Heredity (2013) 110, 138–151
© 2013 Macmillan Publishers Limited All rights reserved 0018-067X/13
www.nature.com/hdy

Heredity

ORIGINAL ARTICLE

Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (*Hieracium*, Asteraceae)

K Krak[1], P Caklová[1], J Chrtek[1,2] and J Fehrer[1]

Phylogeny reconstruction based on multiple unlinked markers is often hampered by incongruent gene trees, especially in closely related species complexes with high degrees of hybridization and polyploidy. To investigate the particular strengths and limitations of chloroplast DNA (cpDNA), low-copy nuclear and multicopy nuclear markers for elucidating the evolutionary history of such groups, we focus on *Hieracium* s.str., a predominantly apomictic genus combining the above-mentioned features. Sequences of the *trn*V-*ndh*C and *trn*T-*trn*L intergenic spacers were combined for phylogenetic analyses of cpDNA. Part of the highly variable gene for squalene synthase (*sqs*) was applied as a low-copy nuclear marker. Both gene trees were compared with previous results based on the multicopy external transcribed spacer (ETS) of the nuclear ribosomal DNA. The power of the different markers to detect hybridization varied, but they largely agreed on particular hybrid and allopolyploid origins. The same crown groups of species were recognizable in each dataset, but basal relationships were strongly incongruent among cpDNA, *sqs* and ETS trees. The ETS tree was considered as the best approximation of the species tree. Both cpDNA and *sqs* trees showed basal polytomies as well as merging or splitting of species groups of non-hybrid taxa. These patterns can be best explained by a rapid diversification of the genus with ancestral polymorphism and incomplete lineage sorting. A hypothetical scenario of *Hieracium* speciation based on all available (including non-molecular) evidence is depicted. Incorporation of seemingly contradictory information helped to better understand species origins and evolutionary patterns in this notoriously difficult agamic complex.

Heredity (2013) **110,** 138–151; doi:10.1038/hdy.2012.100; published online 5 December 2012

Keywords: *Hieracium*; hybridization; incomplete lineage sorting; low-copy nuclear markers; gene tree incongruence; apomixis

INTRODUCTION

Until about a decade ago, phylogenetic studies in plants were mainly based on chloroplast DNA (cpDNA) alone or in combination with nuclear ribosomal DNA (nrDNA). If insufficient resolution was obtained, the general expectation was that species relationships could be clarified by employing more markers. Since then, a large number of studies using many unlinked markers have been published, but gene tree incongruence has been found to be a common phenomenon (Degnan and Rosenberg, 2009), which seems to intensify as the number of markers increases (Rokas and Carroll, 2006). One cause of such incongruence is hybridization (for example, Linder and Rieseberg, 2004; Fehrer *et al.*, 2007), which is one of the major processes in the evolution of vascular plants (Arnold, 1997) and is increasingly being recognized as important for animals as well. Allopolyploidy, which is merging genomes from different species, is frequently involved in hybrid speciation (Rieseberg and Willis, 2007) and can further complicate phylogenetic inference. Thus, the reconstruction of species relationships in highly reticulate, predominantly polyploid groups remains a substantial challenge. While adding more markers may not necessarily result in well-resolved species trees in such cases, different types of markers and their comparison can provide valuable insights into the evolutionary history of such species complexes and may also contribute to our understanding of the evolutionary dynamics of the markers themselves.

Multicopy nuclear genes (for example, nrDNA regions) are subject to concerted evolution, a process that homogenizes variation among the repeats of a gene family. This results in a single sequence per individual and thereby facilitates phylogeny reconstruction. If concerted evolution is unfinished or fails completely, this biparentally inherited marker can be very suitable for the detection of hybrid or allopolyploid origins. However, the unpredictable nature of concerted evolution may also result in chimeric sequences or eliminate the ribotypes of one parent. Furthermore, locus loss or duplication can lead to erroneous conclusions. The benefits and pitfalls of this type of marker, which has been applied in numerous case studies, have been reviewed by Álvarez and Wendel (2003) and Feliner and Rosselló (2007). CpDNA is still the most widely applied type of marker in plant phylogenetics (for a review, see Olmstead and Palmer, 1994). It is usually uniparentally (most often maternally) inherited, haploid and non-recombinant, which makes it easily applicable. For closely related species, the level of variation can be too low to obtain sufficient resolution. In case of hybrid origin, only one of the parents can be identified, that is, hybrids cannot be recognized based on cpDNA alone. Low-copy nuclear markers (LCNMs) currently have an

[1]Institute of Botany, Academy of Sciences of the Czech Republic, Průhonice, Czech Republic and [2]Department of Botany, Faculty of Science, Charles University in Prague, Prague, Czech Republic
Correspondence: Dr J Fehrer, Institute of Botany, Academy of Sciences of the Czech Republic, Zámek 1, Průhonice 25243, Czech Republic.
E-mail: fehrer@ibot.cas.cz

Received 12 June 2012; revised 23 October 2012; accepted 23 October 2012; published online 5 December 2012

increasingly important role in phylogenetic inference in plants. Due to the large number of coding genes, they represent an almost unlimited source of markers. LCNMs are biparentally inherited and less susceptible to concerted evolution than multicopy markers. They tend to be highly variable and have often been enlisted if nrDNA and cpDNA gave incongruent results or poorly resolved relationships. They have also been used successfully to reconstruct allopolyploid origins in plants (for example, Brysting *et al.*, 2007). Drawbacks that may affect LCNMs more severely than other types of markers include paralogy (gene duplication reflecting the history of the gene rather than its inheritance from the most recent common ancestor) and population genetic processes like incomplete lineage sorting (ILS) (the stochastic sorting of alleles following divergence from a polymorphic ancestor), genetic drift or natural selection (that may both result in the loss of alleles). In all these cases, the gene tree does not reflect the species tree. As LCNMs have not yet been used as frequently for phylogenetic and hybrid inference as other types of markers, their evolutionary dynamics are still less well understood and have to be assessed for each particular system to which they are applied. Their potential and drawbacks have been reviewed by Sang (2002), Linder and Rieseberg (2004), and Small *et al.* (2004).

Hieracium s.str. represents a particularly challenging system for phylogeny reconstruction, because it combines abundant hybridization and polyploidization with apomixis, which is asexual production of maternal progeny through seeds (Asker and Jerling, 1992). The combination of these features resulted in huge morphological variation and has led to a major disagreement regarding the number of taxa (500–5000 species depending on taxonomic concept) and their delimitation (reviewed by Stace, 1998). The most comprehensive taxonomic study was published by Zahn (1921–1923). He distinguished 'basic species', defined as being morphologically unique, from 'intermediate species', which combine morphological traits of two or more basic species and are thought to have hybrid origin. The genus consists of perennial herbs with main centers of diversity in the Alps, Pyrenees, Carpathians and Balkan mountains. The genus has experienced extensive hybridization in the past (Fehrer *et al.*, 2009, and references therein). In contrast, recent natural hybridization is very rare (Mráz *et al.*, 2005; Chrtek *et al.*, 2006); natural as well as experimental hybrids (Mráz and Paule, 2006) are either completely female sterile or produce only a few seeds. The basic chromosome number of *Hieracium* (and related genera) is $x = 9$. The same species can comprise different cytotypes that are usually indistinguishable morphologically. Only few diploids occur, and are almost exclusively confined to unglaciated refugia (Merxmüller, 1975). Most taxa are triploid, less often tetraploid and very rarely pentaploid (Schuhwerk, 1996). Polyploids appear to be obligatory apomicts (*Antennaria*-type diplospory), whereas diploids are sexual and self-incompatible (Chrtek *et al.*, 2009).

Recently, Fehrer *et al.* (2009) undertook to investigate the evolutionary history of the genus using the external transcribed spacer (ETS) region as a multicopy nuclear marker and the *trn*T-*trn*L intergenic spacer as a cpDNA marker. An almost complete set of 'basic species', which were considered as the main evolutionary units, was analyzed, including accessions of most known diploid cytotypes. The expectation was that if some of the apomicts had cryptic allopolyploid origins, concerted evolution of the nrDNA marker should be slowed down or even absent (Campbell *et al.*, 1997) and thereby allow the inference of the diploid ancestors (if they still existed), whereas diploid sexuals and autopolyploid apomicts should have homogeneous sequences. Surprisingly, not only some polyploids, but also several diploids showed character additivity of different ETS ribotypes. The unexpected lack of concerted evolution in sexuals also allowed the investigation of hybrid origins of diploids in this case. Individuals with only one ribotype were considered as nonhybrid diploids or autopolyploids; these were used for phylogenetic analyses. The resulting ETS tree showed a deep split of the genus into two major lineages. This division had never been suggested in taxonomic treatments; morphological diversity (within both groups and in general) is large and mostly inconclusive for classifications above the species level. Both major clades comprised central European as well as widespread species, but the distribution of most endemic taxa corresponded to either eastern or western European glacial refugia. We therefore referred to them as 'eastern' and 'western' clade or origin. The two clades were corroborated by significant genome size differences; interclade hybrids showed intermediate genome sizes (Chrtek *et al.*, 2009). Several subclades and species groups could be identified in the ETS tree; they corresponded to geographic distribution or ecological preferences, to a lower degree also to the morphology of the respective taxa. Sequence variation within the major clades as well as within the subclades was very low, indicating recent speciation. Consequently, the parentage of most hybrid individuals could only be attributed to subclades, not to particular parental species. The cpDNA marker showed very low overall variation, which is in keeping with recent speciation, but the species groups identified by ETS generally corresponded to particular haplotypes. Most hybrid accessions showed the haplotype of one of the parental lineages inferred by ETS and reflected the maternal parent; maternal inheritance of cpDNA was ascertained for *Hieracium* by Mráz *et al.* (2005). Thus, the level of variation of both markers was too low to resolve the relationships within and among the species groups. The internal transcribed spacer provided no resolution at all (Krak *et al.*, 2012) so that the information content for the multicopy nrDNA marker cannot be enhanced, but additional cpDNA data and a highly variable LCNM may resolve close species relationships.

In order to better understand the potential of different kinds of markers to elucidate speciation processes and evolutionary history in spite of abundant ancient hybridization, we address the following questions: (1) Do additional markers provide sufficient resolution to infer close species relationships and thereby allow identification of particular parental species in case of hybrid origin? (2) How does a LCNM perform in the detection of hybrid or allopolyploid origins compared with the other markers? (3) Do the three different kinds of markers suggest the same phylogenetic relationships of nonhybrids? (4) Can patterns caused by hybrid origin be distinguished from other reasons for gene tree incongruence? To address these objectives, a second intergenic spacer (*trn*V-*ndh*C) was added to improve the resolution of the cpDNA dataset. Recently, Krak *et al.* (2012) have developed three novel LCNMs for *Hieracium* and related genera; the most variable of these markers, part of the gene for squalene synthase (*sqs*), was used here. LCNMs have so far been applied in only two agamic complexes: cheilanthoid ferns (Grusz *et al.*, 2009) and hawthorn (Lo *et al.*, 2010), and on a small number of taxa only. Thus, to our knowledge, our study is the first attempt to investigate LCNMs in a large agamic complex in which about half of the analyzed individuals had inferred hybrid origins.

MATERIALS AND METHODS

Plant material

An almost complete set of 'basic species' (Zahn, 1921–1923) was investigated. The species were represented by 1–3 samples each; if different cytotypes occurred within a species, diploid populations were included whenever possible. Altogether, 61 *Hieracium* individuals from 47 species included in the study of Fehrer *et al.* (2009) were analyzed. *Hispidella* and diploid *Pilosella* species were chosen as outgroups based on Fehrer *et al.* (2007) and Krak *et al.* (2012). DNA extracts from the previous studies were used. For details on plant origin, see Supplementary Table S1. Ploidy, species group, and hybrid status according to Fehrer *et al.* (2009) are shown in Table 1.

Molecular procedures

The chloroplast *trn*V-*ndh*C intergenic spacer was amplified with the primers *trn*V-a and *ndh*C-a (Figure 1a), modified after Shaw *et al.* (2007). PCRs were performed in 25 μl reactions containing 2.0 mM $MgCl_2$, 0.2 mM of each dNTP, 0.5 mM of each primer, 0.5 unit of Taq DNA polymerase (Fermentas, Ontario, Canada), 1 × Taq buffer with KCl (Fermentas) and a few nanograms of genomic DNA. An initial denaturation step at 94 °C for 3 min was followed by 40 cycles of denaturation (94 °C for 30 s), annealing (52 °C for 30 s), extension (72 °C for 2 min) and a final extension at 72 °C for 10 min. PCR products were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and sequenced at GATC Biotech (Konstanz, Germany). The PCR primers as well as several internal primers (Figure 1a) were used for sequencing.

A region of the *sqs* gene spanning exon 4 through intron 8 (Figure 1b) was amplified by semi-nested PCR according to Krak *et al.* (2012), but PCRs were done in triplicate to reduce PCR drift (Wagner *et al.*, 1994) and to achieve representative proportions of alleles for cloning. PCR products were purified as above and directly sequenced with the PCR primers and internal sequencing primers (Figure 1b). Only four *Hieracium* individuals had uniform direct sequences; one sample showed a single polymorphism. All others showed several additive peaks and/or shifts caused by indels and were cloned. Cloning and subsequent procedures followed Fehrer *et al.* (2009). The same primers (Figure 1b) were used for the sequencing of clones. Depending on variation and number of alleles, 3–18 clones per accession (9 on average) were sequenced.

Data analyses

Sequences were proofread with Chromas Lite 2.1 (Technelysium Pty Ltd, Brisbane, QLD, Australia), aligned and edited manually in BioEdit 7.0.4.1. (Hall, 1999) using the IUPAC ambiguity codes to represent polymorphisms in *sqs* direct sequences. *Trn*V-*ndh*C (obtained in this study) and *trn*T-*trn*L sequences from Fehrer *et al.* (2009) were concatenated; the combined dataset was used for phylogenetic analyses. Indels were coded as single characters as described in Fehrer *et al.* (2007); length variation in poly-A regions was omitted. The alignment containing the indel coding is provided in Supplementary File 1. Phylogenetic analyses were conducted with maximum parsimony (MP) using PAUP* 4.0b10 (Swofford, 2002) and Bayesian inference using MrBayes (Ronquist and Huelsenbeck, 2003). For MP analysis, heuristic searches with 10 random sequence addition replicates, saving no more than 100 trees with length $\geqslant$ 1 per replicate, and TBR branch swapping were performed. Bootstrapping with 1000 replicates was performed with the same settings. For Bayesian analysis, the model of molecular evolution best fitting the data was determined with MrModeltest V2 (Nylander, 2004). A F81 + G model was identified as the best fitting using hierarchical Likelihood Ratio Tests; the basic parameters (one substitution rate and gamma distribution) were used as priors along with the default settings. Chains were computed for 5 million generations, sampling every 1000th tree. All statistical parameters indicated that convergence was reached by this time. The first 1250 trees per run were discarded as burn-in, and the remaining 7502 trees were summarized.

For *sqs*, at first, cloned and direct sequences of the individuals were aligned. Direct sequences are more reliable for identifying the entire allelic variation and to distinguish true variation from PCR artifacts than consensus sequences, especially if an individual contains more than two alleles or if truely distinct alleles are highly similar. Therefore, complete or partial direct sequence reads were used to ensure that all polymorphisms were represented by clones and for the correction of polymerase errors. An example how this can be done even in the presence of indels is shown in Supplementary Figure S1 (see also Kaplan and Fehrer, 2007). Based on these alignments, allelic variation within the sample was examined and recombinant sequences (generated during PCR) were identified by eye. Apparently non-recombinant sequences representing the allelic variation for each accession were included in the total alignment (one representative clone per allele, corrected for polymerase errors). Functionality was assessed by translation of exons in BioEdit (Hall, 1999); variation in exons was very low. Indels occurred only in introns; they were coded according to the simple gap coding method (Simmons and Ochoterena, 2000) as implemented in SeqState (Müller, 2005) and attached to the nexus file as a binary matrix. MP analysis and bootstrapping were performed with the same settings as for the cpDNA data. Computer clusters at the University of Oslo Bioportal (https://www.bioportal.uio.no/) were used for the *sqs* analyses. The complete sequence alignment is provided as Supplementary File 2. The tree resulting from the analysis of this dataset (Supplementary Figure S2) showed massive clustering of very similar *sqs* alleles of different species. In order to present the data in a more lucid way and to decrease computing time, the number of highly similar sequences in each monophyletic group of alleles was reduced. The new alignment is provided as Supplementary File 3. Indel coding and MP analysis for the reduced *sqs* dataset were performed as described above. In addition, Maximum likelihood (ML) analysis was done with RAxML (Stamatakis, 2006) using the raxmlGUI 1.1 software (Silvestro and Michalak, 2010). Partitioned datasets composed of DNA and standard characters cannot be analyzed in RAxML, therefore, 0 and 1 in the binary matrix of coded indels were replaced by A and T, and the modified nexus file was used as an input. ML analysis was performed with the rapid BS algorithm in combination with a ML search. A GTR model of nucleotide substitution with a gamma model of rate heterogeneity with a proportion of invariable sites was applied (corresponding to the best fitting model estimated by MrModeltest), and branch support was determined by 1000 bootstrap replicates using the same settings. Character conflict was assessed by a Neighbor Net approach based on uncorrected P-distances as implemented in Splitstree 4.11.3 (Huson and Bryant, 2006) using the same input file as for RAxML.

RESULTS

cpDNA phylogeny

The length of the amplified *trn*V-*ndh*C region varied from 832–1270 bp due to indel polymorphisms. Maximum sequence divergence within *Hieracium* (with indels treated as single characters) was 2.49% P-distance, compared with 2.39% for *trn*T-*trn*L. The combined alignment with coded indels contained 1734 characters; 133 were variable and 74 parsimony informative.

MP and Bayesian analyses produced trees with similar topologies and branch lengths (Figure 2). *Hieracium* was monophyletic and formed five major clusters (haplogroups A-E) and two individual lineages. All clusters contained diploids; the basalmost Haplogroup A consisted exclusively of diploid taxa. Triploid *H. mixtum* occurred in a basal position as sister to the remaining *Hieracium* taxa, which formed a monophyletic group comprising a separate lineage of triploid *H. naegelianum* and haplogroups B-E; the relationships among these remained unresolved (Figure 2, gray area).

Strong discrepancies between ETS and cpDNA were observed: species with a western origin according to ETS (blue) occurred in three cpDNA lineages (haplogroups B, C and D); those with an eastern origin (red) occurred in five lineages (haplogroups A, B, D, E and the *H. naegelianum* lineage). Furthermore, haplogroups B and D comprised subgroups composed of western as well as eastern taxa. For a detailed comparison of the datasets, see Table 1.

Sqs phylogeny

The length of the amplified region from exon 4 to intron 8 (Figure 1b) was 977–1198 bp. The *sqs* dataset including coded indels contained 1616 characters (of which 152 represented the indels); 471 were variable and 367 parsimony informative. The proportion of parsimony informative sites was 1.6 times higher than for ETS; maximum P-distance within *Hieracium* was 11% compared with

Table 1 ***Hieracium*** **accessions and their ploidy; ETS, cpDNA and** ***sqs*** **clades; and number of** ***sqs*** **alleles per clade**

Taxon	*Accession*	*Ploidy*	*ETS clade (subgroup) and inferred hybrid origin*[a]	*cpDNA haplogroup (subgroup)*[b]	*sqs sequence*[c]	*sqs clade (subgroup)*[d]	*No of sqs alleles/ clades*[e]
H. alpinum	Alp.Ukr	2 ×	Eastern (EA)	A1 (EA)	Alp.Ukr.X1c*	8 (EA)	2/1
					Alp.Ukr.X3c	8 (EA)	
	Alp.Boa2	2 ×	Eastern (EA)	A1 (EA)	Alp.Boa2X2c*	8 (EA)	2/1
					Alp.Boa2X3c	8 (EA)	
H. amplexicaule	1050/1	3 ×	Interclade hybrid (WP-E)	C (WP)	1050_1_X1c	11 (EU + EB)	3/2
					1050_1_X3c*	1a (WP)	
					1050_1_X4c	1a (WP)	
H. bifidum	1213/2	3 ×	Western (W)	B2 (W)	1213X1Nc*	10 (W)	2/1
					1213X3Nc	10 (W)	
H. bracteolatum	1240/2	3 ×	Interclade hybrid (Wx-EU)	E1 (EU)	1240X1c*	7 (Wx?)	4/2
					1240X4c*	7 (Wx?)	
					1240X10c	11 (EU + EB)	
					1240X12c*	11 (EU + EB)	
H. bupleuroides	1212/2	3 ×	Eastern (Epo)	D2 (Epo)	1212X2Lc*	12 (Epo 1)	3/3
					1212X3c	1b (Epo)	
					1212X5c*	14 (Epo 2)	
	1033/3	3 ×	Intraclade hybrid (EU-Epo)	E1 (EU)	1033X1c*	11 (EU + EB)	4/3
					1033X2c	14 (Epo 2)	
					1033X3c	11 (EU + EB)	
					1033X10c*	1b (Epo)	
H. caesium	1231 (plumb)	4 ×	Interclade hybrid (W-EU)	B2 (W)	plumX3c	11 (EU + EB)	5/3
					plumX4c	10 (W)	
					plumX8c*[f]	3 (mainly W)	
					plumX10c	11 (EU + EB)	
					plumX11c	10 (W)	
H. canadense	canad	3 ×	Eastern (EU)	E1 (EU)	canadX3c*	15 (EU)	2/2
					canadX5c*	4c (EU)	
H. candidum	1197/3	3 ×	Intraclade hybrid (WP-W)	C (WP)	1197X4c*	1a (WP)	2/2
					1197X5c*	10 (W)	
H. cerinthoides	1176/2	3 ×	Intraclade hybrid (WP-W)	C (WP)	1176X1Sc	10 (W)	2/2
					1176X2c	1a (WP)	
H. cordifolium	1177/5	2 ×	Intraclade hybrid (WP-W)	C (WP)	1177_5c*	1a (WP)	1 (DS)/1
H. eriophorum	1221/1	2 ×	Eastern (EU)	E1 (EU)	1221X1c*	11 (EU + EB)	2/1
					1221X2c*	11 (EU + EB)	
	1222/2	2 ×	Eastern (EU)	E1 (EU)	1222X1c*	11 (EU + EB)	2/1
					1222X3c*	11 (EU + EB)	
H. glaucum	1230/3 (Gla3)	3 ×	Interclade hybrid (W-Epo)	D2 (Epo)	Gla3X4c	10 (W)	3/2
					Gla3X9c	12 (Epo 1)	
					Gla3X10c	12 (Epo 1)	
H. gouani	1171/4	2 ×	Interclade hybrid (WP-E)	C (WP)	1171_4c*	1a (WP)	1 (DS)/1
H. gymnocephalum	1215/1	2 ×	Interclade hybrid (Wy-Ex)	E2 (?)	1215X3c	4d (?)	4/2
					1215X8c*	4 (mainly EU)	
					1215X10c*	4d (?)	
					1215X13c*	4d (?)	
	1207/2	3 ×	Interclade hybrid (Wy-Ex)	E2 (?)	1207X5c	6a (?)	3/2
					1207X12c*	4 (mainly EU)	
					1207X15c*	6a (?)	
H. gymnocerinthe	1172/4	3 ×	Intraclade hybrid (WP-W)	C (WP)	1172_4c	1a (WP)	1 (DS)/1
H. heterogynum	1250/2 (het)	3 ×	Interclade hybrid (W-Wy-Ex-EU)	E1 (EU)	hetX1c	10 (W)	3/3
					hetX2c*	6a (?)	
					hetX12c*	1b (Epo)	
H. humile	1064/2	4 ×	Western (W)	B2 (W)	1064X2c*[f]	3 (mainly W)	>2/>2
					1064X8c	10 (W)	
	1188/2	3 ×	Western (W)	B2 (W)	1188X2c[f]	3 (mainly W)	2/1
					1188X6c[f]	3 (mainly W)	
H. kittanae	1228/2 (kit)	2 ×	Eastern (EB)	B1 (EB)	kittX2c*	6 (EB)	2/1
					kittX8c	6 (EB)	

Table 1 (Continued)

Taxon	*Accession*	*Ploidy*	*ETS clade (subgroup) and inferred hybrid origin*[a]	*cpDNA haplogroup (subgroup)*[b]	*sqs sequence*[c]	*sqs clade (subgroup)*[d]	*No of sqs alleles/ clades*[e]
H. lachenalii	1160/2	3 ×	Interclade hybrid (W)	E1 (EU)	1160X1c	10 (W)	4/2
					1160X2c*	10 (W)	
					1160X3c	10 (W)	
					1160X4c*	Basal 5–9 (?)	
H. laevigatum	1031/11	3 ×	Interclade hybrid (W-EU)	E1 (EU)	1031X3c	10 (W)	3/2
					1031X7c	11 (EU + EB)	
					1031X9c*	11 (EU + EB)	
H. lawsonii	1175/1	3 ×	Western (WP)	C (WP)	1175X1c	1a (WP)	2/1
					1175X4c	1a (WP)	
H. lucidum	H.lucidum	2 ×	Intraclade hybrid (W-Wx)	D (W)	lucX1c*[f]	2 (?)	2/2
					lucX2c*	Basal 5–9 (Wx?)	
H. mixtum	H.mixt	3 ×	Interclade hybrid (W-E)	? (?)	mixX13c	1a (WP)	3/2
					mixX15c	1a (WP)	
					mixX16c*	Basal 3–15 (?)	
H. murorum	875/1	3 ×	Western (W)	B2 (W)	8751X4c[f]	3 (mainly W)	>2/>2
					8751X6c	10 (W)	
H. naegelianum	1208/2	3 ×	Eastern (EB)	? (?)	1208X2c	11 (EU + EB)	>2/>1
					1208X7c	11 (EU + EB)	
H. olympicum	1206/3 (oly)	3 ×	Interclade hybrid (Wx-EB)	E2 (?)	olyX4c*	9 (Wx?)	3/2
					olyX7c*	6 (EB)	
					olyX8c	6 (EB)	
H. pannosum	1205/1 (pan)	3 ×	Eastern (EB)	B1 (EB)	panX1c*	6 (EB)	3/2
					panX8c*[f]	3 (mainly W)	
					panX10c	6 (EB)	
H. petrovae	1229 (petr)	2 ×	Eastern (EB)	B1 (EB)	petrX1c	6 (EB)	3/2
					petrX2c*	6 (EB)	
					petrX7c*	11 (EU + EB)	
H. pictum	1067/4	3 ×	Western (W)	B2 (W)	1067X4c*	Basal 5 (W)	3/3
					1067X11c	10 (W)	
					1067X16c*	4 (mainly EU)	
H. pilosum	1226/1	3 ×	Eastern (Epo)	D2 (Epo)	12261X3Sc*	Basal 12–13 (Epo 1)	3/2
					12261X5c*	14 (Epo 2)	
					12261X14c	14 (Epo 2)	
	1226/2	3 ×	Interclade hybrid (Wy-Epo)	D2 (Epo)	—	—	—
H. plumulosum	1218/2	2 ×	Interclade hybrid (W-Wy-Ex-E)	E (?)	1218X1c	10 (W)	2/1
					1218X4c	10 (W)	
H. pojoritense	Poi.Rom	2 ×	Intraclade hybrid (EU-EA)	A1 (EA)	poiX3c	11 (EU + EB)	3/2
					poiX4Lc	11 (EU + EB)	
					poiX5c*	10 (W)	
H. porrifolium	1052/9	2 ×	Eastern (Epo)	D2 (Epo)	HQ131843*	12 (Epo 1)	2/1
					1052X17c	12 (Epo 1)	
H. prenanthoides	1252 (prenFra)	2 ×	Interclade hybrid (W-E)	D (W)	prenFra_alt*	4a (?)	2/1
					prenFX8c	4a (?)	
	1161/2	3 ×	Interclade hybrid (W-Wx-E)	D1 (W)	1161_2c1	4a (?)	2 (DS)/1
					1161_2c2	4a (?)	
	1187/1	3 ×	Interclade hybrid (W-E-EU)	D (W)	1187X3Lc*	11 (EU + EB)	3/2
					1187X5c*	4a (?)	
					1187X6c	4a (?)	
H. racemosum	874	3 ×	Interclade hybrid (Wx-EU)	E1 (EU)	874X2Lc	11 (EU + EB)	>3/>2
					874X3c*	15 (EU)	
					874X4c	11 (EU + EB)	
H. ramondii	1173/3	3 ×	Western (WP)	C (WP)	1173_3c	1a (WP)	>1 (DS)/1
H. recoderi	1174/4	2 ×	Western (WP)	C (WP)	1174_4c	1a (WP)	>1 (DS)/1
H. sabaudum	1098/2	3 ×	Interclade hybrid (Wx-EU)	E1 (EU)	1098X6c*	11 (EU + EB)	3/2
					1098X8c	11 (EU + EB)	
					1098X10c*	9 (Wx?)	

Table 1 (Continued)

Taxon	*Accession*	*Ploidy*	*ETS clade (subgroup) and inferred hybrid origin*[a]	*cpDNA haplogroup (subgroup)*[b]	*sqs sequence*[c]	*sqs clade (subgroup)*[d]	*No of sqs alleles/ clades*[e]
H. schmidtii	1025/3	3 ×	Western (W)	B2 (W)	1025X11c	10 (W)	3/2
					1025X12c*	10 (W)	
					1025X14c*	5 (W)	
H. sparsum	1251/1 (spaJCh)	2 ×	Eastern (EB)	A2 (?)	spaJCh1X1c*[f]	14 (mainly E)	2/1
					spaJCh1X2cf	14 (mainly E)	
	spa.sst.2	2 ×	Eastern (EB)	A2 (?)	Spasst1X1c*[f]	14 (mainly E)	2/1
					Spasst1X2c[f]	14 (mainly E)	
H. stelligerum	1233/1	2 ×	Western (W)	B2 (W)	1233X2c*	5 (W)	2/2
					1233X3c	10 (W)	
H. tomentosum	1066/8	2 ×	Western (W)	D1 (W)	1066X1c	10 (W)	2/1
					1066X3c	10 (W)	
H. transylvanicum	tra.Boa	2 ×	Western (W)	B (?)	traBoaX1c*	14 (mainly E)	2/1
					traBoaY1c	14 (mainly E)	
	1077/7	2 ×	Western (W)	B (?)	1077X1c*	14 (mainly E)	2/1
					1077X4c	14 (mainly E)	
H. umbellatum	1021/1	2 ×	Eastern (EU)	E1 (EU)	HQ131842	11 (EU + EB)	2/1
					HQ131841	11 (EU + EB)	
	um.AM.1	2 ×	Eastern (EU)	E1 (EU)	umAm1c	11 (EU + EB)	1 (DS)/1
H. villosum	1029/1	4 ×	Eastern (Epo)	D2 (Epo)	1029X1c*	14 (mainly E)	4/2
					1029X2c	14 (mainly E)	
					1029X3c*	14 (mainly E)	
					1029X4c*	Basal 12–13 (Epo 1)	
	1305/3	3 ×	Interclade hybrid (Wy-Epo)	D2 (Epo)	1305X4c*	13 (Epo 1)	3/2
					1305X6c*	Basal 12–13 (Epo 1)	
					1305X9c*	13 (Epo 1)	
H. virosum	1238/1	3 ×	Eastern (EU)	E1 (EU)	1238X3c*	4b (EU)	3/2
					1238X5c*	4c (EU)	
					1238X8c*	4b (EU)	
	vir.1 (vir.R)	3 ×	Eastern (EU)	E (EU)	virRX1c	11 (EU + EB)	2/1
					virRX4c	11 (EU + EB)	
H. intybaceum	1069/1	2 ×	n.d.	A3 (?)	HQ131846(c8)*	2 (?)	2/1
					HQ131847(c13)	2 (?)	
	InbKaer	2 ×	Outgroup	A3 (?)	InbKaerc*	2 (?)	1 (DS)/1

EA, (eastern) *H. alpinum* lineage; EB, (eastern) Balkan species without evidence for hybrid origin; Epo, (eastern) *H. porrifolium* group; EU, (eastern) *H. umbellatum* group; W, basal western; WP, (western) Pyrenean; Wx, Wy, and Ex, 'unknown' ribotypes: two western lineages and one eastern lineage occurring only in hybrids (potential remnants of extinct ancestors).

[a] Eastern or western origin and subclade (if any) based on ETS are shown as well as intra- or interclade hybrid status according to Fehrer *et al.* (2009).

[b] The assignment of haplogroups is based on the *trn*V-*ndh*C + *trn*T-*trn*L combined dataset (see Figure 2); correspondence to ETS (sub)groups is indicated as far as assignable; unique haplotypes that could not be attributed to any group are indicated by '?'.

[c] Sequences included in the reduced *sqs* dataset (Figure 3) are marked with an asterisk (*); sequence labels correspond to those in Supplementary Files 2 and 3 and Supplementary Figures S2 and S4; sequences with accession numbers are from Krak *et al.* (2012).

[d] Particular alleles and the *sqs* clades in which they occur (Figure 3) are given. ETS (sub)groups were assigned as far as possible according to nonhybrid taxa (according to ETS) included in the particular *sqs* clades; alleles that could not be attributed to any group are indicated by '?'; Epo 1 and Epo 2 refer to the inset of Figure 3.

[e] DS = direct sequence; '>' indicates that more alleles/clades exist according to direct sequencing, but no clones representing them were as yet retrieved.

[f] Intragenomic recombinant sequences, excluded from the phylogenetic analysis shown in Supplementary Figure S4.

4.6% for ETS and *sqs* sequences were about twice as long (ETS: 556–570 bp).

The *Hieracium* accessions were arranged in 15 monophyletic clusters of alleles (Figure 3). Most clusters included alleles of diploids. At the base of the tree, relationships among *Hispidella*, *Pilosella*, *Hieracium* Clade 1 and a group containing the rest of *Hieracium* remained unresolved (yellow area). *Hieracium* Clade 2 was sister to a branch consisting of the majority of *Hieracium* accessions (Clades 3–15). In this core group, only one basal monophyletic lineage was supported (Clade 3) whereas the rest of the relationships among the clades remained unresolved (for example, green area). In contrast, at the tips of the tree, clade support was high (Clade 4 obtained low support, but it contained several well-supported subclades). Most relationships within the delimited clades remained unresolved as well.

Massive clustering of very similar alleles by different taxa was evident in Clades 1a, 10 and 11. In addition, sequences of Clades 4 + 15, 5–9 and 12 + 13 (including taxa at basal positions) were highly similar. Nearly all individuals showed more than one allele; these could be very similar (ending up in the same clade), or divergent (ending up in different clades). Most frequent was a combination of two similar and one divergent allele in triploids (Table 1). A graphical overview of 32 individuals of 29 species with alleles occurring in different clades is given as Supplementary Figure S3. Five individuals were diploid, but only 3 of them were supposed

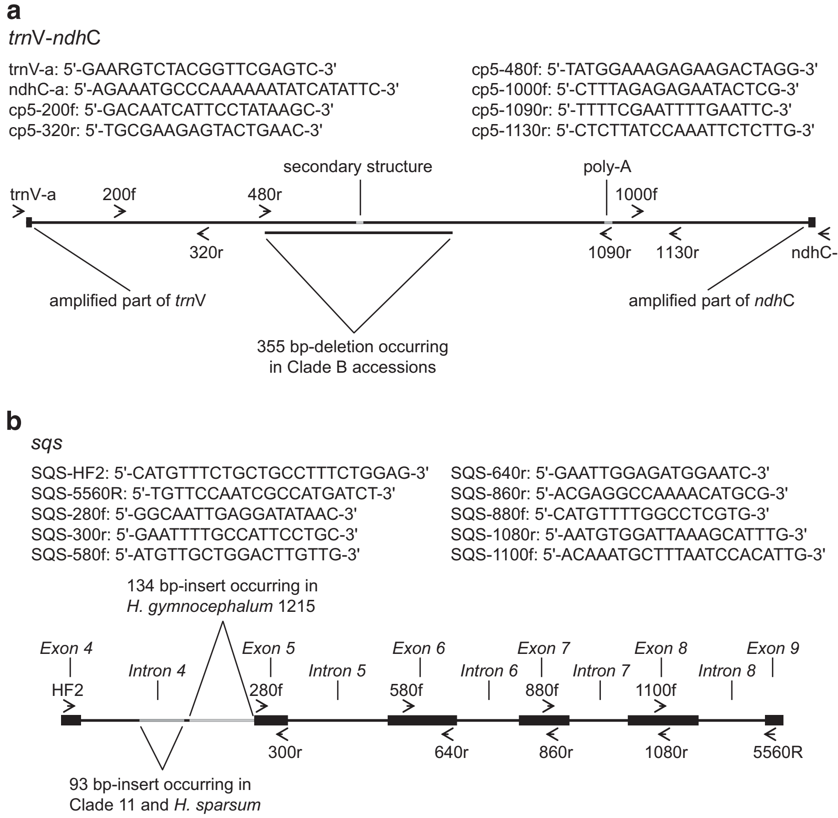
a
trnV-ndhC
trnV-a: 5'-GAARGTCTACGGTTCGAGTC-3'
ndhC-a: 5'-AGAAATGCCCAAAAAATATCATATTC-3'
cp5-200f: 5'-GACAATCATTCCTATAAGC-3'
cp5-320r: 5'-TGCGAAGAGTACTGAAC-3'
cp5-480f: 5'-TATGGAAAGAGAAGACTAGG-3'
cp5-1000f: 5'-CTTTAGAGAGAATACTCG-3'
cp5-1090r: 5'-TTTTCGAATTTTGAATTC-3'
cp5-1130r: 5'-CTCTTATCCAAATTCTCTTG-3'
secondary structure
poly-A
trnV-a
200f
480r
1000f
320r
1090r
1130r
ndhC-a
amplified part of trnV
amplified part of ndhC
355 bp-deletion occurring in Clade B accessions
b
sqs
SQS-HF2: 5'-CATGTTTCTGCTGCCTTTCTGGAG-3'
SQS-5560R: 5'-TGTTCCAATCGCCATGATCT-3'
SQS-280f: 5'-GGCAATTGAGGATATAAC-3'
SQS-300r: 5'-GAATTTTGCCATTCCTGC-3'
SQS-580f: 5'-ATGTTGCTGGACTTGTTG-3'
SQS-640r: 5'-GAATTGGAGATGGAATC-3'
SQS-860r: 5'-ACGAGGCCAAAACATGCG-3'
SQS-880f: 5'-CATGTTTTGGCCTCGTG-3'
SQS-1080r: 5'-AATGTGGATTAAAGCATTTG-3'
SQS-1100f: 5'-ACAAATGCTTTAATCCACATTG-3'
134 bp-insert occurring in H. gymnocephalum 1215
Exon 4
Intron 4
Exon 5
Intron 5
Exon 6
Intron 6
Exon 7
Intron 7
Exon 8
Intron 8
Exon 9
HF2
280f
580f
880f
1100f
300r
640r
860r
1080r
5560R
93 bp-insert occurring in Clade 11 and H. sparsum

Figure 1 Primer sequences, location and structural features of *trn*V-*ndh*C and *sqs*. (**a**) *trn*V-*ndh*C: PCR and internal sequencing primers are listed. Graphics show their locations, the position of a large deletion, a secondary structure and a poly-A region affecting many sequence reads. (**b**) *sqs*: The first two primers are the PCR primers from Krak *et al.* (2012), the others are additional sequencing primers. Graphics show their location, the exon–intron structure and the position of two large inserts in intron 4.

to have hybrid origin (Table 1); 10 out of 27 polyploid individuals were supposed to be autopolyploids (Table 1).

Individuals lacking previous evidence for hybrid origin are indicated on the *sqs* tree (Figure 3) by colors reflecting their respective species groups based on ETS (Table 1). Taxa with western European origin that showed a basal position in the western ETS clade (W) occurred mainly in Clades 3, 5 and 10. *Sqs* alleles of a Pyrenean group (WP) clustered together in Clade 1a. Among taxa with eastern European origin, alleles of the *H. umbellatum* group (EU) belonged to Clades 11, 4 and 15. Members of the *H. porrifolium* group (Epo) fell into Clades 12–14 (and basal positions among them, see also below); one species had additional alleles in Clade 1b. *H. alpinum* alleles (EA) occurred only in Clade 8. Alleles of Balkan species (EB) occurred mainly in Clades 6 and 11.

Tree regions lacking resolution

Both the cpDNA tree (Figure 2) and the *sqs* tree (Figure 3) showed a lack of resolution in basal parts as well as within well-supported clades. For crown groups, this may be due to recent speciation. Concerning the deeper nodes in the cpDNA tree (Figure 2), the polytomy (gray area) can result from insufficient variation even with the use of the combined intergenic spacers and/or from a rapid divergence of these lineages. The variability of *sqs* was very high, therefore, the lack of resolution among basal lineages (Figure 3, yellow and green areas) cannot be attributed to insufficient variation. We therefore investigated the *sqs* data in more detail to search for other explanations.

Closer inspection of the *sqs* alignment (Supplementary File 2) revealed several recombinant sequences. (i) Clade 14 included *H. sparsum* (two individuals, two highly similar alleles each). The first two-thirds of these four sequences corresponded to Clade 11 (including a diagnostic 93 bp-insert, Figure 1b) whereas from exon 7 onwards, they were most similar to alleles from Clades 5–9. Several shared unique characters and the absence of other alleles in both accessions according to direct sequencing refute PCR recombination. Apparently, ML analysis as well as the bootstrapping of the MP analysis were affected by these intragenomic recombinant sequences: the inset in Figure 3 shows two parts of the parsimony strict consensus tree in which sequences of Clades 12 and 13 along with very similar sequences at their base formed one group (lower part of the inset, Epo 1) while the rest of the Epo alleles in Clade 14 formed another group (Epo 2), with *H. sparsum* and *H. transylvanicum* as separate lineages (upper part of the inset). Clade 11 branched off between these groups (see also Supplementary Figure S2). (ii) One allele of *H. lucidum* occurred basal to Clades 5–9, the second clustered with *H. intybaceum* in Clade 2. Intron 5 of the latter contained many diagnostic indels and substitutions of *H. intybaceum* (Supplementary File 2). From exon 7 onwards, this allele was most similar to Clade 3. The two highly divergent alleles of *H. lucidum* were based on three clones each; no PCR recombinants were found in this individual so that this allele also represents an intragenomic recombinant sequence. (iii) Intron 5 of all Clade 3 alleles was shared with Clade 1 whereas intron 6 corresponded to sequences of Clades 4 and 15. In addition, from intron 7 onwards, *H. murorum* and *H. caesium* alleles (based on five and three identical clones,

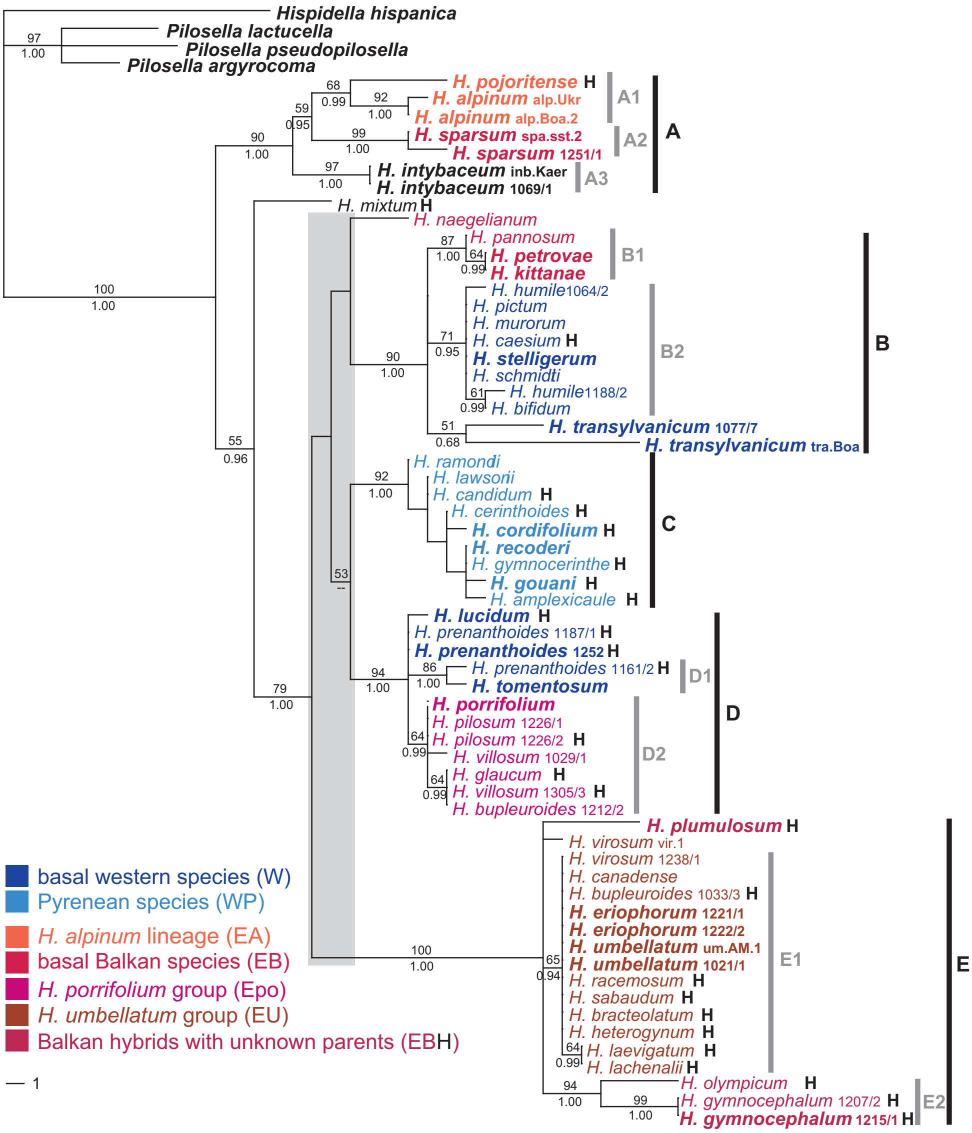

Figure 2 Phylogenetic analyses of combined chloroplast datasets (*trn*T-*trn*L, *trn*V-*ndh*C). One of the 25 most parsimonious trees is shown. The MP trees differ only in details of unsupported relationships within Haplogroup C and the exact placement of *H. transylvanicum* within Haplogroup B. Bootstrap values of the MP analysis are above branches; posterior probability values ⩾0.90 from Bayesian analysis are below branches. The gray area highlights unresolved relationships among five main lineages. Diploid accessions are in bold. Species falling into the major western clade in the ETS tree (Fehrer *et al*., 2009) are in blue; those of the major eastern ETS clade in red. Colors of accessions with hybrid origin ('H' behind the species name) match their maternal parents (if assignable). Different shades of blue and red correspond to species subgroups/ETS subclades; the same abbreviations for these groups are used in Tables 1 and 2, and Figure 3.

respectively) were recombinant with Clade 10. All Clade 3 alleles also shared several unique mutations (resulting in strong bootstrap support). Triploid *H. humile* 1188 contained exclusively Clade 3 alleles; one of them was almost identical to an allele from tetraploid *H. humile* 1064 (see also Supplementary Figure S2). Thus, also for the sequence patterns in Clade 3, PCR recombination can hardly be responsible. (iv) One allele of *H. pojoritense* occurred in Clade 10. Closer inspection showed that this sequence corresponded mostly to others of Clade 10, but exon 4 through intron 5 were more similar to the majority of sequences from other clades. This allele was based on three cloned sequences that were unique (several substitutions and an indel) compared with other taxa, which excludes PCR artifacts, and therefore, we assume a further case of intragenomic recombination.

Additional phylogenetic analyses were carried out excluding these recombinant alleles to test how they affected tree topology and branch support. The results are shown in Supplementary Figure S4. Most of the relationships remained unaffected. Clade 14 (previously comprising *H. sparsum*) lost support and *H. transylvanicum* became separated, as in the original parsimony analysis (inset in Figure 3, Supplementary Figure S2). A new branch with moderate support emerged that contained Clades 11–13, but relationships among them remained unresolved. To visualize character conflict, Neighbor net analyses were performed, successively deleting the recombinant sequences (Supplementary Figure S5). Character conflict was reduced at each step; however, although all clades remained stable, their relationships (corresponding to the yellow and green areas in Figure 3) still remained unresolved. Thus, these recombinant

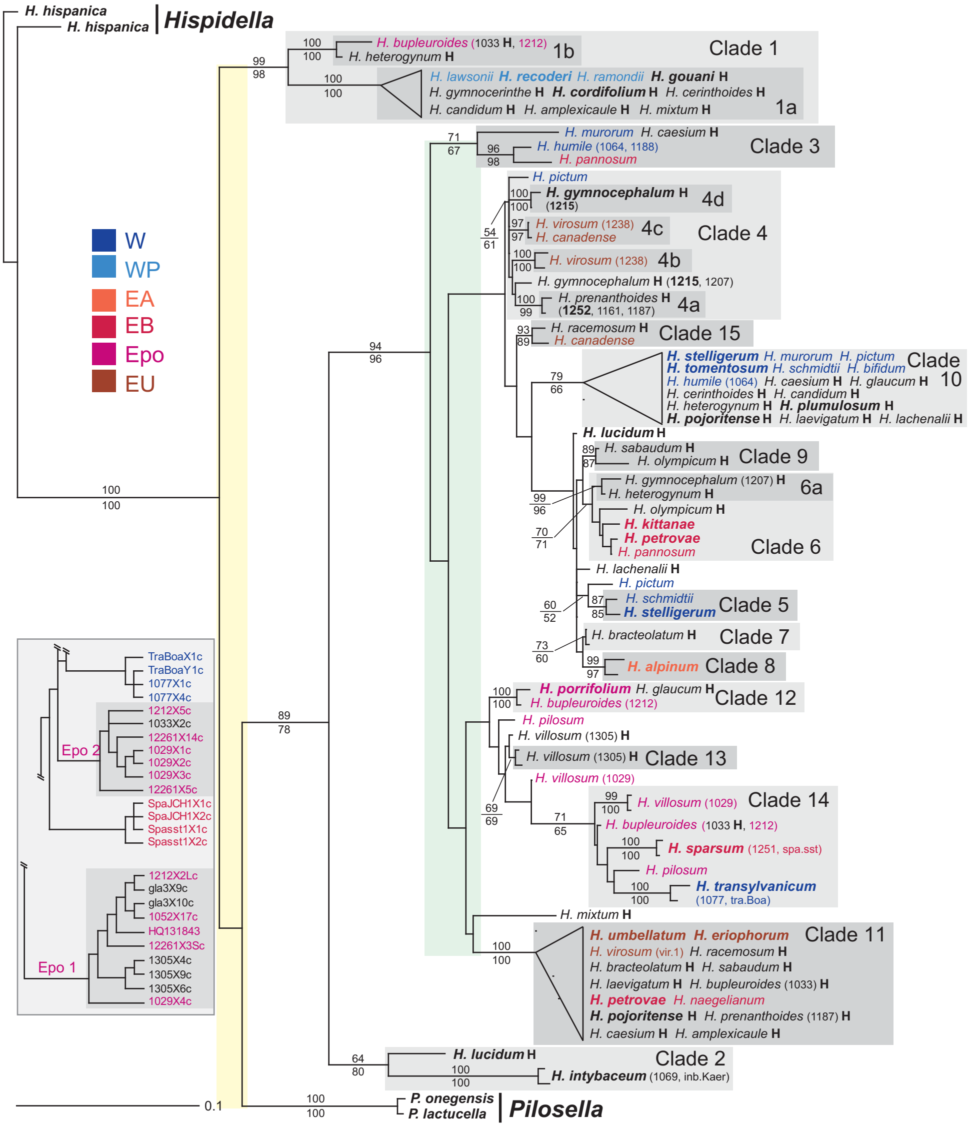

Figure 3 Phylogenetic analyses of the reduced *sqs* dataset. The single most likely RAxML tree is presented with bootstrap values above branches. Bootstrap support from the MP analysis is given below branches. Monophyletic clades with significant bootstrap support are shaded in gray and numbered arbitrarily. The names of all species or individuals falling into the particular clades in the complete dataset are included. If many occurred in the same clade and the relationships within the clade remained unresolved, the branches of the original tree were replaced by triangles (Clades 1b, 10, and 11). The inset to the left of the tree shows parts of the parsimony strict consensus tree of the complete dataset for Clades 12–14 and basal taxa therein (see also Supplementary Figure S2). Diploid accessions are in bold; inferred hybrid origin according to Fehrer *et al.* (2009) is indicated by an 'H' after the species name. Non-hybrid taxa are shown in the same colors as in the cpDNA tree (Figure 2), species groups are indicated by the same abbreviations. The yellow and green areas highlight unresolved relationships among clades referred to in the text.

sequences had surprisingly little effect on tree topology and resolution. Moreover, by excluding them, important information about true allelic variation was lost, including two entire taxa (*H. humile* 1188 and both individuals of *H. sparsum*).

Reassessment of hybrid origins with *sqs*

Previously inferred hybrid origins (Fehrer *et al.*, 2009) were reassessed as far as the assignment of *sqs* clades to species groups (outlined above) was possible for particular individuals (Table 1). Taxa for which combinations of *sqs* alleles corresponded to the anticipated parental lineages included *H. bupleuroides* 1033, *H. candidum*, *H. cerinthoides*, *H. glaucum*, *H. laevigatum*, *H. caesium* and *H. prenanthoides* 1187. In contrast, *sqs* did not provide evidence for the hybrid origin of *H. gymnocephalum* 1215, *H. prenanthoides* 1252 and 1161, *H. plumulosum*, *H. villosum* 1305, *H. gouani*, *H. cordifolium* and *H. gymnocerinthe*. In these cases, *sqs* alleles from only one clade were found, mostly representing the maternal parent.

For a few individuals, *sqs* allele compositions conflicted with previously inferred origins. Some of these might reflect cases of undetected hybrid origins and were therefore investigated in more detail. (i) Diploid *H. transylvanicum* was already suspected of hybrid origin due to its western ETS ribotype, but eastern geographic distribution and a genome size in the range typical of eastern species (Chrtek *et al.*, 2009). CpDNA remained inconclusive (Clade B, Figure 2). *Sqs* alleles of this species were quite divergent from all others. They nested in Clade 14 or its vicinity (Figure 3, Supplementary Figures S4 and S5), which may be considered as molecular evidence for an eastern origin and, taken together, suggest interclade hybrid origin for this species. (ii) Triploid *H. pictum*, supposed to be an autopolyploid (W), contained two western alleles (Clade 10, basal to Clade 5) in combination with an allele from Clade 4, which contained several subclades composed of EU species or interclade hybrids (Figure 3). Among the latter are three *H. prenanthoides* accessions (Subclade 4a). Distribution areas of *H. pictum* and diploid endemic populations of *H. prenanthoides* overlap in the southwestern Alps, therefore, *sqs* might indicate allopolyploid origin for *H. pictum* that was undetectable by ETS due to a lack of variation among basal western ribotypes. All other species of Clade 4 occur on the Balkan Peninsula, in Western Asia, or Canada and are therefore unlikely introgressants. (iii) Triploid *H. heterogynum* showed three divergent *sqs* alleles in Clade 10 (W), Clade 6 (EB), and Clade 1b (Epo); the latter did not fit to either ETS nor cpDNA. Taken together, this would suggest an even more complex parentage (five different lineages) of this hybrid, a result that currently can neither be dismissed nor explained. (iv) Triploid *H. pannosum* (considered as an autopolyploid EB) showed a combination of *sqs* alleles from Clade 6 (EB) and Clade 3 (mainly W). The Clade 3 allele of *H. pannosum* (confined to the Balkans and Anatolia) clustered with triploid and tetraploid accessions of *H. humile* (mainly western Alps). It cannot be excluded that the (unknown or extinct) ancestral diploid populations of these species have hybridized, but their current distribution areas do not support this. (v) Possibly, the intragenomic recombinant alleles (see above) that did not fit the previously inferred parentages of two diploid hybrids may also result from additional or alternative undetected hybridization events. In case of *H. pojoritense* (EA-EU), the partial Clade 10 (W) allele may have been contributed by pollen from a widespread western polyploid species or an interclade hybrid, but there is no concrete evidence to draw firm conclusions. *H. lucidum* (W-Wx) has currently only a single relict population on Sicily (Italy) whereas diploid *H. intybaceum* (for its origin, see below) is distributed in the Alps, and ecological demands as well as geographic distance (almost 1000 km between the nearest populations) of these species suggest that, if introgression between them was responsible for the recombinant (W–*intybaceum*) allele, this should have been a rather ancient event.

Incongruences among ETS, cpDNA and *sqs* datasets

A comparison of the occurrence of nonhybrid taxa in the ETS, cpDNA and *sqs* trees is provided in Table 2. The only species group that formed a clade in all datasets was the Pyrenean group (WP), and two accessions of *H. alpinum* (EA) formed lineages distinct from all other groups, but even WP and EA occurred at very different positions in the individual trees. Partial concordance occurred between ETS and cpDNA; ETS and *sqs* matched only in some individual cases (Table 2).

The division of the genus into two major clades with western or eastern origin revealed by ETS were neither found with cpDNA nor with *sqs*. Although each ETS species (sub)group was reflected by corresponding species clusters in the cpDNA and/or *sqs* trees (Figures 2 and 3), most of these groups were split. For example, basal western species (W) fell into two distinct haplogroups and into three divergent *sqs* clades. Furthermore, the species compositions of these groups differed between cpDNA and *sqs*. Species of the EU and Epo groups occurred in three, often very divergent *sqs* clades whereas the respective species shared group-specific cpDNAs. EB species showed one haplogroup and two individual cpDNA lineages, and their *sqs* alleles fell into two divergent clades and one individual lineage. In contrast, some highly divergent ETS groups were merged in the same monophyletic clusters, although there was no indication for hybrid origin of the respective individuals. For example, two cpDNA haplogroups characteristic of western taxa grouped together with those of eastern species: W and Epo (= D2) accessions co-occurred in Clade D; Clade B comprised EB and W accessions corresponding to subclades B1 and B2 (Figure 2). The very similar sequences of *sqs* Clades 5–9 also comprised W and EB taxa (Figure 3); all of them shared cpDNA haplogroup B. In the basalmost *Hieracium sqs* Clade 1, one allele of *H. bupleuroides*, a triploid Epo species (1b), clustered with alleles of the Pyrenean clade (WP).

Another major incongruence among the three datasets concerned the basalmost lineages and the outgroup. Fehrer *et al.* (2007) have inferred an ancient intergeneric hybridization event between *Hieracium* and *Pilosella* based on incongruent nuclear and chloroplast data. *Hieracium intybaceum* nrDNA was so divergent from *Hieracium* and related genera that this taxon was suitable as an outgroup for phylogenetic analyses of nrDNA datasets (Fehrer *et al.*, 2007, 2009). According to cpDNA (Fehrer *et al.*, 2007; Figure 2), this species clearly belongs to *Hieracium*, a pattern that was attributed to a further

Table 2 Comparison of the three datasets for 'pure' species

ETS (sub)group	*cpDNA haplogroup (sub)group*	*sqs clade*	*Concordance among datasets*
Major western clade (W + WP)	B2 (W), C (WP), D1 (W)	1a (WP), 3 (W), 5 (W), 10 (W)	No
Basal western species (W)	D1: *H. tomentosum* (W)	10 (W)	Yes (1 individual)
	B2 (W)	10 (W), 3 (W), 5 (W)	No
Western Pyrenean (WP)	C (WP)	1a (WP)	Yes
Major eastern clade (E)	A1 (EA), A2 (EB), *H. naegelianum*, B1 (EB), D2 (Epo), E (EU)	1b (Epo), 4 (EU), 15 (EU), 6 (EB), 8 (EA), 12 + 13 + basal (Epo 1), 14 (Epo 2), 11 (EU + EB)	No
Eastern basal Balkan species (EB)	B1 (EB)	6 (EB), 11 (EU + EB)	ETS & cpDNA
	H. naegelianum	11 (EU + EB)	ETS & sqs (1 individual)
	A2 (*H. sparsum*)	14 (*H. sparsum*) = recombinant 6 (EB) + 11 (EU + EB)	ETS & sqs (1 species)
Eastern *H. umbellatum* group (EU)	E (EU)	11 (EU + EB), 4 (EU), 15 (EU)	ETS & cpDNA
Eastern *H. porrifolium* group (Epo)	D2 (Epo)	12 + 13 + basal (Epo 1), 14 (Epo 2), 1b (Epo)	ETS & cpDNA
Eastern *H. alpinum* lineage (EA)	A1 (EA)	8 (EA)	Yes (1 species)

Notes: Abbreviations in parentheses refer to species subgroups as in Table 1 and Figures 2 and 3 except that *H. transylvanicum*, *H. pictum*, and *H. pannosum*, which may also have hybrid origin according to *sqs*, were excluded; Epo 1 and Epo 2 refer to the inset in Figure 3.

chloroplast capture event (Fehrer *et al.*, 2007). *Sqs* may reflect these ancient hybridization events: outgroup and ingroup lineages formed a basal polytomy (yellow area in Figure 3), which also comprised *H. intybaceum* after deleting the intragenomic recombinant *H. lucidum* sequence (see above) from the analysis (Supplementary Figure S4).

DISCUSSION

Phylogenetic analyses of 61 individuals representing 47 basic species of *Hieracium* were performed using combined analyses of two chloroplast intergenic spacers (*trn*V-*ndh*C, *trn*T-*trn*L) and a recently developed marker, the low-copy nuclear gene for squalene synthase (*sqs*) (Krak *et al.*, 2012), which is applied here for the first time in a phylogenetic study. These data were compared with that of previous results for the same accessions based on ETS and *trn*T-*trn*L (Fehrer *et al.*, 2009). The combined cpDNA dataset resulted in a better-resolved tree (Figure 2) compared with analysis of *trn*T-*trn*L alone. However, many species shared the same haplotypes, that is, resolution of close interspecific relationships was not achieved. Variation of *sqs* was about four times as high as for ETS; this marker was therefore expected to be suitable for the identification of previously unresolved relationships. However, most well-supported clades contained highly similar alleles of several species (Figure 3), usually without any further resolution within the clades. This implies that speciation in *Hieracium* was actually so recent that not even this highly variable LCNM was able to resolve the relationships within most species groups.

As in most phylogenetic studies investigating multiple markers (for example, Doyle *et al.*, 2003), incongruence was observed among gene trees. Although clusters of the same crown groups were recognizable with all markers, the deep split of the genus into two major phylogenetic lineages based on ETS was not revealed by the other datasets, and most of the subclades and species groups were either split into several clades or merged into a single one by cpDNA as well as *sqs*, but these groups were not always the same between the two markers. Gene tree incongruence can principally result from three different evolutionary processes: paralogy (gene duplication), hybridization or ILS (Funk and Omland, 2003), which will be addressed in the following sections.

Copy status of *sqs*

Single-copy status (one locus per haploid genome) applies if the number of alleles is not higher than the ploidy, independent of the origin or diversity of alleles within a sample. By 'alleles', we refer to cloned sequences that were corrected for polymerase errors and apparently not recombinant within an accession, that is, reflecting true variation. Alleles were validated by direct sequencing and mostly based on several identical clones (see Materials and Methods).

For the majority of individuals, more than one allele was identified (Table 1; Supplementary Figure S3). Usually, the number of alleles did not exceed the ploidy level. Surplus alleles occurred in seven individuals (Table 1), but the number of *sqs* clades (Table 1, Figure 3) did not exceed the ploidy level. Given the high sequence similarity within the clades, these additional 'alleles' may be attributed to polymerase errors that were not identifiable as such (Speksnijder *et al.*, 2001) or to a failure to decide which of several very similar sequences were recombinant within a sample, especially if the putative recombination point was close to one end of the sequence. The inclusion of sequences into the final alignment was done in a conservative way in order not to dismiss small, but real differences, which may be easily overlooked if alleles are inferred from consensus sequences of clones. Therefore, these few highly similar surplus alleles probably do not represent real differences. If they were true variation, the gene would have been independently duplicated in individual taxa and different clades, which is not very likely. Besides, duplicated alleles occurring at the tips of the tree cannot be responsible for incongruence at deeper nodes. Paralogs resulting from locus duplication in internal branches normally can be observed to form parallel clades composed of the same sets of taxa (for example, Evans and Campbell, 2002). However, nearly every individual had its own unique allele composition (Table 1). No stop codons were found, and all exon–intron junctions were conserved, that is, there is no indication of pseudogenes either. To conclude, paralogs are unlikely for *sqs* in *Hieracium*, but even if the surplus alleles should result from independent duplication events, these would be confined to the tips of the tree and be irrelevant for the question of gene tree incongruence.

Hybrid origin

For almost half of the *Hieracium* accessions investigated, hybrid origin had been inferred based on additive patterns of ETS or, in one case, on incongruence between ETS and *trn*T-*trn*L trees (Fehrer *et al.*, 2009). In several cases, *sqs* confirmed the presumed parental lineages. In other cases, it revealed only one of the parents (usually the maternal one). Potential reasons could be allele loss (in diploids), or inheritance of similar alleles from both parents instead of divergent ones (in apomictic polyploids) or a combination of both, if hybridization was not immediately followed by polyploidization and apomixis. *Sqs* also indicated a few additional cases of potential hybrid origin or a more complex genome composition of some known hybrids. Morphology is inconclusive concerning the potential hybrid origin of *Hieracium* 'basic species' (see Introduction), but some further evidence was provided by the genome size of the hybrid (one case) or the contemporary distribution areas of the putative parental taxa. As the latter information was either supportive, equivocal or contradictory, hybrid origin may not in each case be the correct explanation, if individual *sqs* alleles occurred in the 'wrong' clade, or at least not the only possible one.

Such a high proportion of hybrid versus nonhybrid taxa is very unusual for studies on hybrid or allopolyploid origins. While the large number of reticulation events required detailed pattern analyses across all datasets and careful assessment of all other available information, putative hybrid origin as such actually revealved very little contradiction between the datasets. This is in agreement with the notion of Russell *et al.* (2010) that reticulate evolution will cause more consistent patterns among different markers than other reasons for incongruence. The inference of reticulation in *Hieracium* was probably facilitated by the relatively late occurrence of most events, associated with the emergence of polyploidy and apomixis in *Hieracium* (Fehrer *et al.*, 2009). However, the major conflicts among the trees affected deeper nodes in all three datasets and also affected taxa without any indication of hybrid origin. Thus, despite being abundant, hybridization did not account for the major incongruences among the datasets.

Incomplete lineage sorting

Both ILS and hybridization mostly concern closely related and recently diverged taxa (Degnan and Salter, 2005), and it may be often impossible to distinguish between these processes (Joly *et al.*, 2006). However, ILS also occurs in cases of ancient rapid radiation (Degnan and Rosenberg, 2009) and can affect haploid markers like chloroplast or mitochondrial DNA as well (Funk and Omland, 2003). In some circumstances, branch length information can be used to distinguish between ILS and hybridization (Holder *et al.*, 2001). Whitfield and Lockhart (2007) suggested that where different data sets agree that the same branches are short or obtained low support, this could be used as an indication of rapid radiation.

It must first be determined whether such short internal branches are due to character conflict. This is certainly the case for the *sqs* tree (Figure 3), where intragenomic recombinant sequences involving alleles from divergent clades resulted in a lack of resolution for relationships among the clades. However, exclusion of these sequences did not resolve the relationships among the remaining clades (Supplementary Figures S4 and S5), that is, the two basal polytomies (yellow and green areas in Figure 3) remained. Lack of resolution was also observed in the cpDNA tree as a basal polytomy of five lineages (Figure 2, gray area). Due to the lower variation of cpDNA, this polytomy might be resolvable with additional data. On the other hand, adding new plastid (thus linked) regions may also reinforce the pattern. Based on the available information, the divergence of major lineages that comprised the majority of the taxa in the cpDNA and *sqs* trees might have occurred in rapid succession, an important prerequisite for ILS.

The splitting or merging of ETS species groups (indicated by colors in the cpDNA or *sqs* trees, Figures 2 and 3) are obvious incongruences among the three markers that are unlikely to result from hybridization, if we exclude all individuals that show even the slightest indication of potential hybrid origin. These are taxon names followed by 'H' (previously inferred hybrid origin) in both trees, and additionally, in the *sqs* tree (Figure 3), *H. transylvanicum* (Clade 14), *H. pannosum* (Clade 3), *H. pictum* (Clade 4) because of new putative hybrid status and *H. sparsum* whose occurrence in Clade 14 is an artifact caused by intragenomic recombinant sequences (see above). A few particularly interesting cases of split or merged groups shall be discussed here in more detail. (i) All EU individuals shared the same ETS ribotype and cpDNA haplotype. However, the *sqs* alleles of five individuals of three species fell into Clade 11 (Figure 3, brown) and the alleles of two other individuals of two species fell into Clades 4 and 15, whose sequences are rather similar to one another, but very divergent from those of Clade 11. Furthermore, two individuals of the same species (*H. virosum*) occurred in the divergent clades. These *sqs* alleles do not follow any conceivable pattern and also do not match the geographic distribution of the species or the sampling localities of the individuals (Supplementary Table S1). (ii) Endemic diploids (*H. tomentosum* from the western Alps, *H. porrifolium* from the southeastern Alps) co-occurred in cpDNA haplogroup D. Other endemic diploids (*H. petrovae* and *H. kittanae* from the Balkan Peninsula, *H. stelligerum* from a small relict area in southern France), co-occurred in haplogroup B. *Sqs* Clades 5 and 6, whose sequences are rather similar, contained the same three species. Individuals bearing introgressed alleles should be geographically close to individuals or species from which the allele is derived (Rieseberg, 1998). The complete lack of geographic pattern in these and other cases is a strong argument against hybridization.

Hybridization and ILS can result in exactly the same patterns in a tree (Joly *et al.*, 2006), but ILS predates the speciation event, and hybridization follows it. Naturally, the relative timing is typically unknown so that some uncertainty will always be attached to the assumption which process is more likely to explain the incongruent patterns. If ILS affects mostly the basal lineages as it seems to be the case in the cpDNA and *sqs* trees, and speciation is, according to the combined evidence of all markers, recent in *Hieracium*, ILS is the most likely explanation for the strongly incongruent patterns.

Potentials and limitations of the different markers

Multicopy nuclear marker, ETS (summarized according to Fehrer *et al.*, 2009)—a prominent failure of concerted evolution, even in strictly outcrossing diploids, allowed inference up to four distinct ribotypes per individual. This made this marker very powerful for detecting hybrid origins in *Hieracium*, but it also created a high level of individual-specific noise. The strong correlation of two major species clades with geographic origin and genome size, and species subgroups that 'made sense' in the light of other information, suggested biological relevance of the phylogenetic inference. Therefore, the ETS tree is considered as a good approximation of the species tree. A disadvantage of the marker was the poor resolution of many interspecific relationships.

Low-copy nuclear marker, *sqs*—the main advantage was the high level of variation. Nevertheless, many close relationships remained unresolved due to massive clustering of very similar alleles and very low differentiation within well-supported clades, which is in keeping with recent speciation. Some intragenomic recombinant alleles caused problems in phylogenetic analyses, but provided interesting insights into the molecular evolution of the marker, and excluding them would result in the loss of important information about origins of individuals or entire taxa. *Sqs* provided many insights about hybrid origin (including some novel evidence), but was generally less informative than ETS to detect reticulation. *Sqs* also showed strong indications of ancestral polymorphism and ILS, which allowed new insights into evolutionary patterns in *Hieracium*, but it also made this marker unsuitable to infer phylogenetic relationships.

CpDNA (*trn*T-*trn*L, *trn*V-*ndh*C)—even though sequences of two of the most variable intergenic spacers in plants (Shaw *et al.*, 2007) were combined, the level of variation was still rather low. CpDNA reliably revealed the maternal parent of hybrids whose parental lineages were identifiable by ETS. Species compositions of haplogroups and tree structure were in conflict with other markers and showed strong indications for ILS that affected basal lineages. Therefore, this marker also did not reflect the species tree.

Given the specific characteristics of each marker, when dealing with closely related species or (as in this case) a recently evolved genus, it is not unusual for the origin of some alleles of a gene to predate the speciation events in question. If such is the case, one cannot expect a perfect match between genealogy of the gene and species phylogeny.

A hypothetical evolutionary scenario

We attempt to outline the putative evolutionary history of *Hieracium* by incorporating all available (including conflicting) evidence. The ETS tree will serve as a reference for the reconstruction and relative sequence of events.

The earliest stages of the evolution of the group were already affected by reticulation among related genera. If we consider the split of *Hieracium* into species with eastern or western origin as a landmark, at least those major cpDNA haplogroups and *sqs* clades, in which eastern and western taxa co-occurred, must have predated the divergence of the major ETS ribotypes. The same applies to the species groups EU, Epo, EB and W, because alleles of each group occurred in 2–3 divergent *sqs* clades, and W and EB also showed 2–3 cpDNA lineages each. We consider ILS as the main process leading to incongruence among the datasets at this level. Survival of relatively few diploids in different glacial refugia led to population bottlenecks and divergence of major ETS ribotypes under allopatric conditions. Subsequent explosive speciation occurred within both groups, accompanied by little further ribotype, haplotype and *sqs* divergence. At least some of the crown group lineages recognizable in all three datasets must have existed before the bottleneck (for example, haplogroups C and E; *sqs* Clade 1a), but their branch structures also suggest relatively late speciation. This adaptive radiation, still at the diploid level, may have been accelerated by the retreat of glaciers and the availability of new habitats. The major hybridization events occurred when the isolated species groups came into secondary contact. Merging of the divergent genomes (meanwhile distinct in genome size) may have distorted regular meiosis and resulted in polyploidization. Origin of polyploidy after speciation at the diploid level is likely, because diploids (independent of hybrid origin) account for almost the entire genetic variation in all datasets. Most *Hieracium* populations are triploid, and the emergence of apomixis by which the triploids maintained the ability to reproduce may be directly linked to polyploidization. Once apomictic, the reproductively isolated lineages propagated fixed genotypes and were able to rapidly colonize deglaciated areas. This resulted in the typical pattern of cytotype distribution, which has been described as geographical parthenogenesis (Hörandl, 2009).

Conclusions and outlook

Gene tree incongruence is sometimes considered as a mere nuisance for inferring species phylogenies. However, each gene reflects some aspects of the speciation process, and attempts should be made to better understand this process by integrating conflicting data rather than discard valuable hints about the organismal history. Combined information from single- and multi-copy nuclear and chloroplast markers with their specific strengths and limitations helped to develop a much better understanding of speciation processes and reticulation patterns in the predominantly polyploid apomictic genus *Hieracium*. Geographic distribution, degree of endemism, ploidy, and genome size provided a framework against which the molecular evidence could be evaluated.

Although more genes may not necessarily improve the inference, we plan to apply two additional LCNMs developed by Krak *et al.* (2012) to further assess these patterns. Inclusion of six recently collected diploid taxa will complete the sampling of known diploids.

DATA ARCHIVING

Sequence data have been submitted to GenBank: accession numbers JX129534-JX129745.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The Czech Science Foundation (206/05/0657, P506/10/1363), European Union (SYNTHESYS ES-TAF-1365), Academy of Sciences of the Czech Republic (RVO 67985939), and Ministry of Education, Youth, and Sports of the Czech Republic (0021620828) provided financial support. We thank Barbara Mable and two anonymous reviewers for suggestions to the manuscript, and John Bailey for language corrections.

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

Arnold ML (1997). *Natural Hybridization and Evolution*. Oxford University Press: New York.

Asker SE, Jerling L (1992). *Apomixis in Plants*. CRC Press: Boca Raton, Florida.

Brysting AK, Oxelman B, Huber KT, Moulton V, Brochmann C (2007). Untangling complex histories of genome mergings in high polyploids. *Syst Biol* **56**: 467–476.

Campbell CS, Wojciechowski MF, Baldwin BG, Alice LA, Donoghue MJ (1997). Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Mol Biol Evol* **14**: 81–90.

Chrtek J, Mráz P, Sennikov AN (2006). *Hieracium grofae*—a rediscovered diploid hybrid from the Ukrainian Carpathians. *Biologia* **61**: 365–373.

Chrtek J, Zahradníček J, Krak K, Fehrer J (2009). Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups. *Ann Bot* **104**: 161–178.

Degnan JH, Rosenberg NA (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol* **24**: 332–340.

Degnan JH, Salter LA (2005). Gene tree distributions under the coalescent process. *Evolution* **59**: 24–37.

Doyle JJ, Doyle JL, Rauscher JT, Brown AHD (2003). Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). *New Phytol* **161**: 121–132.

Evans RC, Campbell CS (2002). The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *Amer J Bot* **89**: 1478–1484.

Fehrer J, Gemeinholzer B, Chrtek J, Bräutigam S (2007). Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). *Mol Phylogenet Evol* **42**: 347–361.

Fehrer J, Krak K, Chrtek J (2009). Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise. *BMC Evol Biol* **9**: 239.

Feliner GN, Rosselló JA (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol Phylogenet Evol* **44**: 911–919.

Funk DJ, Omland KE (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst* **34**: 397–423.

Grusz AL, Windham MD, Pryer KM (2009). Deciphering the origins of apomictic polyploids in the *Cheilanthes yavapensis* complex (Pteridaceae). *Amer J Bot* **96**: 1636–1645.

Hall TA (1999). BioEdit, a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Series* **41**: 95–98.

Holder MT, Anderson JA, Holloway AK (2001). Difficulties in detecting hybridization. *Syst Biol* **50**: 978–982.

Hörandl E (2009). Geographical parthenogenesis: opportunities for asexuality. In: Schön I, Martens K, van Dijk P (eds). *Lost Sex—The Evolutionary Biology of Parthenogenesis*. Springer: Dordrecht, Heidelberg, London, New York, pp 161–186.

Huson DH, Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Joly S, Starr JR, Lewis WH, Bruneau A (2006). Polyploid and hybrid evolution in roses east of the Rocky Mountains. *Amer J Bot* **93**: 412–425.

Kaplan Z, Fehrer J (2007). Molecular evidence for a natural primary triple hybrid in plants revealed from direct sequencing. *Ann Bot* **99**: 1213–1222.

Krak K, Álvarez I, Caklová P, Costa A, Chrtek J, Fehrer J (2012). Development of novel low-copy nuclear markers for Hieraciinae (Asteraceae) and their perspective for other tribes. *Amer J Bot* **99**: e74–e77.

Linder CR, Rieseberg LH (2004). Reconstructing patterns of reticulate evolution in plants. *Amer J Bot* **91**: 1700–1708.

Lo EYY, Stefanovic S, Dickinson TA (2010). Reconstructing reticulation history in a phylogenetic framework and potential of allopatric speciation driven by polyploidy in an agamic complex in *Crataegus* (Rosaceae). *Evolution* **64**: 3593–3608.

Merxmüller H (1975). Diploide Hieracien. *Anales del Instituto Botánico A. J. Cavanilles* **32**: 189–196.

Mráz P, Chrtek J, jun, Fehrer J, Plačková I (2005). Rare recent natural hybridization in the genus *Hieracium* s.str.—evidence from morphology, allozymes and chloroplast DNA. *Plant Syst Evol* **255**: 177–192.

Mráz P, Paule J (2006). Experimental hybridization in the genus *Hieracium* s. str.: crosses between diploid taxa. *Preslia* **78**: 1–26.

Müller K (2005). SeqState: Primer design and sequence statistics for phylogenetic DNA datasets. *Appl Bioinf* **4**: 65–69.

Nylander JAA (2004). *MrModeltest v2. Program Distributed by the Author.* Evolutionary Biology Centre, Uppsala University.

Olmstead RG, Palmer JD (1994). Chloroplast DNA systematics: a review of methods and data analysis. *Amer J Bot* **81**: 1205–1224.

Rieseberg LH (1998). Molecular ecology of hybridization. In: Carvalho GR (ed). *Advances in Molecular Ecology.* IOS Press: Burke, Virginia, pp 243–265.

Rieseberg LH, Willis JH (2007). Plant speciation. *Science* **317**: 910–914.

Rokas A, Carroll SB (2006). Bushes in the tree of life. *PLoS Biol* **4**: 1899–1904.

Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Russell A, Samuel R, Klenja V, Barfuss MHJ, Rupp B, Chase MW (2010). Reticulate evolution in diploid and tetraploid species of *Polystachya* (Orchidaceae) as shown by plastid DNA sequences and low-copy nuclear genes. *Ann Bot* **106**: 37–56.

Sang T (2002). Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit Rev Biochem Mol Biol* **37**: 121–147.

Schuhwerk F (1996). Published chromosome counts in *Hieracium.* http://www.botanischestaatssammlung.de/projects/chrzlit.html.

Shaw J, Lickey EB, Schilling EE, Small RL (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Amer J Bot* **94**: 275–288.

Silvestro D, Michalak I (2010). *raxmlGUI: a graphical front-end for RAxML*. Available at http://sourceforge.net/projects/raxmlgui/.

Simmons MP, Ochoterena H (2000). Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol* **49**: 369–381.

Small RL, Cronn RC, Wendel JF (2004). Use of nuclear genes for phylogenetic reconstruction. *Aust Syst Bot* **17**: 145–170.

Speksnijder AGCL, Kowalchuk GA, De Jong S, Kline E, Stephen JR, Laanbroek HJ (2001). Microvariant artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Appl Environ Microbiol* **67**: 469–472.

Stace CA (1998). Sectional names in the genus *Hieracium* (Asteraceae) sensu stricto. *Edinb J Bot* **55**: 469–472.

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Swofford DL (2002). *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods).* Version 4 Sinauer: Sunderland, Massachusetts.

Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P *et al*. (1994). Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Syst Biol* **43**: 250–261.

Whitfield JB, Lockhart PJ (2007). Deciphering ancient rapid radiations. *Trends Ecol Evol* **22**: 258–265.

Zahn KH (1921–1923). *Hieracium* L. In: Engler HGA (ed). *Das Pflanzenreich IV(280). Compositae—Hieracium*. Wilhelm Engelmann: Leipzig. Vol. 76, pp 1–32.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)