Molecular phylogeography of the Woodland Ringlet (*Erebia medusa* [Denis and Schiffermüller] 1775) in Europe

Dissertation zur Erlangung des Grades
Doktor der Naturwissenchaften

am Fachbereich Biologie
der Johannes Gutenberg-Universität in Mainz

Nasera Hammouti

geb. am 29. 12. 1978 in Tarbes, France

Mainz, 2006

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: .....

Contents

1. GENERAL INTRODUCTION 1

2. PHYLOGEOGRAPHY OF THE WOODLAND RINGLET, *EREBIA MEDUSA* BASED ON SEQUENCES OF THE MITOCHONDRIAL COI GENE 5

2.1. Introduction 5

2.2. Material and methods 7
2.2.1. Sampling design 7
2.2.2. Sequencing of mitochondrial DNA 7
2.2.3. Nested clade phylogeographic analysis 8

2.3. Results 10

2.4. Discussion 13

2.5. Summary 17

3. ADJUSTMENT OF RING-SHAPED AMBIGUITIES IN MINIMUM SPANNING NETWORKS FOR RECOMBINATION AND HOMOPLASY: ITS IMPACT ON PHYLOGEOGRAPHIC RECONSTRUCTION 19

3.1. Introduction 19

3.2. Material and methods 21
3.2.1. Haplotype network construction and ring-shaped ambiguities 21
3.2.2. Investigation of recombination and homoplasy 23
a- Resolving ring ambiguities under recombination 23
b- Detection of homoplasy 24
3.2.3. Nested clade phylogeographic analysis 24

3.3. Results 25
3.3.1. Recombination 25
3.3.2. Homoplasy 31

3.4. Discussion 34
3.4.1. The allozyme scenario 34
3.4.2. Adjustment for recombination (scenarios A-F) 36
3.4.3. Adjustment for homoplasy (scenarios G-H) 37
3.4.4. Comparison among adjustments 37
3.4.5. Conclusions 38

3.5. Summary 39

4. PHYLOGEOGRAPHY OF THE WOODLAND RINGLET (*EREBIA MEDUSA*) BASED ON THE HIGHLY VARIABLE CONTROL REGION 40

4.1. Introduction 40

4.2. Material and methods 42
4.2.1. Sequencing of mtDNA 42
4.2.2. Nested clade phylogeographic analysis 43
4.2.3. Solving haplotype networks with ring-shaped ambiguities 44

4.3. Results 45
4.3.1. Geographical distribution of haplotypes 45
4.3.2. Minimum spanning tree 47
a- Identical treatment of indels and substitution 47
b- Differential treatment of indels and substitution 50

4.4. Discussion 55

4.5. Summary 59

5. GENERAL CONCLUSION 60

6. ABSTRACT 63

7. REFERENCES 65

8. APPENDIX 72

9. ACKNOWLEGMENTS ..... FEHLER! TEXTMARKE NICHT DEFINIERT.

1. General introduction

Population structure, defined as the distribution of genotypes in time and space, results from present processes and past history (Hewitt and Butlin 1997). During the Quaternary (1.6 million years), climatic fluctuations are considered as a major historical process influencing the genetic diversity of natural populations of the temperate Northern Hemisphere (Hewitt 1996, 2004). The Croll-Milankovitch theory proposes that these climatic fluctuations are due to several forces such as excentricity, precession, axial tilt and obliquity that together produce the Milankovitch oscillations. Hence regular variations in the earth's orbit around the sun led to a modification of the insolation of the earth which received more energy, transported by the oceanic circulation system. Thus the interaction of orbital variation and currents led to climate changes (Williams et al. 1998). One consequence of the decrease of temperature on the earth is the formation of large ice caps and ice sheets during the cold periods (glacial) which partially melted during the warmer periods (interglacial).

Alternation of glacial and interglacial stages constitutes the ice ages. These climate fluctuations are particularly supported by analyses of carbon and oxygen isotopes, pollen profiles, and animal and plant remains contained in the ice sheet. Four major glaciations occurred during the Quaternary and are known as Günz, Mindel, Riss, and Würm, from the older to the more recent ones (Andersen and Borns 1997). In Europe, the Last Glacial Maximum (LGM) occurred around 18,000 years before present and induced the formation of (i) a large ice sheet covering parts of Britain and northern Europe, and (ii) of ice caps on the top of major mountain ranges such as the Pyrenees, the Alps, and the Caucasus (Frenzel 1973, Nilsson 1983). At the edges of the ice sheets, cold steppes (tundra) covered Europe (Tzedakis et al. 2002).

The severe climatic conditions strongly modified the distribution of animals and plants. They went through successive cycles of range contractions and range expansions. Suitable localities, where temperate fauna and flora could persist during the cold periods, are defined as "glacial refugia". In Europe, the southern peninsulas of Iberia, Apennine, and Balkans constitute the main glacial refugia (Hewitt 1996). Thus, during isolation among refugia, many taxa evolved into different genetic lineages. Climate warming, at the end of each glacial stage, enabled northwards expansion of species ranges out of the refugia (Taberlet et al. 1998). This shaped the genetic structure of populations. Predictions considering the different model of mode of dispersion (leptokurtic, stepping stone, and normal dispersal) assume that rapid continued expansion resulted in an erosion of genetic diversity (founder effect) whereas in

slower expanding populations much more genetic diversity is maintained. Expanding populations contain only a subset of the original gene pool, localised and persisting in the refugia (Hewitt 1999).

In Europe, at least three major typical patterns of genetic variability emerged from postglacial expansion of temperate biota (Hewitt 1999): (i) the 'grasshoper pattern' with postglacial expansion only from the Balkans; the Iberian and Apennine lineages blocked by the Pyrenees and the Alps, (ii) the 'bear pattern' with expansions from south-eastern Europe and Iberia, (iii) the 'hedgehog pattern' showing colonisation of Northern Europe from the three Mediterranean glacial refugia. Recently, it was described a fourth pattern in which only the Iberian lineage not considerably contribute to the postglacial colonisation of Central and Northern Europe (Marbled White butterfly; Habel et al. 2005). After postglacial range expansion, the different genetic lineages met, and these meeting-areas are termed hybrid or suture zones (Hewitt 1996, 1999, 2000, 2001, Taberlet et al. 1998). Considering these concepts, the study of the actual genetic structure at a large geographical scale will allow for the inference of the colonisation routes after the last and glaciation (Würm).

Phylogeography is defined as the „field of study concerned with the principles and processes governing the geographic distributions of genealogical lineages, especially those within and among closely related species" (Avise 2000). It mainly addresses questions of intra-specific relations. Phylogeography has expanded rapidly during the last two decades and is now a fully recognised field of biological research that links phylogenetics to biogeography. It is an integrative discipline based on knowledges from molecular genetics, population genetics, phylogenetics, demography, ethology, and historical geography. The understanding of patterns of population structure enables to analyse other aspects of the biology of an organism in a meaningful context.

Molecular techniques such as PCR, coupled with the development of population genetic concept such as coalescent theory allows the identification of genetic lineages and relevant refugia to infer putative routes of expansion. The Quaternary ice ages are relatively recent events at the geological scale. To detect their effect on lineage divergence, fast evolving markers are necessary. Considering the whole animal genome, one of the fastest evolving regions is the mitochondrial DNA (mtDNA) (2% per Myr in higher primates; Brown et al. 1979). It is a circular genome, maternally transmitted without recombination in most species (Moritz et al. 1987). It contains 24 genes encoding for the translational machinery of the mtDNA itself (two ribosomal RNAs and 22 transfer RNAs) and 13 genes encoding for the subunits of the mitochondrial respiratory chain. The different regions of the mtDNA evolve at

different rates. The cytochrome oxydase subunit one (COI) is one of its most slowly evolving genes compared to the control region, which is the fastest evolving partition of the mitochondrial genome. Therefore, and due to its overall high and regionally different mutation rate, the mtDNA constitutes a powerful molecule for phylogeographic studies of animals allowing to discriminate evolutionary histories of the species and populations through their molecular differentiation.

In contrast to other taxa, only few phylogeographic studies on butterflies are available at a European scale. To best of my knowledge, phylogeographical patterns are described only for: *Aglais urticae* (Vandewoestijne et al. 2003), *Euphydryas aurinia* (Joyce and Pullin 2001), the *Erebia tyndarus* group (Martin et al. 2002), *Erebia triaria*, and *Erebia palarica* (Vila 2004). Molecular biogeographical scenarios for European butterfly species were up to now mainly inferred from allozyme data: *Melanargia galathea* and *M. lachesis* (Satyrinae) (Habel et al. 2005); *Polyommatus icarus* (Schmitt et al. 2003), *Polyommatus coridon* species group (Schmitt and Seitz 2001b, 2002; Schmitt et al. 2002, Schmitt and Krauss 2004), *Maculinea alcon* species group (Bereczki et al. 2005), *Aricia agestis-artaxerxes* complex (Aagaard et al. 2002) (Lycaenidae); *Maniola jurtina* (Schmitt et al. 2005), *Pieris napi* (Porter and Geiger 1995), *Coenonympha hero* (Cassel and Tammaru 2003), *Erebia medusa* (Schmitt 1999, Schmitt and Seitz 2001a) (Nymphalidae). The latter study analysed the genetic pattern of the Woodland Ringlet *E. medusa* ([Denis and Schiffermüller] 1775), a Siberian faunal element (de Lattin 1957, Varga 1977). The expected pattern for this species should be a continuous loss of genetic diversity during its postglacial westwards expansion (founder effect). However, the nuclear data revealed the existence of four major lineages evolving during the past 70,000 years. This particular genetic structure suggests the existence of extra-Mediterranean glacial refugia for this species in Europe. This assumption only relies on nuclear data (allozymes), and a combination of different genetic markers, including maternally inherited ones, should prove the consistency of this biogeographic scenario. Since the allozyme system addresses the variability at the protein level it could be subject to selection (Eanes 1999). In contrast, the circular mt DNA is assumed to evolve selectively neutral.

The Woodland Ringlet belongs to a species-rich genus with Holoarctic distribution. Numerous species occupy alpine and/or arctic habitats. The Palaearctic *E. medusa* is currently distributed from central France and south eastern Belgium over large parts of Central Europe and southern Siberia to the Pacific. It is absent from the Iberian Peninsula, Great Britain, from an area of the North Sea and from Scandinavia (Kudrna 2002, Korschunov and Gorbunov

1995). The ecology and biology of *E. medusa* are well studied (Ebert and Rennwald 1991, Schmitt 1993, 2002). The species is typical for meadows poor in nitrogen and for fallow land. It is a grass feeder in its larval stage, univoltine and, depending on altitude, active between mid May and the end of July.

In my thesis I study the phylogeography of the Woodland Ringlet through the analysis of mitochondrial genes. My aim is to reconstruct a consistent evolutionary history on the basis of a combination of published nuclear and new mitochondrial data. Fractions of two differentially evolving genes, namely the protein coding COI gene and the control region, are used to establish a concise phylogeographical history for this butterfly species over large parts of Europe.

My thesis is composed of three major chapters (chapters 3-5), which can be read independently. Each chapter contains an introduction, description of methods, results and a discussion section; it ends with a short summary. Chapter 5, resumes my general conclusions.

Chapter 2 deals with the phylogeography of the Woodland Ringlet based on sequences of COI gene. I performe nested clade phylogeographic analysis (NCPA: Templeton 1995) to infer an evolutionary scenario considering the genetic pattern from COI. In combination with allozyme data this allows me to reconstruct an improved phylogeographic scenario for the Woodland Ringlet in Europe. Chapter 3 emphasizes the disturbances introduced into phylogeographic reconstruction through recombination and homoplastic base substitutions[1]. Since allozyme data (Schmitt and Seitz 2001a) estimated a Late Pleistocene final genetic structuring for the Woodland Ringlet in Europe, chapter 4 reconstructs the phylogeographical history of the Woodland Ringlet using the fast evolving mitochondrial control region.

[1] Chapter 2 and 3 are in a similar form submitted respectively as Hammouti et al. submitted a and b.

2. Phylogeography of the Woodland Ringlet, *Erebia medusa* based on sequences of the mitochondrial COI gene

2.1. Introduction

Climatic fluctuations during the Quaternary are the major historical processes influencing the genetic diversity of natural populations of the Northern Hemisphere. During the late Pleistocene the increasingly strong climatic fluctuations between cold and warm periods considerably influenced the distribution of most animal and plant species (Hewitt 1996, 2004). In general, two contrasting groups of species can be distinguished in the Palaearctic: (i) species with their refugial phases occurring during the warm interglacial periods with range expansion during the cold phases (i.e., arctic and/or alpine elements) and (ii) species with refugial periods during the ice ages and range expansions during the warm interglacials (i.e., Mediterranean or Siberian elements sensu de Lattin 1967) (Hewitt 2004). Hence, these severe climatic oscillations induced range contractions and expansions of fauna and flora that are expected to have left signatures in the geographical distribution and genetic diversity of extant populations (Avise 2000). Advancing glaciers and permafrost regions destroyed habitats of temperate species, eradicated many of their populations and restricted access to refugia. During periods of retreating ice sheets, populations of these species recolonised more northern latitudes.

New molecular methods enable the investigation of intraspecific population structures through the deduction of geographical genetic variation. The evolving field of phylogeography has rapidly expanded during the last two decades, and is now a fully recognised area of biological research that links phylogenetic reconstruction to biogeography. Avise (2000) defined phylogeography as the "principles and processes governing the geographic distributions of genealogical lineages, especially within and among closely related species". These phylogeographic analyses provide increasing support of the classical Mediterranean differentiation centres (Taberlet et al. 1998, Hewitt 1999, 2000, 2004). However, the hypothesis of Siberian faunal elements expanding throughout Eurasia during the postglacial receives little support from recent scientific studies. No major genetic differentiation should be found in Europe for Siberian elements, and the only likely scenario for these species is the continuous loss of genetic diversity during the postglacial westward expansion (founder effect). In contrast, there is increasing support for survival of such temperate taxa in extra-Mediterranean differentiation centres in Europe, both from fossil

remains (Coope 1970, 1978, 1994, Hertelendy et al. 1992, Füköh et al. 1995, Sümegi and Rudner 2001, Sümegi and Krolopp 2002, Pazonyi 2004, Willis and van Andel 2004) and, to a lesser degree, from genetic analyses (Napolitano and Descimon 1994, Meglécz et al. 1997, Nesbo et al. 1999, Englbrecht et al. 2000, Rafinski and Babik 2000, Schmitt and Seitz 2001a, b, Babik et al. 2004, Hewitt 2004). However, further genetic support of the evolutionary importance of these extra-Mediterranean refugia is needed.

One example for these former "Siberian" elements is the Woodland Ringlet *E. medusa* (de Lattin 1957, Varga 1977). The Woodland Ringlet typically lives in different types of grasslands (SBN 1987, Ebert and Rennwald 1991, Schmitt 2002) and is distributed throughout temperate Eurasia, but is missing in the euatlantic and eumediterranean regions of Europe and in Scandinavia (Tolman and Lewington 1997, Kudrna 2002). Using allozyme data, Schmitt and Seitz (2001a) deduced an evolutionary scenario of at least four major lineages with several subgroups evolving in Europe during the past 70,000 years. This and therefore also strongly supports extra-Mediterranean glacial differentiation centres for this species in Europe. I used mitochondrial sequence data to examine the phylogeography of this Palaearctic butterfly species in Europe.

Such a high complexity of potential historical and recurrent processes in space and time is well addressed by the nested clade phylogeographic analysis (NCPA) as proposed by Templeton et al. (1995). It considers and distinguishes among contemporary (e.g., restricted gene flow) and historical (e.g., past fragmentation, range expansion or colonization) processes. NCPA analysis has become a popular tool for phylogeographic studies since it detects, without *a priori* assumptions, non-random geographical haplotype associations and proposes the best phylogeographical scenario. I superimposed the results of an NCPA analysis of part of the cytochrome oxydase subunit one (COI) mitochondrial gene data to the conclusions based on previous allozyme data (Schmitt and Seitz 2001a). I then infer an improved scenario of the evolutionary history of *E. medusa* in Europe, based on evidence from both nuclear and mitochondrial data.

2.2. Material and methods

2.2.1. Sampling design

Butterflies from 32 populations were netted across Europe from 1996 to 1998 (Schmitt and Seitz 2001a). Nine additional populations were sampled in 2002 (Figure 2.1 and Appendix 1). Samples were stored in liquid nitrogen immediately upon capture.

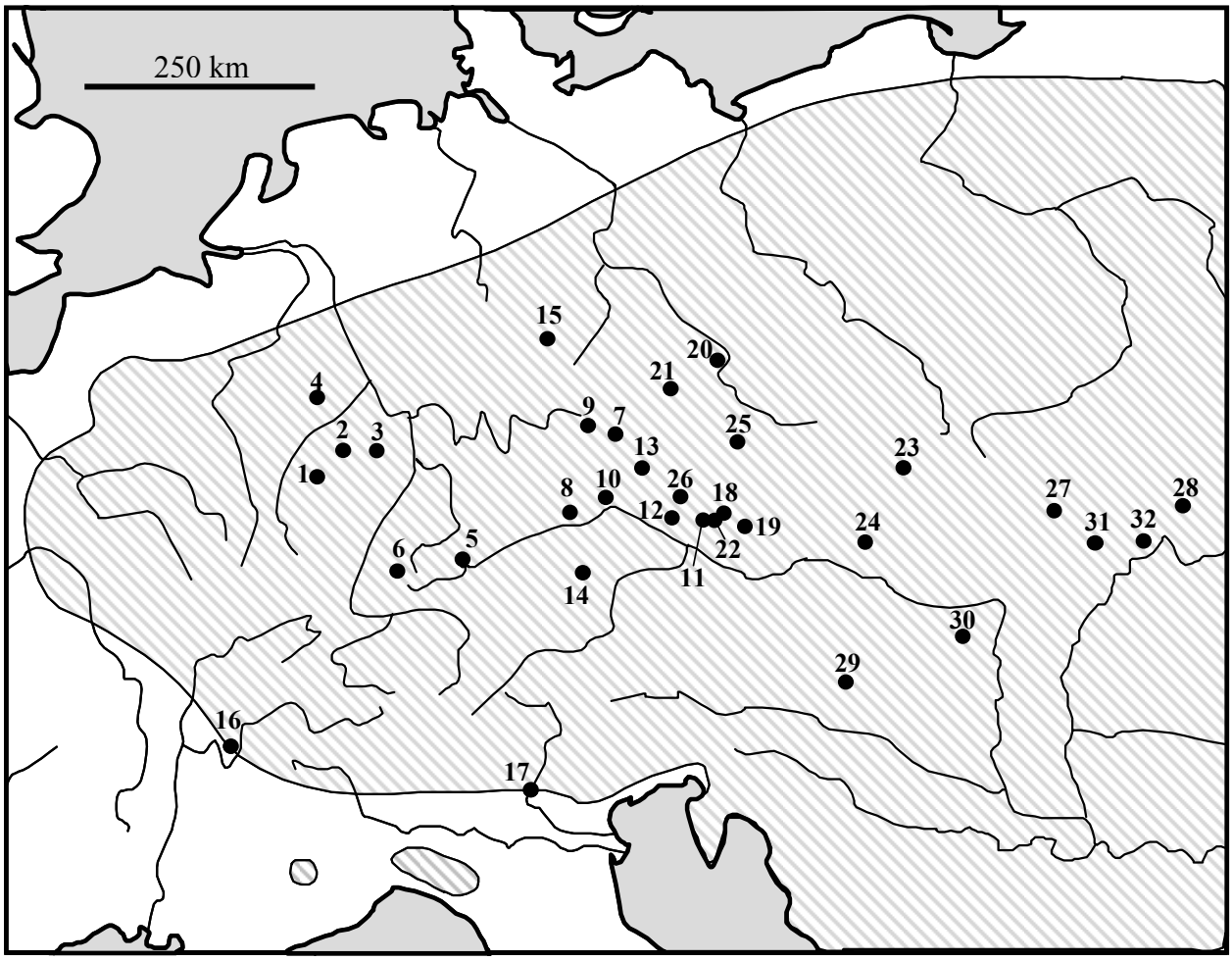


Figure 2.1: Sample locations of *Erebia medusa* in Europe; grey shading represents the distribution area.

2.2.2. Sequencing of mitochondrial DNA

To evaluate mitochondrial DNA (mtDNA) variation, I consistently examined 5 specimens per locality. DNA was extracted from the thorax using the Roche High Pure PCR template preparation kit (Roche Diagnostics GmbH). PCR amplifications were performed in 25 µl volumes containing 1 µl DNA extract, 1 µl of each primer (15 pmol µL-1), 1 µl MgCl2 and 21 µl water (Carl Roth GmbH and Co). The primers C1-J-2183 and TL2-N-3014 (Simon et al.

1994) were used to amplify a circa 800 bp fragment of the COI gene. I focused on the amplification of the second half of COI because it contains a higher rate of variability for this gene (Lunt et al. 1996). The PCR program started with denaturation at 95 °C for 5 min, 35 further cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and a final extension step at 72 °C for 90 s. PCR products were run on a 1.4% agarose gel and checked visually under UV light. Positive PCR products were purified with Roche High Pure PCR purification kit and used for single stranded sequencing with the primer TL2-N-3014 and the following program: denaturation at 96 °C for 1 min, 25 cycles of denaturation at 96 °C for 30 s, annealing at 45 °C for 15 s, and extension at 72 °C for 4 min. PCR products were sequenced with an automatic sequencer ABI 377 A. I finally obtained 529 bp long sequences for all samples that were aligned with the Sequence Navigator software (Applied Biosystems) and subsequently refined by eye.

2.2.3. Nested clade phylogeographic analysis

Nested clade phylogeographic analysis (NCPA), commonly used to analyse intraspecific phylogeography, detects, without *a priori* assumptions on the underlying processes, non-random associations of haplotypes with their geographic location. The method is based on a test of the following null hypothesis: there is no geographical association between the position of a haplotype in a gene tree and its geographical distribution. For significant associations, an inference is made if they are due to e.g., recurrent events such as restricted gene flow or historical events such as fragmentation, range expansion or colonization (Templeton et al. 1995).

First I calculated a minimum spanning haplotype network on the basis of statistical parsimony with TCS 1.18 (Clement et al. 2000). This network was unrooted. I therefore tried to identify the root through the inclusion of two outgroups, *E. gorge* and *E. epiphron*. However, due to a pronounced divergence among these species and *E. medusa* TCS could not link either outgroup to the network. The position of the root was therefore determined *via* a maximum parsimony (MP) analysis. I calculated an MP tree with PAUP* (Swofford 2001), including three further hierarchical outgroups: *Maniola jurtina* and *Coenonympha pamphilus* as representatives of satyrine genera closely related to *Erebia* and *Melitaea latonigena*, a Nymphalidae. I defined the latter as outgroup.

The minimum spanning network was then converted into a nested clade design following the rules described by Templeton et al. (1987) and Templeton and Sing (1993). Distance

measures for any clade X such as clade distance Dc (X), nested clade distance Dn (X), and differences between interior and tip clades (I-T) for Dc and Dn were calculated using Geodis 2.2 (Posada et al. 2000). This program tests at the 5% level of significance the distribution of these distance measures under the null hypothesis of no geographical association, applying 1,000 random distributions of all clade members on their respective locations (Templeton et al. 1995). Finally I inferred the history of current haplotype distribution for clades with non-random association of haplotypes to geographic locations using the revised inference key of Templeton (2005).

2.3. Results

I recorded 16 haplotypes from the 160 individuals sequenced. Haplotypes 01 and 02 are each represented in 26% of the sequenced individuals. The frequencies of haplotypes 03 and 04 are 20% and 10%, respectively, and all remaining haplotypes occur at a rate less than 5%. Some haplotypes such as H06 are local, while others such as H01 or H02 are more widespread. The most common haplotypes are all fairly widespread. Haplotypes were distributed as three distinct groups in Western, Central and Eastern Europe (Figure 2.2), e.g., haplotype 01 is endemic to Western Europe, haplotype 02 to Central Europe and haplotype 04 to Eastern Europe.

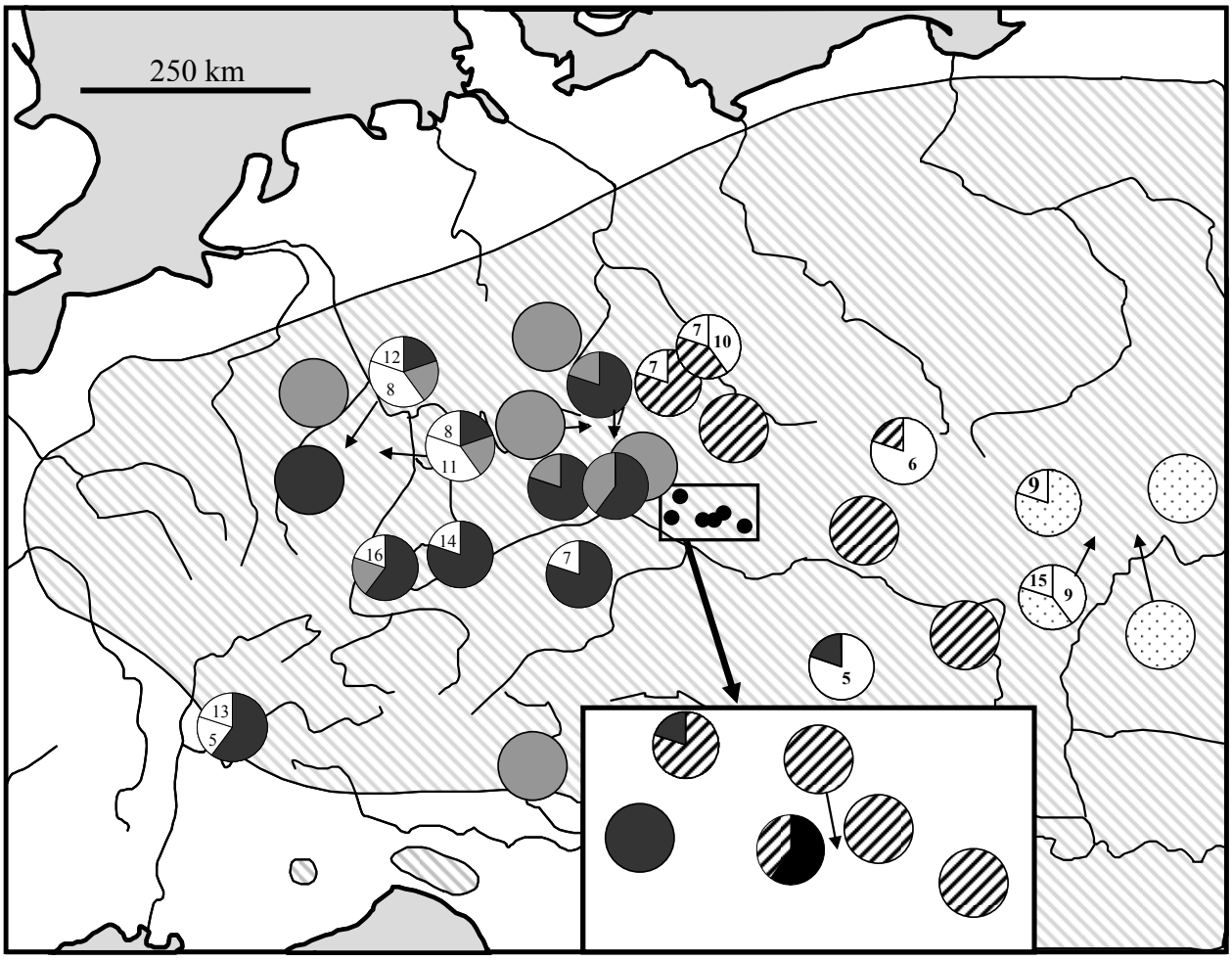


Figure 2.2: Geographic distribution of COI haplotypes of *E. medusa*. Black=haplotype 1, hatched=haplotype 2, grey =haplotype 3, spotted=haplotype 4; all other haplotypes are represented by numbers.

The resulting minimum spanning tree (Figure 2.3), representing the genealogical relationship among haplotypes, connects each haplotype by only one mutational step to its nearest neighbour. Its root was found at an interior position between H09 and an undiscovered or even extinct hypothetical haplotype.

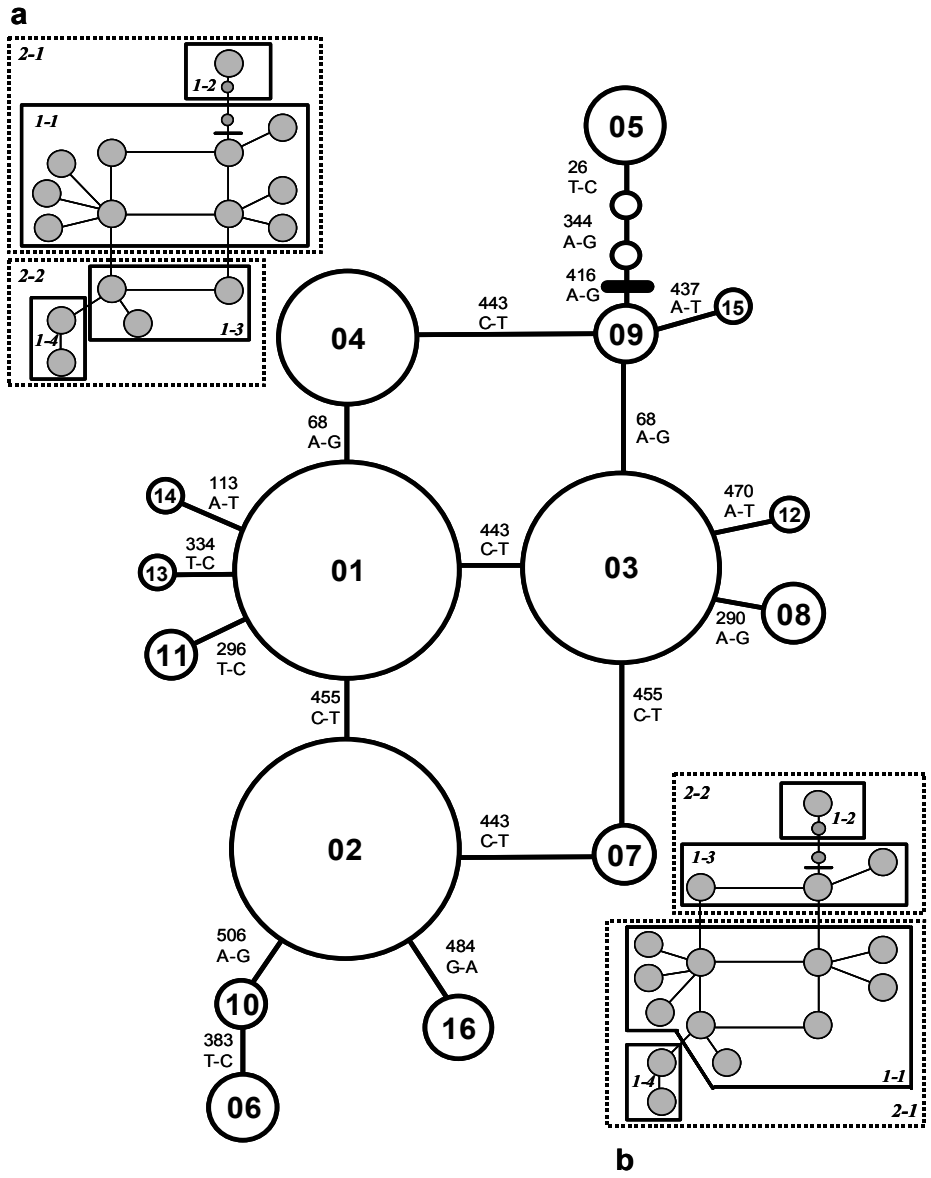


Figure 2.3: Minimum spanning network with a 2 ring ambiguity for *E. medusa* haplotypes; all substitutions are mapped on the tree, small circles represent haplotypes not found; (a) and (b) illustrate the two possible nested clade designs. The root between H09 and a hypothetical haplotype (black bar) was identified *via* maximum parsimony outgroup rooting.

The minimum spanning tree contains a central 2-ring ambiguity involving 6 haplotypes (H04, H09, H01, H03, H02 and H07). Three factors may account for such ring-shaped ambiguities: sequencing mistakes, homoplastic substitutions and recombination (Templeton and Sing 1993). To exclude sequencing artefacts, I carefully checked my sequences at all sites involved

in the ambiguity. Additional backward sequencing with primer C1-J-2183 ensured correct sequencing and interpretation. Hence I presumed that the origin of the ring is from homoplasy or recombination.

Homoplasy, the independent occurrence of the same mutation in non-related lineages, is unlikely to occur when haplotype diversity is rather low, as is the case in most intraspecific analyses. However, in humans mutation is so highly non-random that homoplasy is common even in intraspecific haplotype trees (Templeton et al. 2000). Recombination is considered rare or absent in mtDNA (Moritz et al. 1987), but Lunt and Hyman (1997) has provided evidence for the presence of recombination in the mtDNA of the phytonematode *Meloidogyne javanica*. I could not determine the origin of the ring ambiguity in this case. It is certain, however, that either recombination or homoplasy may constitute a disturbance in the minimum spanning tree inducing alternative evolutionary pathways of equal probabilities that, by definition, are mutually exclusive. Due to this 2-ring ambiguity in the minimum spanning network, one of the special rules of Templeton and Sing (1993) indicating that clades included in a loop have to be nested together and treated as a single clade at the next clade level had to be applied to proceed with the NCPA. This necessitated two alternative nested clade designs (as shown in Figure 2.3) with two separate NCPA trials (Table 2.1).

Table 2.1: Results of the inference key for the ambiguous networks resolved with the additional rules given in Templeton and Sing (1993).

Scenario A

Clade	**Inferred phylogeographic scenario**
clade 1-1	1-2_{d} -$3_{a,b,c}$ -5-6-13-YES → past fragmentation followed by range expansion
clade 1-3	1-2-11-17- inconclusive outcome
clade 2-1	1-2-Tip/interior status cannot be determined-inconclusive outcome
clade 2-2	1-2- Tip/interior status cannot be determined-inconclusive outcome
total clade	1-$2_{a,c,d}$ -3-4-9-NO → allopatric fragmentation (weakened by only one mutational step between 2-1 and 2-2)

Scenario B

Clade	Inferred phylogeographic scenario
clade 1-1	1-2_b –$3_{b,c,d}$-6-13-YES → past fragmentation followed by range expansion
clade 2-1	1-$2_{b,d}$ –3-4- NO → restricted gene flow with isolation by distance
clade 2-2	1-19-20-2-11_b -$YES_{range\ expansion}$-12-NO → contiguous range expansion
total clade	1-2-11_b - $YES_{range\ expansion}$-12-13-YES → past fragmentation followed by range expansion

I subsumed the conclusions of the two NCPA scenarios under an NCPA consensus scenario (see below Table 2.3). I emphasise that this consensus scenario constitutes a fusion at different clade levels of the conclusions resulting in the two clade nesting scenarios A and B. This is only possible because the clade 1-1 and the total clade relate to comparable geographical areas (see discussion chapter 2-4). Inferences at the remaining clade level (1-3, 2-1, 2-2) lead to an inconclusive outcome for clade nesting scenario A.

2.4. Discussion

The present study based on mtDNA showed a notable amount of diversity among the populations of *E. medusa* in Europe. Moreover, the COI haplotype distribution revealed three distinct groups: a Western group, a Central group and an Eastern group. This geographic pattern is in concordance with the allozyme data of Schmitt and Seitz (2001a) which show that samples from eastern France, Germany, northern Italy, Czech Republic, Slovakia and Hungary revealed a complex genetic structure with four major genetic lineages: (i) a western one comprising the samples from eastern France and Germany, (ii) an eastern one composed of the samples from Czech Republic, Slovakia and north-eastern Hungary, (iii) a Pannonic one including samples from western Hungary and (iv) a southern Alps one represented by a sample from the Monte Baldo massif (Figure 2.4).

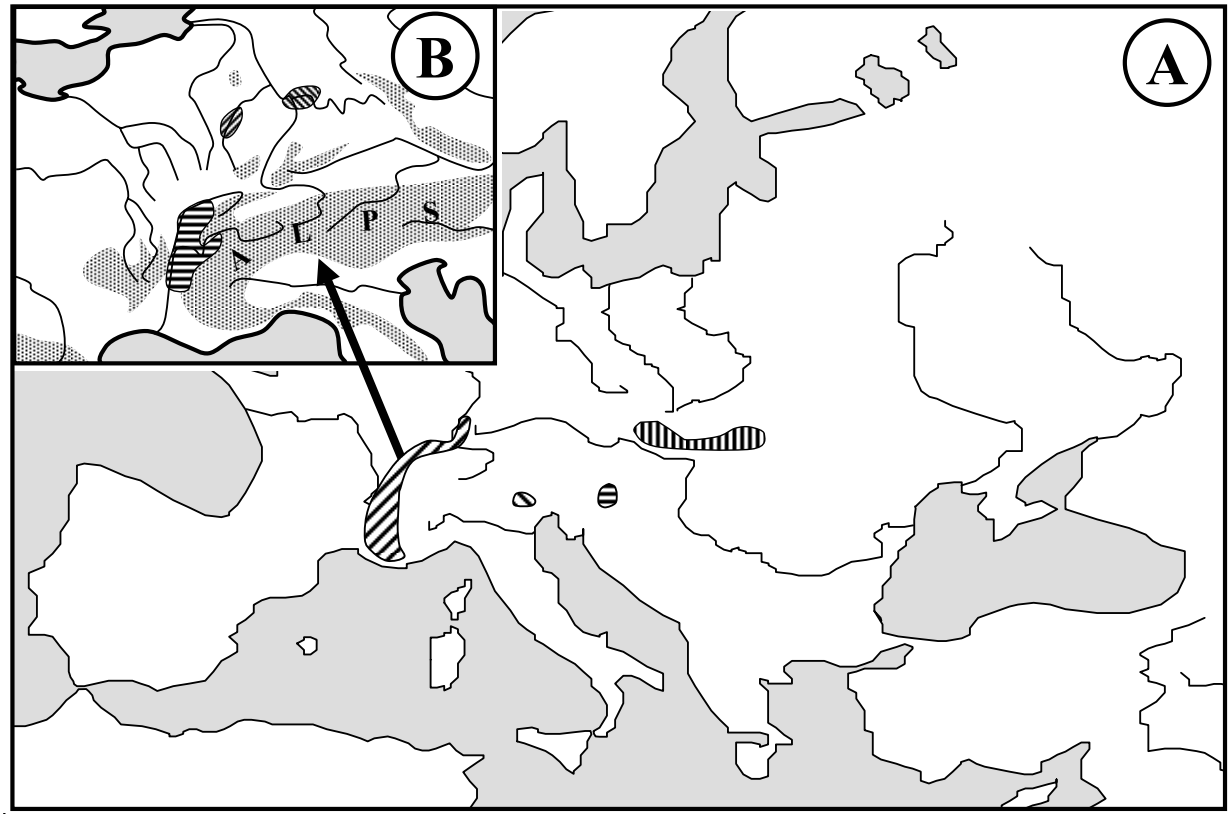


Figure 2.4: (A) Glacial refugia of *E. medusa* and (B) the most probable distribution of the western lineage at the end of the Younger Dryas as supported by allozyme data (from Schmitt and Seitz 2001a, modified due to unpublished further allozyme data). Hatched areas=glacial refugia.

In my study the Monte Baldo sample was part of the Central European haplotype group. The inability of my mtDNA data to detect these recent differentiations at the allozyme level might be due to the fact that the allozyme differentiation is simply based on changes in allele frequency. Most probably, the available time-scale was too short to allow a sufficient resolution to evolve at the mtDNA level. The amount of genetic variation in the allozyme study was high (FST=14.9%), and the fraction of genetic variance distributed among major lineages comprised about two thirds of the total genetic variance. Both the western and eastern lineage showed a pronounced sub-structuring. Furthermore, genetic diversity of the eastern lineage was significantly higher than in the other lineages (Schmitt and Seitz 2001a).

E. medusa is a species that tolerates rather low winter temperatures (Korschunov and Gorbunov 1995). However, cold resistance may not have been the most important factor for its distribution during the last glaciation. Water availability might have been of greater importance (Schmitt and Seitz 2001a). For these two reasons, the species distribution during the last glaciations was restricted to more southern latitudes (e.g., south-eastern Europe) and around glaciated high mountain systems such as the Alps.

The published allozyme data imply that *E. medusa* has had two major differentiation centres (west and north-east of the Alps) and two small centres of differentiation at the southern and south-eastern margins of the Alpine glaciers (Figure 2.4A). The remarkable substructures in the two major lineages imply one or two expansion-retraction cycles during the late Pleistocene after the Last Glacial Maximum (LGM) prior to the final postglacial range expansion with probably three sub-centres in the western lineages (Figure 2.4B). This scenario, based on allozymes, can be summarised to six detectable major events (Table 2.2).

Table 2.2: Phylogeographic scenario of *E. medusa* evolution in Central Europe based on allozyme data (Schmitt and Seitz 2001a); genetic distance according to Nei (1978); ybp: years before present.

Event number	Genetic structure	Genetic distance	Event	Time (ybp)
1	**Fragmentation:** Four major genetic lineages (Western, Southern Alps, Western Hungarian and Eastern)	0.051 – 0.117	vicariance due to the onset of the Würm ice-age	70,000
2	**Fragmentation:** Two sub-lineages in the Western lineage (France+ SW Germany and rest of Germany)	0.032	vicariance during Last Glacial Maximum or Older Dryas	18,000 or 13,500
3	**Fragmentation:** Two sub-lineages in the eastern lineage (Slovakia+ NE Hungary and Czech Republic)	0.036	vicariance during Last Glacial Maximum or Older Dryas	18,000 or 13,500
4	**Fragmentation:** Two secondary sub-lineages in the rest of Germany sub-lineage (W Rhineland-Palatinate+ Saarland and E Rhineland-Palatinate+ Thuringia+ Bavaria)	0.023	Younger Dryas	11,500
5	**Fragmentation:** Two secondary sub-lineages in the Czech Republic sub-lineage (Bohemia against Moravia)	0.026	Younger Dryas or postglacial range expansion	11,500 or younger
6	**Range expansion:** to different degrees in all sub-lineages	< 0.025	postglacial	since 10,500

A first vicariance caused by the onset of the Würm glaciation led to the formation of the four major lineages (Figure 2.4A). Vicariance during the LGM (22 ka bp) or Older Dryas (13.5 ka bp) formed two sub-lineages in the Western lineage (event 2) and in the two Eastern sub-lineages (event 3). Later, during the Younger Dryas (11.5 ka bp), the German sub-lineage most probably fragmented into two secondary sub-lineages (event 4). The authors surmise that two secondary sub-lineages in the Czech Republic have evolved due to (i) vicariance during the same time period or (ii) differentiation during a postglacial expansion process from the more eastern Moravia to the more western Bohemia (event 5). Range expansions to varying degrees in all sub-lineages from 10,500 years bp is assumed (event 6).
I overlaid this allozyme scenario, which is supported by the geographical distribution and relative age of haplotypes of this study, to my NCPA consensus scenario (Table 2.3).

Table 2.3: Consensus phylogeographic scenario of *E. medusa* evolution in Central Europe based on mtDNA data.

Consensus Scenario

Clade	**Inferred phylogeographic scenario**
clade 1-1	Fragmentation followed by range expansion
total clade	Fragmentation

The haplotype composition of the first clade level 1-1 is different in both NCPA scenarios but the fusion of the conclusions from both scenarios is possible because the geographical distribution of the haplotypes bearing this first clade level 1-1 is widely overlapping. Thus most of the haplotypes involved are localised in Central Europe, but scenario B also implies haplotypes of eastern Europe as present in this clade. Hence both scenarios at this clade level (with different statistical values; see Appendix 2) implied past fragmentation followed by range expansion, but it can be interpreted as a consensus conclusion only for the populations from Central Europe. The total clade level inferred allopatric fragmentation for scenario A and past fragmentation followed by range expansion for scenario B. The affected geographical area comprises the entire study area, but the sub-clades relate to different populations. The consensus of both events is again a fragmentation event.
Due to the temporal hierarchy among clade levels, the total clade of the NCPAs is supposed to correspond to the oldest events of the allozyme scenario, i.e., vicariance due to the onset of the last ice age (event 1; Schmitt and Seitz 2001a). The lower clades should address more recent events: (i) LGM (event 2 and 3) at the second clade level and (ii) Younger Dryas and

postglacial (events 4 to 6) at the first clade level. The NCPA consensus scenario supported one of the younger vicariance events and the postglacial range expansion. Hence the consensus conclusion is assumed for Central European populations, while it supports the vicariance in the rest of Germany (event 4). This scenario also agrees with the allozyme scenario at the total level supporting the existence of several differentiation centers during the first phase of the Würm glaciation. However it is not possible here to precisely identify the lineages involved in the fragmentation event. The scenario failed at the second clade level to detect the vicariance and also to deliver additional information for the LGM or Older Dryas (event 2 and 3).

NCPA analysis was recently criticized as producing errors in its inference that may lead to incorrect geographical hypotheses. It was blamed for being unable to accurately infer or distinguish among alternative processes (Knowles and Maddison 2002). Templeton (2004) therefore revised his inference key to account for such type I and type II errors to minimize inference errors. I was aware that the only statistically supported results I gained from NCPA refer to the null-hypothesis of non-random distribution of haplotypes with respect to geography. However, given that the phylogeographic conclusions inferred by my NCPA consensus scenario mainly support the invoked allozyme scenario of *E. medusa* in Europe, my study again supports the validity of NCPA for phylogeographic reconstruction.

The combination of nuclear and mitochondrial data, using NCPA as a tool, allowed us to establish a congruent phylogeographic scenario for *E. medusa* in Europe. Both data sets consistently revealed a distinct genetic structure of *E. medusa* in Europe, with three genetically defined groups occurring in Western, Central and Eastern Europe. Both markers support (i) a first vicariance due to the onset of the Würm glaciation, leading to the formation of major lineages, (ii) the further differentiation of the German population during the Younger Dryas and (iii) the postglacial range expansion since 10,500 years ago.

2.5. Summary

During the Quaternary, drastic climatic oscillations induced range expansions and contractions of temperate fauna and flora, which are expected to have left signatures in the geographical distribution and genetic diversity of extant populations. Hence, an array of different historical and recurrent processes may have shaped today's species distributions. The advent of molecular techniques such as PCR and statistical tools such as nested clade phylogeographic analysis (NCPA) enables the deduction of these processes through analysis

of intraspecific genetic variation. I analyse the genetic variation of the western Palaearctic butterfly species *E. medusa* in Europe to examine its phylogeography in Central Europe.

I sequenced part of the COI mitochondrial gene to examine the phylogeography of the Woodland Ringlet (*E. medusa*) using NCPA. This analysis considers - and distinguishes between - contemporary (e.g., restricted gene flow) and historical (e.g., past fragmentation, range expansion or colonization) processes. It detects without *a priori* assumptions non-random geographical haplotype associations and proposes the best fitting phylogeographic scenario.

The NCPA consensus scenario, resulting from two alternative hypotheses, was overlain to a commonly accepted scenario derived from published allozyme data. I thus infer an evolutionary history of *E. medusa* in Europe based on the total evidence from the nuclear and the mitochondrial scenarios.

My data reveal the existence of three distinct lineages of *E. medusa*, distributed in Western, Central and Eastern Europe, respectively. Both markers supported (*i*) a first vicariance at the onset of the Würm glaciation leading to the formation of major lineages, (*ii*) the differentiation of the German population during the Younger Dryas and (*iii*) postglacial range expansion.

3. Adjustment of ring-shaped ambiguities in minimum spanning networks for recombination and homoplasy: its impact on phylogeographic reconstruction

3.1. Introduction

The application of coalescent theory to phylogeography enables the inference of hypotheses on population history (Avise 2000). Gene genealogies, the central patterns of ancestry and descendance of the coalescence approach, harbour information about population demography through time. In a geographical projection they provide information about population history in space. This historical information is inherent in genealogies and can be extracted in the framework of a temporally and geographically hierarchical analysis of the spatial distribution of genetic variation (Templeton 1998).

This kind of analysis, namely the nested clade phylogeographic analysis (NCPA), gives information about the population history (range expansion, colonization or past fragmentation) and population structure. It relies on a network that connects haplotypes on the basis of statistical parsimony. The haplotypes of the minimum spanning network are nested in clades following a nested hierarchy that reflects an increasing temporal dimension both within and among clade levels (Templeton et al. 1995). The nested design is used to test for associations of haplotypes with geography through the comparison of observed and expected patterns (under different phylogeographic scenarios) using permutation chi-square contingency tests.

During the reconstruction of the 95% plausible sets of haplotype connections, eventually ring-shaped haplotype associations (loops) may appear. They illustrate alternative evolutionary pathways of equal probability which, per definition, are mutually exclusive. Such ambiguities may be caused by recombination, homoplastic mutations or simply through sequencing mistakes (Templeton and Sing 1993). They may constitute a significant problem for the construction of a nested clade hierarchy.

Recombination is a major process by which new genotypes are generated, leading to offspring with sets of genes different from those of either parent. The recombinants comprise a combination of genes from both parental genomes. In animals, recombination is frequent in nuclear DNA, but it is considered rare or absent in mitochondrial DNA (mtDNA) (Moritz et al. 1987). For this reason mitochondrial genes are commonly used in phylogeographic studies.

The requirement for recombination to occur in mitochondria is the rejection of the strict maternal inheritance of mtDNA, supposedly through paternal leakage or doubly uniparental inheritance (DUI).

Homoplasy is defined as the acquisition of the same character state in taxa that arose through independent evolution instead of descent from a common ancestor. It is due to processes such as convergence, parallelism or reversion. Convergent evolution corresponds to the independent acquisition of the same character from different ancestral conditions, and in the case of parallel evolution this acquisition results from the same ancestral state. Reversion is the secondary loss of a derived character, leading back to the ancestral condition (Estoup et al. 2002, Rokas and Holland 2000). In nature, homoplasy occurs frequently and at practically every level of biological organization (Wood et al. 2005). It has been identified in cpDNA (Navascues and Emerson 2005) as well as in mtDNA, where mutations may occur non-randomly (e.g., in hypervariable regions; Herrnstadt et al. 2002). In human mtDNA hypervariable positions constitute 6% of all polymorphic positions (Malyarchuk 2005). If we accept the phenomena of recombination (and the inheritance of the recombinants) and homoplasy in mtDNA, it is necessary to consider their effect on phylogeographic reconstruction.

Using the example of a 2-ring loop ambiguity that emerged in a minimum spanning network from a phylogeographic study of the western Palaearctic butterfly species *E. medusa* (chapter 2), I analyse its impact on NCPA analysis. In the first hypothesis I consider recombination as the source of the ring-shaped ambiguity. I outline a strategy to distinguish between recombinant and parental haplotypes. I then differentially exclude these recombinants from NCPA. In a second approach, I consider homoplasy as the cause of the ambiguity. I identify homoplastically evolving sites and exclude them from phylogeographic analyses. Phylogeographic scenarios from both approaches are finally overlaid on a previously published allozyme-based evolutionary history for the Woodland Ringlet *E. medusa*.

3.2. Material and methods

3.2.1. Haplotype network construction and ring-shaped ambiguities

For the phylogeographic study of *E. medusa* 160 butterflies from 32 populations across Europe were sampled and the cytochrome oxydase subunit one (COI) mitochondrial gene was sequenced for all of them (for details see chapter 2). The geographic haplotype distribution (Figure 3.1) identifies three distinct groups in Western, Central and Eastern Europe.

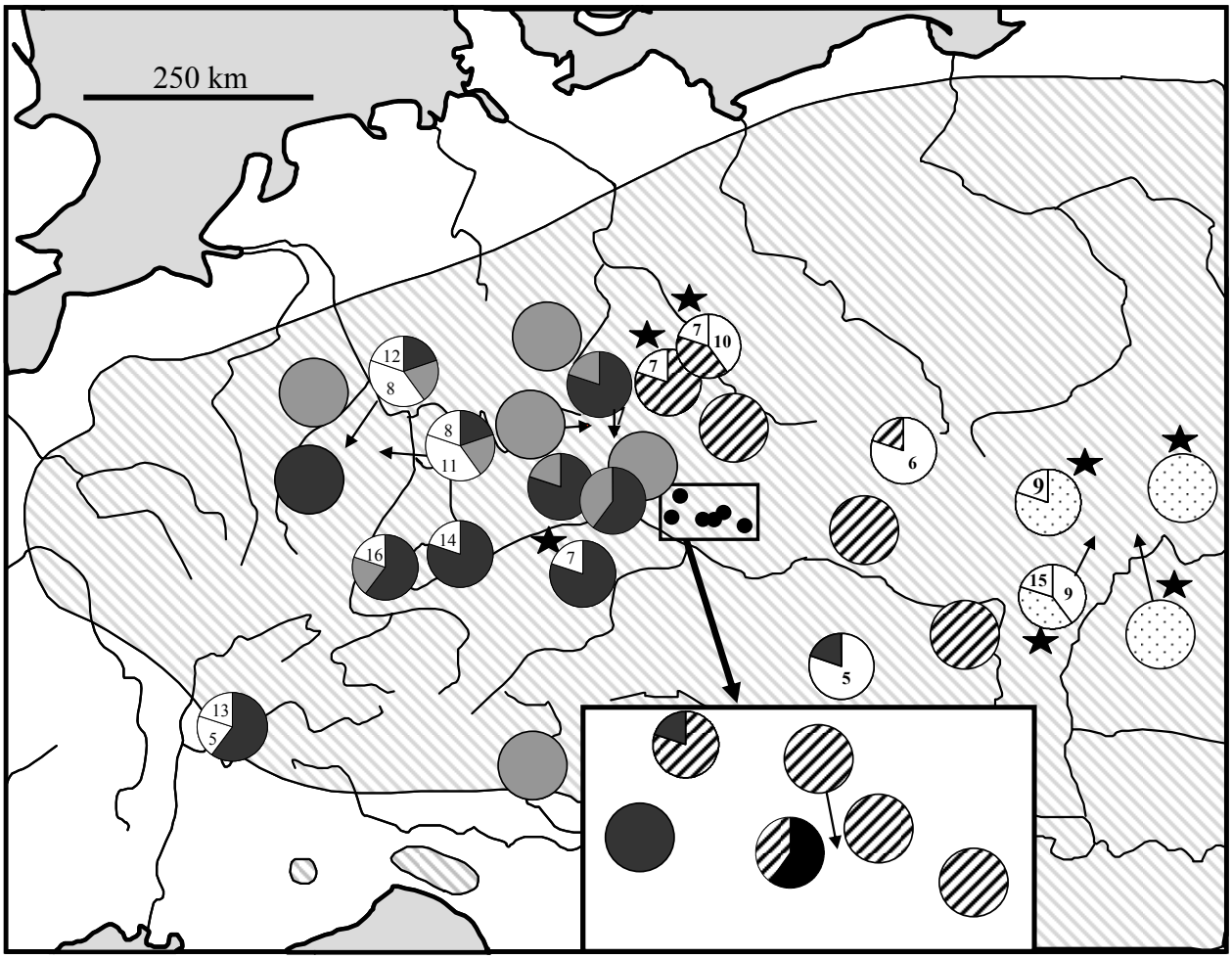


Figure 3.1: Geographic distribution of COI haplotypes of *E. medusa*. Black=haplotype 1, hatched=haplotype 2, grey=haplotype 3, spotted=haplotype 4; remaining haplotypes are marked by numbers; stars indicate recombinant haplotypes.

The minimum spanning tree (Figure 3.2), calculated with the TCS software (Clement et al. 2000), representing the genealogical relationship among haplotypes connects each haplotype by only one mutational step to its nearest neighbour.

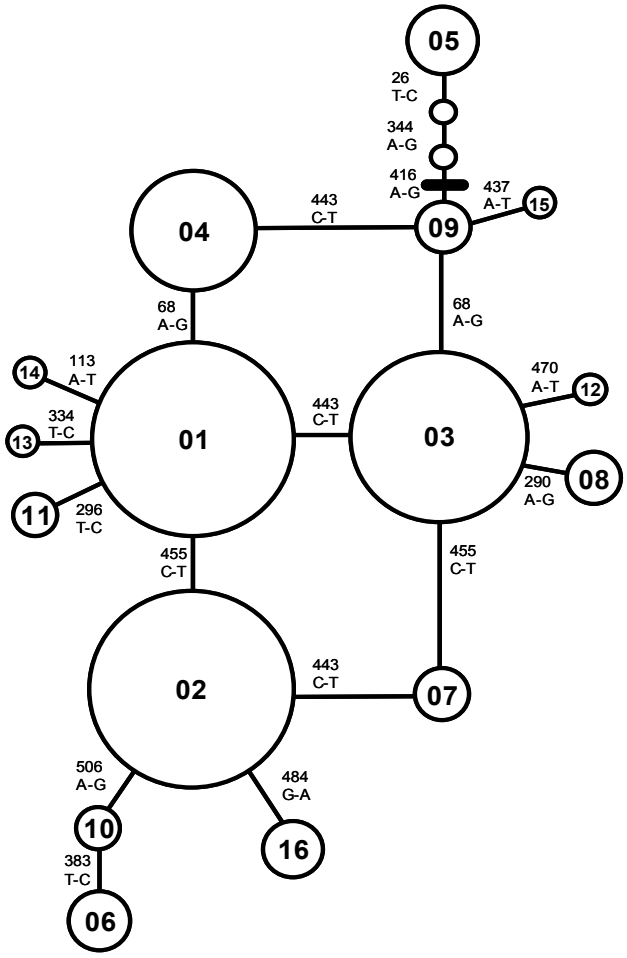


Figure 3.2: Minimum spanning network for *E. medusa* with a 2-ring ambiguity (from chapter 2); circle areas are proportional to the corresponding haplotype frequencies; small unnumbered circles represent hypothetical haplotypes that were not found; substitutions are mapped on the tree; the root between H09 and a hypothetical haplotype (black bar) was identified *via* maximum parsimony outgroup rooting.

It contains a central 2-ring ambiguity, involving six haplotypes (H04, H09, H01, H03, H02 and H07). I tried to identify the root through the inclusion of two outgroups, *E. gorge* and *E. epiphron*. However, due to a pronounced divergence among these species and *E. medusa* TCS could not link either outgroup to the network. The position of the root was therefore determined *via* a maximum parsimony (MP) analysis. I calculated an MP tree with PAUP* (Swofford 2001), including three additional taxa: *Maniola jurtina* and *Coenonympha pamphilus* as representatives of satyrine genera closely related to *Erebia* and *Melitaea latonigena*, a Nymphalidae. I defined the latter as an outgroup. The MP tree rooted the *E.*

medusa network at an interior branch between H09 and an undiscovered or even extinct hypothetical haplotype.

3.2.2. Investigation of recombination and homoplasy

As mentioned above, three factors may account for such ring-shaped ambiguities: sequencing mistakes, homoplastic substitutions and recombination (Templeton and Sing 1993). To exclude sequencing artefacts, I carefully checked my sequences at all sites involved in the ambiguity; additional backward sequencing ensured correct sequencing and interpretation. Consequently, recombination and homoplasy remain as possible explanations for the ring-shaped ambiguity.

a- Resolving ring ambiguities under recombination

To assess recombination I developed a strategy to detect and extract the recombinants from the haplotype network *prior* to phylogeographic inference. I first resolved the ring-shaped ambiguities by applying rules derived from (i) empirical predictions under a neutrality hypothesis and (ii) gene tree analyses with hypothetical recombinant data sets based on simple putative parental sequences under a hypothesis of recombination. These "simulated" data showed that a single loop with two substitutions may be obtained when two parental haplotypes experience one crossing over, resulting in two recombinants (Figure 3.3).

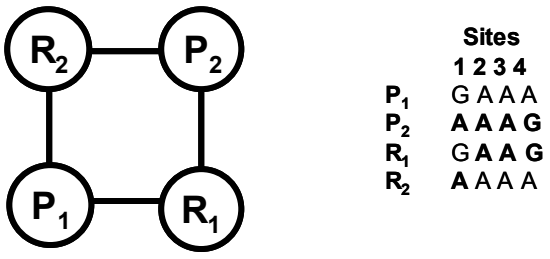


Figure 3.3: Inference of the "position rule" in a one-ring ambiguity; a crossing over of two parental haplotypes P_1 and P_2 between sites 1 and 2 produces recombinant haplotypes R_1 and R_2; a TCS analysis produced a "haplotype ring" with the recombinants sitting in opposite positions.

Moreover, it became obvious that a recombinant is always directly linked to its parents. Consequently, in a four haplotype ring, two recombinants always sit opposite to each other. I name this the "position rule". Further "simulations", however, showed that there exists

another way to obtain a four haplotype ring. It starts with three original haplotypes (potential parents, thus non-recombinant), only two of which act as parental haplotypes. When recombining, they again produce two recombinant haplotypes, however, only one of them corresponds to a new and thus detectable variant. The second recombinant will be identical to the third, non-parental haplotype. In a further step, I performed a "simulation" to show that to obtain a 2-ring ambiguity, such as in my example, three mutational sites have to be involved. Moreover a combination of all possible haplotypes when regarding these three sites results in a total of eight haplotypes. The minimum spanning network calculated with TCS and including these eight haplotypes results in a haplotype cube (for model see results with Figure 3.4).

I applied two additional empirical rules also invoked by Pfenninger and Posada (2002). It is expected that the older haplotypes occupy the interior position in the haplotype network and have a greater number of mutational connections. Under the recombination hypothesis, older haplotypes must be considered as parental since by definition parents must be older than their offspring (including recombinants). I name this the "age rule". Finally, geographically viewed, connection between haplotypes from the same population or region is more likely than connection between haplotypes occurring in distant populations. This is the "geographical rule".

b- Detection of homoplasy

To account for homoplasy I calculated the consistency index (CI; Kluge and Farris 1969) for all variable sites. The CI quantifies the degree to which a character evolves homoplastically on a given tree. It equals 1 when there is no homoplasy and it decreases towards zero as the degree of homoplasy increases. I mapped all variable sites on a neighbour joining (NJ) tree calculated with PAUP* (Swofford 2001), using the best-fitting substitution model (see chapter 2 for details). CI values were calculated with MacClade (Maddison and Maddison 1992). I considered a site as evolving homoplastically when the CI value was below 1.

3.2.3. Nested clade phylogeographic analysis

Nested clade phylogeographic analysis (NCPA), commonly used to analyse intraspecific phylogeography, detects non-random associations of haplotypes with their geographic location without *a priori* assumptions on the underlying processes. The method is based on a

test of the following null hypothesis: there is no geographical association between the position of a haplotype in a gene tree and its geographical distribution. For significant associations, a test determines if they are due to e.g., recurrent events such as restricted gene flow or historical events such as fragmentation, range expansion, or colonization (Templeton et al. 1995).

A minimum spanning network, calculated on the basis of statistical parsimony, is converted into a nested clade design following the rules described by Templeton et al. (1987) and Templeton and Sing (1993). Distance measures for any clade X, such as clade distance D_c (X), nested clade distance D_n (X) and differences between interior and tip clades (I-T) for D_c and D_n, are calculated using Geodis 2.2 (Posada et al. 2000). This program tests at the 5% level of significance the distribution of these distance measures under the null hypothesis of no geographical association, applying 1,000 random distributions of all clade members on their respective locations (Templeton et al. 1995).

I performed NCPA analyse under the hypothesis of recombination and homoplasy to determine the history of current haplotype distribution for clades with non-random associations of haplotypes to geographic locations using the revised inference key of Templeton (2005).

3.3. Results

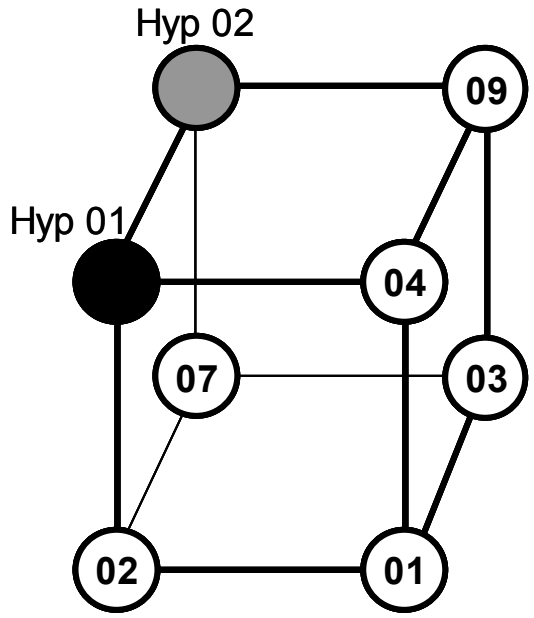
3.3.1. Recombination

To identify the recombinants in my 2-ring loop (Figure 3.2), I further analysed the evolution of H02 and H09. Each of them gave rise to several new haplotypes connected to them with 1-3 mutational steps. Following the "age rule" H02 and H09 are considered as parental. In the lower ring of Figure 3.2 (H01-H02-H03-H07), following the "position rule", H03 must be assigned as parental because H02 itself is, according to the age rule, a parental haplotype; hence, H01 and H07 would be recombinants. In my example H01, the potential recombinant of H02 and H03, continued to evolve in the same manner as H03. Hence, I assumed that the lower haplotype ring arose from three existing haplotypes (H01, H02 and H03) with only H07 being of solely recombinant origin. The H01 population is a compound of recombinant and non-recombinant alleles, among which I cannot distinguish *a priori*. This inevitably implies a

bias in NCPA analysis because one recombinant type will be underestimated (see discussion chapter 3-4).

Following the "age rule" in the upper ring of Figure 3.2 (H01-H03-H04-H09), H09 is a parental haplotype. Support for this comes from its relative position to the supposed root of the network. H01 could again be involved in a recombination event as a parental haplotype. However, the geographic distribution of H01 and H09 does not support this idea. H01 is present in only western populations, whereas H04, its potential recombinant offspring and the second parental haplotype, H09, are only found in the easternmost populations. Thus, due to the "geographical rule" a recombinant origin of H04 involving H01 as a parental haplotype is unlikely.

Regarding the three sites involved in my 2-ring ambiguity, two of the eight possible combined haplotypes (I named them Hyp01 and Hyp02) do not occur in my sample. They are hypothetical haplotypes, currently undiscovered or already extinct. In the resulting minimum spanning network calculated with TCS, the two hypothetical haplotypes fuse the opposite ends of the 2-ring ambiguity to a haplotype cube (Figure 3.4).



	Sites		
	68	443	455
H01	G	C	C
H02	G	C	T
H03	G	T	C
H04	A	C	C
H07	G	T	T
H09	A	T	C
Hyp 01	A	C	T
Hyp 02	A	T	T

Figure 3.4: Minimum spanning network of eight possible haplotypes when three variable sites 68, 443 and 455 are invoked with two possible nucleotides. The black circle represents the hypothetical haplotype Hyp01, the grey circle represents Hyp02; when omitting Hyp01, Hyp02 and their respective connections to other haplotypes, the 2-ring ambiguity of my original minimum spanning tree (Figure 3. 3) emerges.

Following the "position rule", parental haplotypes for H04 could now be either of the pairs H01-H09, H01-Hyp01 or Hyp01-H09. Again applying the "geographic rule" it is not plausible that either haplotype pair where H01 is involved (H01-H09 or H01-Hyp01) could be

the origin of H04. Therefore I consider Hyp01-H09 to be parents of H04. Consequently, Hyp01 is expected to exist or to have existed in a zone east of the sampled area.

Concerning recombinant H07, parental haplotypes could be the pairs H02-H03, Hyp02-H03 or Hyp02-H02. Geographically, H07 is found in the contact zone of H02 and H03, supporting the idea that H07 is a recombinant that arose from crossing over between H02 and H03.

Having identified H04 and H07 as recombinant haplotypes, I excluded them from my alignment, added Hyp01 and recalculated the minimum spanning tree with TCS. The new minimum spanning network was now completely resolved (Figure 3.5).

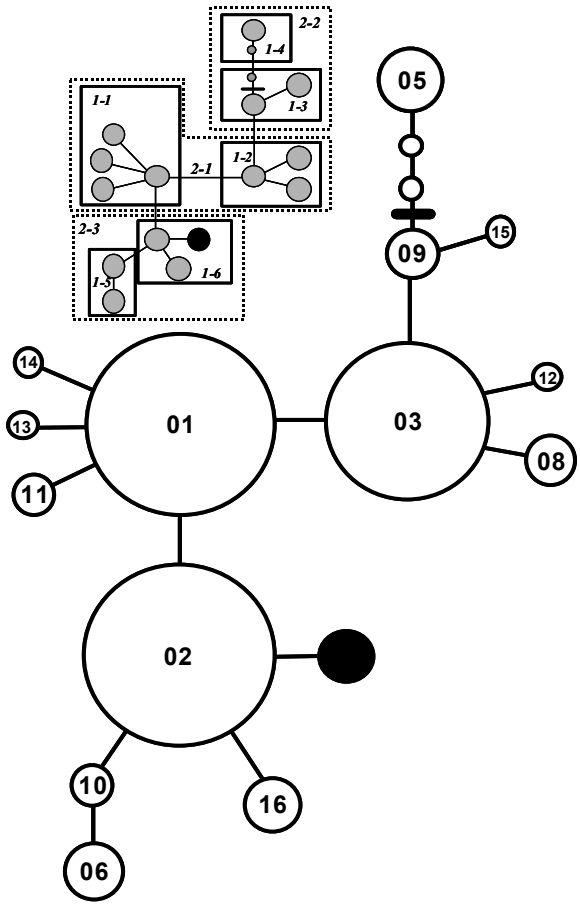


Figure 3.5: Minimum spanning network resolved under the recombination hypothesis; the black circle represents the parental haplotype Hyp01 (not found); circle areas are proportional to the corresponding haplotype frequency; the insert illustrates the nested clade design for scenarios A, B to E and F; the root between H09 and a hypothetical haplotype (black bar) was identified *via* maximum parsimony outgroup rooting.

I did not include Hyp02 because of the "position rule" and in contrast to Hyp01 it must be of a recombination origin. The reason I did not find Hyp02 in my sample may be due to (i) a small sample size, (ii) its existence to the east of the area sampled, or (iii) a crossing over point within a coding triplet, leading to a stop codon or another malfunction of COI.

I performed NCPA on the resolved network excluding the specimens bearing recombinant haplotypes (Figure 3.5; scenario A in Table 3.1). However, omission of recombinants inevitably excludes information through a reduction of my data set (i.e., loss of variation within populations or even loss of populations). I therefore added different methods for recombinant treatment to my analysis. I accounted for the information harboured by the omitted recombinants by assigning them differentially to parental haplotypes. All H04 and H07 individuals were assigned to a single parental haplotype, Hyp01 or H09 for H04 and H02 or H03 for H07. This resulted in four different scenarios, B to E. Because every recombinant contains information from both parental haplotypes, recombinants were equally assigned to each parental haplotype (scenario F). In this scenario parental haplotypes are inevitably geographically linked *via* the location where the recombinants occur. I ran an NCPA for each of these five additional scenarios (Table 3.1; Appendix 3).

Table 3.1: Results of the inference key for the resolved network under the recombination hypothesis.

Scenario A: *Recombinants excluded for the NCPA analyse.*

Clade	Inferred phylogeographic scenario
clade 2-1	1-2-Tip/interior status cannot be determined - inconclusive outcome
clade 2-2	1-19-20-2-11_b- $YES_{range\ expansion}$-12-NO → contiguous range expansion
clade 2-3	1-2-Tip/interior status cannot be determined - inconclusive outcome
total clade	1-$2_{a\ b\ d}$-3-4-NO-restricted gene flow with isolation by distance

Scenario B: *Recombinants included and counted in either parental haplotypes H02 and H09.*

Clade	Inferred phylogeographic scenario
clade 1-6	1-19-20-2-11-17- NO inconclusive outcome
clade 2-1	1-2-Tip/interior status cannot be determined-inconclusive outcome
clade 2-2	1-19-20-2-11_b-$YES_{range\ expansion}$-12-NO → contiguous range expansion
clade 2-3	1-2- Tip/interior status cannot be determined-inconclusive outcome
total clade	1-2_d-3_a-5-$6_{too\ few\ clades}$-7-YES → restricted gene flow/dispersal but with some long-distance dispersal

Scenario C: *Recombinants included and counted in either parental haplotypes H02 and Hyp01.*

Clade	Inferred phylogeographic scenario
clade 1-6	1-19-20-2-11$_{b}$-YES$_{\text{range expansion}}$-12-13-YES → past fragmentation followed by range expansion
clade 2-1	1-2-Tip/interior status cannot be determined-inconclusive outcome
clade 2-2	1-19-20-2-11$_{b\,c}$ - YES$_{\text{range expansion}}$-12-NO → contiguous range expansion
clade 2-3	1-2$_{a,d}$-3-4-NO → restricted gene flow with isolation by distance
total clade	1-2$_{a\,b}$-4-NO-restricted gene flow with isolation by distance

Scenario D: *Recombinants included and counted in either parental haplotypes H03 and H09.*

Clade	Inferred phylogeographic scenario
clade 1-6 (p=0,05)	1-19-20-2-11-17- inconclusive outcome
clade 2-1	1-2-Tip/interior status cannot be determined-inconclusive outcome
clade 2-2	1-19-20-2-11$_{b}$-YES$_{\text{range expansion}}$-12-NO → contiguous range expansion
clade 2-3	1-2-Tip/interior status cannot be determined-inconclusive outcome
total clade	1-2$_{d}$-3$_{a}$-5-6$_{\text{too few clades}}$-7-YES → restricted gene flow/dispersal but with some long-distance dispersal

Scenario E: *Recombinants included and counted in either parental haplotypes H03 and Hyp 01.*

Clade	Inferred phylogeographic scenario
clade 1-6	1-19-20-2-11$_{b}$ - YES$_{\text{range expansion}}$-12-13-YES → past fragmentation followed by range expansion
clade 2-1	1-2-Tip/interior status cannot be determined-inconclusive outcome
clade 2-2	1-19-20-2-11$_{b\,d}$ - YES$_{\text{range expansion}}$-12-NO → contiguous range expansion
clade 2-3	1-2$_{a\,d}$-3-4- NO → restricted gene flow with isolation by distance
total clade	1-2$_{a\,b}$-3-4- NO → restricted gene flow with isolation by distance

Scenario F: *Recombinants counted in <u>both</u> parental haplotypes, with the parental haplotype hyp01 included.*

Clade	Inferred phylogeographic scenario
clade 1-1	1-$2_{c,d}$-$3_{b,c}$-5-$6_{\text{too few clades}}$-7-$\text{YES}_{\text{two of three}}$ → restricted gene flow/dispersal but with some long-distance dispersal
clade 1-2	1-$2_{a,d}$-3-4 → restricted gene flow with isolation by distance
clade 1-5	1-19-20-NO inadequate geographical sampling
clade 1-6	1-2-11_b-$\text{YES}_{\text{range expansion}}$-12-13-YES → past fragmentation followed by range expansion
clade 2-1	1-2-11-17-inconclusive outcome
clade 2-2	1-19-20-2-11_b- $\text{YES}_{\text{range expansion}}$-12-NO → contiguous range expansion
clade 2-3	1-$2_{a,d}$-3-4- NO → restricted gene flow with isolation by distance
total clade	1-2-11_b-$\text{YES}_{\text{range expansion}}$-12-13-YES → past fragmentation followed by range expansion

It is not possible to determine which among these six scenarios (A to F) comes closest to the true evolutionary history of *E. medusa*. Therefore I compiled an NCPA consensus scenario (Table 3.2) from these six scenarios to overlay it to a commonly accepted scenario derived from published allozyme data. Construction of a consensus scenario is possible since all single scenarios (A to F) are based on the minimum spanning tree (Figure 3.5) harbouring the same clade level structure.

Table 3.2: Consensus scenario based on mtDNA under the hypothesis of recombination; number of scenarios with a different statistical value relative to the total number of significant scenarios.

Clade	Phylogeographic inference	Number of scenario
clade 1-6	Fragmentation	3/3[2]
clade 2-2	Range expansion	6/6[1]
clade 2-3	Restricted gene flow	3/3
total clade	Restricted gene flow	5/6

1- Statistical distance values same for scenario C, E and G and for scenario D and F.

2- Each scenarios with different statistical distance values.

3.3.2. Homoplasy

CI values obtained for the variable sites (Table 3.3) showed one site (bp 443) with a CI value of 0.33, which I omitted from my alignment.

Table 3.3: Consistency index (CI) for variable sites of part of the COI mitochondrial gene of *E. medusa*; number of character states and character changes are indicated.

Sites	Mutations	States	Changes	CI
1	26	2	1	1.00
2	68	2	1	1.00
3	113	2	1	1.00
4	290	2	1	1.00
5	296	2	1	1.00
6	334	2	1	1.00
7	344	2	1	1.00
8	383	2	1	1.00
9	416	2	1	1.00
10	437	2	1	1.00
11	**443**	2	3	**0.33**
12	455	2	1	1.00
13	470	2	1	1.00
14	484	2	1	1.00
15	506	2	1	1.00

The resulting minimum spanning network (without site 443) contained only 13 haplotypes (Figure 3.6).

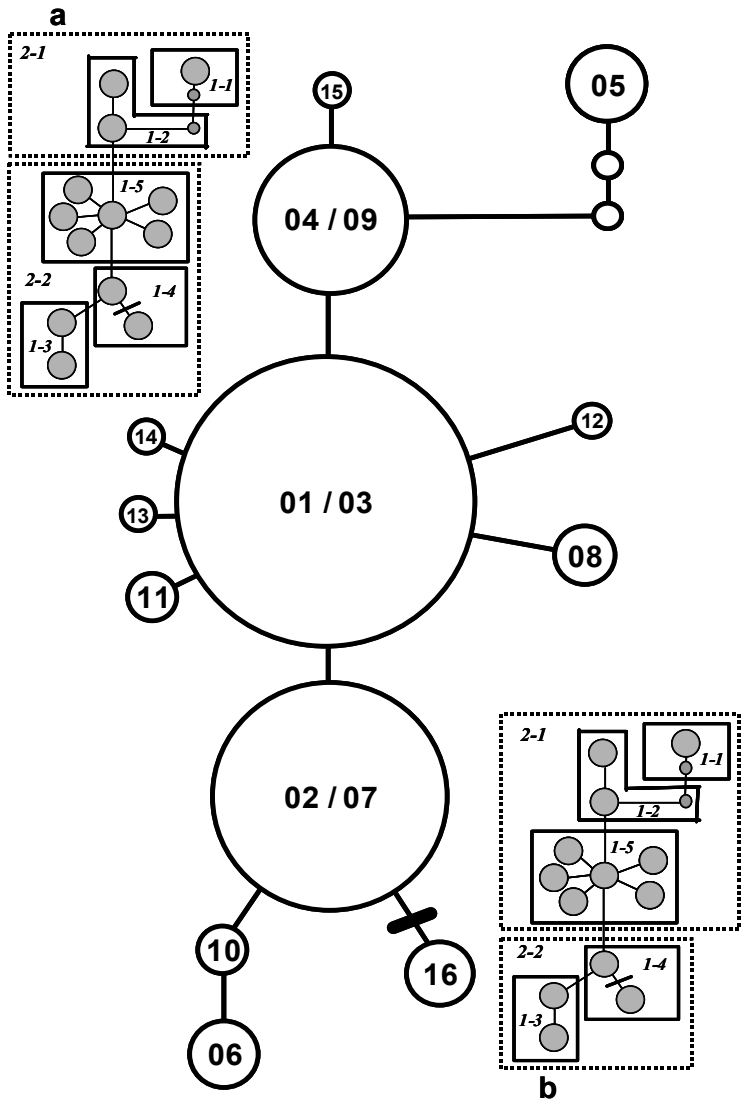


Figure 3.6: Minimum spanning network resolved under the homoplasy hypothesis; circle areas are proportional to the corresponding haplotype frequencies; the inserts illustrate the possible nested clade designs corresponding to the scenario G (a) and H (b) when the homoplasic site 443 was excluded; the root between H02/H07 and H16 (black bar) was identified *via* maximum parsimony outgroup rooting.

This reduction in the number of haplotypes is due to the fact that H03, H07 and H09 became identical to H01, H02 and H04, respectively (Figure 3.7). However, three distinct haplotype groups, one each for Western, Central and Eastern Europe, still exist.

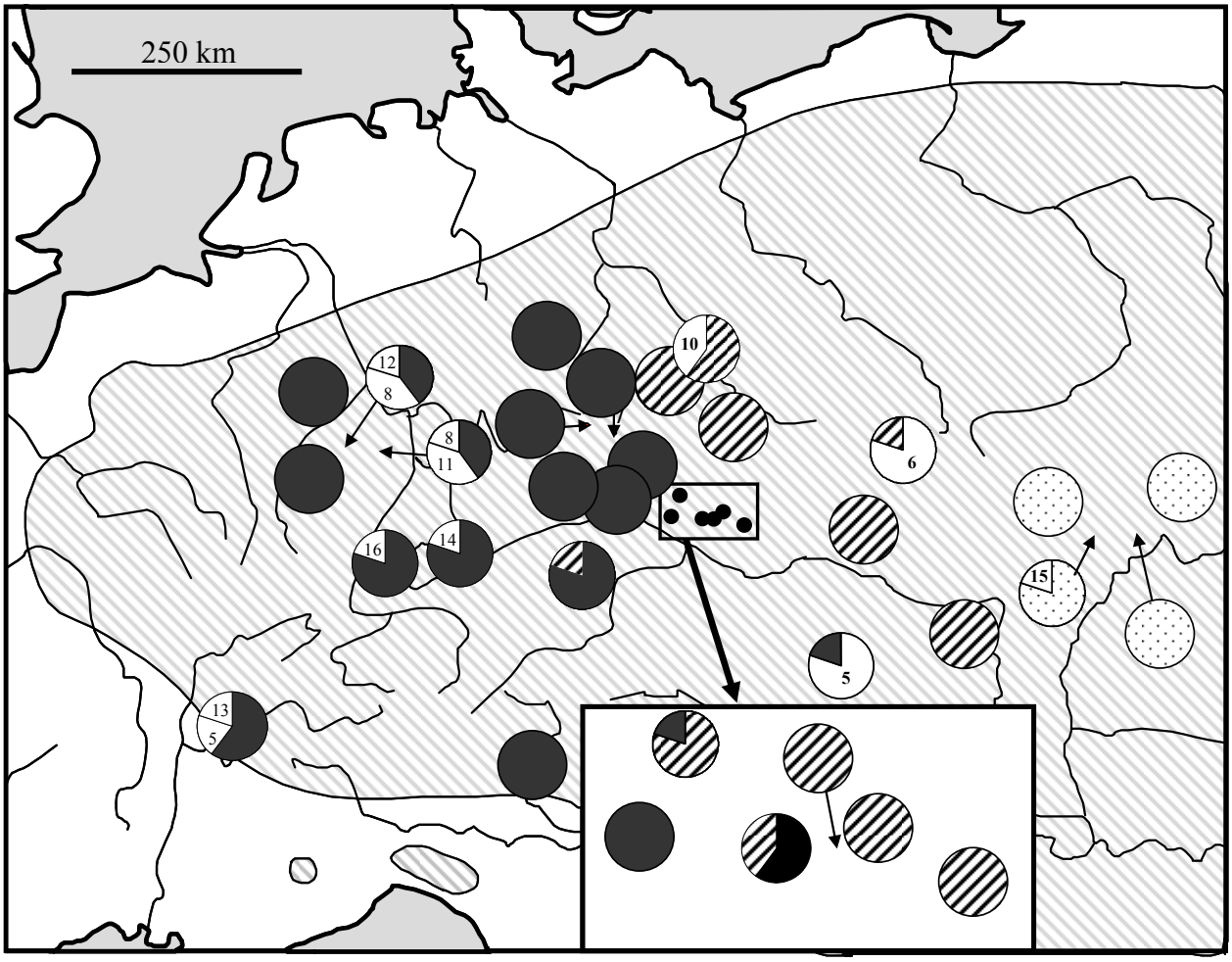


Figure 3.7: Geographic distribution of COI haplotypes of *E. medusa* when the homoplastic site 443 was excluded; black=haplotypes 1 and 3, hatched=haplotypes 2 and 7, spotted=haplotypes 4 and 9; remaining haplotypes are marked by numbers.

The resolved tree without homoplastically evolving sites allowed for two alternative approaches for delimiting 2nd order clades (inlets in Figure 3.6). The two resulting NCPA runs (scenarios G and H) were included in the comparison with previous hypotheses on the evolution of these butterflies (Table 3.4; Appendix 3). Again I tried to conform both scenarios to a consensus scenario but the 2nd clade levels of scenarios G and H harbour different haplotypes.

Table 3.4: Results of the inference key for the resolved network under the hypothesis of homoplasy; the homoplasic site 443 was excluded.

Scenario G

Clade	Inferred phylogeographic scenario
clade 2-1	1-19-20-2-11_b -$YES_{range\ expansion}$-12-NO → contiguous range expansion
clade 2-2	1-2 $_{a,b}$-3-4-NO → restricted gene flow with isolation by distance
total clade	1-2-11_b-$YES_{range\ expansion}$-12-13- YES → past fragmentation followed by range expansion

Scenario H

Clade	Inferred phylogeographic scenario
clade 2-1	1-2-11_b- $YES_{range\ expansion}$-12-13- YES → past fragmentation followed by range expansion
clade 2-2	1-2-Tip/interior status cannot be determined-inconclusive outcome
total clade	1-2-$11_{a,b,c}$- $YES_{range\ expansion}$-12-NO → contiguous range expansion

Moreover, these clades describe non-overlapping geographical areas. Thus no consensus conclusion could be assumed at this clade level. The total clade concerns a comparable geographical area (see discussion chapter 3-4) and thus allows for a consensus conclusion even if the sub-clade composition differs in both scenarios. However, it was not possible to associate the geographical location and the lineages meeting by the consensual event.

3.4. Discussion

3.4.1. The allozyme scenario

The allozyme data of Schmitt and Seitz (2001a) imply that *E. medusa* has had two major differentiation centers (west and north-east of the Alps) and two smaller centers of differentiation at the southern and south-eastern margins of the Alpine glaciers. The remarkable substructures in the two major lineages imply one or two expansion-retraction cycles during the late Pleistocene after the Last Glacial Maximum (LGM) prior to the final postglacial range expansion with most probably three sub-centres in the western lineages

(Schmitt and Seitz 2001a). This young phylogeographic scenario based on allozymes can be summarised with six detectable major events (chapter 2; Table 3.5).

Table 3.5: Phylogeographic scenario of *E. medusa* evolution in Central Europe based on nuclear (Schmitt and Seitz 2001a) and mitochondrial data under the hypothesis of recombination; ybp=years before present.

Event number	Genetic structure	Time (ybp)	Nuclear data	Mitochondrial data
1-a)	**Fragmentation:** Four major genetic lineages (Western, Southern Alps, Western Hungarian and Eastern)	70,000	X	
1-b)	**Restricted gene flow with isolation by distance**			X
2-a)	**Fragmentation:** Two sub-lineages in the Western lineage (France+ SW Germany and rest of Germany)	18,000 or 13,500	X	
2-b)	**Fragmentation:** Two sub-lineages in the eastern lineage (Slovakia+ NE Hungary and Czech Republic)	18,000 or 13,500	X	
2-c)	**Restricted gene flow with isolation by distance**			X
2-d)	**Range expansion**			X
3-a)	**Fragmentation:** Two secondary sub-lineages in the rest of Germany sub-lineage (W Rhineland-Palatinate+ Saarland and E Rhineland-Palatinate+ Thuringia+ Bavaria)	11,500	X	X
3-b)	**Fragmentation:** Two secondary sub-lineages in the Czech Republic sub-lineage (Bohemia against Moravia)	11,500 or younger	X	
3-c)	**Range expansion:** to different degrees in all sub-lineages	since 10,500	X	

A first vicariance caused by the onset of the Würm ice age has led to the formation of the four major lineages. Vicariance during the LGM (22,000 ybp) or Older Dryas (13,500 ybp) formed two sub-lineages in the Western lineage (event 2-a) and two sub-lineages in the Eastern

lineage (event 2-b). Later, during the Younger Dryas (11,500 ybp), the German sub-lineage most probably became fragmented into two secondary sub-lineages (event 3-a). The two secondary sub-lineages in the Czech Republic have evolved due to (i) vicariance during the same time period, or (ii) differentiation during a postglacial expansion process from the more eastern Moravia to the more western Bohemia (event 3-b). Since 10,500 ybp, a range expansion to different degrees in all sub-lineages is assumed (event 3-c).

In the following sections I will complete this allozyme scenario with the additional information given by the NCPA consensus scenario of my ten alternative NCPA scenarios gained under adjustment for recombination and homoplasy, respectively. For comparison with the allozyme scenario, I assume that the total clade of NCPAs corresponds temporally to the oldest event 1-a of the allozyme scenario, the second clade level to the second events 2-a,b and the first clade level to the three more recent events 3-a to 3-c.

3.4.2. Adjustment for recombination (scenarios A-F)

At this point I want to mention that recombinant haplotypes may also arise from *in vitro* recombination (Pääbo et al. 1990). It is assumed that during the PCR reaction the DNA polymerase may jump from one molecule to another, leading to hybrid products. However Ladoukakis and Zouros (2001a) demonstrated that such errors (single nucleotide substitution) occur at a very low frequency. I found 16 and 3 specimens bearing recombinants H04 and H07, respectively. I therefore conclude that the rare event of *in vitro* recombination does not account for the pattern I found.

The clade 1-6 inferred past fragmentation followed by range expansion in scenarios E, G and H, each with different NCPA distance values. The consensus at this clade level is fragmentation followed by range expansion (Table 3.2). The haplotypes bearing this clade are localised in Central Europe, supporting the fragmentation event in the rest of the German sub-lineages of the allozyme scenario.

I inferred for clade 2-2 contiguous range expansion in all six scenarios. NCPA statistical distance values for scenario A are identical to scenarios C and E. The scenarios B and D also have the same statistical distance values. Therefore only three scenarios are to be considered to infer contiguous range expansion that I compiled as a range expansion event in my NCPA consensus scenario. The clade 2-3 inferred restricted gene flow with isolation by distance in scenarios C, E and F, each with different distance values. The retained consensus event is

attributed to restricted gene flow. Both consensus events at the 2nd clade level complete the events 2-a and 2-b of the allozyme scenario.
The total clade level inferred restricted gene flow with isolation by distance over all scenarios except scenario H. These five scenarios (A to E) all have different NCPA statistical distance values. The retained consensus 'restricted gene flow' for this clade level gives additional information to the fragmentation event assumed by the nuclear data.

3.4.3. Adjustment for homoplasy (scenarios G-H)

Scenario G and H, both adjusted for homoplasy, do not allow an NCPA consensus conclusion for the total clade because the inferences resulting at this clade level show no overlap. Therefore it is not possible to compile these scenarios into a consensus scenario. Under this hypothesis no consistent phylogeograhic scenario is assumed because of the gap between the two resulting scenarios. Maybe one of these scenarios reflects - or is close to - the true evolutionary history of *E. medusa,* however, there is no way to distinguish it.

3.4.4. Comparison among adjustments

Any kind of adjustment of mtDNA alignment for disturbing effects such as recombination or homoplasy will remove information from the original data set. However, there is no need to assume that mtDNA data would necessarily give the same and complete answer as nuclear data and *vice versa*. Both will at best exemplify parts of the real history.
Recombination poses a severe problem to phylogeographic analysis. Recombinants harbour information about the evolutionary histories of both parents, thus breaking the pure matrelinear genealogy. I here proposed a way to incorporate information about recombinants into the NCPA analyses to enable haplotype network resolution without introducing too much bias into the phylogeographic reconstruction: (i) identification of recombinants following the "age rule", the "position rule" and the "geographical rule" and (ii) their assignment to both parents. Further analyses of gene trees with ring-like ambiguities may show broader relevance of this approach. I would like to emphasize that adding recombinants simultaneously to both parental haplotypes inevitably links the parents geographically through the localities where recombinants occur. This does not appear to be problematic since geographic affinity of the parents is required for recombination to occur. However, geographic co-occurrence can be assumed only when recombination was recent.

Assuming homoplasy, the exclusion of highly variable sites from an alignment prior to NCPA also leads to incomplete conclusions as compared to allozyme data. This does not necessarily mean that the ring-shaped ambiguities are not due to homoplasy, even if the occurrence of one and the same mutation in non-related lineages is presumed unlikely when haplotype diversity is rather low, as is usually the case in intraspecific analyses. Here I argue that the incomplete NCPA results under adjustment for homoplasy are caused at least partly by the loss of information that is still inherent in the excluded sites.

3.4.5. Conclusions

My results show that the scenario that complements the allozyme scenario is the one under the recombination hypothesis. No consensus scenario under correction for homoplasy is possible. Here I indicate that the genetic variability is lower under the homoplasy hypothesis, but this does not affect the inference within these scenarios that delivers convergent conclusions. This reduces its importance as a possible explanation for the ring-shaped ambiguities, at least in my study.

The assumption that recombination could create the ring ambiguities in my haplotype network is not implausible. Hence, the recombinant haplotype H07 exist only in the contact zone of the central and western haplotype groups, supporting the plausibility of this process to occur. Moreover, Lunt and Hyman (1997) provided evidence for the presence of recombination in the mtDNA of the phytonematode *Meloidogyne javanica* and Thyagarajan et al. (1996) have shown the enzymes involved in this process of recombination to exist in human mitochondria. Ladoukakis and Zouros (2001a) have demonstrated mtDNA recombination in the mussel *Mytilus galloprovincialis,* where heteroplasmy is the rule in males. In another study, based on published sequences, they have shown recombination to exist in three different animal species (a crustacean, an amphibian and a mammal) with standard maternal mtDNA inheritance (Ladoukakis and Zouros 2001b). Mitochondrial recombination was also detected in another mussel species, *Mytilus trossulus* (Burzynski et al. 2003), and also in a vertebrate, the flatfish *Platichthys flesus* (Hoarau et al. 2002). A broad survey of recombination in animal mitochondria concluded that laboratory error may explain some cases of recombination, but also qualified recombination as a process occurring moderately frequently within and between species (Piganeau et al. 2004). Thus, recombination of mtDNA may be more frequent than previously expected (Rokas et al. 2003).

Both recombination and homoplasy may produce ring structures in a haplotype tree that do not allow for prevent an unambiguous reconstruction of a nested clade design as is the necessary prerequisite for NCPA analysis. Hence, the NCPA outcome may still be insufficient when compared with a population history inferred from nuclear data. This does not *a priori* make mtDNA analyses with signs of recombination or homoplasy unsuitable for the reconstruction of evolutionary histories. Rather, statistical methods have to be adapted to deal with recombination and homoplasy occurring in natural populations. As a first step, I think it is necessary to broadly evaluate the impact of recombination and homoplasy on phylogeographic reconstruction and, more specifically, on intraspecific studies using network methods (such as minimum spanning trees) that are preferentially used at the population level (Cassens et al. 2003).

3.5. Summary

Analysis of haplotype networks may become hampered by the appearance of ring-shaped haplotype ambiguities (loops). These may be caused by recombination, homoplastic mutations or simply through sequencing mistakes. Using a 2-ring loop that emerged in a minimum spanning network from a phylogeographic study of the western Palaearctic butterfly species *E. medusa*, I analyse the impact of this loop on nested clade phylogeographic analysis (NCPA). I consider both recombination and homoplasy as possible sources of the ring-shaped ambiguity. I first outline a strategy to distinguish between recombinant and parental haplotypes. In a second approach I suggest a simple procedure for adjusting the alignment for homoplasy prior to network reconstruction. Consensus phylogeographic scenarios from both approaches are finally overlaid on a published allozyme-based evolutionary history for *E. medusa*.

My results show that the scenario that corroborates the allozyme scenario is the recombination hypothesis. No consensus scenario under correction for homoplasy is possible, scaling down its importance as a possible explanation for the emergence of my ring-shaped ambiguities.

4. Phylogeography of the Woodland Ringlet (*Erebia medusa*) based on the highly variable control region

4.1. Introduction

The discipline of pylogeography expanded rapidly during the last two decades due to the invention of new molecular tools and the standardised use of DNA markers. It has developed into an important field of biological research that links phylogenetic reconstruction to biogeography. Avise (2000) defined phylogeography as the "principles and processes governing the geographic distributions of genealogical lineages, especially within and among closely related species". Hence, phylogeography mainly addresses questions of intra-specific relations, for which standard analytical tools, mainly designed for between species or higher taxa phylogenetic analyses, are handicapped by a lack of resolving power.

Phylogeographic studies often invoke the glacial refugia hypothesis (Hewitt 1999, 2000) to interpret patterns of geographical haplotype distribution. During the Pleistocene, severe climatic oscillations induced successive range contractions and expansions of the temperate zone flora and fauna (Avise 2000, Hewitt 2004). In the Palaearctic two different groups of species exist: (i) species considered as arctic and/or alpine element expanding during the cold glacial periods and retreating into refugia during the warm interglacials, (ii) species which during cold periods (glacial periods) survived in refugia from which they expanded as soon as the climate warmed again (i.e., Mediterranean or Siberian elements *sensu* de Lattin, 1967) (Hewitt 2004). Such range oscillations are expected to have left their imprint in the genetic diversity of extant populations since species evolved during isolation in their refugia leading to the formation of divergent genetic lineages.

As a model organisms, I selected the Woodland Ringlet, *E. medusa*. This butterfly is a western Palaearctic species, which typically lives in different types of grassland (SBN 1987, Ebert and Rennwald 1991, Schmitt 2002). It is distributed throughout temperate Eurasia, but is missing in the Euatlantic and Eumediterranean regions of Europe as well as in Scandinavia (Tolman and Lewington 1997, Kudrna 2002). The species was formerly considered a Siberian element which survived the last glaciations in an eastern Palearctic refugium (southern Siberia) and which postglacially re-expanded westward into Europe (de Lattin 1957, Varga 1977). If so, no major genetic differentiation should be found among European *E. medusa*

populations due to a continuous loss of genetic diversity during range expansion (founder effect). However, allozyme data supported an evolutionary scenario of four well differentiated lineages of *E. medusa* existing in Europe, indicating glacial survival in several independent European refugia (Schmitt and Seitz 2001a). Subsequent analyses of partial mitochondrial cytochrome oxydase subunit one (COI) sequences added further support for this multiple refugia hypothesis (chapter 2 and 3). Both nuclear and mitochondrial markers consistently supported: *(i*) a first vicariance at the onset of the Würm glaciation, leading to the formation of several major lineages, (*ii*) further differentiation of German populations during the Younger Dryas, and (*iii*) postglacial range expansions in most of the genetic lineages. In addition, the allozyme data (Schmitt and Seitz 2001a) assumed the colonisation of western-central Europe by the western lineage and the existence of two different sub-lineages in the eastern lineages. The overall phylogeographic scenario suggests a recent differentiation into extant European *E. medusa* populations, probably during the Würm ice age (Schmitt and Seitz 2001a).

In contrast to the protein coding COI gene, which belongs to the more slowly evolving genes of the mitochondrial genome (Simon et al. 1994), the non-coding control region is considered a highly variable mitochondrial region. It therefore should be able to detect marginal divergences of populations. This may make it a marker of choice for the analysis of comparatively recent differentiation events.

The control region is also called the D-loop region in verterbrates. In some inverterbrates, such as insects, it is composed of more than 85% of A/T nucleotides ("A/T rich region"; Zhang and Hewitt 1997). It is believed to be involved in the regulation of transcription and control of mitochondrial DNA (mtDNA) replication (Shadel and Clayton 1997) and constitutes the major non-coding region of this molecule. Like in nuclear non-coding regions, the presumed lack of functional constraints explains its high variability (Simon et al. 1994). Therefore, Zhang and Hewitt (1997) suggested that the control region may be a suitable marker for the study of variation even at the population level. However, use of the control region for phylogenetics and phylogeographic reconstructions remain controversial and has even been questioned in insects. A first application to seven closely related butterfly species (Lycaenidae) showed too little variation to resolve their phylogenetic relationships (Taylor et al. 1993). In contrast, in a more recent study of 29 species ranging across five families of butterflies and two Proxidae moths, the control region proved its value also at the intraspecific level, e.g., among *Erebia* populations (*E. palarica* and *E. triaria*; Vila and Björklund 2004).

In this chapter I try to gain additional insight into the phylogeographic history of the Woodland Ringlet using sequences of the mitochondrial control region. I analyse the genetic variation among populations of *E. medusa* over large parts of central and western Europe and infer phylogeographic scenarios using the nested clade phylogeographic analysis (NCPA). This allows us to detect, without *a priori* assumptions, non-random geographical haplotype associations and to propose the best phylogeographical scenario to explain it (Templeton et al. 1995). Based on a consensus (cross validation) among previously published and new information, my data allow us to describe an improved evolutionary scenario of *E. medusa* in central and western Europe.

4.2. Material and methods

4.2.1. Sequencing of mtDNA

I used the same samples and specimens from 32 populations as in previous chapters 2 and 3 (Appendix 1). Five specimens per locality (except for Klentnice; four specimens) were sequenced. PCR amplification of the control regions was performed in 25 µl volume containing 1 µl DNA extract, 1 µl of each primer (15 pmol μL^{-1}), 1 µl $MgCl_2$ and 21 µl distilled water (Carl Roth GmbH and Co). The PCR program contained a touch down step and started with a denaturation at 95 °C step for 2 min, 35 further cycles: denaturation at 94 °C for 60 s, annealing with a touch down step at 61 °C for 90 s (gradient = -0.2 °C) to a final temperature of 54 °C and a final extension step at 65 °C for 60 s. PCR products were run on a 1.4% agarose gel and checked by visualisation under UV light. The primers J6 (Zhang et al. 1995) and Lep 12S (Taylor et al. 1993) amplified a circa 800 bp fragment. Positive PCR products were purified with Roche High Pure PCR purification kit and used for single stranded sequencing with the specifically designed nested primer Seq Med Met (5'-TATATGAGGTRTGAGCCCAAAAGC) and the following program: denaturation at 96 °C for 1 min, 25 cycles of denaturation at 96 °C for 30 s, annealing at 45 °C for 15 s and extension at 72 °C for 4 min. PCR products were sequenced with an automatic sequencer ABI 377 A. I finally obtained a 405 bp long fragment for all samples. They were aligned using the Sequence Navigator software (Applied Biosystems); the alignment was subsequently refined by eye**.**

4.2.2. Nested clade phylogeographic analysis

Geographical structuring of haplotypes can be due to recurrent events such as restricted gene flow or historical events such as fragmentation, range expansion or colonization (Templeton et al. 1995). Nested clade phylogeographic analysis (NCPA), commonly used to analyse intraspecific phylogeography, identifies processes which best explain non-random geographical association between the position of a haplotype in a gene tree and its geographical distribution. The null hypothesis tested by this method is: there is no such geographical association.

The first step was the calculation of a minimum spanning haplotype network on the basis of statistical parsimony with TCS 1.18 (Clement et al. 2000). This network was unrooted. I therefore tried to identify the root by adding two outgroups, *E. gorge* and *E. epiphron*. However, due to a pronounced divergence among these species and *E. medusa* TCS could not link either outgroup to the network. The position of the root was therefore determined *via* a maximum parsimony (MP) analysis. I calculated a MP tree with PAUP* version 4.0b (Swofford 2001), including further hierarchical outgroups: *Maniola jurtina* and *Coenonympha pamphilus* as representatives of satyrine genera closely related to *Erebia*, and *Melitaea latonigena*, a Nymphalidae. The latter was defined as outgroup.

The minimum spanning network was then converted into a nested clade design. The different factors cited above have different expectations regarding the relationship between the genealogical distances and geographical distances between haplotypes. I calculated two types of geographical distances with Geodis 2.2 (Posada et al. 2000): (i) clade distances which measure how geographically widespread the haplotypes within a clade are and (ii) nested clade distance which measure how far the haplotypes of one clade are from the haplotypes of the sister clades in the higher nesting level. Within each tested clade, statistical comparison of clade and nested clade distances for tip and interior subclades is calculated to look for patterns characteristic of restricted gene flow, fragmentation, range expansion or colonization. This program tests at the 5% level of significance the distribution of these distance measures under the null hypothesis of no geographical association, applying 1,000 random distributions of all clade members on their respective locations (Templeton et al. 1995).

4.2.3. Solving haplotype networks with ring-shaped ambiguities

During the reconstruction of the minimum spanning network eventually ring-shaped ambiguities (loops) may appear. They illustrate alternative evolutionary pathways of equal probability which, per definition, are mutually exclusive. Such ambiguities may be caused by recombination, homoplastic mutations or simply through sequencing mistakes (Templeton and Sing 1993). To exclude sequencing artefacts, I carefully checked my sequences at all sites involved in the ambiguity. Additional backward sequencing with a specifically designed primer E1E2 (5'-CATGATAATCCGAATACAGTTC) also ensured correct sequencing and interpretation.

Ring-shaped ambiguities should be resolved prior to the construction of a hierarchy of nested clades. Therefore, Pfenninger and Posada (2002) suggested rules to decide among alternative evolutionary trajectories, based on three criteria: haplotype frequency, network topology and geographical distribution. The haplotype frequency criterion is based on the expectation that connection to a more frequent haplotype is more likely than connection to a singleton (haplotype with a frequency of one). The topology criterion argues that the connection of any haplotype to an interior haplotype is more likely than its connections to a tip. Finally, the geographical criterion emerges from the idea that connections among haplotype from the same population or region are more likely than those between haplotypes from distant populations. This set of criteria allows selection among mutually exclusive haplotype connections and enables reconstruction of the most probable evolutionary pathway.

I proposed in chapter 3 a different strategy for resolving haplotype rings. Under the assumptions of recombination three rules can be applied to identify recombinant haplotypes in a haplotype network: (*i*) following the age rule, recombinant haplotypes must *per definitionem* be younger than their presumed parents, mirrored in a lower number of descendant haplotypes (age rule), (*ii*) each crossing over must produce two recombinant haplotypes which always sit opposite in a haplotype ring (position rule), (*iii*) recombination between haplotypes can only occur in sympatry, either in terms of populations or regions (geographic rule). These rules allowed in a case study on *E. medusa* COI haplotypes the identification of three recombinant haplotypes, whose exclusion completely resolved two previously existing haplotype rings.

Under the hypothesis of homoplasy, I suggested in chapter 3 exclusion of homoplastically evolving sites prior to haplotype network construction. To identify such sites I suggested calculation of the consistency index CI (Kluge and Farris 1969). This index quantifies the amount of additional substitutions relative to the number of character states needed to explain

evolution of a variable character on a given tree. It equals 1 when there is no homoplasy and it decreases toward zero as the degree of homoplasy increases. In general, my recently proposed rules (chapter 3) try to *a priori* adjust for processes that produce haplotype network ambiguities. Pfenninger and Posada's (2002) rules rather accept ambiguities as they are and help to *a posteriori* find plausible arguments for retaining one haploytpe connection over the other.

In the present chapter, I try to apply Pfenninger and Posada's (2002) rules as well as my rules presented in chapter 3 to resolve the complex haplotype network that emerged from control region analysis. To account for homoplasy I mapped all variable sites on a neighbour joining (NJ) tree calculated with PAUP* version 4.0b (Swofford 2001) and used the best-fitting substitution model (see chapter 3 for details). CI values were calculated with MacClade version 3 (Maddison and Maddison 1992). I considered a site as evolving homoplastically when the CI value was below 1. These treatments resulted in several alternative nested clade designs and hence in alternative spatio-temporal scenarios of haplotype evolution. Such alternatives usually vary only slightly in clade formation and geographical assignment of haplotypes. I therefore assume that strong phylogeographic signals will produce largely consistent inferences, irrespective of the treatment applied for ring resolution.

4.3. Results

4.3.1. Geographical distribution of haplotypes

The control region contained both SNPs and indels. For the 159 sequenced individuals, I obtained 34 haplotypes (24 when not considering indels). Haplotypes H01, H02, and H05 are represented in 19.5%, 13% and 10.5% of the sequenced individuals, respectively. Frequencies of haplotypes H03 and H06 are 6% and 5.5%, respectively. All remaining haplotypes occur at rates less than 5%. Some haplotypes such as H12 are locally restricted, whereas others such as H01 or H02 are widespread (Figure 4.1) with the former being present in twelve populations.

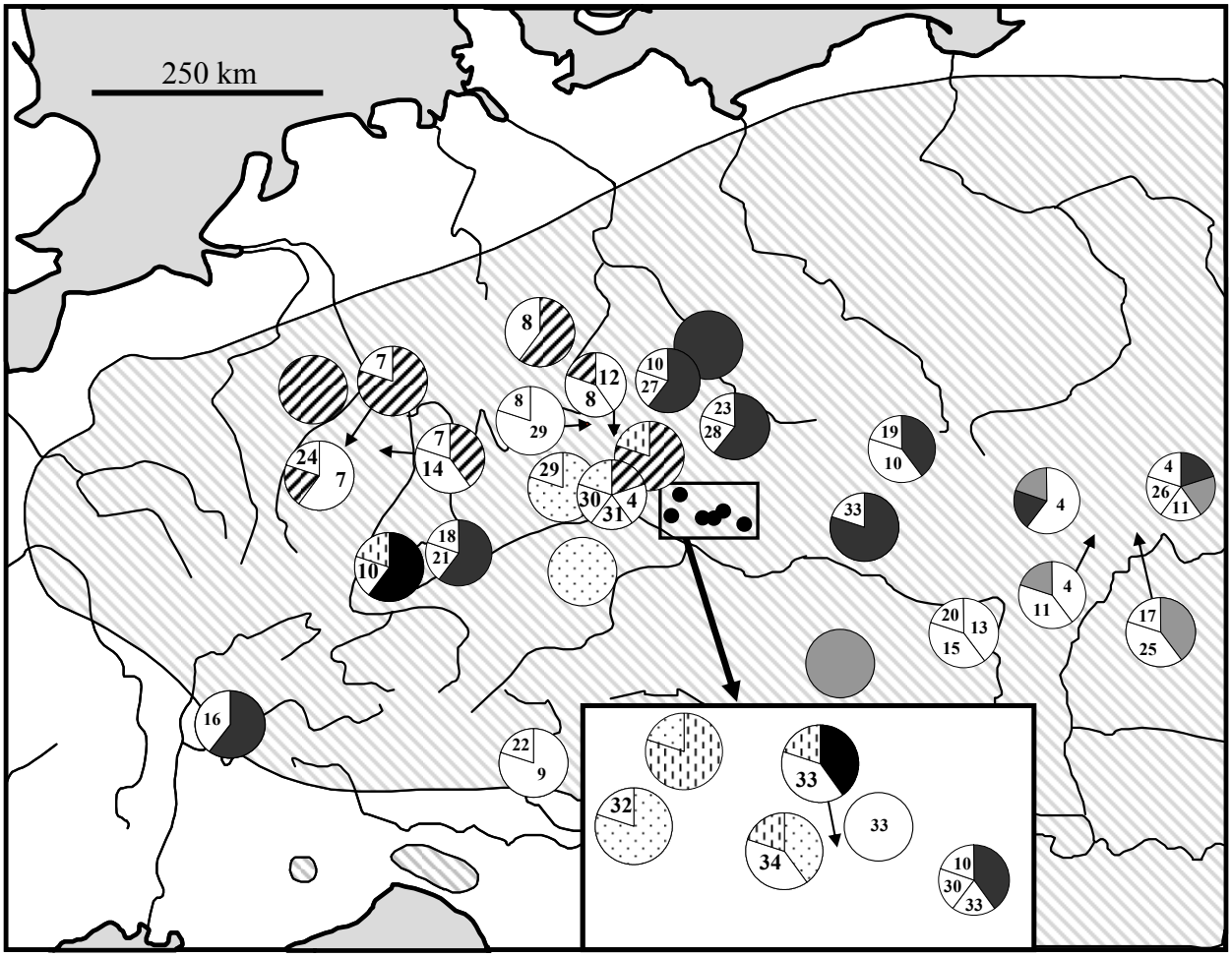


Figure 4.1: Geographic distribution of *E. medusa* control region haplotypes. Black=haplotype 1, hatched=haplotype 2, grey=haplotype 3, spotted=haplotype 5, vertical chequered=haplotype 6; remaining haplotypes are marked by numbers.

There emerges no general pattern of haplotype distribution in Western, Central and Eastern European lineages. However, several haplotypes were endemic for some regions (H02 for Germany; H03 for Hungary and Slovakia; H05 for Bavaria; H06 for the Bohemian-Bavarian border mountains and the Black Forest; H07 for Rhineland-Palatinate and the Saarland (western Germany); H08 for Thuringia and northern Bavaria; H11 for Slovakia; H29 for Bavaria; H30 for Bavaria and southern Bohemia; H33 for the Czech Republic). Some twenty haplotypes were private haplotypes of single populations. Interestingly, all haplotypes found in the population of the Monte Baldo and Gánt are endemic for their respective population.

4.3.2. Minimum spanning tree

a- Identical treatment of indels and substitution

The resulting minimum spanning tree contains eight rings (A to H in Figure 4.2), four of which comprise only three haplotypes and are formed by multiple substitutions at one base position (122, 183 and 191, respectively).

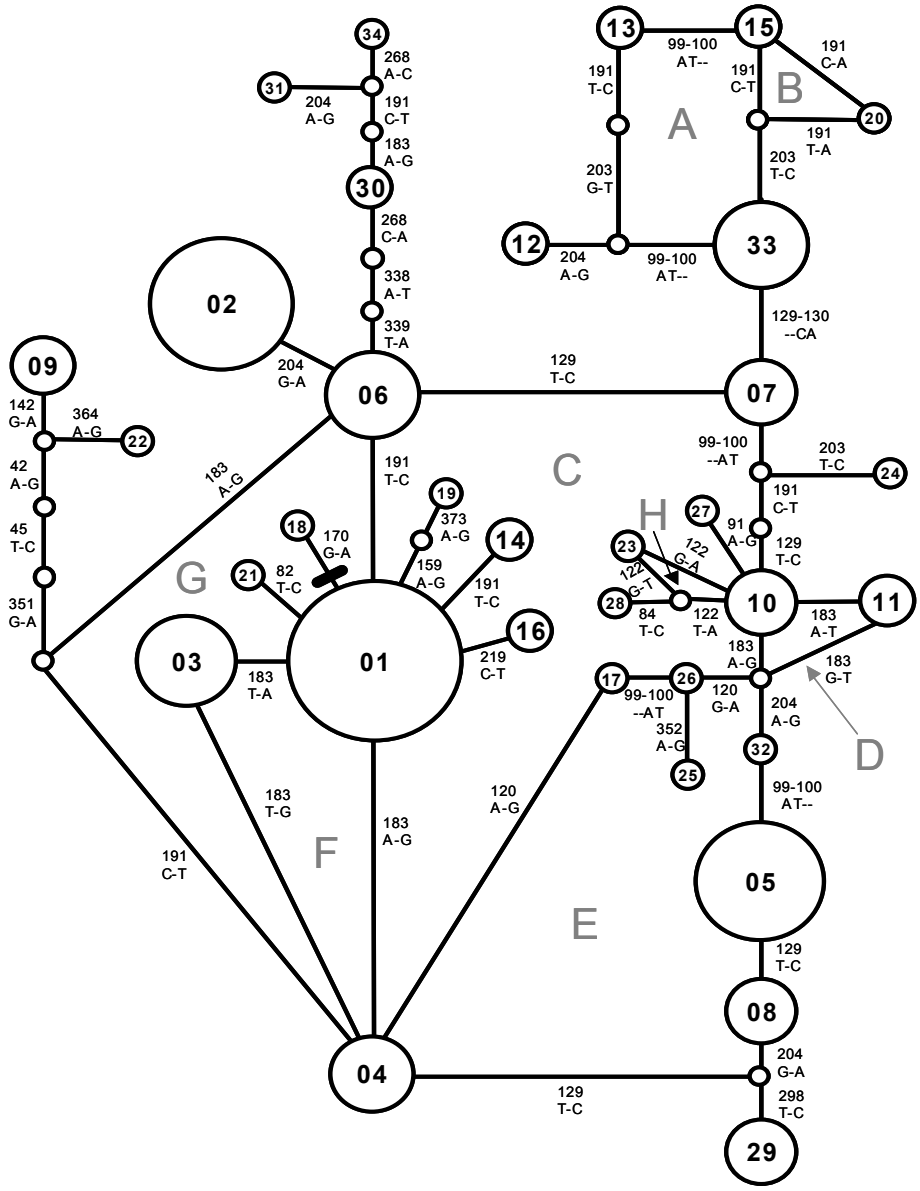


Figure 4.2: Minimum spanning network of *E. medusa* control region haplotypes; circles are proportional in size to the corresponding haplotype frequency; small unnumbered circles represent hypothetical haplotypes not found; substitutions are mapped on the tree and indel positions are fused into one mutation; ring-shaped ambiguities are named A to H. The root between H01 and H18 was identified *via* maximum parsimony outgroup rooting.

These multiple ring-shaped ambiguities in the control region haplotype network obscure a straightforward inference of the evolutionary history of *E. medusa* in Europe. They could not be incorporated into a hierarchy of nested clades through the application of the special rules proposed by Templeton and Sing's (1993). Grouping all haplotypes that are involved in a ring into a single clade would have entailed a large number of alternative 1st order clades, with a continuous augmentation at higher clade levels. I therefore preferred to break ring-shaped ambiguities using different approaches and to extract phylogeographic information from the resulting and simplified haplotype networks.

The rules of Pfenninger and Posada (2002) failed to break most of these ring-shaped ambiguities. In rings A and B (Figure 4.2) the lack of information about the three hypothetical haplotypes does not allow applying any of their criteria. Likewise, rings D and G also each contain one hypothetical haplotype. In rings C and E, H17 could be connected to H04, based on the frequency criterion. However, H26 remains ambiguous because it is linked to a hypothetical haplotype and H17, the latter of which is found at the same frequency as H26 (frequency criterion failed). In ring F, H03 is almost equally frequent as H04, hence the frequency criterion again failed to solve the ambiguity. H01, H03 and H04 emerged together in two populations (Podlesok and Snina), averting the application of the geographical criterion. Moreover, because the interior or tip status of H03 could not be determined, none of the criteria allows to solve this loop.

My own strategy (chapter 3) to solve ring-shaped ambiguities under the assumption of recombination also failed. The several juxtaposed loops did not allow identification of potential recombinants since application of rules cited in chapter 3 ("age rule", "geographical rule" and "position rule") is not possible. The rings A, B, C, D, E, H and G contain hypothetical haplotypes which constitute a lack of information; they can not be identified as being either recombinant or parental. Consequently, in these loops application of my own rules is not possible. In ring F, H01, H03 and H04 appear in the same geographical area but their distribution patterns do not resemble that of a zone of contact. This does not allow the identification of any of these haplotypes as recombinants. However, recombination can not *per se* be excluded as the origin of these ring ambiguities.

Only the approach suggested in chapter 3 for correction under the hypothesis of homoplasy was applicable to my control region data. Eight variable sites have a CI value below one and can therefore be considered homoplastic (Table 4.1).

Table 4.1: Consistency index (CI) for the variable sites of the mitochondrial control region of *E. medusa*. Number of character states and character changes are indicated.

Sites	Mutations	States	Changes	CI
1	42	2	1	1.00
2	45	2	1	1.00
3	82	2	1	1.00
4	84	2	1	1.00
5	91	2	1	1.00
6	**99**	2	11	**0.09**
7	**100**	2	11	**0.09**
8	120	2	1	1.00
9	122	3	2	1.00
10	**129**	3	8	**0.25**
11	**130**	2	5	**0.20**
12	142	2	1	1.00
13	159	2	1	1.00
14	170	2	1	1.00
15	**183**	3	4	**0.50**
16	**191**	3	4	**0.50**
17	203	2	1	1.00
18	**204**	2	3	**0.33**
19	219	2	1	1.00
20	268	2	1	1.00
21	298	2	1	1.00
22	338	2	1	1.00
23	339	2	1	1.00
24	351	2	1	1.00
25	352	2	1	1.00
26	**364**	2	2	**0.50**
27	373	2	1	1.00

Especially CI values for double indels (99-100 and 129-130) are very low (0.09 and 0.20/0.25, respectively) reflecting a high rate of homoplasic evolution. In fact, these AT indels are situated in an A/T-rich region, with more AT tandems connected to them. They

therefore resemble imperfect microsatellites (Goldstein and Schlötterer 1999), which are known to evolve at a high rate. In a next step I therefore excluded either all homoplasic sites or only the microsatellite-like indels and calculated new minimum spanning trees.

b- Differential treatment of indels and substitution

Exclusion of all homoplastic sites resulted in a star-like tree (Figure 4.3), with H01 being the central and almost ubiquitously common haplotype.

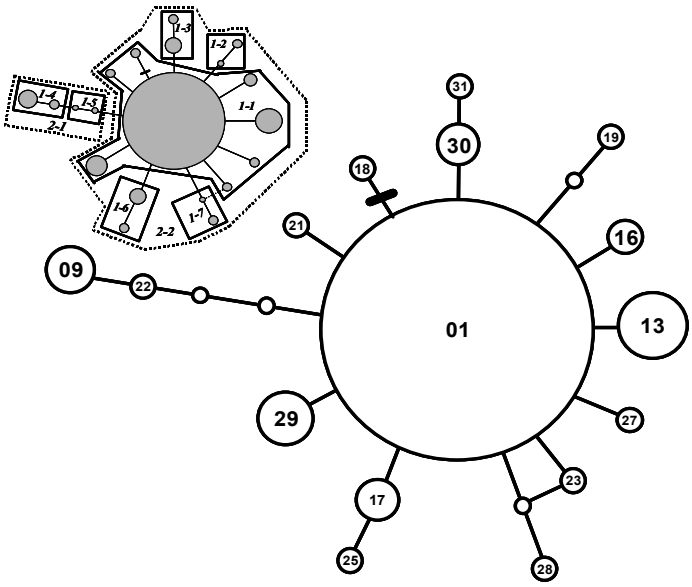


Figure 4.3: Minimum spanning network of *E. medusa* control region haplotypes after deletion of all homoplastic sites; circle areas are proportional to the haplotype frequencies; the insert illustrates the nested clade design. The root between H01 and H18 was identified *via* maximum parsimony outgroup rooting.

A simple nested design was possible, with only one small ring being left. It was solved following the rules of Pfenninger and Posada (2002) assuming the connection of H23 to the more frequent haplotype H01, likely as to the hypothetical haplotype ancestor of H28. An NCPA scenario (A) was inferred from this resolved minimum spanning tree (Table 4.2; for results see Appendix 4).

Table 4.2: Results of the inference key under the hypothesis of homoplasy; exclusion of the eight homoplasic sites.

Scenario A

Clade	Inferred phylogeographic scenario
clade 1-1	1-2-$11_{a,b}$- $YES_{range\ expansion}$-12-13-YES → past fragmentation followed by range expansion
clade 2-2	1-2 $_{a,d}$-3 $_{a,b}$-5-$6_{too\ few\ clades}$-7-YES → restricted gene flow/dispersal but with some long-distance dispersal
total clade	1-19-20-NO → inadequate geographical sampling

Analysing substitutions and indels separately, I calculated two minimum spanning trees: a first one excluding the indels (Figure 4.4) of the sequences and a second one considering only the indels (Figure 4.5 below).

Elimination of only the four sites with tandem indels (99/100 and 129/130) resulted in a more simple tree, however, with still five haplotype rings left (Figure 4.4).

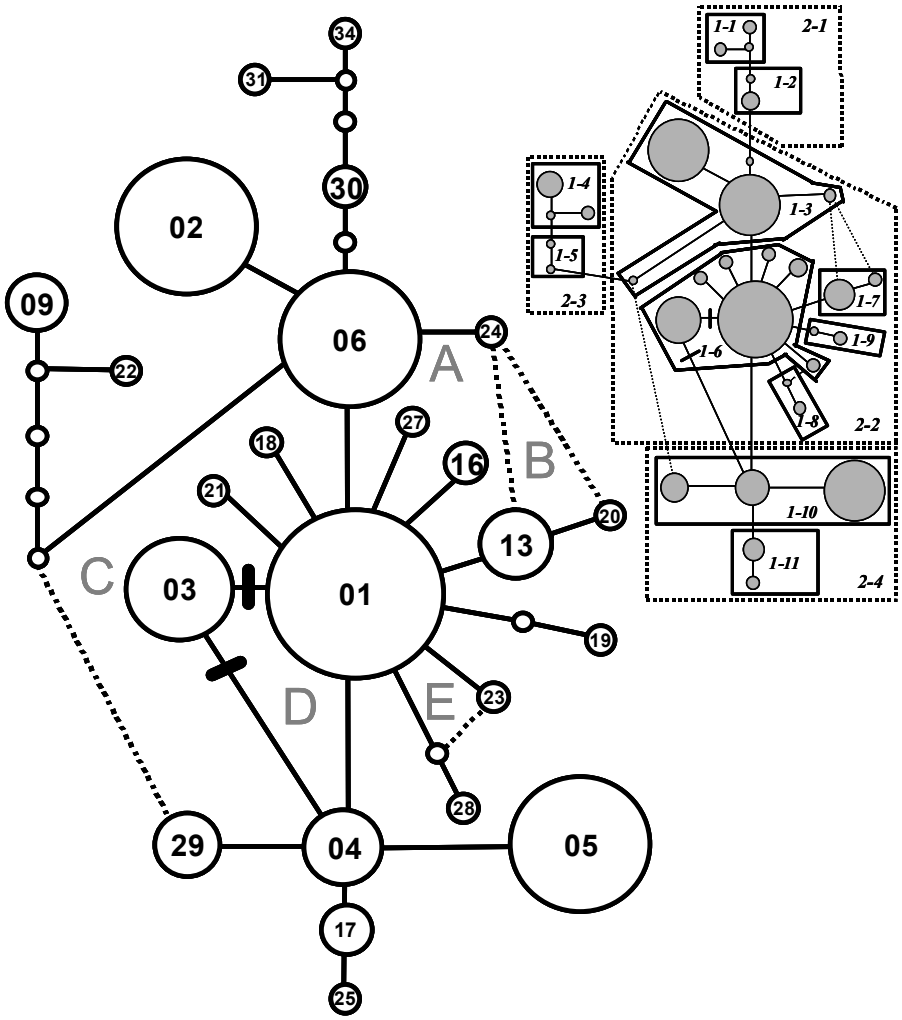


Figure 4.4: Minimum spanning network of *E. medusa* control region haplotypes after omitting the microsatellite-like sites 99/100 and 129/130; circle sizes are proportional to the corresponding haplotype frequencies; dotted lines represent alternative connections that are resolved when applying the rules of Pfenninger and Posada (2002); the insert illustrates the nested clade design. The root between H03 and H01 or H04 was identified *via* maximum parsimony outgroup rooting.

Application of Templeton and Sing's (1993) rules failed here for reasons already mentionned above for the original minimum spanning tree. However, application of Pfenninger and Posada's (2002) rules was now possible. In ring A H24 is linked to the H06, H13 and H20. Following both the frequency and the topology criterion (H06 having a more central position as H13), the link between H06 and H24 seems more likely. In ring B, the H20 is linked to the singleton H24 and to the more frequent H13. The frequency criterion prefers the link between H20 and H13. In loop C, the links between H01-H06 and H01-H04 are both supported by the frequency criterion. Regarding the hypothetical haplotype between H06 and H29 to be the ancestor of the Monte Baldo haplotypes (H09 and H22), the link to the central European H06

seems more likely than to the eastern European H04 *via* H29 (geographical criterion; Figure 4.1). Both alternatives invoke a trans-alpine connection, making both equally likely. Additionally, the frequency criterion reinforces the link between the ancestor of the Monte Baldo haplotypes and H06 since H06 is more frequent than H29. Since the Monte Baldo haplotypes is separated in the gene tree from all other haplotypes by a comparatively large number of steps they constitute an independent second clade with no major impact on the nested clade design and therefore on the NCPA inferences. Loop E is resolved through application of the frequency criterion. The link between H23 and the more frequent haplotype H01 is more likely than a link between H23 and the hypothetical haplotype ancestor of the singleton H28. Finally, only loop D remains unresolved because no criterion (frequency, topology or geographical) allows breaking of any connection between H01, H03 and H04 (see above). Nevertheless, construction of a nested design is possible. However, the status of H03 still remains ambiguous. It could be treated either an interior or a tip haplotype. I therefore considered both possibilities in the NCPA and deduced a scenario (B) from this resolved minimum spanning tree (Table 4.3; for results see Appendix 4).

Table 4.3: Results of the inference key for the ambiguous networks excluding the indels and resolved with the rules given in Pfenninger and Posada (2002).

Scenario B

Clade	Inferred phylogeographic scenario
clade 1-6	1-2-11_b- $YES_{range\ expansion}$-12-13-YES → past fragmentation followed by range expansion
clade 1-10	1-2_a-3-4- NO → restricted gene flow with isolation by distance
clade 2-2	1-2_b-3_b-5-$6_{too\ few\ clade}$-7-YES → restricted gene flow / dispersal but with some long distance dispersal
Clade 2-4	1-2-11_b- $YES_{range\ expansion}$-12-13-YES → past fragmentation followed by range expansion
total clade	1-$2_{a,c,d}$-3-4- NO → restricted gene flow with isolation by distance

Some sites in fact are microsatellite-like dinucleotide indels. Base positions 99-100 and 129-130 are characterised by the presence or absence of the dinucleotides AT and TA, respectively. Position 129 also harbours an additional T-C transition. If I consider only presence or absence of indels (omitting the transition) and calculate a minimum spanning tree that covers all possible presence-absence combinations, a 4-ring ambiguity emerges (Figure 4.1) that is not informative with respect to the evolutionary relationship between these

haplotypes. If I also consider the T-C transition at position 129, two additional combinations occur (Figure 4.2), with the resulting minimum spanning tree remaining unresolved (three haplotype ambiguity). Both approaches clearly show that at least part of the ring-shaped ambiguities in the overall halotype network may well be caused by fast evolving microsatellite-like structures.

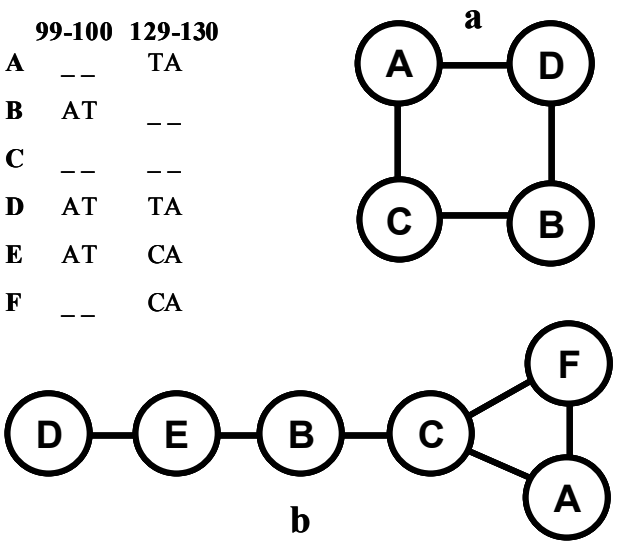


Figure 4.5: Minimum spanning network of *E. medusa* haplotypes based only on the two microsatellite-like sites 99/100 and 129/130; a) haplotype A to D consider only presence/absence of the indels; b) two additional haplotypes E and F exist when considering the T-C transition at position 129.

The geographic distribution of the dinucleotide indels indicates a different geographical pattern for indels and substitutions. Presence-absence of the 99-100 and 129-130 deletions seems to be randomly distributed over the entire study area (Figure 4.6-a,b).

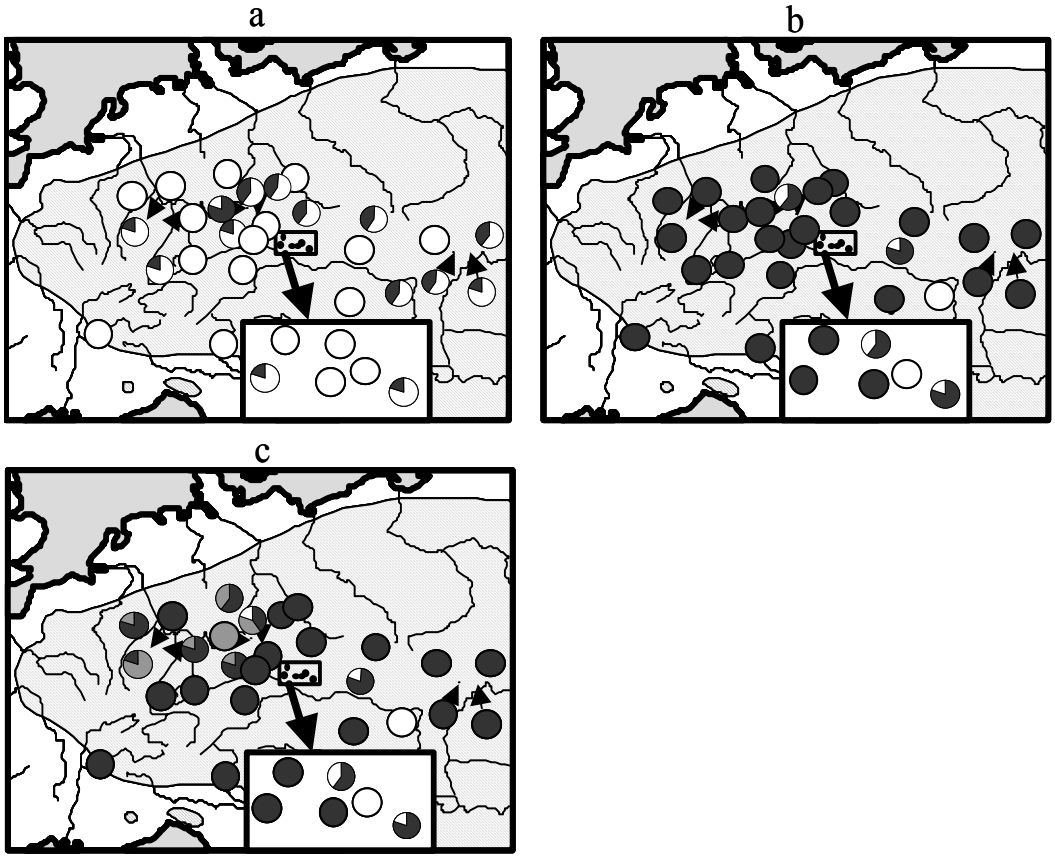


Figure 4.6: Geographic distribution of indels; white = absence of indels (AT or TA), black = presence of indels; grey= presence of the transition; (a) positions 99-100, (b) positions 129-130. (c) Position 129-130 considering the transition for the site 129.

Only the T-C transition at position 129 is restricted to the German populations (Figure 4.6-c). This again argues for a microsatellite-like behaviour of these two sites with frequent change of presence and absence across the whole distribution area. The geographical restriction of the T-C substitution at site 129 indicates that it rather behaves like a SNP with a 'normal' transition rate.

4.4. Discussion

Several ring-shaped haplotype network ambiguities emerged from my control region analysis of European *E. medusa* butterflies that severely hamper deduction of any phylogeographic scenario. Several special rules are available to treat such loops prior to nested clade phylogeographic analysis. However, in my case only few of them proved to be applicable.

In a first approach, I excluded the homoplastic sites and obtained a simplified haplotype network from which hierarchical clade nesting allowed us to deduce phylogeographic

scenario A. In a second approach I specifically focused on microsatellite-like indels. Their omission simplified the original haplotype network in a way that the Pfenninger and Posada (2002) rules became applicable. The resolved gene tree led to the inference of my phylogeographic scenario B.

Due to the temporal hierarchy among clade levels, the total clade of the NCPAs is supposed to correspond to the oldest events of the allozyme scenario, i.e., vicariance due to the onset of the last ice age ca. 70,000 years ago (event 1a,b in Schmitt and Seitz 2001a). The lower clades should address more recent events: Last Glacial Maximum (LGM) (event 2a to 2d) at the second clade level, and Younger Dryas and postglacial (events 3a to 3d) at the first clade level.

At the 2nd clade level geographic areas in my scenarios A and B largely overlap, and for both restricted gene flow with isolation by distance is consistently inferred (Tables 4.2 and 4.3). At the 1st order clade level, populations affected by the fragmentation event in Germany only partially overlapped; in the area of the Czech Republic, such populations are entirely overlapping. Therefore, both scenarios consistently support restricted gene flow during the LGM or Older Dryas (2nd clade level) as well as a more recent fragmentation (albeit only for parts of the German area) followed by range expansion during the Younger Dryas (since 11,500 ybp) (Table 4.4).

Table 4.4: Phylogeographic scenario of *E. medusa* evolution in Central Europe based on nuclear (Schmitt and Seitz 2001a) and mitochondrial data, including COI (chapter 2) and the control region; ybp = years before present.

Event number	**Genetic structure**	**Nuclear data**	**Mitochondrial data**		
Time (ybp)		**Allozymes**	**COI**	**Control region**	
				Scenario A	**Scenario B**
1-a) 70,000	**Fragmentation:** Four major genetic lineages (Western, Southern Alps, Western Hungarian and Eastern)	X			
1-b)	**Restricted gene flow with isolation by distance**		X		X

2-a) 18,000 or 13,500	**Fragmentation:** Two sub-lineages in the Western lineage (France+ SW Germany and rest of Germany)	X			X
2-b)	**Fragmentation:** Two sub-lineages in the eastern lineage (Slovakia+ NE Hungary and Czech Republic)	X			X
2-c)	**Restricted gene flow with isolation by distance**		X	X	X
2-d)	**Range expansion**	X	X		X
3-a) 11,500	**Fragmentation:** Two secondary sub-lineages in the rest of Germany sub-lineage (W Rhineland-Palatinate+ Saaarland and E Rhineland-Palatinate+ Thuringia+ Bavaria)	X	X	X	X
3-b) 11,500 or younger	**Fragmentation:** Two secondary sub-lineages in the Czech Republic sub-lineage (Bohemia against Moravia)	X		X	X
3-c)	**Restricted gene flow with isolation by distance**				X
3-d) since 10,500	**Range expansion:** to different degrees in all sub-lineages	X	X	X	X

Overall, scenario A is less conclusive than scenario B, probably due to the loss of information harboured by the excluded homoplastic sites. Scenario B, namely the exclusion of only the microsatellite-like indels, adds information such as the fragmentation event at the 2nd clade level, which was already inferred from allozyme data (Schmitt and Seitz 2001a).

Comparison with my previous analyses (chapter 2 and 3) shows that general conclusions drawn from the COI data matched by my new control region scenario (Table 4.4). The COI

inference converges with several events deduced from the control region scenario B, such as restricted gene flow during the Würm glaciation and range expansion after the LGM or Older Dryas. The latter is also supported by the allozyme data.

Combination of all available information including already published nuclear data and my two mitochondrial markers allows drawing a much better supported picture of the evolutionary history for *E. medusa* in Europe: (i) a first vicariance, due to the onset of the Würm glaciation, led to the formation of several major lineages, and is mirrored in NCPA by restricted gene flow, (ii) further vicariance led to the formation of two sub-lineages in the western lineage and two sub-lineages in the eastern lineage during the LGM or Older Dryas; this is reflected in NCPA by the restriction of gene flow with isolation by distance during the LGM or Older Dryas, (iii) final vicariance effects resulted in two secondary sub-lineages in the area of Germany and, maybe, in two other secondary sub-lineages in the Czech Republic, (iv) strong range expansions following postglacial warming in most of the genetic lineages. This overall scenario may appear less detailed that that derived from allozyme analysis (Schmitt and Seitz 2001a). The latter discussed in detail spatio-temporal processes as the causes for the contemporary distribution of allozyme alleles of *E. medusa* in Europe. However, in search for consensus I here rely only on scenarios that are statistically supported by several markers. Consequently, the apparent loss of resolution is balanced by an increase in statistical support and hence reliability of inferences.

I currently know only one other phylogeographic study of a lepidopteran species (*Aglais urticae*) using the control region as a marker (Vandewoestijne et al. 2003). In this species haplotype variability of the control region was significantly lower than for COI. The authors impute this to the extreme A/T bias of the mitochondrial control region. They therefore regard it as less useful for population genetic studies of closely related invertebrate taxa since this A/T bias is widespread in this group (Zhang and Hewitt 1997).

Here I could show a notable amount of genetic variability of the control region among populations of *E. medusa* in Europe. In contrast to *A. urticae*, the genetic diversity of this marker is higher than for COI, although its haplotype distribution does not reflect any obvious geographical pattern. When omitting the two homoplastic microsatellite-like sites 99/100 and 129/130, the number of control region haplotypes equals that for COI (15 and 16, respectively). Assuming a general lack of resolving power of the control region therefore seems premature. Application of tools such as correction for homoplasy (chapter 3) and application of Pfenniger's and Posadas's (2002) rules will certainly help interpreting patterns

of geographic control region haplotype distribution even in the presence of ring-shaped ambiguities. Therefore, and despite well-known technical difficulties with sequencing (which is certainly also due to its pronounced A/T bias) the control region may still be a useful tool to investigate intraspecific genetic variation in a phylogeographic context.

4.5. Summary

A phylogeographic scenario for the Woodland Ringlet, *E. medusa*, based on allozymes and part of the COI supports a recent differentiation of European populations. Here I add sequences of the mitochondrial control region which is considered as highly variable and therefore able to detect also marginal divergence of populations. However, the D-loop minimum spanning network contained multiple ring-shaped ambiguities averting a straightforward extraction of the phylogeographic information. I therefore applied two approaches to deal with such loops: (*i*) exclusion of homoplastically evolving sites and (*ii*) exclusion of microsatellite-like indels. I calculated minimum spanning networks for both approaches. Since it was not possible to decide *a priori* which of the resulting phylogeographic scenarios was superior to the other, I compared both to the evolutionary history already known from allozymes for *E. medusa*. I took only those inferences into account that were consistently supported by both approaches.

The combination from the already published nuclear data and from the new mitochondrial data now allows us to draw a more precise picture of the evolutionary history of *E. medusa* in Europe: (*i*) a first vicariance, due to the onset of the Würm glaciation, led to the formation of several major lineages, (*ii*) further vicariance during the LGM, the Older and/or the Younger Dryas in the Western and Eastern lineages and hence restriction of gene flow with isolation by distance during these periods and (iii) final vicariance effects resulted in two secondary sub-lineages in the area of Germany and, maybe, in two other secondary sub-lineages in the Czech Republic, (iv) strong range expansions following postglacial warming in most of the genetic lineages.

My results show that, in the presence of haplotype network ambiguities, differential correction for homoplastically evolving microsatellite-like base position can substantially improve the resolving power of the otherwise problematic control region.

5. General conclusion

Different genetic markers have different sensitivities to the process of evolution. Some evolve at slower rates than others (Pesole et al. 1999) and therefore could reconstruct older levels of of an organism's evolutionary history. Moreover, another important feature for a genetic marker is the mode of transmission. In constrast to nuclear DNA, mitochondrial DNA is mainly maternaly inherited and therefore its evolution is clonal. The mtDNA variation is therefore the result of mutations accumulated in maternal lineages since divergence of the common ancestor. Hence only one genealogy supports the evolution of mtDNA sequences. The best way to reconstruct a phylogeographic history, being as close as possible to the truth history, is therefore to combine information from different types of genetic markers, e.g., from the nuclear and mitochondrial genomes.

I first investigated the evolutionary history of the Woodland Ringlet by sequencing part of the cytochrome oxydase subunit one (COI) gene. These results show the existence of three distinct lineages of *E. medusa*, confined to Western, Central and Eastern Europe, respectively. From the resulting gene tree (haplotype network) ring-shaped ambiguities emerged, necessitating two parallel analyses. Nested clade phylogeographic analysis (NCPA) inferred two alternative scenarios that I compiled into a consensus scenario. In concert with results from nuclear markers (allozyme), COI haplotype distribution supports (*i*) a first vicariance at the onset of the Würm glaciation, leading to the formation of major lineages, (*ii*) the differentiation of German populations during the Younger Dryas and (*iii*) postglacial range expansions.

In a second study, I focused on the impact of the ring-shaped ambiguities on phylogeographic reconstruction of the Woodland Ringlet. Such loop-structures, due to homoplasy or recombination, hamper a straightforward extraction of phylogeographic information harboured in the history of a gene. I therefore resolved the COI haplotype network assuming either recombination or homoplasy as the origin of the ring and inferred a consensus scenario for both alternatives. Especially the scenario corrected for recombination substantially supplemented the allozyme scenario. Accepting the process of recombination to occur in mitochondria (see below), this additional phylogeographic information improves the existing evolutionary history for *E. medusa* in Europe.

In my third study I investigated the phylogeography of *E. medusa* sequencing the highly variable mitochondrial control region. The geographic haplotype distribution of this gene was less clear than in COI, nevertheless again the already known pattern of three distinct lineages

distributed in Western, Central and Eastern Europe, respectively, appeared. Haplotypes found in Monte Baldo Massif (southern Alp) are private to this population. This corresponds to the allozyme data which also showed the existence of a fourth lineage in this area. Despite numerous ring-shaped ambiguities in the control region haplotype network, the extracted phylogeographic informations is in concordance with that extracted from COI and allozyme analyses. It also indicates that both, recombination and homoplasy, have significantly shaped the COI and control region haplotype variation of the Woodland Ringlet mtDNA.

Combining information from independent genetic markers in fact supported a phylogeographic scenario that does not characterise *E. medusa* as a Siberian element. The remarkable genetic structure of this species in Europe implies the existence of several extra-Mediterranean centres of differentiation during the glaciation, contrasting with that typically expected for a Siberian element.

An interesting aspect of my thesis was the differential treatment of ring-shaped ambiguities in mitochondrial haplotype networks. They may severely handicap the extraction of phylogeographic information. One of the possible reasons for their occurrence is recombination, however the existence of this process in mitochondria is still a controversial issue. If we accept recombination to occur in mtDNA, including successful transmission of recombinants to the next generation, it is necessary to consider its effect on gene genealogies, the basis of phylogeographic reconstruction. Since recombinants combine different evolutionary histories, traditional methods of phylogenetic reconstruction that assume a linear evolutionary history are not appropriate (Pfenninger and Posada 2002). Here I developed an alternative approach to exploit phylogeographic information from gene tree: (i) first I resolved the tree by considering both recombination and homoplasy and if possible (ii) I compiled a consensus of alternative phylogeographic scenarios. This strategy has always to be adapted to the number of ring-shaped ambiguities, and the geographical distribution of haplotypes involved in loop formation. It has to be strict with respect to the consensus scenario that is to be inferred.

In conclusion, my approach allowed to substantially adding information to our knowledge on the population history of the Woodland Ringlet. Based on different genetic markers, I deduce for *E. medusa* a pattern that I would not expect for a Siberian faunal element. It comforts the existence of extra-Mediterranean differentiation centres.

Two promising future directions of research directly emerge from my results. First, it would be interesting to study more organisms that are considered Siberian faunal elements to see if their population histories match the hypothesis of extra-Mediterranean refugia. This could

finally lead to a generalized pattern of extra-Mediterranean refugia, as was previously found for the ‘classical’ hedgehog, grasshopper and bear patterns. Second, the potential impact of recombination and homoplasy on the reconstruction of evolutionary histories should be worked out in more detail, including simulation studies.

Finally, studies on other taxa would unravel whether such phylogeographic patterns with extra-Mediterranean refugia as exemplified for *E. medusa* constitute an exception or a general paradigm for many other organisms.

6. Abstract

Phylogeography is a recent field of biological research that links phylogenetics to biogeography through deciphering the imprint that evolutionary history has left on the genetic structure of extant populations. During the cold phases of the successive ice ages, which drastically shaped species' distributions since the Pliocene, populations of numerous species were isolated in refugia where many of them evolved into different genetic lineages. My dissertation deals with the phylogeography of the Woodland Ringlet (*Erebia medusa* [Denis and Schiffermüller] 1775) in Central and Eastern Europe. This Palaearctic butterfly species is currently distributed from central France and south eastern Belgium over large parts of Central Europe and southern Siberia to the Pacific. It is absent from those parts of Europe with mediterranean, oceanic and boreal climates. It was supposed to be a Siberian faunal element with a rather homogeneous population structure in Central Europe due to its postglacial expansion out of a single eastern refugium.

An already existing evolutionary scenario for the Woodland Ringlet in Central and Eastern Europe is based on nuclear data (allozymes). To know if this is corroborated by organelle evolutionary history, I sequenced two mitochondrial markers (part of the cytochrome oxydase subunit one and the control region) for populations sampled over the same area.

Phylogeography largely relies on the construction of networks of uniparentally inherited haplotypes that are compared to geographic haplotype distribution thanks to recent developed methods such as nested clade phylogeographic analysis (NCPA). Several ring-shaped ambiguities (loops) emerged from both haplotype networks in *E. medusa*. They can be attributed to recombination and homoplasy. Such loops usually avert the straightforward extraction of the phylogeographic signal contained in a gene tree.

I developed several new approaches to extract phylogeographic information in the presence of loops, considering either homoplasy or recombination. This allowed me to deduce a consistent evolutionary history for the species from the mitochondrial data and also adds plausibility for the occurrence of recombination in *E. medusa* mitochondria. Despite the fact that the control region is assumed to have a lack of resolving power in other species, I found a considerable genetic variation of this marker in *E. medusa* which makes it a useful tool for phylogeographic studies.

In combination with the allozyme data, the mitochondrial genome supports the following phylogeographic scenario for *E. medusa* in Europe: (i) a first vicariance, due to the onset of the Würm glaciation, led to the formation of several major lineages, and is mirrored in the

NCPA by restricted gene flow, (ii) later on further vicariances led to the formation of two sub-lineages in the Western lineage and two sub-lineages in the Eastern lineage during the Last Glacial Maximum or Older Dryas; additionally the NCPA supports a restriction of gene flow with isolation by distance, (iii) finally, vicariance resulted in two secondary sub-lineages in the area of Germany and, maybe, to two other secondary sub-lineages in the Czech Republic. The last postglacial warming was accompanied by strong range expansions in most of the genetic lineages.

The scenario expected for a presumably Siberian faunal element such as *E. medusa* is a continuous loss of genetic diversity during postglacial westward expansion. Hence, the pattern found in this thesis contradicts a typical Siberian origin of *E. medusa*. In contrast, it corroboratess the importance of multiple extra-Mediterranean refugia for European fauna as it was recently assumed for other continental species.

7. References

Aagaard K, Hindar K, Pullin AS, James CH, Hammarstedt O, Balstad T, Hanssen O (2002) Phylogenetic relationships in brown argus butterflies (Lepidoptera: Lycaenidae: *Aricia*) from northwestern Europe. *Biological Journal of the Linnean Society* **75:** 27-37.

Andersen BG, Borns Jr HW (1997) *The Ice Age World*. Scandinavian University Press, Oslo.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge.

Babik W, Branicki W, Sandera M, Litvinchuk S, Borkin LJ, Irwin JT, Rafinski J (2004) Mitochondrial phylogeography of the moor frog, *Rana arvalis*. *Molecular Ecology* **13:** 1469-1480.

Bereczki J, Pecsenye K, Peregovits L, Varga Z (2005) Pattern of genetic differentiation in the *Maculinea alcon* species group (Lepidoptera, Lycaenidae) in Central Europe. *Journal of Zoological Systematics and Evolutionary Research* **43:** 157-165.

Brown WM, Georges MJr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceeding of the National Academy of Sciences of the USA* **76:** 1967-1971.

Burzynski A, Malgorzata Z, Skibinski D, Wenne R (2003) Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. *Molecular Biology and Evolution* **20:** 388-392.

Cassel A, Tammaru T (2003) Allozyme variability in central, peripheral and isolated populations of the scarce heath (*Coenonympha hero*: Lepidoptera, Nymphalidae): implications for conservation. *Conservation Genetics* **4:** 83-93.

Cassens I, Van Waerebeek K, Best PB, Crespo EA, Reyes J, Milinkovitch MC (2003) The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of networks methods rooting procedures. *Molecular Ecology* **12:** 1781-1792.

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

Coope GR (1970) Interpretations of Quaternary insect fossils. *Annual Review of Entomology* **15:** 97–120.

Coope GR (1978*) Constancy of Insect Species Versus Inconstancy of Quaternary Environments - Diversity of Insect Faunas*. Blackwell Scientific Publications, Oxford.

Coope GR (1994) The response of insect faunas to glacial-interglacial climatic fluctuations. *Philosophical Transactions of the Royal Society of London, Serie B:Biological Sciences* **280:** 313-340.

Eanes WF (1999) Analysis of selection on enzyme polymorphisms. *Annual Review of Ecology, Evolution and Systematics* **30:** 301-326.

Ebert G, Rennwald E (eds.) (1991) *Die Schmetterlinge Baden-Württembergs*. Vol. 2, Verlag Eugen Ulmer, Stuttgart.

Englbrecht CC, Freyhof J, Nolte A, Rassmann K, Schliewen U, Tautz D (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces: Teleostei: Cottidae) suggests a pre-Pleistocene origin of the major Central European population. *Molecular Ecology* **9:** 709-722.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11:** 1591-1604.

Füköh L, Krolopp E, Sümegi P (1995) Quaternary malacostratigraphy in Hungary. *Malacological Newsletter (Gyöngyos)* Supplement **1:** 1-129.

Frenzel B (1973) *Climatic Fluctuations of the Ice Age*. The press of Case western Reserve University, Cleveland and London.

Goldstein B, Schlötterer C (1999) *Microsatellites - Evolution and Applications.* Oxford University Press, Oxford.

Habel JC, Schmitt T, Müller P (2005) The fourth paradigm pattern of post-glacial range expansion of European terrestrial species: the phylogeography of the marbled white butterfly (Satyrinae, Lepidoptera). *Journal of Biogeography* **32:** 1489-1497.

Hammouti N, Schmitt T, Seitz A, Kosuch J, Veith M (Submitted a). Phylogeography of *Erebia medusa* based on DNA sequences of the mitochondrial COI gene.

Hammouti N, Schmitt T, Seitz A, Kosuch J, Veith M (Submitted b). Adjustement of ring-shaped ambiguities in minimum spanning networks for recombination and homoplasy: its impact on phylogeographic reconstruction.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull D, Anderson C, Ghosh S, Olefsky J, Beal F, Davis R, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding region sequences for the major African, Asia and European haplogroups. *American Journal of Human Genetics* **70:** 1152-1171.

Hertelendy E, Sümegi P, Szöör G (1992) Geochronological and paleoclimatic characterisation of Quaternary sediments in the great Hungarian plain. *Radiocarbon* **34:** 833-839.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58:** 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:** 87-112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907-913.

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

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **359:** 183-195.

Hewitt GM, Butlin RK (1997) *Causes and Consequences of Population Structure*. In: Krebs J, Davies, N, eds. *Behavioural Ecology: an Evolutionary Approach*, pp. 350-372. Blackwell Scientific Publications, Oxford.

Hoarau G, Holla S, Lescasse R, Stam W, Olsen J (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichtys flesus*. *Molecular Biology and Evolution* **19:** 2261-2264.

Joyce DA, Pullin AS (2001) Phylogeography of the marsh fritillary *Euphydryas aurinia* (Lepidoptera: Nymphalidae) in the UK. *Biological Journal of the Linnean Society* **72:** 129-141.

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of anurans. *Systematic Zoology* **18:** 1-32.

Knowles L, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology* **11:** 2623-2635.

Korchunov J, Gorbunov P (1995) *Dnevnye babotschki asiatskoij tschasti Roussii (russisch). Spravotschnik*, Jekaterinburg.

Kudrna O (2002) The distribution atlas of European butterflies. *Oedippus* **20:** 1-342.

Ladoukakis ED, Zouros E (2001a) Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Molecular Biology and Evolution* **18:** 1168-1175.

Ladoukakis ED, Zouros E (2001b) Recombination in animal mitochondrial DNA: evidence from published sequences. *Molecular Biology and Evolution* **18:** 2127-2131.

Lattin G de (1957) *Die Lepidopteren-Fauna der Pfalz, I. Teil A. Systematisch-chorologischer Teil. Mitteilungen der Pollichia, pfälzischer Verein für Naturkunde und Naturschutz* III. Reihe **4:** 51-167.

Lattin G de (1967) *Grundriss der Zoogeographie.* Gustav Fischer, Jena.

Lunt DH, Hyman BC (1997) Animal mitochondrial DNA recombination. *Nature* **387:** 247.

Lunt DH, Zhang DX, Szymura JM, Hewitt GM (1996) The insect cytochrome oxidase I gene: evolutionary patterns, and conserved primers for phylogenetic studies. *Insect Molecular Biology* **5:** 153-165.

Maddison WP, Maddison DR (1992) *MacClade: Analysis of Phylogeny and Character Evolution, version 3*. Sinauer, Sunderland.

Martin JF, Gilles A, Lortscher M, Descimon H (2002) Phylogenetics and differentiation among the western taxa of the *Erebia tyndarus* group (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **75:** 319-332.

Malyarchuk BA (2005) Distribution of nucleotide substitutions in human mitochondrial DNA genes. *Russian Journal of Genetics* **41:** 79-84.

Meglecz E, Pecsenye K, Peregovits L, Varga Z (1997) Allozyme variation in *Parnassius mnemosyne* (L.) (Lepidoptera) populations in northeast Hungary: variation within a subspecies group. *Genetica* **101:** 59-66.

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology, Evolution, and Systematics* **18:** 269-292.

Napolitano M, Descimon H (1994) Genetic structure of French populations of the mountain butterfly *Parnassius mnemosyne* L. (Lepidoptera: Papilionidae). *Biological Journal of the Linnean Society* **53:** 325-344.

Navascués M, Emerson C (2005) Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. *Molecular Ecology* **14:** 1333-1341.

Nesbo CL, Fossheim T, Vollestad LA, Jakobsen KS (1999) Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology* **8:** 1387-1404.

Nilsson T (1983) *The Pleistocene: Geology and Life in the Quaternary Ice Age*. Ferdinand Enke Verlag, Stuttgart.

Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *Journal of Biological Chemistry* **265:** 4718-4721.

Pazonyi P (2004) Mammalian ecosystem dynamics in the Carpathian basin during the last 27,000 years. *Paleogeography, Paleoclimatology, Paleoecology* **212:** 295-314.

Pesole G, Gissi C, De Chirico A (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution* **48:** 427-434.

Pfenninger M, Posada D (2002) Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): fragmentation, corridor migration, and secondary contact. *Evolution* **56:** 1776-1788.

Piganeau G, Gardner M, Eyre-Walker A (2004) A broad survey of recombination in animal mitochondria. *Molecular Biology and Evolution* **21:** 2319-2325.

Porter AH, Geiger H (1995) Limitations to the inference of gene flow at regional geographic scales- an example from the *Pieris napi* group (Lepidoptera: Pieridae) in Europe. *Biological Journal of the Linnean Society* **54:** 329-348.

Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology* **9:** 487-488.

Rafinski J, Babik W (2000) Genetic differentiation among northern and southern populations of the moor frog *Rana arvalis* Nilsson in Central Europe. *Heredity* **84:** 610-618.

Rokas A, Holland P (2000) Rare genomic changes as a tool for phylogenetics. *Trends in Ecology and Evolution* **15:** 454-459.

Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution* **18:** 411-417.

SBN (eds.) (1987) *Tagfalter und ihre Lebensräume Arten Gefährdung Schutz Schweiz und angrenzende Gebiete. Band 1*. Fotorotar, Egg. Abgekürzt SBN.

Schmitt T (1993) Biotopansprüche von *Erebia medusa brigonna* Fruhstorfer, 1917 (Rundaugen Mohrenfalter) im Nordsaarland. *Atalanta* **24:** 33-56.

Schmitt T (1999) *Phylogeographie europäischer Tagfalter basierend auf populations genetischen Analysen unter besonderer Berücksichtigung von Polyommatus coridon und Polyommatus icarus (Lepidoptera: Lycaenidae) sowie Erebia medusa und Maniola jurtina (Lepidoptera: Satyrinae).* PhD thesis, University of Mainz, Germany.

Schmitt T (2002) The Biology of *Erebia medusa* ([Denis and Schiffermüller], 1775) in Central Europe (Lepidoptera). *Acta Biologica Debrecina* **24:** 113-129.

Schmitt T, Seitz A (2001a) Intraspecific allozymatic differentiation reveals the glacial refugia and the postglacial expansions of European *Erebia medusa* (Lepidoptera: Nymphalidae). *Biological Journal of the Linnean Society* **74:** 429-458.

Schmitt T, Seitz A (2001b) Allozyme variation in *Polyommatus coridon* (Lepidoptera: Lycaenidae): identification of ice-age refugia and reconstruction of post-glacial expansion. *Journal of Biogeography* **28:** 1129-1136.

Schmitt T, Seitz A (2002) Postglacial distribution area expansion of *Polyommatus coridon* (Lepidoptera: Lycaenidae) from its Ponto-Mediterranean glacial refugium. *Heredity* **89:** 20-26.

Schmitt T, Gießl A, Seitz A (2002) Postglacial colonisation of western Central Europe by *Polyommatus coridon* (Poda, 1761) (Lepidoptera: Lycaenidae): evidence from population genetics. *Heredity* **88:** 26-34.

Schmitt T, Gießl A, Seitz A (2003) Did *Polyommatus icarus* (Lepidoptera: Lycaenidae) have distinct glacial refugia in southern Europe? Evidence from population genetics. *Biological Journal of the Linnean Society* **80:** 529-538.

Schmitt T, Krauss J (2004) Reconstruction of the colonization route from glacial refugium to the northern distribution range of the European butterfly *Polyommatus coridon* (Lepidoptera : Lycaenidae). *Diversity and Distributions* **10:** 271-274.

Schmitt T, Röber S, Seitz A (2005) Is the last glaciation the only relevant event for the present genetic population structure of the meadow brown butterfly *Maniola jurtina* (Lepidoptera: Nymphalidae)? *Biological Journal of the Linnean Society* **85:** 419-431.

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. *Annual Review of Biochemistry* **66:** 409-435.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87:** 650-701.

Sümegi P, Rudner ZE (2001) In situ charcoal fragments as remains of natural wild fires in the upper Würm of the Carpathian basin. *Quaternary International* **76:** 165-176.

Sümegi P, Krolopp E (2002) Quatermalacological analyses for modeling of the upper Weichselian palaeoenvironmental changes in the Carpathian basin. *Quaternary International* **91:** 53-63.

Swofford DL (2001) *PAUP Phylogenetics Analysis Using Parsimony (and Other Methods), version 4.06b*. Sinauer Associates, Sunderland.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453-464.

Taylor MFJ, McKechnie SW, Pierce N, Kreitman M (1993) The lepidopteran mitochondrial control region: structure and evolution. *Molecular Biology and Evolution* **10:** 1259-1272.

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

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* **13:** 789-809.

Templeton AR (2005) *Geodis Inference Key (November 11, 2005)*. http://darwin.uvigo.es/

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117:** 343–351.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134:** 659-669.

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140:** 767-782.

Templeton AR, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF (2000) Cladistic structure within the human *Lipoprotein lipase* gene and its implications for phenotypic association studies. *Genetics* **156:** 1259-1275.

Thyagarajan B, Padua R, Campbell C (1996) Mammalian mitochondria possess homologous DNA recombination activity. *Journal of Biological Chemistry* **271:** 27536-27543.

Tolman T, Lewington R (1997) *Field Guide to Butterflies of Britain and Europe*. Harper Collins Publishers, London.

Tzedakis PC, Lawson IT, Frogley MR, Hewitt GM, Preece RC (2002) Buffered tree population changes in a Quaternary refugium: evolutionary implications. *Science* **297:** 2044-2047.

Vandewoestijne S, Baguette M, Brakefiel PM, Saccherib IJ (2003) Phylogeography of *Aglais urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. *Molecular Phylogenetics and Evolution* **31:** 630-646.

Varga Z (1977) Das Prinzip der areal-analytischen Methode in der Zoogeographie und die Faunenelement-Einteilung der europäischen Tagschmetterlinge (Lepidoptera: Diurna). *Acta Biologica Debrecina* **14:** 223-285.

Vila M (2004) *Effects of Environmental Changes during the Quaternary on the Distribution of Extant Species. The Case of the Butterflies Erebia triaria and E. palarica.* PhD thesis, University of A Coruña, Galicia, Spain.

Vila M, Björklund M (2004) The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). *Journal of Molecular Evolution* **58:** 280-290.

Williams D, Dunkerley D, DeDecker P, Kershaw P, Chappell M (1998) *Quaternary Environnements*. Arnold, London.

Willis KJ, van Andel TH (2004) Trees or no trees? The environments of Central and Eastern Europe during the last glaciation. *Quaternary Science Reviews* **23:** 2369-2387.

Wood TE, Burke JM, Rieseberg LH (2005) Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123:** 157-170.

Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology* **25:** 99-120.

Zhang DX, Szymura JM, Hewitt GM (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution* **40:** 382-391.

8. Appendix

Appendix 1: Sample sites, country and location, geographical coordinates and date of capture are indicated.

Sample	Country	Location	Latitude	Longitude	Date of capture
1	Germany	Haustadt	49°25´N	6°43´E	07.06.1996
2	Germany	Hunsrück Thranenweier	49°42´N	7°05´E	10.06.1996
3	Germany	Hammerstein	49°40´N	7°17´E	06.06.1996
4	Germany	northern Eifel Birgel	50°19´N	6°37´E	08.06.1996
5	Germany	Göllesberg	48°10´N	8°40´E	25.05.1998
6	Germany	Hinterzarten	47°53´N	8°05´E	07.06.1997
7	Germany	Bärnhöhe	49°53´N	12°04´E	09.06.2002
8	Germany	Breitenfurth	48°52´N	11°05´E	04.06.1997
9	Germany	Flughafen Bayreuth	49°59´N	11°38´E	28.05.2002
10	Germany	Kallmünz	49°08´N	11°58´E	30.05.2002
11	Germany	Mauth	48°52´N	13°34´E	14.06.2002
12	Germany	Rusel	48°52´N	13°04´E	31.05.2002
13	Germany	Schönsee	49°31´N	12°33´E	12.06.2002
14	Germany	Oberschleissheim	48°15´N	11°34´E	26.05.1998
15	Germany	Berka vor dem Hainich	51°02´N	10°22´E	10.06.1997
16	France	Maillat	46°07´N	5°32´E	05.06.1998
17	Italy	Monte Baldo	45°34´N	10°41´E	25.07.1996
18	Czech Republic	Blazejovice	48°57´N	13°56´E	01.06.2002
19	Czech Republic	Cesky Krumlov	48°49´N	14°19´E	03.06.2002
20	Czech Republic	Fojtovice	50°43´N	13°50´E	13.06.1997
21	Czech Republic	Holý Vrch	50°23´N	14°58´E	13.06.1997
22	Czech Republic	Hlinisté	48°54´N	13°45´E	14.06.2002
23	Czech Republic	Jakuboyice	50°00´N	16°49´E	02.06.1997
24	Czech Republic	Klentnice	48°51´N	16°38´E	01.06.1997
25	Czech Republic	Stará Hut’	49°46´N	14°11´E	18.06.1997
26	Czech Republic	Spicák	49°08´N	13°13´E	19.06.1997
27	Slovakia	Podlesok	48°28´N	19°13´E	15.06.1997
28	Slovakia	Snina	48°58´N	22°08´E	16.06.1997
29	Hungary	Felsöszölnök	46°52´N	16°10´E	27.05.1998
30	Hungary	Gánt	47°22´N	18°23´E	29.05.1998
31	Hungary	Szölöhegy	48°28´N	20°34´E	24.05.1997
32	Hungary	Komlóska	48°19´N	21°28´E	26.05.1997

Appendix 2: NCPA result for the different scenarios based on the COI mitochondrial gene. Only clades with significant geographic structure are shown; I= interior clade, T= tip clade; χ^2 = observed chi-square; $p_{\chi 2}$ = probability of random χ^2 (10000 permutations) being greater or equal to observed χ^2; D_c = distance within clade; D_n = distance within nested clade; (I-T)/D = interior *vs*. tip clade distances; significant values: ^{S} = smaller than mean distance; ^{L} = larger than mean distance.

Scenario A: Recombinants included in the ambiguous networks resolved with the additional rules given in Templeton and Sing (1993). Alternative 1.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	(I-T)/D_c	(I-T)/ D_n
1-1	I/T	354.06	<0.001			152.81 ^{L}	-326.18 ^{S}
H01	I			203.88 ^{S}	209.16 ^{S}		
H03	I			221.69 ^{S}	242.82 ^{S}		
H04	T			80.39 ^{S}	651.33 ^{L}		
H08	T			6.44 ^{S}	371.47		
H09	I			43.63 ^{S}	587.29 ^{L}		
H11	T			0.00 ^{S}	361.85		
H12	T			0.00	376.28		
H13	T			0.00	579.54		
H14	T			0.00	266.62		
H15	T			0.00	619.76		
1-3		69.41	0.004			16.73	-144.88 ^{S}
H02	I			130.28	125.42 ^{S}		
H07	T			151.39	193.27		
H16	T			0.00	501.43 ^{L}		
2-1		71.08	<0.001			57.17	-84.34
1-1	I			317.83	320.80		
1-2	T			260.66	405.14		
2-2		33.07	0.001			37.23	-31.49
1-3	I			137.75	140.60		
1-4	T			100.52	172.09		
total clade		143.74	<0.001			180.57 ^{L}	157.68 ^{L}
2-1	I			324.74 ^{L}	334.93 ^{L}		
2-2	T			144.17 ^{S}	177.25 ^{S}		

Scenario B: Recombinants included in the ambiguous networks resolved with the additional rules given in Templeton and Sing (1993). Alternative 2.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-1	I/T	431.86	<0.001			142.24	-83.60 [S]
H01	I			203.88	201.48 [S]		
H02	I			130.28 [S]	248.75		
H03	I			221.69	227.40		
H07	T			151.39	203.70		
H08	I			6.44 [S]	337.30		
H11	T			0.00 [S]	327.67		
H12	T			0.00	342.12		
H13	T			0.00	560.22 [L]		
H14	T			0.00	241.61		
H16	T			0.00	292.27		
1-2		6.00	0.062			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28		
1-3		8.1667	0.135			-32.03	-23.22
H04	T			80.39	82.07		
H09	I			43.63	55.39		
H15	T			0.00	23.27		
2-1		87.90	<0.001			133.20 [L]	-93.90
1-1	I			233.72	231.19		
1-2	T			100.52 [S]	325.09		
2-2		25.00	<0.001			-185.52	-325.99 [S]
1-3	I			75.13 [S]	120.50 [S]		
1-4	T			260.66	446.50 [L]		
total clade		147.86	<0.001			-49.65	-298.58 [L]
2-1	T			235.36 [S]	236.05 [S]		
2-2	I			185.70 [S]	534.63 [L]		

Appendix 3: NCPA result for different scenarios based on the COI mitochondrial gene considering recombination (scenarios A to E) and homoplasie (scenario G and F). The clades with significant geographic structure are shown; I= interior clade, T= tip clade; χ^2 = observed chi-square; $p_{\chi 2}$ = probability of random χ^2 (10,000 permutations) being greater or equal to observed χ^2; D_c = distance within clade; D_n = distance within nested clade; (I-T)/D = interior *vs*. tip clade distances; significant values: ^{S} = smaller than mean distance; ^{L} = larger than mean distance.

Scenario A: Network resolved and recombinants excluded.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-1	I/T	48.93	0.213			203.88	-46.84
H01	I			203.88	207.65		
H11	T			0.00	231.65		
H13	T			0.00	439.00		
H14	T			0.00	115.69		
1-2		25.59	0.231			216.86 ^{L}	20.73
H03	I			221.69	226.02		
H08	T			6.44 ^{S}	204.08		
H12	T			0.00	208.90		
1-3		0.44	1.000			43.63	16.36
H09	I			43.63	40.90		
H15	T			0.00	24.54		
1-5		6.00	0.065			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28 ^{L}		
1-6		44.00	0.061			130.28	-379.01 ^{S}
H02	I			130.28	127.09 ^{S}		
H16	T			0.00	506.11 ^{L}		
2-1		59.46	<0.001				
1-1	I			211.72	213.07		
1-2	I			223.71	232.91		
2-2		8.00	0.019			-223.84^S	36.95
1-3	I			36.81 ^{S}	277.97		
1-4	T			260.66	241.02		
2-3		32.95	0.003			35.19	-33.27
1-5	T			100.52	171.98		
1-6	I			135.71	138.71		
total clade		241.47	<0.001			82.60 ^{L}	32.46
2-1	I			221.78 ^{S}	238.70		
2-2	I			257.44	495.74 ^{L}		
2-3	T			142.70 ^{S}	231.66		

Scenario B: Network resolved and recombinants included and counted in either parental haplotypes H02 and H09.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-1	I/T	48.93	0.216			203.88	-46.84
H01	I			203.88	207.65		
H11	T			0.00	231.65		
H13	T			0.00	439.00		
H14	T			0.00	115.69		
1-2		25.59	0.230			216.86 [L]	20.73
H03	I			221.69	226.02		
H08	T			6.44 [S]	204.08		
H12	T			0.00	208.90		
1-3		3.1579	1.000			77.91	54.58
H09	I			77.91	77.86		
H15	T			0.00	23.27		
1-5		6.000	0.075			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28		
1-6		46.00	0.052			132.59	-371.58 [S]
H02	I			132.59	129.85 [S]		
H16	T			0.00	501.43 [L]		
2-1		59.46	<0.001				
1-1	I			211.72	213.07		
1-2	I			223.71	232.91		
2-2		25.00	<0.001			-185.52	-325.99 [S]
1-3	I			75.13 [S]	120.50 [S]		
1-4	T			260.66	446.50 [L]		
2-3		33.07	0.005			37.23	-31.49
1-5	T			100.52	172.09		
1-6	I			137.75	140.60		
total clade		288.74	<0.001			69.18 [L]	-157.68 [L]
2-1	I			221.78 [S]	274.05		
2-2	I			185.70 [S]	534.63 [L]		
2-3	T			144.17 [S]	177.25 [S]		

Scenario C: Network resolved and recombinants included and counted in either parental haplotypes H02-Hyp01

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-1	I/T	48.93	0.215			203.88	-46.84
H01	I			203.88	207.65		
H11	T			0.00	231.65		
H13	T			0.00	439.00		

H14	T			0.00	115.69		
1-2		25.59	0.221			216.86 [L]	20.73
H03	I			221.69	226.02		
H08	T			6.44 [S]	204.08		
H12	T			0.00	208.90		
1-3		0.444	1.000			43.635	16.36
H09	I			43.63	40.90		
H15	T			0.00	24.54		
1-5		6.000	0.060			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28		
1-6		125.00	<0.001			56.92	-188.53 [S]
H02	I			132.59 [S]	177.51 [S]		
H16	T			0.00	614.12 [L]		
Hyp01	T			80.39 [S]	350.55 [L]		
2-1		59.46	<0.001				
1-1	I			211.72	213.07		
1-2	I			223.71	232.91		
2-2		8.00	0.016			-223.84[S]	36.95
1-3	I			36.81 [S]	277.97		
1-4	T			260.66	241.02		
2-3		43.80	<0.001			127.87 [L]	71.65
1-5	T			100.52 [S]	156.45		
1-6	I			228.39	228.04		
total clade		232.86	<0.001			3.49	20.14
2-1	I			221.78 [S]	274.05		
2-2	I			257.44	449.36 [L]		
2-3	T			221.81 [S]	271.24		

Scenario D: Network resolved and recombinants included and counted in either parental haplotypes H03 and H09.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	(I-T)/D_c	(I-T)/ D_n
1-1	I/T	48.93	0.204			203.88	-46.84
H01	I			203.88	207.65		
H11	T			0.00	231.65		
H13	T			0.00	439.00		
H14	T			0.00	115.69		
1-2		27.85	0.332			215.89 [L]	1.98
H03	I			220.72	225.67		
H08	T			6.44 [S]	222.48		
H12	T			0.00	227.31		
1-3		3.1579	1.000			77.91	54.58
H09	I			77.91	77.86		

H15	T			0.00	23.27		
1-5		6.000	0.066			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28		
1-6		44.00	0.043			130.28	-379.01 [S]
H02	I			130.28	127.09 [S]		
H16	T			0.00	506.11 [L]		
2-1		59.42	<0.001				
1-1	I			211.72	212.38		
1-2	I			225.47	235.20		
2-2		25.00	<0.001			-185.52	-325.99 [S]
1-3	I			75.13 [S]	120.50 [S]		
1-4	T			260.66	446.50 [L]		
2-3		32.95	<0.001			35.19	-33.27
1-5	T			100.52	171.98		
1-6	I			135.71	138.71		
total clade		284.38	<0.001			71.70 [L]	154.05 [L]
2-1	I			222.85 [S]	270.90		
2-2	I			185.70 [S]	534.63 [L]		
2-3	T			142.70 [S]	176.79 [S]		

Scenario E: Network resolved and recombinants included and counted in either parental haplotypes H03 and Hyp01.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-1	I/T	48.93	0.216			203.88	-46.84
H01	I			203.88	207.65		
H11	T			0.00	231.65		
H13	T			0.00	439.00		
H14	T			0.00	115.69		
1-2		27.85	0.311			215.89 [L]	1.98
H03	I			220.72	225.67		
H08	T			6.44 [S]	222.48		
H12	T			0.00	227.31		
1-3		0.444	1.000			43.635	16.36
H09	I			43.63	40.90		
H15	T			0.00	24.54		
1-5		6.000	0.057			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28		
1-6		119.00	<0.001			54.61	-180.14 [S]
H02	I			130.28 [S]	176.48 [S]		
H16	T			0.00	623.66 [L]		
Hyp01	T			80.39 [S]	339.94 [L]		

2-1		59.42	<0.001				
1-1	I			211.72	212.38		
1-2	I			225.47	235.20		
2-2		8.00	0.024			-223.84[S]	36.95
1-3	I			36.81[S]	277.97		
1-4	T			260.66	241.02		
2-3		44.22	<0.001			127.00[L]	67.93
1-5	T			100.52[S]	159.45		
1-6	I			227.52	227.38		
total clade		229.19	<0.001			4.95	12.82
2-1	I			222.85[S]	270.90		
2-2	I			257.44	449.36[L]		
2-3	T			221.21[S]	275.16		

Scenario F: Network resolved and recombinants counted in both parental haplotypes, with the parental haplotype *hyp01* included.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	(I-T)/D_c	(I-T)/ D_n
1-1	I/T	97.87	0.003			203.88[L]	-46.84
H01	I			203.88[S]	207.65		
H11	T			0.00[S]	231.65		
H13	T			0.00	439.00[L]		
H14	T			0.00	115.69		
1-2		53.45	0.046			216.49[L]	11.18
H03	I			221.32	226.01		
H08	T			6.44[S]	213.61		
H12	T			0.00	218.44		
1-3		4.36	0.248			73.97	57.95
H09	I			73.97	73.73		
H15	T			0.00	15.77		
1-5		12.00	0.004			0.00	74.64[L]
H06	T			0.00[S]	75.64[S]		
H10	I			0.00	150.28[L]		
1-6		213.00	<0.001			60.01	-261.93[S]
H02	I			131.48[S]	153.61[S]		
H16	T			0.00	569.44[L]		
Hyp01	T			80.39[S]	396.30[L]		
2-1		118.54	0.00				
1-1	I			211.72	212.70		
1-2	I			224.81	234.22		
2-2		34.00	0.00			-191.75	-229.67[S]
1-3	I			68.90[S]	161.07[S]		
1-4	T			260.66	390.74		
2-3		76.84	0.00			97.15[L]	39.73

1-5	T			100.52 ^{S}	157.86		
1-6	I			197.67	197.59		
total clade		456.61	0.00			28.84	81.42 ^{L}
2-1	I			222.37 ^{S}	272.45		
2-2	I			228.62 ^{S}	512.06 ^{L}		
2-3	T			193.58 ^{S}	231.55^{S}		

Scenario G: Network resolved in the hypothesis of homoplasy. Exclusion of the homoplasic site 443. Alternative 1.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	(I-T)/D_c	(I-T)/ D_n
1-2		3.157	1.000			77.91	54.58
H04	I			77.91	77.86		
H15	T			0.00	23.27		
1-3		6.000	0.060			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28		
1-4		46.00	0.070			132.59	-371.58^{S}
H02	I			132.59	129.85^{S}		
H16	T			0.00	501.43^{L}		
1-5		109.05	0.139			213.36^{L}	-18.28
H01	I			215.78^{S}	220.00		
H08	T			6.44^{S}	222.20		
H11	T			0.00^{S}	212.82		
H12	T			0.00	226.88		
H13	T			0.00	456.70^{L}		
H14	T			0.00	130.48		
2-1		25.00	<0.001			-185.52	-325.99^{S}
1-1	T			260.66	446.50^{L}		
1-2	I			75.13^{S}	120.50^{S}		
2-2		201.98	<0.001			90.65	-93.90
1-3	T			100.52^{S}	325.09		
1-4	I			137.75^{S}	234.65		
1-5	I			221.78	229.21		
total clade		147.86	<0.001			49.65	-298.58^{S}
2-1	T			185.70^{S}	534.63^{L}		
2-2	I			235.36^{S}	236.05^{S}		

Scenario H: Network resolved in the hypothesis of homoplasy. Exclusion of the homoplasic site 443. Alternative 2.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-2		3.157	1.000			77.91	54.58
H04	I			77.91	77.86		
H15	T			0.00	23.27		
1-3		6.000	0.071			0.00	74.64
H06	T			0.00	75.64		
H10	I			0.00	150.28		
1-4		46.00	0.064			132.59	-371.58^S
H02	I			132.59	129.85^S		
H16	T			0.00	501.43^L		
1-5		109.05	0.134			213.36^L	-18.28
H01	I			215.78^S	220.00		
H08	T			6.44^S	222.20		
H11	T			0.00^S	212.82		
H12	T			0.00	226.88		
H13	T			0.00	456.70		
H14	T			0.00	130.48		
2-1		177.67	<0.001			-67.63	-84.34
1-1	T			260.66	405.17		
1-2	I			75.13^S	633.73^L		
1-5	I			221.78^S	$244:47^S$		
2-2		33.07	0.005			37.23	-31.49
1-3	T			100.52	172.09		
1-4	I			137.75	140.60		
total clade		135.62	<0.001			-134.05^S	-127.64^S
2-1	T			316.04^L	330.21^L		
2-2	I			181.99^S	201.56^S		

Appendix 4:

NCPA result for the different scenario based on the mitochondrial control region. Only clades with significant geographic structure are shown; I= interior clade, T= tip clade; χ^2 = observed chi-square; $p_{\chi 2}$ = probability of random χ^2 (10000 permutations) being greater or equal to observed χ^2; D_c = distance within clade; D_n = distance within nested clade; (I-T)/D = interior *vs*. tip clade distances; significant values: ^{S} = smaller than mean distance; ^{L} = larger than mean distance.

Scenario A: Exclusion of the putative homoplasic sites.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-1		377.22	<0.001			89.52	-73.53
H01	I			263.70^S	264.93		
H13	T			407.27^L	438.57^L		
H16	T			0.00^S	635.48^L		
H18	T			0.00	323.13		
H21	T			0.00	323.13		
H23	T			0.00	127.36		
H27	T			0.00	214.80		
H29	T			10.24^S	132.61^S		
1-3		1.333	1.000			66.02	0.17
H30	I			66.02	74.24		
H31	T			0.00	74.06		
1-6		0.444	1.000			39.08	14.62
H17	I			39.08	36.64		
H25	T			0.00	22.01		
2-2		177.95	<0.001			231.77^L	-46.47
1-1	I			274.64	273.99		
1-2	T			0.00	287.92		
1-3	T			74.19^S	74.51^S		
1-6	T			32.98^S	626.22^L		
1-7	T			0.00	113.88		
total clade		158.00	<0.001			-277.01^S	134.09
2-1	T			277.01^S	278.68		
2-2	I			0.00^S	412.77		

Scenario B: Haplotype network resolved with the rules given in Pfenninger & Posada (2002). Haplotype H03 considered as interior.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	$(I-T)/D_c$	$(I-T)/D_n$
1-1	I/T	2.000	1.000				
H31	T			0.00	60.64		
H34	T			0.00	60.80		
1-3		43.049	0.085			16.57	42.37^{L}
H02	T			182.24^{S}	202.80^{S}		
H06	I			191.22^{S}	248.05^{L}		
H24	T			0.00	271.99		
1-6		110.99	0.031			255.12	-115.96
H01	I			276.57^{S}	278.83^{S}		
H03	I			195.74^{S}	403.29^{L}		
H16	T			0.00	711.84^{L}		
H18	T			0.00	414.26		
H21	T			0.00	414.26		
H23	T			0.00	119.25		
H27	T			0.00	195.45		
1-7		0.466	1.000			381.43	119.71
H13	I			381.43	367.73		
H20	T			0.000	248.02		
1-10		53.10	<0.001			66.79	312.05^{L}
H04	I			151.55	441.42^{L}		
H05	T			100.96^{S}	124.79^{S}		
H29	T			10.24^{S}	150.43		
1-11		0.444	1.000			39.08	14.62
H17	I			39.08	36.64		
H25	T			0.00	22.01		
2-1		2.000	1.000			26.91	26.91
1-1	T			60.72	60.73		
1-2	I			87.65	87.65		

2-2		225.97	<0.001			6.59	-95.70
1-3	I			226.42^{S}	235.33^{S}		
1-6	I			324.50	349.81^{L}		
1-7	T			350.63	440.32^{L}		
1-8	T			0.00	136.92		
1-9	T			0.00	315.69		
2-4		33.56	0.001			158.79	-320.27^{S}
1-10	I			191.78^{S}	223.75^{S}		
1-11	T			32.98^{S}	544.02^{L}		
total clade		307.82	<0.001			90.15^{L}	67.88^{L}
2-1	T			74.19^{S}	74.54^{S}		
2-2	I			304.82^{L}	303.39^{L}		
2-3	T			0.00^{S}	412.77		
2-4	T			256.59	229.29^{S}		

Scenario B: Haplotype network resolved with the rules given in Pfenninger & Posada (2002). Haplotype H03 considered as tip.

Clade		χ^2	$p_{\chi 2}$	D_c	D_n	(I-T)/D_c	(I-T)/ D_n
1-1	I/T	2.000	1.000				
H31	T			0.00	60.64		
H34	T			0.00	60.80		
1-3		43.049	0.0610			16.57	42.37^{L}
H02	T			182.24^{S}	202.80^{S}		
H06	I			191.22^{S}	248.05^{L}		
H24	T			0.00	271.99		
1-6		110.99	0.0290			255.12	-115.96^{S}
H01	I			276.57^{S}	278.83^{S}		
H03	T			195.74^{S}	403.29^{L}		
H16	T			0.00	711.84^{L}		
H18	T			0.00	414.26		
H21	T			0.00	414.26		
H23	T			0.00	119.25		

H27	T			0.00	195.45		
1-7		0.466	1.000			381.43	119.71
H13	I			381.43	367.73		
H20	T			0.000	248.02		
1-10		53.10	<0.001			66.79	312.05^{L}
H04	I			151.55	441.42^{L}		
H05	T			100.96^{S}	124.79^{S}		
H29	T			10.24^{S}	150.43		
1-11		0.444	1.000			39.08	14.62
H17	I			39.08	36.64		
H25	T			0.00	22.01		
2-1		2.000	1.000			26.91	26.91
1-1	T			60.72	60.73		
1-2	I			87.65	87.65		
2-2		225.97	<0.001			6.59	-95.70
1-3	I			226.42^{S}	235.33^{S}		
1-6	I			324.50	349.81^{L}		
1-7	T			350.63	440.32^{L}		
1-8	T			0.00	136.92		
1-9	T			0.00	315.69		
2-4		33.56	0.001			158.79	-320.27^{S}
1-10	I			191.78^{S}	223.75^{S}		
1-11	T			32.98^{S}	544.02^{L}		
Total clade		307.82	<0.001			90.15 ^{L}	67.88^{L}
2-1	T			74.19^{S}	74.54^{S}		
2-2	I			304.82^{L}	303.39^{L}		
2-3	T			0.00^{S}	412.77		
2-4	T			256.59	229.29^{S}		