Nova Acta Leopoldina NF 92, Nr. 342, 129–145 (2005)

Low-level Taxonomy and Intrageneric Evolutionary Trends in Higher Plants

Guido W. GRIMM, Matthias SCHLEE, Nataliya Y. KOMAROVA, Roman A. VOLKOV, and Vera HEMLEBEN (Tübingen)

With 9 Figures and 2 Tables

Abstract

The nuclear encoded 18S, 5.8S, and 25S ribosomal RNA genes (18–25S rDNA) have been used as molecular markers in numerous systematic studies at all hierarchical levels. In the case of low-level evolution the spacer regions of the 18–25S rDNA have been of major interest. Most recently, several objections came up against the use of rDNA sequences since it belongs to a tandemly organized multigene family. Phylogenies inferred from the same marker were found to be conflicting each other, and the discussion focussed on the impact of pseudogenes and unknown paralogy of sequences that might have been used in such studies. This is especially true for intrageneric studies using the internal and external transcribed spacers (ITS1, ITS2, and ETS). If cloning is applied to assemble an ITS/ETS data basis that comprehensively reflects not only the intrageneric (interspecific), but also intraspecific and intraindividual genetic variability, the objections against 18S–25S rDNA can be cleared out. At hand of different variable, differentiated, and complex ITS and also ETS data sets of the arborescent genera *Acer*, *Fagus*, *Zelkova* and the herbs *Lathyrus* and *Solanum*, it can be demonstrated that especially the detected intraspecific variability of the ITS and ETS is a most valuable source for taxonomic-systematic purposes. In addition, appropriate phylogenetic analyses of such data help to trace molecular evolution and intrageneric differentiation through space and time (e.g. ancient hybridization events, gene flow, general mutational trends). So-deduced phylogenetic hypotheses correlate in detail with other data sources such as morphology, biogeography, the fossil record, and ecology, and hence allow to identify general evolutionary trends within these genera. Such a taxonomic, systematic, and phylogenetic resolution on an intrageneric level is up to now not achieved with any known single-copy and/or only uniparentally inherited molecular marker.

Zusammenfassung

Nucleotidsequenzen von verschiedenen Regionen der kernkodierten ribosomalen 18S, 5,8S und 25S RNA-Gene (18S–25S rDNA) werden in zahlreichen systematischen Studien unterschiedlicher Hierarchie verwendet. Im Falle der intragenerischen Studien kommen vor allem Sequenzen der intern oder extern transkribierten Spacerregionen (ITS und ETS) zur Anwendung. Dabei hat sich gezeigt, daß vermutlich auf Grund der Organisation der kernkodierten Gene – tandemartig angeordnete multiple Kopien (‚multigene family') – Probleme bei der Analyse und Interpretation der Daten auftreten. Insbesondere bei Verwendung desselben Markers kam es zu markanten Inkonsistenzen bei den errechneten phylogenetischen Hypothesen. In diesem Zusammenhang werden vor allem zwei Punkte diskutiert: die nicht bekannte Homologie oder Paralogie der jeweilig sequenzierten Tandemrepeats bzw. das Auftreten von sogenannten „Pseudogenen", welche die phylogenetische Hypothese verzerren. Mittels Klonierung läßt sich jedoch eine Datenbasis aufbauen, die einen Einblick in die intraspezifische und sogar individuelle Variabilität der jeweiligen Mitglieder einer Gattung liefert. An Hand unterschiedlich variabler Datensätze (ITS1, ITS2, 5' ETS) von verschiedenen Baumgattungen (*Acer*, *Fagus*, *Zelkova*) und krautigen Pflanzen (*Lathyrus, Solanum*) kann gezeigt werden, daß die vorhandene genetische Variabilität innerhalb der untersuchten Arten von enormer Bedeutung für taxonomisch-systematische Fragestellungen ist. Ferner lassen sich an Hand der umfassenden Datensätze komplexe phylogenetisch-evolutionäre Prozesse, wie z. B. vorrezente Hybridisierungsereignisse und allgemeine Entwicklungstendenzen, detailliert beschreiben. Die resultierenden molekular-phylogenetischen Rekonstruktionen lassen sich problemlos mit

morphologischen, biogeographischen, ökologischen und fossilgeschichtlichen Aspekten verknüpfen. Eine derartig spezifische und detaillierte taxonomisch-systematische und phylogenetische Auflösung der komplexen intrageneri-schen Zusammenhänge ist mit keinem der derzeit bekannten und benutzen ‚single copy' bzw. nur uniparental ver-erbten (Organell-DNA) molekularen Markern zu erreichen.

1. Introduction

For phylogenetic reconstructions at low taxonomic levels the spacer regions of the nuclear en-coded 18S, 5.8S, and 25S ribosomal DNA (18S–25S rDNA), especially the internal tran-scribed spacers (ITS 1 and 2), are widely used and discussed as valuable molecular tool due to a high rate of molecular evolution. The 18S–25S rDNA, located at the nucleolus organiz-ing region (NOR), consists of tandemly arranged repeated units (Fig. 1) and serves as a con-venient model for the investigation of molecular changes in a multigene family. In higher plants, up to 20,000 copies of 18S–25S rDNA per genome were detected (for review see HEM-LEBEN et al. 1988). Sometimes, more than one rDNA locus per chromosomal set can be found (FRANSZ et al. 1998, for review see VOLKOV et al. 2004). Most recently, several objections came up against the use of rDNA sequences for phylogenetic studies. Phylogenies inferred from the 18S–25S rDNA spacer regions appeared to be conflicting. The theoretical discus-sion was mainly focused on the impact of pseudogenes and the unknown paralogy of se-

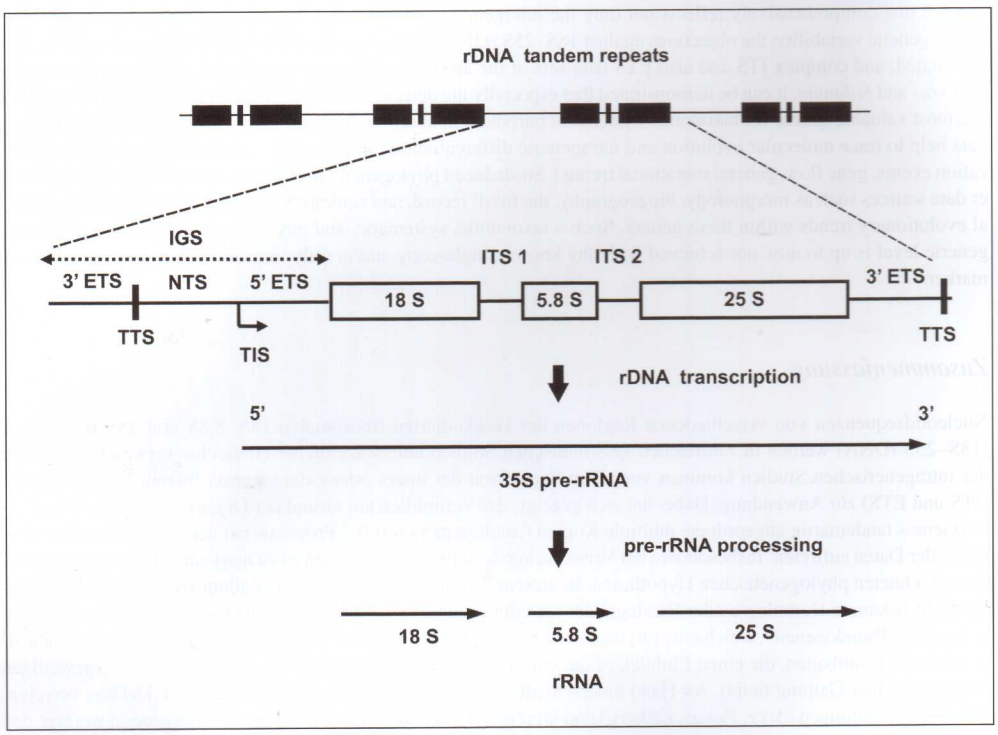

Fig. 1 General organization scheme of the 18S–25S rDNA (modified from VOLKOV et al. 2004). Abbr.: 5.8S, 5.8S rDNA; 18S, 18S rDNA; 25S, 25S rDNA; ETS, external transcribed spacer; IGS, intergenic spacer; ITS, internal tran-scribed spacer; TIS, transcription initiation site; TTS, transcription termination site.

quences used in such studies. In some cases, e.g. for *Quercus* (DENDUANGBORIPANT and CRONK 2001) and *Solanum* (VOLKOV et al. 2003), the problems connected with pseudogenes were documented. This culminated in the conclusion of ÁLVAREZ and WENDEL (2003, p. 429) "that ITS no longer (can) be routinely utilized for phylogenetic analysis, opting instead for using several or more different single-copy nuclear loci". However, the authors acknowledged that "ITS sequence data have and may continue to provide insights into phylogenetic history, polyploidy ancestry, genome relationships, historical introgression, and other evolutionary questions", if the objections arising from paralogs and pseudogenes can be experimentally ruled out or dealt with. This was further underlined by BAILEY et al. (2003), who also addressed this topic with mainly theoretical arguments.

In the case of intrageneric evolution in flowering plants, patterns of differentiation are complex due to reticulation, i.e. ancient and recent hybridization/polyploidization events (Fig. 2). Immediately after allopolyploidization rDNA is bi-parentally inherited (e.g. UCHIMIYA et al. 1983, HEMLEBEN et al. 1998), hence, in an allopolyploid at least 2 paralogous sets of rDNA obtained from both parents should be present (e.g. WISSEMANN 2002, for *Rosa*). Remarkably, in the natural allotetraploid *Nicotiana tabacum* the maternal rDNA got lost and was nearly completely replaced by its paternal counterpart (VOLKOV et al. 1999, LIM et al. 2000), which endured apparent structural rearrangement in the intergenic spacer region (IGS).

Taking into account that several variants of rDNA can be present in the nucleus, cloning of PCR products followed by sequencing of several clones is necessary to assemble a comprehensive data set; in order to identify different levels of intragenomic/-individual variabili-

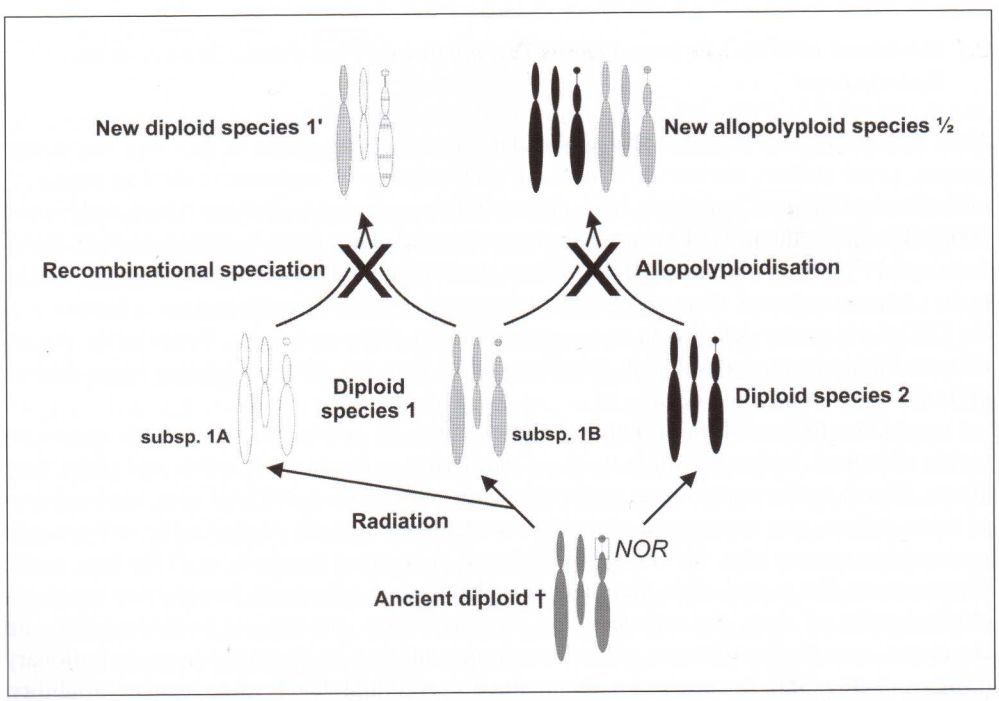

Fig. 2 Diagram exhibiting a simple reticulate evolution and its implication for the current gene pools. Abbr.: NOR, nucleolus organisator region.

ty of rDNA. As consequence, detailed insight into the pathway of intrageneric molecular evolution and differentiation including reticulation can be obtained (GREBENSTEIN et al. 1998, JOBST et al. 1998, STEANE et al. 1999, WISSEMANN 1999, 2002, VON BALTHAZAR et al. 2000, MUIR et al. 2001, DENK et al. 2002, GRIMM 2003, VOLKOV et al. 2003, SCHLEE et al. 2003, DENK and GRIMM 2005, DENK et al. 2005).

In the following paragraphs, we review some of the ongoing molecular work currently undertaken in our lab using the 18S–25S rDNA spacer sequence (ITS and ETS) data as a most valuable source for taxonomic-systematic purposes like the reconstruction of intrageneric phylogeny and identification of ancient and recent horizontal gene flow. In focus are differently variable and complex ITS and also ETS (5' external transcribed spacer, see Fig. 1) data sets of the arborescent angiosperm genera *Acer* (124 species, VAN GELDEREN et al. 1994), *Fagus* (8 species, cf. DENK et al. 2005), *Zelkova* (fam. Ulmaceae, 5 disjunct species, cf. DENK and GRIMM 2005) and the herbal genera *Lathyrus* (~160 species, ASMUSSEN and LISTON 1998, 20 species in sect. *Lathyrostylis*, BÄSSLER 1981), and *Solanum* (234 species in sect. *Petota*, HAWKES 1990). The depicted phylogenetic hypotheses correlate with other data sources such as morphology, biogeography, ecology and the fossil record, and hence, allow identifying general evolutionary trends within these genera. Up to now, such a taxonomic, systematic, and phylogenetic resolution on an intrageneric and inter- and intraspecies level can not achieved with single-copy and/or only uniparentally inherited molecular markers.

2. Patterns of rDNA Spacer Sequence Diversity

2.1 Ribosomal DNA Spacer Sequences as Taxonomic and Phylogenetic Marker at the Generic Level

Now it is commonly accepted that molecular variability can differ in different taxonomic groups. In our studies, obvious differences in accumulation of sequence variability are exemplified in the ITS1 and 2 of genus *Acer* (GRIMM 2003) compared to *Zelkova* (DENK and GRIMM 2005). The ITS sequences of *Acer* are strongly diverged. They contain numerous point mutations and 1–3 bp indels distributed throughout the ITS1 and ITS2; also, several longer indels, up to 15 bp, were found. Generally, length conserved and polymorphic motives alternating in the ITS of *Acer* were detected in the same position in different species. Remarkably, the indels are highly taxon-specific and, therefore, are of high taxonomic-systematic significance (Tab. 1).

In addition to these specific indels, another source of taxonomic information represents highly divergent "oligonucleotide motives" harboring multiple short indels and point mutations. A base-per-base alignment and analysis of such regions is difficult since the evolution of these regions may encompass also indel incidents not directly recognizable in the recent nucleotide sequence (Fig. 3). The example, based on a rather simple 8- or 11-bp long motif, demonstrates, that any possible alignment of such regions is principally biased from a methodological point of view. Nevertheless, the oligonucleotide motives as macrocharacters are conserved, specific for different taxonomic groups and can be evaluated in an evolutionary framework (Fig. 3*D*). In conclusion, due to the observed high levels of molecular variability, unique ITS sequences and oligonucleotide motives define and sustain morphologically distinguished taxa such as subspecies, species, series and sections within genus *Acer*. A set of

Tab. 1 Group-specific indels within the ITS of *Acer* spp. (modified from GRIMM 2003)

	length ITS1	ID1	ID2	ID3	ID4	ID5	length ITS2	ID6	ID7	ID9[1]
Dipteronia sinensis	233bp	–	–	–	–	–	238bp	–	–	**I**
sect. *Acer*	233–242bp	–	–	–	(+)	–	233–242bp	(+)	–	–
ser. *Arguta*	237/238bp	–	–	–	–	–	234bp	–	–	–
sect. *Ginnala*	234/235bp	–	–	–	–	–	250/251bp	–	+	–
ser. *Grisea*	231–233bp	–	–	–	–	+	216–219bp	–	–	–
A. carpinifolium, sect. *Indivisa*	220/227bp	–	(+)	–	–	+	243bp	–	–	–
A. diabolicum, ser. *Lithocarpa*	232/233bp	–	–	–	–	+	235/236bp	–	–	–
sect. *Macrantha*	234–236bp	–	–	–	–	–	227–238bp	–	–	**I/D**
sect. *Rubra*	221/222bp	–	–	–	–	+	233–235bp	–	–	–
A. buergerianum, ser. *Trifida*	230bp	+	–	+	–	+	236bp	–	–	–

Legend: +, group-specific indel (**ID**) always realized at this position; (+), realized in some accessions of the group (series/section).
[1] *Macrantha* genotype 1 exhibits + 6 bp (**I**), *D. sinensis* + 5 bp, in comparison to the consensus nucleotide composition, *Macrantha* genotype 2 lacks 6 bp (D).

divergent sequences as detected for the ITS of *Acer* (Fig. 4) and the ETS of *Solanum* (see below) can be directly used to depict molecular phylogenies with standard computer-based analysis methods.

Compared to *Acer,* the overall level of ITS sequence variability in the genus *Zelkova* is markedly lower. Nevertheless, the data can be successfully used for taxonomic purposes. Prominent length polymorphism is here restricted to a single region within the ITS1, where ITS clones of two taxa (i.e. *Z. schneideriana* and some clones of *Z. serrata*) exhibit a 24-bp long deletion. Analogous to the ITS of *Acer*, the ITS of *Zelkova* is composed of regions, that preferentially accumulated and fixed point mutations, and of conserved regions, in which no changes were detected. However, in contrast to *Acer* markedly less mutations were fixed during the speciation of *Zelkova* and, therefore, the mutations are of high taxonomic weight. Remarkably, most of the point mutations in *Zelkova* have been detected only in some ITS clones, whereas the other clones are identical to the consensus of all *Zelkova* species (DENK and GRIMM 2005). Here, the taxonomic identity is *not* given by a unique ITS sequence; instead the *intraindividual variability* has to be taken into account. Such a less differentiated data set produces phylogenies, wherein sequences of different individuals and species, respectively, partly intermix with each other (Fig. 5*A*). Possible paralogous ITS sequences that interfere with the standard molecular analysis were found in *Z. serrata* and *Z. carpinifolia*. As a result, the ITS accessions of both taxa are either mixed with *Z. schneideriana* and *Z. abelicea/sicula* clones, respectively, or form distinct clades. To clarify this "ambiguous" placement, we defined and correlated series of mutations (oligonucleotide motif variants) found for the ITS of

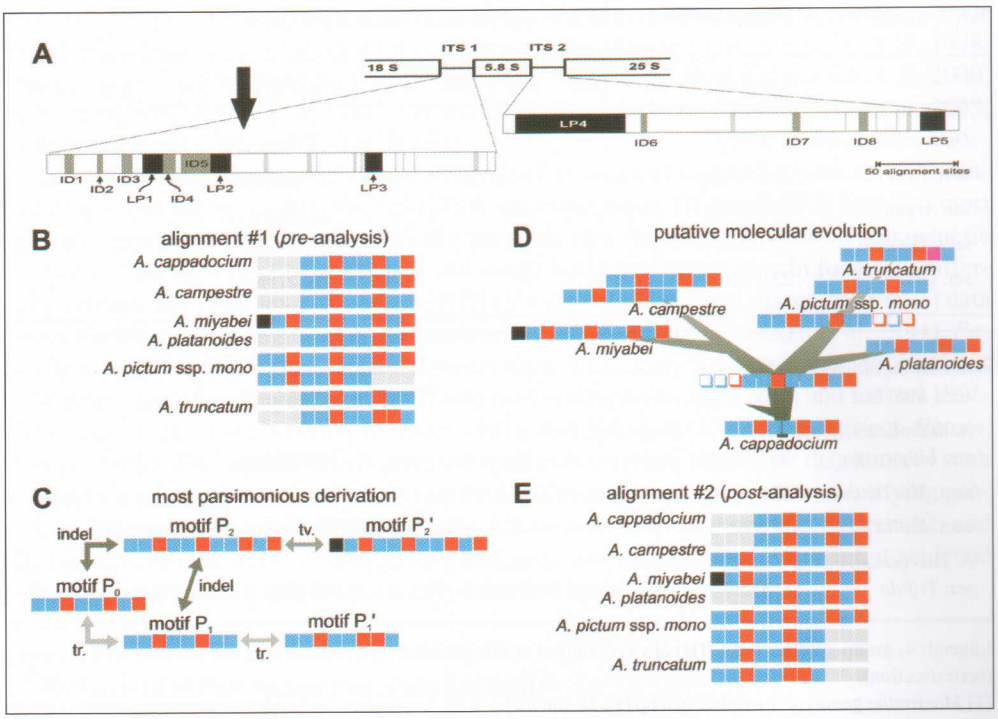

Fig. 3 Difficulties arising from a fixed alignment of indels, an example from the LP2 motif in *Acer* sect. *Platanoidea* (modified from GRIMM 2003). (*A*) Scheme of the ITS; position of the LP2 indicated by the black arrow. (*B*) Alignment with a minimum of alignment gaps meeting MP and ML criteria for individual nucleotide sites. (*C*) Derivation of complete motives meeting MP and ML criteria; 8-bp and 11-bp long motives differ in single mutations from each other. (*D*) Putative molecular evolution of LP2 according to the complete ITS data; the initial 3-bp insertion is later compensated by 3-bp deletions, and (*E*) the accordingly optimized alignment. Each box equals one nucleotide: black = guanine, blue = cytosine, red = thymine. Abbr.: ID, indel; LP, length polymorphic region; tr., transition; tv., transversion.

Zelkova with morphological data and the fossil record. This allowed us (*i*) to identify subrecent introgression between *Z. schneideriana* and *Z. serrata* and (*ii*) to depict ancestral ITS motives (example given in Fig. 6, for details refer to DENK and GRIMM 2005). Furthermore, we showed that the ancestral motives shared between the nowadays western Eurasian taxa *Z. abelicea*, *Z. carpinifolia*, *Z. sicula*, and the Japanese *Z. serrata* and the co-occurring taxon-specific derived motives can be correlated in detail to morphological characteristics found in western Eurasian macrofossil assemblages of *Zelkova*.

Intraindividual polymorphism of ITS is even more pronounced in species of the genus *Fagus* (DENK et al. 2002, GRIMM 2003, DENK et al. 2005). Similar to the situation in *Zelkova*, length polymorphism is restricted to two regions in ITS1 and ITS2, respectively. The occurrence of a 15-bp long indel within the ITS1 in approximately 50 % of all clones representing different individuals of *F. engleriana* and *F. japonica* (forming the subgen. *Engleriana*) is characteristic for these two species and, hence, represent a genetic synapomorphy of the subgenus *Engleriana*. This indel is always accompanied by several point mutations in other parts of the ITS, which are also restricted to this subgenus (GRIMM 2003). Again, the detected intraindi-

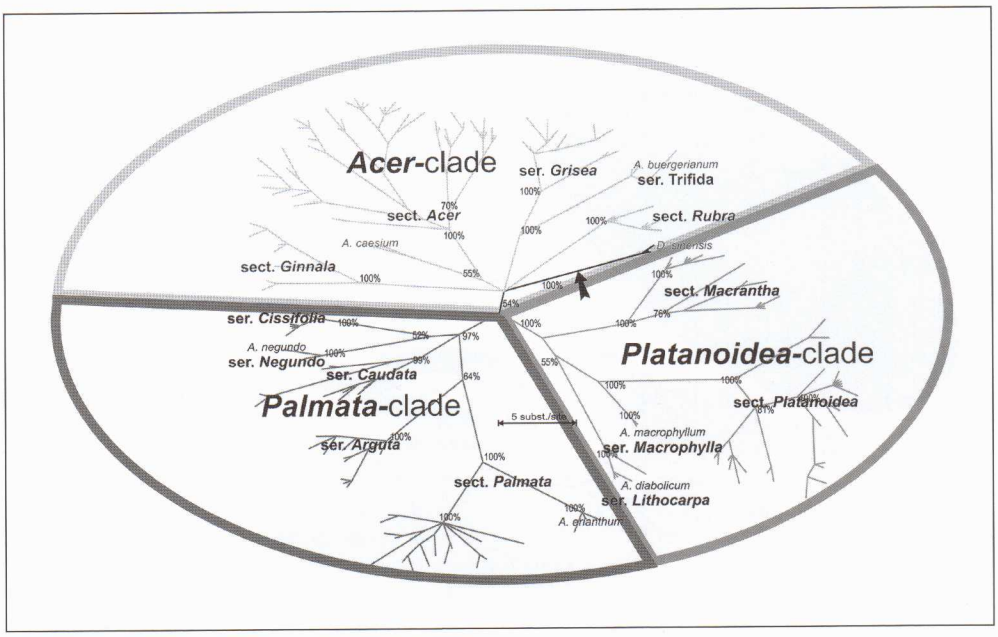

Fig. 4 Maximum likelihood (ML) molecular phylogeny of *Acer* as inferred from ITS sequence data based on a Bayesian inference (BI) analysis (GRIMM 2003). Numbers at nodes indicate posterior probabilities. Branches with < 50 % probability are collapsed.

vidual variability and differentiation are of high taxonomic importance. The encountered intraspecific and -individual divergence in genus *Fagus* is equal to or higher than the mean interspecific divergences. In the case of subgenus *Engleriana*, the detected intraindividual variability is even comparable to the levels of interserial and -sectional molecular differentiation in *Acer* (Tab. 2). On the other hand, nearly identical ITS sequences (≤ 1 % difference) can be found in all Eurasian taxa of the subgenus *Fagus*, although these taxa are morphologically well distinguishable. Consequently, the mode of ITS sequences variability strongly interferes with the phylogenetic reconstruction based on standard molecular analyses; Eurasian species of subgenus *Fagus* are not subdivided, only the positions of subgenus *Engleriana* and the North American *F. grandifolia* are clearly defined (Fig. 5*B*). In addition, individuals of the two endemic East Asian species *F. hayatae* and *F. longipetiolata* show a remarkable polymorphism with sequences that group together with the other Eurasian species, whereas the remaining sequences form a distinct clade. ITS variants related to intraindividual genetic polymorphism may indeed represent paralogs. The question is, do they really confuse phylogeny? A detailed reinvestigation and correlation of the ITS data to the fossil record and an accompanying morphological cladistic analysis (GRIMM 2003, DENK et al. 2005) revealed (*i*) an *ancient genetic polymorphism*, which is still present in modern individuals of *F. hayatae* and *F. longipetiolata* (East Asian relict species, characterized by the lack of derived morphological features), and (*ii*) *horizontal gene flow* between populations of *Fagus* throughout space and time. Thus, the at a first glance somewhat confusing ITS variability in *Fagus* (Fig. 5*B*) can be incorporated into a comprehensive framework taking into account the complex intrageneric differentiation history of the genus. Phases of unhindered horizontal gene flow resulted in intraindividual ge-

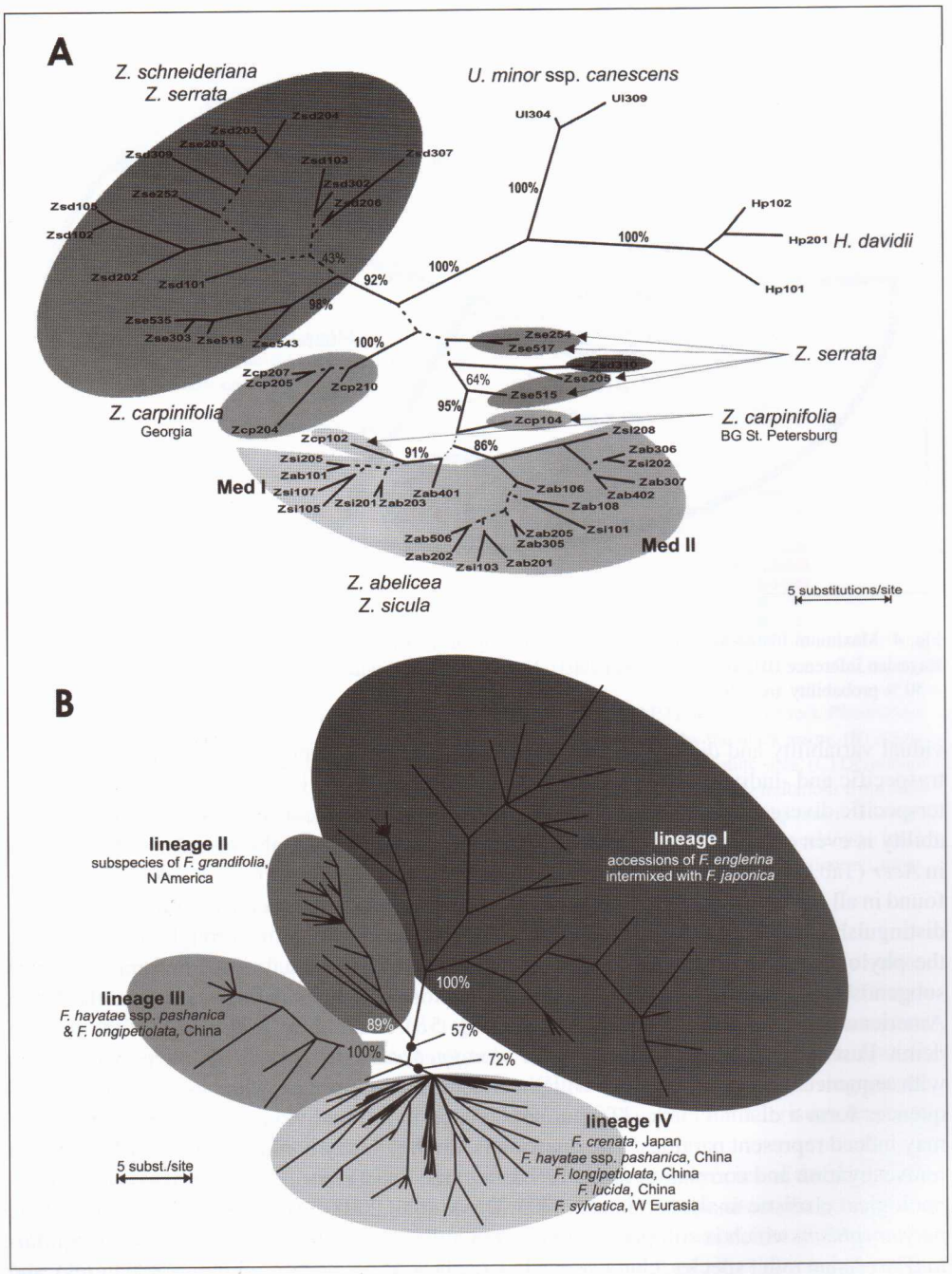

Fig. 5 ML *via* BI phylograms based on "ambiguous" ITS data. (*A*) *Zelkova* (DENK und GRIMM 2005). (*B*) *Fagus* (DENK et al. 2005). "Med I/II" labels refer to two ITS variants realized in individuals of *Z. abelicea* and *Z. sicula*. Numbers at nodes indicate posterior probabilities. Branches with < 50 % probability are collapsed. Abbr.: *H.*, *Hemiptelea*; *U.*, *Ulmus*.

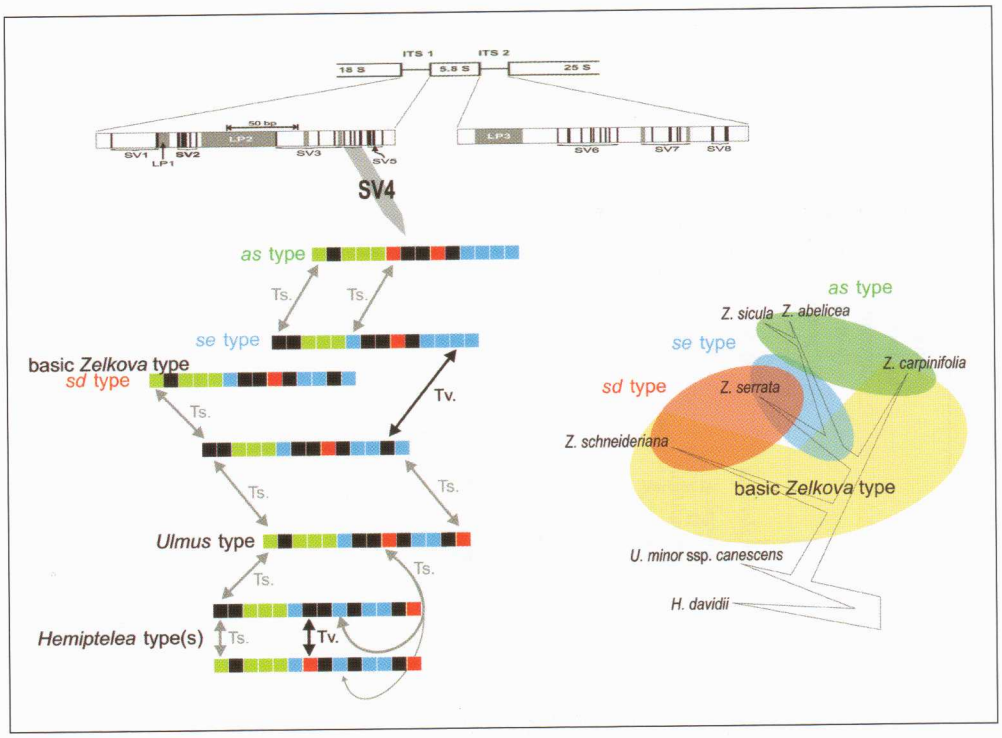

Fig. 6 Derivation of the oligonucleotide motif SV4 in *Zelkova* (modified after DENK and GRIMM 2005). Top, scheme of the ITS of *Zelkova*; bars indicate mutations, labels refer to distinguished oligonucleotide motives. Left, derivation of detected SV2 types serving as well MP as ML criteria. Right, evolution of the SV4 motif in course of species differentiation; the dendrogram is based on molecular (ITS), morphological, and fossil evidence; distribution of motives is indicated by coloured circles. Abbr.: LP, length polymorphism; SV, series of site variabilities.

netic variability reflected by possibly paralogous ITS sequences. This paralogy is not hindering the phylogenetic reconstruction; instead it is a clue to understand the lack of strict morphoclines in macrofossils and the restriction of certain morphological characteristics within recent *Fagus* species. Uniparentally inherited molecular markers such as the cpDNA *mat*K were not able to reproduce the actual reticulate history of this genus and consequently produced molecular phylogenies (STANFORD 1998, MANOS and STANFORD 2001) that contradicted both morphological (SHEN 1992, DENK 2003) and fossil data.

Comparison of the 5'ETS sequences allows successful taxonomy reconstruction in sect. *Petota* of the genus *Solanum*. In the 5'ETS two major structural regions can be distinguished: (*i*) a variable region (VR), demonstrating significant structural rearrangements, and (*ii*) a conservative region (CR), evolving mainly by base substitutions (Fig. 7). In VR, a conservative element (CE) was detected. Ribosomal DNA with one CE in the 5'ETS (ancestral structural variant A) was found in European out-group species (*S. dulcamara*, *S. nigrum*), in non-tuber-bearing South American species of ser. *Etuberosa*, and in Central American tuber-bearing diploids. Duplication of CE occurred in the 5'ETS of South American diploid species of sers. *Commersoniana* and *Circaeifolia* (variant B). All other South American diploids and Central American polyploids of superser. *Rotata* also possess two CE, and additionally two duplica-

Tab. 2 KIMURA-2-parameter distances calculated for *Acer* and *Fagus*

		mean intraspecific diversity[1]	interspecific diversity[1] (group mean) (other groups)	mean interspecific diversity[1]
	D. sinensis	0.001	n/c	0.001
***Acer*-clade**	sect. *Acer*	0.08	0.101	0.109
	sect. *Ginnala*	0.050	n/c	0.050
	ser. *Grisea*	0.006	0.019	0.026
	sect. *Indivisa*	0.005	n/c	0.005
	sect. *Rubra*	0.002	0.004	0.006
	ser. *Trifida*	0.008	n/c	0.008
***Palmata*-clade**	ser. *Arguta*	0.004	0.010	0.015
	ser. *Caudata*	0.049	0.021	0.070
	ser. *Cissifolia*	0.002	0.003	0.005
	ser. *Negundo*	0.010	n/c	0.010
	sect. *Palmata*	0.011	0.067	0.078
***Platanoidea*-clade**	ser. *Lithocarpa*	0.013	n/c	0.013
	ser. *Macrophylla*	0.000	n/c	0.000
	sect. *Macrantha*	0.014[2]	0.041[2]	0.06
	sect. *Platanoidea*	0.009	0.042	0.051

		mean intraspecific diversity[1] (within subgenus)	mean interspecific diversity[1] (other lineages)	mean interspecific diversity[1]
subgen. *Engleriana*	*F. engleriana*	0.091	0.086	0.097
	F. japonica	0.086		
subgen. *Fagus*	*F. grandifolia*	0.014	0.039	0.053
	F. hayatae	0.035	0.040	0.049
	F. longipetiolata	0.034	0.034	
	F. lucida	0.028[3]	0.032	0.056
	F. crenata	0.016	0.030	
	Georgian *F. sylvatica*	0.025	0.029	
	other *F. sylvatica*	0.014	0.028	

[1] Within group averages calculated *via* Kimura 2-parameter substitution model, gamma-distributed (MEGA 2.1®).

[2] In the case of taxa of sect. Macrantha the comparably low intraspecific diversity and the high interspecific diversity is due to the occurrence of two distinct genotypes. Species which can be assigned to the same genotype are basically identical (not differentiated on a molecular level).

[3] Mainly due to interlocality variability

tions around CE1 are present in VR (variant C). Finally, 18S–25S rDNA with six CE in the 5'ETS was found in tomato (*S. lycopersicum*) and related wild species (variant D). Amplification of CE occurred independently at least two times, in "potato"- and "tomato"-lineages (VOLKOV et al. 2003, KOMAROVA et al. 2004). Phylogenetic dendrograms derived from the comparison of the 752-bp portion of 5'ETS upstream of the 18S rRNA coding region (i.e. excluding all major structural rearrangements detected in VR) agree well with the biogeographical and morphological data and showed that the *structural variants A–D* found in the 5'ETS are characteristic for different clades and, therefore, can be used as taxonomically significant diagnostic "macrocharacters" in the sect. *Petota*.

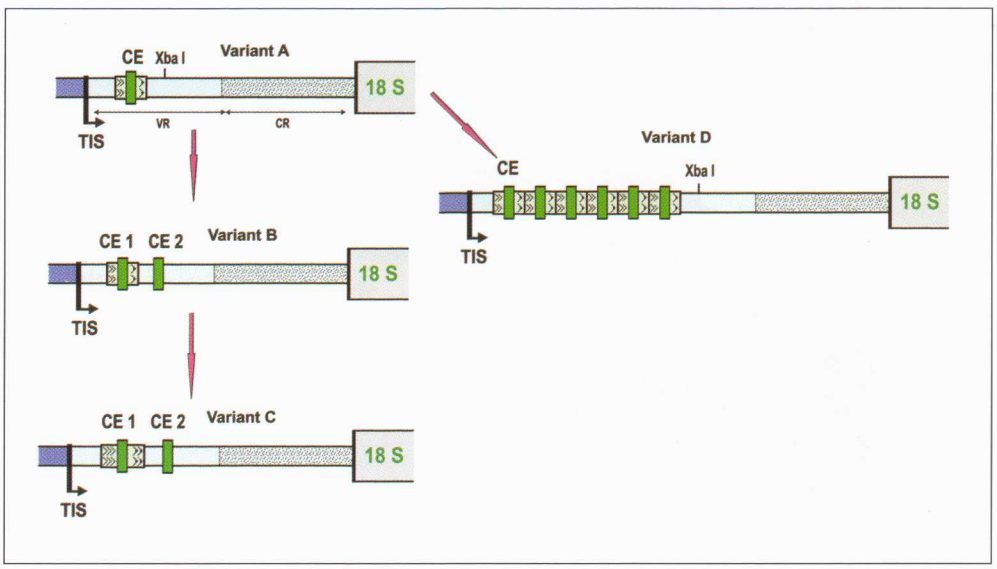

Fig. 7 Independent amplification of CE in the ETS of *Solanum* species. Abbr.: CE, conservative element; TIS, transcription initiation site.

2.2 *Intraspecific Differentiation*

Comparison of rDNA spacer regions is mainly used to describe interspecific differentiation. In combination with cloning, rDNA spacer sequence data can be useful for intraspecific studies. DENK et al. (2002) compared the patterns of ITS variability in individuals of *Fagus sylvatica* from western Eurasia with the lack of decent morphoclines in this taxon (DENK 1999a,b). Our results showed that the overlapping intraindividual variability is in accordance with the morphological variability detected in the respective populations. The combined data indicated a Transcaucasian origin of *F. sylvatica* and a subsequent loss of genetic variability towards central and western Europe, which could be further confirmed by additional molecular and morphological data from the remaining taxa of the genus (DENK 2003, GRIMM 2003, DENK et al. 2005). General observations for the genus *Fagus* clearly point towards a northern Pacific origin of the genus and to East Asia as centre of diversification and modern refuge

for primitive taxa. Transferring the assembled morphological, fossil, and ITS data into this evolutionary framework, it is possible to reconstruct the gene pools (i.e. ITS pools in a strict sense) of extinct ancestral taxa and their modern relatives (Fig. 8). In strong correlation to the fossil record, these gene pools reflect an *ancestral intraspecific differentiation* that gave rise to the extant and morphologically distinct species.

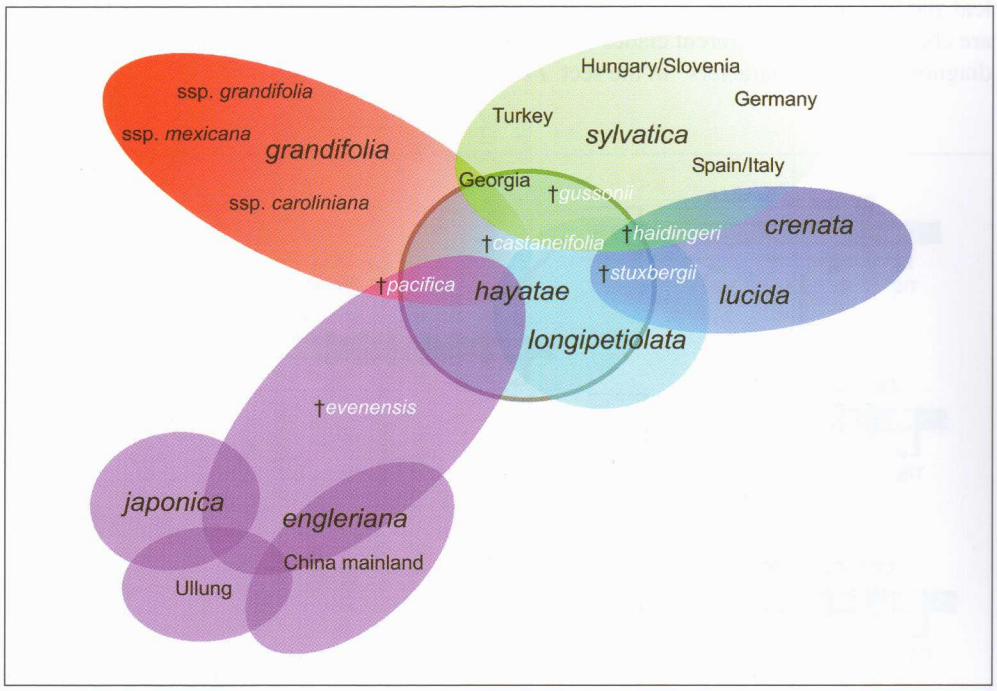

Fig. 8 Ancient and recent ITS pools in *Fagus*. Note that specific ITS molecular characteristics are shared by different taxa. Species' ITS boundaries are indicated by accordingly coloured circles. Assumed position of extinct predecessors (in white font) of modern species is given by the correlation with morphological and palaeobiogeographical evidence (cf. DENK 2003, GRIMM 2003, DENK et al. 2005).

Lathyrus pannonicus, a member of the basal subgenus *Orobus* sect. *Lathyrostylis*, shows a remarkable intraspecific differentiation in regard to phytosociological, morphological, and molecular genetical characteristics (ITS and ETS data) of analyzed populations throughout Europe (SCHLEE et al. 2003, M. SCHLEE, W. SAUER, and V. HEMLEBEN, in prep.). Apparently, an *ancient ITS polymorphism*, which occurred in species of Siberia (subsp. *multijugus*), has been retained in a derived form throughout all sampled populations and each distinguished subspecies of *L. pannonicus*. Molecular peculiarities found in the ITS polymorphs of each population further allow the reconstruction of putative migration pathways. Moreover, intraspecific differentiation related to an ecological shift from dry to wet ecotypes in the Pannonian area is also reflected in the ITS polymorphs. Populations and subspecies which are adapted to more wet ecological conditions share exclusive point mutations and intraindividual site variabilities, which are not found in the dry adapted counterparts. This is confirmed by ETS data

(Fig. 9) from selected individuals: A maximum likelihood analysis places an ecologically intermediate individual of *L. pannonicus* subsp. *varius* from Istria in between and basal to clades comprizing either the dry adapted subspecies *collinus* (including *suevicus*) or the more moist tolerant subspecies *pannonicus*. The latter one was established in course of the ecological shift in the Pannonian area and is suggested to be the ancestor of the further derived subspecies *longestipulatus*, which extended the species' range to the Atlantic and northern Spain, respectively (Fig. 9).

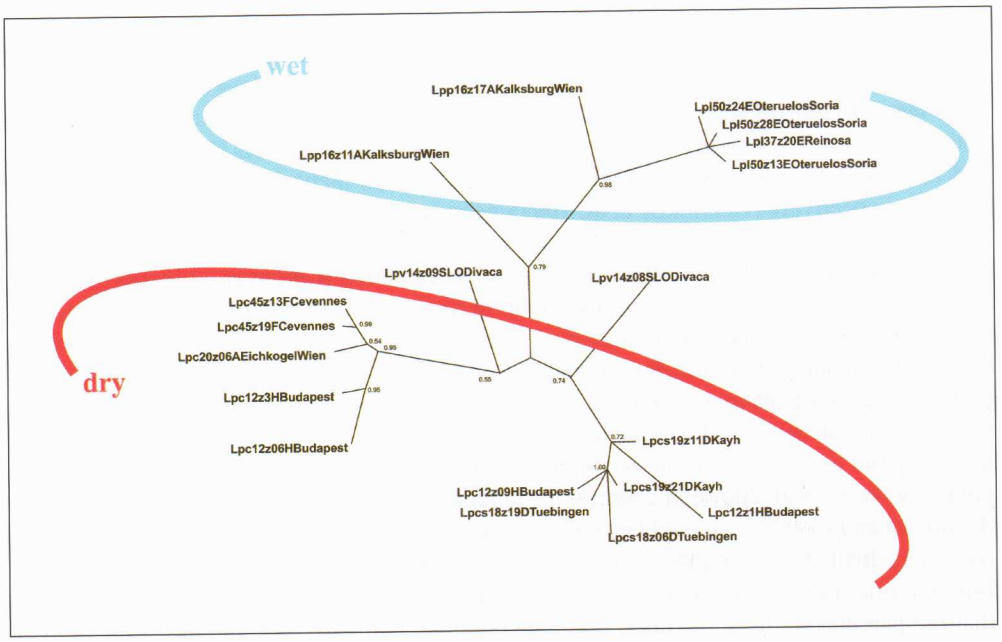

Fig. 9 ML phylogram based on the 5' ETS of individuals of *Lathyrus pannonicus* out of different relevées representing wet and dry stands (inferred via BI). Clone labels: Lpp, subsp. *pannonicus*, Pannonicum; Lpl, subsp. *longestipulatus*, Spain; Lpc(s), subsp. *collinus* including *suevicus*, Central Europe. Numbers at nodes are computed posterior probabilities.

2.3 Identification of Hybrids

Horizontal gene flow as a consequence of interspecific hybridization represents a common evolutionary feature of flowering plants (GRANT 1981). Accordingly, the pathways of speciation are often reticulate and not dichotomous. In the case of recombinational speciation (GRANT 1981) interspecific hybridization occurs at the same (mainly diploid) ploidy level, and a derived hybrid contains a "mix" of chromosomes of both parents (cf. Fig. 2). This could apply to the ITS situation detected for *Zelkova* and *Fagus* (see above), which are considered to comprise mainly diploid species (with the exception of the triploid *Z. sicula*). Respectively, in these genera, intragenomic heterogeneity of rDNA may indicate introgression of genetic information from a related species. Such situation was observed in our investigations of genus *Lathyrus* (M. SCHLEE, W. SAUER, and V. HEMLEBEN, in prep.). In *L. filiformis* most ITS clones

are nearly identical for individuals of different populations. However, in a Moroccan individual 5 clones showed ITS types which are either typical for *L. bauhinii* or for populations of *L. pannonicus* subsp. *longestipulatus* in northern Spain. Contamination (e.g. by alien pollen on the leaf surface) can be fairly ruled out, since (*i*) the studied individual was the only *Lathyrus* specimen in the used collection (Herbarium FB) and (*ii*) *L. bauhinii* and *L. pannonicus* are not found in Morocco. Hence, these data would indicate subrecent horizontal gene flow between *L. bauhinii*, *L. filiformis* and *L. pannonicus* with respect to the ancestors of the *L. filiformis* populations nowadays native to Morocco (data not shown).

In hybrids, comparison of rDNA spacer sequences allows identification of a species donating the rDNA-bearing chromosome. Generally, phylogeny obtained for molecular markers located at different chromosomes should have the same topology if no reticulate evolution takes place. In our studies of diploid *Solanum* species of sect. *Petota* the same phylogeny up to the level of series was produced, both for 5S rDNA spacer sequences (VOLKOV et al. 2001) and 5' ETS of the 18–25S rDNA (VOLKOV et al. 2003), which are located at chromosome I and II, respectively (GEBHARDT et al. 1991). In contrast, ETS sequences of members of the *S. brevicaule* complex (UGENT 1970) never grouped together, but were distributed between other species of ser. *Tuberosa*, illustrating reticulate evolution in the group and demonstrating that the *S. brevicaule* complex is composed of morphologically similar but not always genetically closely related species (VOLKOV et al. 2003).

Another mode of hybridization in plants represents allopolyploidization, which produces polyploid species containing complete chromosomal sets of diploid parents (cf. Fig. 2). In this case comparison of ITS/ETS sequences could be also applied to trace back the origin of the allopolyploid genome as it was shown in our investigations for several polyploid species of potato (VOLKOV et al. 2003). In contrast, the tetraploid *Acer pseudoplatanus* and the octaploid *A. rubrum* and *A. saccharinum* ITS sequences were found to be extremely homogeneous and obviously distinct from all ITS sequence variants found in related diploid *Acer* species; therefore, the origin of the *Acer* polyploids still remains unclear. A possible explanation could be that the parental forms are extinct. It is also possible that the rDNA variants of modern allopolyploids endured intensive concerted evolution, which resulted in homogenization/elimination of repeats of one of the diploid parents and even additional rearrangement of 18S–25S rDNA of the other parent. In this case, only one type of rDNA would be observed, which appears rather uniform. A similar phenomenon was described for the tetraploid tobacco (*Nicotiana tabacum*; VOLKOV et al. 1999). However, the maintenance of the 18S–25S rDNA from both parents was demonstrated in other species, e.g. in *Brassica* (CHEN and PIKAARD 1997) and *Solanum* (VOLKOV et al. 2003), suggesting either a recent origin of the respective allopolyploids and/or different rates of homogenization in different taxonomic groups.

3. Conclusion and Outlook

The rDNA spacer sequences obtained for several plant genera represent most valuable molecular data sets for all kinds of detailed studies addressing various aspects of intrageneric evolution and the origin and differentiation of species. Since PCR and sequencing is a rather easy and fast method, the determination of taxon-specific rDNA based characteristics provides valuable tools for taxonomic and systematic studies. This is especially of importance for the estimation of "number of species", which is the basis to count biodiversity. We have to keep

in mind that certain genera are morphologically variable and include numerous accepted species and infraspecific taxa, although they demonstrate relatively low sequence variability in the rDNA spacer. An interesting example here represents sect. *Petota* of genus *Solanum*, which according to HAWKES (1990) comprises 234 species. Many of these species are morphologically polymorph, but differ from each other only by overlapping character states. Respectively, they can be morphologically distinguished only by using numerical phenetic analysis (SPOONER et al. 2000, 2001). Our data show that such species are also very similar at the molecular level (VOLKOV et al. 2003), supporting the proposal to reduce significantly the number of species recognized in sect. *Petota* (e.g. SPOONER et al. 2000, 2001). Other groups (e.g. *Fagus*) exhibit a relatively high ITS sequence variability, but are morphologically conserved. The example of *Fagus* (cf. Tab. 2) illustrates that in this genus the protection of each Asian individual is more important than that of any European population in regard to preserve the full genetic potential of this genus. The first group (many-species, morphologically variable genera) is always in the focus of biodiversity and systematics, while the latter (few-species, morphologically conserved genera) is often ignored.

Our work with selected plant model groups demonstrates that molecular and morphological data should be combined to evaluate "taxa", in particular to define the boundaries of species. Moreover, the major evolutionary trends documented in the fossil record and (palaeo)biogeographical phenomena, like migration routes, (re-)occupation of niches and ecological shifting, should be considered (HEMLEBEN 1999). Especially, the detection of refugial habitats may help to identify the hotspots of evolution, where processes of speciation are evident and prototypes of modern taxa can be found as we have already shown in case of *Fagus* and *Zelkova.*

To reconstruct reticulate intrageneric evolution, we need information about all presumptive paternal lineages. Accordingly, ITS/ETS sequence comparison allows identification of the 18S–25S rDNA donors, both in diploid hybrids and in allopolyploids. Sequence polymorphism among members of a population may indicate introgression of genetic material of related species. In many cases, both parental rDNA variants could be identified, especially if "young" hybrids/allopolyploids were studied. In addition, rearrangement of genetic material was demonstrated in "old" natural allopolyploids as a result of concerted evolution, which led to homogenization/conversion of parental rDNA. Hence, introgression and allopolyploidization, recent and ancient, respectively, may be traced back on the molecular level using sequence comparison of rDNA spacer sequences. It is obvious that the reconstruction of complex pathways of plant speciation is dependent on as much as possible data reflecting not only interspecific differentiation, but also the intraspecific and intraindividual variability, which can be documented by cloning and sequencing of a large number of clones.

Although the generation of a single mutation is a statistical phenomenon, its fixation at a particular position in any gene region is a function of evolution. This is in particular true, if not a single nucleotide with its limited number of character states (i.e. A, C, G, and T) is addressed but complexes of evolutionary linked mutations. The analysis of distribution of original and derived ITS sequence variants among *Zelkova* species and the subrepeat organization in the ETS of *Solanum* demonstrate that well-sustained hypotheses about phylogeny and molecular evolution can be established without relying solely on computer algorithms and standard molecular base-per-base analysis methods. Instead, molecular systematics should in addition focus on "molecular macrocharacters" like taxon-specific sequence variants, structural rearrangements, repeat patterns, etc., in comparison to complex morphological characters

such as leaf anatomy and flower morphology, which are commonly used for taxonomic and systematic purposes. As demonstrated for *Fagus* and *Zelkova* those basically phenomenological analyses outperform the capabilities of computer-based, mainly statistical methods and were found to correlate perfectly with and to round off the morphological and fossil evidences. Crucial for future studies on the pathways of molecular evolution is a better understanding of the molecular mechanisms that determine mutation patterns. An intriguing question is why mutations do appear preferentially in particular positions, e.g. in ITS, in different taxonomic groups. Especially in the case of intrageneric evolution, the major task for molecular systematics should not be simply the accumulation and evaluation of sequence data, but the tracking of internal and external factors which determine speciation and molecular variability. The identification and subsequent analysis of molecular macrocharacters may be one clue to this problem.

References

Álvarez, I., and Wendel, J. F.: Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phyl. Evol. *29*, 417–434 (2003)

Asmussen, C. B., and Liston, A.: Chloroplast DNA characters, phylogeny, and classification of *Lathyrus* (Fabaceae). Amer. J. Bot. *85*, 387–401 (1998)

Bässler, M.: Revision von *Lathyrus* L. sect. *Lathyrostylis* (Griseb.) Bässler (Fabaceae). Feddes Rep. *92*, 179–254 (1981)

Bailey, C. D., Carr, T. G., Harris, S. A., and Hughes, C. E.: Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol. Phyl. Evol. *29*, 435–455 (2003)

Balthazar, M. von, Endress, P. K., and Qiu, Y.-L.: Phylogenetic relationships in Buxaceae based on nuclear internal transcribed spacers and plastid *ndh*F sequences. Int. J. Plant Sci. *161*, 785–792 (2000)

Chen, Z. J., and Pikaard, C. S.: Transcriptional analysis of nucleolar dominance in polyploid plants: Biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. Proc. Natl. Acad. Sci. USA *94*, 3442–3447 (1997)

Denduangboripant, J., and Cronk, Q. C. B.: Evolution and alignment of the hypervariable arm 1 of *Aeschynanthus* (Gesneriaceae) ITS 2 nuclear ribosomal DNA. Mol. Phyl. Evol. *20*, 163–172 (2001)

Denk, T.: The taxonomy of *Fagus* in western Eurasia, 1: *Fagus sylvatica* subsp. *orientalis* (= *F. orientalis*). Feddes Rep. *110*, 177–200 (1999a)

Denk, T.: The taxonomy of *Fagus* in western Eurasia, 2: *Fagus sylvatica* subsp. *Sylvatica*. Feddes Rep. *110*, 381–412 (1999b)

Denk, T.: Phylogeny of *Fagus* L. (Fagaceae) based on morphological data. Plant Syst. Evol. *240*, 55–81 (2003)

Denk, T., and Grimm, G. W.: Phylogeny and biogeography of *Zelkova* (Ulmaceae s. str.) as inferred from leaf morphology, ITS sequence data and the fossil record. Bot. J. Linn. Soc. *147*, 129–157 (2005)

Denk, T., Grimm, G. W., and Hemleben, V.: Patterns of molecular and morphological differentiation in *Fagus*: implications for phylogeny. Amer. J. Bot. *92*, 1006–1016 (2005)

Denk, T., Grimm, G., Stögerer, K., Langer, M., and Hemleben, V.: The evolutionary history of *Fagus* in western Eurasia: Evidence from genes, morphology and the fossil record. Plant Syst. Evol. *232*, 213–236 (2002)

Fransz, P., Armstrong, S., Alonso-Blanco, C., Fischer, T. C., and Torres-Ruiz, R. A.: Cytogenetics for the model system *Arabidopsis thaliana*. Plant J. *13*, 867–876 (1998)

Gebhardt, C., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschabel, U., Kaufmann, H., Thomson, R. D., Bonierbale, M. W., Ganal, M. W., Tanksley, S. D., and Salamini, F.: RFLP maps of potato and their alignment with the homologous tomato genome. Theor. Appl. Genet. *83*, 49–57 (1991)

Grant, V.: Plant Speciation. 2nd Ed. New York: Columbia Univ. Press 1981

Grebenstein, B., Röser, M., Sauer, W., and Hemleben, V.: Molecular phylogenetic relationships in Aveneae (Poaceae) species and other grasses as inferred from ITS1 and ITS2 sequences. Plant Syst. Evol. *213*, 233–250 (1998)

Grimm, G. W.: Tracing the mode and speed of intrageneric evolution: A phylogenetic case study of genus *Acer* L. and genus *Fagus* L. Ph.D. thesis, Eberhard-Karls-Universität, Tübingen (2003)

Hawkes, J. G.: The Potato Evolution, Biodiversity and Genetic Resources. Washington, DC: Smithsonian Institution Press 1990

Hemleben, V.: Die Bedeutung der Molekularbiologie für die moderne Evolutionsforschung. In: Junker, T., und Engels, E.-M. (Eds.): Die Entstehung der synthetischen Theorie. Beiträge zur Geschichte der Evolutionsbiologie in Deutschland 1930–1950. Band 2, S. 211–221. Berlin: VWB, Verlag für Wissenschaft und Bildung 1999

HEMLEBEN, V., GANAL, M., GERSTNER, J., SCHIEBEL, K., and TORRES, R. A.: Organization and length heterogeneity of plant ribosomal RNA genes. In: KAHL, G. (Ed.): Architecture of Eukaryotic Genes; pp. 371–383. Weinheim: VCH Verlagsgesellschaft mbH 1988

HEMLEBEN, V., ZANKE, C., PANCHUK, I. I., and VOLKOV, R. A.: Repetitive elements as molecular markers in potato breeding. Beiträge zur Züchtungsforschung *4*, 61–66 (1998)

JOBST, J., KING, K., and HEMLEBEN, V.: Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of Cucurbitaceae. Mol. Phyl. Evol. *9*, 204–219 (1998)

KOMAROVA, N. Y., GRABE, T., HUIGEN, D. J., HEMLEBEN, V., and VOLKOV, R. A.: Organization, differential expression and methylation of rDNA in artificial *Solanum* allopolyploids. Plant Mol. Biol. *56*, 439–463 (2004)

MANOS, P. S., and STANFORD, A. M.: The historical biogeography of Fagaceae: Tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. Int. J. Plant Sci. *162*, S77–S93 (2001)

MUIR, G., FLEMING, C. C., and SCHLÖTTERER, C.: Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L.: Mol. Biol. Evol. *18*, 112–119 (2001)

LIM, K. Y., KOVARIK, A., MATYASEK, R., BEZDEK, M., LICHTENSTEIN, C. P., and LEITCH, A. R.: Gene conversion of ribosomal DNA in *Nicotiana tabacum* is associated with undermethylated, decondensed and probably active gene units. Chromosoma *109*, 161–172 (2000)

SCHLEE, M., SAUER, W., und HEMLEBEN, V.: Molekulare und pflanzensoziologische Analyse von pontisch-pannonischen Reliktarten aus wärmebegünstigten Saum-Gesellschaften Süddeutschlands und benachbarter Gebiete. Nova Acta Leopoldina NF Bd. *87*, Nr. 328, 379–387 (2003)

SHEN, C.-F.: A Monograph of the Genus *Fagus* Tournai ex. L. (Fagaceae). Ph.D. thesis, The City University of New York, New York (1992)

SPOONER, D. M., VAN DEN BERG, R. G., and MILLER, J. T.: Species and series boundaries of *Solanum* series *Longipedicellata* (Solanaceae) and phenetically similar species in ser. *Demissa* and ser. *Tuberosa:* implications for a practical taxonomy of Section *Petota.* Amer. J. Bot. *88*, 113–130 (2000)

SPOONER, D. M., VAN DEN BERG, R. G., REVERA-PENA, A., VELGUTH, P., DEL RIO, A., and SALAS-LOPEZ, A.: Taxonomy of Mexican and Central American members of *Solanum* series *Conicibaccata* (sect. *Petota*). Syst. Bot. *26*, 743–756 (2001)

STANFORD, A. M.: The biogeography and phylogeny of *Castanea*, *Fagus*, and *Juglans* based on *mat*K and ITS sequence data. Ph.D. thesis, Chapel Hill (1998).

STEANE, D. A., SCOTLAND, R. W., MABBERLEY, D. J., and OLMSTEAD, R. G.: Molecular systematics of *Clerodendron* (Lamiceae): ITS sequences and total evidence. Amer. J. Bot. *86*, 98–107 (1999)

UCHIMIYA, H., OHGAWARA, T., KATO, H., AKIYAMA, T., HARADA, H., and SUGIURA, M.: Detection of two different nuclear genomes in parasexual hybrids by ribosomal RNA genes analysis. Theor. Appl. Genet. *64*, 117–118 (1983)

UGENT, D.: The potato. Science *170*, 1161–1166 (1970)

VAN GELDEREN, D. M., JONG, P. C. DE, and OTERDOOM, H. J.: Maples of the World. Portland, Oregon: Timber Press 1994

VOLKOV, R. A., BORISJUK, N. V., PANCHUK, I. I., SCHWEIZER, D., and HEMLEBEN, V.: Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*. Mol. Biol. Evol. *16*, 311–320 (1999)

VOLKOV, R. A., ZANKE, C., PANCHUK, I. I., and HEMLEBEN, V.: Molecular evolution of 5S rDNA of *Solanum* species (sect. *Petota*): application for molecular phylogeny and breeding. Theor. Appl. Genet. *103*, 1273–1282 (2001)

VOLKOV, R. A., KOMAROVA, N. Y., PANCHUK, I. I., and HEMLEBEN, V.: Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. *Petota* (genus *Solanum*). Mol. Phyl. Evol. *29*, 187–202 (2003)

VOLKOV, R. A., MEDINA, F. J., ZENTGRAF, U., and HEMLEBEN, V.: Molecular Cell Biology: Organization and molecular evolution of rDNA, nucleolar dominance and nucleolus structure. In: ESSER, K., LÜTTGE, U., BEYSCHLAG, W., and MURATA J. (Eds.): Progress in Botany. Vol. 65, pp. 106–146. Berlin, Heidelberg, New York: Springer 2004

WISSEMANN, V.: Genetic constitution of *Rosa* Sect. *Caninae* (*R. canina, R. jundzillii*) and Sect. *Gallicanae* (*R. gallica*). J. Appl. Bot. – Angewandte Botanik *73*, 191–196 (1999)

WISSEMANN, V.: Molecular evidence for allopolyploid origin of the *Rosa canina*-complex (Rosaceae, Rosoideae). J. Appl. Bot. – Angewandte Botanik *76*, 176 (2002)

Prof. Dr. Vera HEMLEBEN
Lehrstuhl für Allgemeine Genetik
Universität Tübingen
Zentrum für Molekular Biologie der Pflanzen (ZMBP)
Auf der Morgenstelle 28
72076 Tübingen
Germany

Tel.: +49 7071 2973554
Fax: +49 7071 295042
E-Mail: guido.grimm@uni.tuebingen.de