Molecular Systematics, Phylogeography and Evolution of the genus *Cardaminopsis* HAYEK (Brassicaceae), the closest relatives of the model plant *Arabidopsis thaliana* (L.) HEYNH.

Michaela Maria Wernisch
(geb. Matschinger)

Dissertation
Universität für Bodenkultur, Wien 2007

Betreut von:
Univ.-Prof. Dr. rer. nat. Diplom Biologe Marcus A. Koch
Univ.-Prof. Dr. rer. nat. Diplom Geograph Karl-Georg Bernhardt

Contents

General introduction

The former genus *Cardaminopsis* Hayek, or simply "the wild relatives of *Arabidopsis thaliana* (L.) Heynh." as their members are often called, have received wide interest since their inclusion in the genus *Arabidopsis* (DC.) Heynh. (O'KANE & AL-SHEHBAZ 1997, AL-SHEHBAZ et al. 1999). The genus *Arabidopsis* belongs to the Brassicaceae, a family which comprises about 338 genera and some 3709 species (AL-SHEHBAZ et al. 2006). The family includes important crop plants (*Brassica oleracea* L., *Brassica napus* L., *Armoracia rusticana* P. Gaert., B. Mey. & Scherb. and many others), ornamental plants (*Aubrieta* Adans., *Iberis* L., *Lunaria* L., *Arabis* L., *Draba* L. and others) (KOCH & KIEFER 2007) and several model organisms (like *Brassica* L., *Capsella* Medik., *Arabis* and *Boechera* A. Löve & D. Löve) (KOCH et al. 2007), among which *Arabidopsis thaliana* is the most important. This species has become the model plant par excellence for genetic research since its introduction as a model plant for experimental genetics by LAIBACH (1943). A small genome, a short life circle and a small size made the species ideal for several fields of studies, such as genetics and physiology and later molecular and developmental biology (PIGLIUCCI 2002).

According to O'KANE & AL-SHEHBAZ (1997) the genus *Arabidopsis* comprises nine species. Additional taxa were recently taxonomically introduced (SHIMIZU et al. 2005; WARWICK et al. 2006), but their status is subject to discussion (KOCH et al. 2007). The area of distribution of the several taxa is mainly European with two species groups (*A. lyrata* (L.) O'Kane & Al-Shehbaz, *A. halleri* (L.) O'Kane & Al-Shehbaz) expanding into North America and/or East Asia. Diploids occur in *A. halleri* (AL-SHEHBAZ & O'KANE 2002; KOLNIK & MARHOLD 2006) and *A. cebennensis* (DC.) O'Kane & Al-Shehbaz, whereas both diploids and tetraploids are known from *A. lyrata*, *A. arenosa* (L.) Lawalree and *A. neglecta* (Schult.) O'Kane & Al-Shehbaz (AL-SHEHBAZ & O'KANE 2002). In the European distributed *A. arenosa* group, a high morphological and cytological variation with diploids to pentaploids and even aneuploids is observed, which has led to the description of several taxa (MESICEK 1970).

Arabidopsis thaliana and *Cardaminopsis* diverged approximately five million years ago (KOCH et al. 2000). This close relationship allows discussion of scientific questions on a broader scale, especially in molecular ecology and evolution (MITCHELL-OLDS 2001). While *A. thaliana* is annual, mostly self-compatible and weedy, the wild relatives show a broader spectrum of biological traits such as perennial life form, self-incompatibility to self-compatibility, heavy metal tolerance, different levels of polyploidy or potential for hybridization (CLAUSS & KOCH 2006). It has been shown that hybridization, reticulation, and polyploidy have played an important role in speciation and differentiation of the Brassicaceae (MARHOLD & LIHOVA 2006), particularly in Quaternary times during periods which were greatly influenced by glaciation and deglaciation (KOCH & MUMMENHOFF 2006). *Arabidopsis suecica* (Fr.) Norrl. is long known to be a hybrid (HYLANDER 1957) with *Arabidopsis thaliana* as the maternal and *A. arenosa* as the paternal parental species (for further literature see JAKOBSSON et al. 2006). Recently hybridization was detected in *A. lyrata* ssp. *kamchatica* (Fisch. ex DC.) O'Kane & Al-Shehbaz from Japan and Taiwan (SHIMIZU et al. 2005; BECK et al. 2007) with *A. lyrata* and *A. halleri* as parental taxa.

Studies on single species or groups of populations of the genus *Arabidopsis* beside *A. thaliana* concerning biogeographical questions are increasing (JONSELL et al. 1995; CLAUSS et al. 2002; WRIGHT et al. 2003; RAMOS-ONSINS et al. 2004; ANSELL et al. 2004; CLAUSS & MITCHELL-OLDS 2006), but there is no framework comprising all species from their entire distribution range.

The current study was based on a large dataset comprising nearly all species (except *A. pedemontana* (Boissier) O´Kane & Al-Shehbaz) with a representative geographic sampling. Based on sequence analyses of the nuclear marker ITS and the cpDNA marker *trnL*-Intron + *trnL/F*-IGS, phylogenetic trees were drawn, haplotype data were analysed phylogeographically and gene diversity parameters were calculated. In a non-glaciated refuge area near the eastern edge of the Eastern Alps two sympatrically occurring species of the genus were investigated on basis of a detailed population sampling. A complex hybrid zone with past and maybe ongoing hybridization was detected and is subject to further investigations.

Chapter 1

Systematics of the genus *Arabidopsis* with special emphasis on the former genus *Cardaminopsis*

Abstract

The genus *Arabidopsis* with the model plant *A. thaliana* and the included former genus *Cardaminopsis* is subject to broad interest due to the possibility of transferring knowledge from the model plant into its more biologically interesting wild relatives. Several phylogenetic studies dealing with a small subset of species or populations exist. The current study was based on an extensive, representative world-wide sampling of nearly all taxonomic entities. Sequences from the nuclear internal transcribed spacer region (ITS) were used to assess relationship among the several species and subspecies. Five major lineages beside *A. thaliana* could be recognized (*A. halleri*, *A. cebennensis*, *A. lyrata*, *A. arenosa* and *A. croatica*). *A. cebennensis* and *A. halleri* were basal to the remaining lineages. Only within the *A. halleri* group subspecies could be recognized as phylogenetically distinct units with *A. h.* ssp. *dacica* basal and *A. h.* ssp. *gemmifera* most derived. Hybrids between *A. lyrata* and *A. arenosa* and *A. arenosa* and *A. croatica* were detected. The recently proposed allotetraploidy of *A. lyrata* ssp. *kamchatica* with *A. lyrata* and *A. halleri* ssp. *gemmifera* as parental species could be proved for Japan. But there is no evidence for a hybrid origin of the taxon in North America. Taking chloroplast sequence data (*trnL*-Intron + *trnL/F*-IGS) into account, no species-specific distribution patterns of variation could be detected. Either ancestral cpDNA-haplotype diversity which predates separation of the main evolutionary lineage is possible, or reticulation and hybridization among lineages which have transferred cpDNA-haplotypes from one lineage into another.

Introduction

A revision of the genus *Arabidopsis* in the late nineties of the last century excluded about 59 species and included all the members of the former genus *Cardaminopsis* (O'KANE & AL-SHEHBAZ 1997; AL-SHEHBAZ et al. 1999) which resulted in the description of nine species and several subspecies within the newly defined genus *Arabidopsis*. Beside *Arabidopsis thaliana* and *A. suecica*, a hybrid between *A. thaliana* and *A. arenosa*, three main species groups can be recognized: the circumpolar *A. lyrata* group, the mainly European *A. halleri* group with one subspecies extending to East Asia and Japan and the exclusively European *A. arenosa* group, which could extensively expand its distribution range into former glaciated areas of N-Europe in the last one and a half centuries. Additionally three very narrow distributed taxa are known, *A. croatica* (Schott) O'Kane & Al-Shehbaz from Croatia, *A. cebennensis* (DC.) O'Kane & Al-Shehbaz from SE-France and *A. pedemontana* from NW-Italy. Recently new taxa were introduced like *Arabidopsis arenicola* (Richardson ex Hook) Al-Shehbaz, R. Elven, D. Murray & S.I. Warwick (WARWICK et al. 2006), *Arabidopsis kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh and *A. kamchatica* ssp. *kawasakiana* (Makino) K. Shimizu & Kudoh (SHIMIZU et al.

2005). Additionally changes in the subspecies definition in the *A. halleri* group were published (KOLNIK & MARHOLD 2006).

Morphologically the genus can be recognized by the presence of short petiolate leaves, an indumentum of simple trichomes mixed with few-forked ones, usually well-defined basal rosettes, white to lavender or rarely purple flowers and several other characteristics or the absence of those. A key to the species and subspecies is additionally provided by O'KANE & AL-SHEHBAZ (1997) and also found in AL-SHEHBAZ & O'KANE (2002). Nevertheless morphological characters have been demonstrated to be highly homoplasious in the Brassicaceae (e.g. KOCH et al. 2003a). Although the family can be easily recognized by its floral architecture, the division into tribes mainly based on fruit morphology and seed embryology can be subject to considerable convergence and is therefore sometimes taxonomically unreliable (AL-SHEHBAZ et al. 2006). A revision of molecular studies dealing with the Brassicaceae led to a new tribal alignment and placement of the genus *Arabidopsis* in the tribe Camelineae together with genera like *Capsella*, *Erysimum* L., *Neslia* Desv., *Turritis* L. and several others.

Different haploid chromosome numbers are found in the genus. *Arabidopsis thaliana* is characterized by n=5, the amphidiploid *A. suecica* by n=13. Diploid and tetraploid chromosome counts based on n=8 are known in *A. neglecta* and the various subspecies of *A. arenosa* and *A. lyrata* (AL-SHEHBAZ & O'KANE 2002), whereas all subspecies of *A. halleri* show 2n=16 (AL-SHEHBAZ & O'KANE 2002; KOLNIK & MARHOLD 2006). Especially in *A. arenosa* of the Carpathian Mountains high morphological and cytological diversity with diploids to pentaploids and even aneuploids is observed. This led to the description of several taxa with nomina provisoria (MESICEK 1970) which are only weakly defined (O'KANE & AL-SHEHBAZ 1997).

Several studies investigated the relationship of the various taxa within the genus and presented phylogenetic trees estimated with different marker systems but often only few individuals and few taxa (MIYASHITA et al. 1998; KOCH et al. 1999, 2000, 2001; SAVOLAINEN et al. 2000; AL-SHEHBAZ & O'KANE 2002; O'KANE & AL-SHEHBAZ 2003; BECK et al. 2007).

The current study focused on the species beside *A. thaliana* and presented the largest dataset investigated up to now with nearly all species (except *A. pedemontana*) and subspecies defined by O'KANE & AL-SHEHBAZ (1997), based on the analysis of a cpDNA and a nuclear rDNA marker.

Material and Methods

Plant Material

Plant material was obtained from the following herbaria: Biologiezentrum Linz (LI), Herbarium of the ETH Zürich (ZT), Natural History Museum and Herbarium London (BM), Natural History Museum and Herbarium Vienna (W), Herbarium Botanical Institute at Vienna University (WU), the Agriculture and Agri-Food Canada Vascular Plant Herbarium (DAO), the Herbarium of the California Academy of Sciences (CAS), Herbarium, Biology Department Hobart and William Smith Colleges (DH), Herbarium Harvard University (GH) and from the Academy of Sciences Bratislava (SAV). Additional material was collected directly in the field from 1999 – 2004. See Table I (Appendix) for detailed sampling information and collectors. The collected material was deposited at the Herbarium of Heidelberg (HEID). In total 617 accessions were studied (exclusively accessions from different papers). A distribution map of the

studied accessions is shown in Fig. 1-1.

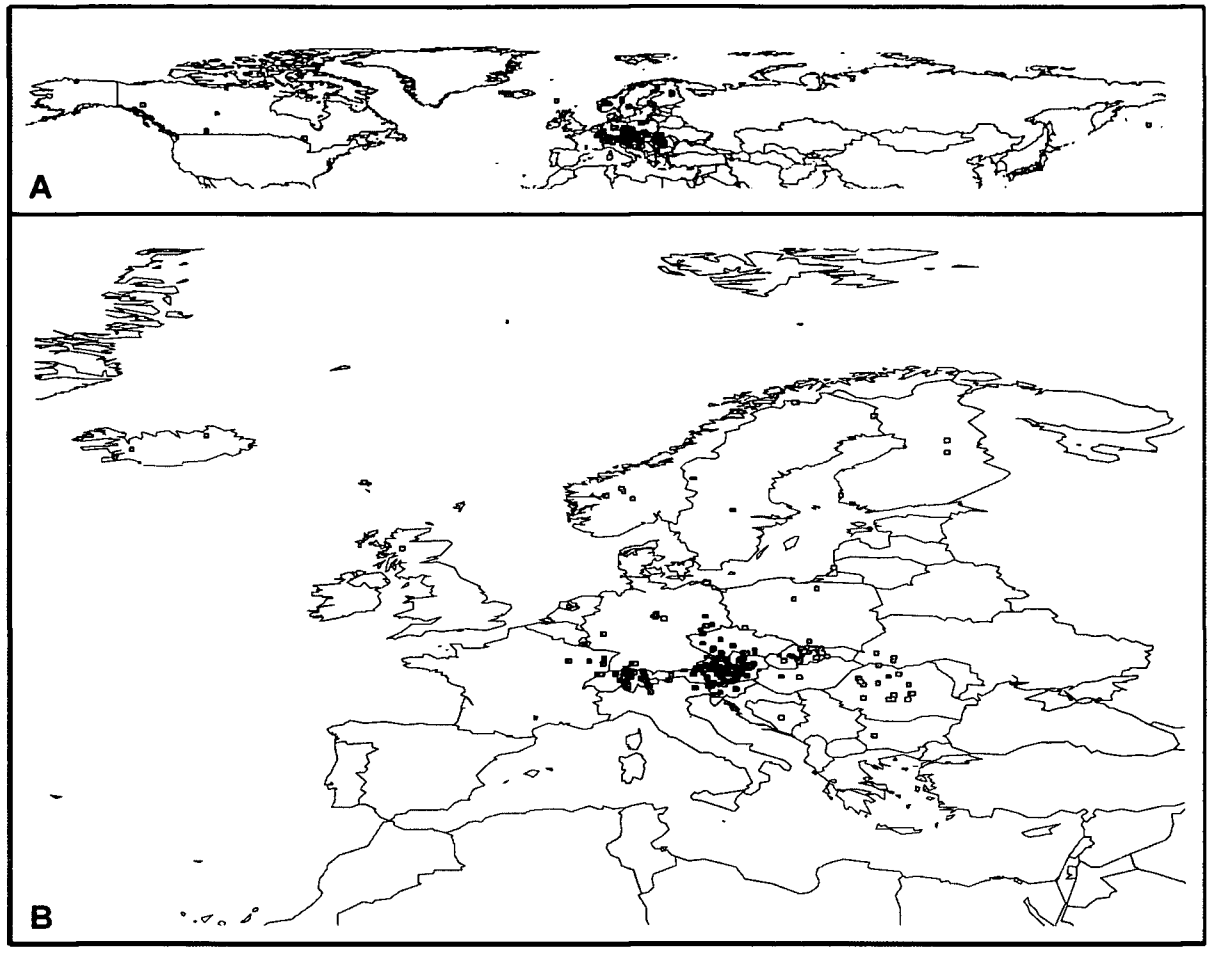

Figure 1-1: Sampling map of the analysed 617 accessions. (A) shows the world-wide sampling. (B) shows the sampling in Europe in more detail.

For information on the various *Arabidopsis* species and their geographic distribution see Tab. 1-1.

Table 1-1: Geographic distribution of the various *Arabidopsis* species (KOCH et al. 2007, in prep.)

Arabidopsis thaliana	native range almost all Europe to central Asia
Arabidopsis arenosa	Europe (differentiation between ssp. *borbasii* and ssp. *arenosa* is difficult)
Arabidopsis cebennensis	SE France
Arabidopsis croatica	Bosnia, Croatica
Arabidopsis halleri	
ssp. *dacica*	Carpathian Mountains, Romania
ssp. *gemmifera*	Russia Far East, NE China, Korea, Japan, and Taiwan
ssp. *halleri*	Europe
ssp. *ovirensis*	Austria only
ssp. *tatrica*	Tatra Mountains, Slovakia
Arabidopsis lyrata	
ssp. *kamchatica*	boreal Alaska, Canada, E Siberia, Russian Far East, Korea, N China, Japan, Taiwan (acc. from Japan are allopolyploids)
ssp. *lyrata*	NE European Russia, Alaska, Canada, parts of the USA
ssp. *petraea*	Europe
A. arenicola	Canada
Arabidopsis neglecta	Carpathian Mountains
Arabidopsis pedemontana	NW Italy
Arabidopsis suecica	Fennoscandinavia and Baltic region

455 accessions, of which the nuclear marker and/or the cpDNA marker system worked, were used for phylogenetic analyses, including all species and subspecies of the former *Cardaminopsis* group (now *Arabidopsis*) with exception of *Arabidopsis pedemontana*. Taxon designation followed AL-SHEHBAZ et al. (1999). Subspecies designation within the *A. halleri* species group followed KOLNIK & MARHOLD (2006). The recently on the species-level introduced *Arabidopsis arenicola* (WARWICK et al. 2006) is discussed.

DNA-extraction and amplification

Total DNA was extracted from comparable amounts of herbarium material as well as from silica-gel dried leaf material following the CTAB protocol (DOYLE & DOYLE 1987) with minor modifications involving grinding of dry leaf tissue in 2-ml tubes using a Retsch swing mill (MM 200), washing of the DNA pellet twice with 70% ethanol and a final RNAse treatment (1 unit). DNA was finally dissolved in 50 µL TE-buffer for long-term storage. 50 µl PCR reactions were performed in a master mix containing 1x TaqBuffer (50 mM KCl, 10 mM Tris-HCl pH 8.3, 15 mM Mg^{2+}; Schott-Eppendorf), 5 µl TaqMaster (Schott-Eppendorf), 0.4 µM of each primer, 0.2 mM of each dNTP, 1 unit Taq DNA polymerase (Schott-Eppendorf), and approximately 2 ng of template DNA using an PTC200 (MJ Research) thermal cycler.

For chloroplast DNA markers thermal cycling started with a denaturation step at 95 °C lasting five min; followed by 35 cycles each comprising 60 s denaturation at 95 °C, 45 s annealing at 42 °C (*trnL*-Intron)/45 °C *trnL/F*-IGS, and 60 s elongation at 72 °C. Amplification ended with an elongation phase at 72 °C lasting ten min, and a final hold at 4 °C. The *trnL*-Intron was amplified using the forward primer 5'-CGA AAT CGG TAG ACG CTA CG-3' and the reverse primer 5'-GGG GAT AGA GGG ACT TGA AC-3' (primer c and d according to TABERLET et al. 1991), which annealed in the first and second exon of the *trnL* gene, respectively. Sequences comprised the complete intron and the second exon of the *trnL* gene. For amplification of the *trnL/F*-IGS primers 5'-GGT TCA AGT CCC TCT ATC CC-3' (primer e according to TABERLET et al. 1991) and 5'-GAT TTT CAG TCC TCT GCT CTA C-3' annealing in the second exon of the *trnL* gene and the *trnF* gene, respectively, were used. Amplified sequences included the complete IGS and the first 18 bases of the *trnF* gene.

The PCR cycling scheme for the ITS was five min at 95 °C; 35 cycles of one min at 95 °C, one min at 38 °C, and one min at 72 °C; 10 min extension at 72 °C; and a final hold at 4 °C. The oligonucleotide sequences, which were used to amplify the complete ITS region, including the 5.8 S rDNA region, had been initially designed by WHITE et al. (1990) and modified by MUMMENHOFF et al. (1997) for ITS4. The forward primer was located at the 3'-end of the 18 S rDNA gene (5'-GGA AGG AGA AGT CGT AAC AAG G-3') and the reverse primer was located at the 5'-end of the 25 S rDNA gene (5'-TCC TCC GCT TAT TGA TAT GC-3'). PCR products spanned the entire ITS1, 5.8 S rDNA, and ITS2 region and were cycle-sequenced directly without cloning. PCR products were checked for length and concentration on 1.5% agarose gels. Only ITS- PCR products were purified using the High Pure PCR product purification kit (Roche). Cycle sequencing was performed using the TaqDyeDeoxy Terminator Cycle Sequencing Kit (ABI Applied Biosystems, Inc.) and the original amplification primers. However, the reverse *trnL/F*-IGS primer was modified adding an additional cytosine to its 3'-end (DOBES et al. 2004). Products were analysed on an ABI 377XL automated sequencer. Cycle sequencing was performed on both strands. In the majority of cases each reaction spanned the complete sequence. The program SeqMan4.0 (DNAstar) was used to edit the complementary sequences.

The *trnL*-Intron and the *trnL/F*-IGS were edited separately and their different

haplotypes were determinated also separately. The *trnL*-Intron haplotypes and the *trnL/F*-IGS haplotypes were submitted to GenBank. DNA sequence data from the ITS region were obtained from a direct sequencing approach. The ITS region is in principle subject to a progress called concerted evolution and multiple copies might indicate either species-specific naturally occurring variation among loci or might demonstrate the result of more recent hybridisation and reticulation (KOCH et al. 2003b). We obtained numerous sequences with ambiguous sites and not totally homogenized ITS copies. Therefore the IUPAC ambiguity symbols were applied to indicate individual single nucleotide polymorphisms, i.e. Y is meant as the presence of A and C, not as an ambiguous reading between A or C. Polymorphic sites were identified on the electrophorograms when two peaks were present at one position and one peak was at most 40% weaker than the other one. Different ITS types were defined and submitted to GenBank. See Table I (Appendix) for detailed information on haplotypes and GenBank accession numbers.

Data analysis: phylogenetic inference and haplotype networks

DNA sequence alignments were performed with MegAlign 4.0 (DNAstar) and BioEdit 7.0.1 and subsequently adjusted by hand. The *trnL*-Intron sequence and the *trnL/F*-IGS were put together with the *trnL*-exon sequence between them (not submitted). For the two combined cpDNA dataset a new haplotype (*trnL*-Intron + *trnL/F*-IGS haplotype) was established. For the several analyses gaps (except polyT's) were coded as additional single characters.

The occurrence of numerous in parallel evolved pseudogenes in the *trnL/F*-IGS (KOCH et al. 2005) leads to non homologous alignments. Therefore the *trn*/F-IGS sequence was excluded from the beginning of the first pseudogene (from position 203 of the correctly aligned *trnL/F*-IGS on to the end of the sequence). Several of the combined *trnL*-Intron + *trnL/F*-IGS haplotypes are therefore summarized. This was done by the program TCS version 1.21 (CLEMENT et al. 2000). These new types, comprising different haplotypes, were then called "suprahaplotypes" (named with capital letters). Additionally a "suprahaplotype" network was constructed with TCS. The setting "gaps = missing" was used, since the gaps had been coded as additional characters at the end of the "suprahaplotype sequence". Two closed suprahaplotype loops occurring in the network were resolved following BITTKAU & COMES (2005): "(i) haplotypes are more likely to be connected to common than to rare haplotypes (the frequency criterion); (ii) haplotypes are more likely to be connected to haplotypes from the same population or region than to haplotypes from distant populations (the geographical criterion)." These predictions are based on coalescent theory (ROSENBERG & NORDBERG 2002).

For the ITS dataset the program TCS version 1.21 (CLEMENT et al. 2000) was used to group the different ITS-types with their high amount of nucleotide polymorphisms. TCS recognized several general ITS-types. Then one representative sequence for each general type with the lowest (or even zero) number of ambiguous sites was selected manually. These sequences were then called "ITS-supratypes" following the term "suprahaplotypes" in the cpDNA dataset. An ITS-supratype network was constructed. Since there were only single nucleotide gaps, the gaps were not coded separately, but the setting gaps = fifth state was used in TCS. The network analysis was first done with a 95% confidential interval which allowed only 3 internal steps and did not add all sequences into one network. Secondly the 90% confidential interval was used (4 internal steps) and finally maximal number of steps were allowed.

A Maximum Likelihood (ML) analysis and a Maximum Parsimony (MP) analysis

were performed for each of the two datasets (ITS-supratypes and cpDNA-suprahaplotypes) using PAUP version 4.0b10 (SWOFFORD 2002). The ML-analyses were performed with the standard input order and changed options like "heuristic search", "random addition of sequences" and "1000 replicates". The appropriate model of sequence evolution for each data set was chosen on the basis of hierarchical likelihood-ratio tests as implemented in Modeltest 3.7 (POSADA & CRANDALL 1998). Of the two appropriate models, which Modeltest 3.7 suggested, the one which uses the Akaike information criterion (AIC) (Akaike 1974) was taken. The model of evolution proposed for the ITS-dataset was GTR+G (Basefreq = (0.2171, 0.2814, 0.2724); Nst = 6; Rmatrix = (1.5126, 2.8065, 3.7894, 0.5713, 6.1211); Rates = gamma; Shape = 0.0933; Pinvar = 0). For the cpDNA-dataset the model F81+G was suggested (Basefreq = (0.3482, 0.1682, 0.1594); Nst = 1; Rates = gamma; Shape = 0.2967; Pinvar = 0). For the MP-analyses constant characters were excluded and the analyses were conducted using the same changed options as for the ML-analyses. Bootstrapping was carried out on 1000 replicates using the "heuristic search" option and "add sequence random". *Arabidopsis thaliana* was used as an outgroup.

Results

ITS

The analysis of 354 world-wide accessions resulted in 102 different ITS-types. This was due to a high number of nucleotide polymorphisms. The program TCS grouped the ITS-types into 23 ITS-supratypes which were coded alphabetically (a-w). These 23 ITS-supratypes were used for further analyses (MP- and ML-analyses and TCS network). The total length of the alignment was 651bp. 60 characters were variable, 20 of them were parsimony informative. The MP-analysis generated 1385 equally parsimonious trees with 72 steps (CI = 0.8889; RI = 0.9036). The 50% majority consensus tree is given in Fig. 1-2.B. The ML-analysis resulted in one tree with a likelihood score of –ln = 1360.60059 which is shown in Fig. 1-2A. Tab. 1-2 gives information concerning the ITS-supratypes, for example in which species and geographic areas (countries) a certain ITS-supratype occurred. The ITS-supratype network as shown in Fig. 1-3 should not be taken to resolve the interspecific relationship of the several *Arabidopsis* species. The reason is that it is as a whole not statistically supported. When the 95% statistical parsimony method of TEMPLETON et al. (1992) was used only the three main species groups *arenosa*, *lyrata, halleri* appeared as three different not connected networks since they were separated from each other and from *A. croatica*, *A. cebennensis* and the outgroup *A. thaliana* by several mutational steps. When the 90% confidence level for network reconstruction was used the *A. halleri* group and the *A. arenosa* group were at least connected. But for solving the intraspecific position of the several ITS-supratypes the single species group networks were used and compared with the phylogenetic trees (ML+MP).

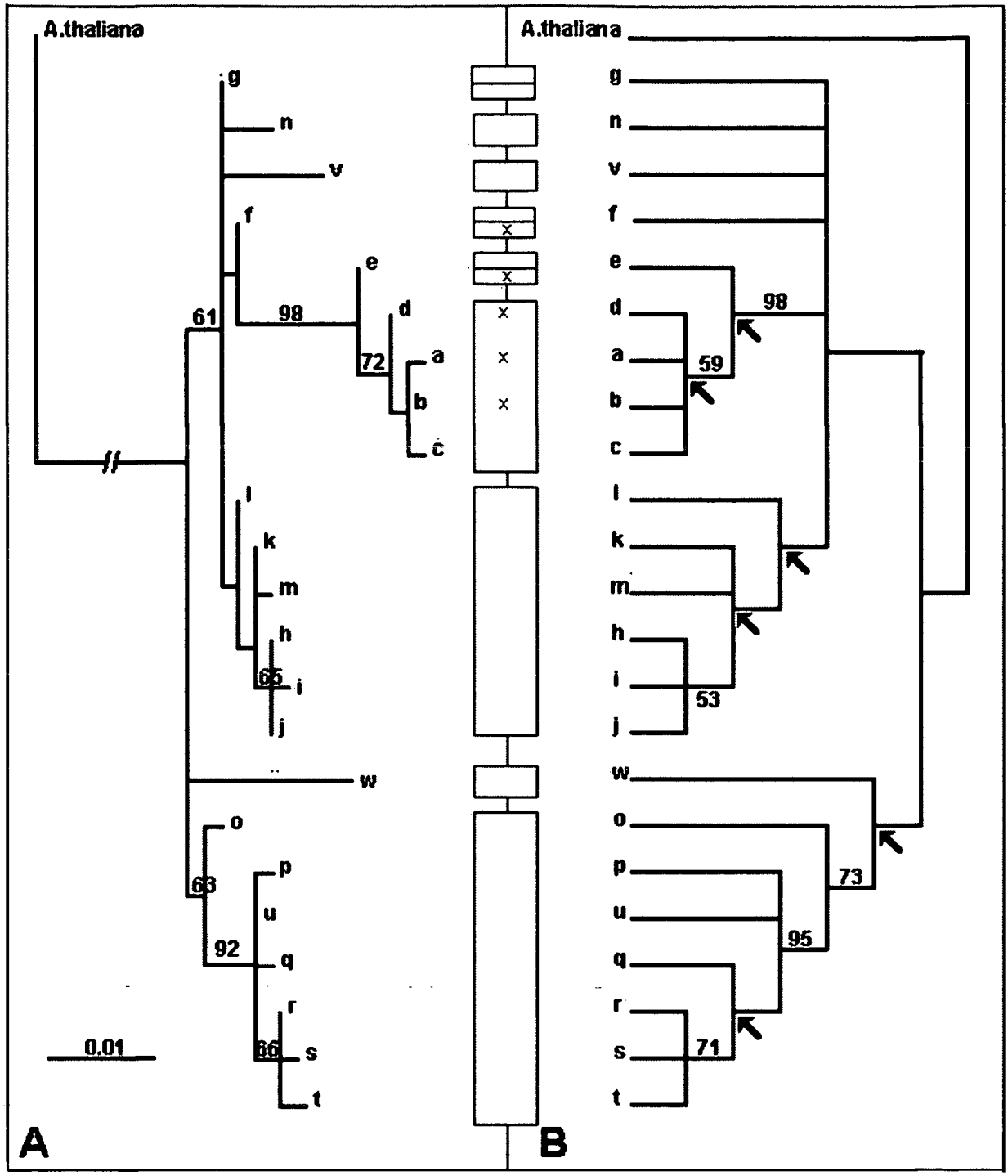

Figure 1-2: Analysis of the ITS-supratypes. Numbers above branches indicate bootstrap percentages where support is > 50%. (A) Maximum likelihood tree estimated with the GTR+F model of substitution. (B) 50% majority rule consensus tree of 1385 most parsimonious trees. Arrows mark groups which are not present in the strict consensus tree.

Phylogenetic position of the different species (using MP and ML)

The ML-tree and the MP-tree were generally congruent with slight differences. There were two (MP) to three (ML) major clades. The first clade comprised the *arenosa*-group, the *lyrata*-group and *A. cebennensis*. Bootstrap support was low (61 % in the ML-analysis and less than 50% in the MP-analysis). The second clade was formed by the *halleri*-group and showed low (63%-ML) to moderate (73%-MP) bootstrap support. In the MP-analysis *A. cebennensis* appeared basal to the *halleri*-group, whereas it formed a separate clade in the ML-analysis. Both placements of *A.*

cebennensis had no good bootstrap support (< 50%). The MP-tree shown in Fig. 1-2.B is the 50% majority rule tree. The strict consensus tree, which is not shown placed *A. cebennensis* in the same way as the ML-tree. Within clade 1 there was a polytomy of two well resolved groups and three to four single ITS-supratypes. One group comprised all *lyrata* ITS-supratypes. Bootstrap support for this group was very high in both analyses (98 %). In the ML-analysis ITS-supratype f appeared basal to the *lyrata*-group with bootstrap support below 50%. This ITS-supratype comprised one *A. croatica* sequence and one sequence, which seemed to be a hybrid between *A. arenosa* and *A. lyrata*. The second well resolved group compriseed the majority of the *arenosa* ITS-supratypes. Single ITS-supratypes of the clade 1 polytomy belonged to *A. croatica* and *A. arenosa.* In one of these ITS-supratypes (g) one ITS-type of *A. croatica* and one ITS-type of *A. arenosa* were grouped.

Phylogenetic signals within the major groups

The phylogenetic position of the several ITS-supratypes within the *halleri* group was congruent in all three analyses (ML, MP and TCS-network). Within this group ITS-supratype o was basal. It comprised only accessions from Romania, which were assigned to subspecies *dacica*, except one individual, which was determined as a *A. halleri* ssp. *halleri* by its collector. This basal position of ITS-supratype o is supported by moderate bootstrap values (63% ML and 73% MP). ITS-supratype u was the most common type and was found all over Europe in ssp. *halleri*, ssp. *tatrica* (Pawl.) Kolnik (all accessions) and ssp. *dacica* (Heuff.) Kolnik (only three accessions). Two rare ITS-supratypes were derived from ITS-supratype u: ITS-supratype q, which was only found in one individual in Italy and ITS-supratype p which occurred in the *A. halleri* ssp. *ovirensis* (Wulfen) O´Kane & Al-Shehbaz accessions from Mount Obir (Carinthia, Austria). The most derived ITS-supratypes (r, s, t) belonged to *A. halleri* ssp. *gemmifera* (Matsum.) O´Kane & Al-Shehbaz from Japan and formed a separate group with moderate bootstrap support (66% ML; 71% MP).

Within the *arenosa*-group there was no phylogenetic signal recognizable. The different species within the group, *A. arenosa*, *A. petrogena* (A. Kern) Kolnik & Marhold, *A. neglecta* and two with nomina provisoria described taxa *Cardaminopsis nitida* and *Cardaminopsis carpatica* (MESICEK 1970), did not have species-specific ITS-supratypes. They shared several ITS-supratypes. There was not even a geographical structuring in the distribution of the ITS-supratypes recognizable. In the network of the *arenosa* group (Fig. 1-3) three ITS-supratypes formed a loop.

In the *lyrata*-group ITS-supratypes a, c and d were only present in Europe, that meant in ssp. *petraea*. ITS-supratype b was present in all three subspecies, ssp. *petraea* (L.) O´Kane & Al-Shehbaz, ssp. *lyrata* and ssp. *kamchatica*. The most basal ITS-supratype e occurred in ssp. *lyrata* and ssp. *kamchatica* accessions from North America and in one individual from Austria which was a hybrid between *A. arenosa* and *A. lyrata* ssp. *petraea*. In the *lyrata*-group network the ITS-supratypes a, b and d formed a loop.

Table 1-2: Species/subspecies assignment and country assignment of the different ITS-supratypes. Different colours correspond to different species/species-groups.

ITS	species	country
a	*A. l.* ssp. *petraea*	Austria, Czech R., Iceland, Norway, Scotland
b	*A. l.* ssp. *petraea*	Austria, Czech R., Germany, Sweden, Norway, Iceland, Faroer
	A. l. ssp. *lyrata*	Aleuten
	A. l. ssp. *kamchatica*	Alaska, Canada, Japan
c	*A. l.* ssp. *petraea*	Iceland
d	*A. l.* ssp. *petraea*	Austria
e	*A. l.* ssp. *lyrata*	Canada
	A. l. ssp. *kamchatica*	Alaska
	A. arenosa (x *A. l.* ssp. *petraea*)	Austria
f	*A. arenosa* (x *A. l.* ssp. *petraea*)	Austria
	A. croatica	Croatia
g	*A. arenosa* (incl. *A. a.* ssp. *borbasii*)	Austria, Germany, Czech R., Slovakia, Switzerland, Sweden, Fennia, Norway, Lithuania, Poland, Romania, Slovenia
	A. neglecta	Poland, Slovakia, Austria?
	A. petrogena	Slovakia
	A. nitida	Slovakia
	A. carpatica	Slovakia
	A. croatica	Croatia
h	*A. arenosa* (incl.*A. a.* ssp. *borbasii*)	Romania, Germany, Luxemburg
	A. nitida	Slovakia
	A. petrogena	Romania
	A. carpatica	Slovakia,
	A. neglecta	Slovakia, Romania
i	*A. neglecta*	Slovakia
j	*A. arenosa*	France, Switzerland, Austria, Bosnia, Germany
	A. neglecta	Romania
	A. carpatica	Slovakia
k	*A. arenosa* (incl. *A. a.* ssp. *borbasii*)	Austria, France, Switzerland
	A. neglecta	Romania
l	*A. arenosa*	Switzerland,
	A. petrogena	Romania
m	*A. arenosa* (incl. *A. a.* ssp. *borbasii*)	Slovenia, Austria, Czech R., Germany, Yugoslavia, Switzerland
n	*A. petrogena*	Slovakia
	A. arenosa	Hungary
o	*A. halleri*	Romania
	A. h. ssp. *dacica*	Romania
p	*A. h.* ssp. *ovirensis*	Austria
q	*A. halleri*	Italy
r	*A. h.* ssp. *gemmifera*	Japan
s	A. h. ssp. *gemmifera*	Japan
t	*A. h.* ssp. *gemmifera*	Japan
u	*A. halleri*	Slovakia, Germany, Austria, Switzerland, Italy, Romania, Poland, Ukraine, Hungary, Slovenia
	A. halleri ssp. *tatrica*	Slovakia
	A. halleri ssp. *dacica*	Romania
v	*A. croatica*	Croatia
w	*A. cebennensis*	France

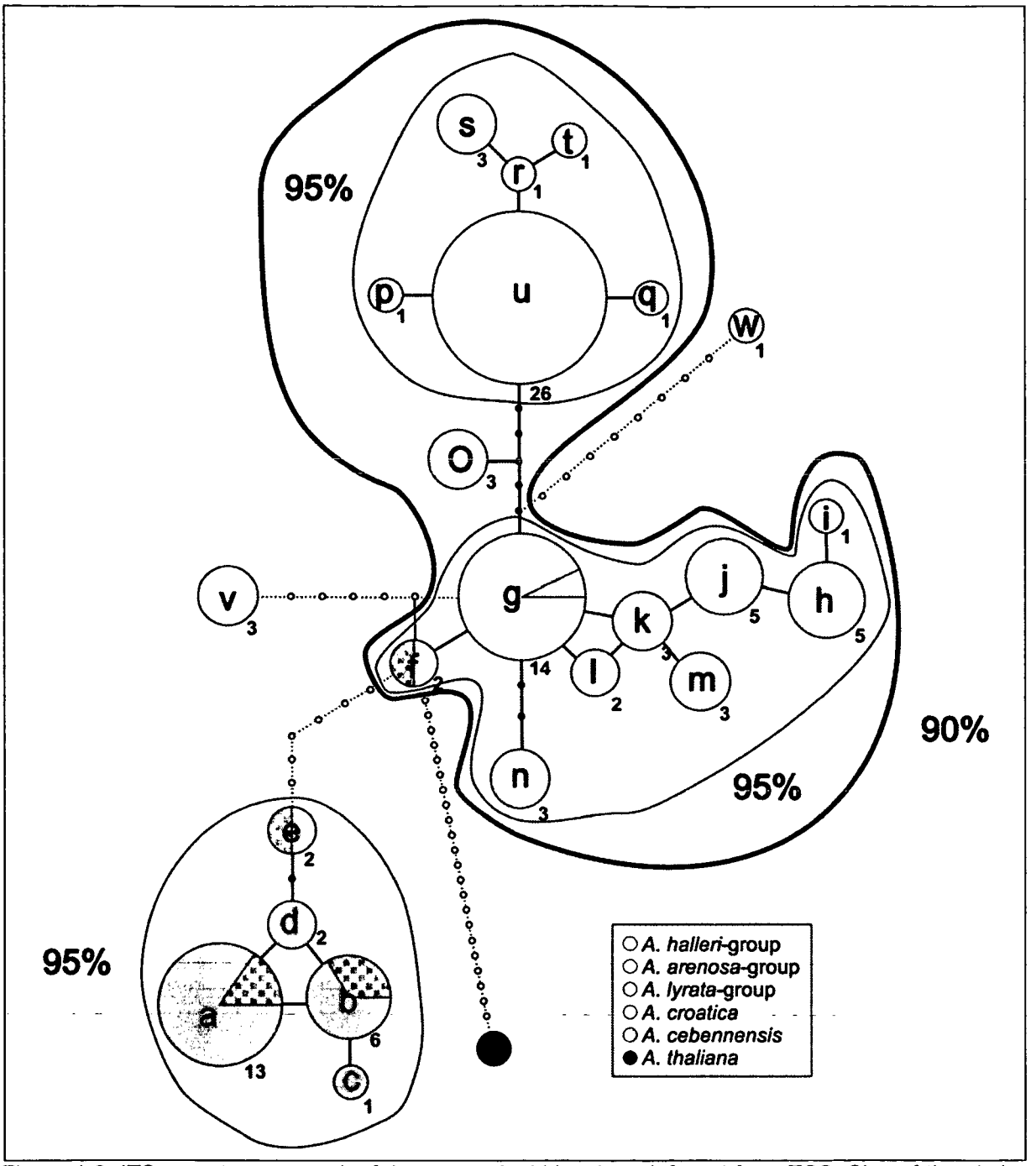

Figure 1-3: ITS-supratype network of the genus *Arabidopsis* as inferred from TCS. Size of the circles corresponds to the number of ITS-types (beside the circle) that are combined due to ambiguities. Using the 95% statistical parsimony method of TEMPLETON et al. (1992) the three main species groups appear as different networks. Using the 90% confidence level for network reconstruction the *A. halleri*-group and the *A. arenosa*-group are connected.

CpDNA

The analysis of 344 accessions of the world-wide dataset resulted in 148 haplotypes. It has to be mentioned that in the Wachau-study (chapter 3) with its detailed sampling on population level additional 31 haplotypes, mostly rare were found, which were not included in this dataset. The length of the haplotypes varied due to presence and absence of several tandem repeated pseudogenes from 863bp to 1386bp. The mutation rate for this region varies between 2.4 to 3.8 x 10^{-8} mutations/site/year and exeeds the normal mutation rate of the entire *trn*L-intron/*trn*L-F spacer region by a factor of 20 (KOCH et al. 2005). After exclusion of the pseudogene region the

haplotypes were grouped together to so called "suprahaplotypes", which were then used for phylogenetic analyses. The total length of the alignment including the outgroup was 767bp. 36 characters were variable and seven characters were parsimony informative. The MP-analysis resulted in 288 trees with 47 steps (CI = 0.9362; RI = 0.8929). The 50% majority rule consensus tree is given in Fig. 1-4B. The ML-analysis resulted in two trees. The better fitting tree was chosen (Fig. 1-4A) The tree had a likelihood score of –ln = 1391.12492.

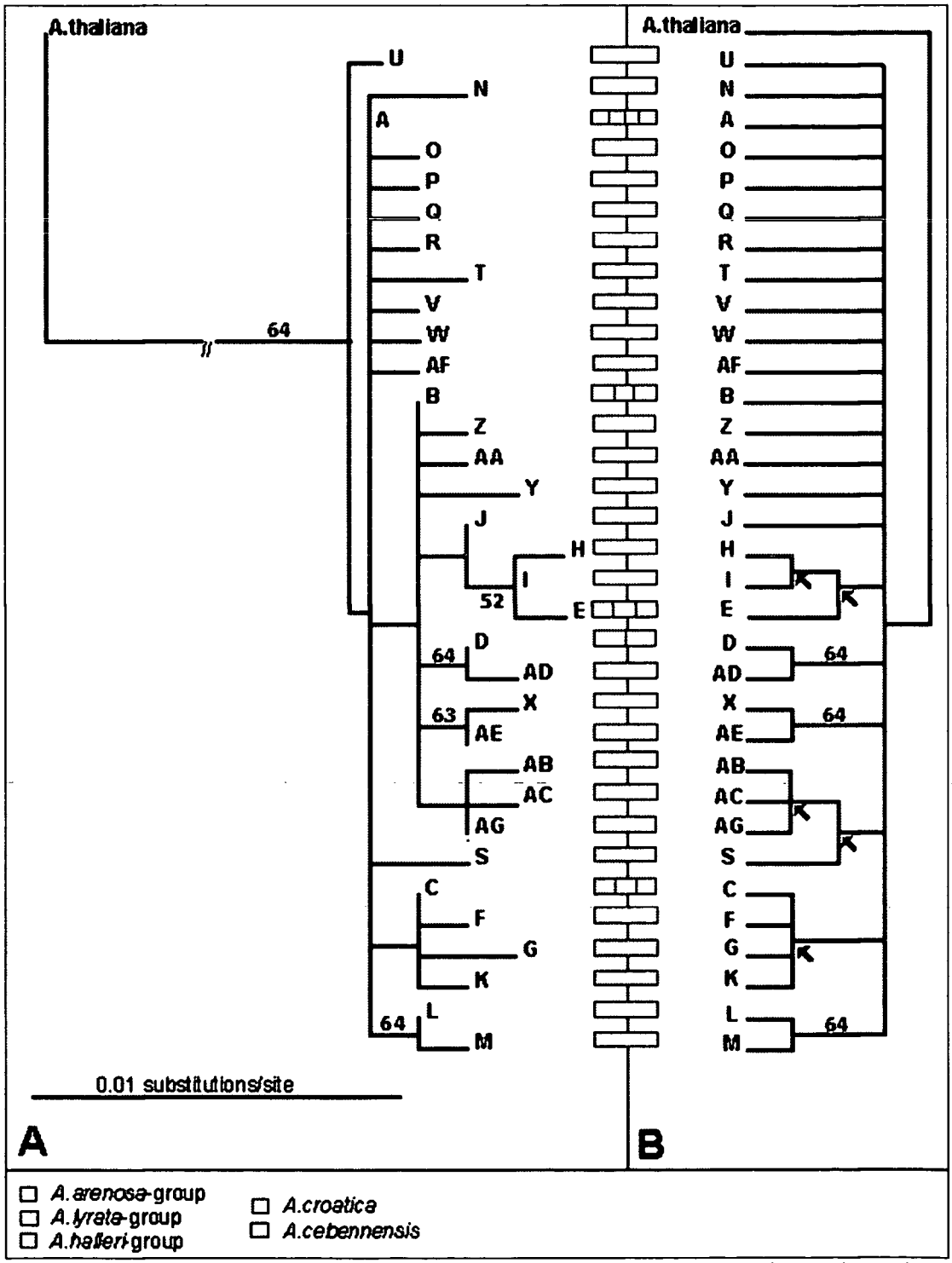

Figure 1-4: Analyses of the cpDNA-suprahaplotypes (capital letters). Numbers above branches indicate bootstrap percentages where support is > 50%. (A) one of the two maximum likelihood trees estimated with the F81+G model of substitution. (B) 50% majority consensus tree of 288 most parsimonious trees. Arrows mark groups not present in the strict consensus tree.

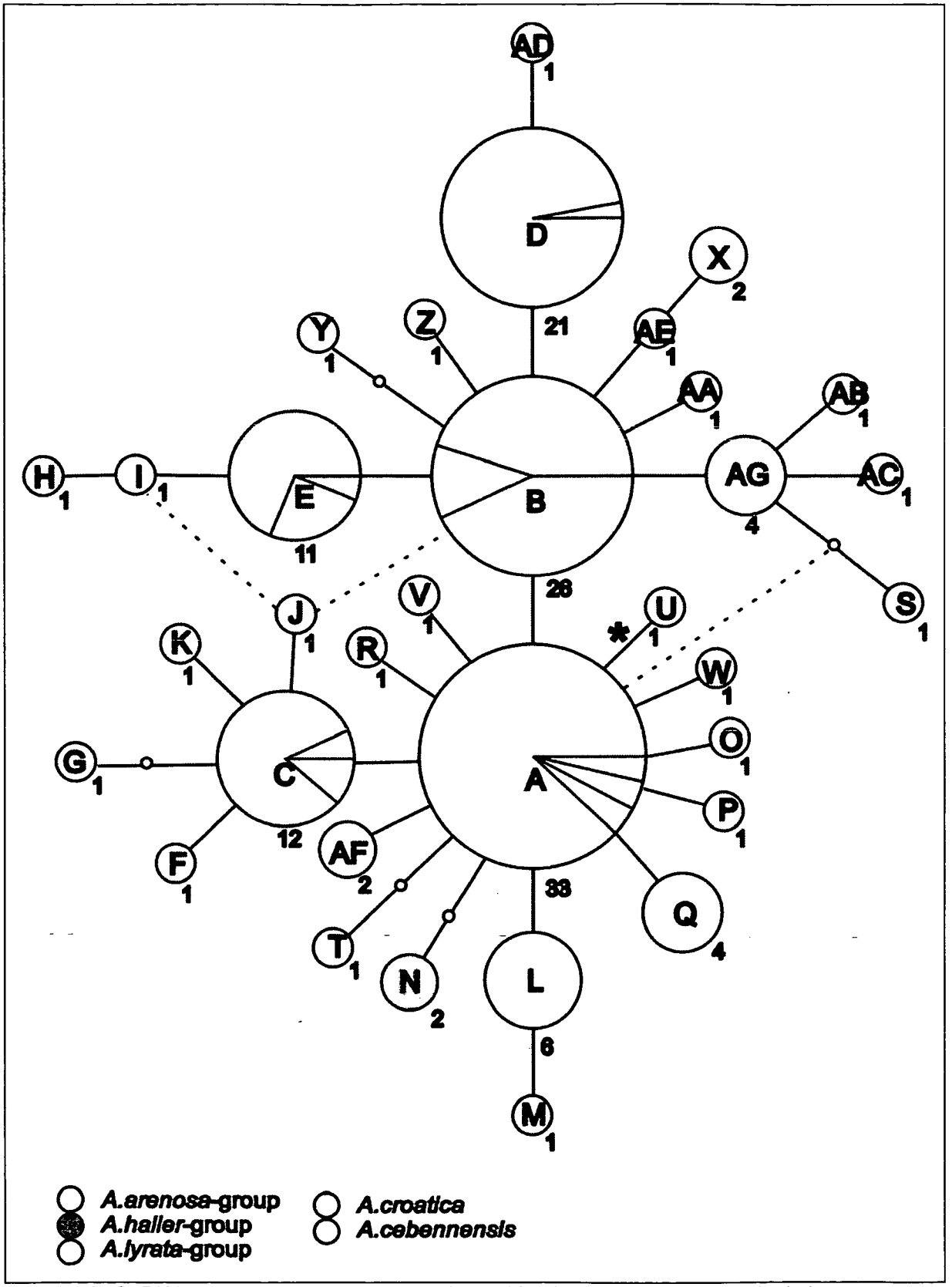

Figure 1-5: CpDNA-suprahaplotype network estimated by the program TCS. Each circle represents a suprahaplotype (named with capital letters) and the size corresponds with the number of haplotypes (below each circle), which are summarized to a suprahaplotype (for details see material and methods). Broken lines show closed suprahaplotype-loops which were resolved following BITTKAU & COMES (2005).

Phylogenetic signal
The ML-tree (Fig. 1-4A) was better resolved than the MP-tree (Fig. 1-4B), but this was not supported by good bootstrap values. In general both trees did not have good bootstrap values (best value 64%). The ITS results were in so far better than the cpDNA results as that the species and species groups formed more or less separated lineages, whereas in the cpDNA data there was no good structure present. This might be due to the high amount of suprahaplotype sharing in the cpDNA data (Fig. 1-5). All the inner suprahaplotypes (A, B, C, E and D), which were regarded as the older/ancestral suprahaplotypes relative to the suprahaplotypes at the tips of the network showed a more or less high amount of suprahaplotype sharing. For geographical assignment of the different suprahaplotypes to the three main species groups see also chapter 2 (Fig. 2-1; Fig. 2-3 and Fig. 2-7).

Discussion

The sampling of the former genus *Cardaminopsis*, which together with *Arabidopsis thaliana* forms a newly defined genus *Arabidopsis* (O'KANE & AL-SHEHBAZ 1997; AL-SHEHBAZ et al. 1999), presented here is considerably larger than in any previous study. For the several analyses more accessions with a representative geographical distribution of all known species and subspecies were used with exception of *A. pedemontana*, due to lack of material. Meanwhile KOCH & MATSCHINGER (2007) presented data of *A. pedemontana*. Additional several species and subspecies, which were not correctly published by MESICEK (1970) but belong to the *arenosa* group were also included into the dataset.

There are several previous molecular studies which included some of the species and subspecies, which were used in the current study. These studies were taken for comparison with the results of the current study. The most extensive studies to compare this dataset with are from O'KANE and AL-SHEHBAZ (Al-SHEHBAZ & O'KANE 2002; O'KANE & AL-SHEHBAZ 2003) based on ITS sequence data. The following studies miss *A. arenosa*-group and *A. cebennensis* (MIYASHITA et al. 1998; KOCH et al. 1999; KOCH et al. 2001), but interestingly some of them are based on different marker systems beside ITS.

The various molecular studies show partly different phylogenetic relationships within the genus *Arabidopsis*, depending on the taxa, phylogenetic markers and algorithm for calculation, used.

Phylogenetic relationships within *Arabidopsis* beside *A. thaliana*
Arabidopsis thaliana was clearly separated from the rest of the genus. Beside *A. thaliana* three major species groups, *A. halleri*, *A. lyrata* and *A. arenosa* were recognized as well as two single species *A. cebennensis* and *A. croatica*.

***Arabidopsis cebennensis, A. pedemontana* and *A. halleri* group**
Phylogenetic analyses of this dataset gave the well separated *A. halleri* group and *A. cebennensis* a basal position relative to the *lyrata*-group, the *arenosa*-group and *A. croatica*. In the ML-analysis *A. cebennensis* and the *halleri* group were sister groups whereas the MP method grouped *A. cebennensis* basal to the A. *halleri*-group. This relationships is also not clear in the molecular studies of AL-SHEHBAZ & O'KANE (2002) and O'KANE & AL-SHEHBAZ (2003). In the first *A. cebennensis* is basal to the rest of the species beside *A. thaliana*, whereas in the second there is a polytomy and the positions of the different species are not resolved. In KOCH & MATSCHINGER (2007) *A.*

pedemontana forms a clade together with *A. cebennensis*. They are separated from each other and their sister clade by several mutational steps which leads the authors to the conclusion that *A. cebennensis* and *A. pedemontana* are relatively old diploids and genetically well-defined species with a relict distribution in SE-France and NW-Italy, respectively.

Within the A. *halleri*-group the subspecies were more or less phylogenetically distinct groups. *A. h.* ssp. *dacica* accessions with ITS-supratype o were basal. This is also supported by AL-SHEHBAZ & O'KANE (2002) and O'KANE & AL-SHEHBAZ (2003). ITS-supratype u was derived from ITS-supratype o. It comprised nearly all *A. h.* ssp. *halleri* accessions and all *A. h.* ssp. *tatrica* accessions. Interestingly three accessions which were determined as *A. h.* ssp. *dacica* were also found within ITS-supratype u. It is possible that they are determined wrongly.

A. h. ssp. *tatrica* is an endemic of the Western Carpathian Mountains with its distribution centre in Slovakia (KOLNIK & MARHOLD 2006). Although it can be more or less recognized morphologically, it shared its ITS-supratype (u) with *A. halleri* ssp. *halleri* (and some individuals of *A. h.* ssp. *dacica*). JONES & AKEROYD (1993) report that intermediates between ssp. *halleri* and ssp. *ovirensis* occur in the Tatra Mountains. The question is, if *A. h.* ssp. *tatrica* is a real subspecies or if it is just within the morphological range of *A. h.* ssp. *halleri*.

The *A. h.* ssp. *ovirensis* accessions from Mount Obir (locus classicus) did not cluster with the *A. halleri* ssp. *dacica* accessions. Formerly both subspecies were treated as one subspecies, namely *A. h.* ssp. *ovirensis* (AL-SHEHBAZ et al. (1999). The accessions from Mount Obir (*A. h.* ssp. *ovirensis*) showed an unique ITS-supratype (p) and were separated from the typical ITS-supratype of *A. h.* ssp. *dacica* accessions (o) by four mutational steps. This is consistent with the findings of KOLNIK & MARHOLD (2006), "...material from the type locality of *A. h.* ssp. *ovirensis* in Austria is well different from the populations occurring in the Eastern and Southern Carpathians.", Consequently they treat the *A. h.* ssp. *ovirensis* accessions from the Carpathian Mountains and the Balkan Peninsula as a different subspecies, namely *A. h.* ssp. *dacica*.

A. h. ssp. *gemmifera* was the most derived subspecies. All the studied accessions of *A. halleri* ssp. *gemmifera* from Japan formed a group, with three different ITS-supratypes (r, s, t). They were derived from common European ITS-supratype u. That means, that they most likely originated from ssp. *halleri*. This is congruent with the finding of AL-SHEHBAZ & O'KANE (2002).

Phylogenetic relationship among *A. arenosa*-group, *A. lyrata-group* and *A. croatica*

The *A. arenosa*-group, the *A. lyrata*-group and *A. croatica* together formed a sister clade to *A. halleri* and *A. cebennensis* (based on the ML-tree). Within this clade, there was a polytomy of an A. *lyrata*-group clade, an A. *arenosa*-group clade, a single A. *arenosa*-group ITS-supratype and two A. *arenosa*-group/*croatica* ITS-supratypes. The three species/species-groups were more or less in contact with one another. *A. arenosa* and *A. lyrata* ssp. *petraea* were connected via hybrids (hybrid ITS-sequences), as well as *A. arenosa* and *A. croatica*. The phylogenetic relationship of the three species/species-groups in the current study differes from that of AL-SHEHBAZ & O'KANE (2002). They show an ITS tree with *A. arenosa* basal to a clade formed by *A. lyrata* and *A. croatica*. Whereas in O'KANE & AL-SHEHBAZ (2003) the three species/species-groups are polytomious. This seems more reliable and is congruent with the results of the present study. A correct and well resolved phylogenetic relationship of *A. croatica*, *A. arenosa*-group and *A. lyrata*-group to each

another cannot be given. A reason might be that the clade is too young and connected via hybrids.

A discrepancy between morphological and molecular data concerning *A. croatica* arises in AL-SHEHBAZ & O'KANE (2002): On the one hand "*A. lyrata* (and its subspecies) form a well supported clade in a sister group relationship of *A. croatica*" and on the other hand "*A. croatica* is morphologically very close to and might represent a subspecies of the earlier published *A. arenosa*." It can be suspected that *A. croatica* arose via hybridization of *A. arenosa* and *A. lyrata*. This is supported by the ITS-supratypes f and g (Tab. 1-2), which *A. croatica* either shared with *A. arenosa* or with a hybrid between *A. lyrata* and *A. arenosa*.

Phylogenetic relationship within *A. arenosa*-group and within *A. lyrata*-group

Six out of ten ITS-supratypes of the *A. arenosa*-group formed a clade. Only one branch of this clade was supported by bootstrap values larger than 50%. Within this clade there was no phylogenetic structure recognizable as it was in *A. halleri*. The different species within the *A. arenosa*-group did not have their own ITS-supratypes, they shared them with nearly every other species within the *A. arenosa*-group. Like AL-SHEHBAZ & O'KANE (2002) described the situation for *A. arenosa* and *A. neglecta*, the same is found here: "The different species must have diverged so recently that paraphyly in the nrDNA ITS gene tree is still seen."

All ITS-supratypes of the *A. lyrata*-group formed a monophyletic clade. Most of the hybrids between *A. lyrata* ssp. *petraea* and *A. arenosa* fell within the *A. lyrata*-group clade (see TCS in material and methods). Within the *A. lyrata*-group there was phylogenetic structure recognizable in contrast to the *A. arenosa*-group, but this structure was not straight forward. The most derived group in the *A. lyrata*-group clade was formed by the ITS-supratypes a, b and c. ITS-supratype b occurred in all three subspecies, whereas ITS-supratype a and c only occurred in ssp. *petraea*. The most basal ITS-supratype e, within the *A. lyrata-group* clade occurred in two N-American accessions from *A. lyrata* ssp. *lyrata* and *A. lyrata* ssp. *kamchatica*. The assignment of these accessions to the two subspecies could be incorrect. Due to AL-SHEHBAZ & O'KANE (2002) the three subspecies of *A. lyrata* can be distinguished, but with some difficulty, especially in areas where their distribution ranges overlap, as it is in N-America. However its is interesting that these two accessions are from N-America and that they are basal in the *A. lyrata*-group clade.

In both species groups, *A. arenosa* and *A. lyrata*, a loop was found within the ITS-supratype network. This might be due to homoplasy or recombination (TEMPLETON et. al. 1992). Contact and mixture of migrating populations with different ITS-supratypes is possible.

Comments on *Arabidopsis kamchatica*, *A. k.* ssp. *kawasakiana* and *A. arenicola*

SHIMIZU et al. (2005) proposed that "the taxon corresponding to *Arabis lyrata* var. *kamchatica* Fisch. ex DC., including samples corresponding to *Arabis kawasakiana* Makino, originated with allotetraploidy between two diploid taxa, *Arabidopsis halleri* ssp. *gemmifera* and *Arabidopsis lyrata*". They therefore recognized *Arabidopsis kamchatica* as a distinct species from *Arabidopsis lyrata*, with the subspecies *Arabidopsis kamchatica* ssp. *kawasakiana*. BECK et al. (2007) analysed one *A. lyrata* ssp. *kamchatica* accession from Japan and one from Taiwan with the nuclear marker system *Atmyb2*. In both double fragments, the smaller part of a clade with *A. lyrata* s.l. sequences and the larger part of clade with sequences from *A. halleri* s.l., were detected, which is strong evidence for hybridisation involving members of these two taxa.

The herein analysed accessions of *Arabidopsis lyrata.* ssp. *kamchatica* and *Arabis kawasakiana* accessions from Japan were characterized by cpDNA-suprahaplotype AD, which derived from the common *A. halleri* suprahaplotype D. The same accessions were defined by ITS-supratype b, which was characteristic for *A. lyrata.* This strongly confirmes the hypothesis that *Arabidopsis lyrata* ssp. *kamchatica/ Arabis kawasakiana* from Japan indeed represents a hybrid between *A. halleri* ssp. *gemmifera* and *A. lyrata.* However, no *A. lyrata* accessions from outside Japan analysed herein were characterized by cpDNA-suprahaplotype AD and thus not favouring any hybridization scenario with *A. lyrata* ssp. *gemmifera.* The hybridization event seemed to be limited to Japan and therefore the American *Arabidopsis lyrata* ssp. *kamchatica* accessions could keep their name but the Japanese *Arabidopsis lyrata* ssp. *kamchatica* and *Arabis kawasakiana* accessions should be named different from them (KOCH & MATSCHINGER 2007)

Arabidopsis arenicola as analysed by WARWICK et al. (2006) is characterized by cpDNA-suprahaplotype A and ITS-supratype e, which supports closest relatedness to *A. lyrata.* Interestingly ITS-supratype e was the most basal ITS-supratype in the ITS network (Fig. 1-3).

Comparison of nuclear and cpDNA data

While the ITS network (Fig. 1-3) provides species-specific distribution patterns of genetic variation, the cpDNA network (Fig. 1-5) gives a different picture. The oldest/inner suprahaplotypes of the network were shared extensively among almost all species lineages as characterized by the ITS analysis. Only the younger haplotypes at the tips of the cpDNA-network were species-specific. This incongruence between nuclear and cpDNA data can be explained in two ways (KOCH et al. (2003b); SCHAAL et al. (1998); KOCH & MATSCHINGER 2007): Firstly, ancestral cpDNA type diversity predates separation of the main evolutionary lineages. This scenario is very likely, and correlates well with the fact, that suprahaplotypes from internal position of the network consist of more haplotypes than those at the tips of the network. However, this hypothesis requires that an old centre of genetic diversity is congruent to a centre of origin of the various lineages. The chloroplast type from *A. cebennensis* is congruent with such a hypothesis, because this haplotype (T) is directly connected to the ancestral suprahaplotype A. Secondly, reticulation and hybridization among lineages have transferred cpDNA types from one lineage into another. There is some evidence that this might be true for *A. arenosa* and *A. lyrata* ssp. *petraea* which form a hybrid zone at the edge of the Eastern Alps in Austria. Plastid capture between the two species was observed (see chapter 3).

Chapter 2

Phylogeography of the three main species groups *A. halleri*, *A. lyrata* and *A. arenosa*

Abstract

This study presents a phylogeographic framework of the three main species groups *A. halleri*, *A. lyrata* and *A. arenosa* on a representative scale which has been lacking up to now. The mesophytic *A. halleri* seems to have suffered most from glaciation in Europe since it showed the lowest number of cpDNA-suprahaplotypes compared to the other species groups. Its centre of diversity and maybe also its centre of origin was found in the non-glaciated parts of the Eastern Alps in Austria. *Arabidopsis lyrata*, the only species group which successfully colonized the northern regions with a circumpolar distribution, is supposed to have colonized North America at least two different times. The species group might have survived the last glaciation in America in Beringia and also southeast of the ice sheet. In Europe periglacial survival in the Northern regions and in the non-glaciated parts of the Eastern Alps of Austria is likely. In *A. arenosa* group the centre of diversity was found in the Carpathian Mountains. The species group also seems to have a good migration ability since it has been neosynanthrop in the formerly glaciated North for more than 100 years and showed the highest amounts of haplotype/suprahaplotype sharing among the regional European groups.

Introduction

Since its presentation by AVISE et al. (1987) phylogeography is an exciting topic, which helps us to understand the evolution of taxa in space and time. First applied to animal taxa, it is now a common field of research in plants. In the Brassicaceae about 100 molecular studies on biogeography and phylogeography are compared and discussed in a recently published review (KOCH & KIEFER 2006). Several factors, like landscape morphology, soil composition, competition and climate show influence on the distribution of species (KOCH & KIEFER 2006), but are overlain by the drastic factor of Pleistocene glaciations and deglaciation cycles. During the Quaternary each species went through many contractions and expansions of range, characterized by extinctions of populations in areas which were glaciated when the temperature decreased, and expansion from refugia, when the temperature increased (Taberlet 1998).

In the genus *Arabidopsis* phylogeographic studies on a broader scale are available for *A. thaliana* (SHARBEL et al. 2000) and *A. l.* ssp. *petraea* (ANSELL 2004). Some publications focus in at least some aspects on this topic or draw conclusions which are interesting for phylogeographic questions (JONSELL et al. 1995; VAN TREUREN et al. 1997; CLAUSS et al. 2002; WRIGHT et al. 2003; KAWABE & MIYASHITA 2003; RAMOS-ONSINS et al. 2004; CLAUSS & MITCHELL-OLDS 2006; CLAUSS & KOCH 2006). Comparing the three main species groups beside *A. thaliana*, most is known about *A. lyrata*, little about *A. halleri* and virtually nothing about *A. arenosa*.

Genetic data from *A. lyrata* ssp. *petraea* demonstrate greater gene diversity for Central European populations (Bavaria) than for Scandinavian and Russian

populations (CLAUSS et al. 2002) and have led to the assumption of periglacial survival of *A. l.* ssp. *petraea* in Central Europe during the last glacial maximum (ANSELL 2004; CLAUSS & MITCHELL-OLDS 2006). High genetic differentiation is shown between populations from Russia and other parts of the northern species area such as Sweden, Norway or Iceland (JONSELL et al. 1995; VAN TREUREN et al. 1997) and periglacial permafrost regions in Central Europe (Clauss et al. 2002; CLAUSS & MITCHELL-OLDS 2006). Several studies suggest a bottleneck of North American *A. lyrata* ssp. *lyrata* based on low level of variation if compared with *A. l.* ssp. *petraea* (WRIGHT et al. 2003; RAMOS-ONSIS et al. 2004).

There are some indications that population level variability of *A. halleri* in Europe is less than of *A. lyrata* (CLAUSS et al. 2002; RAMOS-ONSIS et al. 2004) and DNA variation in *A. halleri* ssp. *gemmifera* is low (KAWABE & MIYASHITA 2003).

From *A. arenosa* we know a high number of different morphotypes and cytotypes from the Carpathians (MESICEK 1970) which are most likely explained by the glacial history of this region (KOCH & CLAUSS 2006).

The aim of the current study was to provide a phylogeographic framework comprising the main species groups *A. halleri*, *A. lyrata* and *A. arenosa* on a representative scale. The study was based on maternally inherited cpDNA sequence data which were compared with nuclear data.

Material and methods

Plant material

In this chapter the accessions mentioned in chapter 1 were used.

DNA-extraction and amplification

For detailed information see chapter 1.

Data analysis

For phylogeographic analyses the accessions were divided into 9 regional groups, based on geography (e.g. mountain ranges, glaciation):

+ glaciated North – comprising the areas of Europe, which were glaciated at the maximum extend of Pleistocene glaciation (Iceland, Norway, Sweden, Fennia, Denmark, coastal areas around the Baltic Sea, N-Ireland, N-Great Britain)
+ America (Alaska, Canada, Northern USA)
+ Japan
+ Central European permafrost region – comprising the region in Europe between the glaciated North and the Alps
+ Western Alps
+ glaciated part (during the Pleistocene) of the Eastern Alps
+ non-glaciated East – comprising the non-glaciated part of the Eastern Alps and the non-glaciated area between the Eastern Alps, the Central European permafrost region and the Western Carpathian Mountains.
+ Western Carpathian Mountains
+ Southern and Eastern Carpathian Mountains

The three main species groups (*halleri*, *lyrata*, *arenosa*) were analysed separately. Genetic diversity was estimated both for suprahaplotypes and haplotypes as haplotype richness (R, the number of different haplotypes corrected for sample size

through rarefaction; EL MOUSADIK & PETIT 1996) and effective diversity (according to GREGORIUS 1978):

$$V_a = \left(\sum_{i=1}^{n} p_i^2\right)^{-1} \qquad p_i = \text{frequency of haplotype i}$$

Nucleotide diversity (NEI 1987) could only be estimated for suprahaplotypes. Genetic differentiation calculations were performed between all pairs of regions and among all regions (on condition that the species occurred in this region) with an AMOVA (analysis of molecular variance) using the program Arlequin (EXCOFFIER et al. 2005). Φ_{ST}, an estimate of differentiation taking into account the molecular distance between haplotypes (number of pairwise differences) and F_{ST}, an estimate of differentiation based only on allele frequencies, were estimated for suprahaplotypes. Private haplotypes (p) were counted and corrected for sample size through rarefaction and given as percentage of R.

Results

The results for the three main species groups are given separately.

3.1. *Arabidopsis halleri*

CpDNA data

Among the 86 analysed accessions of *A.halleri* distributed in Europe and Japan eight suprahaplotypes comprising 42 haplotypes were detected (Fig. 2-1, Tab. 2-1, Tab. 2-2). The most ancestral suprahaplotype was A from Austria (Eastern Alps, non-glaciated East) and Slovakia (Western Carpathian Mountains). Two suprahaplotypes, B and D, were quite common. They were found in more than four regions. The most wide spread suprahaplotype D, occurred in Japan and in all European regions except the Western Alps. D was the only suprahaplotype which was found in Japan.

Nucleotide diversity was highest in the non-glaciated East (0.00215), followed by the SE-Carpathians (0.00142) and the Western Carpathian Mountains (0.001244). In the Western Alps and in Japan nucleotide diversity was zero since only one suprahaplotype was detected.

Nearly all suprahaplotypes were present in the non-glaciated East (seven suprahaplotypes, R = 3.69), except suprahaplotype X, which was private to the SE-Carpathians. The adjacent regions, the Eastern Alps and the Western Carpathian Mountains, showed also quite high numbers of different suprahaplotypes (Eastern Alps - five suprahaplotypes, R = 2.86; Western Carpathian Mountains – four suprahaplotypes, R = 2.77). The third diversity measurement, the effective diversity, was also highest in the non-glaciated East (3.69) followed by the Eastern Alps (2.42) and the Western Carpathian Mountains (2.38). Private suprahaplotypes were found in the non-glaciated East and the SE-Carpathians.

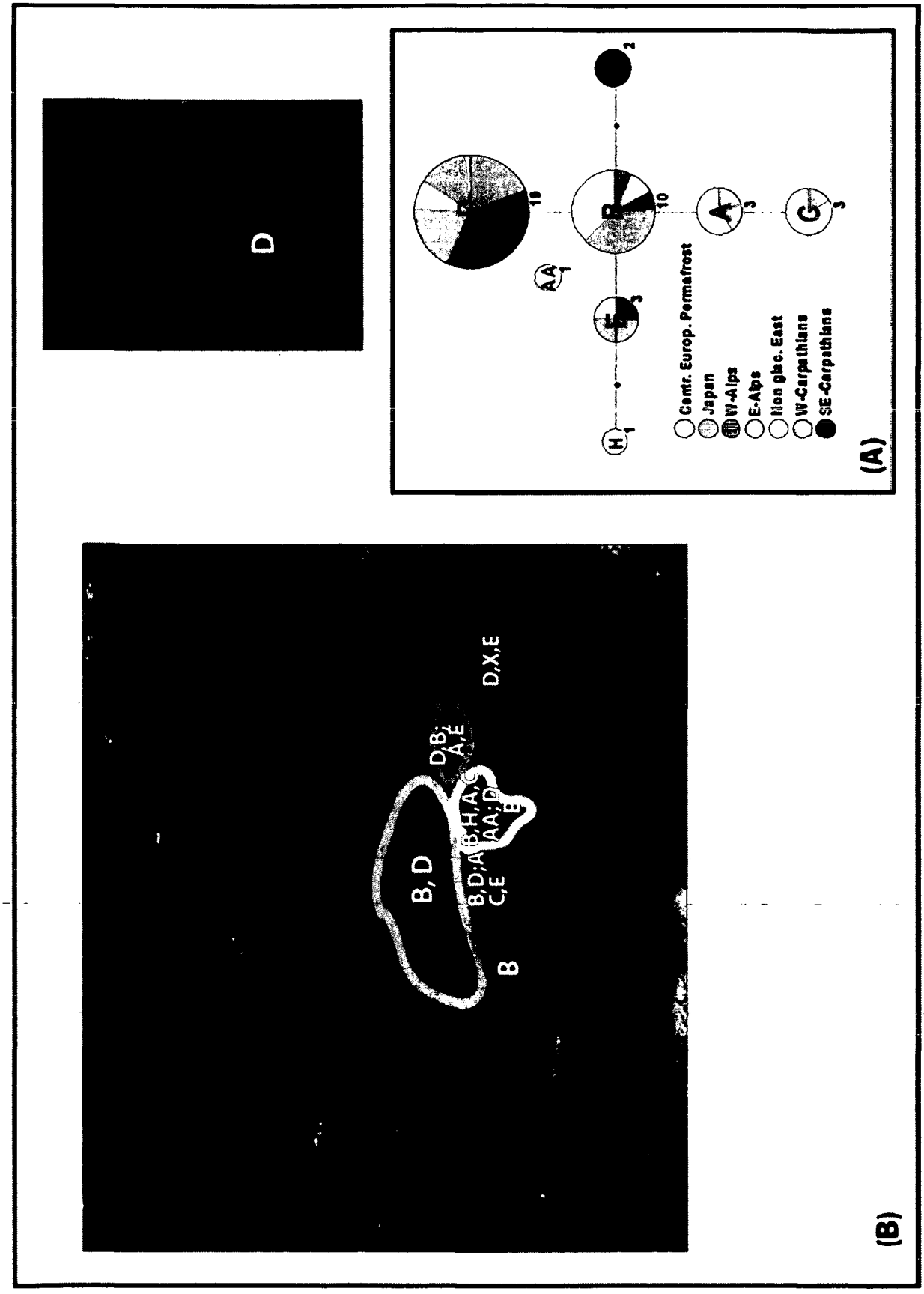

Figure 2-1: CpDNA-suprahaplotypes in *A. halleri agg.* (A) CpDNA-suprahaplotype network of *A. halleri agg.* redrawn from Fig. 1-5. Each circle represents a suprahaplotype. The size of the circle corresponds with the number of haplotypes (number below each circle), which are summarized to a suprahaplotype. (B) Geographic distribution of the cpDNA-suprahaplotypes among the several major regions

Table 2-1: Regional values of *A. halleri* based on cpDNA-suprahaplotypes - sample size (n), number of different haplotypes corrected for sample size 6 (R), effective diversity according to Gregorius (V_a), nucleotide diversity (π) +/- SD and private haplotypes (occur in only one region) in %.

						cpDNA-suprahaplotypes							
Region	n	R	V_a	π	p	A	B	C	D	E	H	X	AA
Central Europ. permafrost reg.	6	2	2	0.000800 +/- 0.000843			3		3				
W-Alps	2	1	1	0			2						
E-Alps	19	2.84	2.42	0.001092 +/- 0.000914		1	11	1	5	1			
non glac. East	24	3.69	3.69	0.002150 +/- 0.001470	31%	3	11	2	1	1	4		2
W-Carpathians	10	2.77	2.38	0.001244 +/- 0.001053		1	2		6	1			
SE-Carpathians	18	2.06	1.57	0.001420 +/- 0.001101	35%				14	1		3	
Japan	7	1	1	0					7				

Haplotype diversity estimates (Tab. 2-2) showed a different picture. The highest number of different haplotypes (12 haplotypes, R = 4.82) and the highest effective diversity (V_a = 8.6) were detected in the Eastern Alps. Quite similar high values were found in the SE-Carpathians (11 haplotypes, R = 4.75, V_a = 6.23) and the non-glaciated East (12 haplotypes, R = 4.62, V_a = 6.26). The Western Alps and Japan again showed the lowest diversity estimates. Private haplotypes were found in all regions. Interestingly 93% of the SE-Carpathian haplotypes were private, in the Western Carpathian Mountains and in Japan 100% of the haplotypes were private.

Table 2-2: Regional values of *A. halleri* based on cpDNA-haplotypes - sample size (n), number of different haplotypes corrected for sample size 6 (R) and effective diversity according to Gregorius (V_a). Private haplotypes (p), which occurr in only one region are given in %. Unique haplotypes, occurring in one individual in one region are given under u.

					cpDNA-haplotypes							
region	n	R	V_a	p	80	14	29	58	91	122	12	55
Centr. Europ. permafr. reg.	6	4	3	75%	3							
W-Alps	2	2	2	50%	1							
E-Alps	19	4.82	8.6	49%	5	1	1	1	1	1		
non glac. E	24	4.62	6.26	46%	8	1	3	1	2		2	
W-Carp.	10	4.33	4.55	100%								
SE-Carp.	18	4.75	6.23	93%						1		2
Japan	7	2.86	2.58	100%								

	cpDNA-haplotypes									
region	79	117	120	121	123	139	142	143	144	u
Centr. Europ. permafr. reg.										3
W-Alps										1
E-Alps	2				2					4
non glac. E						2				5
W-Carp.				3			3			4
SE-Carp.		6	2							7
Japan								3	3	1

Regional suprahaplotype-sharing (Tab. 2-3) was highest between the non-glaciated East and the Eastern Alps. Four suprahaplotypes were shared between the non-glaciated East and the Western Carpathian Mountains and between the Western Carpathian Mountains and the Eastern Alps, respectively. The Central European permafrost region had two suprahaplotypes in common with the Eastern Alps, the non-glaciated East and the Western Carpathian Mountains. Two shared suprahaplotypes were detected between the SE-Carpathians and the Western Carpathian Mountains, the non-glaciated East and the Eastern Alps. Most haplotypes were shared between the two neighbouring regions, the non-glaciated East and the (formerly glaciated) Eastern Alps. The remaining pairs of regions shared at most one haplotype. Interestingly the Western Carpathian Mountains had no haplotype in common with any other region and the SE-Carpathians shared only one haplotype with the Eastern Alps.

Table 2-3: Shared regional cpDNA-suprahaplotypes/haplotypes of *A. halleri*. Number of shared suprahaplotypes between all pairs of regions in the upper part. Number of shared haplotypes between all pairs of regions in the lower part. Suprahaplotypes respectively haplotypes in brakes.

Φ_{st}\F_{st}	Centr. Europ. permafr.reg.	W-Alps	E-Alps	non glac. East	W-Carp.	SE-Carp.	Japan
Centr. Europ. permafr. reg.		**1** (B)	**2** (B,D)	**2** (B,D)	**2** (B,D)	**1** (D)	**1** (D)
W-Alps	**1** (80)		**1** (B)	**1** (B)	**1** (B)	0	0
E-Alps	**1** (80)	**1** (80)		**5** (A,B,C,D,E)	**4** (A,B,D,E)	**2** (D,E)	**1** (D)
non glac. East	**1** (80)	**1** (80)	**5** (14,29,58,80,91)		**4** (A,B,D,E)	**2** (D,E)	**1** (D)
W-Carpathians	0	0	0	0		**2** (D,E)	**1** (D)
SE-Carpathians	0	0	**1** (122)	0	0		**1** (D)
Japan	0	0	0	0	0	0	

Pairwise Φ_{ST} estimates showed that the non-glaciated East was clearly differentiated from Japan, the SE-Carpathians and the Western Carpathian Mountains (Tab. 2-4). The Western Alps and Japan were totally different, because they had no suprahaplotypes in common. Japan was more or less significantly differentiated from all European regions except the SE-Carpathians. Pairwise F_{st} estimates showed a quite similar picture. The Western Alps were significantly differentiated from the SE-Carpathians, whereas the pairwise Φ_{ST} estimates showed a non-significant differentiation.

Table 2-4: Pairwise differentiation among regional groups of *A. halleri*: Φ_{ST} (lower part) and F_{ST} (upper part) are both estimated from cpDNA *trnL*-Intron + *trnL/F*-IGS sequence data. Values corresponding to significant differentiation at the 0.01 level in a permutation test (1000 permutations) are given in bold, values with p<0.03 are indicated with an x. Underlined values correspond to quite large Φ_{ST}/F_{st} values without significant differentiation.

$\Phi_{st}\backslash F_{st}$	Centr. Europ. permafr. reg.	W-Alps	E-Alps	non glac.E	W-Carp	SE-Carp	Japan
Centr. Europ. permafr. reg.		0.14286	-0.05516	0.07475	-0.03937	0.23138	0.43243
W-Alps	0.14286		-0.06308	-0.02878	0.36681	**0.66307**	1x
E-Alps	-0.0156	-0.19847		0.02695	0.11825	**0.36371**	**0.46109**
non glac. E	0.14612	-0.17731	0.07196		**0.18062**	**0.393**	**0.47137**
W-Carpathians	-0.1068	0.13223	0.05836	**0.18924**		0.0379	0.14019
SE-Carpathians	0.02951	0.27627	**0.23198**	**0.33474**	0.01908		0.03991
Japan	0.43243	1x	**0.44859**	**0.43498**	0.16667	0.03659	

ITS data

In *Arabidopsis halleri* agg. seven different ITS-supratypes comprising 36 different ITS-types were detected when analysing 94 accession from Europe and Japan (Fig. 2-1). In Japan, where only ssp. *gemmifera* occurred, most different ITS-supratypes (r, s, t) were found. These three ITS-supratypes did not occur in any other region respectively subspecies. ITS-supratype u was distributed all over Europe but did not occur in Japan. This led to the strongest differentiation of Japan from the European regions among the regional groups in pairwise Φ_{ST} and F_{ST} estimates (Tab. 2-6). The SE-Carpathian Mountains, the non-glaciated East and the Eastern Alps each harboured a second but always different ITS-supratype. Diversity estimates (Tab. 2-5) were highest for the SE-Carpathian Mountains (π = 0.002610, V_a = 1.49) and Japan (π = 0.0001434, V_a = 2.57), followed by the non-glaciated East (π = 0.000245, V_a = 1.18) and the Eastern Alps. No diversity was measured in the Central European permafrost region, in the Western Alps and in the Western Carpathian Mountains, since only common ITS-supratype u was present.

Table 2-5: Regional values of *A. halleri* based on ITS-supratypes - sample size (n), number of different haplotypes corrected for sample size 6 (R) and effective diversity according to Gregorius (V_a).

					ITS-supratypes						
Region	n	R	V_a	π	o	p	q	r	s	t	u
Centr. Europ. permafr. reg.	13	1	1	0							13
W-Alps	1	1	1	0							1
E-Alps	22	1.27	1.1	0.000140 +/- 0.000288			1				21
non glac. East	24	1.45	1.18	0.000245 +/- 0.000388		2					22
W-Carpath.	10	1	1	0							10
SE-Carpath.	18	1.91	1.49	0.002610 +/- 0.001789	5						13
Japan	6	3	2.57	0.001434 +/- 0.001309				2	3	1	

Table 2-6: Pairwise differentiation among regional groups of *A. halleri*: Φ_{ST} (lower part) and F_{ST} (upper part) are both estimated from ITS sequence data. Values corresponding to significant differentiation at the 0.01 level in a permutation test (1000 permutations) are given in bold, values with p<0.03 are indicated with an x. Underlined values correspond to quite large Φ_{ST} values without significant differentiation.

Φ_{st}\F_{st}	Centr. Europ. permafr. reg.	W-Alps	E-Alps	non glac. E	W-Carpath.	SE-Carpath.	Japan
Centr. Europ. permafr. reg.		0	-0.02614	0.00694	0	0.19588	**0.75597**
W-Alps	0		-1	-0.91394	0	-0.52941	0.26667
E-Alps	-0.02614	-1		-0.00235	-0.04195	**0.18477**	**0.74554**
non glac. East	0.00694	-0.91304	0.02681		-0.01033	0.14916x	**0.69284**
W-Carpath.	0	0	-0.04195	-0.01033		0.16667	**0.71576**
SE-Carpath.	0.19588	-0.52941	**0.24973**	**0.25325**	**0.16667**		**0.46995**
Japan	**0.83714**	0.49091	**0.8481**	**0.82676**	**0.80769**	**0.50939**	

Discussion

Arabidopsis halleri is a middle to southeast European distributed plant species of mainly mountainous to subalpine habitats (MEUSEL et al. 1965). Three subspecies are usually recognized, *A. h.* ssp. *halleri*, *A. h.* ssp. *ovirensis* and *A. h.* ssp *gemmifera* of which the later is the only one distributed to E-Asia and Japan (AL-SHEHBAZ & O'KANE 2002). Two further subspecies are described by KOLNIK & MARHOLD (2006): Firstly *A. h.* ssp. *dacica*, which summarizes the formerly *A. h.* ssp *ovirensis* from the Carpathian Mountains and the Balkan, since they are different from the *A. h.* ssp. *ovirensis* from locus classicus in Austria (Mount Obir) and secondly *A. h.* ssp. *tatrica*, an endemic of the Western Carpathian Mountains. From sequence data (see ITS-data in chapter 1) all subspecies, except *A. h.* ssp. *tatrica* were recognized as more or less distinct units.

Concerning the "older" diversity of the cpDNA-suprahaplotypes the centre of diversity of the species group *A. halleri* in Europe was actually found in the non-glaciated East. This region harboured the highest amount of suprahaplotypes and showed the highest nucleotide diversity and effective diversity. Going concentric into any direction from this centre, diversity estimates got reduced with still high values in the adjacent formerly glaciated Eastern Alps and Western Carpathian Mountains, the lowest value in the Western Alps.

"High" levels of DNA variation are often associated with the centre of origin of a species (KOCH et al. 2006)." It is likely that the species evolved in this region which stretches from the Eastern Alps to the Western Carpathian Mountains. In its centre (the non-glaciated East) all suprahaplotypes (except a private one of the SE-Carpathian Mountains - X) were present. The oldest suprahaplotypes of the *A. halleri* network A, B and C (Fig. 2-1) were shared at a high amount with the two other main species groups *A. arenosa* and *A. lyrata* (Fig. 2-5). A possible explanation for this is a common ancient genepool of the progenitors of the several species. From this genepool suprahaplotypes could have been distributed to the three main species

groups. This might be the reason why suprahaplotype A, which was typical for *A. arenosa* and suprahaplotype C, which was typical for *A. lyrata,* were still present in some individuals of *A. halleri.* Suprahaplotype B and E are more common in *A. halleri* and suprahaplotype D, which evolved from B was typical for *A. halleri.* It comprises most of the haplotypes found in *A. halleri.* An alternative explanation is hybridization, especially past hybridization, because from ITS-data no hint for ongoing hybridization was found.

Individuals which carried suprahaplotype D succeeded in colonizing E-Asia and Japan. Since suprahaplotype D was very common in Europe and the three haplotypes which were detected in Japan were different from the haplotypes, which were summarized in suprahaplotype D in Europe, the origin of the colonizers was somewhere in Europe. For two of the three haplotypes detected in E-Asia and Japan ITS-sequences were available and they interestingly also had corresponding different ITS-supratypes (143--> s; 144--> r). Additionally a third different ITS-supratype (t) was found in Japan and Russian Far East, when using data from O'KANE & AL-SHEHBAZ (2003). Comparing the ITS diversity estimates of Japan and Europe (Tab. 2-5) Japan harboured more different ITS-supratypes than any region in Europe. It seems as if the populations in Japan where the different ITS-supratypes evolved have been separated from each other with no contact. According to KAWABE & MIYASHITA (2003) "A recent bottleneck and/or small population size (with low migration) should be considered in the case of *A. halleri* ssp. *gemmifera.* "

After an initial period of range expansion, where the suprahaplotypes spread from their centre of origin into all directions, a period of separation occurred especially for the Western and the SE-Carpathian Mountains, which might have led to the evolution of new subspecies. New haplotypes evolved in both regions. The Western Carpathian Mountains did not share any haplotype with an adjacent region! This might support the idea of KOLNIK & MARHOLD (2006) that the plants occurring in the Western Carpathian Mountains are different and therefore best circumscribed as a different subspecies, *A. halleri* ssp. *tatrica.* In a recent AFLP analyses (Kolnik & Marhold unpub. data) *A. halleri* ssp. *tatrica* appears as a well-defined group with 95% bootstrap support. The SE-Carpathian Mountains only shared one haplotype with another region. Subspecies *dacica* was found in this region and additionally subspecies *halleri*, which also occurred outside the SE-Carpathian Mountains.

Comparing *A. halleri* with *A. arenosa* and *A. lyrata* the low number of different cpDNA-suprahaplotypes was striking. In *A. halleri* only 7.95 different suprahaplotypes were found (if the number of different haplotypes was corrected through rarefaction). In *A. arenosa* 10.24 different haplotypes and in *A. lyrata* 13 haplotypes were found. *Arabidopsis halleri* seems to have suffered most from glaciation. As a mesophytic plant mainly distributed at mountainous to subalpine habitats (MEUSEL et al. 1965) and with no tendency to occupy extreme climates (HOFFMANN 2005) climatic changes must have effected this species group more than for example *A. lyrata*, a cold- and draught-tolerant species with a great ecological flexibility (ANSELL 2004). It is likely that rare suprahaplotypes, which were formerly distributed to the Eastern Alps got extinct because of glaciation. Fig. 2-2 shows that the species did not expand into the formerly glaciated North of Europe and not far into the Central European permafrost region. A reduced migration ability due to loss of ecotypes and genetic variability during the Ice Ages is a possible explanation for this.

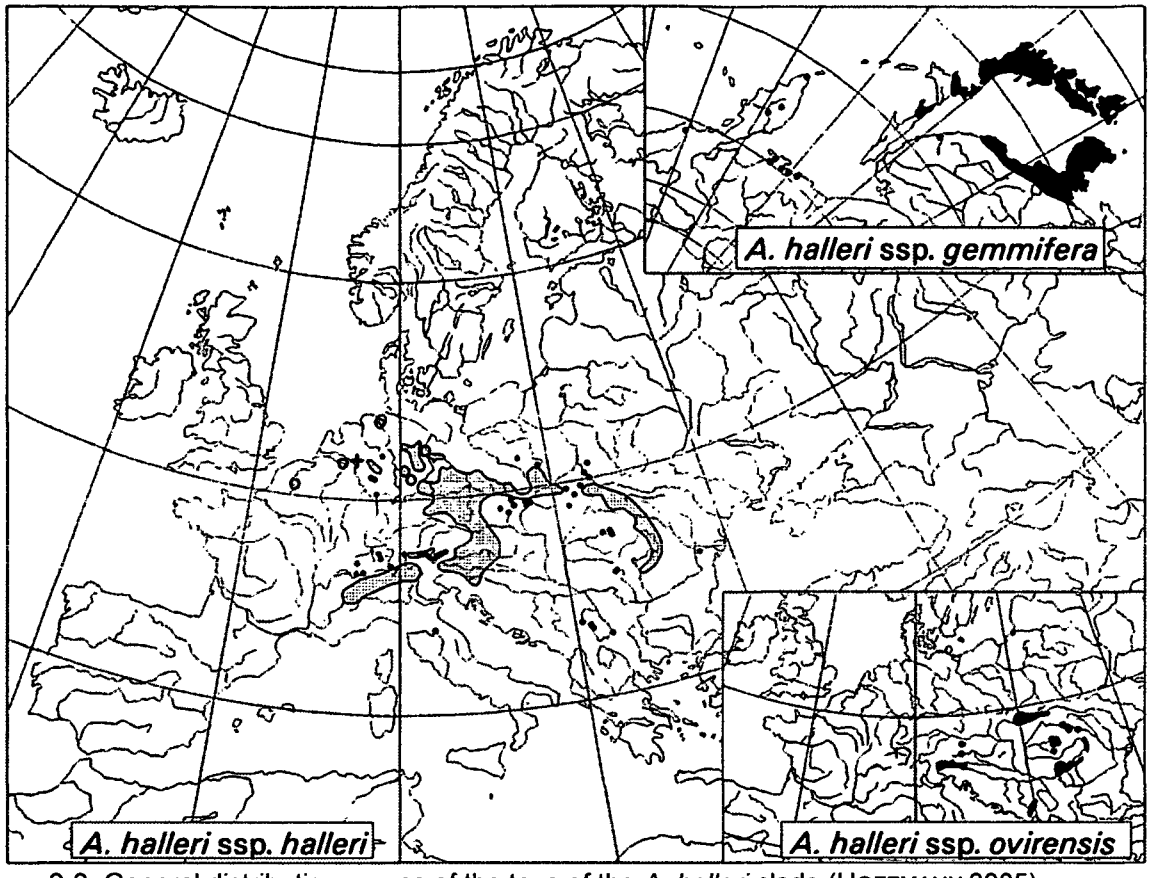

Figure 2-2: General distribution ranges of the taxa of the *A. halleri* clade (HOFFMANN 2005).

It can be assumed that *A. halleri* survived the last glaciation in the non-glaciated East, the Carpathian Mountains and maybe to some extend at the south edge of the Alps in small mountainous refuge areas which were not glaciated. Two accessions from Italy (Card0383, Card0304) with unique cpDNA-haplotypes and one of them even with an unique ITS-supratype (q) can be seen as evidence for a survival south of the Alps. Since the Alps act as a barrier Italian lineages are often isolated (TABERLET et al. 1998) and a strong colonization from this refuge into more northern areas seems not likely.

3.2. *Arabidopsis lyrata*

CpDNA-data

Thirteen different suprahaplotypes comprising 31 haplotypes were detected among the 70 accessions from the distribution range of *A. lyrata* in Europe and North America (Fig. 2-3, Fig. 2-4). The most ancestral suprahaplotype A was found in America, but also in the non-glaciated East. Two major lineages evolved from suprahaplotype A. Derived suprahaplotypes, the tips of the network, were only distributed to Europe. Three rare suprahaplotypes from the non-glaciated East (R, V, AF) were directly connected to the most ancestral suprahaplotype A. In America only the most ancestral (inner) suprahaplotypes A, B and C were found. These three suprahaplotypes were shared by the three main species groups. Suprahaplotype A was an *A. arenosa* typical suprahaplotype and suprahaplotype B was more often found in *A. arenosa* and *A. halleri* than in *A. lyrata*. Tab. 2-7 summarizes suprahaplotype frequencies and diversity indices of *A. lyrata*. The formerly glaciated North of Europe appeared to be the region with highest diversity estimates. Nucleotide diversity (π = 0.002812), effective diversity (V_a = 2.81) and number of different haplotypes corrected for sample size (R = 3.8) were higher than in any other region. The non-glaciated East showed also quite high values of nucleotide diversity and haplotype frequency (π = 0.001798; R = 3.07). Comparing suprahaplotype estimates and haplotype estimates (Tab. 2-7, Tab. 2-8) diversity indices were different. Both regions, the formerly glaciated North and the non-glaciated East showed the same number of different haplotypes corrected for sample size (R = 5.3). The highest effective diversity (V_a = 7.09) however was found in the non-glaciated East. The lowest effective diversity and haplotype frequency for suprahaplotype and haplotype estimates was observed in the Central European permafrost region.

Table 2-7: Regional cpDNA values of *A. lyrata* based on cpDNA-suprahaplotypes - sample size (n), number of different haplotypes corrected for sample size 7 (R), effective diversity according to Gregorius (V_a), nucleotide diversity (π) +/- SD and private haplotypes (occur in only one region) in %.

						cpDNA-suprahaplotypes											
Region	n	R	V_a	π	p	B	C	AG	A	G	J	K	R	S	V	AB	AC	AF
glac. North	11	3.8	2.81	0.002812 +/- 0.001912	50%		2	6		1				1		1		
America	15	2.86	2.71	0.001168 +/- 0.000972	35%	7	3		5									
Centr. Europ. permafr. reg.	7	2	1.32	0.001143 +/- 0.001042			6	1										
non glac. E	37	3.07	2.10	0.001798 +/- 0.001266	55%	1	25		1		1	1	1		1		4	2

Table 2-8: Regional values of *A. lyrata* based on cpDNA-haplotypes - sample size (n), number of different haplotypes corrected for sample size 6 (R) and effective diversity according to Gregorius (V_a). Private haplotypes (p), which are occurring in only one region are given in %. Unique haplotypes, occurring in one individual in one region are given under u.

					cpDNA-haplotypes										
Region	n	R	V_a	p	13	16	29	1	22	30	31	84	87	88	u
glac. North	11	5.32	5.76	83%			2						3	2	4
America	15	4	3.46	78%		3					2	7			3
Centr. Europ. permafr. reg.	7	4	3.27	50%	1		3			2					1
non glac. E	37	5.33	7.09	70%	3	1	12	4	3						14

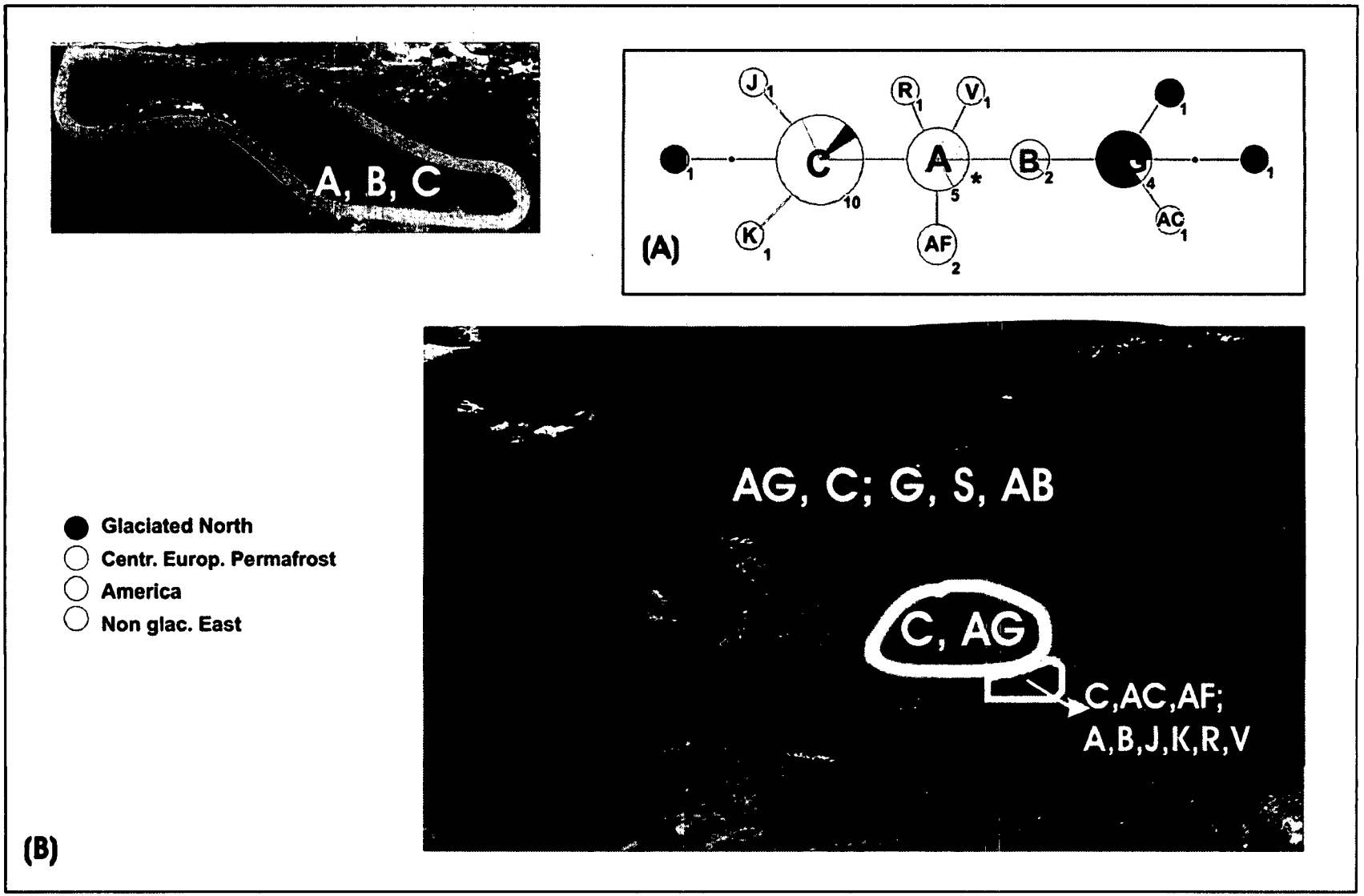

Figure 2-3: CpDNA-suprahaplotypes in *A. lyrata agg.* (A) CpDNA-suprahaplotype network for *A. lyrata agg.* redrawn from Fig. 1-5. Each circle represents a suprahaplotype. The size of the circle corresponds with the number of haplotypes (below each circle), which are summarized to the suprahaplotype. (B) Geographic distribution of the cpDNA-suprahaplotypes among the several major regions.

Figure 2-4: Left: Distribution of cpDNA-suprahaplotypes (A,B,C) resp. haplotypes of *A. lyrata* occurring in North America. Right: Distribution of the ice-cover (white) and tundra (dark grey) in the Northern Hemisphere at the last glacial maximum (from ABBOTT & BROCHMANN 2003 after FRENZEL 1968; FRENZEL et al. 1992)

Regional suprahaplotype sharing (Tab. 2-9) was low in *A. lyrata*. A maximum of three suprahaplotypes was shared among the four different regions. This was the case for the non-glaciated East and America, although these two regions were most distant from each other. Regional haplotype sharing was even lower. The two neighbouring regions, the Central European permafrost region and the non-glaciated East shared two haplotypes, as well as two suprahaplotypes. America had no haplotypes in common with the formerly glaciated North and the Central European permafrost region.

Table 2-9: Shared regional cpDNA-suprahaplotypes/haplotypes of *A. lyrata*. Number of shared suprahaplotypes between all pairs of regions in the upper part. Number of shared haplotypes between all pairs of regions in the lower part. Suprahaplotypes/haplotypes in brakes.

	glac. N	America	Centr. Europ. permafr. reg.	non glac. E
glac.North		**1** (C)	**2** (C,AG)	**1** (C)
America	0		**1** (C)	**3** (A,B,C)
Centr. Europ. permafr. reg.	**1** (29)	0		**1** (C)
non glac. E	**1** (29)	**1** (16)	**2** (13, 29)	

Pairwise Φ_{ST} and F_{ST} estimates showed a clear differentiation for all regions except the non-glaciated East and the Central European permafrost region (Tab. 2-10).

Table 2-10: Pairwise differentiation among regional groups of *A. lyrata*: Φ_{ST} (lower part) and F_{ST} (upper part) are both estimated from cpDNA *trnL*-Intron + *trnL/F*-IGS sequence data. Values corresponding to significant differentiation at the 0.01 level in a permutation test (1000 permutations) are given in bold, values with p<0.03 are indicated with an x.

Φ_{st}\F_{st}	glaciated N	America	Centr. Europ. permafr. reg.	non glac. E
glaciated N		**0.28233**	0.31846x	**0.30988**
America	**0.2593**		**0.3726**	**0.29221**
Centr. Europ. permafr. reg.	0.319x	**0.34863**		-0.01319
non glac. E	**0.33939**	**0.22092**	-0.06141	

ITS data

In *Arabidopsis lyrata* agg. five different ITS-supratypes comprising 19 different ITS-types were detected in Europe and America. Additional five ITS-types, which belong to interspecific hybrids between *A. lyrata* and *A. arenosa*, were assigned to *A. lyrata* agg. by the program TCS. But they were excluded from the following estimations. As for cpDNA data, the number of different ITS-type corrected for sample size (R) and the effective diversity (v_a) were highest in the formerly glaciated North (Tab. 2-11). Regarding ITS nucleotide diversity the highest estimates were found in America (π = 0.002458), but followed by the formerly glaciated North (π = 0.002458). The non-glaciated East however showed the lowest ITS diversity estimates, whereas this was the case for the Central European permafrost region in the cpDNA data.

Table 2-11: Regional values of *A. lyrata* based on ITS-supratypes - sample size (n), number of different ITS types corrected for sample size 6 (R) and effective diversity according to Gregorius (V_a).

					ITS-supratypes				
	n	R	V_a	π	a	b	c	d	e
glac. North	11	2.53	2.28	0.001061 +/- 0.000979	4	6	1		
America	6	2	1.8	0.002458 +/- 0.001943		4			2
Centr. Europ. permafr. reg.	11	1.98	1.86	0.000782 +/- 0.000805	7	4			
non glac. East	34	1.33	1.05	0.000175 +/- 0.000320	32			2	

Pairwise differentiation (Tab. 2-12) was stronger for America and the non-glaciated East (Φ_{ST} = 0.79605; F_{ST} = 0.080659), when estimated from ITS-data than from cpDNA data (Tab. 2-10: Φ_{ST} = 0.22092, F_{ST} = 0.29221). The non-glaciated East was additionally significantly differentiated from all other regional groups.

Table 2-12: Pairwise differentiation among regional groups of *A. lyrata*: Φ_{ST} (lower part) and F_{ST} (upper part) are both estimated from ITS-supratype sequence data. Values corresponding to significant differentiation at the 0.01 level in a permutation test (1000 permutations) are given in bold, values with p<0.03 are indicated with an x. Underlined values correspond to quite large Φ_{ST} values without significant differentiation.

Φ_{st}\F_{st}	glac. North	America	Centr. Europ. permafr. reg.	non glac.E
glac. North		0.08836	0.01159	**0.57068**
America	0.25812x		0.31456	**0.80659**
Centr. Europ. permafr. reg.	0.04474	**0.41765**		**0.33104**
non glac.E	**0.59476**	**0.79605**	**0.33104**	

Discussion

The oldest and most inner suprahaplotypes of the *A. lyrata* cpDNA-network (Fig. 2-3A) A, B and C, which were also present in the two other main species groups *A. halleri* and *A. arenosa* (Fig. 1-5) originated from a common ancient genpool, which might have been located in Middle to Eastern Europe, since most of the variation of the three main species groups was found there. In contrast to *A. arenosa* and *A. halleri*, *A. lyrata* colonized Northern regions most successfully and is today circumpolar distributed (Fig. 2-6). This might have happened in former times (not

after the last glaciation). HULTEN (1937) in ABBOTT & BROCHMANN (2003) proposed that many arctic plants obtained a circumarctic distribution early in the Quaternary period.

If we consider that ancestors of the closest relatives of *A. thaliana* diverged from the *A. thaliana* ancestor approximately five million years ago (KOCH et al. 2000, 2001) and count the mean number of mutational steps in the ITS network (Fig. 1-3) from *A. thaliana* to any other tip of the network (30 steps), we obtain a rough estimate for the age of the inner part of the network of approximately 2 my, which is close to the beginning of the Pleistocene (Quaternary) and its various glaciation and deglaciation cycles (Koch & Matschinger 2007).

The colonization of America from Eurasia could have taken place in at least two separate colonization events. This assumption is based on the fact that the oldest ITS-supratype of the ITS-network (e) was only found in America (except in one individual of hybrid origin between *A. l.* ssp. *petraea* and *A. arenosa* in Lower Austria). Additionally a second, younger ITS-supratype (b) which is separated from ITS-supratype e via 3 mutational steps, was found in America. But in contrast to ITS-supratype e, ITS-supratype b was widely distributed in the species group *A. lyrata* around the world. It is possible that the first colonization (ITS-supratype e) was carried out from Europe directly to America. There is no strong support for this idea. According to Fig. 2-5, ITS-supratype e has its distribution centre in the centre and more to the north-east side of Canada. It does not occur in Asia and in the dataset of WARWICK et al. (2006) a new ITS-supratype was detected which evolved from ITS-supratype e via two mutational steps. This type was located in North Carolina!

Figure 2-5: Distribution of the ITS-supratypes a, b, c, d and e occurring in *A. lyrata* and *A. arenicola*. Additionally to the data of this study, data from WARWICK et al. (2006) and O'KANE & AL-SHEHBAZ (2003) were checked and marked in.

For the second colonization (ITS-supratype b) it is more likely that it took place from Europe via Asia and the Bering street to America. A connected broad distribution range from Eurasia to central North America (Fig. 2-6) supports the second eastward colonization route as well as the fact that ITS-supratype b occurs in the Far East of Russia and at the west side of America but not at the east side.

Taking cpDNA data in account the oldest and most inner suprahaplotypes of the suprahaplotype network A, B and C were found in America; in Europe C was

common but A and B only occurred in the non-glaciated East. Although considering the faster evolving haplotypes, only one out of six haplotypes found in North America was really identical to a haplotype from the non-glaciated East. This identical haplotype (16 belonging to suprahaplotype C) was only distributed to the NW-side of North America (Fig. 2-4). The remaining five haplotypes were private (unique) for North America (Tab. 2-8). That supports the idea, that the colonization of North America by *A. lyrata* from Eurasian progenitors did not take place recently after the last glaciation maximum. It is more likely that it occurred in earlier times since most of the haplotypes were not found anywhere else and had therefore time to evolve independently.

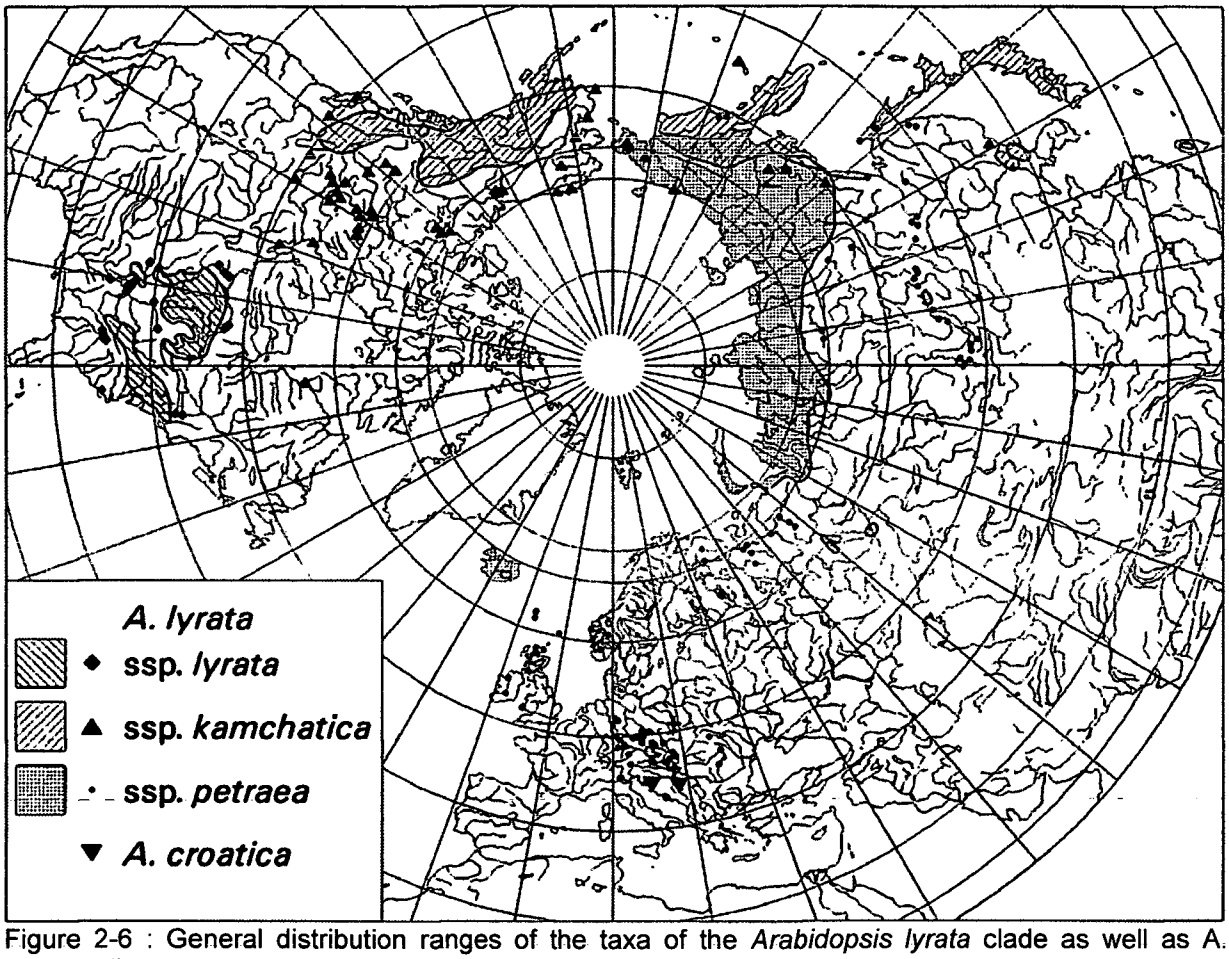

Figure 2-6 : General distribution ranges of the taxa of the *Arabidopsis lyrata* clade as well as A. *croatica* (HOFFMANN 2005).

In the Central European permafrost region and the formerly glaciated North concerning the ancient suprahaplotypes only suprahaplotype C was found. Suprahaplotypes A and B might have not been detected (due to low sample size in the Central European permafrost region) or might have become extinct. From these ancient suprahaplotypes A, B and C new suprahaplotypes evolved which are today mostly found in single regional groups (Fig. 2-3B, Tab. 2-7). Suprahaplotype and haplotype sharing was very low between regional groups and lowest in regard to *A. arenosa* and *A. halleri*. All regional groups (except the non-glaciated East in regard to the Central European permafrost region) were clearly differentiated from each other (with values around 0.3).

The various glaciation and deglaciation cycles might have effected the evolution of the genus as a whole and of the species group *A. lyrata* in particular. In North America only the ancient suprahaplotypes were found and no younger ones as it was

the case in Europe. There might have been new suprahaplotypes in North America which evolved at local sites and could not expand their range (or they were not detected). They might have become extinct and only the common types A, B and C could have survived in refuge areas in America. One possible refuge area might have been located in Bering, which is supposed to be a major northern refuge area for arctic plants throughout the Quaternary (HULTEN 1937 in ABBOTT & BROCHMANN 2003). All three detected American suprahaplotypes were found in this region (Fig. 2-4). From this refuge area suprahaplotype B moved further south-east along mountain ranges. This is supported by the fact that all accessions belonging to suprahaplotype B in America consist only of one haplotype (84). It is unlikely that suprahaplotype A colonized central Canada and the area north of the Great Lakes from Beringia, since it was not detected in the area between. Further did all accessions with suprahaplotype A carry different haplotypes. It is more likely that they survived south of the ice sheet maybe in different refuge areas. The ITS-supratypes might have also survived in different refuge areas. If not, concerted evolution would have acted and they would not be so clearly separated geographically.

Several studies (WRIGHT et al. 2003; RAMOS-ONSIS et al. 2004) point out the low variation of *A. lyrata* ssp. *lyrata* in regard to *A. lyrata* ssp. *petraea* and suggest a bottleneck for *A. lyrata* ssp. *lyrata*. WRIGHT et al. (2003) give three hypothesis for a possible scenario. The first one is very likely: "Polymorphism in the North American populations might have been reduced by a strong population bottleneck. This might have been a recent event associated with glaciation, and postglacial re-colonisation, or it could have occurred during an initial colonization of North America from Europe, followed by a long-term reduction in effective population size." The second one, "population admixture between populations from distinct glacial refugia" could be possible for the Northern parts of Europe.

In Europe diversity was highest in the formerly glaciated North. This region is large and split into distinct units, where *A. lyrata* ssp. *petraea* today occurs. Only suprahaplotype AG was widespread in the formerly glaciated North. It occurred in Iceland, Norway and Sweden. Considerring the work of ANSELL (2004) sequence comparison makes it possibel to find out which of his *trnL/F*-IGS haplotypes correspond to the suprahaplotypes of the current study (Tab. 2-13).

Table 2-13: Corresponding haplotypes of the study of ANSELL (2004) and the current study.

haplotypes of Ansell (2004)	corresponding +/- to suprahaplotypes of the current study
2	C
6	AG (or S but unlikely)
7	AG

Following Fig 7.1 in ANSELL (2004), suprahaplotype AG is additionally wide distributed to Scotland and it occurs in the single population investigated from Whales. His finding of haplotype 6 in Sweden corresponds with the finding of suprahaplotype AG of the current study in Sweden. Beside its broad northern distribution haplotype 6/suprahaplotype AG occurs in an single individual which Ansell found in Austria (Steinhof, Lower Austria near Bernstein).

In the data set of the current study the following suprahaplotypes only occurred in parts of the formerly glaciated North. C only occurred in Iceland, S in Norway, AB in Scotland and G on the Faroer Islands. Following ANSELL (2004, Fig. 7.1) again C also occurs in the single investigated populations from Sweden and Wales and in one population in Scotland. So suprahaplotype C is more widespread than AG. The three rare suprahaplotypes S, AB and G from different parts of the formerly glaciated North

give evidence for a periglacial survival of the *A. lyrata* ssp. *petraea* variation maybe in different refuge areas.

CLAUSS & MITCHELL-OLDS (2006) suppose a persistence of *A. l.* ssp. *petraea* in Central Europe and maintenance of large, genetically diverse populations during recent glacial maxima from their microsatellite data. This surviving Central European gene pool could have expanded to the formerly glaciated North via leading edge expansion and/or admixture with migrants from other glacial refugia. But for them this requires further sampling and fossil evidence.

Sequence data from the chloroplast (the current study) draw a different picture of Central Europe. This region was situated between the Northern ice sheet and the ice sheet of the Alps. It was not glaciated but showed tundral landscape with permafrost soil. Diversity indices of chloroplast data of this region are the lowest among the other regional groups (only two different suprahaplotypes, C and AG were found). This is consistent with the findings of ANSELL (2004). He detected in a sampling of 250 individuals from 25 different populations of the Frankenalp, Bavaria only one haplotype (haplotype 2 which corresponds to suprahaplotype C from this study). Nevertheless it is likely that the species survived the last glaciation in the Frankenalp.

The high diversity estimates for the formerly glaciated North might result only at a small amount from recolonization of a Central European refuge area as it is proposed in CLAUSS & MITCHELL-OLDS (2006). The two regions only have one haplotype in common (Tab. 2-7). It is more likely that several refuge areas nearby the southern edge of the ice sheet are source for the high amount of variation found today in the formerly glaciated North. ABBOTT & BROCHMANN (2003) speculate of a distinct north-eastern glacial refuge area based on the data of two studies (JONSELL et al. 1995; VAN TREUREN et al. 1997), who have analysed a population from Russia, which was very diverged from the Scandinavian populations.

Parallel to this northern periglacial survival there was also survival near the Central European (Alps) ice sheet. At the eastern edge of the Alps *A. lyrata* ssp. *petraea* could survive with large population sizes in a region in Austria which had never been glaciated (non-glaciated East). The region stretches from the "Wachau" in Lower Austria south-eastwards to Styria (where it is only very rare these days). Diversity estimates and suprahaplotype/haplotype frequencies were quite high there. For further and more detailed information see chapter 3.

3.3. Arabidopsis arenosa

Among the 189 accessions of *A. arenosa* distributed in Europe 17 suprahaplotypes comprising 72 haplotypes were detected. Fig. 2-7A shows the relationship of the cpDNA-suprahaplotypes in form of a network. The most ancestral suprahaplotypes were the rare suprahaplotype U from Slovenia and the most common suprahaplotype A, which was distributed all over Europe. The also very frequent suprahaplotype B was only restricted to Austria and Slovenia.

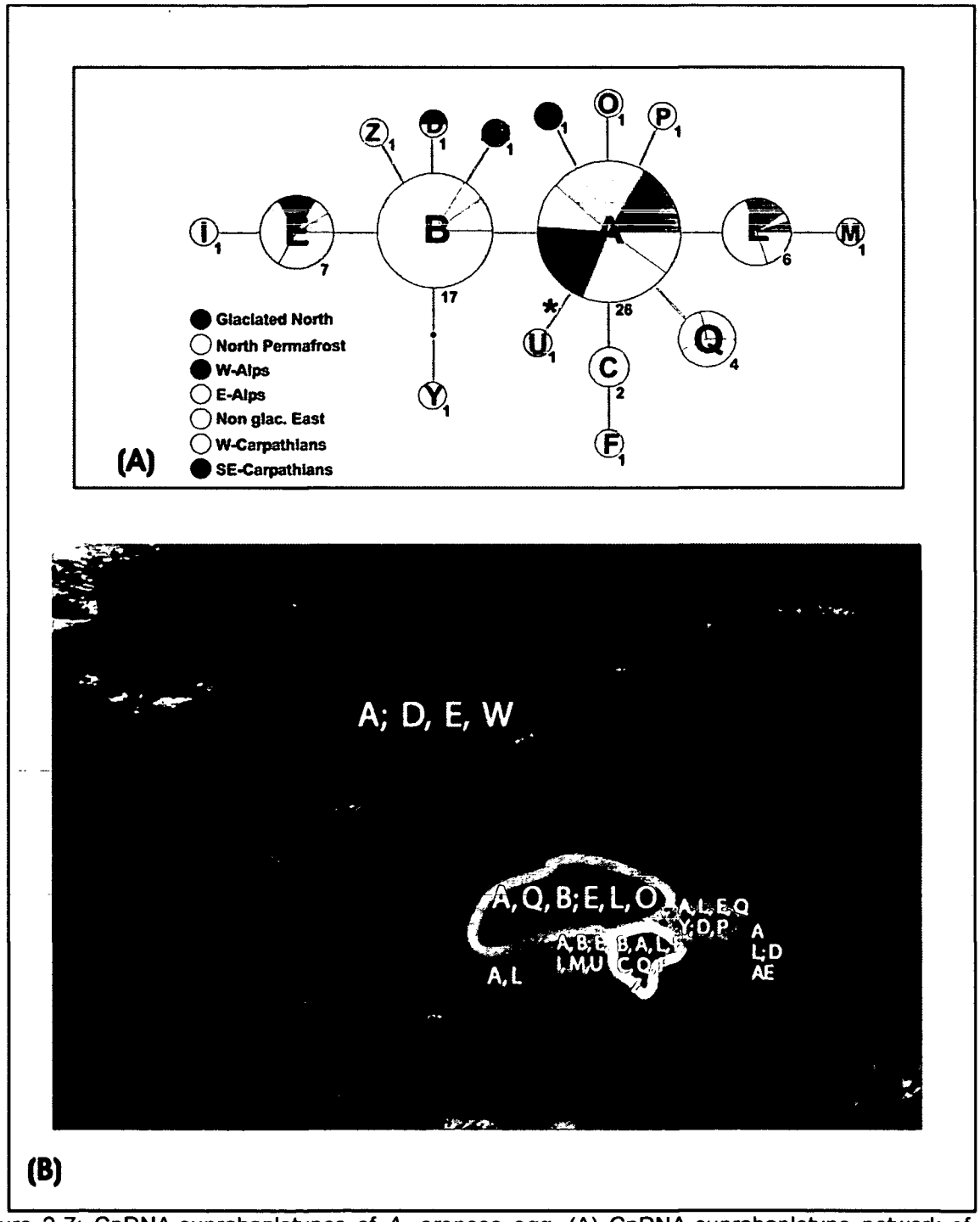

Figure 2-7: CpDNA-suprahaplotypes of *A. arenosa agg.* (A) CpDNA-suprahaplotype network of *A. arenosa agg.* redrawn from Fig. 2-5. Each circle represents a suprahaplotype. The size of the circle corresponds with the number of haplotypes (below each circle), which are summarized to the suprahaplotype. (B) Geographic distribution of the cpDNA-suprahaplotypes among the several major regions.

Nucleotide diversity (Tab. 2-14) was highest in the Western Carpathian Mountains (π = 0.001609) closely followed by the formerly glaciated North (π = 0.001571) and in the non-glaciated East (π = 0.001418). Suprahaplotype frequency and amount of private suprahaplotypes were highest in the formerly glaciated North (R = 4; p = 50%). Similar high values were found in the non-glaciated East (R = 3.82), the Central European permafrost region (R = 3.77) and the Western Carpathian Mountains (R = 3.57). The Western Alps showed the lowest diversity estimates and no private suprahaplotypes were found.

Table 2-14: Regional cpDNA values of *A. arenosa* based on cpDNA-suprahaplotypes - sample size (n), number of different haplotypes corrected for sample size 8 (R), effective diversity according to Gregorius (V_a), nucleotide diversity (π) +/- SD and private haplotypes (occur in only one region) in %.

						cpDNA-suprahaplotypes															
Region	n	R	Va	π	priv	A	B	D	E	L	Q	C	F	I	M	O	P	U	W	Y	Z	AE
glac. North	8	4	2.29	0.001571 +/- 0.001278	50%	5		1	1										1			
Centr. Europ. permafr. reg.	19	3.77	2.64	0.001170 +/- 0.000958	11%	11	2		1	1	3					1						
W-Alps	18	1.44	1.12	0.000148 +/- 0.000281		17				1												
E-Alps	30	2.69	1.66	0.000910 +/- 0.000794	30%	23	3		1					1	1			1				
non glac. E	50	3.82	3.06	0.001418 +/- 0.001059	19%	10	26		3	4	2	3	1								1	
W-Carpath.	42	3.57	2.89	0.001609 +/- 0.001164	15%	22		1	4	10	2						1			2		
SE-Carpath.	22	2.59	1.77	0.000889 +/- 0.000791	14%	16			1	4												1

Considering all haplotypes (Fig. 2-15), the Western Carpathian Mountains harboured the highest amount of different haplotypes, followed by the non-glaciated East and the Central European permafrost region. The effective diversity was extremely high in the non-glaciated East (V_a = 21.93) and a high amount of private haplotypes was found in the non-glaciated East. (73%) The formerly glaciated North showed low diversity estimates when considering haplotypes estimates. Only a high amount of private haplotypes (75%) was found there.

Suprahaplotype and haplotype sharing (Fig. 2-16) was extensive in *A. arenosa*. Up to six haplotypes and five suprahaplotypes respectively were shared among the several regional groups. The non-glaciated East and the Western Carpathian Mountains were in good contact with their neighbouring regions. In the Western Alps suprahaplotype sharing was low, whereas haplotype sharing was quite high. In the formerly glaciated North haplotype sharing was low, whereas suprahaplotype sharing was clearly higher.

Table 2-15: Regional values of *A. arenosa* based on cpDNA-haplotypes - sample size (n), number of different haplotypes corrected for sample size 8 (R) and effective diversity according to Gregorius (V_a). Private haplotypes (p), which are occurring in only one region are given in %. Unique haplotypes, occurring in one individual in one region are given under u.

					cpDNA-haplotypes									
Region	n	R	V_a	p	8	17	18	23	33	59	65	76	77	96
glac. North	8	4	2.29	75%	5									
Centr. Europ. permafr. reg.	19	6.53	8.4	36%	5		1	1	2		1	1		1
W-Alps	18	4.19	2.47	21%	10	1	2		2					
E-Alps	30	5.73	6.82	46%	7	8	2		2				1	
non glac. E	50	6.14	21.93	73%	3	1	1	1	1			1	5	5
W-Carp.	42	6.8	14	41%	2	4	3	1	6	1	2			
SE-Carp.	22	6.21	8.34	51%			1	2	2	1				

	cpDNA-haplotypes															
Region	107	109	110	10	15	20	25	29	61	70	72	83	91	92	95	u
glac. North																3
Permafrost	1									2						4
W-Alps	1															2
E-Alps																10
non glac. E		3		2				2	2			2	2	2		17
W-Carp.		5	3		2						2				2	9
SE-Carp.		3	1			5	3									4

Table 2-16: Shared regional cpDNA-suprahaplotypes/haplotypes of *A. arenosa*. Number of shared suprahaplotypes between all pairs of regions in the upper part. Number of shared haplotypes between all pairs of regions in the lower part. Suprahaplotypes respectively haplotypes in brakes.

	glac.N	Centr. Europ. permafr. reg.	W-Alps	E-Alps	non glac.E	W-Carp	SE-Carp
glac.N		**2** (A,E)	**1** (A)	**2** (A,E)	**2** (A,E)	**3** (A,E,D)	**2** (A,E)
Centr. Europ. permafr. reg.	**1** (8)		**1** (A)	**4** (A,B,E,L)	**5** (A,B,E,L,Q)	**4** (A,E,L,Q)	**3** (A,E,L)
W-Alps	**1** (8)	**4** (8,18,33,107)		**1** (A)	**1** (A)	**1** (A)	**1** (A)
E-Alps	**1** (8)	**3** (8,18,33)	**4** (8,17,18,33)		**4** (A,B,E,L)	**3** (A,E,L)	**3** (A,E,L)
non glac. E	**1** (8)	**6** (8,18,23,33,76,96)	**4** (8,17,18,33)	**5** (8,17, 18, 33, 77)		**4** (A,E,L,Q)	**3** (A,E,L)
W-Carp.	**1** (8)	**4** (8,18,23,33)	**4** (8,17,18,33)	**4** (8,17,18,33)	**6** (8,17,18,23,33,109)		**3** (A,E,L)
SE-Carp.	0	**3** (18,23,33)	**2** (18,33)	**2** (18,33)	**4** (18,23,33,109)	**6** (18,23,33,59,109,110)	

Among the regional groups of *A. arenosa* the non-glaciated East was differentiated strongest from the Western Alps (Tab. 2-17) and additionally clearly differentiated

from the other regional groups (except glaciated North for Φ_{ST}). Among the remaining regions there was differentiation between Western Alps and the glaciated North for both Φ_{ST} and F_{ST} estimates. Concerning only F_{ST} estimates the Western Carpathian Mountains were differentiated from the Western Alps and the Eastern Alps.

Table 2-17: Pairwise differentiation among regional groups of *A. arenosa*: Φ_{ST} (lower part) and F_{ST} (upper part) are both estimated from cpDNA *trnL*-Intron + *trnL/F*-IGS sequence data. Values corresponding to significant differentiation at the 0.01 level in a permutation test (1000 permutations) are given in bold, values with p<0.03 are indicated with an x.

Φ_{st}\F_{st}	glaciated N	Centr. Europ. permafr. reg.	W-Alps	E-Alps	non glac. E	W-Carp.	SE-Carp.
glaciated N		-0.02807	0.16334	0.00085	**0.22683**	-0.00026	-0.00234
Centr. Europ. permafr. reg.	-0.01047		**0.14529**	0.02516	**0.17514**	0.01319	0.02315
W-Alps	0.14961x	0.05268		0.03597	**0.4242**	**0.15434**	0.05769
E-Alps	-0.02049	-0.00115	0.03102		**0.29217**	**0.08629**	0.01863
non glac. E	0.081	**0.15316**	**0.2948**	**0.15738**		**0.21832**	**0.29895**
W-Carpathians	0.00683	0.01303	0.04627	0.02535	**0.15192**		-0.01111
SE-Carpathians	0.03671	0.01741	0.01698	0.01026	**0.20908**	-0.001393	

ITS data

Among the 180 accessions of the *Arabidopsis arenosa* agg. eight different ITS-supratypes were detected comprising 35 different ITS types. One accession, a hybrid with *A. lyrata*, with an additional ITS-supratype (f) was excluded from the following estimations.

Nucleotide diversity (Tab. 2-18) was highest in the Central European permafrost region, followed by the SE-Carpathians and the Western Carpathian Mountains. Most different ITS-supratypes were found in the SE-Carpathians (R = 4.45), in the Western Alps (R = 4.39) and the Central European permafrost region (R = 4.27). ITS-diversity estimates were low for the non-glaciated East and lowest for the formerly glaciated North and the Eastern Alps. In the later two regions only the common ITS-supratype g was found.

Table 2-18: Regional values of *A. arenosa* based on ITS-supratypes - sample size (n), number of different haplotypes corrected for sample size 12 (R) and effective diversity according to Gregorius (V_a).

					ITS-supratypes							
	n	R	V_a	π	g	h	i	j	k	l	m	n
glaciated North	12	1	1	0	12							
Centr. Europ. permafr. reg.	26	4.27	2.84	0.001881 +/- 0.001376	14	2		3	2		5	
W-Alps	20	4.39	3.08	0.001447 +/- 0.001156	10			1	3	4	2	
E-Alps	28	1	1	0	28							
non glac. East	48	2.97	1.63	0.001322 +/- 0.001054	37			2	1		6	2
W-Carpathians	30	3.34	1.78	0.001730 +/- 0.001289	22	4	1	1				2
SE-Carpathians	16	4.45	2.98	0.001779 +/- 0.001356	2	4		1	1	8		

The SE-Carpathian Mountains were for both Φ_{ST} and F_{ST} (Tab. 2-19) significantly differentiated from all other regional groups and strongest from the Eastern Alps and the formerly non-glaciated East. Significant, but low differentiation was found between the Western Carpathian Mountains and the Central European permafrost region, the Western Alps and the Eastern Alps. The Eastern Alps were significantly differentiated from all other regions except the formerly glaciated North.

Table 2-19: Pairwise differentiation among regional groups of *A. arenosa*: Φ_{ST} (lower part) and F_{ST} (upper part) are both estimated from ITS-supratype sequence data. Values corresponding to significant differentiation at the 0.01 level in a permutation test (1000 permutations) are given in bold, values with p<0.03 are indicated with an x.

$\Phi_{st}\backslash F_{st}$	glac. N	Centr. Europ. permafr. reg.	W-Alps	E-Alps	non glac. E	W-Carpath.	SE-Carpath.
glaciated N		**0.18903**	**0.22747**	0	0.05714	0.08019	**0.55669**
Centr. Europ. permafr. reg.	**0.1947**		0.00327	**0.27932**	0.04435	0.04841	**0.23483**
W-Alps	**0.21428**	0.00656		**0.33812**	**0.09174**	0.08593x	**0.14005**
E-Alps	0	**0.28576**	**0.3232**		**0.10354**	**0.14471**	**0.67713**
non glac. E	0.02976	0.06008x	0.05326	0.07176x		0.01707	**0.43005**
W-Carpath.	0.04808	**0.08534**	**0.10034**	**0.10543**	0.04323x		**0.35418**
SE-Carpath.	**0.47134**	**0.17245**	**0.14954**	**0**	**0.32535**	**0.22653**	

Discussion

Among the three main species groups of the genus *Arabidopsis*, *A. arenosa* is the less investigated species group. Due to AL-SHEHBAZ & O'KANE (2002) it harbours two well defined subspecies, *A. a.* ssp. *arenosa* and *A. a.* ssp. *borbasii* (Zapalovicz) O'Kane & Al-Shehbaz. Beside that there are several taxa in East Europe, which were not legitimately published by MESICEK (1970). This splitting into several taxa is due to high karyological and morphological diversity (CLAUSS & KOCH 2006).

The highest number of different suprahaplotypes was found within the non-glaciated East. This might be due to hybridization and reticulation with *A. lyrata*. The suprahaplotypes F and Z were found each in one individual and showed an "hybrid" ITS. Their was also evidence for introgression in suprahaplotype C, which was typically occurring in *A. lyrata*. A hybrid zone was detected in this regional group at the edge of the Eastern Alps, which is outside the range of the last maximum glaciation (Fig. 3-9).

The centre of diversity was actually found in the Western Carpathian Mountains, based on the highest value of nucleotide diversity, a quite high number of different suprahaplotypes and the highest number of different haplotypes. This was in congruence with the findings of MEUSEL et al. (1965), who proposed the Carpathian Mountains as the centre of diversity of this species and the area of origin of ssp. *arenosa*. Following CLAUSS & KOCH (2006) the most likely explanation for the high

diversity of the *A. arenosa*-group in the Eastern Alps and the Balkans is the fact that main parts of this region were never glaciated and served as glacial refuge area for many taxa. Refuge populations subject to genetic isolation and adaptation to local environments could have given rise to divergent local forms.

Due to MEUSEL et al. (1965) it is known, that *A. arenosa* is neosynanthrop in the northern and north-western area of its distribution range. Fig. 2-8 shows the distribution range of *A. arenosa*. From this knowledge it seems strange, that the species-group harboured the highest number of different cpDNA-suprahaplotypes in the formerly glaciated North and additionally showed a high cpDNA nucleotide diversity. However cpDNA-haplotype estimates and ITS-diversity estimates were lowest in this region.

Formerly glaciated areas which were recolonized must not automatically show lower levels in diversity as it is typically expected. Several factors can increase special diversity estimators like secondary contact between migrating populations originating from different refugia (WIDMER & LEXER 2001). An indication for a good migration ability is the fact, that suprahaplotype and haplotype sharing is extensive among the regional groups. This was in contrast to the *A. lyrata* and *A. halleri* group.

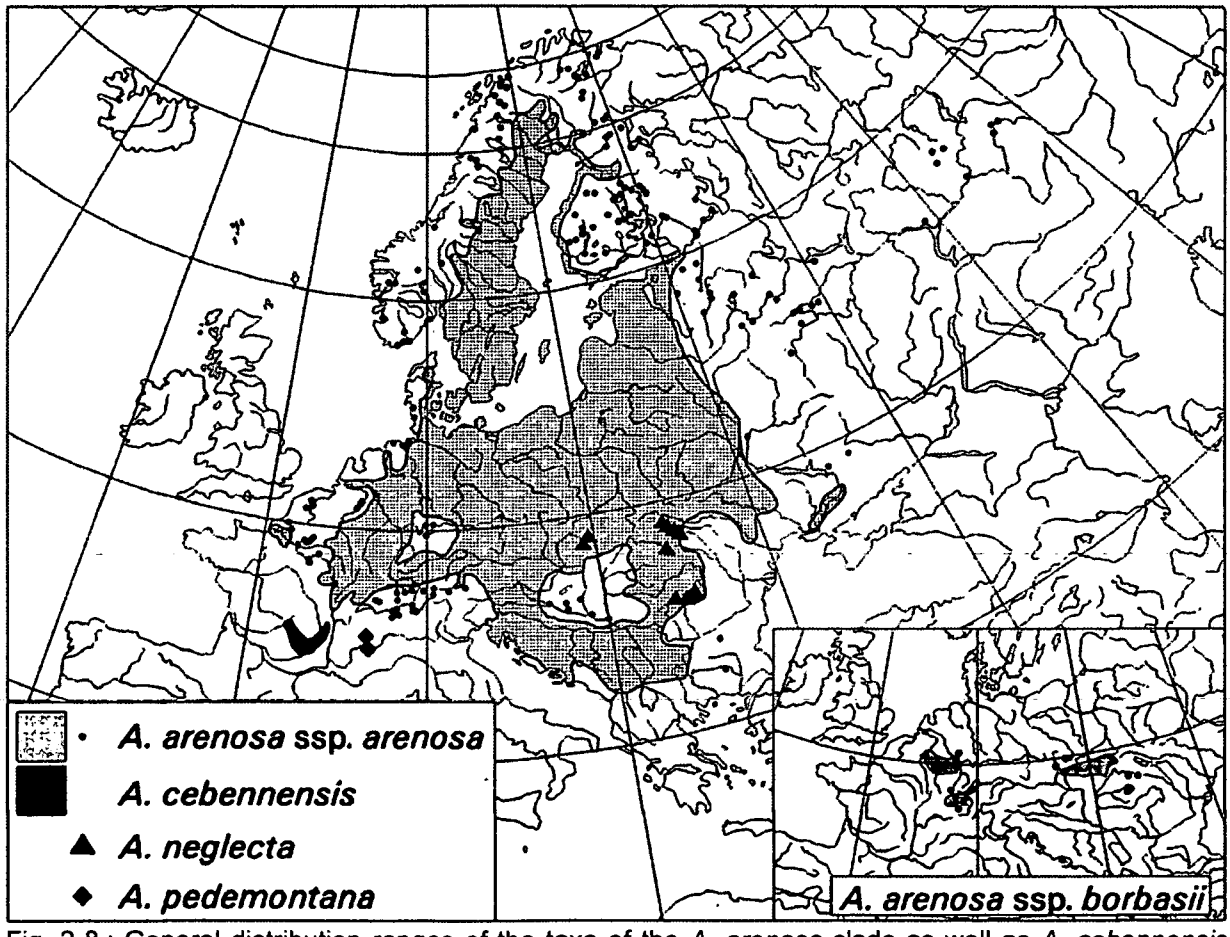

Fig. 2-8 : General distribution ranges of the taxa of the *A. arenosa* clade as well as *A. cebennensis* and *A. pedemontana* (HOFFMANN 2005).

Chapter 3

Genetic analyses of two sympatrically occurring *Arabidopsis* species in a non-glaciated Pleistocene refuge area

Abstract

The non-glaciated Limestone Alps at the edge of the Eastern Alps in Austria have long been known to be a refuge area for several species. In close contact to this area is situated the Wachau, which was also non-glaciated during the last Ice Age. The enclosed Wachau is the valley between Melk and Krems which the river Danube carved out in the silicate bedrock. Tetraploids of *Arabidopsis arenosa* and *A. lyrata* ssp. *petraea* with high morphological variation and morphological intermediates between the two species are found there. Sequence analyses of a cpDNA marker (*trnL*-Intron + *trnL/F*-IGS) revealed a high amount of haplotypes that only occurred in the Wachau (i.e. they are unique) for both species which is evidence for a relatively long in situ evolution. *A. l.* ssp. *petraea* however showed quite a larger number of unique (84%) than common haplotypes (16%) in comparison to *A. arenosa* (45% unique to 55% common haplotypes). Diploids and tetraploids of both species were found south of the Wachau in the non-glaciated Limestone Alps. Sequence analyses of the nuclear marker ITS detected a hybrid zone in the Limestone Alps with several intermediate ITS-sequences. It is assumed that morphologically more "*A. l.* ssp. *petraea*-like" hybrids from this area migrated along the river Traisen into the Wachau. With their enlarged ecological amplitude they were able to colonize the silicate bedrock. A later contact with south migrating *A. arenosa*'s lead to small contact zone with probably ongoing hybridization between *A. arenosa* and *A. l.* ssp. *petraea* in the northern parts of the Wachau.

Introduction

The Lower Austrian edge of the Eastern Alps is known to have been a refuge area during the last glaciation (POLATSCHEK 1966; NIKLFELD 1967, 1970; ZIMMERMANN 1972). The area remained more or less ice-free (only locally glaciated) and contains many (at least 25) absolutely or relatively endemic mountainous plant taxa. These taxa are restricted to azonal, mostly edaphic dry locations and could therefore better resist the ice-age climate (NIKLFELD 1970). *Arabidopsis lyrata* ssp. *petraea* is one of these taxa, which is supposed to have survived the last glaciation at the edge of the Eastern Alps (NIKLFELD 1967, 1970; ZIMMERMANN 1972). An area which is in close contact with the edge of the Eastern Alps is the Wachau. It is the relatively small valley of the Danube river between Melk and Krems. Geologically it belongs to the Bohemian Massif (mostly silicate bedrock) and is separated from the Limestone Alps (limestone, dolomite) by the Alps Vorland, but connected via the river Traisen. It has to be emphasized that the Wachau was also non-glaciated during the last Ice Age and that two species of the genus *Arabidopsis*, *A. l.* ssp. *petraea* and *A. arenosa* are sympatrically distributed in this area. It has been reported that *A. l.* ssp. *petraea* is occasionally difficult to distinguish from *A. arenosa* (POLATSCHEK 1966) and

sometimes mistaken for *A. arenosa* (NIKLFELD 1970). Observations on morphology in the field led the supervisor of this project, M. KOCH, to the assumption of hybridization between the two species (M. KOCH, pers. com.)

The Wachau acts as a contact zone for several species groups. *Pulsatilla grandis* Wenderoth. and *P. vulgaris* Mill. (and probably *P. pratensis* (L.) Mill. – Koch, pers. comment) for example form a gradual hybrid swarm along the Danube from South Germany to Austria of postulated postglacial origin (ZIMMERMANN 1963). A project which analysed this hybrid swarm in the Wachau was able to verify this with genetic markers (KOCH unpub.). A second example is found in the genus *Gentianella* Moench.. Three species (*G. aspera* (Hegetschw. & Heer) Dostal ex Skalicky, Chrtek & Gill, *G. austriaca* (A. & J. Kern) Holub and *G. germanica* (Willd.) Börner) came into contact in the Wachau and seemed to have stabilized as a hybrid complex in several populations (GREIMLER & JANG 2002; M. KOCH unpub.).

Hybridization, reticulation, and polyploidy have played an important role in speciation and differentiation of the Brassicaceae (MARHOLD & LIHOVA 2006), particularly in Quaternary times during periods which were greatly influenced by glaciation and deglaciation (KOCH & MUMMENHOFF 2006). The genus *Arabidopsis* shows potential for hybridization. *A. suecica* is known to be a hybrid between *A. thaliana* and *A. arenosa* with a recent and unique origin between 300 000 and 12 000 years ago (JAKOBSSON 2006 and further literature there). *A. lyrata ssp. kamchatica* is supposed to be of hybrid origin (SHIMIZU et al. 2005; BECK 2007) at least from Japan and Taiwan. Artificial hybrids were generated for several pairs of species (summarized in NASRALLAH et al. 2000).

The aim of the current study was to investigate genetic variation with cpDNA-markers (maternally inherited) and nuclear markers for both species on population level in order to get information on survival in this refuge area and possible hybridization events.

Material and methods

Sampling area

The sampling area "Wachau" (Fig. 3-1, 3-2) is situated in the Danube valley of Lower Austria. It is an area of special interest due to several factors. It was not glaciated during the Ice Age and is today still blessed with a fine climate, since it belongs to the pannonical floral region. This floral region is characterized by warm and dry summers, and low precipitation. The diversity of different habitats is very high in the Wachau. There are deep carved cross running valleys and different substrates mainly gneiss and granite, but also small layers of marble and serpentine. The later with its high amount of heavy metal is responsible for bare slopes with open vegetation. The region is several 100 square kilometres large and geomorphologically and climatically isolated from the surrounding landscape, which makes it more difficult for species to enlarge their distribution range.

Sampling

In spring 2003 (12.04.-16.05) individuals of *Arabidopsis lyrata* ssp. *petraea* and *Arabidopsis arenosa* were collected from 70 populations in the "Wachau" area at an elevation from 210 msm to 680 msm. Two populations (44 and 67) were collected very close to the Wachau (4.5 and 12 km respectively far from the nearest "Wachau"-population). From each population 5 to 20 individuals were collected depending on the population size. All plants were deposited as vouchers at Heidelberg herbarium

(HEID) and a few leaves from each plant were dried in silica gel.

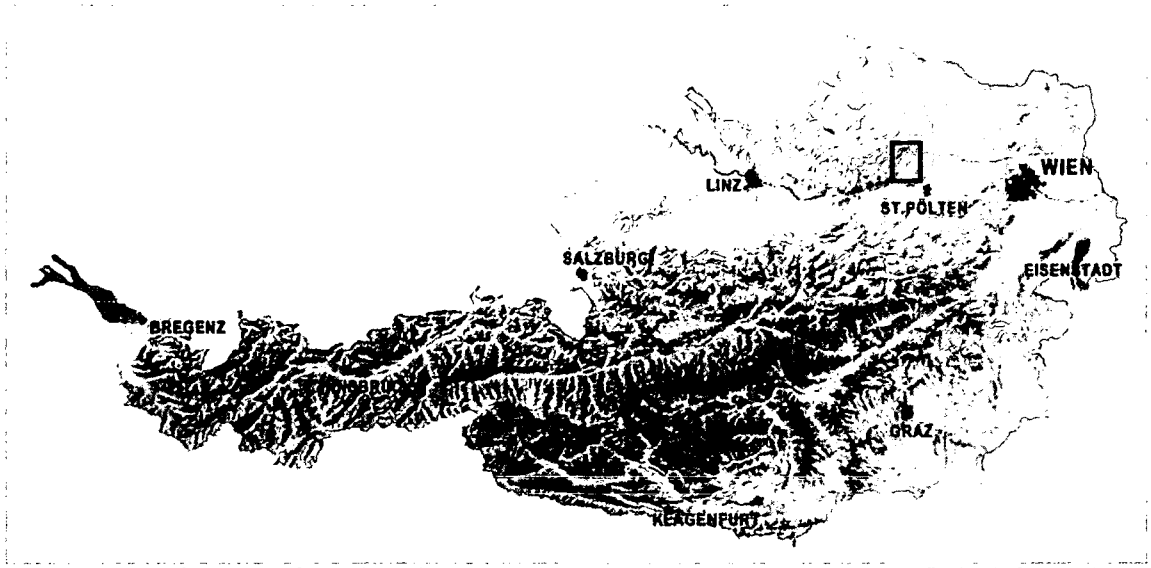

Figure 3-1: Map of Austria (from Austrian map 32.BEV). Sampling area "Wachau" black framed.

DNA-extraction and amplification
See chapter 1.

Data analysis
Data analysis followed chapter 1, but no phylogenetic trees were constructed. Sequences which were gained from the Wachau dataset were compared with all sequences of the world-wide dataset (chapter 1) and new sequences were submitted to GenBank.

Results

Cp-DNA data
In the Wachau (incl. two populations very close to it) the cpDNA marker *trnL*-Intron + *trnL/F*-IGS was analysed for 709 individuals of 72 populations of *A. arenosa* and *A. l.* ssp. *petraea*. *A. arenosa* only occurred between Weißenkirchen and Dürnstein (Fig. 3-2) and was analysed with 123 individuals from 11 populations within the Wachau and two populations very close to it (Senftenberg - 44, Kronsegg - 67). Of the wider distributed *A. l.* ssp. *petraea* 586 individuals from 59 populations were analysed. In toatal 51 different haplotypes (respectively 14 suprahaplotypes) were detected, 18 exclusively in *A. arenosa*, 30 exclusively in *A. l.* ssp. *petraea* and three in both species (Tab. 3-1). From the later three haplotypes, two were assigned to *A. arenosa* (70 and 78), since they were more common in this species in the Wachau or also occurred outside the Wachau in this species or fitted better to this species, if the genuswide haplotype-network (Fig. 1-5) was compared. The third haplotype (11) could be assigned to *A. l.* ssp. *petraea*. Since the sampling size was very different for both species, haplotype numbers were corrected for sample size through rarefaction (Tab. 3-2). With the corrected values *A. arenosa* was identified as the species with the clearly higher number of different haplotypes in the Wachau. The two close *A. arenosa* populations were included in this value, but since they showed no new haplotypes, the value remained uneffected.

Tab. 3-1: Suprahaplotypes and haplotypes found in *A. arenosa* and *A. l.* ssp. *petraea*, in the Wachau with frequencies in percentages. Yellow shaded haplotypes with * are unique, rare haplotypes occur only in one population. Percentage values were calculated for each species separately.

species	suprahaplotype	haplotype	in *A.arenosa*	in *A. l.* ssp. *petraea*	rare	distribution in Austria	distribution world-wide
A. arenosa	A	8	16,3%			Vbg, T, Sbg, Styria, NÖ	F, CH,D,P,CZ,SK,SLO,FIN,SW
		154*	3,3%				
		33	2,4%			T, Sbg	CH,D,P,SK,SLO,R
		17	1,6%		x	N-NÖ, Styria,OÖ	CH,SK,
		64*	0,8%		x		
		34	0,8%		x		R
		150*	0,8%		x		
		153*	0,8%		x		
	B	71*	17,9%				
		96	9,8%			OÖ, N-NÖ, Styria	P(1 in A.halleri)
		155*	0,8%		x		
		156*	0,8%		x		
		76	0,8%		x	OÖ, Styria	
	E	59	13,0%				SK,R
		69*	0,8%		x		
		65	0,8%		x		D,SK
	L	109	0,8%		x	N-NÖ	SK,H,R
	U	57	7,3%				SLO
A. arenosa and *A. lyrata* ssp. *petraea*	A	78*	0,8%	0,2%			
	C	11*	1,6%	2,4%			
	Q	a70	17,9%	1,2%		N-NÖ	CZ
A. lyrata ssp. *petraea*	A	146		1,9%		Styria (1in A. arenosa)	
		174*		0,2%	x		
	B	81		1,0%		NÖ, Styria (1 in A. halleri)	
		159*		0,5%	x		
	C	13		5,6%		NÖ, Styria (1 in A. halleri)	CZ
		29		59,0%		NÖ, NÖ (1 in arenosa+1 in halleri), Styria (1 in A. halleri), K (1 in A. halleri)	CZ,D,Iceland
		151*		0,3%			
		161*		0,9%			
		158*		0,5%	x		
		162*		0,3%	x		
		163*		0,2%	x		
		168*		0,5%	x		
		169*		1,0%	x		
		170*		0,3%	x		
		171*		0,7%	x		
		16		0,3%	x		Canada, Alaska
	AC	1*		13,5%			
		152*		3,2%			
		166*		0,5%			
		165*		0,2%	x		
		172*		0,2%	x		
		178*		0,2%	x		
	AH	157*		0,2%	x		
	AI	160*		2,6%			
		176*		0,2%	x		
	AJ	164*		0,3%	x		
	AK	175*		1,4%	x		
	AL	167*		0,2%	x		
		177*		0,2%	x		
	AM	179*		0,2%	x		

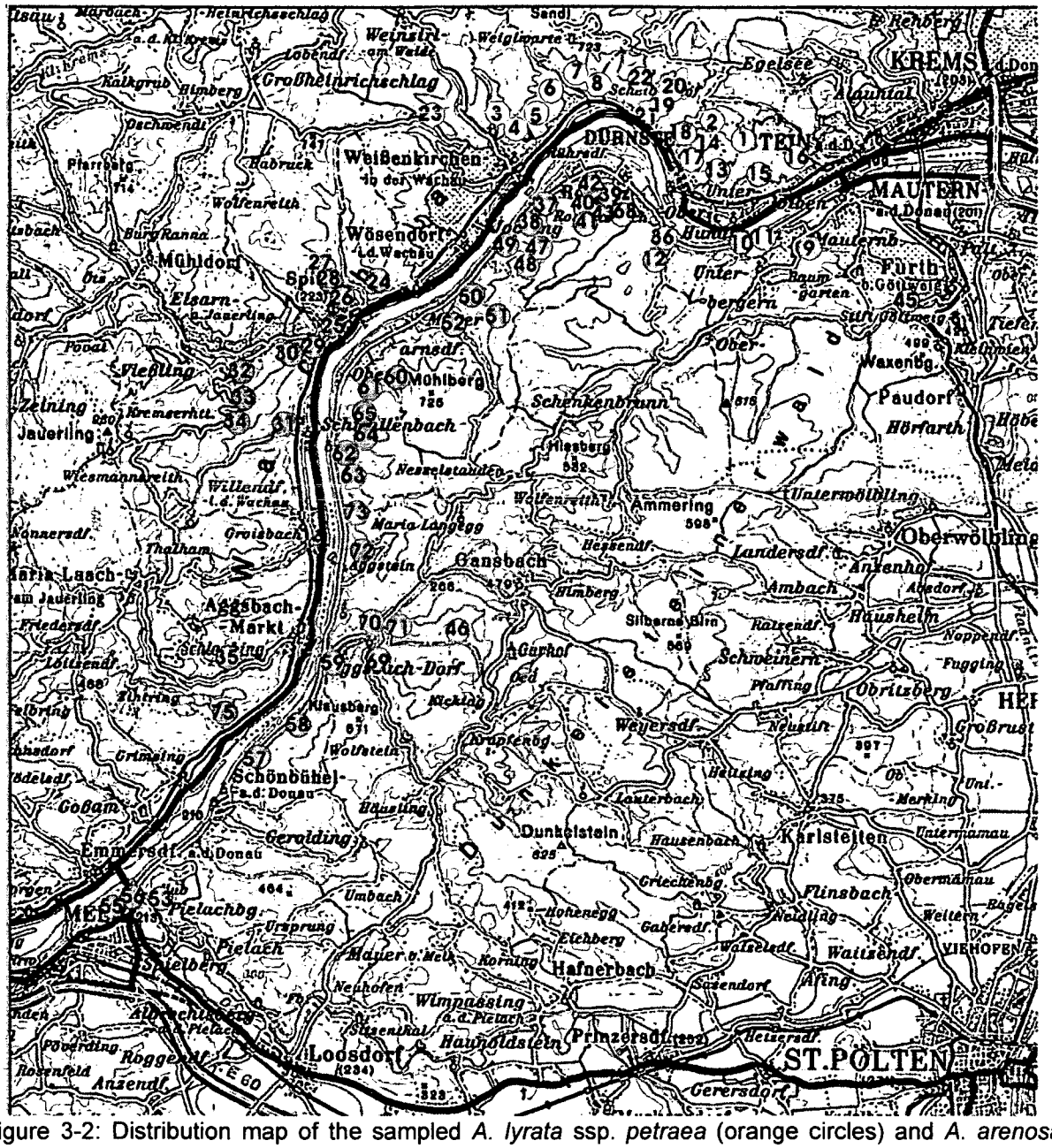

Figure 3-2: Distribution map of the sampled *A. lyrata* ssp. *petraea* (orange circles) and *A. arenosa* (green circles) populations in the Wachau (Danube valley between Melk and Krems) (from Austrian map 32.BEV).

Table 3-2: Haplotype-richness of *A. arenosa* and *A. l.* ssp. *petraea* in the Wachau, Austria – sample size (n), number of counted haplotypes (H), number of haplotypes corrected for sample size 123 (H_{corr}) through rarefaction. Individuals with plastid capture were excluded from the calculation.

Species	n	H	$H_{corr.}$
A. arenosa	121	20	20
A.lyrata ssp. *petraea*	578	31	16,51

In *A. l.* ssp. *petraea* seven unique suprahaplotypes and 26 unique haplotypes (respectively 11 – corrected through rarefaction) were detected (Tab. 3-1, 3-3). That means that these suprahaplotypes/haplotypes were only found in the Wachau, when comparing the Wachau data set with the data from the world-wide sampling (see chapter 2). In *A. arenosa* no unique suprahaplotypes, but at least nine unique haplotypes were found. Many of these unique suprahaplotypes in *A. l.* ssp. *petraea* contained only one or two different haplotypes with low frequencies, but there was also one (AC), which harboured six different haplotypes, with one even very frequent

haplotype (Tab. 3-1). Comparing unique haplotypes with common haplotypes (occur also outside the Wachau) 84% unique haplotypes and 16% common haplotypes were found in *A. l.* ssp. *petraea*, whereas in *A. arenosa* 45% of the haplotypes were unique and 55% are common.

Table 3-3: Unique haplotypes of *A. arenosa* and *A. l.* ssp. *petraea* in the Wachau, Austria – number of sampled individuals with a unique haplotype (n), number of counted unique haplotypes (U), number of unique haplotypes corrected for sampled size 30 (U_{corr}) through rarefaction. Individuals with plastid capture were exclude from the calculation.

Species	n	U	$U_{corr.}$
A. arenosa	30	9	9
A. lyrata ssp. *petraea*	180	26	11

The frequency of haplotypes occurring in more than one population in the Wachau (Fig. 3-3, Tab. 3-1) was totally different for the two species. *A. l.* ssp. *petraea* showed one very frequent common haplotype (29 – with 59%) which occurred in nearly every population. Unique haplotype 1 (13,5%) and the common haplotype 13 (5,6%) were also frequent, but the remaining haplotypes showed values below 5%. A lot of haplotypes were even very rare and occurred only in one population (Tab. 3-1). In *A. arenosa* there were several haplotypes with frequency values between 7,3% and 17,9% and no single but quite frequent haplotype like it was the case in *A. l.* ssp. *petraea*.

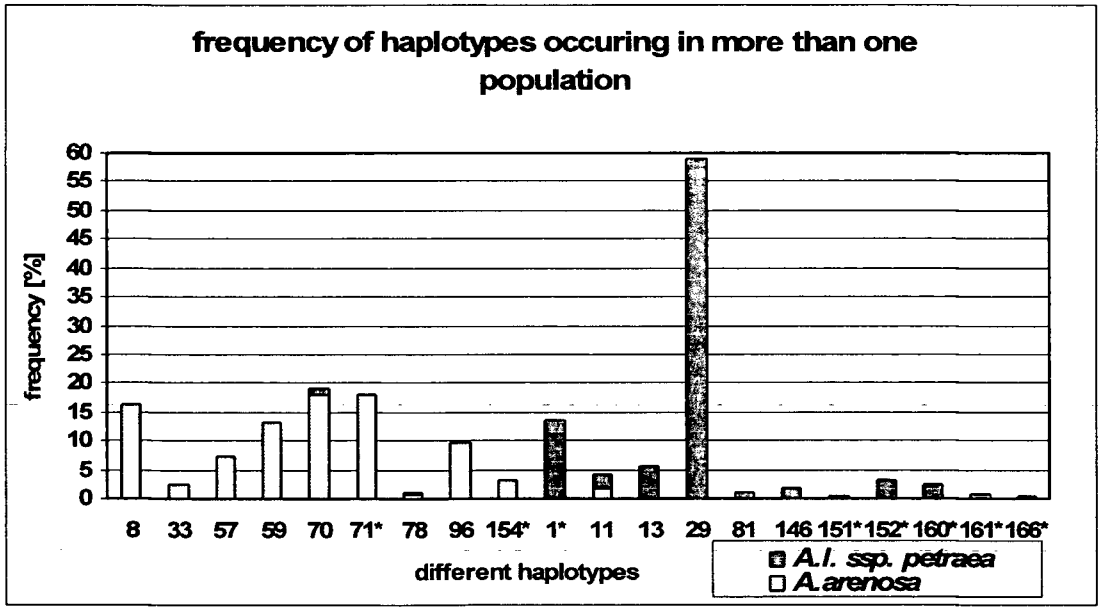

Figure 3-3: Frequency of haplotypes occurring in more than one population. (*) marks haplotypes that are unique for the Wachau. Rare haplotypes (not shown), which occur only in one population, have frequencies lower than 1.63%. Percentage values were calculated for each species separately.

Fig. 3-4 shows the distribution of the detected haplotypes in the Wachau. In the populations of the narrowly distributed *A. arenosa* a maximum of seven different haplotypes was detected. On average 4,5 haplotypes were found in a population of this species (3,9 if the two adjacent populations in the NE are counted). In the wider distributed *A. l.* ssp. *petraea* a maximum of five different haplotypes respectively 2,5 haplotypes on average was detected per population.

The question arose where the five haplotypes of *A. l.* ssp. *petraea*, which not unique to the Wachau, were distributed both in the Wachau and the species world-wide distribution range:

- The most frequent haplotype 29 was distributed in nearly every population of the Wachau, also at the S-edge and the NE-edge of the Wachau. In the world-wide sampling it also occurred in Iceland, Germany, Czech Republic and in Austria in several *A. arenosa* and *A. halleri* individuals of Carinthia, Styria and Lower Austria. In *A. l.* ssp. *petraea* it was detected SE of the Wachau in Lower Austria between Mariazell and Mödling, where several hybrids between *A. arenosa* and *A. lyrata* were identified (Fig. 3-8, Tab. 3-5).
- Haplotype 13 was found in several populations all over the Wachau, also in the populations south of the Danube at the NE-edge of the Wachau. In the world-wide sampling it was detected in the Czech Republic and in the hybrid zone southeast of the Wachau.
- Haplotype 81 was only detected in the middle of the Wachau south of the Danube and in one individual in the Traisen valley near Freiland (hybrid zone) which was from ITS at 90% of its species typical positions a hybrid between *A. l.* ssp. *petraea* and *A. arenosa*. A second individual with this haplotype was found in *A. halleri* of Styria.
- The common haplotype 146 was found in the Wachau only north of the Danube in populations around Dürnstein. In the world-wide sampling it was detected in one individual of *A. arenosa* in Styria, Mixnitz, near the famous last Ice Age refuge area Bärenschützklamm, a quite southern location of *A. l.* ssp. *petraea*.
- The last common haplotype of the Wachau haplotype 16 showed an unusual distribution range. It occurred in one population of the Wachau (population 10 at Ferdinandswarte) and additionally only in Alaska and Canada.

Some of the common haplotypes of *A. arenosa* in the Wachau were found north of the Wachau, from the populations of Senftenberg and Kronsegg and adjacent populations (R. SCHMICKL unpub. data) via the Kamptal up to Hardegg at the Austrian border. In this area five of the 11 common haplotypes were found. Further distribution ranges of the common haplotypes were the Carpathians and other parts of Europe (Tab. 3-1). Unique haplotypes (Fig. 3-5 B) were everywhere in the Wachau. Some of the unique haplotypes were further distributed. For example 1, 11 or 152 in *A. l.* ssp. *petraea* or 71 in *A. arenosa*, but most of the unique haplotypes were only found in one population.

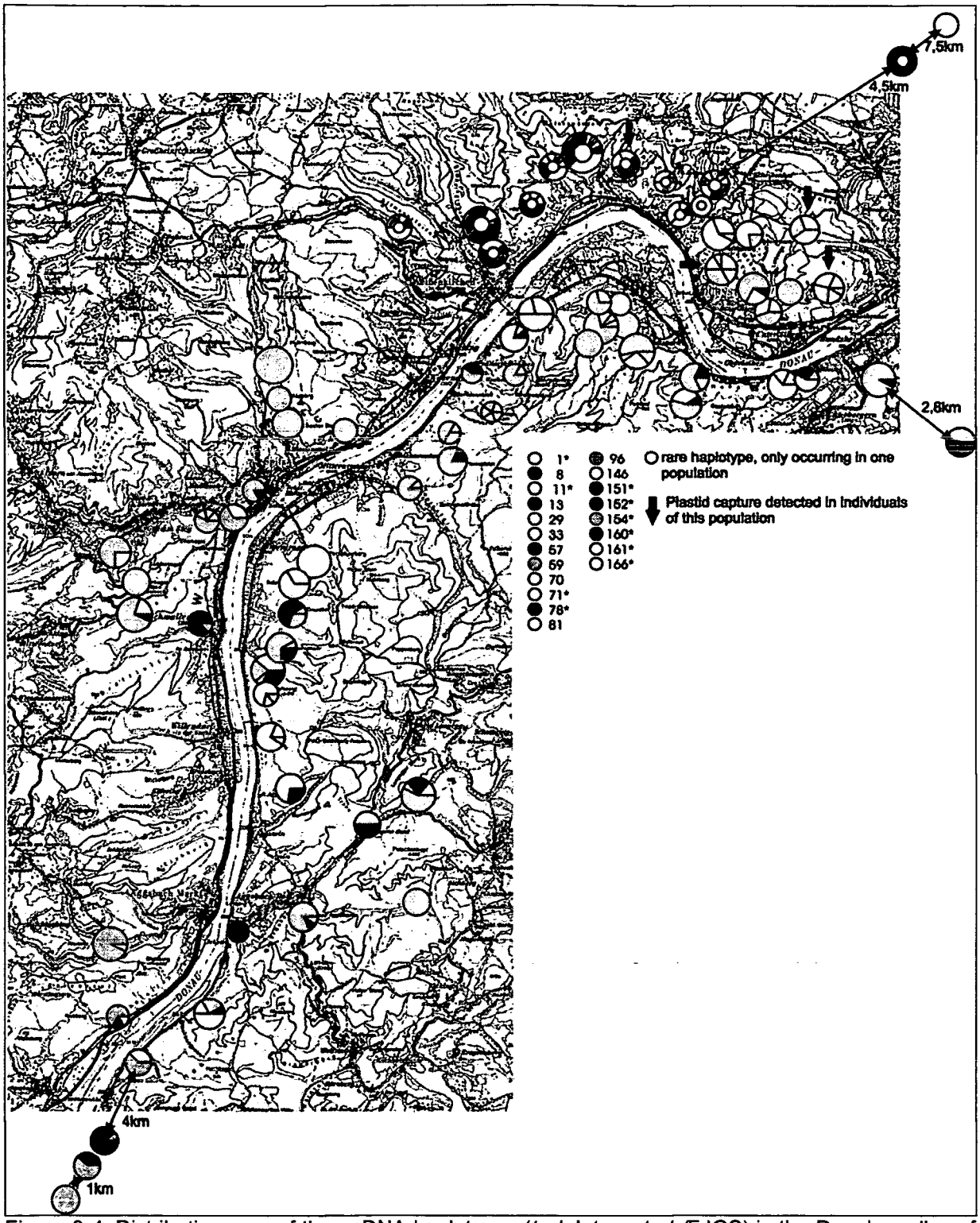

Figure 3-4: Distribution map of the cpDNA-haplotypes (*trnL*-Intron+*trnL/F*-IGS) in the Danube valley of the Wachau plus two adjacent populations. Circles mark populations of *A. l.* ssp. *petraea*, rings mark populations of *A. arenosa*. The size of the circle corresponds with the number of analysed haplotypes. The segments of the circles and rings in different colours symbolize the frequency of the different haplotypes in the single populations. An asterix (*) beside a haplotype number symbolizes an unique haplotype.

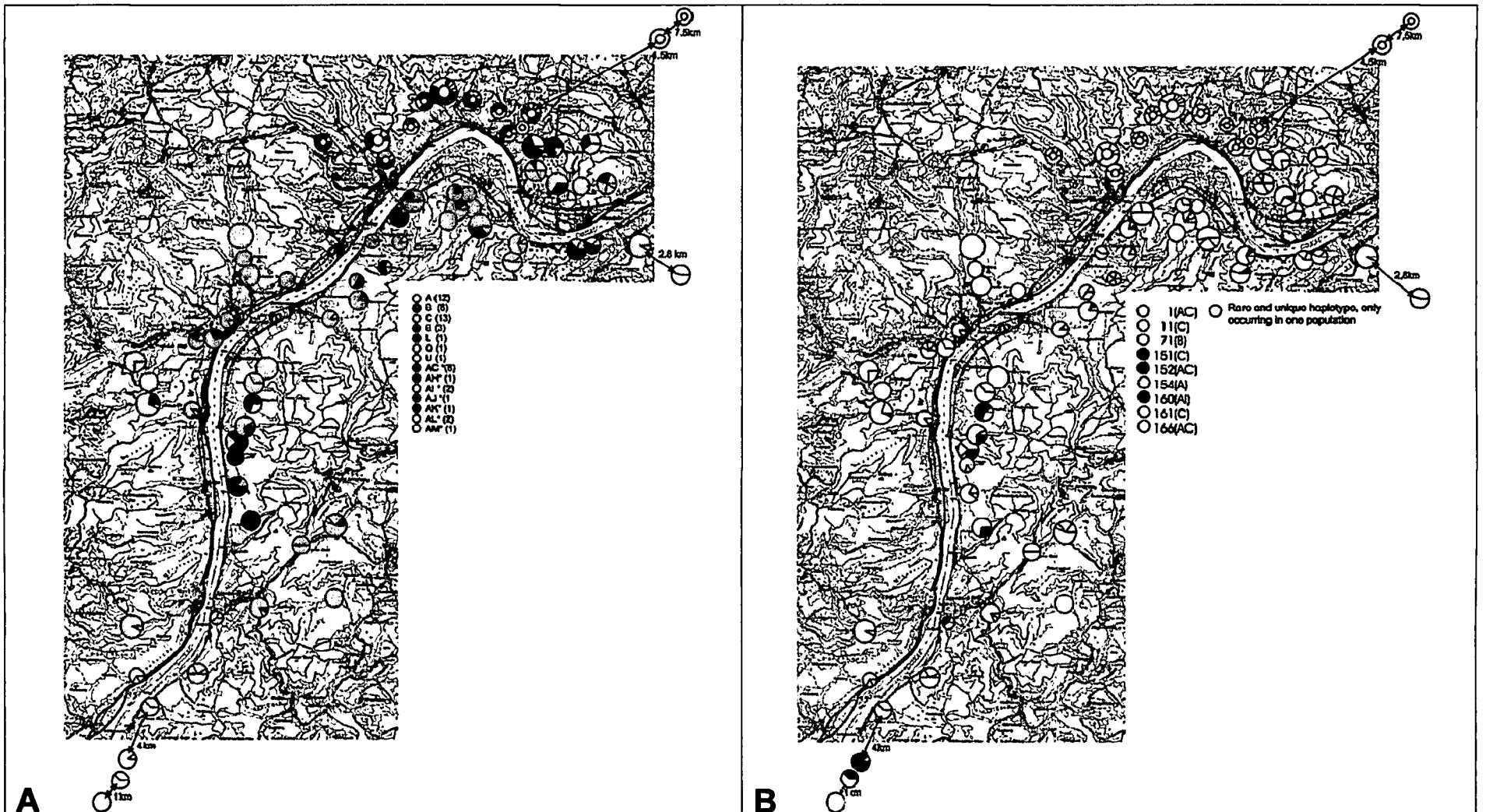

Fig. 3-5: Distribution map of (A) cpDNA-suprahaplotypes, (B) unique haplotypes in the Wachau plus two adjacent populations. Circles mark populations of *A. l.* ssp. *petraea*, rings populations of *A. arenosa*. The size of the circle corresponds with the number of analysed accessions. The segments of the circles and rings in different colours symbolize the frequency of the different suprahaplotypes/unique haplotypes in the single populations. An asterix (*) symbolizes an unique suprahaplotype (A). The number of haplotypes which are combined as a suprahaplotype is found within bracketts.

ITS data

The ITS-marker was analysed for 563 individuals in the Wachau and the adjacent populations of Senftenberg and Kronsegg. Due to technical problems sequences of only 61 individuals of *A. arenosa* were obtained although quite more individuals were sampled. Due to ambiguities 58 different ITS-sequences were detected, 19 belonging to *A. arenosa* and 39 to *A. l.* ssp. *petraea*. In *A. l.* ssp. *petraea* all different ITS types were assigned to ITS-supratype a. In *A. arenosa* the majority was assigned to ITS-supratype g, but also one ITS type to m, one to J and two to two new ITS-supratypes.

A. arenosa and *A. l.* ssp. *petraea* in the Wachau show at ten sites of the ITS-marker species-typical bases (Tab. 3-4). Hybrids between the two species show the typical peaks of both species at these positions (Fig. 3-6).

	79	162	206	212	241	246	406	445	534	569
A. l. ssp. petraea		A	G	A		G			A	C
A. arenosa	C		A	G	C	A	C	C	G	G
Hybrid	Y	W	R	R	Y	R	Y	Y	R	S

T+C = Y
A+T = W
A+G = R
C+G = S

Tab. 3-4: Species differing positions in the ITS marker of *A. arenosa* and *A. l.* ssp. *petraea*. A hybrid between them shows two bases at the species differing positions, which is symbolized with the corresponding ambiguous codes.

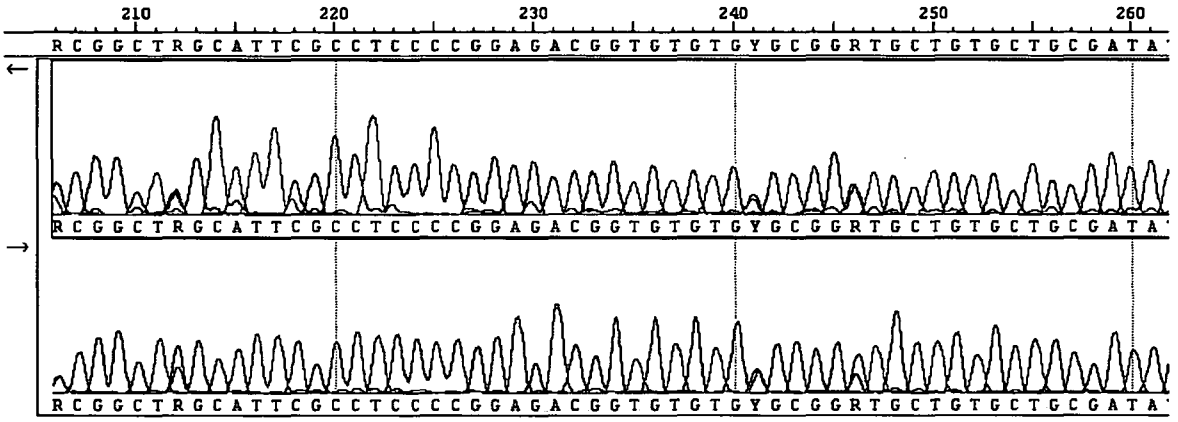

Figure 3-6: Electrophorogram of a hybrid between *A. arenosa* and *A. l.* ssp. *petraea*. At position 206, 212, 241 and 246 two peaks of different bases can be seen.

In the ITS-sequences of the Wachau different amounts of sequence ambiguities at these 10 species typical sites were obtained. The distribution of sequences with differing levels of "hybridization" or "incomplete homogenization of the ITS-copies" is shown in Fig. 3-7A. Most populations in *A. l.* ssp. *petraea* showed individuals with at least one ambiguous position (10%).

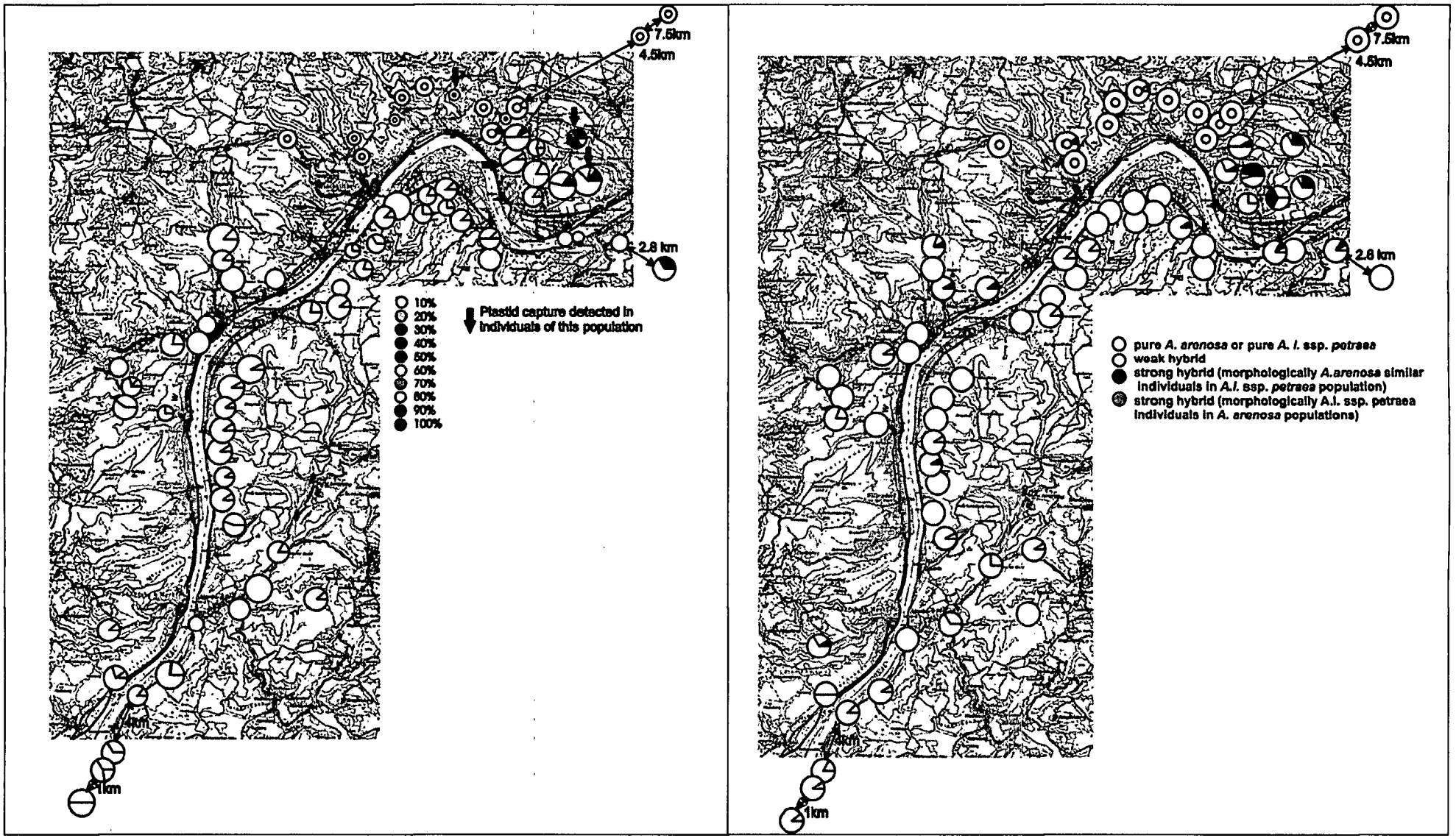

Fig. 3-7: Distribution map of (A) ITS-sequences with different amounts of additative polymorphisms. The size of the circle corresponds with the number of analysed accession. (B) distribution map of morphological hybrids following MALL (2005), redrawn. No correspondence of circle size and number of analysed accessions.

Individuals with more ambiguous positions were found at the S-edge of the Wachau and at the NE-edge of the Wachau. At the later a lot of individuals with different and high amounts of ambiguous positions are found, especially eastwards of Dürnstein, where *A. l.* ssp. *petraea* came in close contact with *A. arenosa*. In *A. arenosa* however only two individuals with sequence ambiguities were detected, both in populations near *A. l.* ssp. *petraea* populations, one even with ambiguities at all species typical positions (100%).

Fig. 3-7B shows the results of the morphometric studies of the Wachau populations, which were done by C. MALL (MALL 2005). In many *A. l.* ssp. *petraea* populations of the Wachau at least weak hybrids, but also strong hybrids were found when analysing morphology. Populations eastwards Dürnstein harboured again many hybrid individuals. In *A. arenosa* mainly pure individuals were detected. Only in two populations strong hybrids were found. But these were different populations than those in Fig. 3-7A, where also hints for hybridization were found in two populations.

Additional interspecific hybrids outside the Wachau

Within the main species groups of Arabidopsis, *A. halleri*, *A. lyrata* and *A. arenosa* hybridization was only detected between the later two when analysing the ITS marker. Among the ITS sequences of the world-wide dataset we detected several sequences with additative polymorphisms at up to ten sites, which were species-specific positions of *A. arenosa* and *A. lyrata* ssp. *petraea*. Individuals which carry sequences with additative polymorphisms (at least five of ten sites) are listed in Tab. 3-5 and their distribution is shown in Fig. 3-8. Interestingly they all occurred in the non-glaciated parts at the eastern edge of the Eastern Alps.

Table 3-5: Hybrids between *A. arenosa* and *A. lyrata* ssp. *petraea* found in the world-wide ITS dataset. The two species differ at ten positions in the ITS alignment. A pure *A. arenosa* and a pure *A. lyrata ssp. petraea* from Austria (hybrids are only found in Austria) are given at the top and at the bottom of the table. Hybrids between the two species show ambiguities at five to ten of the ten species-specific positions in the ITS alignment.

			ITS alignment position										
Acc.Nr.	Species (herb. sheet)	Location	79	162	206	212	241	246	406	445	534	569	hybrid
	pure A. a.	A	C	T	A	G	C	A	C	C	G	G	
813	A. a.	A, Gloggnitz	Y	W	R	R	Y	R	Y	Y	R	S	100%
Card0205	A. a.	A, Kematen	Y	W	R	R	Y	R	Y	Y	R	S	100%
Card0040	A .a	A, Lambach	Y	W	R	R	C	R	Y	Y	R	S	90%
Card0169	A. a./A. l. ssp. p.	A, S Lilienfeld, Freiland	Y	W	R	R	C	R	Y	Y	R	S	90%
Card0209	A. a.	A, Mürztaler Alp.	Y	T	A	G	C	A	Y	Y	R	S	50%
Card0321	A. a.	A, Schwarzau im Gebirge	Y	W	R	R	C	R	Y	Y	G	G	70%
76-05	A. p.	A, Türnitzer Höger	T	A	R	R	Y	R	T	Y	A	S	60%
79-02	A. a.?	A, S Lilienfeld	T	W	R	R	Y	R	T	Y	A	S	70%
	pure A. p.	A	T	A	G	A	T	G	T	T	A	C	

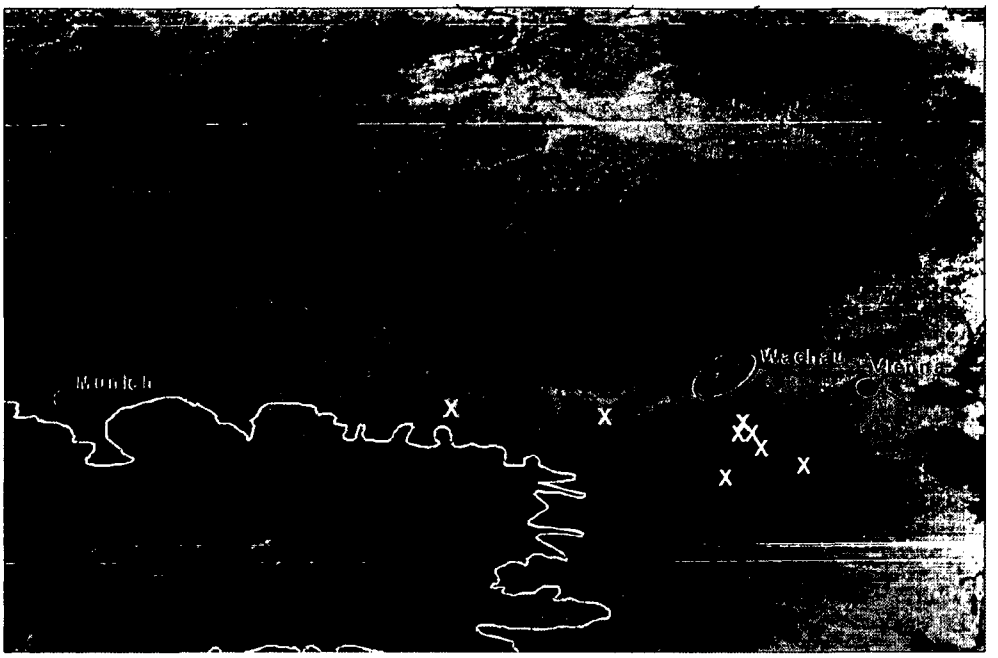

Fig. 3-8: Distribution map of hybrids (at least 50% of the differing positions show additive polymorphisms) between *A. arenosa* and *A. lyrata* ssp. *petraea* at the eastern edge of the Eastern Alps. Hybrids are shown as yellow X, *A. l.* ssp. *petraea* accessions as orange circles and *A. arenosa* accessions as green squares. The white line shows the maximum extend of the glaciers during the last Ice Age. Additionally hybrid individuals are found in the Wachau. For detailed distribution see Fig. 3-7.

Discussion

Both species *A. arenosa* and *A. lyrata* ssp. *petraea* showed high amounts of cpDNA variation in the Wachau, an area which was not glaciated during the last Ice Age. Relatively long in situ evolution in large or contacting populations was assumed for both species, since both harboured many unique haplotypes. Several factors argued for hybridization between the two species. Plastid capture in three populations of *A. l.* ssp. *petraea* near Dürnstein, where the species comes in close contact with *A. arenosa*, were detected and in one population further south (pop. 59). In *A. arenosa* plastid capture was found in at least one population. ITS data and morphometric data (Fig. 3-7) gave evidence for weak hybridization in more or less all *A. l.* ssp. *petraea* populations in the entire Wachau and strong hybridization in the *A. l.* ssp. *petraea* populations near Dürnstein. In *A. arenosa* again at least little evidence for hybridization was found.

Favourite evolutionary scenario

A hybrid zone between the two species was detected in the non-glaciated parts of the Eastern Alps (Fig. 3-8). Although *A. l.* ssp. *petraea* is known as a diploid in most of its distribution range (Norway, Sweden, Russia, Iceland, Scotland, Germany; POLATSCHEK 1966; JONSELL ET AL. 1995; CLAUSS & MITCHELL-OLDS 2006) tetraploids beside diploids are found in this hybrid zone (POLATSCHEK 1966; DOBES & VITEK 2000; HENNRICH 2005). In the Wachau only tetraploids were found in both species (HENNRICH 2005, BALDAUF 2006). In *A. arenosa* diploids beside tetraploids are found in the Carpathian Mountains (M. KOLNIK pers. com.) In Austria outside this hybrid zone only tetraploids are detected (POLATSCHEK 1966; DOBES & VITEK 2000) and in the hybrid zone at least one diploid *A. arenosa* individual (HENDRICH 2005) could be

found. It can be assumed that extensive hybridization followed by reticulation took place between *A. arenosa* and *A. l.* ssp. *petraea* in this area southeast of the Wachau, which resulted in the formation of tetraploid hybrids during the Pleistocene. These more "*A. l.* ssp. *petraea* like" tetraploids then migrated along mountain rivers (e.g. the river Traisen) into the Wachau. Hybrid individuals (accessions Card0169; 79-02, 76-05 from Tab. 3-5), in the Traisen-valley near Freiland and Hohenberg can be taken as evidence of that hypothesis. POLATSCHEK (1966) has already suspected, based on his study on morphology and karyology of *A. arenosa* and *A. l.* ssp. *petraea* at the edge of the Eastern Alps, that the two species are more closely related than presumed previously. He however thought that the tetraploid local populations of *A. l.* ssp. *petraea* in this area resulted from autopolyploidy. This might be true for the populations near Mödling and Baden which are distributed on dolomite and limestone, but not for the populations of the Wachau. A successful colonization of the Wachau with its different substrates (gneiss, granite serpentine) can only be explained with an enlarged ecological amplitude of a hybrid, because outside the Wachau *A. l.* ssp. *petraea* grows on calcerous and dolomite substrates. An argument for a colonization of the Wachau with "*A. l.* ssp. *petraea* like" tetraploids can also be seen in the low amount of common haplotypes (haplotypes which also occur outside the Wachau) in the Wachau. Only five haplotypes (16%) occurred outside the Wachau. They were the first colonizers, but did not stay in contact with their source area. If their would have been an ongoing contact there should not be 84% unique haplotypes in *A. l.* ssp. *petraea* in the Wachau.

In the Wachau *A. l.* ssp. *petraea* grows on extremely flat soil mostly above gneiss, but also serpentine in an open, species-poor *Pinus sylvestris* forest, often together with *Quercus petraea*. A plant community, the *Cardaminopsio petraea-Pinetum* Hübel et Holzner 1977 (MUCINA ET AL. 1993) is even described for the Wachau. *A. arenosa* however does not prefer so extreme habitats. It grows on slightly deeper grounds, in more species-rich *Quercus petraea* forests sometimes together with *Carpinus betulus* or *Pinus sylvestris*. So both species grow in different habitats and in different parts of the Wachau. But near Dürnstein they come in close contact and are only separated by beech-forests, where they do not grow, less than 1 km from each other. In this area a second and maybe still active contact zo ne between *A. l.* ssp. *petraea* and *A. arenosa* is observed (Fig. 3-7). Plastid capture, strong hybrid ITS sequences and morphologically intermediate plants are a strong proof for that. Interestingly introgression mainly occurred from *A. arenosa* into *A. l.* ssp. *petraea* and less the other way around.

Comparing the amount of unique to common haplotypes for both species in the Wachau, it seemed as if *A. arenosa* colonized the Wachau later than *A. l.* ssp. *petraea*. *A. arenosa* only showed 45% unique haplotypes, whereas *A. l.* ssp. *petraea* 84%. A more or less connected distribution range from the Wachau northwards to the Kamptal and up to Hardegg at the northern Austrian border exists and claims for a colonization of the Wachau by *A. arenosa* from the North.

References

ABOTT, R.J. and BROCHMANN, C. (2003): History and evolution of the arctic flora: in the footsteps of Eric Hulten: Mol. Ecol. 12: 299-313.

AKAIKE, H. (1974): A new look at the statistical model identification. IEEE Trans. Auto. Control. 19 (6): 716-723.

AL-SHEHBAZ, I.A.; O'KANE, S.L.JR. and PRICE, R.A. (1999): Generic placement of species excluded from *Arabidopsis* (Brassicaceae). Novon 9: 296-307.

AL-SHEHBAZ, I.A. and O'KANE, S.L.JR. (2002): Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). In: SOMMERVILLE, C.R. and MEYEROWITZ, E.M. (ed): The *Arabidopsis* Book. American Society of Plant Biologists, Rockville, Maryland. http://www.aspb.org/publications/arabidopsis/ (17.10.2003).

AL-SHEHBAZ, I.A.; BEILSTEIN, M.A. and KELLOG, E.A. (2006): Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst. Evol. 259: 89-120.

ANSELL, S.W. (2004): Historical biogeography and population genetics of the plant *Arabidopsis petraea*. PhD thesis. Wolfson College Cambridge.

AVISE, J.C.; ARNOLD, J.; BALL, R.M.; BERMINGHAM, E.; LAMB, T.; NEIGEL, J.E.; REEB, A. and SAUNDERS, N.C. (1987): Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Evol. S. 18: 489-522.

BALDAUF, H. (2006): Retikulate Evolution in glazialen Refugialgebieten: Zytologische Untersuchungen in der Gattung *Arabidopsis* (Brassicaceae). Staatsexamensarbeit. Universität Heidelberg.

BECK, J.B.; AL-SHEHBAZ, I.A.; O'KANE S.L.JR. and SCHAAL, B.A. (2007): Further insights into the phylogeny of *Arabidopsis* (Brassicaceae) from nuclear *Atmyb2* flanking sequence. Mol. Phyl. Evol. 42: 122-130.

BITTKAU, C. and COMES, H.P. (2005): Evolutionary processes in a continental island system: Molecular phylogeography of the Aegean *Nigella arvensis* alliance (Ranunculaceae) inferred from chloroplast DNA. Mol. Ecol. 14: 4065-4083.

CLAUSS, M.J.; COBBAN, H. and MITCHELL-OLDS, T. (2002): Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaceae). Mol. Ecol. 11: 591-601.

CLAUSS, M. and KOCH, M.A. (2006): *Arabidopsis* and its poorly known relatives. Trends Plant Sci. 11: 449-459.

CLAUSS, M. and MITCHELL-OLDS, T. (2006): Population genetic structure of *Arabidopsis lyrata* in Europe. Mol. Ecol. 15: 2753-2766.

CLEMENT, M.; POSADA, D. and CRANDALL, K.A. (2000): TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9: 1657–1660.

DOBES, C. and VITEK, E. (2000): Documented chromosome number checklist of Austrian vascular plants. Museum of Natural History Vienna.

DOBES, C.; MITCHELL-OLDS, T. and KOCH, M.A. (2004): Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii*, *A. x divaricarpa*, and *A. holboellii* (Brassicaceae). Mol. Ecol. 13 (2): 349-370.

DOYLE, J.J. and DOYLE, J.L. (1987): A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem. Bul. 19: 11–15.

EL MOUSADIK, A. and PETIT, R.J. (1996): High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skells] endemic to Morocco. Theoret. Appl. Genet. 92: 832–839.

EXCOFFIER, L.; LAVAL, G. and SCHNEIDER, S. (2005): Arlequin ver.3.0: An integrated software package for population genetics data analysis. Evol. Bioinf. Online. 1: 47-50.

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

GREIMLER, J. and JANG, C.G. (2002): Reticulate relationships between two widespread taxa in the genus *Gentianella*: *G. austriaca* and *G. germanica*. In: Symposium Plant Species-level Systematics: Patterns, Processes and new Applications. Nationalherbarium Netherland, Leiden. The Netherlands, 13.-15.11.2002.

HENNRICH, A. (2005): Hybridartbildung innerhalb der Gattung *Cardaminopsis* in Ostösterreich – Cytologische Variabilität. Staatsexamensarbeit. Universität Heidelberg.

HOFFMANN, M.H. (2005): Evolution of the realized climatic niche in the genus *Arabidopsis* (Brassicaceae). Evolution 59 (7): 1425-1436.

HULTEN, E. (1937): Outline of the history of arctic and boreal biota during the Quaternary period. Lehre J. Cramer, New York.

HYLANDER, N. (1957): *Cardaminopsis suecica* (Fr.) Hiit., a northern amphidiploid species. Bull. Jard. Bot. Etat. Brux. 27: 591-604.

JAKOBSSON, M.; HAGENBLAD, J.; TAVARE, S.; SÄLL, T.; HALLDEN, C.; LIND-HALLDEN, C. and NORDBORG, M. (2006): A unique recent origin of the allotetraploid species: evidence from nuclear DNA markers. Mol. Biol. Evol. 23 (6): 1217-1231.

JONES, B.M.G and AKEROYD, J.R. (1993): *Cardaminopsis*. In Tutin et al. (eds): Flora europaea 1, ed. 2, Cambridge University Press, Cambridge.

JONSELL, B.; KUSTAS, K. and NORDAL, I. (1995): Genetic variation in *Arabis petreaea*, a disjunct species in northern Europe. Ecography 18: 321-332.

KAWABE, A. and MIYASHITA, N.T. (2003): DNA polymorphism in active gene and pseudogene of the cytosolic phosphoglucose isomerase (PgiC) loci in *Arabidopsis halleri* ssp. *gemmifera.* Mol. Biol. Evol. 20(7): 1043-1050.

KOCH, M.; BISHOP, J. and MITCHELL-OLDS, T. (1999): Molecular Systematics and Evolution of *Arabidopsis* and *Arabis.* Plant Biol. 1: 529-537.

KOCH, M.A.; HAUBOLD, B. and MITCHELL-OLDS, T. (2000): Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Molec. Biol. Evol. 17(10): 1483-1498.

KOCH, M.; HAUBOLD, B. and MITCHELL-OLDS, T. (2001): Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. Am. J. Bot. 88 (2): 534-544.

KOCH, M.; AL-SHEHBAZ, I.A. and MUMMENHOFF, K. (2003a): Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). Ann. Missouri Bot. Gard. 90: 151-171.

KOCH, M.; DOBES, C. and MITCHELL-OLDS, T. (2003b): Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Mol. Biol. Evol. 20(3): 338-350.

KOCH, M.A; DOBES, C; MATSCHINGER, M.; BLEEKER, W.; VOGEL, J.; KIEFER, M. and MITCHELL-OLDS, T. (2005): Evolution of the *trn*F(GAA) Gene in *Arabidopsis* relatives and the Brassicaceae family: Monophyletic origin and subsequent diversification of a plastidic pseudogene. Mol. Biol. Evol. 22: 1032-1043.

KOCH, M.A. & KIEFER, C. (2006): Molecules and migration: biogeographical studies in cruciferous plants. Pl. Syst. Evol. 259: 121-142.

KOCH, M.A.; KIEFER, C.; EHRICH, D.; VOGEL, J.; BROCHMANN, C. and MUMMENHOFF, C. (2006): Three times out of Asia minor: the phylogeography of *Arabis alpina* (L.) Brassicaceae. Mol. Ecol. 15: 825-839.

KOCH, M.A. and MUMMENHOFF, K. (2006): Editorial: Evolution and phylogeny of the Brassicacea. Pl. Syst. Evol. 259: 81-83.

KOCH, M.A. and MATSCHINGER, M. (2007): Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. PNAS. 104: 6272-6277.

KOCH, M.A.; WERNISCH, M.M.; SCHMICKL, R. and CLAUSS, M. (2007): *Arabidopsis thaliana*'s wild relatives: an updated overview on systematics, taxonomy and evolution. in prep. for Taxon.

KOLNIK, M. and MARHOLD, K. (2006): Distribution, chromosome numbers and nomenclature conspect of *Arabidopsis halleri* (Brassicaceae) in the Carpathians.

Biologia, Bratislava 66 (1): 41-50.

LAIBACH, F. (1943): *Arabidopsis thaliana* (L.) Heynh. als Objekt für genetische und entwicklungsphysiologische Untersuchungen. Bot. Archiv 44: 439-455.

MALL, C. (2005): Hybridartbildung innerhalb der Gattung *Cardaminopsis* in Ostösterreich. - Untersuchungen zur morphologischen Variabilität. Staatsexamensarbeit. Universität Heidelberg.

MARHOLD, K. and LIHOVA, J. (2006): Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. Pl. Syst. Evol. 259: 143-174.

MESICEK J. (1970): Chromosome counts in *Cardaminopsis arenosa* agg. (Cruciferae). Preslia (Praha) 42: 225-248.

MEUSEL, H.; JÄGER, E. and WEINERT, E. (1965): Vergleichende Chorologie der zentraleuropäischen Flora. G. Fischer Verlag. Jena.

MITCHELL-OLDS, T. (2001): *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. Trends Ecol. Evol. 16(12): 693-700.

MIYASHITA, N.T.; KAWABE, A.; INNAN, H. and TERAUCHI, R. (1998): Intra- and interspecific DNA variation and codon bias of the alcohol dehydrogenase (Adh) locus in *Arabis* and *Arabidopsis* species. Mol. Biol. Evol. 15(11): 1420-1429.

MUCINA, L., GRABHERR, G. and WALLNÖFER, S. (1993): Die Pflanzengesellschaften Österreichs 3: Wälder und Gebüsche. G. Fischer Verlag. Jena.

MUMMENHOF, K. and HURKA, H. (1995): Allopolyploid origin of *Arabidopsis suecica* (Fries) Norrlin: evidence from chloroplast and nuclear genome markers. Bot. Acta 108: 449-456.

MUMMENHOFF, K.; FRANZKE, A. and KOCH, M. (1997): Molecular phylogenetics of *Thlaspi* s. l. (Brassicaceae) based on chloroplast DNA restriction site variation and sequences of the internal transcribed spacers of nuclear ribosomal DNA. Can. J. Bot. 75: 469-482.

NASRALLAH, M.E.; YOGEESWARAN, K.; SNYDER, S. and NASRALLAH, J.B. (2000): *Arabidopsis* species hybrids in the study of species differences and evolution of amphiploidy in plants. Plant Physiol. 124: 1605-1614.

NEI, M. (1987): Molecular evolutionary genetics. Columbia University Press. New York.

NIKLFELD, H. (1967): Das Alter der submediterranen und illyrischen Flora und Vegetation am niederösterreichischen Alpen-Ostrand. Mitt. Int. Ver. Vegetationsk., Ostalp.-Dinar. Sekt. 7: 153-162.

NIKLFELD, H. (1970): Der niederösterreichische Alpen-Ostrand – ein Glazialrefugium montaner Pflanzensippen. Jb. Ver. Schutze Alpenpfl. u. –Tiere 37: 3-54.

O'KANE, S.L.JR.; SCHAAL, B.A. and AL-SHEHBAZ, I.A. (1996): The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences. Syst. Bot. 21: 559-566.

O'KANE, S.L.JR. and AL-SHEHBAZ, I.A. (1997): A synopsis of *Arabidopsis* (Brassicaceae). Novon 7, 323-327.

O'KANE, S.L. JR. and AL-SHEHBAZ, I.A. (2003): Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. Ann. Missouri Bot. Gard. 90: 603-612.

Pigliucci, M. (2002): Ecology and evolutionary biology of *Arabidopsis*. In: SOMMERVILLE, C.R. and MEYEROWITZ, E.M. (ed): The *Arabidopsis* Book. American Society of Plant Biologists, Rockville, Maryland. http://www.aspb.org/publications/arabidopsis/ (17.10.2003).

POSADA, D. and CRANDALL, K.A. (1998): Modeltest: Testing the model of DNA substitution. Bioinf. 14: 817-818.

RAMOS-ONSINS, S.E.; STRANGER, B.E.; MITCHELL-OLDS, T. and AGUADE, M. (2004): Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. Genetics 166: 373-388.

ROSENBERG, N.A. and NORDBORG, M. (2002): Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Gen. 3: 380–390.

SCHAAL, B.A.; HAYWORTH, D.A.; OLSEN, K.M.; RAUSCHER, J.T. and SMITH, W.A. (1998): Phylogeographic studies in plants: problems and prospects. Mol. Ecol. 7: 465-474.

SHARBEL, T.F.; HAUBOLD, B and MITCHELL-OLDS, T. (2000): Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Mol. Ecol. 9: 2109-2118.

SHIMIZU, K.K.; FUJII, S.; MARHOLD, K.; WATANABE, K. and KUDOH, H. (2005): *Arabidopsis kamchatica* (Fisch. ex DC.) K. Shimizu & Kudoh and *A. kamchatica* ssp. *kawasakiana* (Makino) K. Shimizu & Kudoh, new combinations. Acta Phytotax. Geobot. 56: 163-172.

SWOFFORD, D.L. (2002): Phylogenetic analysis using parsimony (and other methods). Version 4. Massachusetts: Sinauer, Sunderlands.

TABERLET, P.; GIELLY, L.; PAUTOU, G. and BOUVET, J. (1991): Universal primers for amplification of three non-coding regions of chloroplast DNA. Pl. Molec. Biol. 17: 1105–1109.

TABERLET, P.; FUMAGALLI, L.; WUST-SAUCY, A.-G. and COSSON, J.-F. (1998): Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7: 453-464.

TEMPLETON A.R.; CRANDALL, K.A. and SING, C.F. (1992): A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics. 132: 619-63.

VAN TREUREN, R.; KUITTINEN, H.; KÄRKKÄINEN, K.; BAENA-GONZALEZ, E. and SAVOLAINEN, O. (1997): Evolution of microsatellites in *Arabis petraea* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana*. Mol. Biol. Evol. 14(3): 220-229.

WARWICK, S.I.; AL-SHEHBAZ, I.A. and SAUDER, C.A. (2006): Phylogenetic position of *Arabis arenicola* and generic limits of *Aphragmus* and *Eutrema* (Brassicaceae) based on sequences of nuclear ribosomal DNA. Can. J. Bot. 84: 269-281.

WHITE, T.J.; BRUNS, T.; LEE, S. and TAXLOR, J. (1990): Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: INNIS, M.; GELFAND, D.; SNINSKY, J.J. and WHITE, T.J. (ed.): PCR protocols: a guide to methods and applications. Academic Press, San Diego.

WIDMER, A. and LEXER, C. (2001): Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. Ecol. Evol. 16 (6): 267-269.

WRIGHT, S.I.; LAUGA, B. and CHARLESWORTH, D. (2003): Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. Mol. Ecol. 12: 1247-1263.

ZIMMERMANN, A. (1972): Pflanzenareale am niederösterreichischen Alpenostrand und ihre florengeschichtliche Deutung. Dissertationes Botanicae 18. Lehre. Verlag J.Cramer. Berlin.

Summary

The combination of the former genus *Cardaminopsis* and the model plant *Arabidopsis thaliana* to the newly defined genus *Arabidopsis* in the late nineties of the last century drew the attention of the scientific community to the new relatives of the model plant. It opened a wide field of expanding knowledge from the model plant with its restricted ecological niche to the very diverse wild relatives. The current PhD-thesis investigated systematics, phylogeography and evolution of these close relatives of *Arabidopsis thaliana* in order to give information on the several species which are already partly used in research but are still unknown in many aspects. The study is based on sequence analyses of nuclear ribosomal DNA (ITS) and maternally inherited cpDNA (*trnL*-Intron + *trnL/F*-IGS) of a representative world-wide sampling.

The first chapter dealt with the systematics of the genus. Beside *A. thaliana* three major species groups (*A. halleri*, *A. lyrata* and *A. arenosa*) and three single species (*A. cebennensis*, *A. croatica* and *A. pedemontana*) were recognized. *A. suecica*, which is known to be a hybrid between *A. thaliana* and *A. arenosa* has been excluded from the analyses. The old diploids *A. cebennensis* and *A. pedemontana* with their restricted geographic distribution range were together with diploid *A. halleri* basal to the remaining species. *A. arenosa* and *A. lyrata* with diploids and tetraploids and *A. croatica* were connected by hybrids. *A. lyrata* ssp. *kamchatica* from Japan was proved to be of hybrid origin with *A. lyrata* and *A. halleri* ssp. *gemmifera* as parental species.

In the second chapter the phylogeography of the three main species groups *A. halleri*, *A. lyrata* and *A. arenosa* was investigated. At the eastern edge of the Eastern Alps in Austria all three species groups showed high amounts of genetic variation. This area was not glaciated during the last Ice Age and served as a refuge area for these species as well as for many other species. The centre of diversity of *A. halleri* was found in this region. In regard to the two other main species groups *A. halleri* showed lower numbers of different haplotypes/suprahaplotypes. As a mesophytic plant mainly distributed at mountainous to alpine habitats with no tendency to occupy extreme climates *A. halleri* seemed to have suffered most from glaciation. *A. lyrata* however grows at dryer or colder habitats and is a successful colonizer of the Northern regions. It seems to have colonized North America more than once and has survived the last glaciation northwest (in Beringia) and southeast of the ice sheet. In Europe the highest nucleotide diversity was found in the formerly glaciated northern parts of Europe which was probably due to periglacial survival of the species and postglacial recolonization of this area from different refuge areas. The centre of genetic diversity of the *A. arenosa* group was found in the Carpathian Mountains. This is in congruence with high morphological and cytological variation, which has led to the description of several taxa.

The third chapter dealt with a hybrid zone which was detected in this non glaciated area at the eastern edge of the Eastern Alps in Austria. Introgression from *A. arenosa* into *A. lyrata* ssp. *petraea* seems to have happened there. These hybrids with their enlarged ecological niche then migrated north into the Wachau, an area with a different substrate which is normally not colonized by *A. lyrata* ssp. *petraea*. This must have happened before the last glaciation, because lots of new cpDNA-suprahaplotypes and haplotypes originated there. Secondary and maybe ongoing contact of this hybrid with south migrating *A. arenosa* was observed in the north part of the Wachau, where the two species meet.

Zusammenfassung

Die Eingliederung der Gattung *Cardaminopsis* in die Gattung *Arabidopsis* in den späten Neuzigern des letzten Jahrhunderts lenkte die Aufmerksamkeit der wissenschaftlichen Gemeinschaft auf die nächsten Verwandten der Modellpflanze *Arabidopsis thaliana*. Das ermöglichte es, das an der ökologisch wenig spannenden Modellpflanze erworbene Wissen auf die viel diverseren Verwandten auszudehnen. Die vorliegende Dissertation beschäftigte sich mit der Systematik, der Phylogeographie und der Evolution dieser nahen Verwandten von *Arabidopsis thaliana* um das Wissen über die einzelnen Arten, die zwar teilweise schon in der Forschung verwendet werden, aber in vielen Aspekten noch ungenügend erforscht sind, zu vermehren. Die Studie basierte auf Sequenzanalysen ribosomaler Kern-DNA (ITS) und mütterlich vererbter Chloroplasten-DNA (*trnL*-Intron + *trnL/F*-IGS) repräsentativer, weltweiter Herkünfte.

Das erste Kapitel beschäftigte sich mit der Systematik der Gattung. Neben *Arabidopsis thaliana* konnten drei große Artengruppen (*A. halleri*, *A. lyrata* und *A. arenosa*) und drei einzelne Arten (*A. cebennenis*, *A. croatica* und *A. pedemontana*) unterschieden werden. *A. suecica*, die ein Hybrid zwischen *A. thaliana* und *A. arenosa* ist, wurde nicht in die Analysen einbezogen. Die geographisch sehr eng verbreiteten alten Diploiden, *A. cebennenisis* und *A. pedemontana* gruppierten sich zusammen mit der diploiden *A. halleri* basal zu den restlichen Arten. *A. arenosa* und *A. lyrata* mit jeweils Diploiden und Tetraploiden, waren zusammen mit *A. croatica* über Hybriden verbunden. Für *A. l.* ssp. *kamchatica* aus Japan konnte bewiesen werden, dass sie hybridogenen Ursprungs ist mit *A. lyrata* und *A. h.* ssp. *gemmifera* als Elternarten.

Im zweiten Kapitel wurde die Phylogeographie der drei großen Artengruppen *A. halleri*, *A. lyrata* und *A. arenosa* untersucht. Alle drei Gruppen zeigten hohe genetische Variation am Ostrand der Ost-Alpen in Österreich. Dieses Gebiet war während der letzten Eiszeit unvergletschert und diente nicht nur für diese Arten als Refugialgebiet. In dieser Region lag nun auch das Diversitätszentrum von *A. halleri*. Im Vergleich zu den zwei anderen Artengruppen konnten in *A. halleri* wesentlich weniger Suprahaplotypen bzw. Haplotypen gefunden werden. Als mesophytische Pflanze mit hauptsächlich montaner bis alpiner Verbreitung und keiner Tendenz extreme Klimate zu besiedeln, scheint *A. halleri* am meisten unter der Eiszeit gelitten zu haben. *A. lyrata* hingegen wächst auch in trockeneren oder kälteren Habitaten und gilt als erfolgreiche Besiedlerin der nördlichen Regionen. Es wurde vermutet, dass sie Nord-Amerika mehr als einmal kolonisiert hat und die Eiszeit dort einerseits im Nordwesten des Eisschildes (in Beringia) und andererseits im Südosten überdauert hat. In Europa wurde die höchste Nukleotid-Diversität im ehemals vergletscherten Nord-Europa gefunden. Grund dafür ist möglicherweise das Überdauern der Art nahe der Gletscher und die Wiederbesiedelung aus verschiedenen Refugialgebieten. Das genetische Diversitätszentrum von *A. arenosa* lag in Übereinstimmung mit hoher morphologischer und cytologischer Variation in den Karpaten. Das hat sich auch in der Beschreibung verschiedener Taxa aus dieser Region niedergeschlagen.

Im dritten Kapitel wurde eine Hybridzone, die im unvergletscherten Bereich am Ostrand der Ost-Alpen entdeckt wurde, untersucht. Es scheint zu Introgression von *A. arenosa* in *A. lyrata* ssp. *petraea* gekommen zu sein. Es scheint der hybridogenen *A.l.* ssp. *petraea* durch ihre vergrößerte ökologische Nische möglich geworden zu sein das silikatische Substrat der nördlich gelegenen Wachau zu besiedeln. Die Art ist ansonsten nur auf Kalk oder Dolomit zu finden. Da in der Wachau sehr viele

einizigarte Haplotypen und auch Suprahaplotypen gefunden wurden, wurde angenommen, dass die Wachau jedenfalls vor der letzten Eiszeit besiedelt wurde. In der Wachau scheint es zu sekundärtem Kontakt und möglicherweise immer noch bestehendem Kontakt mit *A. arenosa* gekommen zu sein, die von Norden her eingewandert ist.

Acknowledgements

This study was supported by a grant to my supervisor Prof. Dr. Marcus Koch from the Austrian Science Foundation (FWF) – P15609. I owe Marcus Koch a dept of gratitude for giving me the chance to work on this exciting topic. Further I want to thank him for continuous support and the motivation he always gave me.
I also want to thank Prof. Dr. Karl-Georg Bernhardt for the opportunity to carry out my thesis at the Institute of Botany, University of Natural Resoureces and Applied Life Sciences, Vienna.

I extend my gratitude to the curators of the Biologiezentrum Linz (LI), Herbarium of the ETH Zürich (ZT), Natural History Museum and Herbarium Vienna (W), Herbarium Botanical Institute at Vienna University (WU), Natural History Museum and Herbarium London (BM), the Agriculture and Agri-Food Canada Vascular Plant Herbarium (DAO), the Herbarium of the California Academy of Sciences (CAS), Herbarium, Biology Department Hobart and William Smith Colleges (DH), Herbarium Harvard University (GH) and from the Academy of Sciences Bratislava (SAV) for providing leaf material for destructive sampling.

I want to thank my college Martin Kolnik, who also likes the Brassicaceae as much as I do, for very nice joint plant collecting trips in Slovakia.

Further thanks goes to Andrea Hennrich, Christina Mall and Holgar Baldauf who investigated my collected plants of the Wachau either karyologically or morphologically in their Staatsexamensarbeiten.

During my year at the Heidelberg Institute of Plant Sciences I got a lot of helping hands from the colleagues at the Institute of Systematics and Biodiversity, Christoph Dobes, Irina Hase, Christiane Kiefer, Markus Kiefer, Judita Lihova and my congenial college and special friend Roswitha Schmickl. I also want to thank the whole team for their warm friendship and the joint activities which helped me not to miss my 700 km distant boy-friend Klemens, now husband, too much. I had a wonderful time in Heidelberg! Thank you very much!

Thanks also goes to Tom Lorimer for some help with the difficulties of the English language.

Special thanks goes to my dear parents, my grandparents and my brothers for their endless love and support.

Finally, my highest expression of gratefulness goes to my beloved husband Klemens, who went through ups and downs of this work together with me and always believed in my succeeding. We have reached our goal!

Publications and manuscripts in preparation

1. KOCH, M.A.; DOBES, C.; **MATSCHINGER, M.**; BLEEKER, W.; VOGEL, J.; KIEFER, M.; and MITCHELL-OLDS, T. (2005): Evolution of the *trnF* (GAA) gene in *Arabidopsis* relatives and the Brassicaceae family: Monophyletic origin and subsequent diversification of a plastidic pseudogene. Mol. Biol. Evol. 22 (4): 1032-1043.

2. KOCH, M.A. and **MATSCHINGER, M.** (2007): Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. PNAS. 104: 6272-6277.

3. KOCH, M.A.; **WERNISCH, M.**; SCHMICKL, R. and CLAUSS, M. (2007): *Arabidopsis thaliana's* wild relatives: an updated overview on systematics, taxonomy and evolution. To be submitted to Taxon.

Present address:

Michaela M. Wernisch
Ferrogasse 15/1
1180 Wien
Austria