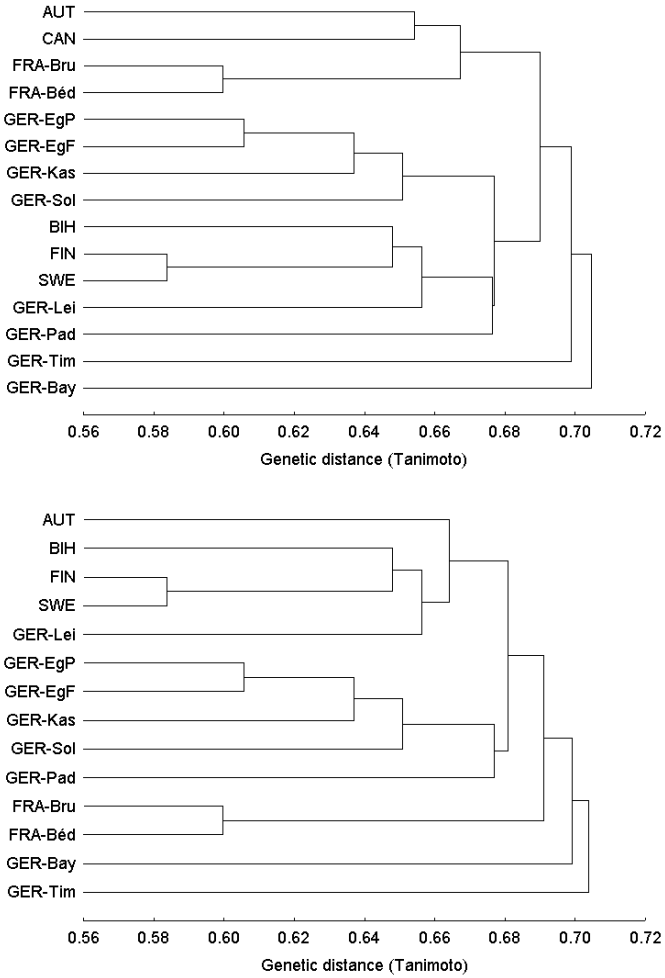

Figure B.17.: WPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Tanimoto distance.

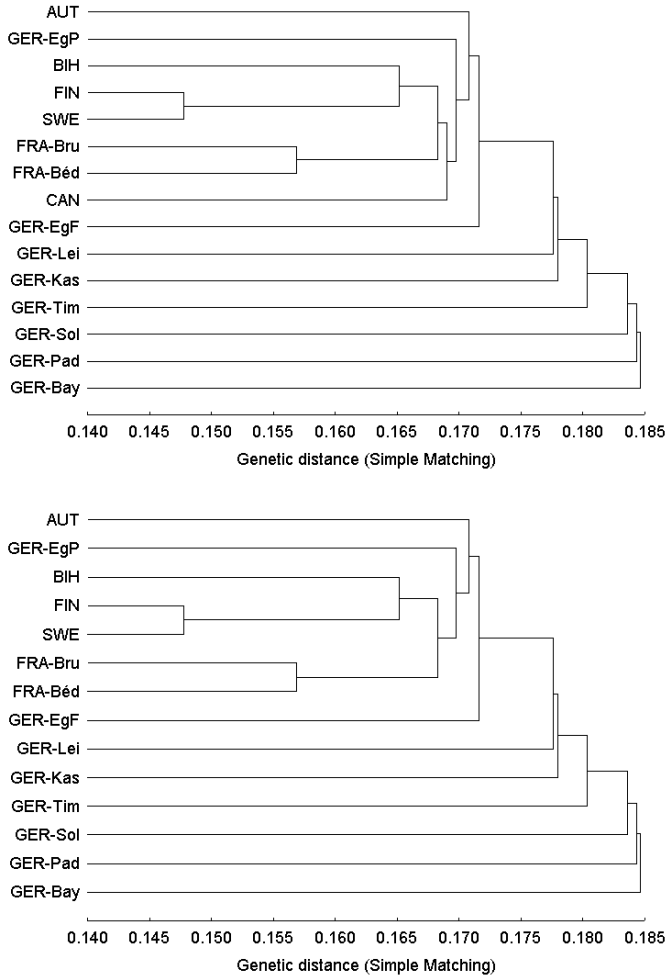

Figure B.18.: Single linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Simple Matching distance.

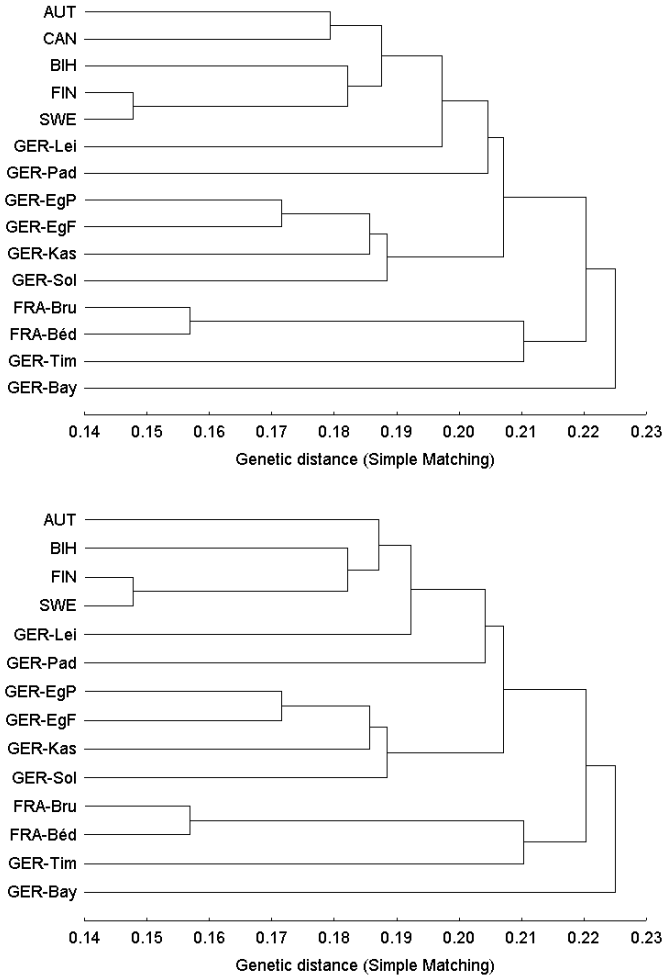

Figure B.19.: Complete linkage dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Simple Matching distance.

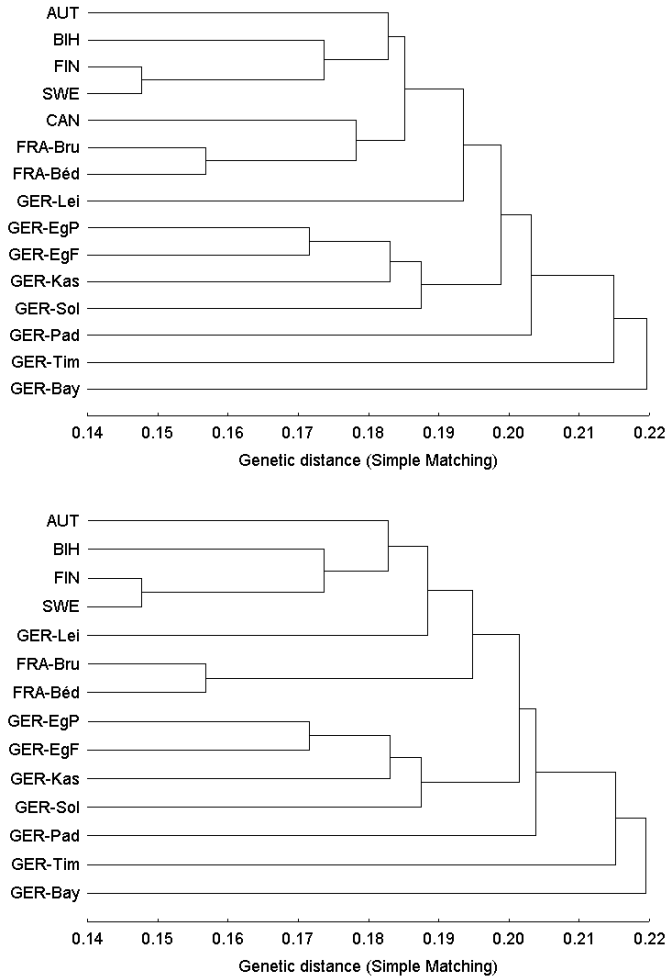

Figure B.20.: WPGMA dendrograms of pairwise mean genetic distances between 15 sampling locations (above) and 14 sampling locations (below, Canada excluded). Genetic distance measure: Simple Matching distance.

Appendix C.

AFLP marker

Table C.1.: AFLP loci and proportion of PRESENCE-phenotype across all samples and at single sampling locations.

Locus (bp)	Proportion of PRESENCE-phenotype in															
	all	FIN	SWE	GER-								FRA-		AUT	BIH	CAN
				Tim	Sol	EgF	EgP	Pad	Kas	Lei	Bay	Bru	Béd			
038.4 ±0.4	0.96	1.00	1.00	1.00	1.00	0.96	0.96	1.00	0.97	0.86	0.96	0.90	1.00	0.97	0.84	0.96
041.6 ±0.4	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.92	1.00	0.96	0.97	1.00	1.00	1.00	1.00
055.7 ±0.3	0.32	0.38	0.48	0.78	0.00	0.22	0.33	0.10	0.14	0.00	0.31	0.97	0.15	0.03	0.19	0.48
061.6 ±0.4	0.24	0.14	0.14	0.59	0.32	0.44	0.15	0.14	0.69	0.10	0.08	0.00	0.04	0.03	0.39	0.04
065.6 ±0.4	0.97	1.00	0.95	0.94	0.84	1.00	1.00	1.00	0.97	0.95	0.92	0.97	1.00	0.97	1.00	1.00
097.2 ±0.3	0.30	0.33	0.33	0.41	0.74	0.19	0.56	0.38	0.42	0.19	0.15	0.83	0.04	0.10	0.00	0.00
101.3 ±0.4	0.89	0.95	0.90	0.84	0.89	0.85	0.89	0.67	0.64	1.00	0.92	1.00	1.00	0.87	0.97	0.96
103.5 ±0.4	0.97	1.00	1.00	0.88	0.89	1.00	1.00	1.00	0.97	0.95	1.00	1.00	1.00	0.93	1.00	1.00
105.9 ±0.4	0.45	0.14	0.14	0.53	0.37	0.37	0.37	0.24	0.58	0.48	0.50	0.86	0.42	0.63	0.35	0.41
115.2 ±0.4	0.16	0.00	0.00	0.13	0.21	0.07	0.04	0.29	0.06	0.52	0.19	0.00	0.27	0.57	0.10	0.00
136.5 ±0.3	0.19	0.05	0.00	0.25	0.26	0.41	0.26	0.10	0.22	0.62	0.04	0.00	0.08	0.27	0.16	0.19
138.7 ±0.4	0.94	1.00	1.00	0.81	0.89	0.89	0.93	1.00	0.94	0.95	0.81	1.00	1.00	0.97	1.00	1.00
140.1 ±0.4	0.83	0.81	1.00	0.84	0.74	0.93	0.89	1.00	0.69	0.90	0.81	0.72	0.73	0.90	0.97	0.59
141.2 ±0.4	0.37	0.57	0.52	0.66	0.37	0.63	0.37	0.62	0.25	0.48	0.38	0.34	0.00	0.07	0.29	0.15
146.2 ±0.3	0.12	0.00	0.00	0.03	0.53	0.00	0.07	0.43	0.00	0.00	0.35	0.00	0.35	0.13	0.03	0.15
148.7 ±0.3	0.07	0.62	0.05	0.00	0.00	0.00	0.04	0.10	0.03	0.14	0.00	0.00	0.19	0.03	0.00	0.07
149.4 ±0.2	0.11	0.05	0.19	0.03	0.00	0.00	0.11	0.05	0.03	0.10	0.12	0.14	0.23	0.30	0.13	0.15
155.1 ±0.3	0.17	0.10	0.00	0.22	0.16	0.26	0.41	0.38	0.28	0.52	0.12	0.00	0.00	0.13	0.03	0.00
156.9 ±0.2	0.06	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.08	0.14	0.00	0.00	0.35	0.07	0.00	0.19
157.9 ±0.4	0.10	0.05	0.05	0.31	0.05	0.04	0.00	0.48	0.08	0.05	0.19	0.00	0.00	0.07	0.00	0.19
160.0 ±0.3	0.21	0.00	0.14	0.06	0.58	0.41	0.41	0.24	0.33	0.38	0.23	0.00	0.00	0.13	0.26	0.11
163.1 ±0.2	0.11	0.00	0.00	0.31	0.05	0.11	0.07	0.14	0.14	0.24	0.12	0.00	0.12	0.03	0.00	0.22
169.2 ±0.3	0.09	0.00	0.00	0.19	0.11	0.00	0.15	0.10	0.00	0.00	0.50	0.00	0.00	0.20	0.00	0.15
172.0 ±0.2	0.16	0.48	0.24	0.16	0.16	0.04	0.26	0.14	0.33	0.10	0.00	0.00	0.00	0.10	0.35	0.04
182.9 ±0.4	0.93	0.95	1.00	0.84	0.89	0.93	0.85	0.95	0.94	0.95	0.92	1.00	0.96	0.90	1.00	0.89
183.9 ±0.2	0.12	0.14	0.00	0.50	0.21	0.37	0.04	0.05	0.06	0.14	0.15	0.03	0.00	0.03	0.00	0.00
188.3 ±0.3	0.29	0.62	0.95	0.47	0.11	0.15	0.37	0.10	0.31	0.00	0.35	0.00	0.08	0.47	0.26	0.19
189.4 ±0.2	0.19	0.29	0.00	0.41	0.26	0.22	0.41	0.10	0.50	0.29	0.04	0.03	0.00	0.00	0.10	0.15
194.8 ±0.3	0.07	0.00	0.00	0.06	0.05	0.00	0.04	0.05	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.44
202.6 ±0.3	0.18	0.00	0.19	0.22	0.42	0.07	0.07	0.05	0.00	0.33	0.42	0.34	0.19	0.13	0.06	0.26
204.8 ±0.3	0.45	0.00	0.19	0.44	0.47	0.48	0.41	0.52	0.67	0.24	0.27	0.83	0.85	0.33	0.32	0.52

Table C.1.: continued

Locus (bp)	Proportion of PRESENCE-phenotype in															
	all	FIN	SWE	GER-								FRA-		AUT	BIH	CAN
				Tim	Sol	EgF	EgP	Pad	Kas	Lei	Bay	Bru	Béd			
209.9 ±0.4	0.89	0.95	0.90	0.63	0.89	0.93	0.96	0.76	0.89	0.95	0.81	1.00	1.00	0.80	0.97	0.96
212.7 ±0.2	0.16	0.14	0.00	0.00	0.42	0.22	0.22	0.10	0.03	0.05	0.27	0.31	0.27	0.10	0.00	0.41
215.2 ±0.3	0.27	0.57	0.00	0.75	0.47	0.33	0.30	0.43	0.22	0.43	0.00	0.03	0.00	0.27	0.19	0.07
217.9 ±0.4	0.49	0.43	0.00	0.28	0.58	0.52	0.67	0.52	0.44	0.76	0.69	0.45	0.31	0.70	0.48	0.52
218.9 ±0.3	0.10	0.00	0.00	0.28	0.00	0.00	0.15	0.19	0.00	0.19	0.00	0.03	0.04	0.00	0.55	0.04
223.2 ±0.4	0.86	0.95	1.00	0.88	0.89	0.96	0.85	0.81	0.83	0.71	0.62	0.93	0.88	0.87	0.77	1.00
227.1 ±0.4	0.20	0.57	0.10	0.06	0.05	0.11	0.15	0.24	0.19	0.00	0.12	0.00	0.12	0.47	0.35	0.37
230.9 ±0.4	0.08	0.00	0.00	0.34	0.00	0.00	0.04	0.14	0.03	0.24	0.12	0.10	0.00	0.00	0.13	0.00
237.7 ±0.2	0.08	0.00	0.00	0.06	0.00	0.04	0.00	0.00	0.00	0.00	0.23	0.31	0.00	0.33	0.00	0.11
238.8 ±0.4	0.15	0.05	0.05	0.06	0.00	0.00	0.04	0.14	0.08	0.24	0.00	0.59	0.12	0.23	0.35	0.22
241.8 ±0.3	0.04	0.00	0.00	0.03	0.16	0.00	0.00	0.05	0.00	0.00	0.08	0.00	0.12	0.10	0.00	0.15
244.2 ±0.2	0.28	0.48	0.00	0.19	0.37	0.56	0.52	0.29	0.42	0.29	0.15	0.00	0.15	0.37	0.19	0.26
246.8 ±0.2	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.10	0.04
248.0 ±0.4	0.09	0.10	0.05	0.13	0.00	0.00	0.00	0.14	0.00	0.00	0.04	0.69	0.04	0.00	0.00	0.15
254.6 ±0.4	0.17	0.14	0.24	0.00	0.05	0.22	0.07	0.24	0.00	0.43	0.00	0.93	0.31	0.00	0.03	0.00
256.6 ±0.2	0.06	0.05	0.00	0.03	0.00	0.07	0.22	0.38	0.00	0.00	0.08	0.00	0.00	0.13	0.00	0.00
258.4 ±0.3	0.05	0.00	0.00	0.00	0.05	0.26	0.15	0.00	0.00	0.05	0.00	0.07	0.12	0.03	0.00	0.07
261.9 ±0.3	0.24	0.00	0.00	0.03	0.58	0.67	0.56	0.48	0.72	0.10	0.23	0.00	0.04	0.03	0.06	0.00
263.2 ±0.4	0.12	0.05	0.14	0.00	0.05	0.00	0.11	0.24	0.11	0.33	0.04	0.03	0.54	0.13	0.13	0.00
265.2 ±0.2	0.03	0.05	0.00	0.00	0.26	0.00	0.00	0.10	0.00	0.00	0.08	0.00	0.04	0.00	0.00	0.04
268.2 ±0.4	0.31	0.76	0.67	0.66	0.00	0.04	0.26	0.24	0.14	0.81	0.62	0.14	0.23	0.03	0.26	0.00
270.7 ±0.2	0.10	0.00	0.00	0.47	0.00	0.00	0.04	0.05	0.00	0.00	0.12	0.31	0.19	0.03	0.00	0.19
274.4 ±0.3	0.15	0.10	0.00	0.06	0.05	0.00	0.00	0.05	0.03	0.00	0.42	0.79	0.62	0.00	0.00	0.11
276.6 ±0.4	0.15	0.00	0.00	0.03	0.58	0.04	0.00	0.10	0.17	0.05	0.38	0.10	0.27	0.13	0.26	0.15
280.8 ±0.2	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.42	0.00	0.12	0.20	0.03	0.33
286.5 ±0.4	0.31	0.19	0.52	0.16	0.58	0.59	0.33	0.29	0.94	0.05	0.19	0.14	0.00	0.23	0.06	0.26
292.5 ±0.3	0.08	0.00	0.00	0.00	0.00	0.11	0.00	0.05	0.19	0.43	0.38	0.00	0.00	0.07	0.00	0.04
296.1 ±0.4	0.33	0.57	0.19	0.16	0.47	0.44	0.44	0.10	0.25	0.86	0.31	0.00	0.04	0.47	0.42	0.41
297.9 ±0.2	0.02	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.03	0.05	0.00	0.00	0.00	0.00	0.00	0.00
300.2 ±0.3	0.12	0.10	0.10	0.06	0.00	0.00	0.04	0.14	0.00	0.00	0.04	0.72	0.35	0.07	0.06	0.15
301.5 ±0.3	0.15	0.19	0.14	0.47	0.00	0.04	0.00	0.33	0.08	0.05	0.04	0.24	0.31	0.03	0.00	0.26
303.1 ±0.2	0.04	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.10	0.00	0.10	0.13	0.04

Table C.1.: continued

Locus (bp)	Proportion of PRESENCE-phenotype in															
	all	FIN	SWE	GER-								FRA-		AUT	BIH	CAN
				Tim	Sol	EgF	EgP	Pad	Kas	Lei	Bay	Bru	Béd			
305.2 ±0.4	0.08	0.00	0.00	0.09	0.00	0.04	0.04	0.00	0.06	0.00	0.08	0.21	0.35	0.03	0.00	0.22
314.5 ±0.4	0.13	0.10	0.10	0.09	0.05	0.07	0.04	0.10	0.03	0.00	0.42	0.21	0.31	0.20	0.03	0.15
316.3 ±0.3	0.04	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.00	0.15	0.00	0.00	0.07
321.0 ±0.4	0.15	0.00	0.00	0.00	0.42	0.48	0.26	0.14	0.67	0.00	0.00	0.00	0.04	0.10	0.00	0.04
328.9 ±0.2	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.31	0.00	0.00	0.00	0.00
337.7 ±0.4	0.11	0.00	0.00	0.41	0.26	0.00	0.11	0.00	0.31	0.05	0.23	0.00	0.00	0.07	0.03	0.11
339.9 ±0.4	0.15	0.00	0.05	0.00	0.00	0.19	0.07	0.33	0.00	0.33	0.00	0.03	0.04	0.47	0.19	0.56
343.2 ±0.2	0.09	0.00	0.00	0.13	0.11	0.00	0.22	0.19	0.00	0.00	0.00	0.00	0.04	0.30	0.10	0.19
345.3 ±0.3	0.09	0.00	0.00	0.09	0.05	0.22	0.11	0.05	0.06	0.14	0.27	0.00	0.15	0.00	0.16	0.04
347.0 ±0.2	0.07	0.14	0.00	0.00	0.16	0.15	0.04	0.05	0.00	0.10	0.00	0.00	0.00	0.00	0.39	0.00
351.6 ±0.3	0.27	0.43	0.33	0.09	0.32	0.07	0.15	0.57	0.64	0.05	0.27	0.10	0.00	0.60	0.16	0.26
353.3 ±0.4	0.12	0.00	0.57	0.00	0.16	0.04	0.26	0.24	0.25	0.00	0.15	0.00	0.00	0.10	0.10	0.07
354.7 ±0.2	0.08	0.05	0.00	0.13	0.05	0.00	0.00	0.10	0.06	0.00	0.23	0.00	0.04	0.07	0.32	0.04
355.7 ±0.4	0.36	0.52	0.14	0.69	0.37	0.59	0.48	0.43	0.22	0.29	0.62	0.17	0.00	0.43	0.19	0.30
358.4 ±0.4	0.17	0.19	0.29	0.19	0.00	0.33	0.26	0.00	0.17	0.00	0.04	0.00	0.00	0.13	0.52	0.26
360.6 ±0.2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.03	0.05	0.00	0.03	0.00	0.00	0.00	0.00
363.8 ±0.4	0.19	0.19	0.00	0.13	0.11	0.33	0.33	0.33	0.17	0.19	0.08	0.28	0.27	0.23	0.00	0.19
365.1 ±0.2	0.14	0.00	0.05	0.00	0.21	0.11	0.33	0.67	0.17	0.14	0.04	0.03	0.12	0.00	0.29	0.00
376.3 ±0.3	0.07	0.05	0.00	0.06	0.05	0.11	0.11	0.10	0.14	0.00	0.27	0.00	0.04	0.00	0.00	0.11
380.1 ±0.3	0.11	0.00	0.10	0.06	0.11	0.26	0.33	0.00	0.17	0.05	0.04	0.03	0.00	0.33	0.00	0.11
383.2 ±0.4	0.02	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.03	0.00	0.08	0.00	0.00	0.07	0.00	0.07
384.3 ±0.2	0.02	0.00	0.00	0.16	0.05	0.00	0.00	0.00	0.03	0.05	0.00	0.00	0.04	0.00	0.00	0.00
386.1 ±0.4	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.19	0.19	0.55	0.15	0.13	0.13	0.00
387.4 ±0.2	0.04	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.05	0.00	0.00	0.00	0.00	0.00	0.00
390.2 ±0.2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.04	0.00	0.00	0.00	0.00	0.00
392.9 ±0.2	0.01	0.05	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
395.2 ±0.4	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.46	0.23	0.00	0.00
401.5 ±0.4	0.21	0.24	0.48	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.58	0.00	0.50	0.63	0.10	0.07
403.0 ±0.2	0.04	0.05	0.33	0.00	0.05	0.04	0.00	0.05	0.03	0.00	0.04	0.03	0.04	0.00	0.03	0.00
406.0 ±0.4	0.10	0.05	0.05	0.22	0.16	0.07	0.00	0.00	0.31	0.10	0.00	0.00	0.00	0.03	0.35	0.04
407.8 ±0.3	0.10	0.05	0.00	0.00	0.26	0.33	0.33	0.05	0.11	0.00	0.23	0.00	0.04	0.00	0.13	0.04
408.9 ±0.4	0.15	0.24	0.57	0.13	0.42	0.00	0.04	0.24	0.00	0.24	0.27	0.00	0.04	0.00	0.13	0.30

Table C.1.: continued

Locus (bp)	Proportion of PRESENCE-phenotype in															
	all	FIN	SWE	GER-								FRA-		AUT	BIH	CAN
				Tim	Sol	EgF	EgP	Pad	Kas	Lei	Bay	Bru	Béd			
410.0 ±0.2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.04
410.8 ±0.2	0.03	0.00	0.00	0.00	0.00	0.15	0.07	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
413.0 ±0.3	0.09	0.10	0.05	0.03	0.26	0.22	0.22	0.05	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00
414.7 ±0.2	0.08	0.52	0.14	0.00	0.11	0.00	0.04	0.24	0.00	0.05	0.15	0.00	0.00	0.00	0.16	0.00
418.0 ±0.2	0.01	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.06	0.05	0.00	0.00	0.00	0.00	0.00	0.00
422.8 ±0.2	0.22	0.43	0.43	0.00	0.26	0.37	0.56	0.10	0.03	0.33	0.38	0.00	0.00	0.27	0.35	0.04
425.5 ±0.4	0.15	0.00	0.00	0.03	0.05	0.22	0.15	0.00	0.03	0.00	0.31	0.03	0.19	0.27	0.29	0.52
431.8 ±0.2	0.47	0.62	0.52	0.63	0.47	0.85	0.81	0.52	0.42	0.71	0.27	0.00	0.35	0.47	0.23	0.37
438.8 ±0.4	0.05	0.00	0.05	0.00	0.00	0.15	0.07	0.10	0.11	0.00	0.00	0.00	0.00	0.07	0.10	0.00
439.8 ±0.3	0.06	0.05	0.00	0.00	0.00	0.07	0.04	0.10	0.14	0.38	0.00	0.00	0.00	0.00	0.16	0.00
441.4 ±0.3	0.02	0.05	0.00	0.06	0.00	0.00	0.00	0.10	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00
442.4 ±0.3	0.01	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
444.1 ±0.2	0.07	0.00	0.00	0.19	0.00	0.19	0.07	0.29	0.03	0.33	0.00	0.00	0.08	0.00	0.00	0.00
445.6 ±0.3	0.06	0.05	0.05	0.00	0.21	0.04	0.00	0.00	0.00	0.05	0.00	0.03	0.19	0.03	0.10	0.19
446.5 ±0.3	0.02	0.00	0.00	0.00	0.05	0.00	0.00	0.05	0.06	0.00	0.00	0.14	0.00	0.00	0.00	0.04
448.8 ±0.3	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00
450.2 ±0.2	0.02	0.00	0.00	0.00	0.26	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
451.2 ±0.3	0.04	0.05	0.29	0.00	0.05	0.04	0.00	0.05	0.11	0.00	0.00	0.00	0.00	0.03	0.00	0.07
453.2 ±0.3	0.02	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00	0.00	0.00
455.5 ±0.2	0.04	0.00	0.00	0.13	0.16	0.00	0.00	0.05	0.03	0.05	0.15	0.00	0.00	0.07	0.00	0.00
461.7 ±0.3	0.03	0.00	0.00	0.00	0.00	0.07	0.07	0.00	0.00	0.05	0.08	0.00	0.08	0.00	0.13	0.00
463.8 ±0.3	0.09	0.14	0.00	0.34	0.00	0.04	0.11	0.05	0.14	0.00	0.19	0.00	0.08	0.07	0.10	0.00
467.0 ±0.4	0.09	0.00	0.05	0.06	0.05	0.00	0.07	0.24	0.03	0.05	0.12	0.00	0.12	0.07	0.42	0.00
469.9 ±0.4	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.03	0.74	0.00
473.2 ±0.2	0.10	0.19	0.00	0.00	0.00	0.15	0.00	0.05	0.11	0.10	0.00	0.34	0.38	0.03	0.00	0.11
474.0 ±0.2	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.00	0.04	0.03	0.00	0.00
475.7 ±0.2	0.02	0.00	0.00	0.00	0.00	0.07	0.07	0.05	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.07
484.3 ±0.4	0.13	0.48	0.33	0.25	0.11	0.15	0.04	0.24	0.03	0.33	0.00	0.00	0.00	0.13	0.13	0.00
491.1 ±0.2	0.07	0.05	0.00	0.28	0.00	0.04	0.04	0.05	0.00	0.19	0.00	0.00	0.00	0.23	0.03	0.04
498.4 ±0.2	0.04	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.03	0.00	0.04	0.10	0.42	0.00	0.00	0.00

Acknowledgements

Many people expressed their interest in this project, made helpful comments at meetings or stimulated my critical thinking with their questions during presentations. If the reader belongs to this group of supporters or disputants who are not mentioned by name below, please be assured that your contributions have not been forgotten and that I appreciate everything you've said or done!

Many thanks to **Dr. Martin Potthoff** for raising this project and passing it over to me. Thank you for struggling with all the proposals, applications and the troubles during the handover of the project. You were always a friendly, helpful and relaxed 'research director' and your pragmatism was always well-timed.

I thank...
Prof. Dr. Bernard Ludwig for supervising this project, reviewing this work and being examiner of my disputation, and also **Prof. Dr. Kurt Weising** for his prompt willingness to be examiner of my disputation.

Prof. Dr. Reiner Finkeldey for working through my dissertation and giving many helpful comments, and for being a kind examiner of my disputation. **Dr. Oliver Gailing** for everything in the lab,

for the comments regarding my dissertation and for being examiner of my disputation. **Everybody at the Department of Forest Genetics in Göttingen** for beeing very helpful and patient with me and the exotic project all the time. Thanks for all the contributions to my presentations in the lab meetings and beyond, and for being partners in the lab and not rivals for getting a PCR machine. Many thanks to **Dr. Elizabeth M. Gillet** for the helpful computer programs and the rapid support with them.

Prof. Dr. Friedrich Beese for the office and the support during the course of this project. Many thanks to **Dr. Jürgen Prenzel** for everything regarding the computers. To **all the nice people from the (former) ibw**: Thanks for your interest in the project, the department's trips and the Christmas celebrations.

Prof. Dr. Rainer Georg Jörgensen for the opportunity to work at the 'Bodenbiologie' in Witzenhausen. Dear **'Bobis'**: During my scarce visits and activities in Witzenhausen I always immediately felt like being at 'my' workgroup. Thanks for being such nice colleagues! Special thanks to **Susanne Beck** for doing all the administrative stuff.

Dr. Guénola Peres, **Dr. Visa Nuutinen** and **Dr. Kevin Butt** for your encouraging words, earthworm deliveries and literature hints.

Tamara Coja, **her familiy** and **Snezana Trifunovic** for the supernatural efforts getting the worms from the Balkans.

Eva Eichinger, **Dr. Stefanie Krück**, **Dr. Alexander Bruckner**, **Dr. Werner Borken**, **Dr. Jan Lagerlöf**, **Christian Löfke**, **Benjamin Stein** and **René Trujillo** for ideas, earthworm sampling, earthworm feeding, lab work and many other things keeping this project running.

The **DFG** for financing parts of this project.

To express some more personal feelings to some people, a few words in German:
Vielen Dank an **die Fechter, die G7 und alle anderen Sportler des Hochschulsports Göttingen** für die erfolgreiche Ablenkung von der Arbeit und die Einsicht, dass der 45ste Treffer, ein gut getimtes Folgeangebot oder eine Vorhand mit perfekter Länge genauso viel Wert sind wie ein vollendeter Absatz der Doktorarbeit — manchmal sogar mehr.

Vielen Dank an **Anja Bergstermann**, **Elli Mertes** und **Daniel Fröhlich** für das richtige Maß an 'Unernsthaftigkeit' ab und an (danke für EM und DIES 2008).

Danke an **meine Eltern** für die Unterstützung, wenn es nötig war. Danke, dass Ihr mich nicht vergessen habt, auch wenn es mit einem Besuch in Kassel mal wieder länger dauerte.

Zusammenfassung

Genetische Struktur in europäischen Populationen des Regenwurms *Lumbricus terrestris*

Die genetische Diversität und genetische Differenzierung des anözischen Regenwurms *Lumbricus terrestris* wurde an 14 Standorten in Europa und an Regenwürmern, die in der Vergangenheit durch europäische Siedler in Kanada eingeführt worden sind, untersucht. Insgesamt wurden 394 *L. terrestris* untersucht.

Die AFLP Methode (Amplified Fragment Length Polymorphism) wurde das erste Mal erfolgreich bei einer Regenwurmart mit sexueller Fortpflanzung eingesetzt. Basierend auf der Reproduzierbarkeit von AFLP Banden wurden 125 polymorphe AFLP Marker genutzt, um genetische Diversität und genetische Differenzierung mit unterschiedlichen Methoden zu berechnen.

Es gab kein einheitliches Muster der genetischen Diversität. Niedrige Werte in Populationen aus Frankreich und Schweden könnten auf den spezifischen Ursprung der Populationen oder geographische Besonderheiten der Untersuchungsstandorte zurückzuführen sein.

Mit dem Monmonier Algorithmus, der Wombling Methode und Clusteranalysen, basierend auf mehreren genetischen Distanzmaßen,

wurden genetische Barrieren im europäischen Untersuchungsgebiet identifiziert, also Regionen mit einer starken Heterogenität der genetischen Information. Diese genetische Barrieren machten eine Definition von Regionen in Europa möglich, die möglicherweise von jeweils identischen Refugialpopulationen nach Ende der letzten Eiszeit in Europa vor etwa 14,000 Jahren besiedelt worden waren. Durch AMOVA (Analysis of Molecular Variance) wurde der Anteil an genetischer Varianz zwischen den Regionen quantifiziert. Dieser war gering, aber signifikant.

Offensichtlich gibt es eine Verbindung zwischen Populationen im Balkan und in Nordeuropa (Finnland und Schweden), während Populationen in Frankreich und möglicherweise Italien stärker vom restlichen Europa isoliert sind. Es existieren starke Hinweise darauf, dass *L. terrestris* auf Ackerbau angewiesen ist, um sich auszubreiten und neue Populationen aufzubauen. Das Ausbreitungsmuster des Ackerbaus während der Neolithischen Revolution vor etwa 9000 Jahren hilft die heutigen Muster in der genetischen Struktur von *L. terrestris* Populationen in Europa zu erklären.

AFLPs und andere molekulare Methoden sind hilfreich, um Ursprung, Verbreitung, genetische Diversität und Populationsdynamik von Bodenorganismen zu erklären, die häufig nur schwer zu untersuchen sind.