Charles University in Prague, Faculty of Science
Department of Botany

Molecular phylogeny and evolutionary trends in *Hieracium* (Asteraceae, Lactuceae)

Molekulární fylogeneze a evoluční trendy v rodě *Hieracium* (Asteraceae, Lactuceae)

Karol Krak

Ph.D. thesis
Prague, May 2012

Supervised by: Dr. Judith Fehrer

Content

Hereby I declare that I made this thesis independently, using the mentioned references. I have not submitted or presented any part of this thesis for any other degree or diploma.

Karol Krak (in Prague 18. 5. 2012)

Author contribution statement
I declare that I made substantial part of the thesis papers (manuscripts). Contribution to each paper is specified below:

1. Judith Fehrer, Karol Krak, Jindřich Chrtek: **Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise.**
Molecular data acquisition, participation on data analyses and manuscript preparation;
Total contribution 40%

2. Jindřich Chrtek, Jaroslav Zahradníček, Karol Krak, Judith Fehrer: **Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups.**
Molecular data acquisition and substantial contribution to data analysis;
Total contribution: 25%

3. Karol Krak, Ines Álvarez, Petra Caklová, Andrea Costa, Jindřich Chrtek, Judith Fehrer: **Development of novel low-copy nuclear markers for Hieraciinae (Asteraceae) and their perspective for other tribes.**
Study design, lab work, data analyses and manuscript preparation;
Total contribution: 80%

4. Karol Krak, Petra Caklová, Jindřich Chrtek, Judith Fehrer: **Reticulate evolution and lineage sorting in *Hieracium* s. str. (Asteraceae): evidence from low-copy nuclear gene and cpDNA**
Lab work, data analyses and interpretation, manuscript preparation;
Total contribution: 80%

Jindřich Chrtek **.....**

Petra Caklová **.....**

Judith Fehrer **.....**

Acknowledgements

I would like to thank to Judith Fehrer for sharing her knowledge on molecular systematics, fruitful discussions and thorough supervision of my work. Jindřich Chrtek, my consultant, was always around, ready to help inconspicuously. His deep knowledge of *Hieracium* taxonomy was essential for successful completion of the thesis. Petra Caklová supported me continuously during those long years, not only as a colleague with substantial contribution to the labwork, but as well as a good friend. I thank Jaroslav Zahradníček for sharing his data on genome size, Siegfried Bräutigam and Franz Schuhwerk for their help with determination of plant material. Patrik Mráz and Pavel Trávníček are acknowledged for valuable comments on the fourth manuscript. I won't be able to submit the thesis without the help of Petr Vít and Tomáš Urfus: they patiently corrected the countless typing errors and provided excellent technical assistance during the finalization.
I would like to express my thanks to my family and Martina for their love, patience and encouragements. With them even the worst working day turned to a nice evening.
I'm grateful to all those that provided plant material: S. Bräutigam, F. Krahulec, A. Krahulcová, G. Mateo, P. Mráz, M. Niketić, Z. Szeląg, P. Trávníček, T. Tyler, V. Vladimirov, J. Zahradníček.
The work was supported by the Czech Science Foundation (Projects No. 206/05/0657 and P506/10/1363) and SYNTHESYS program of the EU (Project No ES-TAF-1365).

Summary

The hawkweed subgenus *Hieracium* s. str. is notoriously known for its extreme morphological variability and variation in ploidy levels that is associated with differences in modes of reproduction. Extensive past hybridization is supposed for the subgenus, but recent hybridization was evidenced only in few cases. The subgenus attracts the attention of botanists already for more than a century. Therefore the species diversity is largely examined and the taxonomy of the subgenus is well elaborated, although several contradictory taxonomic concepts exist. However the relationships among the species are unknown and haven't been studied yet. The investigation of these relationships from a phylogenetic perspective using molecular approaches was the main aim of the presented thesis. Basic species (both diploid and polyploid), representing morphologically unique taxa, that are supposed to be the basic evolutionary units of the subgenus were studied. The sequences of two intergenic spacers of the cpDNA (*trn*T-*trn*L and *trn*V-*ndh*C) and the external transcribed spacer of the nuclear ribosomal DNA (nrDNA ETS) were analyzed. Moreover, three new low-copy nuclear markers with higher variability than nrDNA and cpDNA markers were developed and their suitability for phylogenetic studies in *Hieracium* s. str. was evaluated. One of these, a part of the gene coding for squalene synthase (*sqs*) was included to the study as the second nuclear marker. Based on ETS, two previously unrecognized major evolutionary lineages were found, that correlated well with the geographic origin and distribution of the species. The observed pattern in genome size variation was in high correlation with this phylogenetic pattern. On the other hand, the cpDNA and *sqs* phylogenies were incongruent with the ETS, in respect of these major groups. However, they reflected well the subgroups identified within the major ETS clades. This pattern together with the low interspecific variability of the ETS indicates that *Hieracium* s. str. is a group with rather recent speciation. The level of interspecific hybridization found greatly exceeded the previous expectations: almost half of the studied accessions was evidenced to be of hybrid origin. Moreover, hybrid origin was inferred not only for the polyploids, but it concerned diploids as well. Genetic material representing most probably already extinct parental lineages was identified in some of the hybrid accessions. Hybridization in the subgenus is probably even more abundant as described above, but its detection was obscured by population level (genetic drift, incomplete lineage sorting) or intragenomic (concerted evolution) processes.

Introduction

Driving forces of plant evolution

Fundamental sources of variation in living organisms are associated with structural changes in genetic material – mutations. Due to these changes, new alleles related to novel functional traits arise in populations. Hence, mutations together with gene flow (migration), genetic drift and natural selection are acting as engines of speciation, the process of novel species formation. However, in the plant kingdom, speciation is shaped by additional processes more extensively than in other groups of organisms. These factors include hybridization, polyploidization and apomixis.

Hybridization

Hybridization is defined as crossing between individuals belonging to different species or genetically divergent populations within a species (Soltis & Soltis 2009). Despite recognised already by Linné (e.g. Rieseberg 1997), the role of hybridization in plant evolution was underestimated for a long time (Arnold 1997). Nowadays is already known that hybridization may have various evolutionary consequences including the increase of intraspecific genetic diversity, origin of genetic adaptation as well as origin of new ecotypes or species (Rieseberg 1997).

Hybridization in higher plants is quite widespread. About 25% of all plant species are known to hybridize at least with one other species. This number may be underestimated, because hybrid origin is often challenging to prove (Mallet 2007).

Interspecific hybridization introduces evolutionary novelties by combinations of traits typical for parental species, developed in separation, into a single individual. Due to the contribution of previously isolated (and therefore diverse) genotypes, confirmed hybrids were supposed to have higher genetic diversity compared to their parents. However, according to recent studies, considerable variation in this aspect exists among particular hybrid species (Rieseberg 1997). In natural populations, newly arising hybrids are often threatened by extinction. They are markedly distinct from their parents; posses lower fitness, lack an adaptive history to any ecological niche and suffer from sterility caused by chromosomal incompatibilities (Mallet 2007). On the other hand, hybrids often exhibit novel or extreme characters compared to their parents, referred to as transgressive segregation (Rieseberg 1997, Seehausen 2004). These changes can enable hybrids to utilise novel niches (e.g., to colonize habitats not occupied by the parental taxa; Seehausen 2004) and allow selection to act in favour of their establishment. Similarly, the sterility of hybrids can be overcome by various mechanisms involving polyploidization or apomixis (see below).

Polyploidy

Polyploidization (whole genome duplication) is another substantial mechanism of plant speciation. Its evolutionary importance is underlined by the fact that ancient polyploidization was evidenced in almost all plant lineages (Jiao et al. 2011, Soltis & Soltis 2009). The increasing attention of biologists in the last few decades is going hand in hand with the development of new karyological, cytogenetical and molecular methods enabling the study of mechanisms, patterns and processes behind polyploid formation.

The most common mechanism of polyploid origin is likely through unreduced gametes (Ramsey & Schemske 1998, Thompson & Lumaret 1992, Harlan & De Wet 1975). Polyploids may arise directly by fusion of two unreduced gametes, or by fusion of one reduced and one unreduced gamete. The latter scenario requires two evolutionary steps: at first, a triploid lineage is established, and secondly, higher polyploids are formed from the triploids (triploid bridge) by backcrosses to diploids or crosses between triploids. This process is complicated by a high sterility of triploid seeds (so-called 'triploid block'; Levin 1975, Ramsey & Schemske 1998) and lower fertility of triploids. In contrast, a crucial role of triploids in the formation of higher polyploids was proved, e.g., in *Chamerion angustifolium* (Burton & Husband 2001) and *Hieracium echioides* (Peckert & Chrtek 2006). The second major mechanism concerns hybridization between intersterile diploid taxa and subsequent polyploidization of the hybrids (e.g. Mallet 2007).

Two types of polyploidy are traditionally recognised based on the origin of the duplicated genome. Autopolyploidy is caused by a duplication of a chromosome set within a single plant or by crossing of different individuals within a species. This type of polyploidy is considered to be less frequent. However, its detection is often difficult, therefore the current assumption on autopolyploidy in plants is likely to be underestimated (Grant 1981, Soltis & Rieseberg 1986, Soltis et al. 2007). Allopolyploidy is a merging of two or more distantly related genomes (e.g., belonging to different species) into a single one. Hence allopolyploidy is one of the means of hybrid speciation, and it appears to be much more common than homoploid hybridization (hybridization where the parents and their hybrid have the same ploidy level) in nature (Rieseberg & Willis 2007). As indicated by this abundance, polyploidy has several positive effects on establishment and stabilization of hybrids. Diploid hybrids often suffer from sterility. Their nuclei harbour two structurally distinct sets of parental chromosomes that are unable to pair with each other during meiosis which results in the development of malfunctioning gametes (Grant 1981, Ramsey & Schemske 1998). On the contrary, in the allopolyploid genome, the parental chromosome sets are duplicated and pairing of homologous chromosomes is possible. For many hybrids, polyploidy therefore represents a way to escape from sterility (Ramsey & Schemske 1998). Due to genome duplication, polyploidization can prevent the back-crossing of the hybrids to their parents (Rieseberg & Willis 2007, Soltis & Soltis 2009). Such a reproduction barrier is very efficient as it develops instantaneously and persists even in case of sympatric occurrence of the parental taxa with their hybrids.

Apomixis

Apomixis is defined as asexual production of maternal progeny through seeds (Asker & Jerling 1992, Savidan 2007). Three essential processes are characteristic for this mode of reproduction: the avoidance of reductional meiosis in embryo sac formation, development of the unfertilized egg cell, and the capability to either produce endosperm autonomously or fertilization-dependent (Koltunow 1993, Richards 2003, Bicknell & Koltunow 2004). Basically two major types of apomixis can be distinguished based on the mode of embryo sac formation: (i) diplospory, when the embryo sac is produced from the megaspore mother cell, like in sexual reproduction, but meiosis is avoided in the first cell division and unreduced megaspores are formed by mitosis; (ii) apospory, characterized by formation of the embryo sac from a somatic cell, usually from a nucellar cell.

The occurrence of apomixis in flowering plants is not random; it certainly prevails in some families (Poaceae, Rosaceae, Asteraceae; Richards 2003). Due to the above-mentioned variation at the developmental level, it is hypothesized that it has evolved several times independently even within these families (Richards 2003, Savidan 2007). Apomixis is closely related to polyploidy and hybridization. It has been described for 126 plant genera (Carman 1997), and almost all known apomicts are polyploids. They usually possess high heterozygozity and form agamic complexes with their diploid sexual relatives (Savidan 2007). From the biosystematic point of view, apomicts represent difficultly classifiable units and are the focus of debates on taxonomic treatments and species delimitation (recognition of large agamospecies *vs.* narrow microspecies, etc.; see e.g., Hörandl 1998, Stace 1998).

Apomicts are particularly associated with arctic and boreal conditions, suggesting that they are more successfull colonizers of remote areas, deglaciated after the last ice age, than outcrossing sexuals. Their advantage may stem from the ability to establish from single disseminules and a lack of homozygosity and imbreeding depression (Mogie & Ford 1988).

Obligatory apomixis is restricted only to some diplosporous species (Richards 2003), whereas the vast majority of apomicts is facultative and capable to reproduce sexually to a certain degree. Besides, even those species where sexual reproduction was not evidenced are often able to produce functional pollen (Mogie 1992) and could be involved in crosses with their sexual relatives as paternal parents. Due to these features, apomicts should not be underestimated as 'dead ends of evolution', but considered as potential sources of variation in microevolutionary processes.

Molecular markers and plant evolution

The study of plant evolution has largely benefited from the use of molecular biological approaches that facilitate the study of genetic information contained in the molecules of DNA or proteins. Molecular data have several advantages in comparison with other data types. They are universal – each living organism possesses DNA and proteins so that the inference of evolutionary relationships even among distantly related groups of organisms, where only a limited number of comparable morphological characters exists, is feasible. Furthermore, molecular approaches generate an immense number of characters for comparison even for organisms where other types of characters are sparse (e.g., if morphology is highly reduced). Molecular characters are discrete, and their particular states are well-defined and easily distinguishable from each other – DNA sequences are composed of four nucleotide bases; protein sequences are composed of 21 amino acids; the study of tandemly repeated DNA regions is based on the existence of alleles containing exact numbers of repetitions; polymorphism in restriction site analysis is due to the presence or absence of a position recognized by a particular endonuclease, etc. In contrast, the analysis of other characters, e.g. morphological or chemical, may be less straightforward as they are often not discrete, or are semiquantitative. Genetic processes are essential for evolution – DNA- or protein-based studies facilitate to incorporate information directly related to these processes. There is no doubt that morphology and physiology of living organisms have a genetic basis, but the particular phenotypes are dependent on other factors as well (e.g., they may be affected by environmental conditions or interactions at the genomic or epigenetic level) that may complicate the interpretation of these data from an evolutionary point of view.

A variety of well established molecular approaches is available for the purpose of molecular phylogenetic studies. Among these, sequencing of DNA is the most widely used. Studies focusing on different taxonomic levels require markers with different evolutionary rates. Species level phylogenies that aim to resolve relationships among closely related taxa require markers with high evolutionary rates. Therefore, non-coding regions such as intergenic spacers and introns are usually targeted in these kinds of studies. Both, the organellar as well as the nuclear genomes offer suitable candidates for these purposes. However, each of these markers has its own advantages and drawbacks stemming from their mode of inheritance and evolutionary dynamics.

The most widely applied molecular marker in plant phylogenetics that has also been used for the longest time is the chloroplast DNA (cpDNA). One of the main features of the cpDNA is its uniparental inheritance. In higher plants, cpDNA is inherited maternally in most cases, but strict paternal (e.g. in gymnosperms; Nealy et al. 1989) or biparental inheritance (e.g., in *Silene vulgaris*; McCauly et al. 2007, or *Helianthus verticillatus;* Ellis et al. 2008) was already described for some groups. Furthermore, cpDNA is haploid and non-recombinant which eliminates intraindividual variation, reduces the amount of intraspecific and intrapopulation variability and decreases the fixation time of cpDNA haplotypes in comparison with diploid and recombining genomes (Birky 1988). Although the substitution rates of the chloroplast genome are lower in comparison with the nuclear genome (Olmstead & Palmer 1994, Drouin et al. 2008), several highly variable regions, suitable for phylogenies at species level or even below, have been described (Taberlet et al. 1991, Small et al. 1998, Shaw et al. 2005, 2007). Moreover, the chloroplast genome is structurally conserved so that the arrangement of genes and their order is highly similar even among distantly related taxa (Olmstead & Palmer 1994), facilitating the development of universal primers flanking the variable introns and intergenic spacers (Taberlet et al. 1991, Shaw et al. 2005, 2007). All these attributes, together with the high abundance in the plant organism, makes cpDNA an extremely user-friendly research tool.

However, the particular drawbacks of cpDNA stem from the same features as its advantages. In case of hybrid origin of some of the investigated taxa, only one parental lineage will be recovered by cpDNA due to its uniparental inheritance. This marker will fail to recover the other parent of the putative hybrids, and even worse, the fact that the particular taxa originated via hybridization will remain undetected, obscuring an important process of the speciation of the group studied (Rieseberg & Soltis 1991). Haploidy, uniparental inheritance and the lack of recombination also have the consequence that cpDNA is transmitted to the next generation in one ‘piece’ (as one locus). Therefore, even if sevaral non-coding cpDNA regions are used, they cannot be considered as independent sources of phylogenetic information (Olmstead & Palmer 1994). Due to introgression or ancient hybridization, often, geographic patterns are reflected by cpDNA data rather than the phylogeny of the studied species (Whittemore & Schaal 1991). This so-called ‘chloroplast capture’ is facilitated by the relatively short fixation times of the cpDNA haplotypes (Birky 1988, Rieseberg & Soltis 1991) and by the fact that cpDNA genomes are being exchanged more often than the nuclear genes between species (*Quercus*: Whitemore & Schaal 1991, *Helianthus*: Dorado et al. 1992, reviewed in Rieseberg & Soltis 1991, Doebley 1989).

Sequences of the nuclear genome have the potential to counterbalance the main disadvantages of the cpDNA: they are biparentally inherited and have higher rates of evolution (Small et al. 2004). The most widely used nuclear marker for phylogenies at low taxonomic level is the internal transcribed spacer of the nuclear

ribosomal DNA (nrDNA ITS). Its popularity was demonstrated in the review by Álvarez & Wendel (2003), according to which ITS has been used in 66% of plant phylogenetic studies published between 1998 and 2003. This popularity has several reasons: (i) nrDNA is a multi-copy marker that is present in the genome in several hundreds to thousands of copies arranged in one or more tandemly repeated arrays, therefore it is easy to amplify; (ii) the occurrence of highly conservative coding regions adjacent to ITS enabled the design of universal primers (White et al. 1990), and (iii) it is realatively short (500-700 bp) and interrupted in the middle by another conserved region (5.8S rDNA) suitable for primer design. That, hand in hand with its high abundance in the genome, facilitates amplification even from partly degraded or ancient material like herbarium specimens (Blattner 1999, Álvarez & Wendel 2003).

Like other gene families, nrDNA regions are subjected to concerted evolution (Elder & Turner 1995). This process tends to homogenize the variation within the gene family, accumulated by independent mutations, due to mechanisms termed unequal crossing-over and high-frequency gene conversion, so that all (or nearly all) repeat copies within a genome (as well as within a species) are often genetically uniform. This process was long believed to be the main feature of nrDNA (Baldwin et al. 1995, Elder & Turner 1995). Recent studies revealed that this mechanism does not operate as absolutely as had been thought. Concerted evolution can be unfinished or fail completely in case of some hybrids or allopolyploids. Such information can even be of advantage with respect to the evolutionary history of hybrids as both parental linages may still be recovered from the hybrid genome. However, the situation is more complex. The evolutionary fate of the parental copies in the hybrid genome can vary from case to case. As reviewed in Álvarez & Wendel (2003), three scenarios may occur: (i) the parental copies may persist unchanged; (ii) the parental copies may recombine to various degree, and a mixture of chimeric and unrecombined parental sequences can be sampled, and (iii) one of the parental copies can be eliminated completely.

The factors affecting the patterns of concerted evolution are still not well understood. It seems that the number and location of the nrDNA arrays may be important (Wendel et al. 1995, Álvarez & Wendel 2003). Furthermore, the mode of reproduction (Dover 1989), generation time, and the time since hybridization (Álvarez & Wendel 2003) may have an impact on this process. Accordingly, in asexually reproducing taxa (meiosis is considered essential for recombination) or in taxa with long generation times as well as in recent hybrids, completely or almost completely unhomogenized parental sequences should be present (Álvarez & Wendel 2003, Calonje et al. 2009). However, already Dover (1989) noted that concerted evolution can be promoted even via mitosis, but with slower rates than during meiosis. Despite this, unhomogenized sequences are often found in recent or younger hybrids (e.g., *Tragopogon*: Soltis & Soltis 1991, Soltis et al. 1995; *Nicotiana rustica*: Kovařík et al. 2008), and ancient intraindividual variation has been observed as well (Winteraceae: Suh et al. 1993). On the other hand, nearly complete homogenization was observed in recent hybrids even after a few generations (*Armeria*: Aguilar et al. 1999).

Besides the aforementioned effects, another important fact about nrDNA should be kept in mind when this marker is intended to be used. In the plant genome, more than one nrDNA locus could be present. The number of these loci may even differ among closely related taxa. And they may be different in respect of their origin as well. Some of these loci may be homologous across the studied set of taxa in which case their sequence variation will represent the evolutionary relationships among the organisms from which they were isolated (so-called orthologous sequences). On the other hand, in some of the studied genomes, some of these loci

may become duplicated. In such a case, sequences are not reflecting the speciation process, but the duplication event, and their similarity with sequences from other taxa is not orthologous, but paralogous. The distinction between paralogy and orthology is never easy, and in synergy with concerted evolution, it may become almost impossible.

According to the above-mentioned experiences, concerted evolution is a highly unpredictable process that could seriously affect the reconstruction of phylogenetic relationships. The inclusion of paralogs (or pseudogenes) into phylogenetic analysis may result in erroneous estimates of relationships or incongruence (Doyle 1992). Similarly, recombinant sequences will be placed in a basal position to either of the parents and may cause a decrease in bootstrap support for the clades containing the parental sequences (McDade 1992). And finally, completed homogenization towards one of the parental sequences will eliminate important evidence on hybrid origin or speciation (Baldwin et al. 1995, Wendel et al. 1995) in a similar way like cpDNA. Despite all these inconveniences, ITS remains a highly valuable source of phylogenetic information but careful evaluation of the observed patterns is done in order to fully use its potential (Feliner & Rosselló 2007). The principles described here also apply to other nrDNA regions, e.g., the external transcribed spacer (ETS) or the 5S non-transcribed spacer (5S-NTR), but ITS has been much more frequently investigated.

Low-copy nuclear genes (LCNG) represent another set of markers used for low taxonomic level phylogenies. These markers are based in non-coding regions (e.g., introns or UTR regions) of functional protein coding genes. LCNGs are meanwhile being widely used, especially in cases where other markers (cpDNA and nrDNA) have low resolution or give conflicting signals (Small et al. 2004). This is frequently the case, if close relationships or hybrid speciation are being examined.

In comparison with the above-mentioned markers, LCNGs possess several advantages. Like all nuclear markers, LCNGs are biparentally inherited. Because of low copy numbers (in ideal case only a single copy is occurring in the genome), they are much less prone to concerted evolution (Small et al. 2004). Due to the huge number of coding genes present in plant genomes, they represent a potentially unlimited source of markers. Another feature of great importance is that multiple unlinked loci can be used, which are required for independent phylogenetic reconstruction, e.g. in cases when hybridization and reticulation or lineage sorting caused by rapid diversification of polymorphic ancestors have occurred in the evolution of a particular group.

On the other hand, there are several serious limitations that may discourage scientists to use LCNG markers. Their evolutionary dynamics is not uniform; in different plant groups, independent gene duplication and deletion events may occur, and the copy number of a particular gene may be highly variable among these groups. For example, the gene for alcohol dehydrogenase (*Adh)* was proven to be single-copy in most plant species, but in *Gossipium*, it forms a small gene family with 7 paralogs (Small et al. 2004). The variation in copy number is again connected with the orthology/paralogy problem (Sang 2002, Hughes et al. 2004, Small et al. 2004). Often the non-coding intron regions may not exhibit higher variation than the nrDNA (Hughes et al. 2004) and therefore they won't bring additional resolution required for elucidation of close relationships. Moreover, their variation may differ considerably in various groups (Sang 2002, Small et al. 2004), likewise, the overall variability of different LCNGs within the same sample set may differ (Senchina et al. 2003, Hughes et al. 2004). Due to the low copy number, these markers may be susceptible to a greater extent to population genetic processes (e.g., genetic drift, selection),

eliminating alleles from the studied genomes or causing allelic non-monophyly (incomplete lineage sorting) within the species in comparison with nrDNA (Linder & Rieseberg 2004). According to the described features, LCNGs should be treated as lineage-specific rather than as universal markers.

Despite all inconveniences listed above, the potential of LCNG markers will most probably stay unreduced. However, additional effort in terms of time and finances has to be invested to perform pilot studies aiming at revealing the evolutionary dynamics and phylogenetic signal of the candidate loci (Small et al. 2004). Pilot studies should be based on a small number of (preferably diploid) individuals that cover the genetic variation of the studied group as comprehensively as possible. The existence of reference phylogenies based on nrDNA or cpDNA can further contribute to assist data analyses and interpretation. Based on such a study, appropriate gene(s) can be selected that can be applied for phylogenetic investigation, involving detailed sampling of accessions from the studied group.

Hieracium s. l.

Within the family Asteraceae, *Hieracium* L. has been classified as a member of the tribe Lactuceae Cass. (e.g., Bremer 1994).

The genus *Hieracium* s. l. comprises perennial herbs with one to numerous leafless or foliaged stems terminated by one to numerous capitula. Leaves are shaped from entire to deeply dentate or lobed. Involucres are formed by several irregularly imbricate rows of linear to lanceolate bracts. Individual ligules are usually yellow. Achenes are covered with 10-13 ribs of narrowly obconical shape (never beaked), carrying 1 or 2 rowed pappi. The basic chromosome number in the genus is x = 9 (Zahn 1921-1923, 1922-1930, Sell & West 1976, Bräutigam & Schuhwerk 2002, Chrtek 2002, 2004). The genus is distributed in temperate areas of Europe, Asia, and North America, and in Central and South American mountains. Main centres of *Hieracium* diversity are located in Europaean mountain ranges (Alps, Pyrenees, Carpathians, Sudeten Mts., mountains of the Balkan Peninsula), in American mountains and westernmost Asia.

Hieracium s. l. is famous for its high morphological diversity, and the opinions on its taxonomy are almost as diverse as the genus itself: the number of recognized species varies from 500 to almost 10 000 (Willis 1973) among the different taxonomic concepts. Two main directions are traditionally distinguished in 'Hieraciology'. Species concepts elaborated by Scandinavian and British authors (Norrlin 1888, Elfstrand 1893, Norrlin 1912, Pugsley 1948, Sell & West 1976, Sell & Murrell 2006, Tennant & Rich 2008, Tyler 2011) and followed by botanists from the former Soviet countries (Juxip 1960, Shlyakov 1989) are based on the recognition of microspecies, i.e., morphologically and ecologically discrete forms are given species rank. On the contrary, Central European taxonomists (e.g., Nägeli & Peter 1885, Zahn 1921-1923, 1922-1930) are supporting a broader definition of species. Within each species, numerous taxa of lower rank were described. Generally, the subspecies recognised by the Central European botanists are taxonomically equivalent to the species of the Scandinavian and British concepts. According to Sell (1987), the basic difference among these two schools of thought is that: “The Central European botanists believe they can tell how these species originated and the Scandinavian and British botanists do not think this is possible.”

The most comprehensive taxonomic treatment was published by Zahn (1921-23). According to his monograph, the genus includes 756 species further divided into many subspecies (the index of his work contains approximatelly 18,000 names). Based on morphology, he distinguished two classes of species: (i) Basic species (*species principales collectivae*, Hauptarten) that posses unique sets of morphological characters and are therefore well distinguishable from each other, and (ii) Intermediate species (*species intermediae collectivae*, Zwischenarten) that combine characters typical for two or more basic species in their morphology and are supposed to be of hybrid origin.

Zahn divided the genus into four subgenera, namely subgen. *Euhieracium* (= *Hieracium* s.str.), subgen. *Pilosella* (Hill) S.F.Gray, subgen. *Stenotheca* Fries (correct name: subgen. *Chionoracium* Dum.), and subgen. *Mandonia* (Schultz-Bip.) Zahn. The subgenus *Mandonia* is nowadays merged with subgen. *Chionoracium* by most authors. The subgenera markedly differ from each other with respect to morphology, breeding systems, DNA content, geographic distribution, and ecological demands. Furthermore, there is no evidence for recent hybridization among the subgenera, although hybridization is (or was in the past) frequent within the genus (see below). Species of the rather well-defined subgen. *Pilosella* are often treated as a separate genus (Sell & West 1976, Bräutigam & Greuter 2007). Similarly, a separate genus *Stenotheca* Monn. was described and is accepted by some authors (Sennikov & Illarionova 2002).

Hieracium s. str.

Hieracium s. str. comprises perennial herbs with one to few leafless or foliaged stems. Basal leaves form rosettes or may be missing completely. The leaf margin is entire to deeply dentate, and at least the basal leaves are usually distinctly petiolate. The margins of receptacle pits are more or less dentate, sometimes with more or less long fimbriae. Ligules are yellow, sometimes dingy yellow, rarely greenish. Achenes are 2.5-5 mm long, 10-ribbed; ribs are apically confluent in an obscure ring. The pappus hairs are arranged into two unequal series (Sell 1987). The subgenus is distributed in temperate regions of Europe, Asia, northernmost Mediterranean Africa, and North America (and introduced to several other regions, e.g. to New Zealand). The subgenus has a broad ecological amplitude and can be found in diverse kinds of different habitats from lowland meadows through forest and forest margins to alpine highlands (Zahn 1921-1923).

The estimated number of species highly varies depending on the taxonomic concept (Sell 1987). According to Zahn (1921-1923), the subgenus comprises 500 species. Only a small part of the species is diploid; the vast majority of the taxa are tri- and tetraploids (Merxmüller 1975, Schuhwerk 1996, Schuhwerk & Lippert 1998, Chrtek et al. 2007); pentaploids are extremely rare (Chrtek et al. 2004, Tennant & Rich 2008). Aneuploidy is considered to be very rare (for reference see Schuhwerk 1996). Diploids are confined to certain geographic areas that mostly represent never glaciated refuges. They have been reported from Southwest Europe, the Eastern Carpathians and the Balkan Peninsula (e.g. Merxmüller 1975, Chrtek 1996, Mráz 2003, Castro et al. 2007). On the other hand, polyploids are widespread also in areas that have been covered by ice sheets during the last glacial age (Merxmüller 1975). For some taxa, diploid as well as polyploid cytotypes were detected (e.g. Schuhwerk & Lippert 1998, Chrtek et al. 2007).

Variation in reproductive modes was recognized in *Hieracium* s. str. as well. The diploid species (and diploid cytotypes of species, if diploids as well as polyploids were recorded) are considered to be obligatory sexual with regular micro- and macrosporogenesis (Gustafsson 1947) and self-incompatible (Bergman 1941, Rosenberg 1927). However, the presence of heterospecific pollen on stigma can induce the breakdown of self-incompatibility under some circumstances (a process called mentor effect; Mráz 2003; Mráz & Paule 2006). The polyploid species are apomictic. Diplospory of the *Antennaria*-type was described for the subgenus (Bergman 1941, Noyes 2007). So far, there is no evidence for residual sexuality in the polyploid taxa, therefore they are considered to be obligatory apomictic (Mráz & Paule 2006, Zahradníček 2008).

The high number of intermediate species nowadays recognized as distinct taxa could serve as evidence for extensive hybridization occuring in the past. The majority of the diploid sexuals are considered as basic species (Mráz et al. 2005). Therefore, hybridization involving these taxa is most probably the source of the huge morphological variation observed in the subgenus. Hybrid forms arising from these crosses were stabilized by polyploidization and apomixis. On the contrary, recent hybridization in *Hieracium* s. str. is highly restricted. Despite successful crossing experiments among diploids (Mráz & Paule 2006) as well as diploids and polyploids (where the latter acted as pollen donors; Mráz & Tomčíkova 2004), only a very few recent hybrids are found in nature (Chrtek, personal communication), and only two cases have been published so far, namely *Hieracium krasanii* (Mráz et al. 2005, 2011) and *H. grofae* (Chrtek et al. 2006). Moreover these hybrids are often sterile and produce only small amount of viable pollen.

Previous molecular and cytological studies on *Hieracium*

While much attention was paid to examine *Hieracium* species diversity, very little is known about their relationships. Only two molecular phylogenetic studies dealing with *Hieracium* s. l. have been published so far (Fehrer et al. 2007; Gaskin & Wilson 2007). Although both studies include some representatives of *Hieracium* s. str., neither of them was primarily focused on the relationships within this subgenus. The work of Fehrer et al. (2007) used sequences of two cpDNA markers (*trn*T-*trn*L intergenic spacer and *mat*K gene) and the nrDNA ITS to elucidate the relationships within the subgenus *Pilosella*. However, a representation of taxa from the subtribe Hieraciinae (including several species of *Hieracium* s. str.) was sampled that allowed revealing some patterns concerning *Hieracium* s. str. as well. In the ITS analyses, it formed a monophyletic group together with subgen. *Chionoracium*, whereas this relationship was not supported based on the cpDNA data, the reason of which was a chloroplast capture event involving a *Pilosella* lineage. Most of the relationships among the investigated *Hieracium* s. str. remained unresolved due to low genetic variation and apparently very recent speciation. Gaskin & Wilson (2007) investigated the phylogenetic relationships of the native North American *Hieracium* s. str. and *Chionoracium* taxa to the introduced *Pilosella* species based on the *trn*T-*trn*F and *pet*N-*psb*M cpDNA spacers. In concordance with Fehrer et al. 2007, *Hieracium* s. str. taxa were found to be polyphyletic with these markers, and the relationships among them were not completely resolved. The results of these studies suggested that additional and more variable markers have to be used in order to reconstruct the phylogenetic relationships within *Hieracium* s. str. in more detail.

In contrast to molecular genetic approaches, flow cytometry is widely used to investigate *Hieracium* s. l. (Bräutigam & Bräutigam 1996, Suda et al. 2007, Trávniček et al. 2011). Several studies confirmed that variation in genome size exists among *Hieracium* s. str. and *Pilosella* (Suda et al. 2007) as well as within both of these subgenera (Bräutigam & Bräutigam 1996, Suda et al. 2007, Zahradníček 2008). In *Hieracium* s. str., up to 1.22- fold variation of 1Cx values among species was observed (Zahradníček 2008). Not much is known about the origin of this variation, and the lack of information on species relationships precluded its investigation from the phylogenetic perspective. Therefore, the data obtained in the frame of the diploma thesis of J. Zahradníček (the diploma thesis was part of the same research project as the presented PhD. thesis) were used in combination with molecular data to evaluate the significance of the phylogenetic signal of the observed genome size variation. The results of this combined effort are part of both, the diploma thesis of J. Zahradníček as well as the presented PhD. thesis.

Aims of the thesis

The present work is the first attempt to elucidate the interspecific relationships within *Hieracium* s. str. using multiple molecular markers. Due to the extremely high number of recognized species, it is focused on Zahn's (1921-1923) basic species (both diploid and polyploid). This approach is supposed to cover the entire genetic variation and to delimit the major evolutionary lineages within the subgenus and therefore to allow identification of basic patterns and processes of speciation. The concrete aims of the thesis can be summarized as follows:

1.) Identify the basic phylogenetic relationships among the species of *Hieracium* s. str. using cpDNA and nrDNA markers and compare it with the existing sectional classification (Paper 1)

2.) Investigate whether the well documented variation in genome size could be correlated to the subgenus' phylogeny (Paper 2)

3.) Develop highly variable low-copy nuclear markers for further phylogenetic studies of *Hieracium* s. l. and related groups (Paper 3)

4.) Assess the extent of hybridization among the major evolutionary lineages of *Hieracium* s. str. and infer the origin of polyploids and the patterns of speciation using a multigene approach (Paper 4)

.

References

Aguilar J.F., Rossello J.A. & Feliner G.N. (1999): Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). – Molecular Ecology 8: 1341–1346.

Álvarez I. & Wendel J.F. (2003): Ribosomal ITS sequences and plant phylogenetic inference. – Molecular Phylogenetics and Evolution 29: 417–434.

Arnold M. L. (1997): Natural hybridization and evolution. – Oxford University Press, New York.

Asker S.E. & Jerling L. (1992): Apomixis in plants. CRC Press, Boca Raton, Florida.

Baldwin B.G., Sanderson M.J., Porter J.M., Wojciechovski M.F., Campbell C.S. & Donoghue M.J. (1995): The ITS region of the nuclear ribosomal DNA – a valuable source of evidence on angiosperm phylogeny. – Annals of the Missouri Botanical Garden 82: 247–277.

Bergman B. (1941): Studies on the embryo sac mother cell and its development in *Hieracium* subgen *Archihieracium*. – Svensk Botanisk Tidskrift 35: 1–42.

Bicknell R.A. & Koltunow A.M. (2004): Understanding apomixis: Recent advances and remaining conundrums. – Plant Cell 16: 228 –245.

Birky C.W. (1988): Evolution and variation in plant chloroplast and mitochondrial genomes. – In: Gottlieb L.D. & Jain S.K. (eds.), Plant Evolutionary Biology, pp. 23–53. Chapman and Hall, London.

Blattner F.R. (1999): Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. – Biotechniques 27: 1180.

Bräutigam S. & Greuter W. (2007): A new treatment of *Pilosella* for the Euro-Mediterranean flora [Notulae ad floram euro–mediterraneam pertinentes No. 24]. – Willdenowia 37: 123–137.

Bräutigam S. & Schuhwerk F. (2002): *Hieracium* L. – In: Jäger E. J. & Werner K. (eds.), Rothmaler, Exkursionsflora von Deutschland 4. Gefässpflanzen, Kritischer Band, 9. Auflage, pp. 709–734, Spektrum Akademischer Verlag, Heidelberg & Berlin.

Bräutigam S. & Bräutigam E. (1996): Determination of the ploidy level in the genus *Hieracium* subgenus *Pilosella* (Hill) S.F. Gray by flow cytometric DNA analysis. – Folia Geobotanica 31: 315–321.

Bremer K. (1994): Asteraceae: Cladistics and classification. Timber Press, Portland, Oregon.

Burton T.L. & Husband B.C. (2001): Fecundity and offspring ploidy in matings among diploid, triploid and tetraploid *Chamerion angustifolium* (Onagraceae): consequences for tetraploid establishment. – Heredity 87: 573–582.

Calonje M., Martin-Bravo S., Dobes C., Gong W., Jordon-Thaden I., Kiefer C., Kiefer M., Paule J., Schmickl R. & Koch M.A. (2009): Non-coding nuclear DNA markers in phylogenetic reconstruction. – Plant Systematics and Evolution 282: 257–280.

Carman J. G. (1997): Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. – Biological Journal of the Linnean Society 61: 51–94.

Castro M., Mateo G. & Rosseló J.A. (2007): Chromosome numbers in *Hieracium* and *Pilosella* species (Asteraceae) from the Iberian Peninsula and the Balearic Islands. – Botanical Journal of Linnean Society 153: 311–320.

Chrtek J. jun. (1996): Chromosome numbers in selected *Hieracium* species (Compositae) in the Sudeten Mts and West and Ukrainian East Carpathians. – Fragmenta Floristica et Geobotanica 41(2): 783–790.

Chrtek J. jun (2002): *Hieracium* L. – In. Kubát K. (ed.), Klíč ke květeně České republiky, pp. 706–732, Academia, Praha.

Chrtek J. jun (2004): *Hieracium* L. – In: Štěpánek J. & Štěpánková J. (eds.), Květena České republiky 7, pp. 540–701, Academia, Praha.

Chrtek J. jun., Mráz P. & Severa M. (2004): Chromosome numbers in selected species of *Hieracium* s.str. (*Hieracium* subgen. *Hieracium*) in the Western Carpathians. – Preslia 76: 119–139.

Chrtek J., Mráz P. & Sennikov A.N. (2006): *Hieracium grofae* – a rediscovered diploid hybrid from the Ukrainian Carpathians. – Biologia, Bratislava, 61(4): 365–373.

Chrtek J., Mráz P., Zahradníček J., Mateo G. & Szeląg Z. (2007): Chromosome numbers and ploidy levels of selected species of *Hieracium* s. str. (Asteraceae). – Folia Geobotanica 42: 411–430.

Doebley J.F. (1989): Molecular evidence for a missing wild relative of maize and the introgression of its chloroplast into *Zea perennis*. – Evolution 43: 1555–1558.

Dorado O., Rieseberg L.H. & Arias D.M. (1992): Chloroplast DNA introgression in southern California sunflowers. – Evolution 46: 566–572.

Dover G.A. (1989): Linkage disequilibrium and molecular drive in the rDNA gene family. – Genetics 122: 249–252.

Doyle J.J. (1992): Gene trees and species trees: Molecular systematics as one-character taxonomy. – Systematic Botany 17: 144–163.

Drouin G., Daoud H. & Xia J. (2008): Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. – Molecular Phylogenetics and Evolution 49: 827–831.

Elder J.F. & Turner B.J. (1995): Concerted evolution of repetitive DNA sequences in eukaryotes. – Quarterly Review of Biology 70: 297– 320.

Elfstrand M. (1893): *Hieracia alpina* aus den Hochgebirgsgegenden des Mittleren Skandinaviens. – Uppsala.

Ellis J.R., Bentley K.E. & McCauley D.E. (2008): Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower *Helianthus verticillatus*. – Heredity 100: 574–580.

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). – Molecular Phylogenetics and Evolution 42: 347–361.

Feliner G.N. & Rosselló J.A. (2007): Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. – Molecular Phylogenetics and Evolution 44: 911–919.

Gaskin J.F. & Wilson L.M. (2007): Phylogenetic relationships among native and naturalized *Hieracium* (Asteraceae) in Canada and United States based on plastid DNA sequences. – Systematic Botany 32: 478–485.

Grant V. (1981): Plant speciation. Columbia University Press, New York.

Gustafsson Å. (1947): Apomixis in higher plants II. The causal aspect of apomixis. Acta University Lund N. F. Adv. 43: 69–179.

Harlan J.R. & de Wet J.M.J. (1975): On Ö. Winge and a prayer: the origins of polyploidy. – Botanical Review 41: 361–390.

Hörandl E. (1998): Species concepts in agamic complexes: Applications in the *Ranunculus auricomus* complex and general perspectives. – Folia Geobotanica 33: 335–348.

Hughes C.E., Eastwood R.J. & Bailey C.D. (2004): From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. – Philosophical Transactions of the Royal Society B – Biological Sciences 361: 211–225.

Jiao Y.N., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P. E., Tomsho L.P., Hu Y., Liang H.Y., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H., Leebens-Mack J. & de Pamphilis C.W. (2011): Ancestral polyploidy in seed plants and angiosperms. – Nature 473: 97–110.

Juxip A.J. (1960): Jastrebinka – *Hieracium* L. – In: Šiškin B.K. & Bobrov E.G. (eds), Flora SSSR 30: 1–698, Nauka, Moskva et Leningrad.

Koltunow A.M. (1993): Apomixis: Embryo sacs and embryos formed without meiosis or fertilization in ovules. – Plant Cell 5: 1425–1437.

Kovařík A., Dadejová M., Lim Y.K., Chase M.W., Clarkson J.J., Knapp S. & Leitch A.R. (2008): Evolution of rDNA in *Nicotiana* allopolyploids: A potential link between rDNA homogenization and epigenetics. – Annals of Botany 101: 815–823.

Levin D.A. (1975): Minority cytotype exclusion in local plant populations. – Taxon 24: 35–43.

Linder C.R. & Rieseberg L.H. (2004): Reconstructing patterns of reticulate evolution in plants. – American Journal of Botany 91: 1700–1708.

Mallet J. (2007): Hybrid speciation. – Nature 446: 279–283.

McCauley D.E., Sundby A.K., Bailey M.F. & Welch M.J. (2007): Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): Evidence from experimental crosses and natural populations. – American Journal of Botany 94: 1333–1337.

McDade L.A. (1992): Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. – Evolution 46: 1329–1346.

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

Mogie M. (1992): The evolution of asexual reproduction in plants. – Chapman & Hall, London.

Mogie M. & Ford H. (1988): Sexual and asexual *Taraxacum* species. – Biological Journal of the Linnean Society 35: 155–168.

Mráz P. (2003): Mentor effects in the genus *Hieracium* s. str. (Compositae, Lactuceae). – Folia Geobotanica 38: 345–350.

Mráz P. & Tomčíková D. (2004): Experimental hybridization in the genus *Hieracium* s. str. – crosses between diploid *H. umbellatum* and triploid *H. sabaudum*. – Thaiszia – Journal of Botany 14, Suppl. 1: 15–16.

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 Systematics and Evolution 255: 177–192.

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

Mráz P., Chrtek J. & Fehrer J. (2011): Interspecific hybridization in the genus *Hieracium* s. str.: evidence for bidirectional gene flow and spontaneous allopolyploidization. – Plant Systematics and Evolution 293: 237–245.

Nägeli C. & Peter A. (1885): Die Hieracien Mitteleuropas. – Monographische Bearbeitung der Piloselloiden mit besonderer Berücksichtigung der mitteleuropäischen Sippen, München.

Nealy D.B., Marshall K.A. & Sederoff R.D. (1989): Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don Endl. – Proceedings of the National Academy of Sciences of the USA 86: 9347–9349.

Norrlin J.P. (1912): Nya nordiska Hieracia, II. – Acta Societas Pro Fauna Et Flora Fennica 36(4): 1–127.

Norrlin J.P. (1888): Bidrag till Hieracium-floran i Skandinavska halföns mellersta delar. – Acta Societas Pro Fauna Et Flora Fennica 3(4): 1–117.

Noyes R.D. (2007): Apomixis in the Asteraceae: diamonds in the rough. – Functional Plant Science and Biotechnology 1: 207–222.

Olmstead R.G. & Palmer J.D. (1994): Chloroplast DNA systematics: a review of methods and data analysis. – American Journal of Botany 81: 1205–1224.

Peckert T. & Chrtek J. jun. (2006): Mating interactions between coexisting diploid, triploid and tetraploid cytotypes of *Hieracium echioides* (Asteraceae). – Folia Geobotanica 41: 323–334.

Pugsley H.W. (1948): A prodromus of the British Hieracia. – Journal of Linnaean Society (Botany) 54: 1–356.

Ramsey J. & Schemske D.W. (1998): Pathways, mechanisms and rates of polyploid formation in flowering plants. – Annual Reviews in Ecology and Systematics 29: 467–501.

Richards A.J. (2003): Apomixis in flowering plants: an overview. – Philosophical Transactions of the Royal Society of London Series B – Biological Sciences 358: 1085–1093.

Rieseberg L.H. (1997): Hybrid origin of plant species. – Annual Reviews in Ecology and Systematics 28: 359–389.

Rieseberg L.H. & Soltis D.E. (1991): Phylogenetic consequences of cytoplasmatic gene flow in plants. – Evolutionary Trends in Plants 5: 65–81.

Rieseberg L.H. & Willis J.H. (2007): Plant speciation. – Science 317: 910–914.

Rosenberg O. (1927): Die semiheterotypische Teilung und ihre Bedeutung fur die Entstehung verdoppelter Chromosomenzahlen. – Hereditas 8: 305–338.

Sang T. (2002): Utility of low-copy nuclear gene sequences in plant phylogenetics. – Critical Reviews in Biochemistry and Molecular Biology 37: 121–147.

Savidan Y. (2007): Apomixis in higher plants. – In: Hörandl et al. (eds.), Apomixis: Evolution, Mechanisms and Perspectives, pp. 15–25, A.R.G. Panter Verlag, Rubeol, Lichtenstein

Schuhwerk F. & Lippert W. (1998): Chromosomenzahlen von *Hieracium* (Compositae, Lactuceae) Teil 2. – Sendtnera 5: 269–286.

Schuhwerk F. (1996): Published chromosome counts in *Hieracium*. http://www.botanischestaatssammlung.de/index.html?/staff/schuhwerk.html.

Seehausen O. (2004): Hybridization and adaptive radiation. – Trends in Ecology and Evolution 19: 198–207.

Sell P.D. (1987): An introduction to the study of the British *Hieracia*, 1. History and classification. – Watsonia 16: 365–371.

Sell P.D. & Murrell G. (2006): Flora of Great Britain and Ireland: Campanulaceae-Asteraceae. – Cambridge University Press, Cambridge.

Sell P.D. & West C. (1976): *Hieracium* L. – In: Tutin T.G. et al. (eds), Flora Europaea 4: 358–410, Cambridge University Press, Cambridge.

Senchina D.S., Álvarez I., Cronn R.C., Liu B., Rong J.K., Noyes R.D., Paterson A.H., Wing R.A., Wilkins T.A. & Wendel J.F. (2003): Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. – Molecular Biology and Evolution 20: 633–643.

Sennikov A.N. & Illarionova I.D. (2002): Carpological studies in Asteraceae – Cichorieae, 1. subtribe Hieraciinae. – Komarovia 2: 97–125.

Shaw J., Lickey E.B., Beck J.T., Farmer S.B., Liu W.S., Miller J., Siripun K.C., Winder C.T., Schilling E. & Small R.L. (2005): The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. – American Journal of Botany 92: 142–166.

Shaw J., Lickey E.B., Schilling E.E. & Small R.L. (2007): Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. – American Journal of Botany 94: 275–288.

Shlyakov R.N. (1989): Jastrebinka – *Hieracium* L. – In: Tzvelev N.N. (ed), Flora evropejskoj časti SSSR 8: 140–300, Nauka, Leningrad.

Small R.L., Ryburn J.A., Cronn R.C., Seelanan T. & Wendel J.F. (1998): The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. – American Journal of Botany 85: 1301–1315.

Small R.L., Cronn R.C. & Wendel J.F. (2004): Use of nuclear genes for phylogenetic reconstruction. – Australian Systematic Botany 17: 145–170.

Soltis D.E. & Rieseberg L.H. (1986): Genetic consequences of autopolyploidy in *Tolmiea* (Saxifragaceae). – Evolution 43: 586–594.

Soltis D.E., Soltis P.S., Schemske D.W., Hancock J.F., Thompson J.N., Husband B.C. & Judd W.S. (2007): Autopolyploidy in angiosperms: have we grossly underestimated the number of species? – Taxon 56: 13–30.

Soltis P.S., Plunkett G.M., Novak S.J. & Soltis D.E. (1995): Genetic variation in *Tragopogon* species: Additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae). – American Journal of Botany 82: 1329–1341.

Soltis P.S. & Soltis D.E. (1991): Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. – Systematic Botany 16: 407–413.

Soltis P.S. & Soltis D.E. (2009): The role of hybridisation in plant speciation. – Annual Reviews of Plant Biology 60: 561–588.

Stace C.A. (1998): Species recognition in agamosperms – the need for pragmatic approach. – Folia Geobotanica 33: 319–326.

Suda J., Krahulcová A., Trávníček P., Rosenbaumová R., Peckert T. & Krahulec F. (2007): Genome size variation and species relationships in *Hieracium* subgen. *Pilosella* (Asteraceae) as inferred by flow cytometry. – Annals of Botany 100: 1323–1335.

Suh Y.B., Thien L.B., Reeve H.E. & Zimmer E.A. (1993): Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of the ribosomal DNA in Winteraceae. – American Journal of Botany 80: 1042–1055.

Taberlet P., Gielly L., Pautou G. & Bouvet J. (1991): Universal primers for amplification of three non–coding regions of chloroplast DNA. – Plant Molecular Biology 17: 1105–1109.

Tennant D. & Rich T. (2008): British alpine hawkweeds. – Botanical Society of the British Isles, London.

Thompson J.D. & Lumaret R. (1992): The evolutionary dynamics of polyploid plants: origins, establishment and persistence. – Trends in Ecology and Evolution 7: 302–307.

Trávníček P., Dočkalová Z., Rosenbaumová R., Kubátová B., Szeląg Z. & Chrtek J. (2011): Bridging global and microregional scales: ploidy distribution in *Pilosella echioides* (Asteraceae) in central Europe. – Annals of Botany 107: 443–454.

Tyler T. (2011): *Hieracium* sect. Oreadea (Asteraceae) in Sweden – from a complete mess to a preliminary taxonomic classification. – Nordic Journal of Botany 29(5): 538–589.

Wendel J.F., Schnabel A. & Seelanan T. (1995): Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). – Proceedings of the National Academy of Sciences of the USA 92: 280–284.

Willis J.C. (1973): A dictionary of flowering plants and ferns, ed. 8. – Cambridge University Press, Cambridge.

White T.J., Bruns T.D., Lee S. & Taylor J. (1990): Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M. A. et al. (eds.), PCR Protocols, A Guide to Methods and Applications, Academic Press, California.

Whittemore A.T. & Schaal B.A. (1991): Interspecific gene flow in oaks. – Proceedings of the National Academy of Sciences of the USA 88: 2540–2544.

Zahn K. H. (1921–1923): *Hieracium* L. – In: Engler H.G.A. (ed.) Das Pflanzenreich: Regni vegetabilis conspectus. IV, 280, Compositae – *Hieracium*. Band 76: 1–32, Wilhelm Engelmann, Leipzig.

Zahn K.H. (1922–1930): *Hieracium* L. – In: Ascherson P. & Graebner. (eds.), Synopsis der Mitteleuropäischen Flora 12(1): 1–492, Bornträger, Leipzig.

Zahradníček J. (2008): Velikost genomu a evoluční trendy rodu *Hieracium* L. – Ms. Thesis (Dipl. práce, depon. in: Knihovna katedry botaniky PřF UK, Praha).

BMC Evolutionary Biology

Research article

Open Access

Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise

Judith Fehrer*[1], Karol Krak[1] and Jindřich Chrtek Jr[1,2]

Address: [1]Institute of Botany, Academy of Sciences of the Czech Republic, Zámek 1, 25243 Průhonice, Czech Republic and [2]Department of Botany, Faculty of Science, Charles University Prague, Benátská 2, 12801 Prague, Czech Republic

Email: Judith Fehrer* - fehrer@ibot.cas.cz; Karol Krak - krak@ibot.cas.cz; Jindřich Chrtek - chrtek@ibot.cas.cz

* Corresponding author

Published: 22 September 2009

BMC Evolutionary Biology 2009, **9**:239 doi:10.1186/1471-2148-9-239

Received: 15 April 2009
Accepted: 22 September 2009

This article is available from: http://www.biomedcentral.com/1471-2148/9/239

Abstract

Background: *Hieracium* s.str. is a complex species-rich group of perennial herbs composed of few sexual diploids and numerous apomictic polyploids. The existence of reticulation and the near-continuity of morphological characters across taxa seriously affect species determination, making *Hieracium* one of the best examples of a 'botanist's nightmare'. Consequently, its species relationships have not previously been addressed by molecular methods. Concentrating on the supposed major evolutionary units, we used nuclear ribosomal (*ETS*) and chloroplast (*trnT-trnL*) sequences in order to disentangle the phylogenetic relationships and to infer the origins of the polyploids.

Results: Despite relatively low interspecific variation, the nuclear data revealed the existence of two major groups roughly corresponding to species with a Western or Eastern European origin. Extensive reticulation was mainly inferred from the character additivity of parental *ETS* variants. Surprisingly, many diploid species were of hybrid origin whilst several polyploid taxa showed no evidence of reticulation. Intra-individual *ETS* sequence polymorphism generally exceeded interspecific variation and was either independent of, or additional to, additive patterns accounted for by hybrid origin. Several *ETS* ribotypes occurred in different hybrid taxa, but never as the only variant in any species analyzed.

Conclusion: The high level of intra-individual *ETS* polymorphism prevented straightforward phylogenetic analysis. Characterization of this variation as additive, shared informative, homoplasious, or unique made it possible to uncover the phylogenetic signal and to reveal the hybrid origin of 29 out of 60 accessions. Contrary to expectation, diploid sexuals and polyploid apomicts did not differ in their molecular patterns. The basic division of the genus into two major clades had not previously been intimated on morphological grounds. Both major groups are thought to have survived in different glacial refugia and to have hybridized as a result of secondary contact. Several lines of evidence suggest the data is best explained by the presence of an extinct range of variation and a larger diversity of ancestral diploids in former times rather than by unsampled variation. Extinct diversity and extensive reticulation are thought to have largely obscured the species relationships. Our study illustrates how multigene sequences can be used to disentangle the evolutionary history of agamic complexes or similarly difficult datasets.

Background

Agamic complexes, usually consisting of few diploid and many polyploid taxa - the latter reproducing by apomixis to various degrees - are characterized by large numbers of species often exhibiting more or less continuous morphological variation. Consequently species delimitation is a challenging task, and even in the presence of expert taxonomic knowledge, does not always lead to satisfactory results. Therefore, the molecular study of species relationships in such groups is seriously hampered, and few attempts have been made to reconstruct the evolutionary history of such complexes [1].

Three or four subgenera of hawkweeds (*Hieracium* L. s.l.) have traditionally been recognized. In this study, we will focus on *Hieracium* subgen. *Hieracium* (*Hieracium* s.str.). For a review of taxonomic treatments, see [2]. *Hieracium* s.str. is a highly diverse group (500-8000 species depending on taxonomic concept) known for its notorious taxonomic complexity, which is associated with a variation in ploidy level, breeding system and supposed history of extensive hybridization [3,4]. The group consists of perennial herbs distributed mainly in temperate areas of Europe, Asia, and North America. Main centers of *Hieracium* diversity are located in the European mountains (e.g., Alps, Pyrenees, Carpathians, Balkan Peninsula) and in westernmost Asia. The species occupy forests, forest margins, various grasslands, and rocks. Polyploid (triploid and tetraploid, very rarely pentaploid) taxa [5,6] with asexual reproduction via parthenogenetic development of the unreduced egg cell (*Antennaria*-type diplospory) prevail. Polyploids are near-obligate apomicts that produce seed asexually (corresponding to the maternal genotype), but can produce pollen via meiosis to different degrees. Sexual reproduction is rare and restricted to diploid species ($2n = 18$). Diploids are mainly confined to unglaciated refugia; polyploids are also widespread in areas that had been covered by ice sheets [7]. Many taxa comprise populations of different ploidy level.

The most complete taxonomic treatment was published by Zahn [3]. He used a broad species concept and distinguished two different kinds of species: (i) 'basic' species (*species principales collectivae*, 'Hauptarten') having a unique set of morphological characters; and (ii) 'intermediate' species (*species intermediae collectivae*, 'Zwischenarten'), which combine the morphological characters of two or more 'basic' species and are generally thought to be of hybrid origin. Following this concept, we expected the diploid 'basic' species to represent major evolutionary units from which the multitude of polyploids arose. In addition, we suspected polyploid 'basic' species (i.e., taxa presumed to be of non-hybrid origin) to comprise further basal lineages, especially as the number of recent diploids is rather low (about 20 out of several hundred macrospecies). Thus, in order to cover as much as possible of the overall genetic variation present in the subgenus, we focused on all 'basic' species irrespective of their ploidy.

The 45S rDNA cistron continues to be the most popular nuclear region for species-level phylogenetic studies of plants [8] despite its acknowledged flaws [9]. The 5'-*ETS* region situated upstream of the 18S rRNA gene [10] was chosen as a nuclear marker because of its higher variability compared to *ITS* in the Asteraceae [11,12]. *ITS* showed hardly any resolution in *Hieracium* s.str. [13]. While the majority of publications based on *ETS* are focused on phylogeny reconstruction, several studies have also used this region to infer hybrid or allopolyploid origin, often in combination with *ITS* [14-21].

In sexually reproducing diploids not introgressed by other species, one would expect sequences of this multicopy region to be more or less homogeneous at the intra-individual and intraspecific level [22]. Since concerted evolution [23] tends to be slowed down in asexually reproducing organisms [24-26], apomictic allopolyploids should have retained the *ETS* variants of their respective diploid progenitors. Additivity patterns revealed by direct sequencing of nrDNA have been used to identify hybrid or allopolyploid origin in a wide range of plant groups [27-35], in one case even in a triple hybrid [36]. We therefore adopted a direct sequencing approach complemented by cloning of selected accessions.

Analysis of the *trnT-trnL* intergenic spacer of chloroplast DNA was also performed in order to establish a phylogeny of maternal lineages and to identify the maternal parent of hybrid accessions. We have recently confirmed maternal transmission of cpDNA for *Hieracium* s.str. [37].

This paper reports the first molecular phylogeny of *Hieracium* subgen. *Hieracium* based on a sampling of most of the assumed major evolutionary units. It demonstrates the unexpected hybrid origin of many of these including diploid species, provides evidence of extinct ancestral diversity and discusses the occurrence of extensive intra-individual polymorphism found in most diploid and polyploid accessions. In order to decipher the phylogenetic signal in spite of these abundant polymorphisms, we distinguished between character additivity, shared informative variation, and noise by detailed character state analysis.

Methods

Sampling

Zahn's [3] basic framework was used for a meaningful taxon sampling for phylogenetic analysis, as complete sampling was impractical due to the high species numbers and the unclear delimitation of 'intermediate' species. Up

to three samples of each accessible 'basic', i.e., supposedly non-hybridogenous species of *Hieracium* s.str. were collected. If ploidy level varied within a species, diploid populations were included whenever available. The taxonomic concept generally follows Zahn [3] with a few exceptions: The species concept of section *Cerinthoidea* (Iberian and Pyrenean taxa) was adopted from Mateo [38], and *H. plumulosum* (*H. waldsteinii* s.l.) was treated as a separate species. Two newly described diploid species from the Balkans (*H. kittanae* and *H. petrovae*; [39,40]), diploid *H. pojoritense* - considered as an endemic species of the Eastern Carpathians by Nyárády [41], and *H. mixtum* - which was first treated as a 'basic' species by De Retz [42], followed by other authors - were also included. Altogether, we analyzed 60 accessions of 46 species. For the remaining 'basic' species *sensu* Zahn (*H. fuscocinereum* Norrlin, *H. laniferum* Cav., and *H. schmalhausenianum* Litv. & Zahn), only herbarium material more than 30 years old was available, in which the DNA was too degraded for amplification. Details of all accessions are included in the Additional file 1: Origins of individual accessions.

As outgroups for the *ETS* analyses, the most closely related genera *Pilosella*, *Hispidella*, *Andryala*, and '*Hieracium*' *intybaceum* were used [13]. For cpDNA analysis, only *Hispidella* and a *Pilosella* species with 'original' chloroplast haplotype were used as outgroups. This was because some of the *Pilosella* species, *Andryala*, and '*Hieracium*' *intybaceum* have captured chloroplasts derived from *Hieracium* due to ancient intergeneric hybridizations [13]. Vouchers of all specimens are deposited in the herbarium of the Institute of Botany in Průhonice (PRA).

Molecular methods

DNA was isolated from fresh or CTAB-conserved leaves as described in [43]. The *ETS* region of the nuclear ribosomal DNA was PCR-amplified using the primers Ast-8 and 18 S [44]. PCR products were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and sequenced (GATC Biotech, Konstanz, Germany). For each sample, both strands were directly sequenced using the PCR primers. Sequences containing more than one indel or otherwise difficult to read were cloned. Prior to cloning, the PCR products were excised from 1% agarose gels and purified with the Zymoclean Gel DNA Recovery kit (Zymoresearch, Orange, CA). The gel-purified fragments were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) following the manufacturer's instructions, but downscaled to half reactions. Approximately 24 colonies per sample were transferred into 20 μl ddH_20 and denaturated at 95°C for 10 min. They served as templates for subsequent PCR amplifications for sequencing. Using the Ast-8 primer, 3-15 clones per individual were sequenced (eight on average). For the *trnT-trnL* intergenic spacer of chloroplast DNA, PCR amplification, sequencing and indel coding were done as described previously [13]. Sequence alignments were done by eye in BioEdit [45]; they were unambiguous for both markers due to low overall variation.

Treatment of ETS *sequences prior to phylogenetic analysis*

All *ETS* sequences contained a certain proportion of polymorphic sites. Polymorphisms were distinguished from sequencing artifacts when they occurred in both reading directions. As overall interspecific sequence divergence was low, if only single short (1-2 bp) indels were present, polymorphic sites were determined from direct sequences by reading both strands on to the indel position, and after it by confirming them in both directions via peak subtraction. Where multiple or longer indels obscured sequence reads, the samples were cloned, and the clones were used to aid peak subtraction in the electropherograms obtained by direct sequencing in order to infer polymorphic sites in regions whose readability was strongly affected by indels. Polymorphisms were represented by the IUPAC ambiguity codes.

Preliminary phylogenetic analyses of *ETS* revealed little structure apart from a monophyly of *Hieracium* [46]; all species emerged from a single basal polytomy, and the few subclades found were poorly supported. However, visual inspection of the alignment suggested the existence of two major species groups. Many accessions showed additive characters (superimposed peaks and a diagnostic indel) corresponding to positions that differed between these main groups, suggesting hybrid origin. These accessions were removed from the phylogenetic analyses. The resulting dataset was inspected for the occurrence of further additive characters that might be indicative of hybridization within each of the two species groups. Several such accessions were detected and also excluded prior to final phylogenetic analyses.

Cloned sequences - mostly from hybrids - revealed in several cases two major *ETS* variants corresponding to sequences of both major species groups. However, some were recombinant according to visual inspection, one showed a novel hybrid sequence, some showed single nucleotides occurring on the 'wrong' ribotype, and in case of strongly biased ratios of different ribotypes (or at single polymorphic positions), not all underrepresented sequences (or character states) were retrieved (Additional file 2: Patterns of *ETS* recombination). Also, the number of polymerase errors approximated the low interspecific variation in some cases. Therefore, cloned sequences were not included in the phylogenetic analyses (but see below).

In the preliminary parsimony and Bayesian analyses of the dataset from which all suspected hybrid sequences had been deleted, the effect of the remaining polymorphic sites on tree topology and branch support were assessed by including each accession twice, namely one sequence with all polymorphisms included, and a second one reflecting the major sequence type in which ambiguities were 'resolved' towards the overrepresented character state in case the latter represented at least about 70% of the total signal. In the resulting trees, both sequences of the same accession always appeared together except for *H. villosum* 1029 and *H. lucidum* where they differed slightly in their placement (not shown). As these were minor effects and because of the higher information content of the dominant sequence types in both cases, for the final phylogenetic analyses, only major *ETS* sequences were used for all accessions. This also reduced computing time as well as the number of unique indels in the alignment as they were typically present in lower amounts within an individual.

In outgroup species, *Pilosella* samples produced several equally strong bands in PCR amplification that could not be eliminated by optimizations. The shortest fragment was homologous to the 5'-*ETS* of other species while the longer fragments contained duplications (variable number of large subrepeats, not shown) which had also been found in other Asteraceae [12,47]. Therefore, *Pilosella* was represented in the phylogenetic analyses by three cloned sequences of the short fragment. In *Hispidella*, intra-individual polymorphisms included one indel adjacent to two substitutions. One allele was dominating so that the second one could be inferred from direct sequencing. Both variants were included in the phylogenetic analyses.

In order to represent the complex features of the *ETS* dataset as comprehensively as possible, direct sequences of all accessions were submitted to GenBank in duplicates: (i) with all polymorphic sites included, and (ii) corrected for overrepresented character states (e.g., the major sequence types used for phylogenetic analyses). In addition, all cloned sequences were submitted; some were corrected for polymerase errors (substitutions found in only one clone and neither accounted for by direct sequencing nor present in any other taxon). Corrected clones are indicated, e.g., as 'clone 1c' in the sequence description line of the submitted sequences; recombinant sequences are indicated as such in the notes. [GenBank:EU821363-EU821419, EU867566-EU867709, and FJ858089-FJ858133 (*ETS*), and GenBank:EU867710-EU867763 (*trnT-trnL*)]. Eight additional *trnT-trnL* sequences (for two outgroup and six *Hieracium* accessions) were adopted from [13].

Phylogenetic analyses

Maximum parsimony (MP), maximum likelihood (ML) (PAUP* V4.0b10, [48]), and Bayesian inference (MrBayes V3.1.2, [49,50]) were applied for phylogenetic analyses of the *ETS* and *trnT-trnL* datasets.

For MP analysis of the *ETS* region, single-base gaps were treated as a 5th character state, longer indels as a single character, and remaining ambiguous bases as polymorphisms. Heuristic searches were performed with 1,000 random sequence addition replicates, saving no more than 100 trees of length greater than or equal to 1 per replicate and TBR branch swapping. Bootstrapping was done with the same settings, but without branch swapping. Prior to ML analysis, the model of molecular evolution best fitting to the data was determined with Modeltest version 3.5 [51]. A HKY+G model was found in hierarchical Likelihood Ratio Tests (hLRTs) which was used for ML analyses with the estimated parameter settings. Heuristic searches were done with one random addition sequence replicate and TBR branch swapping; 1,000 bootstrap replicates were performed without branch swapping. For Bayesian analyses, the same basic model parameters determined by Modeltest (two substitution rates and gamma distribution) were used. Two replicate analyses with four chains each were performed with the default parameters and computed for 3 million generations, sampling every 1,000th tree. All statistical parameters indicated that convergence was reached. The first 1,000 trees per run were discarded as burn-in, and the remaining 4,002 trees were summarized.

For analyses of cpDNA, MP analysis was performed as described above, treating indels as single events. As sequence divergence was low, character state changes were mapped onto the branches by hand, and homoplasies were identified based on the alignment. For ML analysis, the F81 model determined in hLRTs was applied. Accordingly, Bayesian analyses were run using one substitution rate and equal rates as priors, all else as indicated above.

Evaluation of intra-individual polymorphism

Intra-individual polymorphism in multicopy sequences results from the presence of more than one ribotype within a particular genome accompanied by incomplete homogenization of the different variants by concerted evolution. If meaningful patterns can be distinguished from stochastic ones, intra-individual polymorphisms can add considerably to the understanding of evolutionary processes. As potentially meaningful polymorphisms we considered (i) sites that consisted of different character states (nucleotides or indels) that were monomorphic in other sequences of the dataset (character additivity), and (ii) sites showing the same alternative character states in more than one accession, but one character state was

missing in the rest of the dataset (shared polymorphisms). The latter can either be meaningful (in the case of shared ancestral polymorphism) or not (if they are homoplasious). We distinguished putative homoplasious polymorphisms according to (i) geography - accessions for which recent or past overlap of distribution areas is highly unlikely; (ii) phylogeny - polymorphism-sharing accessions belonging to divergent clades, or hybrids of certainly different origin; and (iii) singularity - no further evidence for grouping these particular accessions (or accession groups) could be found. In most cases, more than one of these criteria applied. Unique (accession-specific) polymorphic sites represented a further kind of uninformative variation. A comprehensive list of sites along with the accessions in which the polymorphisms occurred and their interpretation according to the criteria described above is given in Additional file 3: Summary of intra-individual polymorphisms. To facilitate reproducibility of these inferences, an alignment with the corresponding positions is also provided (Additional file 4: Alignment of *Hieracium ETS* sequences).

Identification of further lineages

Detailed analysis of shared polymorphisms revealed several groups of accessions that were characterized by different sets of consistently shared polymorphisms. In order to analyze the origin of this variation in more detail, all samples for which meaningful patterns emerged after analysis of intra-individual polymorphism were also cloned and sequenced as above. Clones were inspected visually, and all non-recombinant clones containing further signal (i.e., exceeding accession-specific variation, see Additional file 2: Patterns of *ETS* recombination) were subjected to additional phylogenetic analyses under the same conditions as described above.

Results

Intra-individual polymorphism in the ETS *region*

A surprisingly high level of intra-individual polymorphism in *ETS* sequences was found in many accessions irrespective of ploidy, ranging from one to 37 per sample (12 on average). In the whole ingroup (Additional file 4: Alignment of *Hieracium ETS* sequences), intra-individual polymorphisms concerned 196 aligned positions, exceeding the number of variable sites based on substitutions (120). At 30 of these 196 positions, two different kinds of polymorphism were found (e.g., Y and K); at five further positions, even three different ones were observed (e.g., Y, K and S). Thus, the total number of different intra-individual polymorphisms was 236 out of 573 aligned characters. A complete list of polymorphic positions, their occurrence in particular accessions, and their classification as unique, shared (homoplasious or informative), or additive is given in Additional file 3: Summary of intra-individual polymorphisms.

At most of the polymorphic sites (unique and shared), one of the alternative character states was lacking across the whole dataset while true character additivity was comparatively rare (Figure 1a). A large proportion of the polymorphisms was accession- or species-specific. Most of this variation was uninformative, but a few unique polymorphisms showed character additivity (Figure 1b) and indicated hybrids of unique origin (see below). Among the polymorphisms shared between accessions of different species, some were shared among apparently unrelated species (according to the phylogeny), or between reasonable groups of accessions and single or multiple outliers. These polymorphisms were considered as homoplasious according to the criteria specified in the Methods. 'Reasonable groups of accessions' refers to assemblages that were supported by more than one shared polymorphism and usually also by other evidence (e.g., synapomorphic substitutions, phylogenetic position, geography). In several of these cases, character additivity applied to one of the 'reasonable' subgroups (Figure 1b) which often comprised many accessions. Thus, meaningful information could in some cases also be retrieved from homoplasious positions by a detailed evaluation of the patterns. Shared polymorphisms for which no obvious evidence for homoplasy was found were considered as potentially informative. A comparably large part of them was additive (Figure 1b). Typically, the same or similar sets of accessions shared further polymorphisms and were usually also supported by other evidence (e.g., phylogenetic position, geography). Thus, informative intra-individual polymorphism could be retrieved from an excess of uninformative or even misleading variation by detailed character state analyses. Additive and shared informative variation was crucial for inferring reticulation patterns or added to the phylogenetic information content of the data (see below).

ETS *phylogeny*

Interspecific variation of the *ETS* region in *Hieracium* was rather low; maximal sequence divergence based on substitutions was 4.6% *p*-distance (compared to 1.9% in *ITS*, not shown). Seven short indels (1-10 bp) occurred, six of them represented intra-individual polymorphisms, only one differed between accessions.

Phylogenetic analyses of the dataset from which all inferred hybrid accessions had been excluded based on character additivity resulted in basically the same tree under different optimality criteria (Figure 2). A basal split into two well-supported major clades was found, one composed mostly of species with Western, the other one of taxa with mainly Eastern European distribution. Widespread or Central European taxa fell into one or other cluster. Both major clades were characterized by basal polytomies from which a few well-supported subclades emerged.

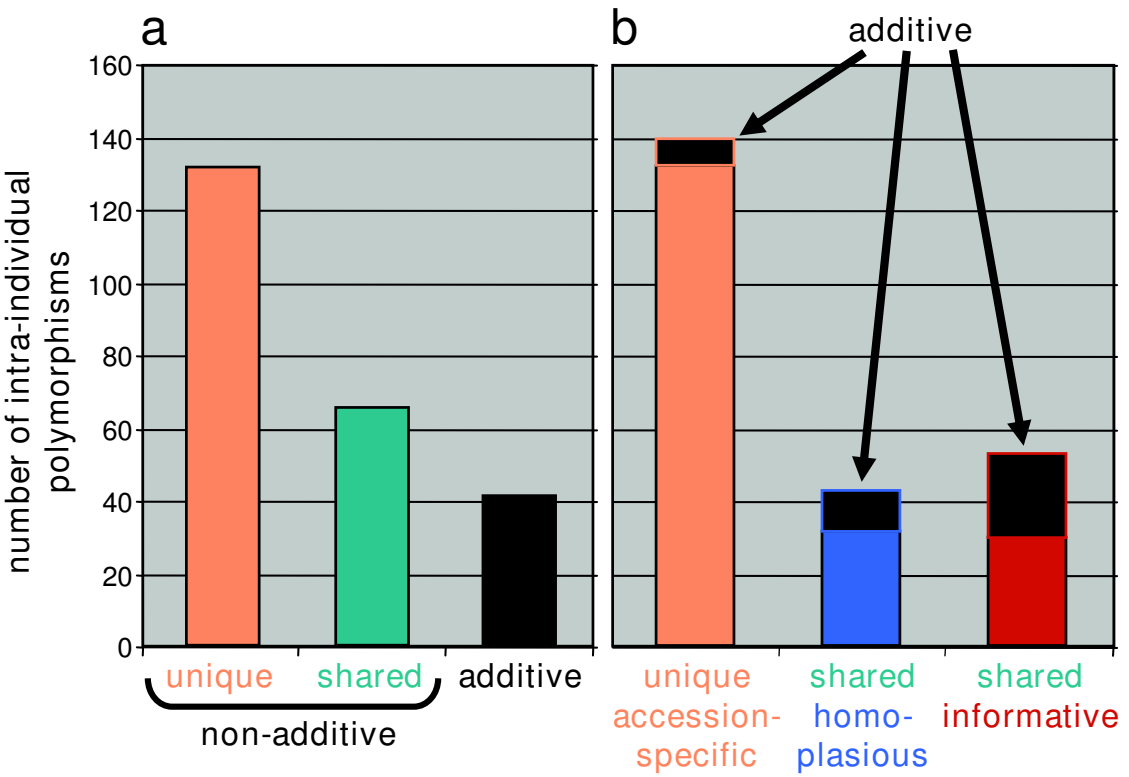


Figure 1
Categories of intra-individual polymorphism. The large amount of uninformative (unique) and potentially misleading (homoplasious) versus meaningful (additive, shared informative) variation is illustrated. a) Discrimination of non-additive (one of the character states is missing in the whole dataset, e.g., T and Y are present, but no C) from additive polymorphisms (e.g., T, Y as well as C are present). b) Further discrimination of polymorphisms shared by two or more accessions into homoplasious or informative variation, and distribution of polymorphisms that show character additivity for all categories.

All species of the 'Eastern' group that occurred at unresolved positions are from Southeastern Europe, mostly from the Balkans (*H. naegelianum*, *H. pannosum*, *H. petrovae*, *H. sparsum*, *H. kittanae*) (Figure 2). One lineage emerging from the basal polytomy of the 'Eastern' clade consisted of two diploid accessions of *H. alpinum*. While *H. alpinum* s.l. is a widespread species, diploid populations occur only in the Eastern and Southern Carpathians [52] consistent with an eastern origin of that species. The '*H. porrifolium*' subclade consisted of the alpine species *H. pilosum*, *H. villosum*, *H. bupleuroides*, and *H. porrifolium*. The first three are Central European polyploid taxa; *H. porrifolium* is the only diploid species in this subclade. This taxon is restricted to the Southeastern Alps - a known glacial refuge area. Species of the '*H. porrifolium*' group occur only in rock crevices on limestone and have a very similar ecology. The best supported group ('*H. umbellatum*' subclade) consisted of the tall-growing perennials *H. virosum*, *H. umbellatum*, *H. eriophorum*, and *H. canadense*. Their distribution extends from Siberia to North America. The kind of involucrum in these species, an important character complex for *Hieracium* taxonomy, resembles that of other species with exclusively Eastern European distribution. *Hieracium umbellatum* is the most widely distributed diploid.

In the 'Western' cluster, subclades emerging from the basal polytomy were a branch composed of three Pyrenean taxa ('Pyrenean' subclade) and a lineage consisting of two accessions of *H. transylvanicum*. All species with exclusively or mainly Pyrenean or Western Alpine distribution belonged to the 'Western' clade, some of them may be remnants of previously much larger populations. In addition, diploid populations of *H. prenanthoides* are also

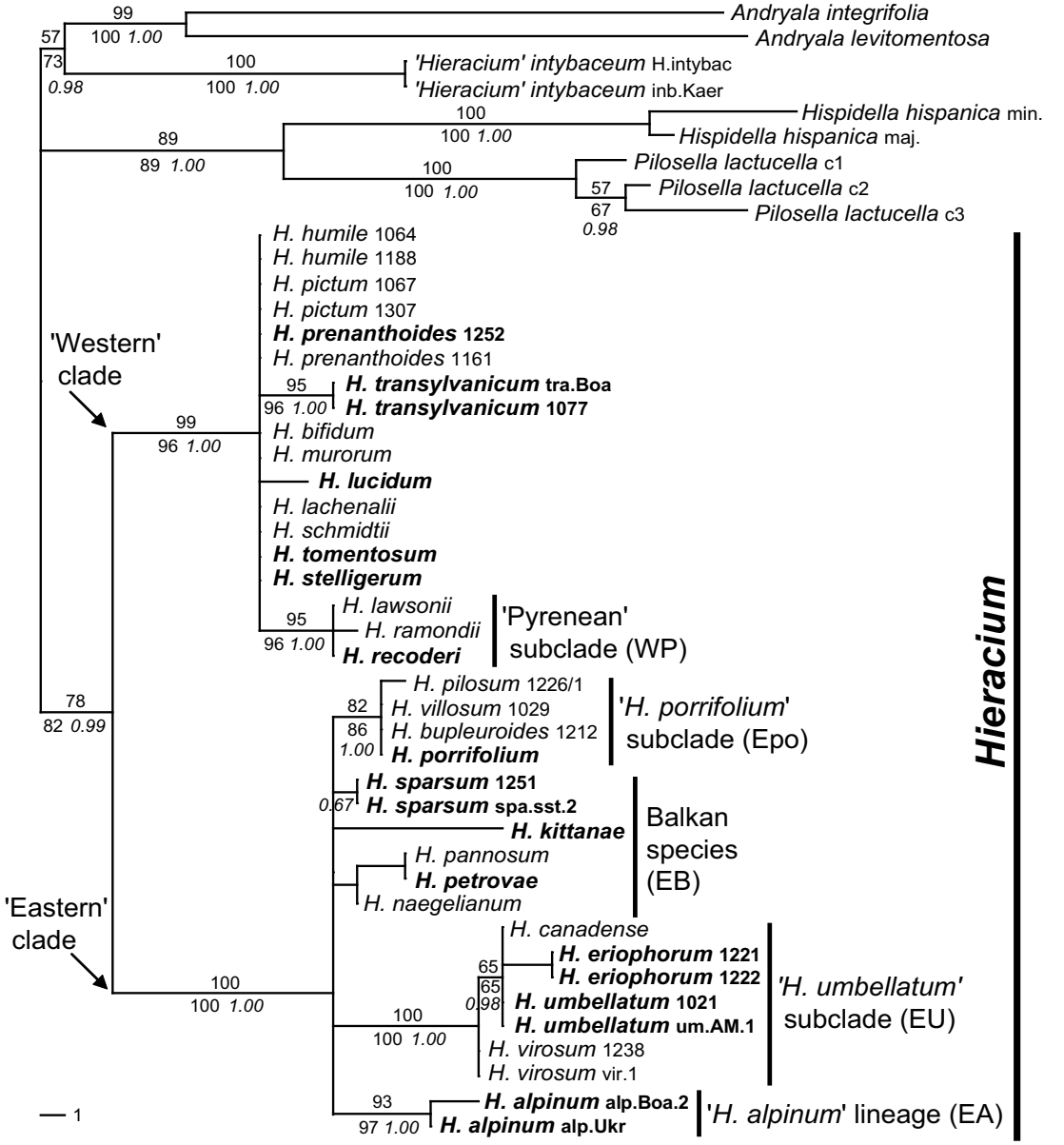


Figure 2

Molecular phylogeny of major evolutionary units of *Hieracium* based on *ETS* sequences. One out of 256 equally parsimonious trees is shown (417 steps, ri = 0.947, ci = 0.940; 153 variable characters of which 89 were parsimony informative) with bootstrap support indicated above the branches. The strict consensus tree topology corresponds to branches with support values in the MP analysis. Bootstrap values for ML and posterior probabilities for Bayesian analyses are given below the branches. Diploid *Hieracium* species are indicated in boldface. Four accessions (*H. prenanthoides* 1252 and 1161, *H. lachenalii*, *H. lucidum*) are included in these analyses in spite of their inferred hybrid origin, because *ETS* sequences of the first three were 'Western' apart from 2-3 polymorphic sites, and the dominant sequence of *H. lucidum* was also 'Western'.

restricted to the Southwestern Alps, indicating western origin. One species, *H. lucidum*, is endemic to Northwestern Sicily and is only known from a single relict population. Widespread or Central European species whose accessions fell into the 'Western' clade were *H. schmidtii*, *H. murorum*, and *H. bifidum*. The accession of the widespread species *H. lachenalii* used, was apparently introgressed by a species from the ('Eastern') '*H. umbellatum*' subclade as indicated by its chloroplast DNA (see below). The only species not fitting into the 'Western' clade according to its distribution was *H. transylvanicum* which occurs in the Carpathians, the Eastern Alps and the Northern Balkans. Its large genome size (about 10% higher than that of 'Western' species) may also suggest an eastern origin (for details, see [46]). We therefore cannot exclude that this is a similar case to *H. lachenalii*, however, the chloroplast haplotype of *H. transylvanicum* was unique and did not allow an assignment to any ('Eastern'?) taxon.

Diploid accessions - all were sexual and self-incompatible - were spread all over the tree and occupied basal as well as derived positions within the major clades (Figure 2). The same was true for polyploid (triploid or tetraploid) accessions, which were experimentally confirmed as apomictic. Accessions included in the 'Western' cluster had significantly lower DNA content (1Cx values corrected for ploidy) than those of the 'Eastern' clade (3.66 ± 0.21 pg versus 4.03 ± 0.19 pg, see Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features). Details of reproduction, ploidy, and genome size are given in a parallel paper [46]. Thus, the major clades revealed by phylogenetic analysis of the *ETS* region coincide with geographic patterns and with significant genome size differences (~10% on average).

Chloroplast haplotypes

Variation of the *trnT-trnL* intergenic spacer of chloroplast DNA was expectedly lower (max. 2.12% *p*-distance) than that of the nuclear *ETS*. Figure 3 summarizes the results of the phylogenetic analyses for this marker. The tree also allows to determine the number of substitutions and indels that distinguish between particular sequences. Three homoplasious mutations occurred. Two of these were substitutions in rather basal positions (G → T, G → A) that accounted for major rearrangements in equally parsimonious trees. These branches were not supported in any of the analyses. In contrast, a 7 bp-deletion was always derived from haplotypes of either *H. lucidum* or *H. umbellatum* (or sequences identical to these), with a minimum of three substitutional steps (as in the tree shown), or with up to five steps separating *H. lachenalii*/*H. laevigatum* from *H. tomentosum*/*H. prenanthoides* 1161. Both lineages found significant support in the Bayesian analysis. As this mutation involved a tandem repeat, we assume parallel deletion events, especially as part of the same motif was independently lost in *H. mixtum*. For the assignment of haplotype groups, homoplasious substitutions were ignored.

Bootstrap support was generally low - if haplotypes differ only by a single mutation, it cannot be higher on principle [53]. We therefore concentrated only on particular haplotypes (often identical sequences) and on conservative inferences we considered as unequivocal. No species relationships were inferred from this marker. Despite the low bootstrap support, all but the three above-mentioned mutations were free of homoplasy (which is reflected by very high retention and consistency indices in the parsimony analysis, see Figure 3) and can therefore be considered as diagnostic. Posterior probabilities in Bayesian analysis were usually significant (≥ 0.95).

The inferred haplotype groups (Figure 3) generally matched the particular clades or species groups found in the *ETS* analyses (Figure 2). Taxa of the 'Pyrenean' lineage were subdivided into two haplotype groups separated by a single substitution; *H. naegelianum* had a unique haplotype and was excluded from the 'Balkan' group. Shared chloroplast haplotypes (usually identical sequences) allowed inference of the maternal origin of most hybrid accessions and always matched the subclades of one parent previously inferred from the *ETS* dataset (see also below). In addition, a few rather derived unique chloroplast haplotypes occurred, especially in hybrid accessions whose exact parentages could not be inferred from the *ETS* data. Only in two cases did cpDNA clearly contradict the phylogenetic position of the species revealed by *ETS*. One was *H. lachenalii*, which showed a 'Western' *ETS*, but its cpDNA corresponded to species of the '*H. umbellatum*' clade, a derived 'Eastern' lineage. The taxon also shows some morphological features of species belonging to the '*H. umbellatum*' group. However, no trace of such an introgression was found in the *ETS*. The genome size of *H. lachenalii* was about 5% higher than the usual values of other 'Western' clade species (Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features). Therefore, chloroplast capture indicative of cryptic hybridization between a diploid maternal parent of the '*H. umbellatum*' clade and a 'Western' pollen donor with almost complete homogenization of *ETS* towards the 'Western' parent (see also Table 1) seems to be the most likely explanation for the discrepancy. The other case was *H. sparsum*. The cpDNA haplotypes of both analyzed accessions were almost identical and unique, they differed by many mutations from all other Balkan species. Furthermore, they were apparently derived from an '*H. alpinum*' haplotype. This result was less easy to interpret. It is possible that either a chloroplast capture event occurred very early in the history of the species or that *H. sparsum* is derived from an unknown maternal species with '*H. alpi-*

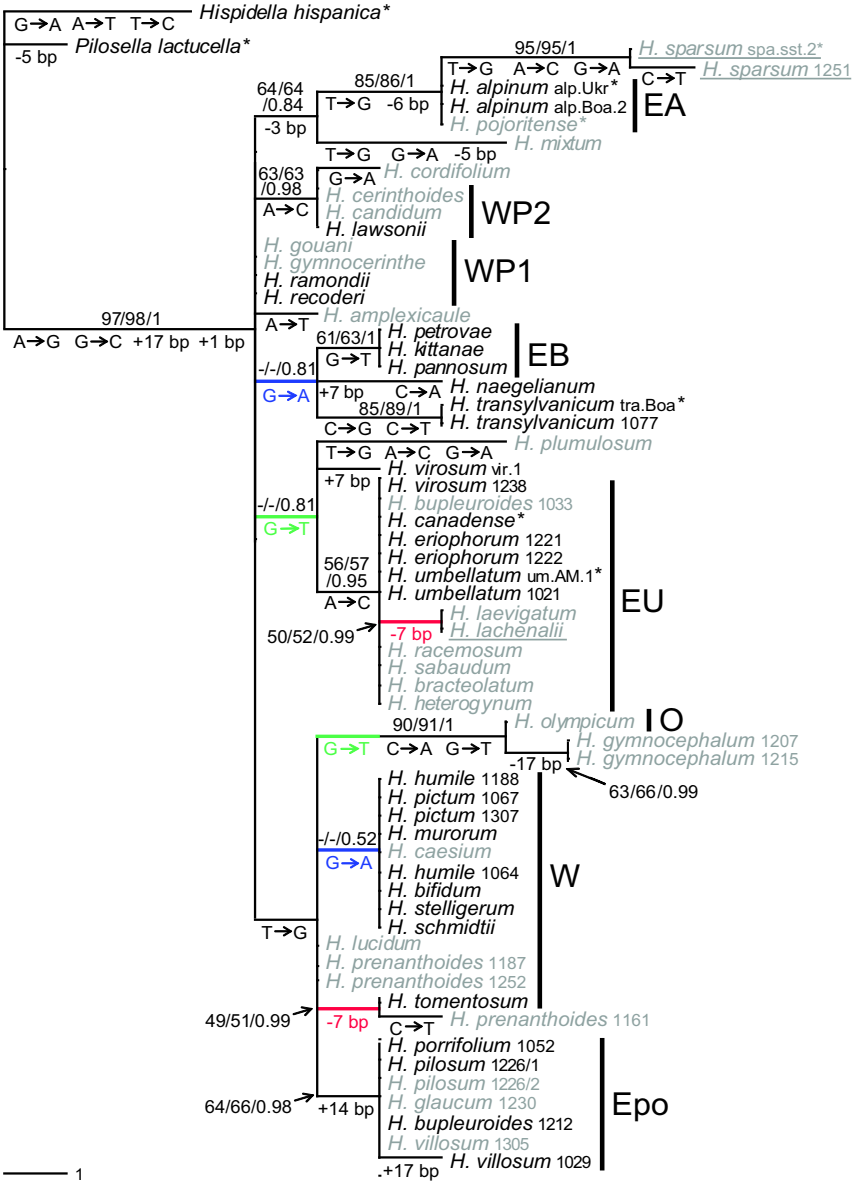


Figure 3

Phylogenetic analysis of the chloroplast *trnT-trnL* intergenic spacer. One out of 60 equally parsimonious trees is shown (44 steps, ri = 0.972, ci = 0.988; 37 variable characters of which 21 were parsimony informative) whose topology exactly matches that of ML and Bayesian analyses including unsupported groups. Character state changes are shown below the respective branches, identical colors indicate homoplasious mutations. Bootstrap values for MP/ML and Bayesian posterior probabilities are above branches. Hybrid accessions are in grey, those for which hybrid origin was only implied by cpDNA are underlined. Asterisks after species names: sequences adopted from [13]. Acronyms of particular haplotype groups refer to Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features; a subset corresponds to the subclades and species groups in Figure 2.

Table 1: Positions distinguishing 'Eastern' and 'Western' clades and signatures of selected hybrid accessions

Clades and particular accessions	Diagnostic sites differing between major clades (position in alignment)														
	187	**194**	**197**	**233**	**244**	**252**	**262**	**281**	**307**	**368**	**446**	**461**	**478**	**486**	**488**
'Western' clade	**G**	**C**	**T**	**C**	**G**	**C**	**T**	**T**	**G**	**C**	**G**	**T**	**T**	**A**	**T**
'Eastern' clade	*T*	*A*	*C*	*T*	*T*	*T*	-	*C*	*A*	*T*	*A*	*A*	*C*	*T*	*A*
Interclade hybrids[1]	K	M	Y	Y	K	Y	-/T	Y	R	Y	R	W	Y	W	W
H. prenanthoides 1252	**G**	**C**	**T**	**C**	**G**	**C**	**T**	**T**	**G**	**C**	**G**	**T**	Y	W	W
H. prenanthoides 1161	**G**	**C**	**T**	**C**	**G**	y	**T**	**T**	**G**	**C**	**G**	**T**	Y	W	W
H. prenanthoides 1187	k	m	y	y	k	y	-/t	y	r	y	r	w	Y	W	W
H. prenanthoides 1187 dominant peaks	**G**	**C**	**T**	**C**	**G**	**C**	**T**	**T**	**G**	**C**	**G**	**T**	Y	W	W
H. lachenalii	**G**	**C**	**T**	**C**	**G**	y	**T**	**T**	r	**C**	**G**	**T**	**T**	**A**	**T**
H. mixtum dominant peaks	**G**	**C**	*C*	**C**	**G**	**C**	-	**T**	*A*	*T*	**G**	**T**	*C*	*T*	*A*
H. plumulosum dominant peaks	*T*	*A*	*C*	*T*	*T*	*T*	-	*C*	*A*	*T*	*A*	*A*	*C*	*T*	*A*

[1] Interclade hybrids that show character additivity at all 15 sites comprise *H. amplexicaule*, *H. bracteolatum*, *H. caesium*, *H. glaucum*, *H. gouani*, *H. gymnocephalum* (both accessions), *H. heterogynum*, *H. laevigatum*, *H. mixtum*, *H. olympicum*, *H. pilosum* 1226/2, *H. plumulosum*, *H. prenanthoides* 1187, *H. racemosum*, *H. sabaudum*, and *H. villosum* 1305.
Lower case letters indicate unequal representation of superimposed peaks.
'Western' character states are in boldface; 'Eastern' in italics.

num' cpDNA (for more details, see also Additional file 1: Origins of individual accessions).

A compilation of accessions, their geographic origin, ploidy/DNA content, *ETS* features, and cpDNA variant (as depicted in Figures 2 and 3) is given in Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features.

Reticulation within 'Eastern' and 'Western' clades

Two accessions showing within-'Eastern' clade hybrid origin were identified by additivity patterns. One was the diploid *H. pojoritense*, which was comprised of an '*H. umbellatum*' group sequence (resulting in seven additive characters at positions 232, 326, 349, 425, 426, 459, and 479) and the '*H. alpinum*' ribotype (responsible for another four additive sites at positions 120, 123, 180, and 340; Additional file 4: Alignment of *Hieracium ETS* sequences); the latter *ETS* variant was strongly predominating in the hybrid. Its chloroplast DNA also corresponded to *H. alpinum* (Figure 3). The second was an accession of triploid *H. bupleuroides* (1033) that turned out to be 'pure' *H. bupleuroides* introgressed by the '*H. umbellatum*' group: Character additivity for the seven diagnostic sites of the latter plus two sites reflecting the synapomorphic substitutions of the '*H. porrifolium*' clade (positions 254 and 367) indicated members of these two subclades as parents. In addition, at two of the '*H. umbellatum*' clade diagnostic sites, a 'pure' accession of *H. bupleuroides* (1212) showed unique polymorphisms whose alternative character states differed from those of the '*H. umbellatum*' group. The particular positions (425, 426) in *H. bupleuroides* 1033 showed the corresponding triple peaks. This plant resembled *H. umbellatum* in some morphological characters, and its chloroplast DNA also corresponded to the '*H. umbellatum*' haplotype. Another taxon with putative hybrid origin within the 'Eastern' group was *H. sparsum* (both analyzed accessions). In this case, the only evidence for potential reticulation is based on the discrepancy between nuclear and chloroplast DNA (see above).

Four within-'Western' clade accessions of putative hybrid origin were identified; all had similar origins. Species of the 'Pyrenean' subclade differed by 2-3 synapomorphic substitutions (at positions 171, 259, and 485) from the majority of 'Western' clade species that occurred at unresolved positions. Four Pyrenean species showed the corresponding character additivity while lacking patterns that would link them to any other taxa: *H. cerinthoides*, *H. cordifolium*, *H. candidum*, and *H. gymnocerinthe*. Their cpDNA haplotypes were either identical to those of taxa of the 'Pyrenean' subclade or were slightly derived from these (Figure 3).

Hybrids between major clades: a multitude of origins

Most hybrid accessions proved to have originated from different combinations of 'Eastern' *and* 'Western' clade species. The major clades differed by 14 substitutions and one diagnostic 1 bp-indel. These sites were additive in the interclade hybrids (Table 1). In most cases, either equal or biased amounts of the respective character states of one or the other major group were found throughout the entire sequence. Thus, relative peak height and/or relative

number of clones usually corresponded to parental *ETS* variants present in either equal or different amount in the hybrid accessions.

Apart from the general contribution of 'Eastern' and 'Western' lineages, one or both of the respective parental taxa of the interclade hybrid accessions could often be narrowed down to particular subclades, sometimes even to the genotype of a particular parental accession: In addition to the 15 polymorphisms representing the differences between the major clades, contribution from the '*H. umbellatum*' subclade to interclade hybrids could be inferred from seven additional diagnostic characters (see above). This was the case for *H. racemosum*, *H. sabaudum*, *H. laevigatum*, *H. bracteolatum*, and *H. caesium*. All of these except *H. caesium* ('Western' cpDNA haplotype) also had a chloroplast haplotype corresponding to that of the '*H. umbellatum*' clade. In *H. heterogynum*, only two of the seven additive characters were present in *ETS*. The cpDNA haplotype of that accession also corresponded to the '*H. umbellatum*' haplotype. Almost complete 'Western' *ETS*, but '*H. umbellatum*' cpDNA was found in *H. lachenalii* (see also above). Another hybrid taxon, *H. glaucum*, was composed of an unidentified 'Western' clade parent and a member of the ('Eastern') '*H. porrifolium*' group which also donated the cpDNA. The same pattern was found in one accession of each *H. pilosum* and *H. villosum* (1226/2 and 1305), whose 'pure' accessions (1226/1 and 1029) belonged to the '*H. porrifolium*' clade. An additive polymorphism (position 21) occurring in the hybrid accession *H. pilosum* 1226/2 reflected a unique substitution of *H. pilosum* 1226/1 which was collected from the same site. The genome constitution of the analyzed accessions of *H. amplexicaule* and *H. gouani* was a combination of an unidentified 'Eastern' clade taxon with a member of the 'Pyrenean' subclade. The cpDNA of *H. gouani* corresponded to one of the Pyrenean haplotypes, the *H. amplexicaule* sequence was probably derived from that variant (Figure 3). For four accessions of three hybridogenous taxa with contribution of both clades (*H. gymnocephalum*, *H. olympicum*, and *H. plumulosum*), no particular species subgroup could be identified for either parent based on character additivity. They also had unique chloroplast haplotypes.

Two particularly interesting cases with respect to *ETS* patterns were the interclade hybrids *H. prenanthoides* and *H. mixtum*. For *H. prenanthoides*, three accessions were analyzed. One of the triploids (1187) was a hybrid between the major clades with the 'Western' ribotype strongly predominating (Table 1) and four out of seven additional polymorphisms reflecting the '*H. umbellatum*' clade (see also cloned sequences for this accession, Additional file 2: Patterns of *ETS* recombination). The other two accessions seemed at first to have an ordinary 'Western' sequence. However, while the diploid 1252 and the triploid 1161 did not show additivity at 12 out of 15 positions distinguishing between the clades, the last three positions at the 3'-end of the *ETS* were strongly additive (equal representation of both character states). In the triploid 'obvious' hybrid 1187, these same three sites showed equal amounts of superimposed peaks at these positions compared to the strongly biased proportion of 'Eastern' and 'Western' ribotypes in the rest of the sequence (Table 1). Thus, it seems that the diploid represents an interclade hybrid that had lost most of the 'Eastern' ribotype, probably by intragenomic recombination. The same signature of strong character additivity at the 3'-end plus several *H. prenanthoides*-specific polymorphic sites occurred in both triploid accessions indicating that the diploid *H. prenanthoides* gave rise to both triploids. Accession 1187 appears to have originated by subsequent hybridization with the '*H. umbellatum*' lineage. Another interclade hybrid, *H. mixtum*, was unique in showing a mixture of strongly predominating 'Eastern' *or* 'Western' character states (see also Additional file 2: Patterns of *ETS* recombination). Strongly overrepresented character states alternated seven times between the diagnostic 'Eastern' and 'Western' character states (Table 1) which indicates a novel hybrid sequence most probably generated by gene conversion.

Inferences from shared informative polymorphisms

Apart from intra-individual polymorphism reflecting *ETS* character state additivity and thus shared by several accessions because of similar hybrid origin, some shared informative variation (Figure 1b) also existed at positions where one of the alternative character states was missing in the dataset. In this section, we will focus on examples where several such polymorphisms (classified as shared informative) were shared by different taxa in order to retrieve meaningful information exceeding the results inferred from character additivity and phylogenetic analyses.

Species of the '*H. porrifolium*' clade were, in addition to synapomorphic substitutions, also linked by shared polymorphisms. Some of the latter indicate a closer relationship between *H. porrifolium* and *H. bupleuroides* (positions 194 and 375) as well as between *H. villosum* and *H. pilosum* (positions 351 and 369) which corresponds to the morphology and sectional classification of these species.

Shared polymorphic sites between a particular accession of *H. umbellatum* (um.AM.1) and *H. eriophorum* (positions 393 and 412) could indicate that the latter species (endemic to Southwestern France) is a young derivative from a similar genotype or from a common ancestral one with *H. umbellatum*. This would also explain the eastern origin inferred from phylogenetic analysis despite the Western European distribution of *H. eriophorum*. The

same two polymorphisms were seen in *H. pojoritense* whose *ETS* sequence was additive for the '*H. umbellatum*' clade and the *H. alpinum* lineage. They narrowed down the paternal parent of *H. pojoritense* to either *H. eriophorum* (which can be excluded on its geographic distribution and endemic status) or to a genotype similar to *H. umbellatum* um.AM.1.

Apart from synapomorphic substitutions, the 'Pyrenean' subclade also differed from most 'Western' species by 2-3 shared polymorphisms (positions 171, 317, and 376). The same also occurred in four Pyrenean taxa that occupied an intermediate position between basal 'Western' species and the 'Pyrenean' subclade and could represent ancestral variation. According to character additivity, they may have hybrid origin (see above). However, an alternative explanation could be that character states identifying the 'Pyrenean' lineage originated as polymorphisms retained in the four putative hybrids. The taxa with more homogeneous *ETS* could then be younger derivatives in which the alternative nucleotides became fixed by concerted evolution. In any case, shared polymorphisms along with geographic vicinity, rather restricted distribution ranges and 'Pyrenean' chloroplast haplotypes indicate a particularly close relationship of these seven taxa.

Species from the Balkans did not cluster together in phylogenetic analyses, because they mostly shared intra-individual polymorphisms rather than substitutions. These patterns were scattered across the species (Figure 4) and produced nearly all possible combinations of Balkan taxa in equally parsimonious trees (not shown). Those with only one particular nucleotide at a given polymorphic position usually showed the consensus (plesiomorphic) character state. Only four apomorphic substitutions occurred in these species; all were reflected by additive polymorphisms in other Balkan species. Most polymorphisms (21), however, were non-additive. Shared (probably ancestral) polymorphisms and geographic distribution indicate a close relationship among these species, and for the interclade hybrid *H. olympicum*, they imply a Balkan species as the most likely 'Eastern' parent, the *H. sparsum* pattern being the best match (Figure 4, Additional file 2: Patterns of *ETS* recombination).

Non-additive, hybrid-specific variation

The *ETS* of the relict diploid *H. lucidum* showed a major ribotype corresponding to the majority of 'Western' species, but also contained a pattern of three small additional peaks and a frameshift caused by a 2 bp-indel (at positions 169, 180, 196-197, 358). The same set of intra-individual polymorphisms occurred in *H. prenanthoides* 1161

Species	Position in alignment																							
	46	49	123	154	162	189	254	256	260	292	311	332	344	357	385	390	434	445	458	468	475	482	501	503	514
H. kittanae	G	Y	T	R	C	Y	Y	Y	T	T	K	A	Y	G	Y	R	G	Y	C	R	G	A	Y	T	S
H. pannosum	G	Y	T	A	C	Y	T	Y	T	Y	T	A	Y	G	T	A	R	C	C	A	R	R	Y	Y	G
H. petrovae	G	C	T	A	C	C	T	T	T	C	T	A	Y	G	T	A	A	C	C	A	R	R	C	C	G
H. naegelianum	K	C	T	A	C	C	T	Y	T	T	T	A	C	G	T	A	R	C	C	A	G	A	Y	T	G
*H. sparsum**	G	C	T	G	C	C	Y	T	T	T	K	A	C	G	Y	R	R	C	C	R	G	A	C	Y	S
H. olympicum[1]	G	C	T	G	C	C	Y	T	T	T	T	A	C	G	Y	R	R	C	C	R	G	A	C	T	S
H. plumulosum[1]	K	C	Y	R	Y	C	Y	T	W	T	Y	M	C	K	T	A	R	Y	Y	A	G	A	C	T	G
*H. gymnocephalum**[1]	G	C	Y	R	Y	C	Y	T	W	T	Y	M	Y	K	T	A	A	C	Y	A	G	A	C	T	G
H. heterogynum[1,2]	G	C	T	R	Y	C	T	T	W	T	Y	A	C	G	T	A	A	C	Y	A	G	A	C	T	G
others = consensus	G	C	T	G	C	C	T	T	T	T	T	A	C	G	T	A	A	C	C	A	G	A	C	T	G

Figure 4

Shared polymorphisms in species from the Balkans. * Two accessions of this species are identical for these characters (except for *H. gymnocephalum* 1215 with C instead of Y at position 344); [1] Interclade hybrids; [2] The 'unknown Western 2' polymorphisms (light violet) were not found in direct sequencing, but are represented by a single cloned sequence and are probably present in very low amounts in this genome. Dark orange: apomorphic substitutions of Balkan species; light orange: shared substitutions of Balkan species and hybrids; green: polymorphisms shared only by Balkan interclade hybrids, but not reflected by 'pure' Balkan species ('unknown Eastern'); light violet: polymorphisms shared with the (non-Balkan) interclade hybrid accessions *H. villosum* 1305 and *H. pilosum* 1226/2, but not reflected by other Balkan species ('unknown Western 2'). At position 254, C occurs on the 'unknown Western 2' ribotype in *H. plumulosum* and *H. gymnocephalum*, but on the 'Eastern' ribotype in *H. olympicum* and must therefore be a parallelism. The scattered and inconclusive distribution of polymorphisms and substitutions in 'pure' Balkan species are reflected by the intra-individual patterns in cloned sequences of *H. kittanae*, their most polymorphic representative. For more details, see Additional file 2: Patterns of *ETS* recombination.

(a cryptic interclade hybrid, see above), the non-consensus character states also being present in low amounts. Four interclade hybrids (*H. racemosum*, *H. sabaudum*, *H. bracteolatum*, and *H. olympicum*) showed these alternative character states in higher amounts so that they had a higher probability of being recovered by cloning. The non-consensus character states turned out to be situated on a particular physical sequence (3 out of 6 clones in *H. racemosum*, 1 out of 3 in *H. sabaudum*, 2 out of 5 in *H. bracteolatum*, and 1 out of 8 clones in *H. olympicum* plus 2 recombinant ones). We refer to this pattern as the 'unknown Western 1' ribotype. Given the diversity of origins of these six taxa as inferred from molecular markers and the large geographic distances between the sampling sites, it is unlikely that shared ancestral variation was responsible for the observed pattern in this case. More likely, the 'unknown Western 1' ribotype they have in common reflects one of the parents of all these hybrids (unsampled or extinct) while their second parent can vary. In *H. prenanthoides* 1161 (see above), this involved a subsequent hybridization with a third parent (i.e. with the 'unknown Western 1').

The maternal parent of the interclade hybrids *H. pilosum* 1226/2 and *H. villosum* 1305 belonged to the ('Eastern') '*H. porrifolium*' clade like the 'pure' accessions of the same taxa. Their exact paternal ('Western') parent could not be identified. Both hybrid accessions shared four polymorphisms (at positions 123, 254, 332, and 357) whose non-consensus character states were situated on 'Western' *ETS* variants (designated as 'unknown Western 2', see also Additional file 2: Patterns of *ETS* recombination). These polymorphisms/ribotypes were also shared with *H. gymnocephalum* (both accessions) and *H. plumulosum* (Figure 4) which were interclade hybrids from the Balkans with unknown 'Eastern' *and* 'Western' parents. One cloned sequence of *H. heterogynum*, an interclade hybrid with '*H. umbellatum*' maternal origin, showed this pattern as well. Exceptionally, these polymorphisms were not apparent in direct sequencing suggesting that this ribotype must be very rare (< 5%) in the genome of this accession and that it was probably retrieved by chance.

The latter three taxa shared another four polymorphisms (positions 162, 260, 311, 458) that were situated on 'Eastern' strands; none of these were shared with any 'pure' Balkan species (Figure 4). We refer to this pattern as the 'unknown Eastern' ribotype (see Additional file 2: Patterns of *ETS* recombination). It was confined to interclade hybrids with Balkanian distribution. Each of the 'unknown Eastern' ribotypes also contained one or two 'Balkan' polymorphisms suggesting that these variants probably arose from a Balkan species.

All three 'unknown' ribotypes occurred only in hybrids, but never as the only *ETS* variant in any other species. This is consistent with previously unidentifiable 'Western' or 'Eastern' parental subclades in all these taxa. This kind of shared variation cannot easily be considered as homoplasious, because each concerned a unique set of four intra-individual polymorphisms that were shared by particular groups of accessions and did not contradict other patterns. We therefore included all non-recombinant sequences of these variants into further phylogenetic analyses (Additional file 6: *ETS* phylogeny with ribotypes present only in hybrids). These ribotypes formed three subclades emerging from the basal 'Western' or 'Eastern' polytomies and showed similar divergences to the previously identified subclades composed of 'pure' species. The combination of direct sequencing and cloned sequences further revealed that *H. plumulosum* (diploid) comprised a total of four identifyable ribotypes: Besides the 'unknown Western 2' and the 'unknown Eastern' variant, an almost pure 'Western' sequence was found, and an ordinary 'Eastern' ribotype could also be inferred (Additional file 2: Patterns of *ETS* recombination). Likewise, *H. heterogynum* (triploid) comprised three identifyable ribotypes, but in addition had a cpDNA haplotype corresponding to a fourth lineage, the '*H. umbellatum*' group (plus two of its diagnostic substitutions in the *ETS*). A compilation of all reticulation events inferred from *ETS* patterns (including the three 'unknown' lineages) and chloroplast haplotypes is included in Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features.

Reticulation in Hieracium *s.str*

To sum up, hybrid origin was unexpectedly inferred for a total of 29 out of 60 *Hieracium* accessions. Most of them (20) were interclade hybrids of which 17 showed complete character additivity for the two major clades; two others showed partial hybrid signatures in *ETS*, and one combined an almost exclusively 'Western' *ETS* with an 'Eastern' chloroplast haplotype. Four accessions showed evidence of multiple hybridization/introgression events. Within the 'Eastern' clade, two hybrid accessions were identified by *ETS* character additivity, and a further two were inferred from putative chloroplast capture. Within the 'Western' clade, four putative hybrids were identified by character additivity, and another species comprised an undifferentiated 'Western' and an 'unknown Western 1' *ETS* ribotype. Diploid hybrids occurred at all levels: four were interclade hybrids, two occurred within the 'Western' clade, and three within the 'Eastern' clade. One diploid comprised four different *ETS* ribotypes which suggests the presence of at least two nrDNA loci per haploid genome.

Altogether, at least 17 different combinations of parental lineages were found in the accessions with hybrid origin. They included three ribotypes that occurred only in

hybrids. A graphical summary of the inferred intra- and interclade reticulation patterns is provided in Figure 5. A more detailed assessment and discussion of the origin of each species/accession, including some ecogeographic, morphological, and floristic information are given in Additional file 1: Origins of individual accessions.

Discussion

NrDNA patterns revealed by direct sequencing and cloning

In our experience ([54], Krak et al., unpubl. data) and that of other labs [32,55], the relative ratios of parental sequence types in PCR amplicons are usually reproducible. Also, in the present study, relative proportions of cloned sequences - generated from separately amplified PCR products - roughly corresponded to relative amounts of superimposed peaks observed in direct sequencing. Furthermore, relative amounts of parental *ETS* alleles in interclade hybrids could be biased towards the 'Eastern' *or* the 'Western' variants (Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features) and corresponded to total genome size of the respective accessions [46]. This suggests that backcrossing or unequal genome composition (e.g., diploid and haploid gametes contributing to a triploid) generally had a stronger influence on different *ETS* ratios in hybrids than concerted evolution. Reproducibility of relative amounts of amplified parental variants, biased ratios occurring in either direction, and matching genome sizes argue against the confounding effects of PCR selection or drift [56] in our data.

The reliability of direct sequencing is also highlighted by a comparison of the strength of particular signals at polymorphic positions in direct sequencing with recombination patterns in cloned sequences. Direct sequencing was usually superior in cases of heavily biased representation of different ribotypes (for examples, see Additional file 2: Patterns of *ETS* recombination), because it can sample variation across all nrDNA loci present in the genome. Rare alleles - contributing less than about 5% of the total signal in electropherograms in our case - are equally difficult to identify by cloning [22,32], especially if there is no prior information that such variation exists. While we cannot exclude that cloning of additional accessions might have revealed still further ribotypes, the example of *H. kittanae* - a 'pure' diploid species with the highest number of polymorphisms (15) - which reflected the inconclusive patterns of Balkan species (Figure 4) by an equally inconclusive pattern at the intra-individual level (Additional file 2: Patterns of *ETS* recombination) raises doubts that further extensive cloning would have retrieved much new information. Besides, in cloned sequences, polymerase errors, intragenomic and/or PCR recombination, single substitutions on the 'wrong' strand or hybrid-specific sequences, the latter probably resulting from gene conversion, have to be disentangled from 'original' or major parental ribotypes (see Methods and Additional file 2: Patterns of *ETS* recombination). This becomes an impossible task if these variants do not differ by a sufficiently high number of diagnostic characters. On the other hand, cloning was necessary to resolve or confirm sequence reads affected by multiple indels and to attribute the derived character states of shared informative polymorphisms to particular physical sequences (e.g., for the 'unknown' ribotypes).

Intra-individual polymorphism in the ETS

The earlier view that nrDNA repeats are generally homogeneous within a species [22] has led to the widespread assumption that intra-individual polymorphism of nrDNA is the exception rather than the rule [57]. More recently however, researchers have become increasingly aware of the problems associated with the complex and often unpredictable behavior of this multicopy marker [9]. By 2003, Bailey et al. [58] can already cite 22 reports of intra-individual polymorphism. Meanwhile, this is thought to be a common phenomenon in plants [59], and a certain degree of intra-individual polymorphism can usually be found whenever it is looked for, though it is still widely ignored [60].

The most frequently observed reasons for *prominent* intra-individual polymorphism of nrDNA are hybrid/allopolyploid origin [27,28,30,32,33,61,62] or the occurrence of pseudogenes [63-67]. We consider the latter possibility as unlikely for *Hieracium* because overall sequence variation was very low, the GC content of all accessions fell within a very narrow range (50.4-51.9%), and long indels were missing. Other criteria for the recognition of pseudogenes are not easily applicable, because *ETS* - unlike the *ITS* region - does not comprise conserved genes, and comparative data on secondary structure are also not available. Reticulation as one source of the extensive intra-individual polymorphism in *Hieracium* is supported by the fact that 17 different combinations of ribotypes accounted for a considerable number of the sequence polymorphisms observed in individual accessions; most of them involving 'Eastern' *and* 'Western' variants. All accessions with more or less homogeneous *ETS* fell into one or the other clade (Figure 5). However, one could also argue that multiple paralogous ribotypes within individual genomes might reflect ancestral polymorphism and lineage sorting [68] rather than reticulation. This would require that the ancestral population of the whole subgenus already comprised all ribotypes as divergent *ETS* alleles or paralogs. Accessions with composite genomes could then be understood as heterozygous individuals derived from various combinations of original *ETS* ribotypes already present in the ancestral population. The 'pure' species would then be homozygous for one of the divergent *ETS* ribotypes, have suffered locus loss, or their ribotypes became homoge-

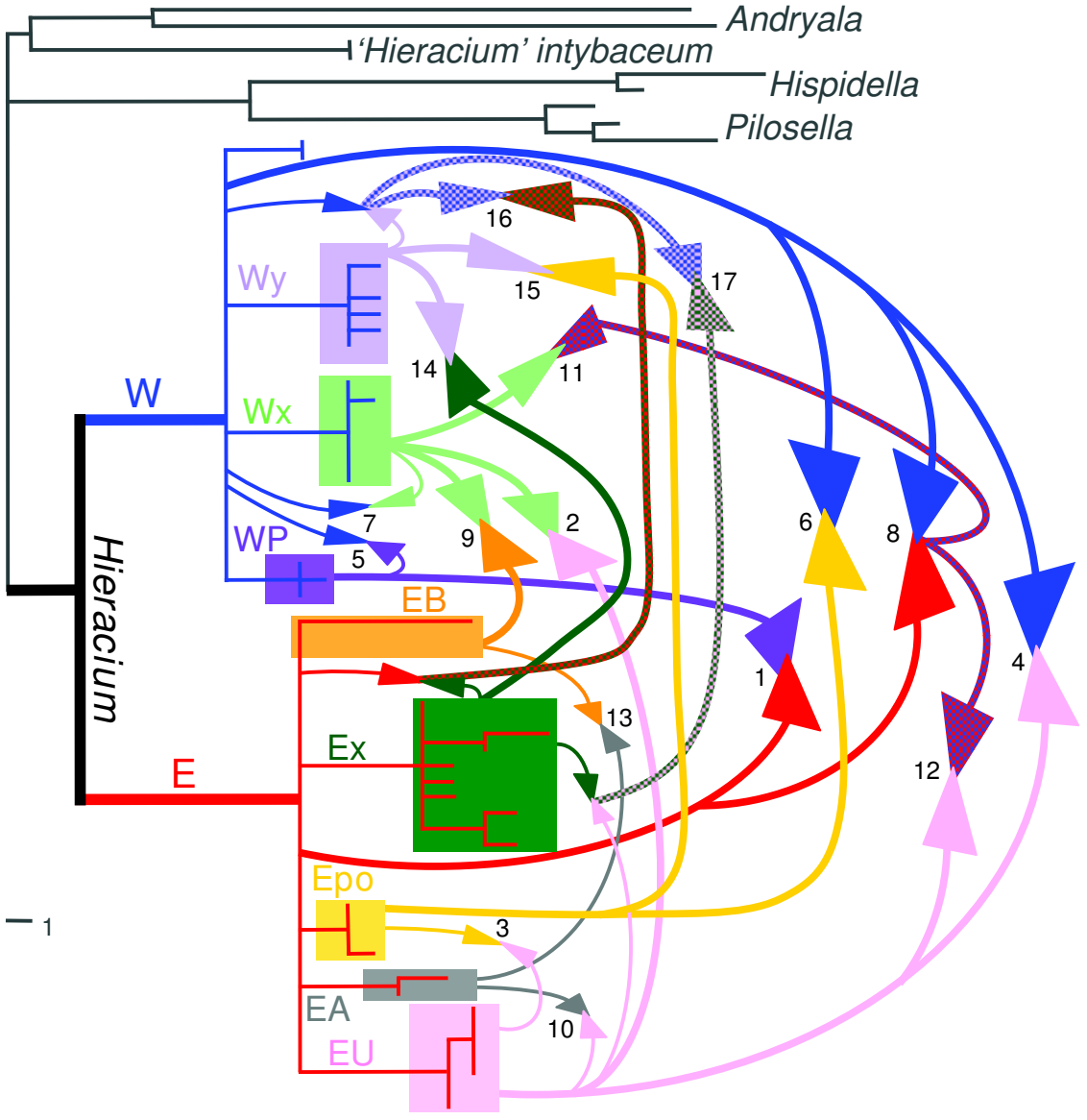


Figure 5 *(see legend on next page)*

Figure 5 *(see previous page)*
Reticulation in *Hieracium*. A graphical summary of reticulation based on *ETS* features and (in two cases) on chloroplast capture is shown on a simplified *ETS* tree (unsupported branches collapsed) that includes also the ribotype lineages occurring only in hybrids (Wx, Wy, Ex). W - 'Western', Wx - 'unknown Western 1', Wy - 'unknown Western 2', WP - 'Pyrenean', E - 'Eastern', EB - Balkan species, EU - '*H. umbellatum*' clade, Ex - 'unknown Eastern', EA - '*H. alpinum*' lineage, Epo - '*H. porrifolium*' clade. Detailed results of this analysis are given in Additional file 6: *ETS* phylogeny with ribotypes present only in hybrids. Small arrows/thin lines represent hybridizations within each clade, larger arrowheads/bold lines indicate involvement of both major clades. If no particular subclade could be inferred for a given combination, arrows start at the polytomies of the 'Western' or 'Eastern' clades. Multiple hybridizations are indicated by patterned arrows. They reflect particular ribotype combinations, not any order of subsequent hybridizations. Numbers near the arrowheads refer to particular hybrid accessions: 1 - *H. amplexicaule*, *H. gouani*; 2 - *H. bracteolatum*, *H. racemosum*, *H. sabaudum*; 3 - *H. bupleuroides* 1033, 4 - *H. caesium*, *H. lachenalii*, *H. laevigatum*; 5 - *H. candidum*, *H. cerinthoides*, *H. cordifolium*, *H. gymnocerinthe*; 6 - *H. glaucum*; 7 - *H. lucidum*; 8 - *H. mixtum*, *H. prenanthoides* 1252, 9 - *H. olympicum*; 10 - *H. pojoritense*; 11 - *H. prenanthoides* 1161, 12 - *H. prenanthoides* 1187, 13 - *H. sparsum* (2); 14 - *H. gymnocephalum* (2), 15 - *H. pilosum* 1226/2, *H. villosum* 1305, 16 - *H. plumulosum*; 17 - *H. heterogynum*.

nized towards particular variants by concerted evolution. The phylogeny (Figure 2) would in that case represent a gene tree rather than a species tree. However, the geographic pattern along with the divergent genome sizes observed for the major clades, the general congruence of cpDNA haplotypes with *ETS* subclades/species groups (Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features), and the correspondence of ecological preferences or morphological similarities of species belonging to particular (e.g., '*H. porrifolium*', '*H. umbellatum*') subclades strongly argue in favor of true species relationships. In contrast, lineage sorting is expected to result in randomly distributed ribotypes following divergence from a polymorphic ancestor. Only the incongruent patterns observed in Balkan species (Figure 4) show features that are best explained by ancestral polymorphism and lineage sorting (see Results). For most of our data, however, this is not the most parsimonious explanation. Instead, we consider hybridization among members of divergent lineages to be a more likely interpretation, at least for the additive patterns.

In many cases, a considerable number of intra-individual polymorphisms were superimposed on the variation due to additivity. While shared informative polymorphisms contained additional phylogenetic signal (see Results), most of the non-additive variation was unique/accession-specific (Figure 1), i.e., phylogenetically uninformative. These polymorphisms were either randomly scattered on different physical strands or tended to have accumulated in a particular ribotype (e.g., 'Eastern' or 'Western', see Additional file 2: Patterns of *ETS* recombination). This is consistent with the assumption that they resulted from mutations occurring on different nrDNA repeats and had spread in the genome to some degree - in order to be detected at all in direct sequencing, they have to reach a threshold of at least 5% of all copies - but failed to be completely homogenized by concerted evolution (similar to retained parental variants in most hybrids). The large amount of unique or apparently randomly distributed (homoplasious) intra-individual variation (Figure 1) may indicate partial saturation with intra-individual polymorphisms suggesting slow concerted evolution relative to mutation rates. The occurrence of two or even three different kinds of polymorphism at 35 positions points in the same direction. At one of these positions (254), C as a derived character state evidently arose twice independently; once as part of the 'unknown Western 2' ribotype, and once on the 'Eastern' alleles of Balkan species (Figure 4, Additional file 2: Patterns of *ETS* recombination).

Due to the complex patterns and abundance of intra-individual variation, prior to phylogenetic analysis of the *ETS* dataset, noise had to be distinguished from meaningful variation, and reticulation had to be identified as such, because hybrids are known to confound phylogenetic analysis when they remain undetected [69,70]. Almost half of all accessions analyzed showed evidence of reticulation. Even restricting the analyses to diploids was not an option in this case, because diploids were equally polymorphic on average and also included interclade hybrid accessions. In fact, the highest number of polymorphisms occurring in any accession (37) was found in a diploid (a hybrid containing four divergent ribotypes plus many unique polymorphisms) while the lowest numbers (1 or 2 polymorphisms) occurred in some triploids and tetraploids of presumed autopolyploid origin. Although a number of programs are currently available that can deal with intra-individual variation and reticulation [59,71,72], no algorithm can possibly distinguish between meaningful and homoplasious patterns if they occur at the same aligned positions, or identify hybridization from partial additivity or from novel hybrid sequences (e.g., *H. prenanthoides*, *H. mixtum*). Similarly, in a recent study on *Fagus* where a high level of intra-individual polymorphism in the *ITS* region was detected [73], only detailed visual investigation of the variability patterns allowed the inference of the basic phylogenetic pat-

terns. Also, in a study of sexual *Erigeron* species that showed frequent intra-individual polymorphism in *ITS* and *ETS* sequences, a fair degree of handwork had to be done to identify diagnostic/additive patterns, to distinguish informative from uninformative variation, and to detect recombination in cloned sequences [44]. However, reticulation explained most of the observed variation in that case - in contrast to our dataset where the majority of intra-individual variation was uninformative or even misleading.

Basic patterns of Hieracium *evolution*

The basal split into a 'Western' and an 'Eastern' clade suggested by *ETS* was not predicted by morphology nor suggested in any taxonomic treatment. In the light of our data, we suggest several reasons for that failure: (i) both clades include species (groups) with unique morphology, (ii) both comprise species with large or Central European distribution areas, (iii) the large number of interclade hybrids has confounded the picture, and (iv) several morphologically defined taxa had multiple origins and are probably inadequately circumscribed. Dating of this basal divergence is not feasible because of a lack of fossils; fossil pollen cannot be attributed to particular genera in the Asteraceae subfamily Lactuceae [74]. However, there are some indications that suggest a Quaternary timescale, at least for speciation within both major clades (see below).

Low overall *ETS* sequence divergence indicates that, despite rather high morphological diversity even within each clade/subclade, speciation was not accompanied by significant molecular differentiation. Similar observations have been made in other plant genera [75,76]. *ETS* sequence divergence in many other Asteraceae genera is considerably greater [12,77-80]. Exceptions are genera that have a restricted geographic range, were sampled only from a restricted area, underwent rapid recent speciation, experienced population bottlenecks, or a combination of these [12,81,82]. The first two reasons do not apply here as our sampling covered the entire, rather large distribution area of *Hieracium* (North America to Siberia, Scandinavia to the Mediterranean). An indication of rapid recent speciation is the extremely low level of *ITS* variation in *Hieracium* s.str. ([13], and unpublished data). While taxonomic levels are not strictly comparable between groups, even *ITS* sequence divergence, although often markedly lower than *ETS* variation, has proved useful for the inference of plant relationships at species and genus level in numerous studies. This might suggest a rather recent divergence even of the two major lineages identified by the *ETS* region. An especially rapid diversification of lineages could be reflected by the basal polytomies of both major clades. The coincidence of the distribution of many diploid and endemic taxa with known glacial refuge areas [83,84] may indicate that these are remnants of originally much larger populations or higher species diversity. The Quaternary in Europe was characterized by many range expansions/contractions related to climatic oscillations which could have resulted in repeated population bottlenecks [85]; this may help to explain the low genetic diversity in our data. After the retreat of the Pleistocene ice sheets, rapid speciation within the 'Eastern' and 'Western' lineage may have been facilitated by the availability of vast areas open to colonization [86]. Subsequently, sympatric speciation by hybridization among these previously separated lineages - accompanied or followed by polyploidization or introgression at the diploid level - may have provided a quick way to adapt to new ecological conditions and to invade new habitats. The obvious correspondence of the distribution areas of many non-hybrid endemics and diploids with known glacial refuge areas and the abundance of interclade hybrids suggest extensive reticulation after the two lineages came into secondary contact. This is in accordance with Stebbins' secondary contact hypothesis [86], which is corroborated by numerous case studies (e.g., [28,87,88]), albeit mostly at the intraspecific level. Above species level, a similar east-west differentiation that may be related to survival in different glacial refuge areas was found in *Hieracium* subgenus *Pilosella* although in that case it mainly concerns divergent chloroplast haplotypes [89].

In the wake of polyploidization, the onset of apomictic reproduction may also be linked to climatic change during or after the Ice Ages [89]. Most diploids are nowadays confined to the southern part of the distributional range of the genus from where the regions further to the north were later re-colonized, which is also a common pattern in other apomictic plants [90]. The largest sequence divergence was found among and within diploids. There is no indication that polyploidy as such has added much to the overall genetic variation. However, polyploidization, frequent hybrid origin, and fixation and spread of genotypes by apomixis have generated the huge number of taxa which makes *Hieracium* one of the largest plant genera, even if a broad species concept is adopted. The large number of accessions with interclade hybrid origin shows that members of the two major clades were not yet reproductively isolated when the reticulation occurred despite their differences in genome size which are reflected by the genome sizes of their hybrids [46].

Indications for ancient hybridization

Even taxa having composite genomes according to our data have unique morphology - reflecting Zahn's [3] definition of a 'basic' species - which is why their hybrid identity had not been assumed before. However, contemporary hybridization is thought not to play a large role in *Hieracium* s.str. [37]. In the following, we discuss

several lines of evidence for ancient hybridization with a special focus on diploids with hybrid origin.

Geography

Many diploid *Hieracium* species are not sympatric and/or are ecologically isolated. Diploid hybridogenous accessions (or species) that are particularly unlikely to have formed recently are *H. gouani*, *H. gymnocephalum* 1215, *H. lucidum*, *H. plumulosum*, *H. pojoritense*, *H. prenanthoides* 1252, and *H. sparsum*. All occur in areas that have never been glaciated, or only partly. Some of these (*H. lucidum*, diploid *H. prenanthoides*) have particular relict character, i.e., their populations are most likely remnants of a previously much larger distribution, or their localities are typical refugia (*H. gouani*, *H. pojoritense*, *H. plumulosum*). In case of *H. lucidum*, pollen producing polyploids like *H. crinitum* (treated as a subspecies of *H. racemosum* by Zahn [3]), could have played a role in past hybridization events [91]. Nowadays, no population of any other *Hieracium* species occurs in the neighborhood of *H. lucidum*. Thus, recent hybridization events among diploids are highly unlikely.

Molecular patterns

ETS signatures of diploid accessions of *H. prenanthoides* and *H. gymnocephalum* were also found in polyploids of the same species indicating that these diploids with hybrid origin have given rise to widespread polyploids assigned to the same taxon. *Hieracium gymnocephalum* is a morphologically very variable and mostly triploid aggregate - our accession 1215 represents the first report of a diploid in this species [46] - whose variation is thought to be caused by past hybridization or introgression [92]. According to *ETS*, *H. sparsum* occurs at an unresolved position within the 'Eastern' clade, and two accessions from different localities were nearly identical in their molecular features. Thus, genetically, *H. sparsum* s.str. behaves like a 'good species' which is in accordance with morphological and floristic observations. While we analyzed mostly diploid accessions of these taxa, *H. sparsum* s.l., *H. prenanthoides* s.l., and *H. plumulosum* (*H. waldsteinii* s.l.) are predominantly polyploid species and are supposed to have been involved in the formation of many 'intermediate' taxa - some of them widespread -, which is suggestive of extensive ancient hybridization, particularly of the diploid progenitors with hybrid origin.

Reproduction

All diploid accessions analyzed here are sexuals with normal seed production [46] while a rare recent natural hybrid was sterile [37]. Experimental hybridization of diploids also resulted in mainly or completely seed-sterile hybrid progeny [93]. Whilst we did not analyze the breeding system of *H. lucidum* - because of its critically endangered status, no living plants were collected - isozyme analyses showed some variation, which is in accordance with sexual reproduction and also with a small population size over a rather long period of time [91]. Apomicts of *Hieracium* have very variable male fertility, ranging from almost complete sterility to normal pollen fertility [52,94,95]. Triploid apomicts, which represent the vast majority of *Hieracium* species, can produce haploid pollen and could therefore fertilize diploid accessions (resulting in diploid hybrid progeny). This would be the only way to overcome the geographic isolation of most recent diploids and to produce significant numbers of hybrids nowadays. However, pollen from other species can lead to a breakdown of self-incompatibility resulting in autogamy. This so-called mentor effect has been demonstrated to be especially strong for pollen from apomictic triploids in *Hieracium* s.str. [95]. Nowadays, autogamy enforced by contact with foreign pollen is a very efficient way to avoid hybridization.

Finally, since most of these *Hieracium* species had been described prior to 1900, many by Linnaeus himself, it is unlikely that hybridizations that gave rise to 'basic' species occurred within the last few hundred years.

Thus, the unanticipated hybridizations involved in the formation of many of the 'basic' species may actually be rather ancient, and may even date back to the early Holocene (see above). This would also provide a reasonable time frame for the formation and spread of the large number of 'intermediate' *Hieracium* species, whose morphology suggests a hybrid origin involving two or more of the 'basic' species.

Evidence for the contribution of variation from extinct taxa

Several lines of evidence suggest a contribution from either unsampled taxa or from extinct lineages that have left molecular traces in the accessions analyzed. We consider contribution from extinct forms more likely, for the following reasons.

The 'unknown Western 1' ribotype, occurring in six hybrid species, at least four of which have different origins, was not found as the only *ETS* variant in any existing species. The respective accessions were collected from sites as distant as Sicily, Poland and the Balkans. The 'unknown Western 1' ribotype may have originated from a lineage/species that was either once widespread, given the collection sites of the hybrid accessions in which it was found, or its genome has spread along with apomictic polyploids after hybridizations in a Western European region. Similarly, the 'Western' parent of six interclade hybrids represented by five different species corresponded to the 'unknown Western 2' lineage. This ribotype also has a large geographic distribution according to the sampling

sites (France, Southeastern Alps, Southern Balkans). A further ribotype present only in hybrids ('unknown Eastern') occurred in four accessions of three species and was restricted to a particular part of the Balkans (Albania, Montenegro).

Chloroplast DNA haplotypes of several hybridogenous accessions (*H. plumulosum*, *H. olympicum*/*H. gymnocephalum*, *H. mixtum*) were unique and fairly divergent from others given the low overall variation of this marker (Figure 3), i.e., a candidate maternal parent is missing. This could be an indication that more diploids existed than are known today, since apomictic polyploids can only act as pollen donors and therefore cannot contribute to cpDNA diversity. Even allowing for the accumulation of mutations in the cpDNA after ancient hybridizations does not explain why the majority of hybrid chloroplast haplotypes were identical to one of the parental groups inferred from the *ETS*. Almost identical cpDNA haplotypes occurred in *H. olympicum* ('unknown Western 1'/Balkan) and *H. gymnocephalum* ('unknown Western 2'/'unknown Eastern'). Given their genomic composition according to *ETS*, they seem to share a maternal progenitor that does not correspond to any of the identified ribotype lineages.

Whilst insufficient sampling can never be excluded, it is unclear what the identity of the species providing the missing variation could possibly be or where they should occur. The only 'basic' species not included in our study were *H. fuscocinereum*, a Northern European polyploid species with some similarities to *H. murorum*; *H. schmalhausenianum*, a Caucasian endemic with unknown ploidy; and *H. laniferum*, a diploid occurring in parts of Spain. The latter is the only candidate that could theoretically consist exclusively of one of the 'unknown Western' *ETS* ribotypes, but it could well be of hybrid origin itself, like half the other 'basic' species or be genetically indistinguishable from the majority of 'pure' species of the 'Western' clade. In any case, these three unsampled taxa certainly cannot account for all the missing variation, i.e., the second 'unknown Western' and the 'unknown Eastern' ribotype, and at least three divergent hybrid chloroplast haplotypes (even if one of them corresponds to an 'unknown' *ETS* lineage). There is also no explanation why 20% (12 out of 60) of the accessions collected from distant geographic sites should comprise these ribotypes, but the ancestral species have always been missed by chance - for the other subclades whose ribotypes also occurred in hybrids, 'pure' candidate parental species were present. Unsampled intraspecific variation as a potential source of the missing variation is also unlikely. Taxa with more than one accession analyzed showed either (i) the same patterns, independent of ploidy, 'pure' species status, or inferred hybrid origin, or (ii) one accession represented a 'pure' species while others were hybrids involving that species and other taxa. Also, the majority of *Hieracium* populations are polyploid and, thus, are unlikely as predecessors. We therefore consider it more likely that extinct species have left their molecular traces in the genomes of taxa with rather ancient hybrid origin.

Species number also argues for the existence of a large range of extinct diversity. Out of the 46 'basic' species analyzed, which were initially considered as the main evolutionary units, only ten diploid species did *not* show molecular evidence of hybrid origin. Theoretically, ten parental species can produce 45 different hybrid combinations (or 90, if reciprocal crosses are considered as different hybrids). However, at least 32 diploids would be needed to explain the species numbers in *Hieracium*, if only macrospecies were taken into account (500). As it is very unlikely that each possible parental combination would be realized in nature, the original number of diploids should probably be much higher.

Morphological indication of potential hybrid origin involving particular recent species is also missing from the analyzed taxa, a fact that is reflected by their treatment as 'basic' species which by definition do not show evidence of character combinations of other species. In contrast, for the multitude of 'intermediate' taxa not analyzed here, morphology does suggest putative parents. Experimental hybridization of several diploid *Hieracium* species showed that hybrid progeny of the same cross can exhibit a large range of morphological variation, but generally, hybrids were either intermediate between their parents or more similar to the maternal one [93]. Thus, hybrid origin from extant parents should usually be detectable. Taken together with the relict character of most diploids, this points to a greater ancestral species diversity now reduced by extinctions. This could - among several other factors - contribute to explaining some of the taxonomic problems in *Hieracium*: As morphological characters of extinct species are naturally unknown, hybrid origin of extant taxa can easily go undetected due to the lack of one or both parents. Even hybridization of the same two parents may result in strikingly dissimilar phenotypes, depending on the particular contribution of gametes. One of the best known examples is the hybrid origin of three different sunflower species from the same parental species combination [96]. However, as long as the parental species and their morphological diversity are known, correct determination of polymorphic hybrids by an expert taxonomist is still feasible [36,97]. However, if only one parent is extinct, even morphological identification of the same hybrid taxon may become impossible.

Our findings add to an increasing number of studies in which the contribution of extinct parent(s) was inferred from molecular data [30,98-103].

Conclusion

The initial aim of this study was to disentangle relationships and species origins of the taxonomically highly complex *Hieracium* s.str. for the first time using molecular markers. The particular features of the *ETS* dataset, characterized by abundant intra-individual variation, most of which was uninformative and some even misleading, required a very detailed inspection of the nature of this variation in order to retrieve the phylogenetic signal and to identify the unexpected hybrid origins of roughly half of all accessions analyzed (Figure 5). Whether the complex data structure is the result of the particular history and extensive apomixis in *Hieracium*, or whether similar molecular patterns will be found in other groups, remains to be seen. The approach used here for the analysis of intra-individual polymorphism might act as a model for the study of other agamic complexes and other similarly challenging patterns of molecular data.

Authors' contributions

JF conceived of the study, did the data analyses and interpretations, and wrote the manuscript. KK was responsible for molecular data acquisition, provided part of the Methods section and critically revised the molecular data files and the manuscript. JC was responsible for collection and determination of the plants and provided chromosome counts, genome size data, information on breeding systems, and taxonomic background for the paper. All authors read and approved the final manuscript.

Additional material

Additional file 1

***Origins of individual accessions**. A detailed assessment of the origin of each species/accession based on* ETS *features (phylogeny, shared polymorphisms), cpDNA haplotype, ploidy and genome size is given. The file also includes some ecogeographic, morphological and floristic information and a table containing source information for all accessions including the outgroup.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-239-S1.PDF]

Additional file 2

***Patterns of* ETS *recombination**. Cloned sequences of nine interclade hybrid accessions (*H. olympicum, H. mixtum, H. prenanthoides *1187*, H. caesium, H. pilosum *1226/2*, H. villosum *1305*, H. gymnocephalum *1215*, H. heterogynum, H. plumulosum*) and* H. kittanae*, the non-hybrid accession with the highest number of intra-individual polymorphisms, are shown. Cloned sequences are compared with intra-individual polymorphisms and dominant character states revealed by direct sequencing. All cloned accessions with recombinant sequences are included. Patterns of gene conversion and the distribution of accession-specific polymorphisms across the cloned sequences can be traced. Color coding of diagnostic character states as in Figure 5.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-239-S2.PDF]

Additional file 3

***Summary of intra-individual polymorphisms**. All intra-individual polymorphisms are listed according to their position in the alignment along with the accessions in which they occur, discriminating additive, unique, and shared variation. Among the shared polymorphisms, homoplasious and informative ones are identified according to specified criteria.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-239-S3.PDF]

Additional file 4

***Alignment of* Hieracium ETS *sequences**. A FASTA file is given comprising two sequences for each accession (i) one with all polymorphic sites included (lower case letters indicate weak [i.e., second peak small] or indel polymorphisms): the sequence name is composed of the abbreviated species name and the number of the accession/plant, and (ii) one resolved for the major sequence in case of skewed ratios: same label, but preceded by an 'M'. Additionally, cloned sequences are given for the respective accessions: same label, but preceded by the clone number. The positions in the alignment correspond to Table 1, Figure 4, and Additional files 2 and 3.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-239-S4.ZIP]

Additional file 5

***Species/accessions, their origin, cytotype,* ETS *and cpDNA features**. This table summarizes the information about individual accessions.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-239-S5.PDF]

Additional file 6

***ETS phylogeny with ribotypes present only in hybrids**. The file contains the phylogenetic analyses on which Figure 5 is based. One out of 1402 equally parsimonious trees is shown (1752 steps, ri = 0.952, ci = 0.982; 182 variable characters of which 95 were parsimony informative) with bootstrap support indicated above the branches. The strict consensus tree topology corresponds to the branches with support values. Bootstrap values for ML and posterior probabilities for Bayesian analyses are given below the branches. Diploid* Hieracium *species are indicated in boldface. All lineages comprise diploids (the 'unknown Western 1' ribotype also occurs in low amounts in diploid* H. lucidum *which was not cloned). The Wx, Wy, and Ex lineages are represented by all non-recombinant clones of these ribotypes. Identical clones of the same accession were included only once (see also Additional file 2: Patterns of* ETS *recombination). The four hybrid accessions still maintained in Figure 2 were excluded, and also* H. lachenalii *and* H. sparsum *because of inferred chloroplast capture. Figure 5 uses the basic structure of this tree on which all inferred reticulation events were mapped.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2148-9-239-S6.PDF]

Acknowledgements

We cordially thank people who provided plant material: Siegfried Bräutigam, František Krahulec, Patrik Mráz, Marjan Niketić, Zbigniew Szeląg, Torbjørn Tyler, Vladimir Vladimirov, and Jaroslav Zahradníček. Petra Caklová is acknowledged for excellent lab assistance, Laura Farrel and John P. Bailey for language corrections, and F. Krahulec for generous support. We are

also grateful for discussions with S. Bräutigam, P. Mráz, T. Tyler, and Franz Schuhwerk about particular species (Additional file 1: Origins of individual accessions). Václav Mahelka and three anonymous reviewers are acknowledged for their suggestions that have helped us to improve the paper. The Grant Agency of the Czech Republic (206/05/0657), the Academy of Sciences of the Czech Republic (AV0Z60050516), and the Ministry of Education, Youth and Sports of the Czech Republic (0021620828) provided financial support.

References

1. Hörandl E, Paun O, Johansson JT, Lehnebach C, Armstrong T, Chen LX, Lockhart P: **Phylogenetic relationships and evolutionary traits in *Ranunculus* s.l. (Ranunculaceae) inferred from *ITS* sequence analysis.** *Mol Phylogenet Evol* 2005, **36:**305-327.
2. Stace CA: **Sectional names in the genus *Hieracium* (Asteraceae) sensu stricto.** *Edinb J Bot* 1998, **55:**417-441.
3. Zahn KH: **Compositae -*Hieracium*.** In *Das Pflanzenreich IV/280* Edited by: Engler A. Leipzig: W. Engelmann:1921-1923.
4. Sell PD: **An introduction to the study of the British Hieracia, 1. History and classification.** *Watsonia* 1987, **16:**365-371.
5. **Published chromosome-counts in *Hieracium*** [http:www.botanik.biologie.uni-muenchen.de/botsamml/projects/chrzlit.html]
6. Schuhwerk F, Lippert W: **Chromosomenzahlen von *Hieracium* (Compositae, Lactuceae) Teil 2.** *Sendtnera* 1998, **5:**269-286.
7. Merxmüller H: **Diploide Hieracien.** *Anal Inst Bot Cavanilles* 1975, **32:**189-196.
8. Nieto Feliner G, Rosselló JA: **Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants.** *Mol Phylogenet Evol* 2007, **44:**911-919.
9. Álvarez I, Wendel JF: **Ribosomal ITS sequences and plant phylogenetic inference.** *Mol Phylogenet Evol* 2003, **29:**417-434.
10. Volkov RA, Komarova NY, Hemleben V: **Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression.** *Syst Biodivers* 2007, **5:**261-276.
11. Baldwin G, Markos S: **Phylogenetic utility of the external transcribed spacer (ETS) of 18-26S rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae).** *Mol Phylogenet Evol* 1998, **10:**449-463.
12. Linder CR, Goertzen LR, Heuvel BV, Francisco-Ortega J, Jansen RK: **The complete external transcribed spacer of 18S-26S rDNA: Amplification and phylogenetic utility at low taxonomic levels in Asteraceae and closely allied families.** *Mol Phylogenet Evol* 2000, **14:**285-303.
13. Fehrer J, Gemeinholzer B, Chrtek J Jr, Bräutigam S: **Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae).** *Mol Phylogenet Evol* 2007, **42:**347-361.
14. Andreasen K, Baldwin BG: **Nuclear ribosomal DNA sequence polymorphism and hybridization in checker mallows (Sidalcea, Malvaceae).** *Mol Phylogenet Evol* 2003, **29:**563-581.
15. Beardsley PM, Schoenig SE, Whittall JB, Olmstead RG: **Patterns of evolution in Western North American *Mimulus* (Phrymaceae).** *Amer J Bot* 2004, **91:**474-489.
16. Hoggard GD, Kores PJ, Molvray M, Hoggard RK: **The phylogeny of *Gaura* (Onagraceae) based on ITS, ETS, and trnL-F sequence data.** *Amer J Bot* 2004, **91:**139-148.
17. Soltis DE, Soltis PS, Pires JC, Kovařík A, Tate JA, Mavrodiev E: **Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons.** *Biol J Linn Soc* 2004, **82:**485-501.
18. Weeks A, Simpson BB: **Molecular genetic evidence for interspecific hybridization among endemic Hispaniolan *Bursera* (Burseraceae).** *Amer J Bot* 2004, **91:**976-984.
19. Devos N, Raspé O, Jacquemart AL, Tyteca D: **On the monophyly of *Dactylorhiza* Necker ex Nevski (Orchidaceae): is *Coeloglossum viride* (L.) Hartman a *Dactylorhiza*?** *Biol J Linn Soc* 2006, **152:**261-269.
20. Suárez-Santiago VN, Salinas MJ, Romero-García AT, Garrido-Ramos MA, de la Herrán R, Ruiz-Rejón C, Ruiz-Rejón M, Blanca G: **Polyploidy, the major speciation mechanism in *Muscari* subgenus *Botryanthus* in the Iberian Peninsula.** *Taxon* 2007, **56:**1171-1184.
21. Timme RE, Simpson BB, Linder CR: **High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18S-26S ribosomal DNA external transcribed spacer.** *Amer J Bot* 2007, **94:**1837-1852.
22. Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ: **The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny.** *Ann Miss Bot Garden* 1995, **82:**247-277.
23. Arnheim N: **Concerted evolution of multigene families.** In *Evolution of genes and proteins* Edited by: Nei M, Koehn R. Sunderland, Mass.: Sinauer; 1983:38-61.
24. Campbell CS, Wojciechowski MF, Baldwin BG, Alice LA, Donoghue MJ: **Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae).** *Mol Biol Evol* 1997, **14:**81-90.
25. Gandolfi A, Bonilauri P, Rossi V, Menozzi P: **Intraindividual and intraspecies variability of ITS1 sequences in the ancient asexual *Darwinula stevensoni* (Crustacea: Ostracoda).** *Heredity* 2001, **87:**449-455.
26. Mes THM, Cornelissen AWCA: **Non-homogenized ITS regions in the parasitic nematode *Cooperia oncophora*.** *Parasitology* 2004, **129:**213-222.
27. Kim K-J, Jansen RK: **Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, Asteraceae): Additional information from internal transcribed spacer sequences of nuclear ribosomal DNA.** *Plant Syst Evol* 1994, **190:**157-185.
28. Sang T, Crawford DJ, Stuessy TF: **Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution.** *Proc Natl Acad Sci USA* 1995, **92:**6813-6817.
29. O'Kane SL Jr, Schaal BA, Al-Shebaz IA: **The origins of *Arabidopsis suecica* (Brassicaceae) as indicated by nuclear rDNA sequences.** *Syst Bot* 1996, **21:**559-566.
30. Vargas P, McAllister HA, Morton C, Jury SL, Wilkinson MJ: **Polyploid speciation in *Hedera* (Araliaceae): phylogenetic and biogeographic insights based on chromosome counts and ITS sequences.** *Plant Syst Evol* 1999, **219:**165-179.
31. Whittall J, Liston A, Gisler S, Meinke RJ: **Detecting nucleotide additivity from direct sequences is a SNAP: An example from *Sidalcea* (Malvaceae).** *Plant Biol* 2000, **2:**211-217.
32. Rauscher JT, Doyle JJ, Brown AHD: **Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex.** *Mol Ecol* 2002, **11:**2691-2702.
33. Fuertes Aguilar J, Nieto Feliner G: **Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (*Armeria*, Plumbaginaceae).** *Mol Phylogenet Evol* 2003, **28:**430-447.
34. Nieto Feliner G, Gutiérrez Larena B, Fuertes Aguilar J: **Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae).** *Ann Bot* 2004, **93:**189-200.
35. Lorenz-Lemke AP, Muschner VC, Bonatto SL, Cervi AC, Salzano FM, Freitas LB: **Phylogeographic inferences concerning evolution of Brazilian *Passiflora actinia* and *P. elegans* (Passifloraceae) based on ITS (nrDNA) variation.** *Ann Bot* 2005, **95:**799-806.
36. Kaplan Z, Fehrer J: **Molecular evidence for a natural primary triple hybrid in plants revealed from direct sequencing.** *Ann Bot* 2007, **99:**1213-1222.
37. Mráz P, Chrtek J Jr, Fehrer J, Plačková I: **Rare recent natural hybridization in *Hieracium* s.str. - evidence from morphology, allozymes and chloroplast DNA.** *Plant Syst Evol* 2005, **255:**177-192.
38. Mateo G: **Aportaciones al conocimiento del género *Hieracium* en España, X. Novedades para el Pirineo catalán.** *Fl Montiber* 2005, **31:**62-69.
39. Vladimirov V: **A new diploid *Hieracium* (Asteraceae: Lactuceae) from Bulgaria.** *Bot J Linn Soc* 2003, **143:**213-218.
40. Vladimirov V, Szeląg Z: **A new diploid species of *Hieracium* sect. *Pannosa* (Asteraceae) from Bulgaria.** *Bot J Linn Soc* 2006, **150:**261-265.
41. Nyárády EI: ***Hieracium* L.** In *Flora Republicii Populare Romîne Volume 10*. Edited by: Nyárády EI. Bucureşti: Editura Academiei Republicii Populare Romîne; 1965:214-746.

42. De Retz B: ***Hieracium*.** In *Flore descriptive et illustrée de la France par l'abbé H. Coste. troisième supplément* Edited by: Jovet P, de Vilmorin R. Paris: Librairie Scientifique et Technique Albert Blanchard; 1975:244-297.
43. Štorchová H, Hrdličková R, Chrtek J Jr, Tetera M, Fitze D, Fehrer J: **An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution.** *Taxon* 2000, **49:**79-84.
44. Noyes RD: **Intraspecific nuclear ribosomal DNA divergence and reticulation in sexual diploid *Erigeron strigosus* (Asteraceae).** *Amer J Bot* 2006, **93:**470-479.
45. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis suite.** *Nucl Acids Symp Ser* 1999, **41:**95-98.
46. Chrtek J Jr, Zahradníček J, Krak K, Fehrer J: **Genome size in *Hieracium* subgen. *Hieracium* (Compositae) strongly correlates with major phylogenetic groups.** *Ann Bot* 2009, **104:**161-178.
47. Markos S, Baldwin BG: **Structure, molecular evolution, and phylogenetic utility of the 5' region of the external transcribed spacer of 18S-26S rDNA in *Lessingia* (Compositae, Astereae).** *Mol Phylogenet Evol* 2002, **23:**214-228.
48. Swofford DL: **PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods).** Version 4. Sunderland, Mass.: Sinauer; 2002.
49. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** *Bioinformatics* 2001, **17:**754-755.
50. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572-1574.
51. Posada D, Crandall KA: **Modeltest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14:**817-818.
52. Mráz P, Chrtek J, Šingliarová B: **Geographical parthenogenesis, genome size variation and pollen production in the arctic-alpine species *Hieracium alpinum*.** *Bot Helv* 2009, **119:**41-51.
53. Kluge AG, Wolf AJ: **Cladistics: what's in a word?** *Cladistics* 1993, **9:**183-199.
54. Mahelka V, Fehrer J, Krahulec F, Jarolímová V: **Recent natural hybridization between two allopolyploid wheatgrasses (*Elytrigia*, Poaceae): Ecological and evolutionary implications.** *Ann Bot* 2007, **100:**249-260.
55. Fuertes Aguilar J, Rosselló JA, Nieto Feliner G: **Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae).** *Mol Ecol* 1999, **8:**1341-1346.
56. Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, Wagner GP, Bartels J, Murtha M, Pendleton J: **Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift.** *Syst Biol* 1994, **43:**250-261.
57. Mayol M, Rosselló JA: **Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*.** *Mol Phylogenet Evol* 2001, **19:**167-176.
58. Bailey CD, Carr TG, Harris SA, Hughes CE: **Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes.** *Mol Phylogenet Evol* 2003, **29:**435-455.
59. Göker M, Grimm GW: **General functions to transform associate data to host data, and their use in phylogenetic inference from sequences with intra-individual variability.** *BMC Evol Biol* 2008, **8:**86.
60. Harris DJ, Crandall KA: **Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): Implications for phylogenetic and microsatellite studies.** *Mol Biol Evol* 2000, **17:**284-291.
61. Quijada A, Liston A, Robinson W, Alvarez-Buylla E: **The ribosomal ITS region as a marker to detect hybridization in pines.** *Mol Ecol* 1997, **6:**995-996.
62. Wichman SR, Wright SD, Cameron EK, Keeling DJ, Gardner RC: **Elevated genetic heterogeneity and Pleistocene climatic instability: inferences from nrDNA in New Zealand *Coprosma* (Rubiaceae).** *J Biogeogr* 2002, **29:**943-954.
63. Buckler ES, Holtsford TP: ***Zea* ribosomal repeat evolution and substitution patterns.** *Mol Biol Evol* 1996, **13:**623-632.
64. Kita Y, Ito M: **Nuclear ribosomal ITS sequences and phylogeny in East Asian *Aconitum* subgenus *Aconitum* (Ranunculaceae), with special reference to extensive polymorphism in individual plants.** *Plant Syst Evol* 2000, **225:**1-13.
65. Manen J-F: **Are both sympatric species *Ilex perado* and *Ilex canariensis* secretly hybridizing? Indication from nuclear markers collected in Tenerife.** *BMC Evol Biol* 2004, **4:**46.
66. Ruggiero MV, Procaccini G: **The rDNA ITS region in the Lessepsian marine angiosperm *Halophila stipulacea* (Forssk.) Aschers. (Hydrocharitaceae): Intragenomic variability and putative pseudogenic sequences.** *J Mol Evol* 2004, **58:**115-121.
67. Harpke D, Peterson A: **Non-concerted ITS evolution in *Mammillaria* (Cactaceae).** *Mol Phylogenet Evol* 2006, **41:**579-593.
68. Pamilo P, Nei M: **Relationships between gene trees and species trees.** *Mol Biol Evol* 1988, **5:**568-583.
69. Nieto Feliner G, Fuertes Aguilar J, Rosselló JA: **Can extensive reticulation and concerted evolution result in a cladistically structured molecular dataset?** *Cladistics* 2001, **17:**301-312.
70. Soltis DE, Mavrodiev EV, Doyle JJ, Rauscher J, Soltis PS: **ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids.** *Syst Bot* 2008, **33:**7-20.
71. Huber KT, Oxelman B, Lott M, Moulton V: **Reconstructing the evolutionary history of polyploids from multilabeled trees.** *Mol Biol Evol* 2006, **23:**1784-1791.
72. Joly S, Bruneau A: **Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: An example from *Rosa* in North America.** *Syst Biol* 2006, **55:**623-636.
73. Grimm GW, Denk T, Hemleben V: **Coding of intraspecific nucleotide polymorphisms: a tool to resolve reticulate evolutionary relationships in the ITS of beech trees (*Fagus* L., Fagaceae).** *Syst Biodivers* 2007, **5:**291-309.
74. Tomb AS: **Pollen morphology in tribe Lactuceae (Compositae).** *Grana* 1975, **15:**79-89.
75. Kelch DG, Baldwin BG: **Phylogeny and ecological radiation of New World thistles (*Cirsium*, Cardueae-Compositae) based on ITS and ETS rDNA sequence data.** *Mol Ecol* 2003, **12:**141-151.
76. Ford KA, Ward JM, Smissen RD, Wagstaff SJ, Breitwieser I: **Phylogeny and biogeography of *Craspedia* (Asteraceae: Gnaphalieae) based on ITS, ETS and psbA-trnH sequence data.** *Taxon* 2007, **56:**783-794.
77. Chan R, Baldwin BG, Ornduff R: **Goldfields revisited: A molecular phylogenetic perspective on the evolution of *Lasthenia* (Compositae: Heliantheae sensu lato).** *Int J Pl Sci* 2001, **162:**1347-1360.
78. Ekenäs C, Baldwin BG, Andreasen K: **A molecular phylogenetic study of *Arnica* (Asteraceae): Low chloroplast DNA variation and problematic subgeneric classification.** *Syst Bot* 2007, **32:**917-928.
79. Mitsui Y, Chen ST, Zhou ZK, Peng CI, Deng YF, Setoguchi H: **Phylogeny and biogeography of the genus *Ainsliaea* (Asteraceae) in the Sino-Japanese region based on nuclear rDNA and plastid DNA sequence data.** *Ann Bot* 2008, **101:**111-124.
80. Tkach NV, Hoffmann MH, Roser M, Korobkov AA, von Hagen KB: **Parallel evolutionary patterns in multiple lineages of arctic *Artemisia* L. (Asteraceae).** *Evolution* 2008, **62:**184-198.
81. Roberts RP, Urbatsch LE: **Molecular phylogeny of *Ericameria* (Asteraceae, Astereae) based on nucler ribosomal 3 ' *ETS* and *ITS* sequence data.** *Taxon* 2003, **52:**209-228.
82. Fehlberg SD, Ranker TA: **Phylogeny and biogeography of *Encelia* (Asteraceae) in the sonoran and peninsular deserts based on multiple DNA sequences.** *Syst Bot* 2007, **32:**692-699.
83. Comes HP, Kadereit JW: **The effect of quaternary climatic changes on plant distribution and evolution.** *Trends Plant Sci* 1998, **3:**432-438.
84. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF: **Comparative phylogeography and postglacial colonization routes in Europe.** *Mol Ecol* 1998, **7:**453-464.
85. Hewitt GM: **Some genetic consequences of ice ages, and their role in divergence and speciation.** *Biol J Linn Soc* 1996, **58:**247-276.
86. Stebbins GL: **Polyploidy and the distribution of the arctic-alpine flora: new evidence and a new approach.** *Bot Helv* 1984, **94:**1-13.
87. Nettel A, Dodd RS, Afzal-Rafii Z, Tovilla-Hernández C: **Genetic diversity enhanced by ancient introgression and secondary contact in East Pacific black mangroves.** *Mol Ecol* 2008, **17:**2680-2690.

88. Kappas I, Baxevanis AD, Maniatsi S, Abatzopoulos TJ: **Porous genomes and species integrity in the branchiopod *Artemia*.** *Mol Phylogenet Evol* 2009, **52:**192-204.
89. Fehrer J, Krahulcová A, Krahulec F, Chrtek J Jr, Rosenbaumová R, Bräutigam S: **Evolutionary aspects in *Hieracium* subgenus *Pilosella*.** In *Apomixis: Evolution, Mechanisms and Perspectives Volume 147*. Edited by: Hörandl E, Grossniklaus U, van Dijk P, Sharbel T. Regnum Vegetabile. Königstein: Koeltz; 2007:359-390.
90. Hörandl E: **The complex causality of geographical parthenogenesis.** *New Phytol* 2006, **171:**525-538.
91. Di Gristina E, Geraci A, Raimondo FM: **Biosystematic investigation on *Hieracium symphytifolium* (Asteraceae).** *Bocconea* 2006, **19:**275-286.
92. Niketić M, Vladimirov V, Mráz P: **Chromosome numbers and taxonomic-chorological notes on selected species of *Hieracium* s.str. (Asteraceae) from Montenegro.** *Phytol Balcan* 2006, **12:**85-97.
93. Mráz P, Paule J: **Experimental hybridization in the genus *Hieracium* s. str.: crosses between diploid taxa.** *Preslia* 2006, **78:**1-26.
94. Rosenberg O: **Die semiheterotypische Teilung und ihre Bedeutung für die Entstehung verdoppelter Chromosomenzahlen.** *Hereditas* 1927, **8:**305-338.
95. Mráz P: **Mentor effects in the genus *Hieracium* s.str. (Compositae, Lactuceae).** *Folia Geobot* 2003, **38:**345-350.
96. Rosenthal DM, Schwarzbach AE, Donovan LA, Raymond O, Rieseberg LH: **Phenotypic differentiation between three ancient hybrid taxa and their parental species.** *Int J Plant Sci* 2002, **163:**387-398.
97. Kaplan Z, Fehrer J: **Evidence for the hybrid origin of *Potamogeton* × *cooperi* (Potamogetonaceae): traditional morphology-based taxonomy and molecular techniques in concert.** *Folia Geobot* 2004, **39:**431-453.
98. Roelofs D, van Velzen J, Kuperus P, Bachmann K: **Molecular evidence for an extinct parent of the tetraploid species *Microseris acuminata* and *M. campestris* (Asteraceae, Lactuceae).** *Mol Ecol* 1997, **6:**641-649.
99. Moore MJ, Francisco-Ortega J, Santos-Guerra A, Jansen RK: **Chloroplast DNA evidence for the roles of island colonization and extinction in *Tolpis* (Asteraceae: Lactuceae).** *Amer J Bot* 2002, **226:**23-33 [http://www.amjbot.org/cgi/content/abstract/89/3/518].
100. Ritz CM, Schmuths S, Wissemann V: **Evolution by reticulation: European dogroses originated by multiple hybridization across the genus *Rosa*.** *J Heredity* 2005, **96:**4-14.
101. Hedge SG, Nason JD, Clegg JM, Ellstrand NC: **The evolution of California's wild radish has resulted in the extinction of its progenitors.** *Evolution* 2006, **60:**1187-1197.
102. Popp M, Oxelman B: **Origin and evolution of North American polyploid *Silene* (Caryophyllaceae).** *Amer J Bot* 2007, **94:**330-349.
103. Brysting AK, Oxelman B, Huber KT, Moulton V, Brochmann C: **Untangling complex histories of genome mergings in high polyploids.** *Syst Biol* 2007, **56:**467-476.

Annals of Botany **104**: 161–178, 2009
doi:10.1093/aob/mcp107, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups

Jindřich Chrtek Jr[1,2,*], Jaroslav Zahradníček[2], Karol Krak[1] and Judith Fehrer[1]

[1]*Institute of Botany, Academy of Sciences of the Czech Republic, CZ-252 43 Průhonice, Czech Republic and* [2]*Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, CZ-128 01, Prague, Czech Republic*

Received: 22 September 2008 Returned for revision: 22 December 2008 Accepted: 30 March 2009 Published electronically: 11 May 2009

• *Background and Aims Hieracium* subgenus *Hieracium* is one of the taxonomically most intricate groups of vascular plants, due to polyploidy and a diversity of breeeding systems (sexuality vs. apomixis). The aim of the present study was to analyse nuclear genome size in a phylogenetic framework and to assess relationships between genome size and ploidy, breeding system and selected ecogeographic features.
• *Methods* Holoploid and monoploid genome sizes (C- and Cx-values) of 215 cultivated plants from 89 field populations of 42 so-called 'basic' *Hieracium* species were determined using propidium iodide flow cytometry. Chromosome counts were available for all analysed plants, and all plants were tested experimentally for their mode of reproduction (sexuality vs. apomixis). For constructing molecular phylogenetic trees, the external transcribed spacer region of nuclear ribosomal DNA was used.
• *Key Results* The mean 2C values differed up to 2·37-fold among different species (from 7·03 pg in diploid to 16·67 in tetraploid accessions). The 1Cx values varied 1·22-fold (between 3·51 and 4·34 pg). Variation in 1Cx values between conspecific (species in a broad sense) accessions ranged from 0·24% to 7·2%. Little variation (not exceeding the approximate measurement inaccurracy threshold of 3·5%) was found in 33 species, whereas variation higher than 3·5% was detected in seven species. Most of the latter may have a polytopic origin. Mean 1Cx values of the three cytotypes (2*n*, 3*n* and 4*n*) differed significantly (average of 3·93 pg in diploids, 3·82 pg in triploids and 3·78 pg in tetraploids) indicating downsizing of genomes in polyploids. The pattern of genome size variation correlated well with two major phylogenetic clades which were composed of species with western or eastern European origin. The monoploid genome size in the 'western' species was significantly lower than in the 'eastern' ones. Correlation of genome size with latitude, altitude and selected ecological characters (light and temperature) was not significant. A longitudinal component was only apparent for the whole data set, but absent within the major lineages.
• *Conclusions* Phylogeny was the most important factor explaining the pattern of genome size variation in *Hieracium sensu stricto*, species of western European origin having significantly lower genome size in comparison with those of eastern European origin. Any correlation with ecogeographic variables, including longitude, was outweighed by the divergence of the genus into two major phylogenetic lineages.

Key words: Apomixis, chromosome numbers, Compositae, genome size, hawkweeds, *Hieracium* subgenus *Hieracium*, mode of reproduction, nuclear DNA content, phylogeny, polyploidy.

INTRODUCTION

Genome size has become a widely studied phenomenon, beginning in the 1950s when large differences in the nuclear content of different organisms were detected (e.g. Swift, 1950; Laurie and Bennett, 1985; Bennett and Leitch, 1995). Large differences in DNA content can be caused by several mechanisms. It has been found that nuclear DNA content is primarily influenced by the proportion of non-genic repetitive DNA, much of which is generated by transposable elements (Flavell *et al.*, 1977; Barakat *et al.*, 1997). In particular, it has been found that retrotransposon copy number can vary among genomes (Arumuganathan and Earle, 1991; Vicient *et al.*, 1999; Kalendar *et al.*, 2000; Piegu *et al.*, 2006; Wicker and Keller, 2007; Hawkins *et al.*, 2008). Decrease in genome size can result from a higher overall rate of deletions than insertions, selection against transposable elements, unequal crossing over and illegitimate recombination (Morgan, 2001; Petrov, 2002; Wendel *et al.*, 2002; Ma *et al.*, 2004; Bennetzen *et al.*, 2005).

Correlations between genome size and specific life traits, most importantly life history and breeding systems, have been documented. Selfers were found to have smaller Cx-values than related outcrossers (Labani and Elkington, 1987; Govindaraju and Cullis, 1991; Albach and Greilhuber, 2004). Annuals, especially weedy species, tend to have lower genome size in comparison with related perennials (Bennett, 1972; Rejmanek and Richardson, 1996; Bennett *et al.*, 1998; Garnatje *et al.*, 2004; Grotkopp *et al.*, 2004), probably due to an association of annual life history with selfing (e.g. Albach and Greilhuber, 2004). Relationships between genome size and ecological factors are less clear (see, for example, Knight *et al.*, 2005). Correlations between genome size and frost resistance in the British flora (MacGillivray and Grime, 1995), elevation in some groups of *Centaurea* (Bancheva and Greilhuber, 2006), *Veronica* (Albach and Greilhuber, 2004), *Dactylis* (Reeves *et al.*, 1998) and

* For correspondence. E-mail chrtek@ibot.cas.cz

Berberis (Bottini *et al.*, 2000) and continentality and habitat conditions (moisture) in *Cirsium* (Bureš *et al.*, 2004) have already been documented. There are also more or less close associations between genome size and cell size and leaf anatomical traits (Bennett, 1972; Edwards and Endrizzi, 1975; Knight *et al.*, 2005; Sugiyama, 2005; Beaulieu *et al.*, 2008), cell cycle duration (e.g. Rees *et al.*, 1966; Bennett *et al.*, 1983; Lawrence, 1985), seed mass (e.g. Knight *et al.*, 2005; Beaulieu *et al.*, 2007) and photosynthetic rate (e.g. Knight *et al.*, 2005).

Polyploids often have smaller Cx-values than their diploid relatives (e.g. Leitch and Bennett, 2004; Weiss-Schneeweiss *et al.*, 2006). These decreases correlate with a mutational bias towards deletion over insertions (Petrov, 2002), and illegitimate recombination has been shown to eliminate retrotransposon sequences (Bennetzen, 2002; Devos *et al.*, 2002; Ma *et al.*, 2004). However, exceptions of this downsizing pattern have been found, e.g. in the genus *Orobanche* (tetraploid *O. transcaucasica*). It was hypothesized that such polyploids are relatively young and that there was not enough time for a substantial reduction in nuclear DNA content (Weiss-Schneeweiss *et al.*, 2006).

Genome size alone is of little value as a phylogenetic indicator at higher taxonomic levels, but can be helpful in infrageneric classification assessments, species delimitation or hybrid identification (Keller *et al.*, 1996; Buitendijk *et al.*, 1997; Morgan *et al.*, 1998; Thalmann *et al.*, 2000; Zonneveld, 2001; Šiško *et al.*, 2003; Bureš *et al.*, 2004; Baack *et al.*, 2005; Záveský *et al.*, 2005; Suda *et al.*, 2007). An important issue that is still largely neglected in the literature, mostly due to a lack of comparative analyses between DNA sequence and genome size data sets, is the understanding of how genome size variation is linked with species evolution (but see Wendel *et al.*, 2002; Albach and Greilhuber, 2004; Grotkopp *et al.*, 2004; Jakob *et al.*, 2004; Weiss-Schneeweiss *et al.*, 2006). Species relationships were therefore assessed based on the external transcribed spacer (ETS) of nuclear ribosomal DNA in order to relate genome size variation to their evolutionary history.

Hieracium subgenus *Hieracium* is distributed in temperate regions of Europe, Asia, Mediterranean Africa and North America and has been introduced to several other regions, e.g. New Zealand. The genus is suitable for the study of genome size variation due its remarkable diversity in ploidy (coupled with breeding systems), habitat preferences and distribution of particular species. Polyploid (triploid, tetraploid and rarely pentaploid, $x = 9$) taxa with asexual reproduction through parthenogenetic development of the unreduced egg cell (*Antennaria*-type diplospory) prevail in this group, i.e. they are (near-)obligate apomicts (e.g. Nogler, 1984). Sexual reproduction is rather rare and restricted to diploid species (Schuhwerk, 1996; Chrtek *et al.*, 2004). The species occupy forests, forest margins, various grasslands and rocks from the lowlands to the alpine belt.

Species concepts in *Hieracium* have long been a matter of discussion (e.g. Schuhwerk, 2003). The Central European school of 'hieraciology' (founded by Nägeli and Peter in the 19th century) accepts a broad species definition (species are then divided into subspecies, varieties, etc.), whereas Scandinavian, British and Russian botanists follow a narrow species concept, i.e. nearly all morphologically recognizable forms are treated at species rank ('microspecies'). We follow the Central European concept because, in our opinion, it better reflects the situation across the whole distribution area, especially in central and southern Europe where most diploids occur, from which the apomictic polyploids are thought to be derived. According to this concept, approx. 500 species (in the broad sense) are accepted (Zahn, 1921–1923; and species described since that time), being either so-called 'basic' or 'intermediate' taxa. The latter share morphological characters of two or more basic species and are supposed to be of hybridogenous origin (hybrids stabilized by agamospermy). Basic species (about 45, including diploids and polyploids) are tentatively considered as main units of species evolution in *Hieracium*.

Here, a nuclear DNA content analysis of 42 basic species of *Hieracium* subgenus *Hieracium* (*sensu* Zahn, 1921–1923, with a few exceptions, see Materials and methods) is reported. The following questions were addressed: (*a*) how does the level of intraspecific variation in holoploid and monoploid genome sizes relate to the circumscription of species *sensu* Zahn? (*b*) how does monoploid genome size (Cx) relate to ploidy (diploids, triploids and tetraploids), i.e. is there evidence for downsizing of genomes in polyploids? (*c*) is there any congruence between the phylogenetic structure and the pattern of genome size variation? and (*d*) how does nuclear genome size relate to selected ecogeographic features (latitude, longitude, altitude, temperature and light)?

MATERIALS AND METHODS

Plant material

Two hundred and fifteen samples from 89 populations of 42 *Hieracium* species were collected in the field (or grown from seeds in a few cases) throughout Europe and transferred to the experimental garden of the Institute of Botany in Průhonice (Table 1; for details of sample localities see Supplementary Data, available online). Taxon sampling was restricted to so-called 'basic', supposedly non-hybridogenous species, generally following Zahn (1921–1923) with a few exceptions. The species concept of section *Cerinthoidea* followed Mateo (2005), *H. plumulosum* (*H. waldsteinii sensu lato*) was treated as a separate species, and two newly described diploid Balkan species (*H. kittanae* and *H. petrovae*; Vladimirov, 2003; Vladimirov and Szeląg, 2006) were included. Complete analysis covering all recognized and mostly (allo)polyploid hybridogenous species (approx. 500 'broad' species) was not feasible, and interpretation of estimated genome sizes would be extremely complicated due to the often unknown origin of polyploids and the reticulate patterns of variation.

For diploid, sexually reproducing species and for agamospermous polyploids with a rather small distribution area, one or two populations were chosen. For sexual diploids with large geographic areas and for more widely distributed agamospermous polyploids, two to six populations were selected. The number of plants analysed per population varied from two in agamospermous species with likely clonal population structure (e.g. Shi *et al.*, 1996; Mráz *et al.*, 2001; Štorchová *et al.*,

TABLE 1. *Accession origin and genome size*

Species	Locality (no. of plants) (S): plant cultivated from seed	2*n*	2C (pg) mean ± s.e.	2C (pg) range	1Cx (pg)	1Cx (pg) species mean ± s.e.	1Cx (pg) species range	1Cx species variation (%)	ETS clade†
H. alpinum L.	Ukraine: Chornohora (4)	18	7·90 ± 0·01	7·88–7·92	3·95	3·95 ± 0·01	3·94–3·97	0·76	E
	France: Hautes Alpes (1)	27	11·87	–	3·96				n.d.
H. amplexicaule L.	Austria: Hohe Tauern, Fragant (2)	27*	10·70 ± 0·01	10·7–10·71	3·57	3·60 ± 0·02	3·54–3·68	3·95	X(W)
	Austria: Hohe Tauern, Mallnitz (1)	27*	10·74	–	3·58				n.d.
	Spain: prov. Gerona (1)	27*	10·8	–	3·6				n.d.
	France: Hautes Alpes (1)	27	10·62	–	3·54				n.d.
	Italy: Rhaetian Alps (2)	36*	14·66 ± 0·07	14·59–14·72	3·67				n.d.
H. bifidum Kit.	Slovakia: Roháče (1)	27	10·67	–	3·56	3·53 ± 0·01	3·51–3·56	1·42	W
	Czech Rep.: distr. Beroun	27	10·52 ± 0·01	10·52–10·53	3·51				n.d.
	Czech Rep.: Krkonoše Mts.	27	10·62 ± 0·01	10·61–10·63	3·54				n.d.
H. bracteolatum Sibth. & Sm.	Greece: Pilion (2) (S)	27	12·39 ± 0·08	12·31–12·47	4·13	4·13 ± 0·03	4·10–4·16	1·46	X
H. bupleuroides C.C.Gmel. I.	Slovakia: Biele Karpaty (2)	27*	11·95 ± 0·05	11·91–12·01	3·99				E(H)
H. bupleuroides C.C.Gmel. II.	Slovakia: Chočské vrchy (2)	27*	12·00 ± 0·02	11·96–12·00	3·99				n.d.
	Slovakia: Roháče (2)	27	11·73	–	3·91				E
	Austria: Dachstein massif (1)	27*	11·61 ± 0·05	11·56–11·66	3·87				n.d.
	Slovakia: Slovenský raj (1)	27*	12·03	–	4·01				n.d.
	Austria: Allgäuer Alpen (1)	27	11·63	–	3·88				n.d.
	H. bupleuroides s.l. mean					3·95 ± 0·06	3·85–4·01	4·20	
H. caesium (Fr.) Fr.	Sweden: prov. Gotland (3)	36*	14·66 ± 0·12	14·52–14·89	3·67	3·68 ± 0·01	3·64–3·72	2·20	X(W)
	Sweden: prov. Gästrikland (3)	36*	14·75 ± 0·05	14·66–14·83	3·69				n.d.
H. cerinthoides L.	Spain: Pyrenees (2)	27*	10·68 ± 0·03	10·66–10·71	3·56	3·56 ± 0·01	3·55–3·57	0·56	W(H)
H. cordifolium Lapeyr.	Andorra: Pyrenees (6)	18*	7·18 ± 0·02	7·11–7·22	3·59	3·59 ± 0·01	3·56–3·61	1·40	W(H)
H. eriophorum St.-Amans	France: Landes, Labenne (6) (S)	18*	8·51 ± 0·03	8·45–8·61	4·25	4·27 ± 0·01	4·22–4·31	2·13	E
	France: Landes, Vieux-Boucau-les-Bains (6) (S)	18*	8·55 ± 0·03	8·44–8·61	4·27				n.d.
H. glaucum All.	Slovenia: Julijske Alpe, Podklanec (2)	27*	11·31 ± 0·03	11·29–11·34	3·77	3·79 ± 0·01	3·76–3·81	1·33	n.d.
	Slovenia: Julijske Alpe, Zadnjica (2)	27*	11·42 ± 0·00	11·42–11·42	3·81				n.d.
	Slovenia: Julijske Alpe, izvir Soče (2)	27	11·39 ± 0·05	11·34–11·44	3·8				X
H. gouani Arv.-Touv.	Spain: prov. Gerona (8)	18*	7·10 ± 0·01	7·07–7·12	3·55	3·55 ± 0·00	3·54–3·56	0·56	X(W)
H. gymnocephalum Griseb. ex Pant.	Albania: Jezerce (2)	18	8·44 ± 0·01	8·43–8·45	4·22	4·23 ± 0·01	4·22–4·23	0·24	X
H. gymnocerinthe Arv.-Touv. & G.Gaut.	Spain: Serra del Cadí (2)	27*	10·6 ± 0·02	10·58–10·61	3·53	3·54 ± 0·01	3·53–3·54	0·28	W(H)
H. heterogynum (Froel.) Guterm.	Montenegro: Lovčen (2) (S)	27	12·52 ± 0·02	12·52–12·55	4·17	4·18 ± 0·01	4·17–4·18	0·24	X
H. humile Jacq.	Austria: Dachstein massif (2)	36*	14·25 ± 0·01	14·24–14·26	3·56	3·55 ± 0·01	3·54–3·57	0·85	W
	France: Corbières (2)	27	10·64 ± 0·01	10·61–10·66	3·55				W
H. kittanae Vladimirov	Bulgaria: Rodopi (3)	18*	8·41 ± 0·04	8·38–8·46	4·21	4·21 ± 0·01	4·19–4·23	0·95	E
H. lachenalii Suter	Czech Rep.: Křivoklátsko (2)	27*	11·25 ± 0·01	11·27–11·29	3·76	3·75 ± 0·01	3·73–3·76	0·80	n.d.
	Czech Rep.: distr. Znojmo (2)	27*	11·22 ± 0·03	11·19–11·24	3·74				X(W)
	Czech Rep.: distr. Praha-east (2)	27	11·24 ± 0·06	11·20–11·32	3·75				n.d.
H. laevigatum Willd.	Czech Rep.: Brdy (2)	27*	12·05 ± 0·02	12·01–12·08	4·02	4·05 ± 0·02	3·97–4·14	4·28	X
	Czech Rep.: Hradec Králové (2)	27*	12·00 ± 0·08	11·94–12·10	4				n.d.
	Germany: Kamenz (2) (S)	27	12·41 ± 0·01	12·40–12·42	4·14				n.d.
H. lawsonii Vill.	France: Corbières (2)	27	10·76 ± 0·02	10·74–10·78	3·59	3·62 ± 0·03	3·58–3·68	2·79	W
	France: Briançon (1)	36	14·71	–	3·68				n.d.
H. murorum L.	Czech Rep.: Plzeň (2)	27*	10·62 ± 0·06	10·55–10·68	3·54	3·56 ± 0·01	3·52–3·59	1·99	W
	Czech Rep.: Doupovské hory (2)	27*	10·67 ± 0·05	10·63–10·72	3·56				n.d.
	Czech Rep.: Český kras (2)	27*	10·74 ± 0·02	10·72–10·76	3·58				n.d.
H. naegelianum Pančić	Montenegro: Durmitor Mts (2)	27*	10·89 ± 0·08	10·81–10·97	3·63	3·63 ± 0·03	3·60–3·66	1·67	E
H. olympicum Boiss.	Bulgaria: Kaloferska Planina (2)	27*	12·27 ± 0·05	12·22–12·33	4·09	4·05 ± 0·01	4·04–4·06	0·50	X

Continued

TABLE 1. *Continued*

Species	Locality (no. of plants) (S): plant cultivated from seed	2*n*	2C (pg) mean ± s.e.	2C (pg) range	1Cx (pg)	1Cx (pg) species mean ± s.e.	1Cx (pg) species range	1Cx species variation (%)	ETS clade†
H. pannosum Boiss. I	Bulgaria: Trojanska Planina (2)	27*	11·71 ± 0·05	11·66–11·77	3·9				E
H. pannosum Boiss. II	Greece: Peloponnesos (2) (S)	36	16·67 ± 0·01	16·66–16·67	4·17				n.d.
	H. pannosum s.l. mean					4·04 ± 0·08	3·89–4·17	7·20	
H. petrovae Vladimirov & Szeląg	Bulgaria: Rodopi (1)	18*	7·9	–	3·95				E
H. pictum Pers.	France: Montegenèvre (2)	27	10·78 ± 0·27	10·5–11·05	3·59	3·59 ± 0·04	3·50–3·68	5·14	n.d.
	France: Briançon (2)	27	10·75 ± 0·13	10·62–10·88	3·58				W
H. piliferum Hoppe	Austria: Gurktaler Alpen (2)	36	15·57 ± 0·04	15·54–15·60	3·89	3·91 ± 0·02	3·86–4·01	3·89	E-int.
	Austria: Reisseck-Gruppe (1)	27	11·58	–	3·86				
	Italy: Alps, Spluga (1)	27	11·58	–	3·86				
	France: Hautes Alpes (2)	27	11·95 ± 0·07	11·89–12·02	3·99				
H. pilosum Schleich. ex Froel. I.	Slovenia: Julijske Alpe (1)	27	11·57		3·86				E
H. pilosum Schleich. ex Froel. II.	Slovenia: Julijske Alpe (1)	27	11·8		3·93				X
	H. pilosum s.l. mean					3·90 ± 0·04	3·86–3·93	1·81	
H. plumulosum A.Kern.	Montenegro: Mrtvica canyon (1)	18*	8·59	–	4·29				X(E)
H. porrifolium L.	Austria: Karawanken (6)	18*	7·76 ± 0·01	7·7–7·79	3·88	3·89 ± 0·01	3·85–3·93	2·08	E
	Austria: Karawanken (1)	18	7·7	–	3·85				n.d.
	Slovenia: Julijske Alpe (6)	18*	7·82 ± 0·02	7·74–7·85	3·91				n.d.
H. prenanthoides Vill. I.	Poland: Karkonosze (2)	27	10·82 ± 0·03	10·78–10·85	3·61				X(W)
	France: Hautes Alpes, La Grave (1)	18	7·11	–	3·55				X(W)
	France: Hautes Alpes, Briançon (2)	18	7·29 ± 0·01	7·24–7·24	3·64				n.d.
H. prenanthoides Vill. II.	Andorra: Canillo (2)	27*	11·41 ± 0·02	11·40–11·43	3·8				X(W)
	H. prenanthoides s.l. mean					3·67 ± 0·04	3·56–3·81	7·02	
H. racemosum Waldst. & Kit. ex Willd.	Czech Rep.: distr. Znojmo (2)	27	12·41 ± 0·02	12·39–12·44	4·14	4·11 ± 0·06	4·08–4·15	1·72	X
	Czech Rep.: Ústí nad Orlicí (2)	27	12·26 ± 0·01	12·26–12·26	4·09				n.d.
	Slovakia: Gemer (1) (S)	27	12·24	-	4·08				n.d.
H. ramondii Griseb.	Andorra: Encamp (2)	27*	10·63 ± 0·07	10·56–10·7	3·54	3·54 ± 0·03	3·51–3·57	1·71	W
H. recoderi De Retz	Spain: prov. Barcelona (8)	18*	7·09 ± 0·01	7·00–7·09	3·53	3·53 ± 0·01	3·50–3·68	1·43	W
H. sabaudum L.	Czech Rep.: Praha (2)	27	12·51 ± 0·05	12·47–12·56	4·17	4·17 ± 0·02	4·12–4·23	2·67	n.d.
	Germany: Oberlausitz (2)	27	12·65 ± 0·03	12·62–12·68	4·22				X
	Czech Rep.: České středohoří (2)	27	12·36 ± 0·01	12·35–12·37	4·12				n.d.
H. schmidtii Tausch	Czech Rep.: České středohoří (2)	27*	10·64 ± 0·05	10·59–10·69	3·55	3·54 ± 0·01	3·52–3·56	1·14	n.d.
	Czech Rep.: České středohoří (2)	27*	10·58 ± 0·00	10·61–10·61	3·54				W
	Czech Rep.: Křivoklátsko (2)	27	10·60 ± 0·01	10·60–10·61	3·53				n.d.
H. sparsum Friv.	Bulgaria: Vitoša (1)	18	8·15	-	4·08	4·03 ± 0·03	3·99–4·08	2·56	E(H)
	Bulgaria: Pirin (2)	18	8·01 ± 0·03	7·98–8·04	4·01				
H. stelligerum Froel.	France: Ardèche (3)	18*	7·03 ± 0·06	6·91–7·14	3·51	3·51 ± 0·03	3·47–3·57	2·89	W
H. tomentosum L.	France: Alpes Maritimes (8)	18*	7·48 ± 0·02	7·41–7·58	3·74	3·75 ± 0·01	3·71–3·79	2·16	W
	France: Briançon (1)	27	11·27	–	3·76				n.d.
	France: Hautes Alpes (2)	27	11·25 ± 0·08	11·17–11·33	3·78				n.d.
H. transylvanicum Heuff.	Ukraine: Marmaros'ki Al'py (8)	18*	8·56 ± 0·01	8·52–8·59	4·28	4·28 ± 0·00	4·26–4·30	0·94	W
H. umbellatum L.	Poland: Baltic coast, Jantar (1)	18	8·34	–	4·17	4·26 ± 0·01	4·17–4·30	3·12	E
	Czech Rep.: Praha (8)	18*	8·54 ± 0·01	8·48–8·59	4·27				n.d.
	Germany: Nordfriesland (2) (S)	18	8·48 ± 0·03	8·45–8·52	4·24				n.d.
H. villosum Jacq. I.	Slovakia: Strážovské vrchy (2)	36	15·71 ± 0·04	15·66–15·75	3·93				E
H. villosum Jacq. II.	France: Hautes Alpes (1)	27	11·6	–	3·87				X(E)
	H. villosum s.l. mean					3·91 ± 0·02	3·87–3·94	1·81	
H. virosum Pall.	Russia: Rostov-na-Donu (2) (S)	27	13·06 ± 0·03	13·02–13·09	4·35	4·34 ± 0·00	4·33–4·36	0·69	E
	Russia: Altajskij kraj (3) (S)	27	13·00 ± 0·01	12·98–13·03	4·34				E
'Hieracium' intybaceum All.	Italy: Rhaetian Alps (1)	18	7·5	–	3·75	3·76 ± 0·01	3·74–3·79	1·34	

	Italy: Alpi Orobie (2)	18	7·53 ± 0·07	7·48–7·58	3·77
	Austria: Ötztaler Alpen (2)	18	7·53 ± 0·07	7·52–7·54	3·77
Pilosella lactucella (Wallr.) P.D.Sell & C.West	Germany: Erzgebirge (1)	18	4·07*		2·04
Andryala integrifolia L.	Spain: Andalusia (1)	n.d.	n.d.		
Andryala levitomentosa (Nyár.) P.D.Sell	Romania: Pietrosul Bogolini (1)	18	5·31§		2·66
Hispidella hispanica Barnades ex Lam.	Spain: Sierra de Guadarrama (1)	18#	~4·00		~2·00

* From Chrtek *et al.* (2007).

† E, 'eastern' clade; W: 'western' clade; X, interclade hybrid; X(E) and X(W), interclade hybrids with predominant 'eastern' or 'western' sequence variant; E(H) and W(H), hybrids within 'eastern' or 'western' clade; hybrid origin of *H. sparsum* and *H. lachenalii* inferred from plastid DNA (J. Fehrer *et al.*, unpubl. res.). E-int. indicates ETS character additivity between the 'eastern' clade and '*Hieracium*' *intybaceum* indicative of hybrid origin.

‡ See also Suda *et al.* (2007).

§ With *Pilosella lactucella* standard.

Luque (1981); Elena-Rosselló *et al.* (1985).

2002) to eight in supposedly genetically variable populations of sexual diploids. Two species, *H. petrovae* and *H. plumulosum*, were represented by a single plant due to their rarity in the field or because of cultivation problems. When two or more ploidies have been reported for a species, this diversity was covered as far as possible. Voucher specimens of all samples are deposited in the herbarium PRA.

Chromosome numbers and breeding system

At least two plants per population were checked for their chromosome number using the method described in Chrtek *et al.* (2007); counts for selected accessions have been published (Chrtek *et al.*, 2007). The mode of reproduction was also tested, generally following Gadella (1987) and Krahulcová and Krahulec (1999). In diploids (in which sexual reproduction was expected), randomly selected capitula were bagged at the bud stage and tested for late-acting autogamy (in the absence of active pollination); results were compared with control capitula from the same plant in open pollination treatments. In polyploids (in which agamospermy was expected), the upper part of the capitulum was cut off at the bud stage (emasculation) and the number of 'full' achenes was counted as a measure of seed set and compared with the number of achenes from untreated capitula of the same plant. Percentages of 'full' achenes after emasculation in particular plants/species are available upon request.

Genome size estimation

Genome size was determined by flow cytometry using a Partec CyFlow cytometer equipped with a green (532 nm) solid-state laser. *Zea mays* 'CE–777' (2C = 5·48 pg; Lysák and Doležel, 1998) and *Pisum sativum* 'Ctirad' (2C = 8·85 pg; Doležel *et al.*, 1994; Suda *et al.*, 2007) were used as internal standards for diploid and polyploid species. The modified two step-procedure described by Otto (1990) was employed for sample preparation. Intact leaf tissue (approx. 1 cm^2) of the analysed species and an appropriate quantity of the internal standard were co-chopped with a sharp razor blade in a plastic Petri dish with 1 mL of ice-cold Otto I buffer (0·1 M citric acid, 0·5% Tween 20) as the nuclear isolating solution. The suspension was filtered through a 42-μm nylon filter and centrifuged at 15 *g* for 5 min. The supernatant was discarded, and the pellet was resuspended in 100 μL fresh Otto I buffer. Samples were incubated for at least 10 min at room temperature and mixed with 1 mL Otto II buffer (0·4 M Na_2HPO_4) supplemented with propidium iodide as the fluorochrome, RNase IIA (both at a concentration of 50 μg mL^{-1}) and β-mercaptoethanol (2 μL mL^{-1}). Samples were stained for 5 min at room temperature before measurement. Usually, 5000 nuclei were analysed for each sample. Nuclear genome size was calculated as a linear relationship between the ratio of 2C peaks of sample and standard. Each plant was measured at least three times on different days by the same operator to eliminate potential artefacts. If the difference between the three measurements exceeded 2%, the value was discarded, and the sample was re-analysed. The coefficients of variation (CVs) of G_0/G_1 peaks did not exceed 5% (with two exceptions in *Hieracium* samples).

Molecular methods

A representative subset of 49 *Hieracium* accessions was selected for phylogenetic analysis. As the outgroup, species of the most closely related genera *Andryala*, *Hispidella* and *Pilosella* (sometimes treated as a subgenus of *Hieracium*) and '*Hieracium*' *intybaceum* were chosen according to previous results (Fehrer *et al.*, 2007). Although the latter species was traditionally placed in *Hieracium* subgenus *Hieracium*, molecular data (ITS sequences) suggested it belongs to an older isolated lineage clearly separated from a cluster formed by *Hieracium s.l.* and its closely related genera *Andryala* and *Hispidella* (Fehrer *et al.*, 2007), which is also supported by the present data based on the ETS region.

Total genomic DNA was extracted from fresh or CTAB-preserved material using a sorbitol extraction method (Štorchová *et al.*, 2000). The ETS region of the nuclear ribosomal DNA was amplified using the primers Ast-8 and 18 S (Noyes, 2006). PCR amplifications were done in 25-μL reactions containing 2·5 mM $MgCl_2$, 0·2 mM of each dNTP, 0·5 μM 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 35 cycles of denaturation (94°C for 30 s), annealing (55°C for 30 s) and extension (72°C for 40 s) steps, and a final extension at 72°C for 10 min. The PCR products were purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and sequenced (GATC Biotech, Konstanz, Germany). Both strands were sequenced using the PCR primers. For one accession, *Pilosella lactucella*, direct sequencing was not successful and, therefore, the amplified fragment was subcloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions, but downscaled to half reactions. Three clones of this sample were sequenced (GATC Biotech), using the primer Ast-8.

Molecular data analyses

Sequence chromatograms were inspected by eye. In many accessions intra-individual polymorphism, i.e. more than one allele of the ETS region, was detected. Polymorphic sites were represented by the NC-IUB ambiguity symbols (e.g. R for A or G).

Initial sequence alignment was done with Clustal X (Thompson *et al.*, 1997) and further edited manually in Bioedit (Hall, 1999). It was unambiguous due to low overall variation. Visual inspection of the alignment revealed the existence of two major groups within *Hieracium s.s.*, and a proportion of accessions were identified as hybrids among these groups according to the additive pattern of polymorphic sites (these are referred to as 'interclade hybrids'). Furthermore, the *Hieracium piliferum* accession analysed was identified as a hybrid between one of the major groups and '*Hieracium*' *intybaceum*. Sequences were submitted to GenBank (EU821362–EU821419).

Bayesian (MrBayes V3·1·2; Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) and maximum likelihood (RAxML V7·0·3; Stamatakis, 2006) analyses were performed (*a*) on the complete data set and (*b*) on a modified data set excluding interclade hybrids and *H. piliferum*. Bayesian analyses were run with two nucleotide substitution rates and gamma distribution. This corresponds to a HKY + G model found in hierarchical Likelihood Ratio Tests as the model of molecular evolution best fitting to the data as implemented in Modeltest V3·5 (Posada and Crandall, 1998). Two parallel runs with four chains each were used for both analyses, sampling every 1000th tree. The analysis of the complete data set was computed for 2 million generations until convergence. The first 500 trees per run were discarded as burn-in and the remaining 1501 trees per run were summarized. For the modified data set, 1 million generations were sufficient for reaching convergence, the first 250 trees per run were discarded as burn-in and the remaining 751 trees per run were summarized. Maximum likelihood analyses were done using the rapid BS algorithm in combination with a maximum likelihood search, using the GTR model of nucleotide substitutions with a gamma model of rate heterogeneity and 1000 bootstrap replicates.

Statistical analyses of genome size

Intraspecific variation in morphologically defined species *sensu* Zahn was assessed and species with variation exceeding the approximate inaccuracy threshold of 3·5% [following Suda *et al.* (2007) in *Hieracium* subgenus *Pilosella*] were marked (*H. amplexicaule*, *H. bupleuroides*, *H. laevigatum*, *H. pannosum*, *H. pictum*, *H. piliferum* and *H. prenanthoides*). In *H. bupleuroides* and *H. prenanthoides*, differences in ETS sequences among accessions corresponding with the differences in genome size were found (J. Fehrer *et al.*, unpubl. res.) which were caused by hybridization with other species. These heterogeneous species were therefore split into more natural units (accessions or groups of accessions) according to genome size and treated separately in the following analyses. The same was done for *H. pannosum*, in which the accessions differed distinctly in genome size (and in ploidy), and for *H. pilosum* and *H. villosum*, in which differences in ETS sequences due to introgression were found between accessions (although genome size variation was rather low in these cases). On the other hand, morphologically homogeneous species with high intrapopulation variation in genome size (*H. pictum*) and species with unclear patterns of genome size variation (*H. amplexicaule*, *H. laevigatum* and *H. piliferum*) were not split. The units after splitting held the name of the broad species and were numbered (I, II) (Table 1). For convenience, they are referred to as species in the following paragraphs. Forty six taxa were recognized after the split (Table 1). To test the correlation between monoploid DNA amount (1Cx) and chromosome number, the Spearman rank order correlation coefficient and the one-way ANOVA were used with a matrix of all samples ($2n = 18$, 27 and 36).

Evolution of genome size was investigated on a sample of trees using the generalized least-squares method implemented in BayesTraits (Pagel and Meade, 2007). For this purpose, the last 751 trees from each run of the Bayesian analysis of the modified data set (without hybrids) were sampled and merged into one file. All 1502 trees were rooted manually with the outgroup (*Andryala integrifolia*, the taxon used as

outgroup in the Bayesian analysis), using the program Dendroscope (Huson *et al.*, 2007). Analyses were conducted on three different data sets: (1) on a complete set of these species; (2) separately for species of the western clade; and (3) separately for species of the eastern clade. All genome size data were $\log_{10}$ transformed prior to analyses. Two models of trait evolution were compared, using likelihood ratio statistics (Huelsenbeck and Rannala, 1997) or BayesFactor. Model A (drift model) corresponds to the standard constant-variance random-walk model, and model B is a directional random-walk model (Pagel, 1999, 2004). The scaling parameters lambda (λ), kappa (κ) and delta (δ) were optimized for 1Cx-values. λ assesses the contribution of the phylogeny to a character, κ scales branch lengths and can be used to test punctual vs. gradual modes of trait evolution, and δ scales the overall path length in the phylogeny. Values of 1·0 correspond to the null model (tree topology and branch lengths accurately describe models A and B). To test whether a model with estimated scaling parameters is a better fit than the null model where all scaling parameters are set to 1 (i.e. a strict Brownian motion model) the likelihood ratio test or BayesFactor was used. Two different methods of analysis were used, namely maximum likelihood and Monte Carlo Markov chain (MCMC).

How the variation in genome size matched the two major lineages of *Hieracium* (named 'western' and 'eastern') suggested by molecular phylogenetic analyses of the ETS region was tested further. In addition, genome sizes of interclade hybrid accessions were analysed. Five informal groups were recognized and used for these analyses, restricting the genome size data set to the accessions for which sequence data were available (Table 1): (*a*) 'western', corresponding to the phylogenetically distinguished 'western' group and containing 'pure' accessions as well as hybrid/hybridogenous accessions within the 'western' group [W and W(H)]; (*b*) 'eastern', corresponding to the phylogenetically distinguished 'eastern' group and containing 'pure' accessions and hybrid/hybridogenous accessions within the 'eastern' group [E and E(H)]; (*c*) hybrid/hybridogenous accessions between 'western' and 'eastern' clade species with about equal contribution of ETS variants from each parent (X); (*d*) interclade hybrid accessions with strongly dominating 'western' ETS sequence type [X(W)]; and (*e*) interclade hybrid accessions with strongly dominating 'eastern' group ETS [X(E)] (Table 1). For the purpose of this analysis, no distinction was made between 'pure' and intraclade hybrids (*a*, *b*) because of the low sequence variation within each clade and because their genome sizes did not differ. Two comparisons were performed: (1) the 'western' and 'eastern' groups with a group comprising all E–W hybrids; and (2) the 'western' and 'eastern' groups with the three different groups of hybrids (*c*–*e*) specified above. More details about ETS sequence features and the identification of particular hybrids will be presented elsewhere (J. Fehrer *et al.*, unpubl. res.). Both comparisons were conducted separately with and without *H. transylvanicum* (which fell into the phylogenetically defined western lineage, but has a genome size and geographic range congruent with the 'eastern' group; see Discussion). The correlation between genome size and phylogenetic pattern (five groups, see above) was tested.

The Spearman non-parametric rank order correlation coefficient was used in testing whether DNA amounts correlated with selected Ellenberg's indicator values, namely for light and temperature (Ellenberg *et al.*, 1992). Mean 1Cx values for species *sensu* Zahn were used for this analysis; only a subset of central European species (for which these values are available) was chosen. Genome size variation was also tested against altitudinal and geographical position (longitude and latitude) for (*a*) the complete set of accessions (mean accessions' 1Cx values were used), and (*b*) excluding accessions of widely distributed species (*H. bifidum*, *H. lachenalii*, *H. laevigatum*, *H. murorum*, *H. sabaudum* and *H. umbellatum*) for which the results are strongly affected by the collection site of the samples.

The only significant correlations of genome size variation with other parameters concerned phylogeny and longitude. In order to distinguish between these two factors, three correlation tests concerning longitude were performed for each pairwise comparison, constrained to accessions for which sequence data were available: (1) across all accessions from the 'western' and 'eastern' clades; (2) within the 'western' clade only; and (3) within the 'eastern' clade only. If the correlation is significant across all species, but not significant within either of the two clades analysed separately, this would argue for a connection between phylogenetic relationships and genome size variation. If, however, significant correlations are found for each of these tests, a relationship of genome size to longitude independent of species relationships would be supported. Data were analysed using the statistical package 'Statistica for Windows 6·0' (StatSoft, 1984–2002).

RESULTS

Chromosome counts and mode of reproduction

Chromosome numbers for plants of 43 populations analysed in the present paper belonging to 28 species were published elsewhere (Chrtek *et al.*, 2007), and counts for the remaining 46 populations are presented here (Table 1). A new ploidy is reported for *H. gymnocephalum* ($2n = 18$). Other counts confirmed previously published chromosome numbers. All diploids studied were found to be sexual and allogamous, and all polyploids (3*x*, 4*x*) were agamospermous (data not provided).

Flow cytometry

Flow cytometric analyses yielded high-resolution histograms with CVs of G_0/G_1 peaks for *Hieracium* samples ranging from 0·83 to 5·76% (mean 2·28%), the values for internal reference standards were 0·97 to 5·0% (mean 2·19%). Generally, CVs of *Pisum sativum* were lower than those of *Zea mays*.

Nuclear DNA content: within-species variation

Intraspecific variation was assessed in 40 of 42 species *sensu* Zahn. Variation within accessions (populations) was generally low (Table 1). Mean values with standard errors (ranges for 2C and means for 1Cx genome sizes) for each population and

mean values with standard errors and ranges for 1Cx for each species are summarized in Table 1. Variation in 1Cx values between conspecific accessions (in both homo- and multiploid species) ranged from 0·24% in *H. gymnocephalum* and *H. heterogynum* to 7·2% in *H. pannosum*. Variation exceeding the approximate measurement inaccuracy threshold of 3·5% was detected in seven species, namely *H. amplexicaule*, *H. bupleuroides*, *H. laevigatum*, *H. pannosum*, *H. pictum*, *H. piliferum* and *H. prenanthoides*.

However, variation in the more naturally delimited (without heterogeneity in ETS sequences and inter-population genome size; see Materials and methods) *H. bupleuroides* was below the threshold of 3·5% variation. Variation within the separately treated populations of morphologically heterogeneous *H. pannosum* was also below 3·5% (Table 1). In further paragraphs, these 'narrower' taxa (a total of 46 taxa) are used.

C-values in the total set of 'basic' species

The mean 2C values differed up to 2·37-fold among different species (from 7·03 pg in diploid *H. stelligerum* to 16·67 in a tetraploid accession of *H. pannosum*). The 1Cx values varied 1·22-fold between 3·51 pg in *H. stelligerum* and 4·34 pg in *H. virosum* (mean 1Cx value of 3·87, s.d. 0·27; Fig. 1). The 1Cx values of diploids (including diploid accessions of multiploid species, means for species/cytodemes) varied 1·22-fold between 3·51 pg in *H. stelligerum* and 4·29 pg in *H. plumulosum* (mean 1Cx value 3·92 pg, s.d. 0·30), in triploids [including triploid accessions of multiploid species (means for species/cytodemes)] 1·23-fold between 3·53 pg in *H. bifidum* and 4·35 pg in *H. virosum* (mean 1Cx value of 3·81 pg, s.d. 0·25), and in tetraploids 1·17-fold between 3·56 pg in *H. humile* and 4·17 pg in *H. pannosum* II (mean 1Cx value of 3·79, s.d. 0·19).

Correlation between genome size, ploidy and breeding system

Diploids differed significantly in their 1Cx values from both triploids ($t = 2{\cdot}71$, d.f. = 196, $P = 0{\cdot}007$) and tetraploids ($t = 2{\cdot}01$, d.f. = 109, $P = 0{\cdot}047$), but triploids did not differ significantly from tetraploids ($t = 0{\cdot}72$, d.f. = 119, $P = 0{\cdot}476$) (Fig. 2). The value of the Spearman non-parametric rank order correlation coefficient was $r = -0{\cdot}179$, $P = 0{\cdot}009$. The mean 1Cx value was 3·93 pg in diploids, 3·82 pg in triploids and 3·78 pg in tetraploids, suggesting a trend towards smaller genome size with increasing ploidy.

In multiploid species, there was no general trend to either genome downsizing or upsizing. In *H. prenanthoides* 2*x*/3*x* there was 2·47% upsizing, in *H. villosum* 3*x*/4*x* 0·06% downsizing, in *H. tomentosum* 2*x*/3*x* 0·36% upsizing, in *H. humile* 3*x*/4*x* 0·36% upsizing and in *H. alpinum* 2*x*/3*x* 0·17% upsizing.

Comparison between 1Cx values of sexually reproducing plants (i.e. all diploids; polyploids were exclusively apomictic) and apomicts (triploids and tetraploids) revealed significant differences at $\alpha = 0{\cdot}01$ (*t*-test, $t = 3{\cdot}04$, d.f. = 213, $P = 0{\cdot}003$); the mean 1Cx value in sexuals was 3·93 pg, whereas in apomicts it was 3·82 pg, corresponding to the value for triploids due to the low number of tetraploid accessions.

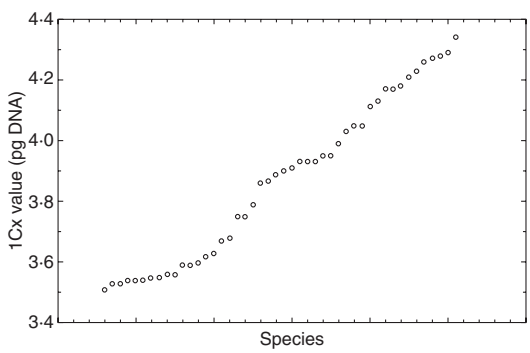


FIG. 1. 1Cx-value variation in 46 taxa of *Hieracium* subgenus *Hieracium*.

Molecular phylogenetics of Hieracium

Analysis of ETS data including all sequenced accessions resulted in the same tree with both methods. It indicates monophyly of *Hieracium* subgenus *Hieracium*, but species relationships remained completely unresolved as reflected by a large polytomy with only two small subclusters that received low support (Fig. 3). However, as two major species groups could be identified by visual inspection of the alignment and many sequences showed additive patterns indicative of hybridization involving both groups, these accessions were deleted from subsequent analysis, because reticulation is known to collapse branches (Feliner *et al.*, 2001; Soltis *et al.*, 2008). With the reduced data set, a clear separation into two major clades with strong statistical support was found with both methods (Fig. 4). These lineages were designated 'eastern' and 'western' clade because they contained species of predominantly eastern or western European origin. A large number of accessions (18) showed ETS variants of both clades in either equal proportion or with the 'eastern' or 'western' sequence type dominating as indicated in Fig. 4. Details of these analyses will be given in a parallel paper (J. Fehrer *et al.*, unpubl. res.).

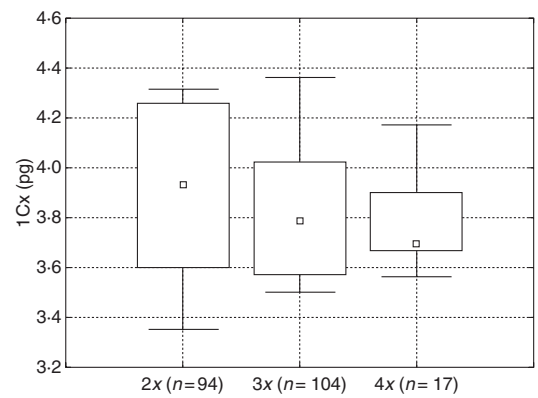


FIG. 2. Variation of genome size among diploids, triploids and tetraploids (all samples). 1Cx values of all accessions are shown. Differences between diploids and triploids, and between diploids and tetraploids are significant. The box indicates the interquartile (25–75%) range, the small square within the box is the median. The whiskers indicate minimum and maximum values.

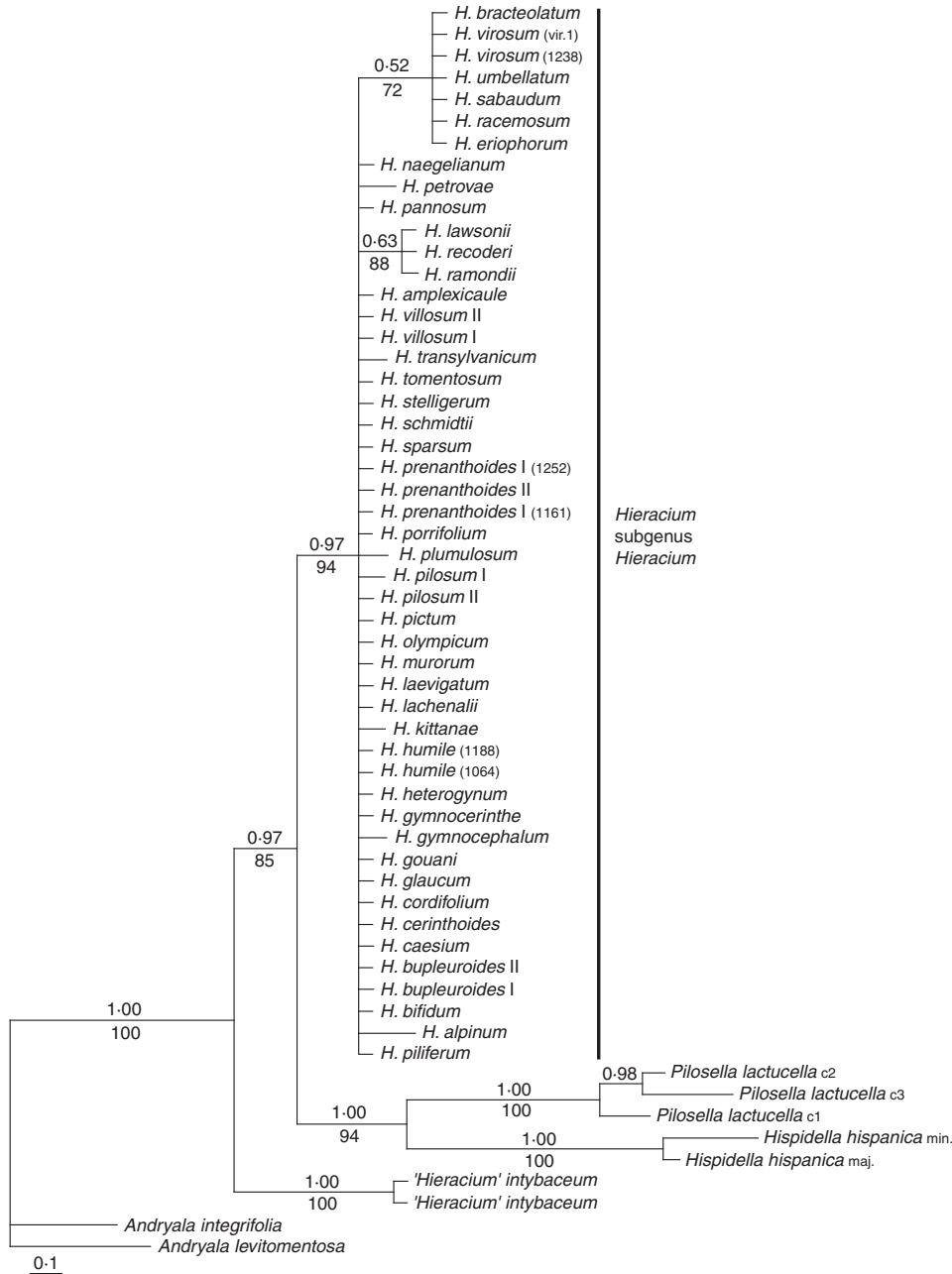


FIG. 3. Phylogenetic analysis of ETS sequences based on all accessions. A Bayesian consensus tree of 3002 saved trees is shown with posterior probabilities above branches. The maximum likelihood tree has the same topology; bootstrap values are indicated below branches. *Hieracium* subgenus *Hieracium* (=*Hieracium sensu stricto*) is monophyletic, but species relationships are completely unresolved when hybrids are included in the analysis. Support for the two subclusters is low.

Correlation of genome size with phylogenetic signal

The 'western' clade included 15 accessions: 2C values ranged from 7·03 pg in diploid *H. stelligerum* to 14·25 pg in a tetraploid accession of *H. humile*; 1Cx values ranged from 3·51 pg in *H. stelligerum* to 4·28 pg in *H. transylvanicum* (mean ± s.d.: 3·61 ± 0·19 pg; with *H. transylvanicum*

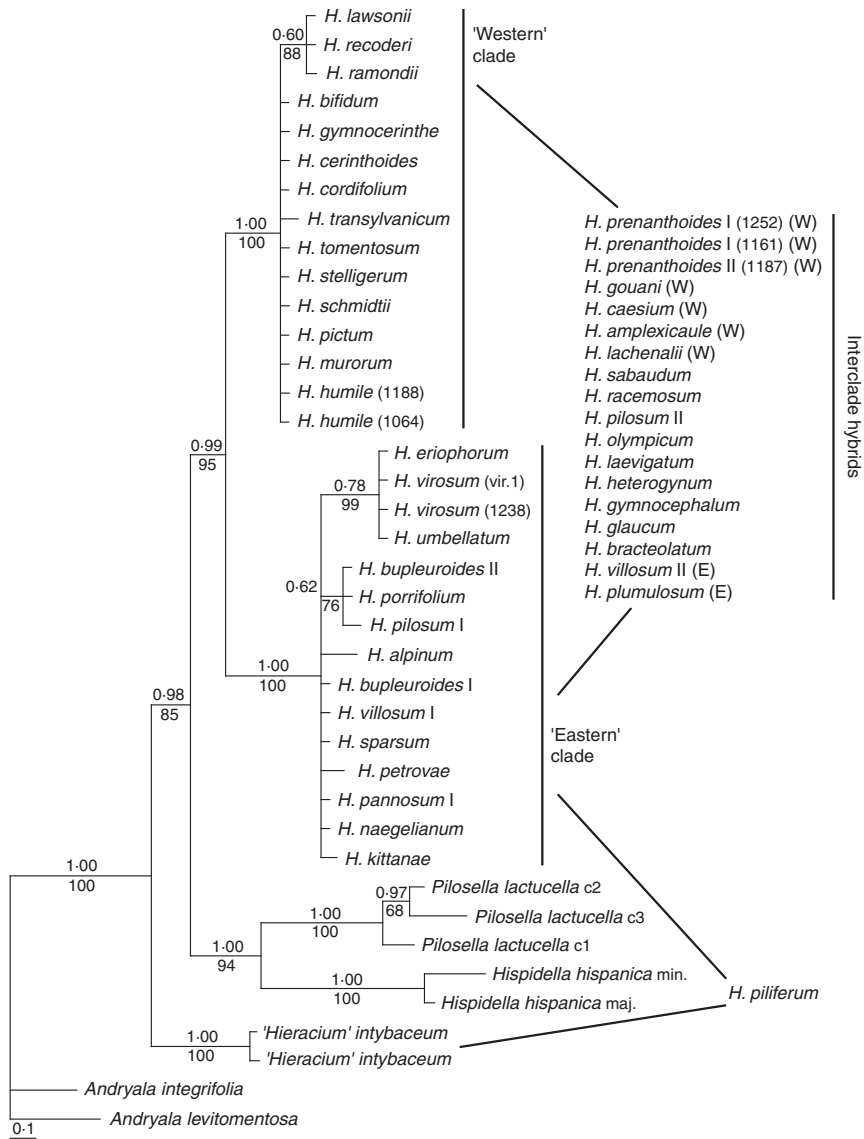


FIG. 4. Phylogenetic analysis of ETS sequences excluding interclade hybrid accessions. A Bayesian consensus tree of 1502 saved trees is shown with posterior probabilities above branches. The maximum likelihood tree has the same topology; bootstrap values are indicated below branches. After exclusion of hybrids based on character additivity, two major groups are resolved (referred to as 'eastern' and 'western' clades). Hybrid accessions composed of parents from both clades (interclade hybrids) are listed to the right. (W), Interclade hybrids with predominant 'western' ETS type; (E), interclade hybrids with predominant 'eastern' ETS type. *Hieracium piliferum* is a hybrid between '*Hieracium*' *intybaceum* and an 'eastern' clade taxon. For details about accessions, see Table 1 and Supplementary Data, available online.

excluded: up to 3·74 pg in *H. tomentosum*, 3·57 ± 0·06 pg). The 'eastern' clade also comprised 15 accessions: 2C values ranged from 7·78 pg in diploid *H. porrifolium* to 15·71 pg in a tetraploid accession of *H. villosum*; 1Cx values ranged from 3·63 pg in *H. naegelianum* to 4·35 pg in *H. virosum* (4·02 ± 0·20 pg). Significant differences in 1Cx values were found between the clades at $\alpha = 0{\cdot}001$ (Student's *t*-test), with ($t = -5{\cdot}71$, d.f. = 28, $P < 0{\cdot}001$) and without ($t = -8{\cdot}23$, d.f. = 27, $P < 0{\cdot}001$) *H. transylvanicum*.

Differences in 1Cx values between accessions of the 'western' (W) and 'eastern' (E) clades and of interclade hybrid accessions (X) are significant, independent of the inclusion of *H. transylvanicum* ($F = 13{\cdot}79$, d.f. = 45, $P < 0{\cdot}001$ with *H. transylvanicum*, $F = 20{\cdot}87$, d.f. = 44,

$P < 0{\cdot}001$ without *H. transylvanicum*; Fig. 5A). However, *post hoc* comparison (Scheffé test) revealed only two groups at $\alpha = 0{\cdot}05$, the first comprising all 'western' accessions, and the second embracing 'eastern' and 'hybrid' accessions. Thus, 'eastern' and 'hybrid' accessions do not differ significantly from each other. Significant differences were also found between five groups, i.e. after splitting the bulk of hybrids into three groups, namely hybrids with intermediate position (X) and hybrids with strongly dominating 'western' [X(W)] or 'eastern' [X(E)] ETS sequences ($F = 17{\cdot}07$, d.f. $= 43$, $P < 0{\cdot}001$ with *H. transylvanicum*, $F = 28{\cdot}86$, d.f. $= 42$, $P < 0{\cdot}001$ without *H. transylvanicum*). The Scheffé test revealed only two groups at $\alpha = 0{\cdot}05$, the first including W and X(W) accessions, the second X, X(E) and E accessions (Fig. 5B). A significant correlation (Spearman rank coefficient $r = 0{\cdot}705$, $P < 0{\cdot}001$) between phylogenetic signal and hybrid origin [all five groups – W, E, X, X(W) and X(E)] and the pattern of genome size variation was found. *Hieracium piliferum* (1Cx = 3·9) occupies an isolated position, and it was identified as a hybrid between an 'eastern' clade taxon and '*Hieracium*' *intybaceum* (1Cx = 3·76).

Evolution of genome size

Maximum likelihood method. For the complete data set (all species), a directional model of evolution (model B) did not result in significantly higher likelihood scores than the drift model of evolution (model A; 74·856 vs. 75·746), indicating that there is no general trend to either genome size increase or decrease. Scaling parameters leading to the highest likelihood for 1Cx values were $\lambda = 0{\cdot}908$, $\delta = 0{\cdot}819$ and $\kappa = 1{\cdot}035$ in model A and $\lambda = 0{\cdot}701$, $\delta = 0{\cdot}637$ and $\kappa = 1{\cdot}201$ in model B, respectively. For both models, the values of scaling parameters did not differ significantly from 1 (the null expectation, LR test), indicating that the phylogenetic tree correctly predicts the pattern of covariance among species on the trait (1Cx) and that there is no evidence of accelerated evolution.

For the western clade, likelihood scores of models A and B did not differ significantly (44·061 vs. 44·790). The maximum likelihood values for λ (<1; 0·550 in model A, 0·523 in model B) show a role of adaptive response to some external factors. Values of δ and κ are >1 in both models (not shown) indicating that longer paths contribute more to 1Cx evolution (accelerated evolution as time progresses) and that longer branches contribute more to the trait evolution. Likelihoods of the null model (with scaling paramaters set to 1·0) are significantly lower in both models, indicating that scaling parameters improve the fit of the data to the models.

For the eastern clade, likelihood scores of models A and B also did not differ significantly (40·915 vs. 42·141). Scaling parameters are not significantly different from 1 (the null expectation, Brownian motion, data not shown) indicating that the phylogenetic tree correctly predicts the pattern of covariance among species and that there is no evidence of accelerated evolution.

MCMC method. For the complete data set, comparison of harmonic means of log maximum likelihoods of models A and B showed a somewhat higher value in the latter (82·544 vs. 84·670). The model with estimated scaling parameters is a better fit than the null model (with scaling parameters set to 1) for both models A and B, showing that the scaling parameters improve the fit of the data to the model. The values of λ did not differ significantly from 1, and relatively high values of κ (3·379 and 3·606, respectively) indicate that longer branches contribute more to genome size evolution.

For the western clade only, the harmonic mean of model B is also higher than that of model A (49·888 vs. 45·494) and the harmonic mean of the null model is significantly lower than that for the model with estimated scaling parameters. Scaling parameters are similar to those found with the maximum likelihood method, i.e. λ and $\delta < 1$, $\kappa > 1$ (for interpretation see above).

For the eastern clade, harmonic means of models A and B do not differ significantly. Values of λ and δ do not differ from 1, higher values of κ (1·988 in model A and 2·212 in model B) again indicate accelerated rates of evolution within long branches.

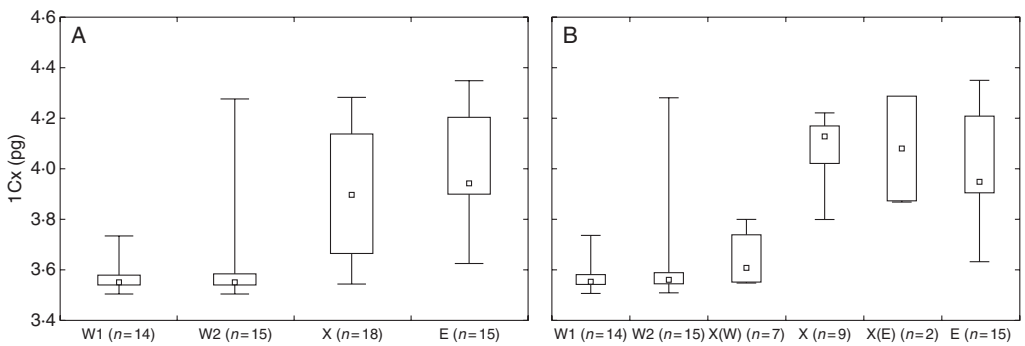


FIG. 5. Correlation of 1Cx values with phylogeny. Only accessions for which sequence data were available are included. (A) W1, 'western' clade accessions without *H. transylvanicum*; W2, 'western' clade accessions including *H. transylvanicum*; X, interclade hybrid accessions; E, accessions of the 'eastern' clade. (B) W1, W2 and E as before, hybrids divided into those with predominant 'western' [X(W)], equal (X) and predominant 'eastern' [X(E)] ETS sequence composition (see also Table 1). The box indicates the interquartile (25–75%) range, the small square within the box is the median. The whiskers indicate minimum and maximum values.

Correlation between genome size and ecogeographic features

The genome size of particular accessions was significantly correlated with their geographic position (longitude) in a west–east direction, both in the complete set of accessions (Spearman rank coefficient $r = 0{\cdot}562$, $P < 0{\cdot}001$; Fig. 6A) and after exclusion of widely distributed species ($r = 0{\cdot}617$, $P < 0{\cdot}001$; i.e. without *H. bifidum*, *H. lachenalii*, *H. laevigatum*, *H. murorum*, *H. sabaudum* and *H. umbellatum*; Fig. 6B). The correlation was stronger in the second case due to the strong dependence on the part of the geographic area from which the target plants of widespread species were sampled.

No correlation between genome size and latitude ($r = 0{\cdot}049$, $P = 0{\cdot}646$) or genome size and altitude ($r = -0{\cdot}224$, $P = 0{\cdot}034$) was found (complete set of accessions, results not shown). Also, no significant correlation was found between genome size and selected ecological parameters (Ellenberg's indicator values), namely temperature ($r = 0{\cdot}194$, $P = 0{\cdot}427$) and light ($r = -0{\cdot}236$, $P = 0{\cdot}331$) in a subset of species occurring in central Europe (results not shown).

Distinction between longitudinal and phylogenetic correlation

In order to determine whether the increase in genome size towards the east/'eastern' clade is based on geographic distribution or on species relationships, the longitudinal correlation was re-analysed for those accessions for which molecular data were available. Even stronger correlation was found between longitude and genome size in a set of accessions with known ETS sequences independent of including ($r = 0{\cdot}656$, $P < 0{\cdot}001$) or excluding ($r = 0{\cdot}688$, $P < 0{\cdot}001$) accessions of widely distributed species (Fig. 7A, B). In contrast, no significant correlation was found either among species of the 'western' ($r = 0{\cdot}161$, $P = 0{\cdot}567$) or among species of the 'eastern' ($r = 0{\cdot}394$, $P = 0{\cdot}245$) clade when tested separately (Fig. 7C and D). These results reveal that the evolutionary history due to eastern or western origin of the species is the dominant parameter affecting genome size in *Hieracium* rather than longitude.

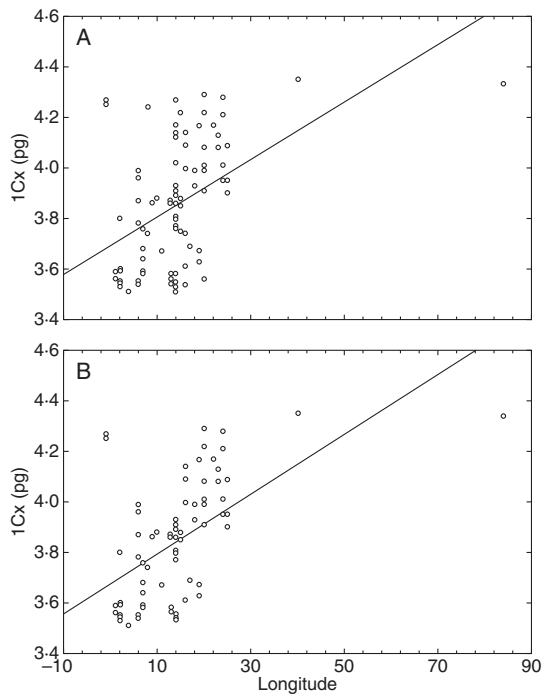


FIG. 6. Distribution of 1Cx values versus longitudinal position of collection sites: (A) based on the complete set of accessions/populations (Spearman rank coefficient $r = 0{\cdot}562$, $P < 0{\cdot}001$); (B) based on a subset after excluding accessions of widely distributed species (*H. bifidum*, *H. lachenalii*, *H. laevigatum*, *H. murorum*, *H. sabaudum* and *H. umbellatum*; $r = 0{\cdot}617$, $P < 0{\cdot}001$).

DISCUSSION

Chromosome numbers and mode of reproduction

Chromosome numbers for plants from 46 populations belonging to 26 species are published here for the first time, and counts for the remaining accessions have been published elsewhere (Chrtek *et al.*, 2007). Among the new data, a new ploidy (diploid) is reported for *H. gymnocephalum* (*s.l.*); previously reported counts (Niketić *et al.*, 2006) refer only to triploids ($2n = 3x = 27$). After *H. petrovae*, this is the second diploid count within section *Pannosa*. Also worth mentioning is the diploid ($2n = 2x = 18$) count for *H. prenanthoides* from the western Alps. Although this number had been reported from the same area in the 1960s (Favarger, 1969; Löve, 1969), it has not been confirmed until now. The remaining chromosome numbers correspond to previously published counts for the target species [cf. Schuhwerk (1996) and other standard reference manuals]. Analysis of the mode of reproduction confirmed the pattern already published for selected species and suggested it to be generally valid throughout the genus: diploid species reproduce sexually and are allogamous whereas polyploids are agamospermous.

Intraspecific genome size variation

Variation beyond arbitrary fluctuation (3·5%) was found in seven species *sensu* Zahn, namely *H. piliferum* (3·89%), *H. amplexicaule* (3·95%), *H. bupleuroides* (4·2%), *H. laevigatum* (4·28%), *H. pictum* (5·14%), *H. prenanthoides* (7·02%) and *H. pannosum* (7·2%). All are agamospermous polyploids, two of them including two cytodemes (*H. amplexicaule*, $3x$ /$4x$; *H. prenanthoides*, $2x$ /$3x$). As only a subset of populations was used for phylogenetic analysis, these results need to be interpreted with caution. In several cases in which sequence data were obtained for more than one accession, multiple origins of a given taxon, namely in *H. bupleuroides*, *H. pilosum*, *H. prenanthoides* and *H. villosum*, were found, which allows a plausible explanation of genome size variation at least for *H. prenanthoide*s. In this species, sequences of three accessions, one being diploid and two triploid, were analysed. All three showed identical signatures of an ancient interclade hybridization already affecting the diploid (J. Fehrer *et al.*, unpubl. res.). The triploids resulted from different subsequent hybridizations, one involving a 'western' and one an 'eastern' lineage. This fits well with

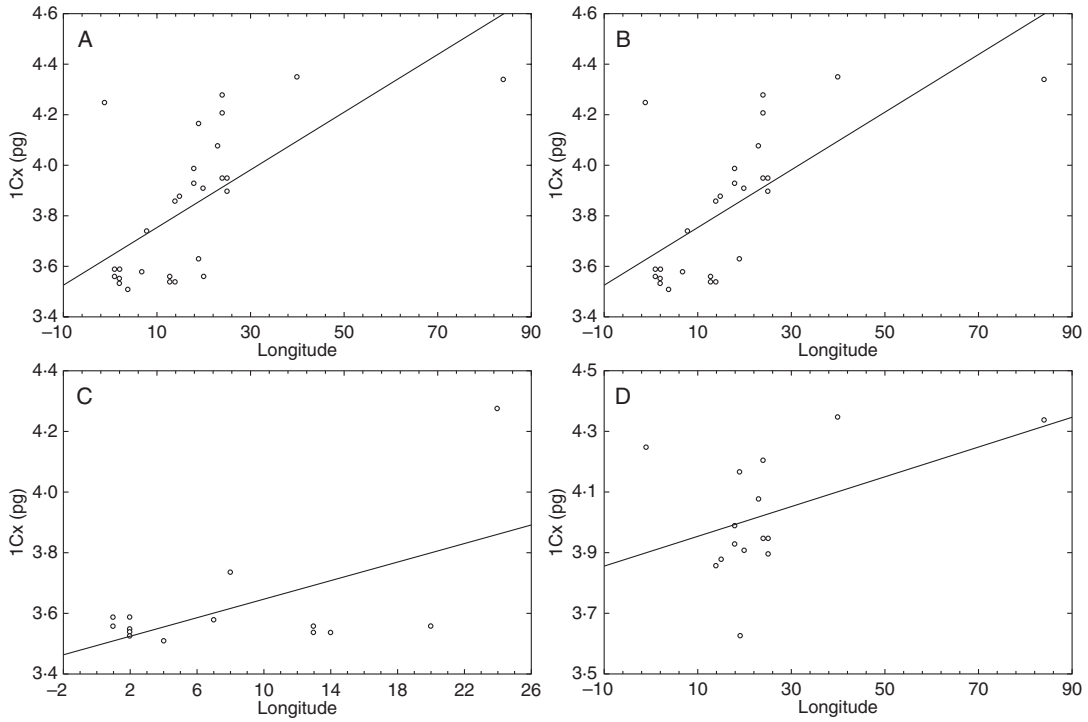


FIG. 7. Longitudinal component of genome size variation for accessions of known phylogenetic origin: (A) based on a complete set of 'western' and 'eastern' accessions/populations analysed by molecular methods (excluding interclade hybrid accessions; Spearman rank coefficient $r = 0{\cdot}656$, $P < 0{\cdot}001$); (B) based on a subset after excluding accessions of widely distributed species (*H. bifidum*, *H. murorum* and *H. umbellatum*; $r = 0{\cdot}688$, $P < 0{\cdot}001$); (C) based on a subset of accessions with 'western' ETS type ($r = 0{\cdot}161$, $P = 0{\cdot}567$); (D) based on subset of accessions with 'eastern' ETS type ($r = 0{\cdot}394$, $P = 0{\cdot}245$).

the 1Cx values: whereas values of the diploid and 'western' introgressed triploid populations ranged from 3·56 pg to 3·67 pg, the 1Cx value of the 'eastern' introgressed triploid was higher (3·81 pg). *Hieracium laevigatum*, *H. amplexicaule* and *H. piliferum* are, according to the ETS sequences, hybrids/hybridogenous types (at least the analysed accessions), and higher intraspecific variation might reflect recurrent polytopic origins. High inter-population variation in *H. pannosum* could also be related to multiple origins as suggested by the different ploidies of the analysed accessions. The high variation in *H. pictum* cannot be explained by the present data. Broader sampling for molecular analysis in these species might reveal hybrid accessions that have not yet been discovered. Thus, unequivocal evidence for intraspecific genome size variation was not found in 'good' species of subgenus *Hieracium*.

Similar results were obtained for *Hieracium* subgenus *Pilosella* in which the majority of wild species/cytotypes possess constant nuclear DNA amounts (variation in fluorescence intensity lower than 3·5%). Nevertheless, higher divergence was observed in six cytotypes belonging to three 'intermediate' species (the same terms as in subgenus *Hieracium*, i.e. hybrids/hybridogenous species) and among genetically variable F_1 offspring of experimental crosses between hexaploid *H. rubrum* and tetraploid *H. pilosella* (Suda *et al.*, 2007).

Intraspecific genome size variation in 'non-hybridogenous' species recently became a matter of debate (Murray, 2005). Whereas many of the examples of variation have been shown to be artefacts of measurement methods (e.g. Teoh and Rees, 1976; Greilhuber, 1998, 2005), there are some reports documenting C-value variation where appropriate controls and standards have been used (Reeves *et al.*, 1998; Hall *et al.*, 2000; Moscone *et al.*, 2003; Pecinka *et al.*, 2006).

Interspecific genome size variation

The present estimates of nuclear DNA content in *Hieracium* subgenus *Hieracium* are the first to be published for this group; all other data on *Hieracium* available so far refer to species of (subgenus) *Pilosella* (Bennett and Leitch, 2005), genomes of which are considerably smaller than those of subgenus *Hieracium* (see below). Holoploid (2C) genome sizes in *Hieracium* species included in the present study ranged 2·37-fold from 7·03 pg to 16·67 pg (mean 2C value 10·16 pg, median 10·61 pg). As almost all so-called 'basic' species were investigated, the present results should cover the genome size variation within the subgenus. A few rare pentaploid hybridogenous (i.e. not 'basic') taxa exist which were not analysed, and thus the upper limit could be higher. Variation in Cx values among species is relatively high (up to about 20%), but more or less continuous.

In *Hieracium* subgenus *Pilosella* with the same basic chromosome number ($n = 9$), holoploid (2C) genome size differs 4·33-fold and ranges from 3·53 pg to 15·30 pg (Suda *et al.*, 2007). However, *Pilosella* has more extensive variation in ploidy, ranging from diploid to octoploid. Monoploid genome sizes (1Cx values) in subgenus *Hieracium* ranged 1·22-fold from 3·51 pg to 4·29 pg (mean 2C value 3·86 pg, median 3·85 pg), whereas genome size in *Pilosella* is distinctly lower (it varies 1·23-fold from 1·72 pg to 2·16 pg). Thus, *Pilosella* has consistently about half the DNA content compared with *Hieracium*. The reasons for these large differences among closely related groups (Fehrer *et al.*, 2007) are unclear at the moment. Chromosomes of subgenus *Hieracium* are distinctly larger than those of *Pilosella* (no quantitative assessments available). Accumulation of repetitive sequence elements as in other plant groups might be one of the causes, but insights into *Hieracium* genomes are still lacking.

Genome size and ploidy

Diploid hawkweeds differ significantly in their 1Cx values from both triploids and tetraploids, but the latter do not differ from each other. This might indicate general downsizing of genomes in polyploid hawkweeds. However, there is no general trend to either downsizing or upsizing within multiploid species. Their origin remains to be elucidated in many cases and could involve autopolyploid origin as well as participation of another taxon (introgression which cannot always be detected by morphology). In Asteraceae, a similar situation was documented in the genus *Centaurea s.l.* in four multiploid (consisting of diploid and tetraploid cytotypes) species: downsizing was found in two species, upsizing in one species and equal monoploid genome size in one species (Bancheva and Greilhuber, 2006). However, downsizing of the genome after polyploidization is widely supposed to be a general trend in angiosperms (Kellogg and Bennetzen, 2004; Leitch and Bennett, 2004; Weiss-Schneeweiss *et al.*, 2006), as seems to be the case for *Hieracium*. The present data also suggest that species of autopolyploid origin might have more uniform genome size (and morphology) than allopolyploids of multiple (hybrid) origin.

Genome size and phylogeny

Genome size distribution basically matches two phylogenetically defined major lineages, i.e. a 'western' and an 'eastern' group (Figs 4 and 5). It thus reveals a strong correlation of nuclear DNA content with the basal evolutionary divergence of *Hieracium* subgenus *Hieracium*. Both clades include sexual diploids, agamospermous triploids and rare tetraploid apomicts. A similar pattern has also been observed for the geographic ranges. Both groups comprise local endemics (e.g. *H. stelligerum* and species of section *Cerinthoidea* in the 'western' group and *H. kittanae*, *H. petrovae* and *H. eriophorum* in the 'eastern' group) and widely distributed species (e.g. *H. murorum* and *H. bifidum* in the 'western' and *H. umbellatum* in the 'eastern' group). Differences of genome size between the two clades were also significant when only diploid or triploid accessions were compared (despite indication for some genome downsizing in polyploids), and thus all cytodemes were analysed together.

As mentioned above, *H. transylvanicum* falls into the western lineage but has a genome size and geographic range congruent with the 'eastern' group. Two alternative scenarios for its origin can be proposed: (1) the species has an eastern origin as suggested by its current distribution and DNA content and shows some ancient introgression from western species, some of which are widespread. Its ETS sequence then became completely homogenized towards the western type by concerted evolution (Arnheim, 1983); and (2) the species originated in western Europe, spread towards the east, the original populations became extinct probably during the Ice Ages and only the eastern populations survived in an eastern glacial refuge like the Carpathian basin. In this case, the high DNA content may be due to other reasons than phylogenetic relationships. Another accession of a 'western' clade species, *H. lachenalii*, has plastid DNA matching some 'eastern' clade species which suggests ancient introgression (J. Fehrer *et al.*, unpubl. res.). Correspondingly, its DNA content is also slightly higher than that of most other species of the 'western' clade. *Hieracium eriophorum*, a local endemic of the Atlantic coast near Arcachon in western France, is another species with an incongruent geographic range and position in the phylogenetic tree. Despite its western European distribution, it is most likely derived from widespread *H. umbellatum*, a species belonging to the same 'eastern' clade (Fig. 4). The morphology of *H. eriophorum* could therefore be interpreted as a local adaptation to sand dunes along the sea coast. The lowest genome size within the 'eastern' clade ($2C = 10{\cdot}89$ pg, $1Cx = 3{\cdot}63$ pg) was detected in triploid *H. naegelianum*. The distribution of this species fits well with other 'eastern' species as it occurs in the Balkan Peninsula and in the Abruzzi Mountains in central Italy, mostly in refugial areas. With respect to morphology, it is the only species of *Hieracium* subgenus *Hieracium* with long underground stolons, which enable the plant to spread vegetatively. Although no evidence of introgression from a 'western' species is apparent from the present molecular data, its occurrence in Italian glacial refuges could be indicative of past contacts and introgression from which only an unusually small genome size is left. Its plastid DNA is unique, and its relationships with other 'eastern' clade species are unresolved.

Hybrid (hybridogenous) 'basic' species

Fourteen 'basic' species were found to be of hybrid origin between 'eastern' and 'western' clade species (Fig. 4): *H. amplexicaule*, *H. bracteolatum*, *H. caesium*, *H. glaucum*, *H. gouanii*, *H. gymnocephalum*, *H. heterogynum*, *H. lachenalii*, *H. laevigatum*, *H. olympicum*, *H. plumulosum*, *H. prenanthoides*, *H. racemosum* and *H. sabaudum*. In addition, individual accessions of *H. pilosum* and *H. villosum* also had this type of hybrid origin. Based on the known mean 1Cx values for the 'western' and 'eastern' groups (3·61 and 4·02 pg, respectively), intermediate genome sizes of the previously mentioned species might be expected. However, hybrid accessions were more similar to the 'eastern' species group and significantly different from the

'western' group. The median 1Cx values of intermediate hybrid taxa and those with dominating 'eastern' ETS sequence were even higher than those for both clades (Fig. 5B). Potential interpretations could be to assume extinct parents with higher genome sizes or, more likely, an increase in DNA content in species of hybrid origin in comparison to their parents, as has been documented in *Helianthus* by Baack *et al.* (2005). The present results also show that hybrids/hybridogenous types with strongly dominating 'western' type ETS (e.g. *H. amplexicaule*, *H. caesium*, *H. gouani*, *H. lachenalii* and *H. prenanthoides*) have similar DNA content in comparison with 'western' species, which is significantly different from hybrids with equal contribution of 'eastern' and 'western' parents. This might suggest repeated backcrossing towards 'western' species at the diploid level before genomes became fixed by apomixis. Intermediate hybrids with dominant 'eastern' ETS, as expected, did not differ significantly from 'intermediate' hybrids or 'eastern' species.

One accession of *H. piliferum*, a basic species distributed in European mountains was found to be a hybrid between an unidentified 'eastern' taxon and '*H.*' *intybaceum* by character additivity in ETS sequences. Comparing 1Cx values of '*H.*' *intybaceum* (3·76 pg, Table 1), *H. piliferum* (3·91 pg) and members of the two ('western' and 'eastern') major clades, past hybridization between '*H.*' *intybaceum* and an 'eastern' species might be expected, which is also congruent with the results of the molecular analyses. Intraspecific genome size variation beyond the arbitrary fluctuation in *H. piliferum* could be indicative of multiple origins.

Evolution of genome size

Two different approaches implemented in BayesTraits for continuous data were applied, namely the maximum likelihood method and the MCMC approach, and two models of evolution (A, drift model; B, directional model) were compared. In most comparisons, our data fit (higher log-likelihoods and harmonic means) model B better, indicating that there is some but no strong trend to genome size increase.

For the complete data set and the maximum likelihood method, the values of scaling parameters did not depart significantly from the default values (1·0) suggesting that tree topology and branch lengths accurately describe the constant variance random-walk model A or B. Thus, genome size is evolving, as expected, given the tree topology, fitting well with the basal split into the two clades. Using the MCMC method, more changes on longer branches were indicated (longer branches contribute more to genome size evolution).

The tempo and mode of evolution differ between the western and eastern clade. In the western clade, values of scaling parameters depart from the default value (1·0) indicating that genome size evolution has not followed the topology or the branch lengths. Phylogenetic history has lower impact ($\lambda < 1$, presumed adaptive response to some external pressures) in this case which may also be reflected by the almost complete lack of resolution of species within the western clade (Fig. 4). Nevertheless, longer paths and branches contribute more to 1Cx evolution (accelerated evolution as time progresses). In contrast, for the eastern clade, the maximum likelihood method revealed that topology and branch lengths accurately describe the constant variance random-walk model A or B. However, using the MCMC method, accelerated rate of evolution in long branches is indicated.

Studies using a phylogenetic approach to evaluate the directionality of genome size evolution revealed both DNA decrease and increase and often different tendencies in genome size diversification in different phylogenetic lineages (e.g. Wendel *et al.*, 2002; Jakob *et al.*, 2004; Caetano-Annolés, 2005; Price *et al.*, 2005). Ancient genome size enlargement followed by more or less drastic parallel reduction in the main phylogenetic lineages was found in *Festuca* (Šmarda *et al.*, 2008).

Genome size and ecogeographic features

In order to identify further components of genome size variation for *Hieracium*, correlations were tested with a number of other factors. A significant, positive correlation was found between 1Cx value and longitude of sampling sites, both in the complete set of accessions and in a restricted set without species with large distribution areas for which the results depend strongly on sampling (Fig. 6). Restriction of these analyses to accessions analysed by molecular data showed that these correlations were even stronger when accessions of ambiguous origin were excluded (Fig. 7A, B). As no significant correlation was found within either of the two clades ('western' and 'eastern'; Fig. 7C, D), it can be concluded that the basal divergence into two phylogenetic lineages is most likely the determining factor of genome size variation in *Hieracium* (or vice versa) rather than longitudinal distribution.

In *Hieracium* subgenus *Pilosella* (Suda *et al.*, 2007), a longitudinal component of genome size distribution was also found: the highest 1Cx values were detected in *H. echioides*, a species distributed mainly in steppic habitats in Asia and eastern Europe (and well differentiated from the remaining species by the absence of a basal leaf rosette at flowering time). However, no comparison with species relationships is available. An opposite relationship between genome size distribution and geographic ranges has been observed in the genus *Cirsium* (Bureš *et al.*, 2004). At the intraspecific level, a geographically correlated variation in DNA content with an increase towards the east has been documented, e.g. in several taxa of *Koeleria* (Pecinka *et al.*, 2006), but no correlation was found in *Sesleria albicans* (Lysák *et al.*, 2000). Thus, there does not seem to be a general trend in genome size variation in relation to longitude. The same holds for a relationship between genome size and latitude, where positive, negative, or non-significant correlations with genome size have been found (reviewed in Knight *et al.*, 2005).

Altitude was also studied. The correlation between this parameter and genome size has been a matter of debate in the past years, and divergent results have been obtained. In subgenus *Hieracium*, genome size variation does not depend on altitude. However, the present data had to be based on altitudes of the sampling sites and are therefore biased by this selection, especially in species occurring across a large range of different altitudes (e.g. *H. bifidum*). Similarly, no correlation between genome size and altitude in Asteraceae was found in *Cirsium*

(Bureš *et al.*, 2004) or *Artemisia* and *Tripleurospermum* (García *et al.*, 2004, 2005). In other families, no or even a negative correlation between genome size and altitude was found by Creber *et al.* (1994), Reeves *et al.* (1998) and Vilhar *et al.* (2002) (all on intraspecific variation in *Dactylis glomerata*). On the other hand, an increase in genome size with higher altitute was found in *Centaurea s.s.* (Bancheva and Greilhuber, 2006) and in some groups of grasses (Bennett, 1976; Laurie and Bennett, 1985; Rayburn and Auger, 1990). Thus, altitudinal genome size variation also seems to be dependent on the particular plant group analysed (Knight *et al.*, 2005). In *Hieracium*, there are specifically montane or alpine taxa, but they are distributed in the Pyrenees, the Alps or the Balkan mountains, i.e. in western, central and eastern European regions, and therefore, the geographic origin of the species strongly dominates any altitudinal genome size variation that might be found.

Also no correlation was found between Cx values and two selected approximate ecological parameters published for German plant species by Ellenberg *et al.* (1992), namely light and temperature. However, the use of Ellenberg's indicator values for *Hieracium* is ambiguous. Many *Hieracium* species have large ecological amplitudes and therefore these approximate values could indeed be useful indicators, but these values are only available for central European species, and therefore species confined to either western or eastern Europe had to be excluded.

Many published correlations between ecogeographical factors (and others such as life form, etc.) and genome size must be interpreted with caution, as phylogenetic information is lacking. Albach and Greilhuber (2004) showed quite different correlations between selected factors (habitat, life history and breeding system) and genome size and DNA C-values in the genus *Veronica* if using a simple statistical test without phylogenetic information or more sophisticated methods incorporating phylogeny (independent contrast, GLSM). Furthermore, the use of linear regression analysis could obscure patterns in relationships if they are not linear. Knight and Ackerly (2002), Knight *et al.* (2005) and Beaulieu *et al.* (2007) used quantile regression analysis and showed that, although the relationship between genome size and a particular parameter was poor for species with small genomes, as genome size increased, the relationship became increasingly significant.

Conclusions

Genome size variation in *Hieracium* subgenus *Hieracium* is congruent with the phylogenetic pattern, with species of putative western European origin having significantly lower genome size than those of eastern European origin. Consequently, a significant longitudinal correlation can also be inferred. Separate analyses of closely related species (i.e. within each phylogenetic clade) clearly show that despite considerably overlapping scales, no significant geographic component is apparent. Thus, in *Hieracium*, any correlation of genome size with longitude, and with other ecogeographic variables such as latitude, altitude, light and temperature, is outweighed by the basal phylogenetic divergence into species of eastern or western European origin.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org/ and provide detailed information about the sample localities.

ACKNOWLEDGEMENTS

We thank S. Bräutigam, R. Douzet, E. Forey, J. Hadinec, P. Ignatova, A. Krahulcová, F. Krahulec, M. Král, K. Marhold, P. Mráz, M. Niketić, A. N. Sennikov, Z. Szeląg, T. Tyler, V. Vladimirov and V. Zavadil for providing us with plant material of several species; G. Mateo, J. A. Rosselló and B. Vreš for their generous help with the field sampling and G. Mateo, F. Schuhwerk and Z. Szeląg for the determination of some plants. We are grateful to P. Trávníček for expert help in the flow cytometry laboratory, P. Caklová for most of the molecular laboratory work, J. Loureiro and J. Suda for critically reading the manuscript and particularly S. Bräutigam and P. Mráz for discussions and many valuable comments. Three anonymous reviewers greatly helped to improve the manuscript. The work was supported by the Czech Science Foundation (grant no 206/05/0657) and partly also by the Academy of Sciences of the Czech Republic (AV0Z60050516) and Ministry of Education, Youth and Sports of the Czech Republic (grant no 0021620828).

LITERATURE CITED

Albach DC, Greilhuber J. 2004. Genome size variation and evolution in *Veronica*. *Annals of Botany* **94**: 897–911.

Arnheim N. 1983. Concerted evolution of multigene families. In: Nei M, Koehn R, eds. *Evolution of genes and proteins*. Sunderland, MA: Sinauer, 38–61.

Arumuganathan K, Earle ED. 1991. Nuclear DNA content of some important plant species. *Plant Molecular Biology Reports* **9**: 208–218.

Baack EJ, Whitney KD, Rieseberg LH. 2005. Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytologist* **167**: 623–630.

Bancheva S, Greilhuber J. 2006. Genome size in Bulgarian *Centaurea s.l.* (Asteraceae). *Plant Systematics and Evolution* **257**: 95–117.

Barakat A, Carels N, Bernardi G. 1997. The distribution of genes in the genomes of Gramineae. *Proceedings of the National Academy of Sciences of the USA* **94**: 6857–6861.

Beaulieu JM, Moles AT, Leitch IJ, Bennett MD, Dickie JB, Knight CA. 2007. Correlated evolution of genome size and seed mass. *New Phytologist* **173**: 422–437.

Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA. 2008. Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytologist* **179**: 975–986.

Bennett MD. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society London, Series B, Biological Sciences* **181**: 109–135.

Bennett MD. 1976. DNA amount, latitude and crop plant distribution. In: Jones K, Brandham PE. eds. *Current chromosome research*. Amsterdam: North-Holland, 151–158.

Bennett MD, Leitch IJ. 1995. Nuclear DNA amounts in angiosperms. *Annals of Botany* **76**: 113–176.

Bennett MD, Leitch IJ. 2005. Plant DNA C-values Database (release 5·0, December 2004). http://www.kew.org/cval/homepage.html (last accessed 16 September 2008).

Bennett MD, Heslop-Harrison JS, Smith JB, Ward JP. 1983. DNA density in mitotic and meiotic metaphase chromosomes of plants and animals. *Journal of Cell Science* **63**: 173–179.

Bennett MD, Leitch IJ, Hanson L. 1998. DNA amounts in two samples of angiosperm weeds. *Annals of Botany* **82** (Suppl. A): 121–134.

Bennetzen JL. 2002. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 29–36.

Bennetzen JL, Ma JX, Devos KM. 2005. Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95**: 127–132.

Bottini M, Greizerstein E, Aulicino M, Poggio L. 2000. Relationships among genome size, environmental conditions and geographical distribution in natural populations of NW Patagonian species of *Berberis* L. (Berberidaceae). *Annals of Botany* **86**: 565–573.

Buitendijk JH, Boon EJ, Ramanna MS. 1997. Nuclear DNA content in twelve species of *Alstroemeria* L. and some of their hybrids. *Annals of Botany* **79**: 343–353.

Bureš P, Yi-Feng Wang, Horová L, Suda J. 2004. Genome size variation in central European species of *Cirsium* (Compositae) and their natural hybrids. *Annals of Botany* **94**: 353–363.

Caetano-Annolés G. 2005. Evolution of genome size in the grasses. *Crop Science* **45**: 1809–1816.

Chrtek J Jr, Mráz P, Severa M. 2004. Chromosome numbers in selected species of *Hieracium s.str.* (*Hieracium* subgen. *Hieracium*) in the Western Carpathians. *Preslia* **76**: 119–139.

Chrtek J Jr, Mráz P, Zahradníček J, Mateo G, Szeląg Z. 2007. Chromosome numbers and DNA-ploidy levels of selected species of *Hieracium s.str.* (Asteraceae). *Folia Geobotanica* **42**: 411–430.

Creber HMC, Davies MS, Francis D, Walker HD. 1994. Variation in DNA C value in natural populations of *Dactylis glomerata* L. *New Phytologist* **128**: 555–561.

Devos K, Brown J, Bennetzen J. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* **12**: 1075–1079.

Doležel J, Doleželová M, Novák F. 1994. Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biologia Plantarum* **36**: 351–357.

Edwards GA, Endrizzi JL. 1975. Cell size nuclear size and DNA content relationships in *Gossypium. Canadian Journal of Genetics and Cytology* **17**: 181–186.

Elena-Rosselló JA, González-Zapatero MA, Navarro F. 1985. Notas cariológicas sobre algunos oroendemismos ibéricos. *Lazaroa* **8**: 91–96.

Ellenberg H, Weber HE, Düll R, Wirth V, Werner W, Paulissen D. 1992. Indicator values of plants in Central Europe. *Scripta Botanica* **18**: 1–258.

Favarger C. 1969. Notes de caryologie alpine V. *Bulletin de la Société Neuchâteloise des Sciences Naturelles* **92**: 13–30.

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

Feliner NG, Aguilar JF, Rosselló JA. 2001. Can extensive reticulation and concerted evolution result in cladistically structured molecular dataset? *Cladistics* **17**: 301–312.

Flavell RB, Rimpau J, Smith DB. 1977. Repeated sequence DNA relationships in four cereal genomes. *Chromosoma* **63**: 205–222.

Gadella TWJ. 1987. Sexual tetraploid and apomictic pentaploid populations of *Hieracium pilosella* (Compositae). *Plant Systematics and Evolution* **157**: 219–246.

García S, Sanz M, Garnatje T, Kreitschitz A, McArthur ED, Vallès J. 2004. Variation of DNA amount in 47 populations of the subtribe Artemisiinae and related taxa (Asteraceae, Anthemideae): karyological, ecological, and systematic implications. *Genome* **47**: 1004–1014.

García S, Inceer H, Garnatje T, Vallès J. 2005. Genome size variation in some representatives of the genus *Tripleurospermum. Biologia Plantarum* **49**: 381–387.

Garnatje T, Vallès J, García S, *et al.* 2004. Genome size in *Echinops* L. and related genera (Asteraceae, Cardueae): karyological, ecological and phylogenetic implication. *Biology of the Cell* **96**: 117–124.

Govindaraju DR, Cullis CA. 1991. Modulation of genome size in plants: the influence of breeding systems and neighbourhood size. *Evolutionary Trends in Plants* **5**: 43–51.

Greilhuber J. 1998. Intraspecific variation in genome size: a critical reassessment. *Annals of Botany* **82** (Suppl. A): 27–35.

Greilhuber J. 2005. Intraspecific variation in genome size in angiosperms: identifying its existence. *Annals of Botany* **95**: 91–98.

Grotkopp E, Rejmanek M, Sanderson MJ, Rost TL. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. *Evolution* **58**: 1705–1729.

Hall SE, Dvorak WS, Johnston JS, Price HJ, Williams CG. 2000. Flow cytometric analysis of DNA content for tropical and temperate new world pines. *Annals of Botany* **86**: 1081–1086.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis suite. *Nucleic Acids Symposium Series* **41**: 95–98.

Hawkins JS, HU G, Rapp RA, Grafenberg JL, Wendel JF. 2008. Phylogenetic determination of the pace of transposable element proliferation in plants: *copia* and LINE-like elements in *Gossypium. Genome* **51**: 11–18.

Huelsenbeck JP, Rannala B. 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276**: 227–232.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**: 460.

Jakob SS, Meister A, Blattner FR. 2004. The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Molecular Biology and Evolution* **21**: 860–869.

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proceedings of the National Academy of Sciences of the USA* **97**: 6603–6607.

Keller ERJ, Schubert I, Fuchs J, Meister A. 1996. Interspecific crosses of onion with distant *Allium* species and characterization of the presumed hybrids by means of flow cytometry, karyotype analysis and genomic *in situ* hybridization. *Theoretical and Applied Genetics* **92**: 417–424.

Kellogg EA, Bennetzen JL. 2004. The evolution of nuclear genome structure in seed plants. *American Journal of Botany* **91**: 1709–1725.

Knight CA, Ackerly DD. 2002. Genome size variation across environmental gradients: a quantile regression analysis. *Ecology Letters* **5**: 66–76.

Knight CA, Molinari NA, Petrov DA. 2005. The large genome size constraint hypothesis: evolution, ecology and phenotype. *Annals of Botany* **95**: 177–190.

Krahulcová A, Krahulec F. 1999. Chromosome numbers and reproductive systems in selected representatives of *Hieracium* subgen. *Pilosella* in the Krkonoše Mts (the Sudeten Mts). *Preslia* **71**: 217–234.

Labani RM, Elkington TT. 1987. Nuclear DNA variation in the genus *Allium* L. (Liliaceae). *Heredity* **59**: 119–128.

Laurie DA, Bennett MD. 1985. Nuclear DNA content in the genera *Zea* and *Sorghum*: intergeneric, interspecific and intraspecific variation. *Heredity* **55**: 307–313.

Lawrence ME. 1985. *Senecio* L. (Asteraceae) in Australia: nuclear DNA amounts. *Australian Journal of Botany* **33**: 221–232.

Leitch I, Bennett MD. 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651–663.

Löve Á. 1969. IOPB Chromosome number reports XXII. *Taxon* **18**: 433–442.

Luque T. 1981. Números cromosómicos para la flora española. *Lagascalia* **10**: 225–256.

Lysák MA, Doležel J. 1998. Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia* **51**: 123–132.

Lysák MA, Rostková A, Dixon JM, Rossi G, Doležel J. 2000. Limited genome size variation in *Sesleria albicans. Annals of Botany* **86:** 399–403.

MacGillivray CW, Grime JP. 1995. Genome size predicts frost-resistance in British herbaceous plants – implication for rates of vegetation response to global warming. *Functional Ecology* **9**: 320–325.

Ma J, Devos KM, Bennetzen JL. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genomic Research* **14**: 860–869.

Mateo G. 2005. Aportaciones al conocimiento del género *Hieracium* en España. X. Novedades para el Pirineo catalán. *Flora Montiberica* **31**: 62–69.

Morgan ER, Burge GK, Seelye JF, Hopping ME, Grant JE. 1998. Production of inter-specific hybrids *Limonium perezii* (Stapf) Hubb. and *Limonium sinuatum* (L.) Mill. *Euphytica* **102**: 109–115.

Morgan MT. 2001. Transposable element number in mixed mating populations. *Genetical Research* **77**: 261–275.

Moscone EA, Baryani M, Ebert I, Greilhuber J, Ehrendorfer F, Hunziger AT. 2003. Analysis of nuclear DNA content in *Capsicum* (Solanaceae) by flow cytometry and Feulgen densitometry. *Annals of Botany* **92**: 21–29.

Mráz P, Chrtek J Jr, Kirschner J. 2001. Genetic variation in the *Hieracium rohacsense* group (*Hieracium* sect. *Alpina*). *Phyton (Horn)* **41**: 269–276.
Murray BG. 2005. When does intraspecific C-value variation become taxonomically significant? *Annals of Botany* **95**: 119–125.
Niketić M, Vladimirov V, Mráz P. 2006. Chromosome numbers and taxonomic-chorological notes on selected species of *Hieracium s.str.* (Asteraceae) from Montenegro. *Phytologia Balcanica* **12**: 85–97.
Nogler GA. 1984. Gametophytic apomixis. In: Johri BM ed. *Embryology of angiosperms.* Berlin: Springer, 475–518.
Noyes RD. 2006. Intraspecific nuclear ribosomal DNA divergence and reticulation in sexual diploid *Erigeron strigosus* (Asteraceae). *American Journal of Botany* **93**: 470–479.
Otto F. 1990. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA. *Methods in Cell Biology* **33**: 105–110.
Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.
Pagel M. 2004. *Continuous. Manual*. Available at http://www.evolution.rdg.ac.uk/BayesTraits.html.
Pagel M, Meade A. 2007. *BayesTraits*. Computer program and documentation available at http://www.evolution.rdg.ac.uk/BayesTraits.html.
Pecinka A, Suchánková P, Lysak MA, Trávníček B, Doležel J. 2006. Nuclear DNA content variation among Central European *Koeleria* taxa. *Annals of Botany* **98**: 117–122.
Petrov DA. 2002. DNA loss and evolution of genome size in *Drosophila. Genetica* **115**: 81–91.
Piegu B, Guyot R, Picault N, *et al.* 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research* **16**: 1262–1269.
Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
Price HJ, Dillon SL, Hodnett G, Rooney WL, Ross L, Johnston JS. 2005. Genome evolution in the genus *Sorghum* (Poaceae). *Annals of Botany* **95**: 219–227.
Rayburn AL, Auger JA. 1990. Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. *Theoretical and Applied Genetics* **79**: 470–474.
Rees H, Cameron FM, Hararika MH, Jones GH. 1966. Nuclear variation between diploid angiosperms. *Nature* **211**: 828–830.
Reeves G, Francis D, Davies MS, Rogers HJ, Hodkinson TR. 1998. Genome size is negatively correlated with altitude in natural populations of *Dactylis glomerata. Annals of Botany* **82**: 99–105.
Rejmanek M, Richardson DM. 1996. What attributes make some plant species more invasive? *Ecology* **77**: 1655–1661.
Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
Schuhwerk F. 1996. *Published chromosome counts in* Hieracium. http://www.botanik.biologie.uni-muenchen.de/botsamml/projects/chrzlit.html (last accessed 16 September 2008).
Schuhwerk F. 2003. Some thoughts on the taxonomy of *Hieracium. Berichte der Bayerischen botanischen Gesellschaft* **72**: 193–198
Shi Y, Gornall RJ, Draper J, Stace CA. 1996. Intraspecific molecular variation in *Hieracium* sect. *Alpina* (Asteraceae), an apomictic group. *Folia Geobotanica Phytotaxonomica* **31**: 305–313.
Soltis DE, Mavrodiev EV, Doyle JJ, Rauscher J, Soltis PS. 2008. ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids. *Systematic Botany* **33**: 7–20.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
Statsoft 1984–2002. *Statistica for Windows*. Tulsa: Statsoft Inc.
Suda J, Krahulcová A, Trávníček P, Rosenbaumová R, Peckert T, Krahulec F. 2007. Genome size variation and species relationships in *Hieracium* sub-genus *Pilosella* (Asteraceae) as inferred by flow cytometry. *Annals of Botany* **100**: 1323–1335.
Sugiyama S. 2005. Developmental basis of interspecific differences in leaf size and specific leaf area among C3 grass species. *Functional Ecology* **19**: 916–924.
Swift H. 1950. The constancy of deoxyribose nucleic acid in plant nuclei. *Proceedings of the National Academy of Sciences of the USA* **36**: 643–654.
Šiško M, Ivanič A, Bohanec B. 2003. Genome size analysis in the genus *Cucurbita* and its use for determination of interspecific hybrids obtained using the embryo-rescue technique. *Plant Science* **165**: 663–669.
Šmarda P, Bureš P, Horová L, Foggi B, Rossi G. 2008. Genome size and CG content evolution of *Festuca*: ancestral expansion and subsequent reduction. *Annals of Botany* **101**: 421–433.
Štorchová H, Hrdličková R, Chrtek J Jr, Tetera M, Fitze D, Fehrer J. 2000. An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* **49**: 79–84.
Štorchová H, Chrtek J Jr, Bartish IV, Tetera M, Kirschner J, Štěpánek J. 2002. Genetic variation in agamospermous taxa of *Hieracium* sect. *Alpina* (Compositae) in the Tatry Mts (Slovakia). *Plant Systematics and Evolution* **235**: 1–17.
Teoh SB, Rees H. 1976. Nuclear DNA amounts in populations of *Picea* and *Pinus* species. *Heredity* **36**: 123–137.
Thalmann C, Guadagnuolo R, Felber F. 2000. Search for spontaneous hybridization between oilseed rape (*Brassica napus* L.) and wild radish (*Raphanus raphanistrum* L.) in agricultural zones and evaluation of the genetic diversity of the wild species. *Botanica Helvetica* **111**: 107–119.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876–4882.
Vicient CM, Suoniemi A, Anamthawat-Jónsson K, *et al.* 1999. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum. The Plant Cell* **11**: 1769–1784.
Vilhar B, Vidic T, Jogan N, Dermastia M. 2002. Genome size and the nucleolar number as estimators of ploidy level in *Dactylis glomerata* in the Slovenian Alps. *Plant Systematics and Evolution* **234**: 1–13.
Vladimirov V. 2003. A new diploid *Hieracium* (Asteraceae: Lactuceae) from Bulgaria. *Botanical Journal of the Linnean Society* **143**: 213–218.
Vladimirov V, Szeląg Z. 2006. A new diploid species of *Hieracium* sect. *Pannosa* (Asteraceae) from Bulgaria. *Botanical Journal of the Linnean Society* **150**: 261–265.
Weiss-Schneeweiss H, Greilhuber J, Schneeweiss GM. 2006. Genome size evolution in holoparasitic *Orobanche* (Orobanchaceae) and related genera. *American Journal of Botany* **93**: 148–156.
Wendel JF, Cronn RC, Johnston S, Price HJ. 2002. Feast and famine in plant genomes. *Genetica* **115**: 37–47.
Wicker T, Keller B. 2007. Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. *Genome Research* **17**: 1072–1081.
Zahn KH. 1921–1923**.** *Hieracium.* In: Engler A ed. Das Pflanzenreich 75, 76, 77, 80, 82 (IV/280). Leipzig: Wilhelm Engelmann.
Záveský L, Jarolímová V, Štěpánek J. 2005. Nuclear DNA content variation within the genus *Taraxacum* (Asteraceae). *Folia Geobotanica* **40**: 91–104.
Zonneveld BJM. 2001. Nuclear DNA content of all species of *Helleborus* (Ranunculaceae) discriminate between species and sectional division. *Plant Systematics and Evolution* **229**: 125–130.

American Journal of Botany

American Journal of Botany: e74–e77. 2012.

AJB Primer Notes & Protocols in the Plant Sciences

Development of novel low-copy nuclear markers for Hieraciinae (Asteraceae) and their perspective for other tribes[1]

Karol Krak[2,5], Inés Álvarez[3], Petra Caklová[2], Andrea Costa[3], Jindřich Chrtek[2,4], and Judith Fehrer[2]

[2]Institute of Botany, Academy of Sciences of the Czech Republic, Zámek 1, 252 43 Průhonice, Czech Republic; [3]Real Jardín Botánico de Madrid, Consejo Superior de Investigaciones Científicas (CSIC), Plaza de Murillo 2, E-28014 Madrid, Spain; and [4]Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, 128 01 Prague, Czech Republic

- *Premise of the study:* The development of three low-copy nuclear markers for low taxonomic level phylogenies in Asteraceae with emphasis on the subtribe Hieraciinae is reported.
- *Methods and Results:* Marker candidates were selected by comparing a *Lactuca* complementary DNA (cDNA) library with public DNA sequence databases. Interspecific variation and phylogenetic signal of the selected genes were investigated for diploid taxa from the subtribe Hieraciinae and compared to a reference phylogeny. Their ability to cross-amplify was assessed for other Asteraceae tribes. All three markers had higher variation (2.1–4.5 times) than the internal transcribed spacer (ITS) in Hieraciinae. Cross-amplification was successful in at least seven other tribes of the Asteraceae. Only three cases indicating the presence of paralogs or pseudogenes were detected.
- *Conclusions:* The results demonstrate the potential of these markers for phylogeny reconstruction in the Hieraciinae as well as in other Asteraceae tribes, especially for very closely related species.

Key words: Asteraceae; gamma-glutamylcysteine synthetase; glycine hydroxymethyltransferase; Hieraciinae; low-copy nuclear markers; squalene synthase.

Despite disadvantages such as incomplete lineage sorting, existence of paralogs, and lack of universality, low-copy nuclear genes are widely used in plant phylogenetics. Their higher rates of evolution and the presumed lack of homogenization via concerted evolution make them suitable markers, especially at low taxonomic levels where nuclear ribosomal DNA (nrDNA) and chloroplast DNA (cpDNA) provide only poor resolution (e.g., Sang, 2002).

Subtribe Hieraciinae (Cichorieae, Asteraceae) is an example of a plant group with relatively recent speciation and extensive reticulation. Fehrer et al. (2007) revealed incongruence between nrDNA and cpDNA as a result of ancient intergeneric hybridizations. Moreover, the internal transcribed spacer (ITS) of nrDNA was not sufficiently variable to resolve the close relationships within the genera. These results prompted us to develop alternative markers that would be more variable and therefore helpful in the elucidation of these phylogenetic relationships. Marker design was done with broader applicability in mind, and the candidate genes were tested for cross-amplification in representatives of eight other Asteraceae tribes.

METHODS AND RESULTS

Candidate genes were selected using the method developed by Álvarez et al. (2008), based on a bioinformatic comparison of a *Lactuca* L. complementary DNA (cDNA) sequence database (http://compgenomics.ucdavis.edu) with a database of Asteraceae sequences from GenBank. Approximately 3000 *Lactuca* sequences matched significantly with Asteraceae sequences. This data set was reduced by considering only matches with unambiguously annotated Asteraceae genes of described proteins and by excluding known large gene families. Two thirds of the remaining data were randomly selected, and the *Lactuca* cDNA sequences characteristic for each match were blasted to the entire GenBank database to verify the identity of the genes. For 20 of these genes, GenBank sequences from a wider range of taxa were aligned to identify conserved regions suitable for primer design. Exon/intron boundaries were identified by comparing the *Lactuca* cDNA sequences with genomic sequences of *Arabidopsis thaliana* (L.) Heynh. Overall variability, occurrence of paralogs, and the possibility of their discrimination by specific primers were evaluated. Information on copy number in various plant groups was searched in the literature, and five candidate genes were finally selected: gamma-glutamylcysteine synthetase (*gsh1*), glycine hydroxymethyltransferase (*shmt*), squalene synthase (*sqs*), ferrochelatase, and delta 1-pyrroline-5-carboxylate synthase (*p5cs*). Test amplifications for Hieraciinae showed that only *gsh1*, *shmt*, and *sqs* yielded PCR products > 500 bp. Cross-amplification success in other Asteraceae, exon/intron proportions, and their variability in nine Hieraciinae accessions were investigated. Twenty species from nine Asteraceae tribes were analyzed (Appendix 1). Plants were cultivated at the Institute of Botany in Průhonice, and voucher specimens were deposited at PRA. All individuals used for this study were diploids—confirmed either by chromosome counts or exclusively diploids

[1]Manuscript received 23 August 2011; revision accepted 21 September 2011.

The authors thank Z. Münzbergová and H. Pánková for the diploid *Aster amellus* sample, A. Krahulcová for the *Hieracium onegense* accession, and P. Jurkovský for DNA extractions during the early stages of this work. An anonymous reviewer is highly acknowledged for valuable comments to the earlier versions of the manuscript. The Czech Science Foundation (206/05/0657 and P506/10/1363), the European Union (SYNTHESYS ES-TAF-1365), the Academy of Sciences of the Czech Republic (AV0Z60050516), and the Ministry of Education, Youth, and Sports of the Czech Republic (0021620828) provided financial support.

[5]Author for correspondence: krak@ibot.cas.cz

doi:10.3732/ajb.1100416

Table 1. Primer sequences and reaction condition variables for the selected candidate genes in Asteraceae.

Gene[a]	Primer name	Primer sequences (5′–3′)	$MgCl_2$ (mM)	T_a (°C)
ferrochelatase	Ferro-2244F	TCTTGGAGGACCAGAGACACTT	4	Touchdown 63–50
	Ferro-2969R	CGCATTGCAATGTAGACATTAGCA		
p5cs	P5CS-462F	TAGGAGCACTTTGCGAGCAG	4	Touchdown 55–45
	P5CS-1475R	CCAGAATATACGAGAAGAATCC		
gsh1	GSH-4668F	CCATGGAGGAGGTTATGTGCAT	3	Touchdown 55–45
	GSH-6683R	GTTCCTCAAATACAGGGTCC		
gsh1–seminested PCR	GSH-4668F	CCATGGAGGAGGTTATGTGCAT	1.5	65
	GSH-HR3	TCCAGAAGCTTCTCTGCTGGAGTT		
shmt	SHMT-260F	GTGATGCAAGCAGTTGGATC	4	Touchdown 55–45
	SHMT-828R	AATCTGTAAGGCATAGTCTCG		
shmt–nested PCR	SHMT-HF1	TATCCAGGTGCTCGATACTATGGTG	1	61
	SHMT-HR1	CCGCAGATATCTTCTTTGTATCAGTCT		
sqs	SQS-3122F	GTTCTCATGGACCAGTTCCA	2	Touchdown 55–45
	SQS-5560R	TGTTCCAATCGCCATGATCT		
sqs–seminested PCR	SQS-HF2	CATGTTTCTGCTGCCTTTCTGGAG	2	65
	SQS-5560R	TGTTCCAATCGCCATGATCT		

Note: T_a = annealing temperature.

[a] While *ferrochelatase* and *p5cs* were omitted from further testing due to low amplicon size in the Hieraciinae, the primer sequences and information on reaction conditions are provided for these two loci as they may amplify longer products in other Asteraceae.

were recorded for the particular species at the localities from which they were collected (Goldblatt and Johnson, 1979; Chrtek et al., 2007; Fehrer et al., 2007). Hieraciinae were selected according to the basic structure of an ITS tree (Fehrer et al., 2007), which was used as a reference phylogeny.

DNA was extracted from leaves according to Štorchová et al. (2000). PCRs were performed in 20 or 50 μL reactions containing 0.5 U of *Taq* DNA polymerase (Fermentas, St. Leon-Rot, Germany), 1× *Taq* buffer with KCl (Fermentas), 0.2 mM of each dNTP, 0.25 μM of each primer, and 12.5–50 ng of genomic DNA. $MgCl_2$ concentrations and primer sequences are given in Table 1. Cycling conditions using a Mastercycler (Eppendorf, Hamburg, Germany) were 95°C for 4 min; followed by 21 cycles of 95°C for 30 s, 55°C for 30 s (–0.5°C in each subsequent cycle), and 72°C for 2.5 min; an additional 14 cycles at 45°C annealing temperature; and a final extension step at 72°C for 20 min.

In most cases, the yield of the amplified products was insufficient for sequencing; therefore, internal primers were designed for the Hieraciinae, and nested or seminested PCRs were carried out as described above using 30–35 cycles and a small aliquot of PCR product (for primer sequences, $MgCl_2$ concentrations, and annealing temperatures, see Table 1). The products were cloned as described in Fehrer et al. (2009). For all *Hieracium* species, 5–24 clones per accession were sequenced; for other genera, cloning was only performed if the direct sequence was not unambiguously readable, and up to four clones were sequenced. Amplification of the ITS region followed Fehrer et al. (2007). Sequencing was done at GATC Biotech (Konstanz, Germany).

Table 2. Features of the three novel low-copy nuclear markers in Asteraceae. In cases where more than one sequence type was detected for an accession, the features of the different copies are separated by slashes.

		Amplicon size (bp)			Proportion of introns (%)		
Tribe	Species	*gsh1*	*shmt*	*sqs*	*gsh1*	*shmt*	*sqs*
Millerieae	*Galinsoga parviflora*	920/n.d.	865/686/696	821	66/n.d.	65/56/57	39
Cardueae	*Carduus crispus*[a]	—	—	—	—	—	—
	Cirsium acaule	—	1140	~1200	—	74	n.d.
Anthemideae	*Artemisia campestris*	1431	586/586/588	1341/1111/1087/990	80	52/49/49	63/57/54/64
Eupatorieae	*Eupatorium cannabinum*	782	790	877	61	62	43
Inuleae	*Inula hirta*[a]	—	—	—	—	—	—
Madieae	*Arnica montana*	888	658	806	66	54	38
Astereae	*Aster amellus*	1060	722	1321	70	58	62
Heliantheae	*Zinnia elegans*	920	628	799	66	52	37
	Helianthus annuus	836	736	760	64	59	34
Cichorieae	*Cichorium intybus*	611/594	832/855	1612	48/49	64/64	69
Cichorieae–Hieraciinae	*Andryala pinnatifida*	649	646	871	62	68	46
	Hispidella hispanica	715/715	696	1093	63/63	70	57
	Hieracium porrifolium[b]	826/822	746	1087	62/62	60	54
	Hieracium umbellatum[b]	825/828	748	1174/1192	62/62	60	57/58
	Hieracium intybaceum[b]	791	819	1003/1003	60	63	49/49
	Hieracium scabrum[c]	803/803	743	1090	61/61	60	54
	Hieracium cf. *guatemalense*[c]	787	756	1092	60	60	54
	Hieracium onegense[d]	776	758	1060/1060	59	60	53
	Hieracium lactucella[d]	778/788	816	1062/1062	59/60	63	53

Note: n.d. = not determined due to incomplete sequencing of the amplicon.

[a] The negative results for *Inula* and *Carduus* may be caused by secondary compounds coextracted with the DNA rather than by primer mismatches. Another DNA extraction method could yield better results.

[b] *Hieracium* subgen. *Hieracium*.

[c] *Hieracium* subgen. *Chionoracium*.

[d] *Hieracium* subgen. *Pilosella*.

Sequences were proofread in Chromas Lite 2.01 (Technelysium Pty. Ltd., Brisbane, Queensland, Australia) and aligned in BioEdit 7.4.0.1 (Hall, 1999). This program was also used to identify pseudogenes by searching for premature stop codons. To assess the variability and phylogenetic signal of the novel markers, phylogenetic analyses of the Hieraciinae samples were carried out for each gene. Prior to analysis, indels were coded according to the simple gap coding method as implemented in SeqState 1.4.1 (Müller, 2005) and included as additional characters. Maximum parsimony analysis was performed in PAUP* 4b10 (Swofford, 2002) using heuristic searches with tree bisection-reconnection branch swapping and 10 random addition sequence replicates. Bootstrapping was done using 1000 replicates with the above-mentioned settings.

No more than two sequence variants per accession were observed in the Hieraciinae samples in all three low-copy genes (Table 2). This intraindividual variation is the result of heterozygosity rather than locus duplication, because sequence variants from one individual belonged to the same clades (Fig. 1, Appendices S1–S4, see Supplemental Data). The low-copy markers had a considerably higher proportion of parsimony informative sites than ITS: 2.1 times in *shmt*, 2.99 times in *sqs*; and 4.5 times in *gsh1* (Table 3).

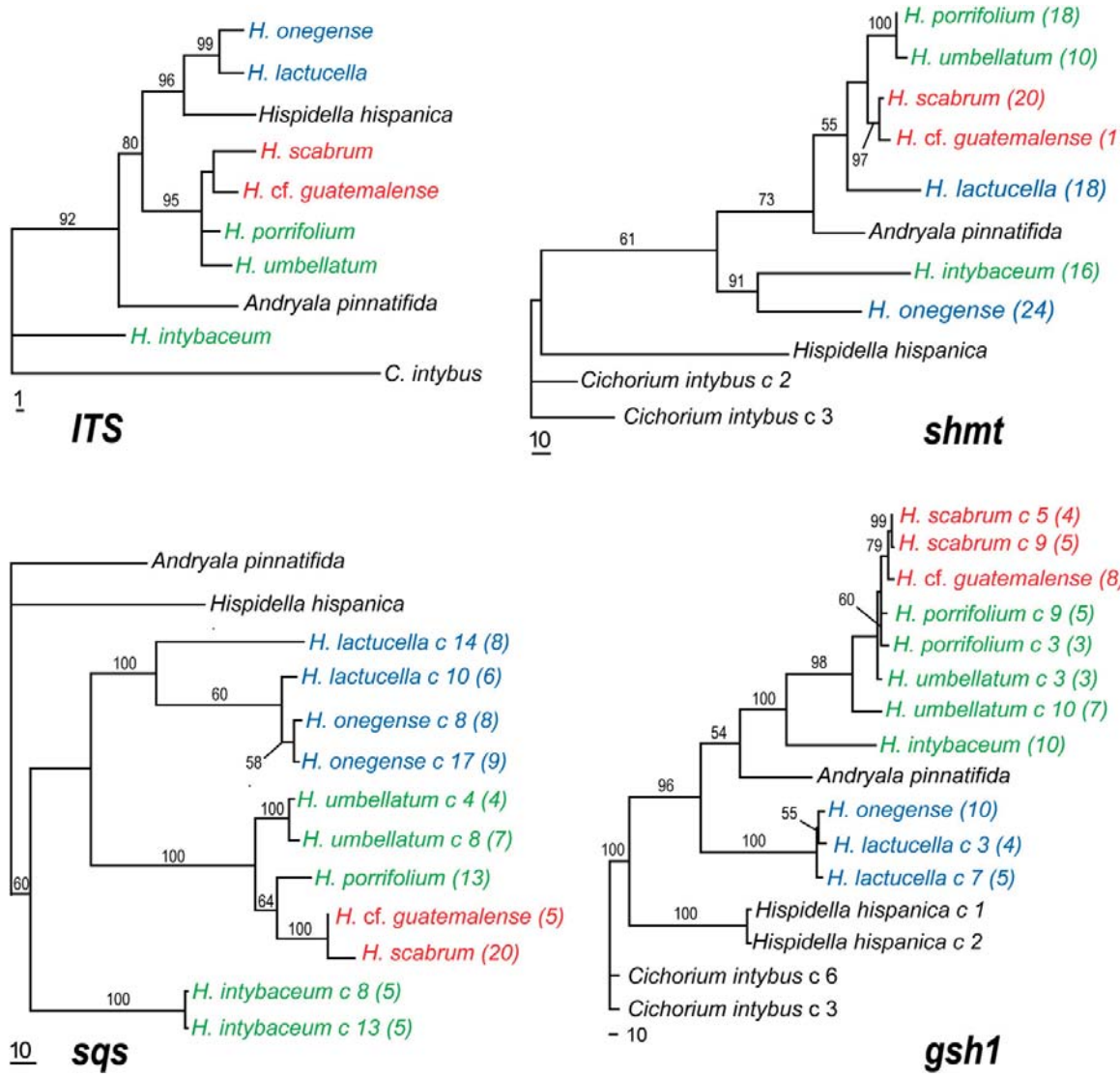


Fig. 1. Phylogenetic analyses of the three low-copy nuclear (LCN) markers and an ITS reference phylogeny for the Hieraciinae. For each of the three LCN markers, the single most parsimonious tree found in the analyses is presented. For ITS, one randomly selected tree from the four most parsimonious ones is presented. In the strict consensus tree, the single unsupported branch was collapsed. For tree statistics, see Table 3. Bootstrap support is indicated above the branches. Blue: *Hieracium* subgen. *Pilosella*, red: *Hieracium* subgen. *Chionoracium*, green: *Hieracium* subgen. *Hieracium*. Clone variants (if more than one per accession) are indicated by, e.g., 'c 14'; the number of identical clones (if more than one) is given in parentheses. For *sqs*, *C. intybus* sequences were not alignable. Note also the different scale for ITS in comparison with the other three markers.

Table 3. Variability and phylogenetic signal of the novel low-copy nuclear markers estimated in the Hieraciinae and compared to the ITS data set.

Data set	Aligned characters	Variable characters	Parsimony informative characters	Pars. inform. chars./ total chars. (%)	No. of MP trees	Tree length	RI	CI
ITS	714	101	32	4.48	4	121	0.869	0.934
shmt	894	243	84	9.40	1	325	0.649	0.858
sqs	1306	314	175	13.40	1	401	0.837	0.868
gsh1	1065	354	215	20.19	1	426	0.911	0.923

Note: CI = consistency index; MP = most parsimonious; RI = retention index.

Cross-amplification was successful in seven out of eight other Asteraceae tribes tested. In addition, positive amplification was observed in the Senecioneae (I. Álvarez, unpublished data). The proportion of introns ranged from 34–80% for different markers/taxa (Table 2). In most cases, one to two alleles were observed (Table 2). Four variants of *sqs* were found in *Artemisia campestris* L.; high exon sequence variation suggested locus duplication in this case. Both divergent paralogs might be used for phylogeny reconstruction in this group. The same plant also showed three *shmt* variants, two of which were pseudogenes. Three variants of *shmt* were also identified in *Galinsoga parviflora* Cav.; these seemed to represent functional copies, however, their orthology/paralogy cannot be estimated based on these data.

CONCLUSIONS

The three newly developed markers amplified in species of eight Asteraceae tribes and proved to be low- or single-copy in most cases. Thus, the primers described here can be used for a broad selection of taxa covering a wide range of variation within the large Asteraceae family. Moreover, in the majority of all accessions tested, only a single sequence variant occurred so that direct sequencing was often possible, which greatly facilitates phylogenetic studies. Occasional paralogs and pseudogenes could be readily identified and should not present a major problem. However, due to the relatively low number of sequenced clones, we cannot exclude that some alleles were missed. In any case, pilot studies for each marker prior to phylogenetic investigation of particular plant groups are recommended (Sang, 2002). According to a pilot study performed on the subtribe Hieraciinae, which contains genera with particularly closely related species, the novel markers showed considerably higher variation than ITS. These results highlight their usefulness, especially for phylogenies at low taxonomic level.

LITERATURE CITED

Álvarez, I., A. Costa, and G. Nieto Feliner. 2008. Selecting single-copy nuclear genes for plant phylogenetics: A preliminary analysis for the Senecioneae (Asteraceae). *Journal of Molecular Evolution* 66: 276–291.

Chrtek, J., P. Mráz, J. Zahradníček, G. Mateo, and Z. Szeląg. 2007. Chromosome numbers and DNA ploidy levels of selected species of *Hieracium* s. str. (Asteraceae). *Folia Geobotanica* 42: 411–430.

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

Fehrer, J., K. Krak, and J. Chrtek. 2009. Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Cichorieae, Asteraceae): Disentangling phylogenetic signal, reticulation, and noise. *BMC Evolutionary Biology* 9: 239.

Goldblatt, P., and D. E. Johnson. 1979 onward (continuously updated). Index to plant chromosome numbers. Missouri Botanical Garden. Website http://mobot.mobot.org/W3T/Search/ipcn.html [accessed 25 April 2006].

Hall, T. A. 1999. BioEdit, a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.

Müller, K. 2005. SeqState: Primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* 4: 65–69.

Sang, T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.

Štorchová, H., J. Hrdličková, J. Chrtek Jr., M. Tetera, D. Fitze, and J. Fehrer. 2000. An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* 49: 79–84.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, Massachussets, USA.

Appendix 1. Voucher information for Asteraceae species used in this study. Information presented: species (voucher specimen/herbarium), collection locale, *gsh1/shmt/sqs*/ITS GenBank accession, chromosome number.

Galinsoga parviflora Cav. (*JC9385/PRA*), Central Bohemia, Sadská; HQ131784/HQ131805-7/HQ131834/–, 2*n* = 16. ***Carduus crispus*** L. (*JC9383/PRA*), Central Bohemia, Praha-Troja; –/–/–/–; 2*n* = 16. ***Cirsium acaule*** Scop. (*JC9384/PRA*), Eastern Bohemia, Chlumec nad Cidlinou; –/HQ131808/–/–; 2*n* = 34. ***Artemisia campestris*** L. (*JC9213/PRA*), Central Bohemia, Praha-Troja; HQ131778/HQ131809-11/HQ131827-30/–; 2*n* = 36. ***Eupatorium cannabinum*** L. (*JC9215/PRA*), Central Bohemia, Hradistko, Kersko; HQ131781/HQ131798/HQ131835/–; 2*n* = 20. ***Inula hirta*** L. (*JC9216/PRA*), Central Bohemia, Cholin; –/–/–/–; 2*n* = 16. ***Arnica montana*** L. (*JC9217/PRA*), Northern Bohemia, Krkonoše Mtns.; HQ131780/HQ131802/HQ131836/–; 2*n* = 36. ***Aster amellus*** L. (*JC9404/PRA*), Central Bohemia, Malíč; HQ131779/HQ131801/HQ131826/–; 2*n* = 18. ***Zinnia elegans*** Jacq. (*JC9401/PRA*), Central Bohemia, Průhonice; HQ131782/HQ131803/HQ131833/–; 2*n* = 24. ***Helianthus annuus*** L. (*JC9218/PRA*), Central Bohemia, Mělník; HQ131783/HQ131804/HQ131832/–; 2*n* = 34. ***Cichorium intybus*** L. (*JC9402/PRA*), Central Bohemia, Průhonice; HQ131776-7/HQ131799-800/HQ131831/AJ633451; 2*n* = 18. ***Andryala pinnatifida*** Ait. (*158131/GLM*), Spain, Canary Islands, La Gomera; HQ131794/HQ131820/HQ131849/AJ633386; 2*n* = 18. ***Hispidella hispanica*** Barnades ex Lam. (*CN 2460/M*), Spain, Sierra de Guadarrama; HQ131797/JF519822/HQ131819/HQ131848/AJ633433; 2*n* = 18. ***Hieracium porrifolium*** L. (*JC9187/PRA*), Austria, Kärnten, Karawanken Mtns.; HQ131785-6/HQ131813/HQ131843/HQ131823; 2*n* = 18. ***Hieracium umbellatum*** L. (*JC9180/PRA*), Poland, Baltic sea coast; HQ131787-8/HQ131814/HQ131841-2/HQ131822; 2*n* = 18. ***Hieracium intybaceum*** All. (*JC9177/PRA*), Italy, Alps, Passo Tonale; HQ131792/HQ131812/HQ131846-7/HQ131821; 2*n* = 18. ***Hieracium scabrum*** Michx. (*JC9178/PRA*), USA, Michigan; HQ131790-1/HQ131815/HQ131844/HQ131825; 2*n* = 18. ***Hieracium* cf. *guatemalense*** Standl. & Steyerm. (*JC9405/PRA*), Guatemala, Volcan Tajomulco; HQ131789/HQ131816/HQ131845/HQ131824; 2*n* = 18. ***Hieracium onegense*** (Norrl.) Norrl. (*JC9403/PRA*), Bulgaria, Western Balkan Range, Berkovitsa; HQ131796/HQ131818/HQ131837-8/AJ633396; 2*n* = 18. ***Hieracium lactucella*** Wallr. (*JC9175/PRA*), Germany, Upper Lusatia, Jonsdorf; HQ131793/HQ131795/HQ131817/HQ131839-40/AJ633389; 2*n* = 18.

Reticulate evolution and lineage sorting in *Hieracium* s. str. (Asteraceae): evidence from low-copy nuclear gene and cpDNA [a,b]

Karol Krak[1], Petra Caklová[1], Jindřich Chrtek[1,2] and Judith Fehrer[1]

[1] *Institute of Botany, Academy of Sciences of the Czech Republic, Zámek 1, 25243 Průhonice Czech Republic*
[2] *Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, 12801 Prague, Czech Republic*

Author for correspondence:

Karol Krak, Institute of Botany, Academy of Scieneces of the Czech Republic, Zámek 1, 252 43, Průhonice, Cech Republic

e-mail: karol.krak@ibot.cas.cz

Telephone: +420271015386

Fax: +420271015105

Running title: Reticulate evolution in *Hieracium*

[a] The manuscript will be submitted to a special issue of Heredity by the end of May 2012. The formatting of the text was done according to the instruction for authors of this journal.

[b] Sequence alignments presented as supplementary files 1, 2 and 3 are provided in the electronic version of the PhD. thesis, available on the CD attached to each hard-copy of the work.

Acknowledgements

The authors thank P. Mráz and P. Trávníček for helpful comments on the manuscript. The Czech Science Foundation (206/05/0657 and P506/10/1363), the European Union (SYNTHESYS ES-TAF-1365), the Academy of Sciences of the Czech Republic (AV0Z60050516), and the Ministry of Education, Youth, and Sports of the Czech Republic (0021620828) provided financial support.

Abstract

The genus *Hieracium* s.str. is characterized by extensive past hybridization, polyploidy and apomictic reproduction. Diploids are rare; the majority of taxa are triploid apomicts. In comparison with other case studies dealing with polyploid speciation, the genus represents a peculiar system: Previous phylogenetic analyses based on nrDNA (ETS) and cpDNA (*trn*T-L) revealed unexpected hybrid origin of about half of the investigated accessions independent of ploidy as well as extinct diploid ancestors. Rather low overall variation indicated relatively recent speciation, but the ETS tree also showed a deep split of the genus into two major lineages that corresponded to genome size and geographic distribution of the species. Both major groups are thought to have survived in different glacial refugia and to have hybridized as a result of secondary contact.

In order to better resolve relationships within the major clades, we added a second, more variable cpDNA (*trn*V-*ndh*C) and a newly developed low-copy nuclear (LCN) marker (squalene synthase, *sqs*). *Sqs* sequence divergence largely exceeded the variation of nr and cpDNA. Most accessions were heterozygous, but the number of *sqs* alleles did not exceed the ploidy level, suggesting orthology and single-copy status except for a single, apparently duplicated allele. While crown groups were roughly equivalent, basal relationships were largely incongruent with all markers. The major lineages of the ETS tree were absent. Instead, several rather divergent alleles and massive allele sharing among taxa were observed. This pattern can be explained by a rapid and rather ancient diversification of the alleles with lineage sorting. The resolution of relationships was further complicated by the hybrid origin of many taxa, but alleles of putative hybrids were often shared with their putative parental taxa / lineages.

These results demonstrate that species trees in closely related and reticulate groups might be approximated by combining evidence from various sources, including seemingly contradictory information. For *Hieracium* s.str., we (1) consider the nrDNA phylogeny as a ‘backbone’ for the interpretation of species origin and relationships because of its congruence with geographic and genomic patterns, (2) use LCN phylogenies to elucidate relationships that are beyond the resolution power of nrDNA, and (3) incorporate information from cpDNA to infer maternal lineages.

Key words: Hieracium, incomplete lineage sorting, low-copy nuclear markers, reticulate evolution, tree incongruence

Introduction

The hawkweed subgenus *Hieracium* s. str. represents undoubtedly one of the major challenges for plant taxonomists. Together with the two other subgenera of *Hieracium,* subgen. *Pilosella* and *Chionoracium* and the genera *Andryala* and *Hispidella*, it is classified as a member of the Lactuceae subtribe Hieraciinae (e.g., Bremer, 1994; Fehrer *et al*., 2007; Krak and Mráz 2009). *Hieracium* s. str. consists of perennial herbs distributed mainly in temperate areas of Europe, Asia and North America. The European mountain ranges (Alps, Carpathians and Pyrenees) and Westernmost Asia are considered as the main centers of the subgenus' diversity. The taxonomic complexity of *Hieracium* s. str. is associated with variation of ploidy level, reproductive mode and extensive past hybridization (Fehrer *et al*., 2009). However recent hybridization is very rare, only two cases of recent natural hybridization have been published so far (*H. krasanii* – Mráz *et al.,* 2005, 2011 and *H. grofae* – Chrtek *et al*., 2006). The basic chromosome number is x=9; most of the recognized species are polyploids, mainly triploids and tetraploids, rarely pentaploids (Merxmüller, 1975; Schuhwerk, 1996; Schuhwerk & Lippert, 1998; Chrtek *et al*. 2004, 2007; Tennant & Rich 2008). Diploids are rare and confined to unglaciated refugia (Merxmüller, 1975; Chrtek, 1996; Mráz, 2003; Castro *et al*., 2007). Polyploids are believed to be obligatory apomicts, whereas diploids are sexual (Chrtek *et al.*, 2009). As a consequence of these processes, huge morphological variation with high numbers of partly overlapping morphotypes can be observed today. This situation resulted in a major disagreement regarding the number of recognized taxa (500 – 8000 species depending on the taxonomic concept) as well as their delimitation (for review on the taxonomic treatments see Willis, 1973). The most comprehensive taxonomic study of the subgenus was published by Zahn (1921-1923). He divided the taxa into two major groups: (i) Basic species (*species principales collectivae*) that are morphologically well distinguishable from each other and (ii) Intermediate species (*species intermediates collectivae*) that combine morphological traits characteristic for two or more basic species; these are believed to be of hybrid origin. Despite the contradictions in taxonomy, the interspecific relationships within this group were left unattended for a long time. Recently Fehrer *et al*. (2009) undertook to investigate the evolutionary history of the subgenus by molecular methods. In that study, sequences of the external transcribed spacer of the nuclear ribosomal DNA (nrDNA ETS) and the *trn*T-*trn*L intergenic spacer of the chloroplast DNA (cpDNA) were analyzed in 60 accessions representing 46 basic species. Low overall genetic variation indicated relatively recent speciation. The ETS tree also showed a deep split of the genus into two major lineages corresponding to taxa with an Eastern or Western European origin, further supported by significant differences in genome size (Chrtek *et al*., 2009). An unexpectedly high level of hybridization among as well as within the major groups was revealed. Hybrid origin concerned almost half of the accessions, including diploids. However, the relationships within the major clades as well as the exact origin of the hybrid taxa remained largely unresolved.

Interspecific hybridization has been recognized as one of the major processes in the evolution of land plants (Arnold, 1997; Mallet, 2007). It is ubiquitous – about 25% of all plant species are known to hybridize with at least one other species (Mallet, 2007). Detecting hybridization among relatively recent and closely related taxa under natural conditions can often be challenging. In the last decades, the determination of hybrid origin has become based predominantly on the analyses of molecular data. If biparentally inherited markers (localized in the nuclear genome)

are used, hybrid origin is usually inferred by the presence of alleles belonging to two (or in some cases more – Kaplan and Fehrer, 2007) parental taxa or by the incongruences among results obtained from different datasets (Linder and Rieseberg, 2004; Fehrer *et al.*, 2007).

Rapid development of novel technologies and decreasing costs of DNA sequencing are going hand in hand with increasing availability of genomic resources even for non-model or agriculturally unimportant organisms (Wu *et al*, 2006; Ueno *et al*., 2010). These achievements enable the use of non-universal and lineage specific molecular tools in biosystematics and related areas. The increased use of low-copy nuclear markers (LCNM) for phylogenetic inference in plants is a good example of this progress. Their contribution to this field has been extensively reviewed (Sang, 2002; Small *et al*., 2004). Their applicability for species level studies within agamic complexes or groups with allopolyploidy and reticulate evolution have been repeatedly reported (e. g. Brysting *et al*., 2007; Grusz *et al*., 2009; Willeyard *et al*., 2009; Kelly *et al*., 2010; Lo *et al.,* 2010; Russel *et al*., 2010). On the other hand, population genetic processes such as incomplete lineage sorting, genetic drift and gene duplications or deletion resulting in undetermined paralogy may hamper results of the analyses (Linder and Rieseberg, 2004). Despite these inconveniences and the effort that often has to be invested to generate these markers, LCNM are usually the tools of choice in cases where incongruent results or poorly resolved relationships were obtained by other markers (e.g., nrDNA and cpDNA). We have recently developed such markers for studying the Hieraciinae in more detail (Krak *et al.*, 2012).

The main aim of the present study is to complement the study of Fehrer *et al.* (2009) by additional molecular data (a second, more variable cpDNA region and a low-copy nuclear marker) in order to: (i) infer further relationships within the major phylogenetic lineages; (ii) identify potentially undetected cases of hybridization; and (iii) better understand patterns of molecular evolution and their implications on speciation within *Hieracium* s.str. In addition, the relative performance of the two nuclear (one low-copy and one multi-copy) markers for the reconstruction of phylogenetic relationships in a closely related and highly reticulate group of taxa will be compared.

Material and methods

Plant material

The present study was performed on the same plant material as the previous studies of Chrtek *et al.*, (2009) and Fehrer *et al.*, (2009). For details on the origin of plant material see Fehrer *et al.,* 2009.

Molecular analyses

For all studied ingroup accessions, DNA extracts from the study of Fehrer *et al.* (2009) were used. For outgroup taxa, DNA extracts from the studies of Fehrer *et al.* (2007; *Hispidella hispanica*, *Hieracium argyrocomum*) and Krak *et al.* (2012; *H. lactucella*, *H. onegense*) were used.

The chloroplast *trn*V–*ndh*C intergenic spacer was amplified using primers *trn*V-a and *ndh*C-a that were modified after Shaw *et al.* (2007) in order to fit better for Asteraceae. PCR amplifications were done in 25 ul 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 x 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) and extension (72°C for 2 min) steps, and a final extension at 72°C for 10 min. PCR products were purified using the Qiaquick purification kit (Qiagen, Hilden, Germany) and sequenced at GATC Biotech (Konstanz, Germany). PCR primers as well as several internal primers (Tab. 1.) were used for sequencing.

Amplification of the *sqs* by semi-nested PCR was done according to Krak *et al.* (2012) with the exception that all PCRs were done in triplicates in order to eliminate the effects of PCR drift and to achieve a more representative proportion of the alleles for cloning. The PCR products were purified as above and directly sequenced with the PCR- and internal sequencing primers (Tab. 2.). Only four samples were homozygous; two showed a single polymorphism. The rest of the accessions (51) was cloned. Cloning and subsequent procedures followed Fehrer *et al.* (2009). The same primers (Table) were used for the sequencing of clones. In general, 3-18 clones per accession (9 on average) were sequenced (GATC Biotech), depending on variation and number of alleles to be retrieved.

Data analyses

Sequence chromatograms were proofread, and ambiguous base callings were corrected manually using Chromas Lite 2.1 (Technelysium Pty Ltd, Brisbane, Australia). Sequences were aligned using Clustal 2.0.11 (Larkin *et al*., 2007), and the alignments were edited manually in BioEdit 7. 0. 4. 1. (Hall, 1999). The sequence alignments are provided as Supplementary Files 1-3.

The data for the *trn*V–*ndh*C intergenic spacer were merged with the *trn*T-*trn*L datas*et al*ready available for the same accessions (Fehrer *et al.*, 2009), and the combined dataset was used for phylogenetic analyses. Insertions/deletions were coded following Fehrer *et al.* (2007). Maximum parsimony (MP) using PAUP* 4.0 b10 (Swofford, 2002) and Bayesian analysis using MrBayes (Ronquist and Huelsenbeck, 2003), were used to infer phylogenetic relationships for the analyzed sequences. For MP analysis, heuristic searches with 10 random sequence addition replicates, saving no more than 100 trees with length greater or equal to 1 per replicate and TBR branch swapping were performed. Bootstrap analysis with 1,000 replicates was performed with the same settings. For Bayesian analysis, the model of molecular evolution best fitting to the data was determined with MrModeltest v2 (Nylander 2004). A F81+G nucleotide substitution model was found in hierarchical Likelihood Ratio Tests (hLRTs) as optimal for the analyzed dataset. Two replicate analyses with four chains each were performed with the default parameters and computed for 5 million generations, sampling every 1,000th tree. All statistical parameters indicated that convergence was reached. The first 1,250 trees per run were discarded as burn-in, and the remaining 7,502 trees were summarized.

For *sqs* separate alignments of cloned sequences for each accession were done, at first. Based on these alignments, allelic variation was examined, and recombinant (resulting most probably during PCR) or redundant sequences were removed. Sequences representing the allelic variation of each accession were then

aligned together. Polymerase errors were corrected manually based on the direct sequence and/or according to multiple clones of the same allele. Direct sequences were also used to confirm that all polymorphisms were accounted for, i.e., that all alleles present in a sample had been identified. Search for pseudogenes was done by *in silico* translation of the exon sequnces in BioEdit (Hall, 1999). Insertions/deletions were coded according to the simple gap coding method as implemented in SeqState (Müller, 2005) and attached to the nexus file generated from the sequence alignment as an additional binary matrix. MP analysis with the same settings as for the analysis of cpDNA data was performed. Computer clusters available at the University of Oslo Bioportal (https://www.bioportal.uio.no/) were used to conduct the analyses.

In order to present the data on a more lucid way, the number of sequences in each monophyletic group recognized based on the results of the preliminary analysis was reduced. The sequences retained for the reduced dataset were selected to represent the overall variability within each group, in order to keep the relationships among the groups unchanged. Indel coding and MP analysis on the reduced dataset was performed as described above. For Bayesian analysis, a GTR+I+G model of molecular evolution was selected as best fitting the data according to the hLRTs. Two replicate analyses with four chains each were performed with the default parameters and computed for 12 million generations, sampling every 1,000th tree. All statistical parameters indicated that convergence was reached. The first 3,000 trees per run were discarded as burn-in, and the remaining 18,002 trees were summarized.

Results

cpDNA phylogeny

The combined alignment of the *trn*T-*trn*L and *trnV-ndhC* intergenic spacers with coded indels contained 1734 characters. MP and Bayesian analyses resulted in consensus trees with similar topologies (Fig. 1). The ingroup taxa (all *Hieracium* s. str.) accessions were resolved into five clusters. The basal most lineage (Clade A) was formed by Eastern Europaean taxa and *H. intybaceum* and could be further subdivided into three subgroups (clades A1-A3). Clade A1 is formed by *H. alpinum* and *H. pojoritense*; clade A2 is composed of two accessions of *H. sparsum*, and two accessions of *H. intybaceum* formed the monophyletic group A3. All other taxa formed a monophyletic group except for the haplotype of *H. mixtum* which was placed into a basal position to all other *Hieracium* s. str. taxa. These were further subdivided into five lineages: (i) a lineage composed of a single accession of *H. naegelianum*; (ii) Clade B with two *H. transylvanicum* accessions at the base, group B1 composed of taxa from the Balkan Peninsula and B2, a group of seven Western Europaean and widely distributed species; (iii) Clade C comprised the Pyrenean accessions; (iv) Clade D consisted of *H. lucidum* and two accessions of *H. prenanthoides* (1252 and 1161) at the base, and subgroups D1 – formed by the third *H. prenanthoides* accession and *H. tomentosum* – and D2 comprising the Eastern Europaean taxa *H. porrifolium*, *H. glaucum*, *H. bupleuroides* 1033, *H. pilosum* and *H. villosum*; (v) Clade E where the rest of the taxa clustered.

Sqs phylogeny

The *sqs* dataset including the additional binary matrix of coded indels contained 1483 characters. Based on the analyses of the *sqs* dataset, the *Hieracium* s. str. accessions were arranged into 14 monophyletic clusters (the analyses of the reduced dataset are presented in Fig. 2; the tree based on the analyses of the complete dataset is presented in Supplementary File 4). *Hispidella hispanica* was placed at the base of the tree. The following three major ingroup lineages were recognized: (i) a lineage composed from subgen. *Pilosella* accessions, (ii) *Hieracium* s. str. accessions with Pyrenean origin together with two accessions of *H. bupleuroides* formed the second lineage – designated as Clade 1, and (iii) a group containing the rest of the *Hieracium* s. str. accessions. Within the last group, Clade 2 composed of *Hieracium lucidum* and *Hieracium intybaceum* was recognized as the sister lineage to the core group formed by the majority of *Hieracium* s. str. accessions. In this core group, two monophyletic lineages were resolved. The first is represented by Clade 3; the second is formed by Clades 4-14. The relationships among the latter clades were supported only partly (usually they gained significant support only in the Bayesian analysis); therefore the basal relationships among them are considered as somewhat ambiguous. On the other hand, most of the clades were well supported by both MP and Bayesian analyses (bootstrap values >70, posterior probabilities >0.94). The only exception was Clade 6 that obtained low support. The relationships within the delimited clades were not supported at all. Some alleles of nine taxa did not group with either of these clades.

Composition of Hieracium s. str. clades

In almost all *Hieracium* s. str. clades, alleles belonging to more than one accession / species were found. Only two clades were formed by alleles of the same accessions: Clade 7 was composed of two alleles of *H. bracteolatum*, and in Clade 13, two alleles of *H. villosum* accession 1305 clustered together. The position of these alleles could be therefore considered as basal to the other clades. Taxon specific (i.e., containing exclusively alleles of accessions belonging to the same taxon) clades were identified in three cases. The first is Clade 4 formed by the alleles identified in the three *H. prenanthoides* accessions. The second is 'Clade 8'composed of *H. alpinum* alleles. The third case is Clade 1b that comprised alleles of two *H. bupleuroides* accessions.

The two major phylogenetic lineages composed of taxa with Eastern or Western Europaean origin as described by Fehrer *et al.* (2009), based on analysis of ETS, were not recovered by *sqs*. However, most of the *Hieracium* s. str. clades identified by *sqs* were formed exclusively by alleles of Eastern or Western taxa and their respective interclade hybrids (assignment to eastern, western and interclade hybrids is adopted from Fehrer *et al.* (2009) and is indicated for all accessions in Tab. 1.). Alleles of Western species occurred in the same clade with alleles of the Eastern species in only three cases. In Clade 10, one allele of *H. pojoritense*, a species with Eastern European origin, clustered together with alleles of several Western species. In Clade 14, alleles of the *H. transylavanicum* (all alleles of both analyzed accessions), a species with western ETS ribotype, grouped with alleles of Eastern species. In Clade 3, an allele of the Balkan species *H. pannosum* grouped with

alleles of Western species. Interestingly, clades formed exclusively by alleles shared among ‘interclade hybrids’ were recognized as well (Clades 4 and 9).

Several subgroups were recognized by Fehrer *et al.* (2009) within the lineage of Eastern taxa. In the representatives of the *H. umbellatum* group – composed of *H. umbellatum*, *H. eriophorum*, *H. canadense*, and *H. virosum* – alleles that grouped in ‘Clade 11’ were identified. The only exceptions were *H. virosum* 1238m and *H. canadense*, that fell at the base of the tree. Moreover, these accessions belong to the same cpDNA haplogroup – haplogroup E, most of the taxa of subgroup E1 (Fig. 1). The members of the *H. porrifolium* group – *H. porrifolium*, *H. bupleuroides*, *H. pilosum*, and *H. villosum*’ – shared the same cpDNA haplogroup (D2), but were not homogeneous regarding the occurrence of *sqs* alleles. The alleles of these four species were found in three clades (and some of the alleles of *H. villosum* and *H. pilosum* fell to the base of the tree), and each of these accessions had a unique combination of alleles. Clades 12 and 1b contained exclusively alleles of accessions belonging to the *H. porrifolium* group. In all accessions from this subgroup, except the diploid *H. porrifolium*, additional alleles belonging to Clade 14 were identified. However, in this clade, alleles of taxa with different ETS ribotypes (Balkan or Western) could be found in addition to the *H. porrifolium* group members. The situation was even more complex within the Balkan subgroup of Eastern taxa. This group was not monophyletic according to ETS (Fehrer *et al.,* 2009), but defined only by a basal position and shared polymorphisms, and it was neither monophyletic with cpDNA and *sqs.* Alleles from Clade 6 seem to be characteristic for these species. However, solely clade 6 alleles were observed only in one accession (*H. kittanae*). They rather occurred in combination with alleles from clades 11 (*H. petrovae*) or 3 (*H. pannosum*). The accessions with at least one clade 6 allele belonged to cpDNA haplogroup B1. Another two taxa, *H. naegelianum* and *H. sparsum*, were included by Fehrer *et al.* (2009) to the Balkan subgroup. In these taxa, neither Clade 6 alleles, nor the cpDNA haplogroup B1 were detected. Instead, either alleles from clade 11 with unique, basal cpDNA (*H. naegelianum*) or alleles from clade 14 in combination with cpDNA haplogroup A2 (*H. sparsum*) were observed. Therefore, at least three lineages could be recognized within the Balkan taxa. In case of *H. pannosum*, previously undetected hybrid origin might be assumed due to the occurrence of a Western allele (belonging to clade 3) in this species.

In contrast to the Eastern taxa, only one well defined subgroup could be characterized in the Western species. Alleles of all Pyrenean taxa fell into Clade 1a and shared the cpDNA haplogroup C, which is in accordance with the ETS data (Fehrer *et al.*, 2009). The rest of the Western species could not be subdivided into subgroups based on the analyses of ETS. According to *sqs*, the alleles of Western species formed three distinct clades (Clade 3, 5, 10). Alleles from Clade 10 were identified in all accessions but *H. humile* 1188, where solely alleles from Clade 3 were found. Alleles exclusively from Clade 10 were identified in one accession only – *H. tomentosum*. This accession was also distinct from the rest of the Western species with respect to the cpDNA. It belonged to the haplogroup D1, whereas the other Western accessions formed haplogroup B2. The remaining Western taxa contained alleles from Clade 10 in combination with alleles from Clades 3 (*H. humile* 1064 and *H. murorum*) or 5 (*H. pictum*, *H. stelligerum*, *H. schmidtii*).

Number of alleles and their distribution across the sqs clades

Sqs exceeded 2 – times the variability of the cpDNA and ETS (Fehrer *et al.* 2009) datasets. For the majority of accessions, more than one allele was identified (Fig. 3., Tab. 3.). In most cases, the number of alleles did not exceed the number expected according to ploidy level. Higher numbers of alleles were observed only in 5 taxa (two diploids, two triploids and one tetraploid, see Tab. 3.) However, the number of clades identified even for these accessions did not exceed the respective ploidy level. Given the low variability within the clades, these additional alleles could be attributed rather to polymerase errors that were not identified as such or to cloning artifacts (as described by Speksnijder *et al.*, 2001) than locus duplication. More over no stop codons were found by *in silico* translation of the exon prats, indicating that functional copies were identified in all accessions.

Alleles from more than one clade were identified in both diploids and polyploids. Accessions with assumed hybrid origin (based on the results of Fehrer *et al.*, [2009] and the present study) dominated these numbers (see Tab. 4.). More than one half of the hybrid accessions (55.2%), irrespective of ploidy level, contained alleles from more than one clade, whereas in non-hybrids, this number was much lower (31.0%). The highest proportion of supposed non-hybrid accessions with alleles from two or more clades could be found among the Western clade species. Here, 45.5% of the taxa possesed alleles from two clades. If ploidy level is taken into account along with the origin of a particular accession, than according to expectation, the highest proportion of accessions with alleles from only one clade occurred in the non-hybrid diploids (87.5%) whereas polyploids with hybrid origin had the highest representation of genomes combining alleles from at least two clades (68.4% in triploids, 100% in tetraploids).

Major incongruences among the data sets

The major incongruence among the three datasets, namely *sqs*, cpDNA, and ETS from the study of Fehrer *et al.* (2009) concerns the existence of major phylogenetic lineages. Whereas two groups corresponding to genome size and geographic origin were revealed by the ETS dataset, these groups were not retained by the cpDNA and *sqs* data. The *sqs* showed a rather high number of well supported clades that roughly corresponded to the subclades identified by the ETS data (Tab. 3.). However, the two nuclear markers differed even in respect of these clusters; taxa belonging to different subgroups according to the ETS shared *sqs* alleles from the same clade (e.g., *H. naegelianum* and *H. petrovae* with taxa from the *H. umbellatum* subgroup – see below). Five phylogenetic lineages were identified in the combined cpDNA dataset which had a much higher resolution than the previous analyses based only on the *trn*T-*trn*L region. Again, several of them corresponded with the subgroups defined according to the ETS (see Tab. 3.). However, haplogroups characteristic for Western taxa grouped together with those of Eastern accessions in two cases: (i) in clade B, the haplogroup B1 found in accessions with Balkan origin formed a sister group to haplogroup B2 that is typical for Western taxa and some interclade hybrids; (ii) in Clade D, Western accessions fell to the base of the group or into haplogroup D1, which is a sister group to Clade D2 composed of Eastern species from the *H. porrifolium* subgroup and interclade hybrids.

Incomplete lineage sorting and hybridization

Sqs alleles from Clade 10 in combination with those from Clade 3 or Clade 5 were found in most of the Western species. The low resolution and absence of further phylogenetic structure within this group of taxa could be attributed to reticulate hybridization or incomplete lineage sorting. Given the fact that these species share the same cpDNA haplogroup (B2) and ETS ribotypes, it is almost impossible to distinguish between these two scenarios. The same is true for *H. naegelianum* and *H. petrovae*, Eastern taxa placed to the Balkan subgroup based on the ETS data, that possessed alleles in Clade 11 together with alleles identified in representatives of the *H. umbellatum* subgroup.

Clear evidence of hybrid origin of several taxa was proposed by Fehrer *et al.* (2009) based on ETS and/or incongruence among the ETS and cpDNA datasets. The proposed origin of several taxa was confirmed by the present study as well (Tab. 1.). These include: *H. bupleuroides* 1033, a hybrid among Eastern *H. porrifolium* and *H. umbellatum* subgroups; *H. candidum* and *H. cerinthoides*, Western species of hybrid origin among a representative of the Pyrenean subgroup and other Western taxa; *H. glaucum*, an interclade hybrid combining ancestral genomes from the Eastern *H. porrifolium* group and an unknown Western taxon, and *H. caesium*, another case of hybridization among a Western linege and the Eastern *H. umbellatum* subgroup. In all these cases, a combination of *sqs* alleles characteristic for the anticipated parental lineages were found. The presence of the Eastern *sqs* allele from Clade 14 in *H. transylvanicum* represents direct molecular evidence for its potential hybrid origin. Before, this was supported only by the incongruence between a typical Eastern lineage-like genome size (and geographic distribution) and a Western ETS ribotype.

Alternative pathways of origin of *H. heterogynum*, *H. pannosum* and *H. pojoritense* compared to those published based on the ETS and *trn*T-*trn*L data, could be proposed based on the present study. In case of *H. heterogynum* in which 3-4 different ribotypes had been found, the *sqs* data indicate a more complex pattern. The presence of an allele from Clade 6 points to a probable involvement of a Eastern taxon from the Balkan subgroup, in addition to a Western species (traces of which were found in ETS and to which the Clade 10 *sqs* allele belongs) and a representative of the *H. umbellatum* subgroup to which also the cpDNA haplotype of this species corresponded. Previously undetected hybrid origin of *H. pannosum* could be inferred from the combination of *sqs* alleles belonging to Clade 6 (characteristic for Eastern taxa with Balkan origin) and Clade 3 (occurring in Western species). Similarly, probable hybrid origin among the two major lineages could be evidenced from the presence of Clade 10 (Western) and Clade 11 (Eastern, predominantly *H. umbellatum* subgroup) alleles in the genome of *H. pojoritense.* Originally, hybridization among two Eastern taxa from the *H. alpinum* (which cpDNA was found in *H. pojoritense*) and *H. umbellatum* subgroups was proposed. However a contribution of some western species is clear from the *sqs* data. The most probable scenarios for the origin of *H. pojoritense* will require hybridization between the *H. alpinum –H. umbellatum* intraclade hybrid with an unknown western accession, or backcrossing of the Western-*H. alpinum* interclade hybrid to some member of the *H. umbellatum* group.

In contrast to the above mentioned taxa, the *sqs* data did not provide direct evidence on the hybrid origin of *H. mixtum*, *H. plumulosum*, *H. amplexicaule*, *H. gouanii*, *H. cordifolium*, and *H. gymnocerinthe*. In these cases, *sqs* alleles from only one clade, representing in most cases (with exception of *H. plumulosum* and *H. mixtum*) the maternal parent, were found. Therefore, the main evidence of their hybrid origin is still based on the ETS dataset. The interclade hybrids *H. bracteolatum*, *H. racemosum* and *H. gymnocephalum* 1207 contained distinct *sqs* alleles in their genomes. However, only one parental lineage (again the maternal one) was identified unambiguously in these taxa. The alleles representing the second lineage had a basal position in the terminal group of the *sqs* tree, therefore the connection to any other evolutionary lineage could not be supported by the *sqs* data. In case of *H. villosum* 1305 and *H. gymnocephalum* 1215, only *sqs* alleles with basal position were identified.

Interesting are the findings concerning the interclade hybrids *H. olympicum*, *H. sabaudum* and *H. prenanthoides* (all three accessions). The first two of them contained Eastern *sqs* alleles (characteristic for the Balkan and *H. umbellatum* subgroups, found in *H. olympicum* and *H. sabaudum*, respectively) in combination with alleles from Clade 9. In this clade, no alleles of other species occurred. Both shared an 'unknown' Western ETS ribotype, and it might be speculated if the Clade 9 allele reflects the extinct Western parent. Similarly, the accessions of *H. prenanthoides* contained *sqs* alleles from Clade 4 that formed a monophyletic group. In accessions 1252 and 1161, only these alleles were found and alleles belonging to the Eastern group of taxa were not identified. In accession 1187, the presence of a third genome apart from the unidentified Eastern and Western ancestor, belonging to a taxon from the *H. umbellatum* subgroup was recognized based on the analyses of ETS. The involvement of the latter group is further supported by the presence of *sqs* alleles from clade 11.

Discussion

Phylogenetic analysis on 58 accessions representing 46 basic species of *Hieracium* s. str. was performed using the *trn*V-*ndh*C intergenic spacer and the low-copy nuclear gene *squalene synthase* (*sqs*). These date were compared to the results obtained by the study of Fehrer *et al.* (2009). We expected that this approach would facilitate in a more comprehensive phylogeny of the subgenus that would allow for identification of previously unresolved relationships. The combined dataset of the two chloroplast markers definitely brought higher resolution and several monophyletic groups were identified. However, both new datasets were incongruent with the results obtained by the ETS. They fail to reveal the deep split into two major phylogenetic lineages as described by Fehrer *et al.* (2009) and clearly monophyletic lineages (described as sobgroubs by these authors) were split into several clades based on the analysis of *sqs*.

Detection of hybrid origin

The present study as well as the study of Fehrer *et al*. (2009) was based exclusively on basic species. These are considered to be pure species according to morphology and hybrid origin hasn't been proposed for any of them so far. However for 29 of these accessions hybrid origin was proposed by Fehrer *et al.* (2009) based on additive patterns of ETS polymorphism, phylogenetic analysis of cloned

sequences, or incongruences among the ETS and *trn*T-*trn*L. Hybrid combinations proposed by these authors have been confirmed by the results of the present study in five cases including hybrids among taxa from the major clades as well as within them. In three cases (*H. heterogynum*, *H. pojoritense* and *H. pannosum*), an alternative scenario of their origin could be proposed when results of the *sqs* dataset are considered in addition to those published in the aforementioned work. In another six cases (*H. mixtum*, *H. plumulosum*, *H. amplexicaule*, *H. gouanii*, *H. cordifolium* and *H. gymnocerinthe*), alleles from only one of the anticipated parental lineages were found. In case of *H. gymnocerinthe*, *H. gouanii* and *H. cordifolium*, well readable direct sequences with low levels of intraindividual polymorphism, readily identified as members of the Pyrenean subgroup of Western taxa, were included to the analyses. The second allele might not be sampled in these accessions due to PCR drift. However, this explanation is not very probable, because two of these accessions are diploids, and the PCRs were carried out in triplicates. Effects of population genetic processes such as genetic drift or selection (maybe not directly on *sqs*, but some locus in close proximity) that may affect genes with low-copy number more strongly than rDNA (Linder and Rieseberg, 2004) may serve as an alternative explanation. The latter scenario could apply for the other three accessions as well. In case of *H. laevigatum* and *H. lachenalii*, hybrid origin was proposed based on the incongruence among the rDNA and plastid datasets. In these cases, only the same parental lineage as detected by the ETS was revealed by *sqs*. A similar incongruence among rDNA and genome size was observed for *H. transylvanicum*. For this taxon, only one *sqs* allele, but not from the same lineage as with rDNA, was detected. Again, processes involved at population level may be responsible for the observed patterns. In the first two cases, these acted in the same direction as concerted evolution on the rDNA whereas in the latter case they went in the opposite direction. In the interclade hybrids *H. prenanthoides*, *H. olympicum*, *H. sabaudum*, *H. bracteolatum*, *H. racemosum*, *H. gymnocephalum* and *H. villosum* 1305, *sqs* alleles with a basal position or from clades represented exclusively by sequences of more than one interclade hybrid accession (Clade 4 – found in all accessions of *H. prenanthoides* and Clade 9 – present in *H. olympicum* and *H. sabaudum*), present alone or in combination with alleles belonging to one of the assumed parental lineages, were found. As these alleles could not be directly connected to any others, positively identified for some of the 'pure' species, the *sqs* could not provide additional support for the data obtained by Fehrer *et al.* (2009). Thus, given the number of congruencies in determination of hybrid combinations, rDNA seems to be clearly superior for the detection of reticulate evolution in *Hieracium* s. str. In addition, Fehrer *et al.* (2009) were able to identify signatures of extinct variation that resided exclusively in the rDNA of interclade hybrids. Some of these accessions contained *sqs* alleles from Clades 4 and 9, indicating that these alleles may be of similar origin than the unknown ribotypes of rDNA, supporting the hypothesis that already extinct taxa contributed considerably to the evolution of the subgenus. If this information is taken into account, than the contribution of *sqs* could not be considered lower than that of the ETS.

Sqs allele sharing among 'pure' accessions: additional evidence on hybridization or incomplete lineage sorting?

In 25 out of 58 accessions (43%), *sqs* alleles belonging to more than one clade, each comprising alleles of several taxa, were found. For analyses based on low-copy nuclear markers, such a pattern is not rare according to the available literature records, especially in cases when species rich groups with recent speciation are investigated (e.g., Peng and Wang, 2008; Russel *et al.*, 2010). Basically, the patterns could be attributed to either population genetic and stochastic processes such as incomplete lineage sorting, gene duplication/deletion, recombination or reticulate evolution (Linder and Rieseberg, 2004). It is very challenging to distinguish among these possibilities. Direct evidence that these patterns are concerning could be obtained only by comprehensive sampling of the studied taxa, that would allow for the estimation of certain population genetic parameters as coalescence time, tests of neutrality, estimate of effective population size. Using this approach, Willeyard *et al*. (2009) were able to detect introgression among species of *Pinus* sect. *Ponderosa* and distinguish it from the background caused by other processes. It has to be kept in mind, however, that the situation in predominantely apomictic species may not be strictly comparable. Once formed, mostly by allopolyploidization, the accessions are fixed, and any subsequent divergence of alleles can only occur by somatic mutations. In our case, this may explain the slight differences of sequences within *sqs* clades after the basic alleles were inherited from the parental taxa. More frequently, hypotheses on interspecific hybridization are based solely on incongruence of multiple unlinked loci. In these cases, the assumption that different loci will be affected to different extent by incomplete lineage sorting whereas reticulate evolution will cause more consistent patterns among the markers, is usually taken into account (Russel *et al*., 2010). Sequences of nrDNA spacers (both ITS and ETS) are widely used as additional markers to complement LCNM datasets. However, these are often affected by concerted evolution, that may delete information of one of the parental lineages (Álvarez and Wendel, 2003; Feliner and Rossello, 2007). In *Hieracium* s. str., concerted evolution is almost completely lacking, even in diploid hybrids that are strictly outcrossing, and mutation accumulation causes noise obscuring greatly the phylogenetic signal. However the hybrid origin of many accessions could be identified based on the remaining signal and additionally on shared polymorphisms (Fehrer *et al*., 2009). The accessions with hybrid origin comprised more than half of those that were not monophyletic in the *sqs* trees. More obscured are the reasons for allele sharing among taxa without other evidence of hybrid origin. This observation is especially frequent within the Western clade accessions. Our sampling based on mostly one accession per species does not allow for approaches used by Willeyard *et al.* (2009), therefore, incomplete lineage sorting of alleles from a highly polymorphic ancestor (or ancestors) has to be considered along with the hypothesis of reticulate evolution involving these taxa. The interesting sharing of *sqs* alleles of Balkan and Western accessions, the most basal and unresolved groups in the ETS tree, may be indicative of lineage sorting. Both groups include diploids that are endemic to the respective areas and have certainly never come into contact, therefore hybrid origin with respect to these shared *sqs* alleles is highly improbable. The same might be true for the earliest splits of *sqs* alleles (*Pilosella* clade, Clades 1 and 2). If the sharing of Western and Balkan alleles is a result of deep coalescense and lineage sorting, these events probably predated the split of the ribotypes in the

Eastern and Western lineages that survived in different glacial refuge areas. On the other hand, hybridization among the taxa from the two major phylogenetic lineages is extensive and is likely associated with secondary contact after the retreat of glaciers (Fehrer *et al.*, 2009). Given that, it is likely that in more closely related taxa that probably survived the last glacial maximum in the same refugia, hybridization could be even more frequent and may not have been detected by ETS, a marker that showed substantially lower variation (the proportion of parsimony informative sites was 1.6 times lower) than *sqs*. Only one non-hybrid taxon from the Western group, *H. humile*, was represented by two accessions. Despite its morphological uniformity, these accessions share only partly their *sqs* alleles. Moreover, the variation in cpDNA clusters *H. humile* 1188 to *H. bifidum* rather than with the other accession of the same species (*H. humile* 1064). These observations could serve as additional, albeit rather poor, evidence of reticulate evolution within the Western group. Similarly, the occurrence of alleles from Clade 11 (specific to the *H. umbellatum* subgroup) and Clade 14 (specific to the *H. porrifolium* subgroup) in accessions from the Balkan subgroup that are accompanied by unique cpDNA haplotypes could speak in favor of hybridization. Therefore the hypothesis involving reticulation within the major clades could be favored slightly more than that solely lineage sorting is responsible for the pattern revealed by the *sqs* phylogeny. Based on the present data, we can only try to draw preliminary conclusions from a comparison of the markers, the species' distribution, ploidy, genome size and (very limited) morphology. An increased number of LCNM and an increased number of accessions per species may shed more light on the evolution of this intricate plant group.

References

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.

Bremer K (1994). *Asteraceae: Cladistics and classification*. Timber Press: Portland.

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.

Castro M, Mateo G, Rosseló JA (2007). Chromosome numbers in *Hieracium* and *Pilosella* species (Asteraceae) from the Iberian Peninsula and the Balearic Islands. *Bot J Linn Soc* **153:** 311–320.

Chrtek J jun (1996). Chromosome numbers in selected *Hieracium* species (Compositae) in the Sudeten Mts and West and Ukrainian East Carpathians. – *Fragmenta Floristica et Geobotanica* **41:** 783–790.

Chrtek J jun, Mráz P, Severa M (2004). Chromosome numbers in selected species of *Hieracium* s.str. (*Hieracium* subgen. *Hieracium*) in the Western Carpathians. *Preslia* **76:** 119–139.

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

Chrtek J, Mráz P, Zahradníček J, Mateo G, Szeląg Z (2007). Chromosome numbers and ploidy levels of selected species of *Hieracium* s. str. (Asteraceae). *Folia Geobot* **42:** 411–430.

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.

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.

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

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis suite. *Nucl Acids Symp Ser* **41:** 95–98.

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

Kelly LJ, Leitch AR, Clarkson JJ, Hunter RB, Knapp S, Chase MW (2010). IntragenicRecombinationEvents and Evidence for Hybrid Speciation in Nicotiana (Solanaceae). *Mol Biol Evol* **27:** 781–799.

Krak K, Mráz P (2009). Trichomes in the tribe Lactuceae (Asteraceae) - taxonomic implications. *Biologia* **63:** 616–630.

Krak K, Álavarez 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. *Am J Bot* e74–e77

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*. 23: 2947–2948.

Linder CR, Rieseberg LH (2004). Reconstructing patterns of reticulate evolution in plants. *Am 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.

Mallet J (2007). Hybrid speciation. *Nature* **446:** 279–283.

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

Mráz P (2003). Mentor effects in the genus *Hieracium* s. str. (Compositae, Lactuceae). *Folia Geobot* **38:** 345–350.

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, Chrtek J, Fehrer J (2011). Interspecific hybridization in the genus Hieracium s. str.: evidence for bidirectional gene flow and spontaneous allopolyploidization. – *Plant Syst Evol* **293:** 237–245.

Müller K (2005). SeqState - primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* **4:** 65-69

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

Peng D, Wang XQ (2008). Reticulate evolution in *Thuja* inferred from multiple gene sequences: Implications for the study of biogeographical disjunction between eastern Asia and North America. *Mol Phylogenet Evol* **47:** 1190–1202.

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). Reticulateevolution 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/index.html?/staff/schuhwerk.html.

Schuhwerk F, Lippert W (1998). Chromosomenzahlen von Hieracium (Compositae, Lactuceae) Teil 2. *Sendtnera* **5:** 269–286.

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. *Am J Bot* **94:** 275–288.

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.

Štorchová H, Hrdličková R, Chrtek J jr, Tetera M, Fitze D, Fehrer J (2000). An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* **49:** 79–84.

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

Tennant D, Rich T (2008). *British alpine hawkweeds*. Botanical Society of the British Isles: London.

Ueno S, Le Provost G, Léger V, Klopp C, Noirot C, Frigerio JM *et al*. (2010). Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* **11:** 650

Willyard A, Cronn R, Liston A (2009). Reticulateevolution and incomplete lineage sorting among the ponderosa pines. *Mol Phylogenet Evol* **52:** 498–511.

Willis J.C. (1973): *A dictionary of flowering plants and ferns*, 8-th edn. Cambridge University Press: Cambridge.

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006). Combining Bioinformatics and Phylogenetics to Identify Large Sets of Single-Copy Orthologous Genes (COSII) for Comparative, Evolutionary and Systematic Studies: A Test Case in the Euasterid Plant Clade. *Genetics* **174:**1407–1430.

Zahn KH. (1921–1923). *Hieracium* L. In: Engler H.G.A. (ed.) Das Pflanzenreich: Regni vegetabilis conspectus. IV, 280, Compositae – Hieracium. Band 76: 1–32, Wilhelm Engelmann, Leipzig.

Tables

Table 1. Primers for the cpDNA intergenic spacer *trn*V-*ndh*C, newly developed in frame of the present study.

Primer name	Sequence	Used for
trnV-a	GAARGTCTACGGTTCGAGTC	Amplification
ndhC-a	AGAAATGCCCAAAAAATATCATATTC	Amplification
cp5-200f	GACAATCATTCCTATAAGC	Sequencing
cp5-320r	TGCGAAGAGTACTGAAC	Sequencing
cp5-480f	TATGGAAAGAGAAGACTAGG	Sequencing
cp5-1000f	CTTTAGAGAGAATACTCG	Sequencing
cp5-1090r	TTTTCGAATTTTGAATTC	Sequencing
cp5-1130r	CTCTTATCCAAATTCTCTTG	Sequencing

Table 2. Primers for the *sqs*, newly developed in the present study.

Primer name	Sequence	Used for
SQS-280f	GGCAATTGAGGATATAAC	Sequencing
SQS-300r	GAATTTTGCCATTCCTGC	Sequencing
SQS-580f	ATGTTGCTGGACTTGTTG	Sequencing
SQS-640r	GAATTGGAGATGGAATC	Sequencing
SQS-860r	ACGAGGCCAAAACATGCG	Sequencing
SQS-880f	CATGTTTTGGCCTCGTG	Sequencing
SQS-1080r	AATGTGGATTAAAGCATTTG	Sequencing
SQS-1100f	ACAAATGCTTTAATCCACATTG	Sequencing

Table 3. Basic overview on cpDNA, ETS and *sqs* results for each analyzed accession. Origin ETS and ETS subclades are referring to the origin of the particular accessions as described according to Fehrer *et al.,* 2009

Taxon	Accession No	Ploidy	cpDNA haplogroup	Origin ETS	ETS subclade	No of *sqs* alleles	*sqs* sequence	*sqs* clade No	No of *sqs* clades identified
Hieracium alpinum	Alp.Ukr	2x	A1	Eastern	Ealp	2	Alp.Ukr.X1c Alp.Ukr.X3c	8 8	1
	Alp. Boa2	2x	A1	Eastern	Ealp	2	Alp.Boa2 X3c Alp.Boa2 X2c	8 8	1
Hieracium aplexicaule	1050/1	3x	C	Interclade hybrid	Wpyr-E	2	1050_1_X4c	1A 1A	1
Hieracium bifidum	1213/2	3x	B2	Western	n.d.	2	1213X1Nc 1213X3Nc	10 10	1
Hieracium bracteolatum	1240/2	3x	E1	Interclade hybrid	Wx-Eumb	4	1240X10c 1240X1c 1240X12c 1240X4c	11 7 11 7	2
Hieracium bupleuroides	1212/2	3x	D2	Eastern	Eporr	3	1212X3c 1212X2Lc 1212X5c	1B 12 14	3
	1033/3	3x	E1	Eastern	Eumb-Eporr	3	1033X3c 1033X2c 1033X10c	11 14 1B	3
Hieracium caesium	1231 (plumb)	4x	B2	Interclade hybrid	W-Eumb	5	plumX10c plumX3c plumX4c plumX11c plumX8c	11 11 10 10 3	3
Hieracium canadense	canad	3x	E1	Eastern	Eumb	DS		BASE	1

Taxon	Accession No	Ploidy	cpDNA haplogroup	Origin ETS	ETS subclade	No of sqs alleles	sqs sequence	sqs clade No	No of sqs clades identified
Hieracium candidum	1197/3	3x	C	Western	Wpyr-W	2	1197X5c 1197X4c	10 1A	2
Hieracium cerinthoides	1176/2	3x	C	Western	Wpyr-W	2	1176X2c 1176X1Sc	1A 10	2
Hieracium cordifolium	1177/5	2x	C	Western	Wpyr-W	DS	1177_5c	1A	1
Hieracium eriophorum	1221/1	2x	E1	Eastern	Eumb	DS	1221_1c	11	1
	1222/2	2x	E1	Eastern	Eumb	2	1222X1c 1222X3c	11 11	1
Hieracium glaucum	1230/3(Gla3)	3x	D2	Interclade hybrid	W-Eporr	3	Gla3X1c Gla3X2c Gla3X4c	12 12 10	2
Hieracium gouani	1171/4	2x	C	Interclade hybrid	Wpyr-E	DS	1171_4c	1A	1
Hieracium gymnocephalum	1215/1	2x	E2	Interclade hybrid	Wy-Ex	DS	1215_1_gcl	BASE	1
	1207/2	3x	E2	Interclade hybrid	Wy-Ex	2	1207X3c 1207X5c	4 6	2
Hieracium gymnocerinthe	1172/4	3x	C	Western	Wpyr-W	DS	1172_4c	1A	1
Hieracium heterogynum	1250/2 (het)	3x	E1	Interclade hybrid	W-E	2	hetX2c hetX1c	6 10	2
Hieracium humile	1064/2	4x	B2	Western	n.d.		1064X7c 1064X2c	10 3	2
	1188/2	3x	B2	Western	n.d.	2	1188X2c 1188X6c	3 3	1
Hieracium intybaceum	1069/1	2x	A3			2	1069_c8 1069_c13	2 2	1
	InbKar	2x	A3			1	InbKaerc	2	1
Hieracium kittanae	1228/2 (H.kit)	2x	B1	Eastern	Ebalc	2	kittX8c kittX2c	6 6	1

Taxon	Accession No	Ploidy	cpDNA haplogroup	Origin ETS	ETS subclade	No of *sqs* alleles	*sqs* sequence	*sqs* clade No	No of *sqs* clades identified
Hieracium laevigatum	1031/11	3x	E1	Western	W-Eumb	3	1031X3c 1031X4c 1031X5c	10 10 10	1
Hieracium lachenalii	1160/2	3x	E1	Western	W-Eumb	3	1160X1c 1160X2c 1160X3c	10 10 10	1
Hieracium lawsonii	1175/1	3x	C	Western	Wpyr	2	1175X1c 1175X4c	1A 1A	1
Hieracium lucidum	H.lucidum	2x	D	Western	W-Wx	2	lucX2c lucX1c	BASE 2	2
Hieracium mixtum	H.mixt	3x	unique, basal to B,C, D,E	Interclade hybrid	W-E	2	mixX7c mixX6c	1A 1A	1
Hieracium murorum	875/1	3x	B2	Western	n.d.	2	8751X4c 8751X6c	3 10	2
Hieracium naegelianum	1208/2	3x	unique, basal to B,C,D	Eastern	Ebalc	2	1208X2c 1208X7c	11 11	1
Hieracium olympicum	1206/3 (oly)	3x	E2	Interclade hybrid	Wx-Ebalc	3	olyX4c olyX1c	9 6	2
Hieracium pannosum	1205/1 (pan)	3x	B1	Eastern	Ebalc	3	panX8c panX1c panX10c	3 6 6	2
Hieracium petrovae	1229 (petr)	2x	B1	Eastern	Ebalc	3	petrX1c petrX2c petrX7c	6 6 11	2
Hieracium pictum	1067/4	3x	B2	Western	n.d.	3	1067X2c 1067X1c 1067X6c	5 base 10	3

Taxon	Accession No	Ploidy	cpDNA haplogroup	Origin ETS	ETS subclade	No of *sqs* alleles	*sqs* sequence	*sqs* clade No	No of *sqs* clades identified
Hieracium pilosum	1226/1	3x	D2	Eastern	Eporr	3	12261X5c 12261X2Sc 12261X3Sc	14 BASE BASE	3
Hieracium plumulosum	1218/2	2x	E	Interclade hybrid	W-E	2	1218X1c 1218X4c	10 10	1
Hieracium pojoritense	Poi.Rom	2x	A1	Eastern	Eumb-Ealp	3	poiX4Lc poiX3c poiX5c	11 11 10	2
Hieracium porrifolium	1052/9	2x	D2	Eastern	Eporr	1	HQ131843	12	1
Hieracium prenanthoides	1252(prenFRA)	2x	D	Interclade hybrid	W-E	2	PrenF_alternative PrenFX8c	4 4	1
	1161/2	3x	D1	Interclade hybrid	W-E-Wx	2	1161_2c1 1161_2c2	4 4	1
	1187/1	3x	D	Interclade hybrid	W-E-Eumb	3	1187X3Lc 1187X1c 1187X2c	11 4 4	2
Hieracium racemosum	874	3x	E1	Interclade hybrid	Wx-Eumb	3	874X2Lc 874X4c 874X3c	11 11 BASE	2
Hieracium ramondii	1173/3	3x	C	Western	Wpyr	DS	1173_3c	1A	1
Hieracium recoderi	1174/4	2x	C	Western	Wpyr	DS	1174_4c	1A	1
Hieracium sabaudum	1098/2	3x	E1	Interclade hybrid	Wx-Eumb	3	1098X6c 1098X8c 1098X10c	11 11 9	2
Hieracium schmidtii	1025/3	3x	B2	Western	n.d.	3	1025X14c 1025X11c 1025X12c	5 10 10	2

Taxon	Accession No	Ploidy	cpDNA haplogroup	Origin ETS	ETS subclade	No of *sqs* alleles	*sqs* sequence	*sqs* clade No	No of *sqs* clades identified
Hieracium sparsum	1251/1(spaJCH)	2x	A2	Eastern	Ebalc	2	spaJCH1X1c	14	1
							spaJCH1X2c	14	
	spa.sst.2	2x	A2	Eastern	Ebalc	2	Spasst1X1c	14	1
							Spasst1X2c	14	
Hieracium stelligerum	1233/1	2x	B2	Western	n.d.	2	1233X3c	10	2
							1233X2c	5	
Hieracium tomentosum	1066/8	2x	D1	Western	n.d.	2	1066X3c	10	1
							1066X1c	10	
Hieracium transylvanicum	traBoa	2x	B	Western	n.d.	2	TraBoaX1c	14	1
							TraBoaY1c	14	
	1077/7	2x	B	Western	n.d.	2	1077X1c	14	1
							1077X2c	14	
Hieracium umbellatum	1021/1	2x	E1	Eastern	Eumb	2	10211_c8	11	1
							10211_Y2	11	
	um.AM.1	2x	E1	Eastern	Eumb	1	umAm1c	11	1
Hieracium villosum	1029/1	4x	D2	Eastern	Eporr	4	1029X3c	14	2
							1029X2c	14	
							1029X1c	14	
							1029X4c	BASE	
	1305/3	3x	D2	Interclade hybrid	Wy-Eporr	4	1305X1c	BASE	2
							1305X6c	BASE	
							1305X2c	13	
							1305X4c	13	
Hieracium virosum	1238/1	3x	E1	Eastern	Eumb	DS	1238c	BASE	1
	virR	3x	E	Eastern	Eumb	2	virRX4c	11	1
							virRX1c	11	

Table 4. Proportion of taxa with *sqs* alleles from one or more clade, correlated to origin (based on all three markers) and ploidy level.

	Hybrids		non-hybrids		non-hybrids Eastern		non-hybrids Western		ploidy level (hybrids+nonhybrids)	
	number	%	number	%	number	%	number	%	number	%
accessions total	29	100,0	29	100,0	18	100,0	11	100,0	58	100,0
diploids total	9	100,0	14	100,0	11	100,0	3	100,0	23	100,0
triploids total	19	100,0	13	100,0	6	100,0	7	100,0	32	100,0
tetraploids total	1	100,0	2	100,0	1	100,0	1	100,0	3	100,0
accessions alleles 1 clade	13	44,8	20	69,0	14	77,8	6	54,5	33	56,9
diploids	7	77,8	12	85,7	10	90,9	2	66,7	19	82,6
triploids	6	31,6	8	61,5	4	66,7	4	57,1	14	43,8
tetraploids	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0
accessions alleles >1 clade	16	55,2	9	31,0	4	22,2	5	45,5	25	43,1
diploids	2	22,2	2	14,3	1	9,1	1	33,3	4	17,4
triploids	13	68,4	5	38,5	2	33,3	3	42,9	18	56,3
tetraploids	1	100,0	2	100,0	1	100,0	1	100,0	3	100,0

Figure legends

Fig. 1. Phylogeny of *Hieracium* s. str. based on the combined datasets of *trn*T-*trn*L and *trn*V-*ndh*C cpDNA intergenic spacers. One of the 24 most parsimonious trees is presented. Bootstrap values of the MP analyses higher than 50 (red numbers above or below branches) and posterior probability values higher than 0.90 from Bayesian analysis (light blue numbers above or below the branches) are presented for each group. The origin of each accession as inferred in Fehrer *et al.* (2009) is highlighted: Western species are in blue, Eastern species are in red, and interclade hybrids in orange. Diploid accessions are given in boldface.

Fig. 2. Phylogeny of *Hieracium* s. str. based on the analyses of the *sqs* dataset. The consensus tree of the Bayesian analysis is presented. Posterior probability values higher than 0.90 (light blue numbers above or below branches) and bootstrap values from the MP analysis (red numbers above or below branches) higher than 50% are presented for each group. Monophyletic groups (clades) are marked by gray and yellow rectangles, and the names of taxa resolved within the particular clades are placed into the rectangles. The clade names corresponding to the text in the Results section are in orange. A broad overview on the origin of the taxa is given by orange symbols next to the particular clade name (E = Eastern, W = Western, H = Interclade hybrid). Taxa that were not resolved within the clades and fell to the base of the terminal groups are given in gray.

Fig. 3. Graphical representation of accessions containing alleles from more than one clade, based on the *sqs* tree from Fig. 2. The inferred origin of each accessions is given: W = Western, E = Eastern, H = Interclade hybrid, Wh = Intraclade hybrid – western, Eh = Intraclade hybrid – eastern.

Fig. 1.


Hispidella
Pilosella
pojoritense
alpinum alp.Ukr
alpinum alp.Boa2
sparsum spa.sst2
sparsum 1251
intybaceum Inb.Kaer
intybaceum 1069
A1
A2
A3
A
mixtum
naegelianum
pannosum
petrovae
kittanae
B1
humile 1064
pictum
murorum
caesium
stelligerum
schmidtii
humile 1188
bifidum
B2
transylvanicum 1077
transylvanicum tra.Boa
B
ramondii
lawsonii
candidum
cerinthoides
cordifolium
recoderi
gymnocerinthe
gouani
amplexicaule
C
lucidum
prenanthoides 1187
prenanthoides 1252
prenanthoides 1161
tomentosum
D1
porrifolium
pilosum 1226/1
villosum 1029
glaucum
villosum 1305
bupleuroides 1212
D2
D
plumulosum
virosum vir1
virosum 1238
canadense
bupleuroides 1033
eriophorum 1221
eriophorum 1222
umbellatum um.AM.1
umbellatum 1021
racemosum
sabaudum
bracteolatum
heterogynum
laevigatum
lachenalii
E1
olympicum
gymnocephalum 1207
gymnocephalum 1215
E2
E
1

Fig. 2.

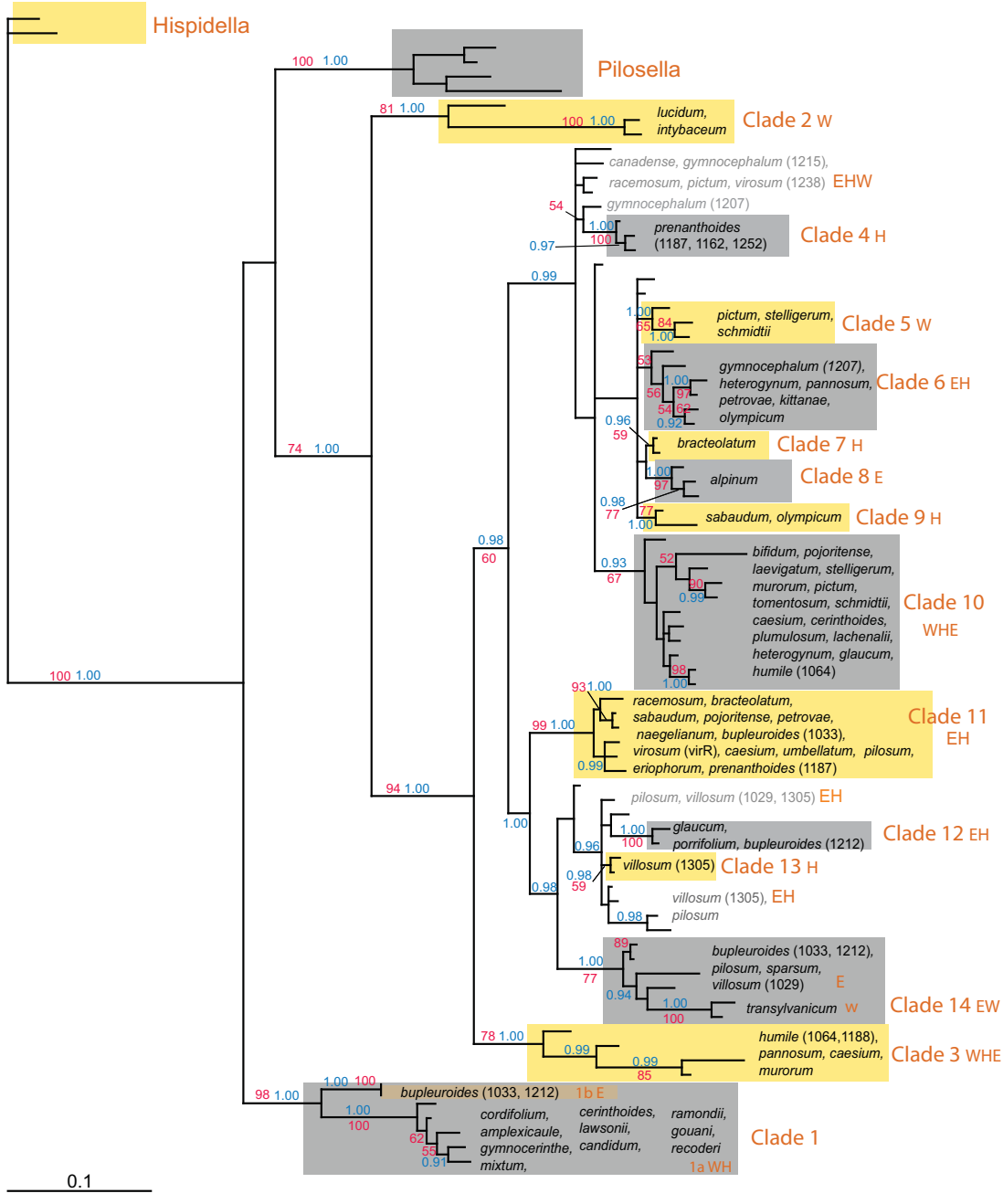
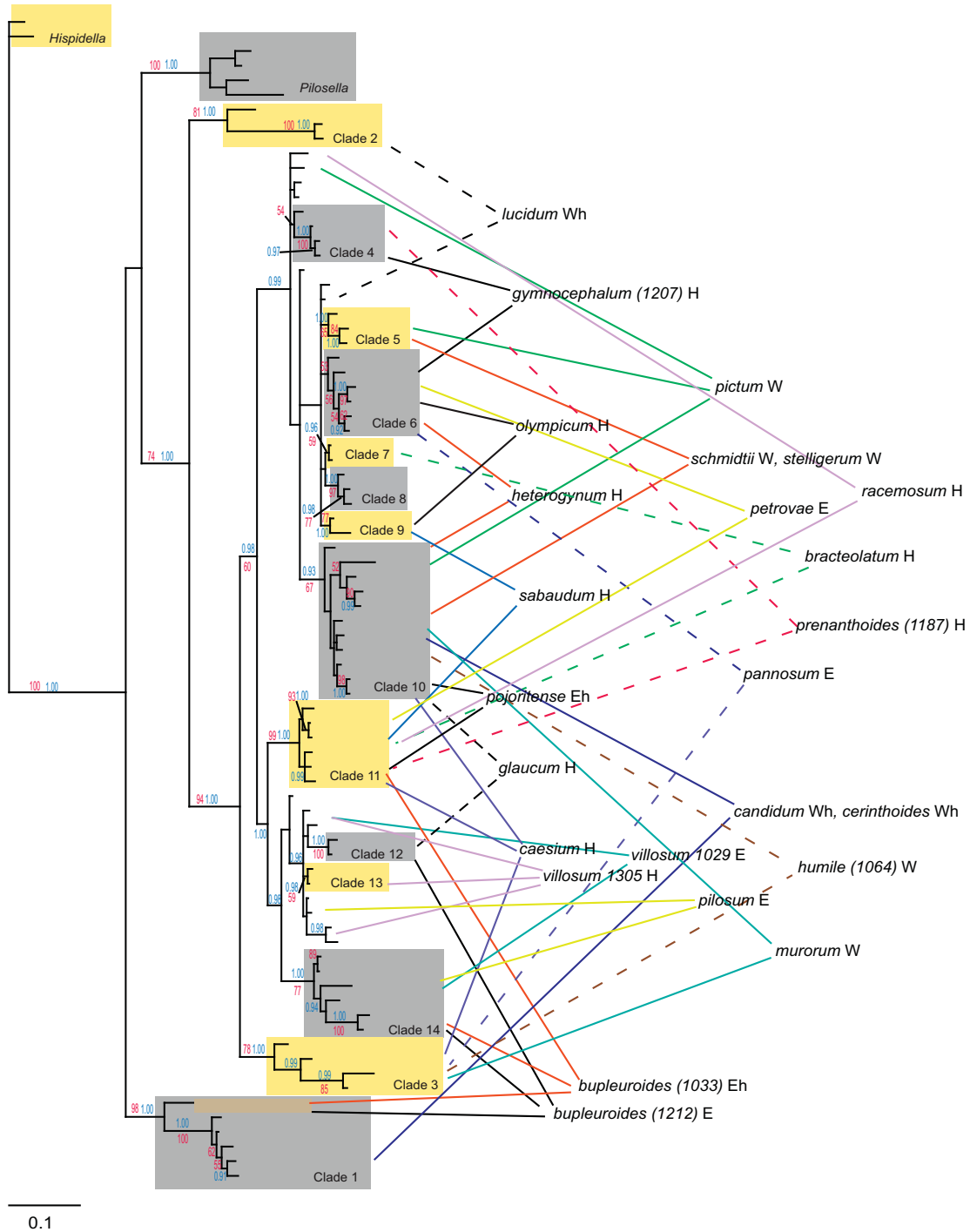

Hispidella
Pilosella
lucidum, intybaceum
Clade 2 W
canadense, gymnocephalum (1215), racemosum, pictum, virosum (1238) EHW
gymnocephalum (1207)
prenanthoides (1187, 1162, 1252)
Clade 4 H
pictum, stelligerum, schmidtii
Clade 5 W
gymnocephalum (1207), heterogynum, pannosum, petrovae, kittanae, olympicum
Clade 6 EH
bracteolatum
Clade 7 H
alpinum
Clade 8 E
sabaudum, olympicum
Clade 9 H
bifidum, pojoritense, laevigatum, stelligerum, murorum, pictum, tomentosum, schmidtii, caesium, cerinthoides, plumulosum, lachenalii, heterogynum, glaucum, humile (1064)
Clade 10 WHE
racemosum, bracteolatum, sabaudum, pojoritense, petrovae, naegelianum, bupleuroides (1033), virosum (virR), caesium, umbellatum, pilosum, eriophorum, prenanthoides (1187)
Clade 11 EH
pilosum, villosum (1029, 1305) EH
glaucum, porrifolium, bupleuroides (1212)
Clade 12 EH
villosum (1305)
Clade 13 H
villosum (1305), pilosum
EH
bupleuroides (1033, 1212), pilosum, sparsum, villosum (1029)
E
transylvanicum
w
Clade 14 EW
humile (1064,1188), pannosum, caesium, murorum
Clade 3 WHE
bupleuroides (1033, 1212)
1b E
cordifolium, amplexicaule, gymnocerinthe, mixtum,
cerinthoides, lawsonii, candidum,
ramondii, gouani, recoderi
1a WH
Clade 1
0.1

Fig. 3.


Hispidella
Pilosella
Clade 2
Clade 4
Clade 5
Clade 6
Clade 7
Clade 8
Clade 9
Clade 10
Clade 11
Clade 12
Clade 13
Clade 14
Clade 3
Clade 1
lucidum Wh
gymnocephalum (1207) H
pictum W
olympicum H
schmidtii W, stelligerum W
racemosum H
heterogynum H
petrovae E
bracteolatum H
sabaudum H
prenanthoides (1187) H
pannosum E
pojoritense Eh
glaucum H
candidum Wh, cerinthoides Wh
caesium H
villosum 1029 E
humile (1064) W
villosum 1305 H
pilosum E
murorum W
bupleuroides (1033) Eh
bupleuroides (1212) E
0.1

Conclusions

Despite much attention was paid to the taxonomy of *Hieracium* s. str. and the species diversity is investigated fairly well, the relationships among the species haven't been addressed so far. The presented thesis represents the first attempt to reconstruct the phylogenetic relationships within the subgenus using molecular approaches.

Due to the high number of described species, complete sampling was not possible. Instead, Zahn's basic species (diploids as well as polyploids) that are morphologically unique and well distinguishable and may therefore represent the major evolutionary lineages of the subgenus were sampled for the study. Altogether, 46 species represented by 58 accessions have been included.

Initially, the basic phylogenetic relationships within *Hieracium* s. str. were investigated based on the analyses of nrDNA ITS and ETS and the *trn*T-*trn*L region of the cpDNA. The variation in the ITS was very low and revealed no phylogenetic pattern. Therefore, the data for this marker are not presented in the thesis, but will be published later in a paper on the molecular evolution of this marker. The variation in the ETS was slightly higher, and two major phylogenetic lineages composed of taxa with Eastern or Western European origin were identified. Several well supported lineages were further resolved within the Eastern clade. Within the Western clade, only one monophyletic subgroup, formed by Pyrenean accessions, was found. Most of the western accessions had unresolved positions at the base of the clade. In contrast to the low interspecific variation, extremely high intraindividual polymorphism was observed in most of the accessions independent of the ploidy level, indicating a lack of sequence homogenization via concerted evolution. However, in several accessions, even polyploid ones, completely homogenized sequences were found. Interspecific hybrids were identified based on character state additivity of the sites showing intraindividual polymorphism. Hybrid origin was inferred for half of the analyzed accessions, including many diploid taxa. Further, hybrids were identified based on incongruence of the nuclear and chloroplast datasets. Hybridization among as well as within the major clades was evident. Altogether, 17 different parental combinations of these hybrids were described. However, this represents only a conservative estimate; the real extent of hybridization is most probably higher, but remained undetected due to the low interspecific variation that usually allowed only an assignment to subclades, but not to individual species. Several ETS ribotypes not matching any of the sequences identified in non-hybrid accessions were found in the genomes of several hybrids. Similarly, several polyploid hybrids with unique cpDNA haplotypes were found that did not correspond to those of any diploid taxon. These 'unknown' ribotypes and haplotypes may be attributed to unsampled or extinct variation.

Variation in genome size within the subgenus was previously reported in the diploma thesis of J. Zahradníček. Therefore, we investigated whether this variation reflects the revealed phylogeny or whether it rather correlates with other, mainly geographic (latitude, altitude) and ecological (light intensity, temperature) patterns. Phylogeny was found to be the most important factor explaining the patterns of variation in genome size. No significant correlation with ecology or geography was found.

The still relatively low variability in the ETS indicated that additional nuclear markers with higher variability had to be applied to issues that could not be solved in frame of the previous study. Therefore, novel low-copy nuclear markers were developed for further phylogenetic study of the subgenus. Candidate genes were selected based on the comparison of an expressed sequence tags (EST) library of *Lactuca* with Asteraceae sequences available from public sequence databases. Three low-copy nuclear genes – *squalene synthase* (*sqs*), *gamma-glutamylcysteine synthetase* (*gsh*1) and *glycine hydroxymethyltransferase (shmt*) – were included to a pilot study in which their variability, phylogenetic signal and evolutionary dynamics (pseudogenization or locus duplication) were investigated on a representative set of accessions from the subtribe Hieraciinae. All three genes showed to be single-copy within this sample set, and all had substantially higher variation than ITS. Furthermore, successful cross-amplification in members of seven other tribes of the Asteraceae was performed, demonstrating that these markers can be used beyond the Hieraciinae as well.

Finally, one of the newly developed low-copy nuclear markers (*sqs*) and an additional cpDNA region (the *trn*V-*ndh*C intergenic spacer) were analyzed on the same set of accessions as the ETS. Both, the combined cpDNA dataset (composed of the previously used *trn*T-*trn*L and the newly analyzed *trn*V-*ndh*C regions) and *sqs* were incongruent with the results based on the ETS with respect to the major phylogenetic groups. Instead, they rather reflected the subgroups recognized within these major groups. In several cases, the cpDNA haplogroups of the western accessions formed monophyletic groups with those identified for the eastern accessions. According to the *sqs*, 14 monophyletic groups were resolved within the subgenus, but the relationships among these were hardly supported at all. Furthermore, more than half of the accessions were polyphyletic and contained alleles from more than one clade. These included several accessions where hybrid origin had not been proposed based on ETS. Additional hybridization or incomplete lineage sorting or a combination of these could explain these patterns, but with the currently available data, neither of these hypotheses can be confirmed or rejected. Most of these were accessions with basal western position according to ETS, or members of the *H. porrifolium* and Balkan subgroups of the Eastern clade. On the other hand, several hybrid accessions were monophyletic with respect to their *sqs* alleles, probably due to genetic drift or selection (not necessarily acting directly on *sqs*, but on some locus in close physical proximity), eliminating one of the parental alleles from the hybrid genomes. Several clades composed exclusively of alleles isolated from hybrid accessions were identified. In these, alleles mainly from accessions that contained the ‘unknown’ ribotypes were placed. Similarly, unique cpDNA haplotypes were found in hybrids or polyploids.

The low variation of the ETS and the absence of deeper phylogenetic relationships in the *sqs* and cpDNA data may be explained by relatively recent speciation of *Hieracium* s. str. Extensive hybridization was previously anticipated for the subgenus, concerning especially the intermediate species. This study involved solely basic species, for which hybrid origin was not previously anticipated based on their morphology. Initially, we expected that some of the polyploids might be hybrids. However, the extent of hybridization revealed by this study by far exceeded all expectations. Particularly surprising is the fact that even several diploid accessions

were found to be of hybrid origin. Therefore, interspecific hybridization and allopolyploidy can be considered as the major driving forces of speciation in the subgenus.

Unique genetic material was found in hybrid accessions by all three markers. With high probability it could represent extinct diploid parental lineages of these accessions, indicating that the diversity within the subgenus was much higher in the past than is observed nowadays. From this perspective, even those taxa that were considered as major evolutionary units rather represent intermediate derivates of this immense past variation.

Professional Curriculum Vitae

Personal data

Karol Krak
25.2.1981, Rožňava, Slovakia

Education

since 2004	Ph.D. study in Botany, Department of Botany, Faculty of Science, Charles University in Prague
1999–2004	MSc. study in Botany and Plant Physiology, Faculty of Science, P. J. Šafárik University, Košice, Slovakia

Employment

since 3/2009	Researcher, Depeartment of Genetic ecology, Institute of Botany ASCR, Průhonice, Czech Republic
7/2008 – 2/2009	Assistant for forensic DNA analyses, Forensic DNA Service s. r. o., Prague, Czech Republic
10/2004 – 9/2008	PhD. student, Department of Taxonomy, Institute of Botany ASCR, Průhonice, Czech Republic

Research interest

- Molecular marker development (qPCR assays, SSR markers, low-copy genes)
- Polyploid and hybrid speciation
- Quantitative dynamics of AM fungal communities

SCI Publications

Krak K, Janoušková M, Caklová P., Vosátka M., Štorchová H. (2012): Intraradical dynamics of two coexisting isolates of the arbuscular mycorrhizal fungus Glomus intaradices sensu lato as estimated by real-time PCR in mitochondrial DNA. *Appl Environ Microb* 78: 3630–3637.

Krak K, Álavarez 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. *Am J Bot* e74–e77.

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.

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.

Krak K, Mráz P (2009): Trichomes in the tribe Lactuceae (Asteraceae) - taxonomic implications. *Biologia* 63: 616–630.

Non-SCI Publications

Krak K (2011): Vyuužití sekvencí DNA při studiu evoluce rostlin. *Zprávy České Botanické Společnosti*, *Praha* 46, Mater. 25: 95–125.

Conference reports

Krak K, Chrtek J, Caklová P, Fehrer J. (2012): Reconstruction of phylogenetic relationships in a highly reticulate group with deep coalescence and recent speciation (*Hieracium*, Asteraceae). International Conference on Polyploidy, Hybridization and Biodiversity, 7–10 May 2012, Pruhonice, Czech Republic (oral presentation).

Khodlová Z, Trávníček P, **Krak K**, Krejčíková J (2012): Polyploid Speciation in Anthoxanthum: Far beyond a Simple Solution. International Conference on Polyploidy, Hybridization and Biodiversity, 7–10 May 2012, Pruhonice, Czech Republic (poster).

Šmerda J, **Krak K**, Fehrer J, Bureš P, Zedek F, Zahradníček J, Chrtek J (2012): TY3-gypsy Copy Number is Correlated with Genome Size in the Subtribe *Hieraciinae* (Lactuceae, Asteraceae). International Conference on Polyploidy, Hybridization and Biodiversity, 7–10 May 2012, Pruhonice, Czech Republic (poster).

Mandák B, Trávníček P, Paštová L, **Krak K** (2012): Is hybridization involved in the evolution of the *Chenopodium album* aggregate? Molecular Ecology, 4–7 February 2012, Vienna, Austria (poster).

Douda J, Doudová J, Drašnarová A, Hadincová V, Jahodová Š, **Krak K**, Zákravský P, Mandák B (2012): Analysis of DNA variation and fossil records in *Alnus glutinosa* and *A*. *incana*: changes in distribution ranges over time. Molecular Ecology, 4–7 February 2012, Vienna, Austria (poster).

Janoušková M, **Krak K**, Štorchová H, Caklová P, Vosátka M (2011): Discrimination and quantification of two *Glomus intraradices* isolates in plant roots by Real-Time PCR assays in mitochondrial DNA. Ecology of Soil Microorganims. Microbes as Important Drivers of soil processes, 27 April – 1 May 2011, Prague, Czech Republic (poster)

Fehrer J, **Krak K**, Chrtek J (2011): When low copy nuclear genes are unable to produce species trees despite large genetic variation: Lessons from a highly reticulate plant group (*Hieracium*). XVIII International Botanical Congress, 23–30 July 2011, Melbourne, Australia (oral presentation J. Fehrer).

Paštová L, **Krak K**, Mandák B (2011): Evolution of the *Chenopodium album* group. XVIII International Botanical Congress, 23–30 July 2011, Melbourne, Australia (poster).

Dvořáková K, Urfus T, **Krak K**. Vít P (2010): Hybridization and microevolutionary relationships among Central European Diphasiastrum species. 19th International Symposium Biodiversity and Evolutionary Biology of the German Botanical Society (DBG), 16–19 September 2010, Vienna, Austria (poster).

Janoušková M, **Krak K**, Štorchová H, Caklová P, Müller K, Vosátka M (2010): Application of Real-Time PCR to studying dynamics of AM fungal communities after inoculation. Sixth National Symposium and Third Iberoamerican Meeting on Mycorrhizal Symbiosis, 6–10 September 2010, Tlaxcala, Mexico (poster).

Krak K, Fehrer J, Caklová P, Chrtek J (2010): Three new low-copy nuclear markers for low level systematic studies in the Asteraceae – development and preliminary results of the phylogenetic analysis of *Hieracium* subgen. *Hieracium* and *Hieracium* subgen. *Chionoracium.* 11th International Hieracium workshop, 27–29 January 2010, Pruhonice, Czech Republic (oral presentation).

Fehrer J, **Krak K**, Chrtek J (2009): Extensive ancient hybridization in sexual and apomictic hawkweeds (Hieracium, Asteraceae) and evidence for extinct diversity. International Conference on Polyploidy, Hybridization and Biodiversity, 17–21 May 2009, San Malo, France (oral presentation J. Fehrer)

Krak K, Fehrer J, Chrtek J (2008): Reticulation, speciation, and patterns of molecular evolution in *Hieracium* (Asteraceae). Systematics 2008, Goettingen Germany (oral presentation)

Grant projects

Project leader

2006 Development of novel low-copy markers for asteraceae (EU SYNTHESYS ES-TAF-1365)

Researcher

2011-2014 Postglacial colonization of Alnus glutinosa and Alnus incana: analysis of DNA variation and fossil records (GA CR P504/11/0402, project leader B. Mandak)

2010-2013 Phylogeny of subtribe Hieraciinae (Asteraceae) - a model example of contrasting evolutionary strategies in closely related lineages (GA CR GAP506/10/1363, project leader J. Chrtek)

2009-2012 Coexistence of native and inoculated arbuscular mycorrhizal fungi in the roots of host plants (GA CR GA526/09/0838, project leader M. Vosátka)

2009-2012 The role of hybridization in plant evolution – In situ hybridization technique (GA CR GA206/09/1126, project leader B Mandak)

2005-2007 Molecular phylogeny and evolutionary trends in *Hieracium* (Asteraceae, Lactuceae) (GA CR GA206/05/0657, project leader J. Chrtek)

Supplementary files to Paper 1, Fehrer et al.: **Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (*Hieracium*, Lactuceae, Asteraceae): disentangling phylogenetic signal, reticulation, and noise.**

Only those additional files are included to this section, that are essential for understanding the subsequent chapters of the thesis. The complete additional information of the article is available at: http://www.biomedcentral.com/1471-2148/9/239

Additional file 1: Origins of individual accessions

In the following, we concentrate on species as taxonomic entities. Each was represented by only one or a few samples, and therefore, the discussion of the molecular data applies only to the particular accessions analyzed. In some cases, we could show particular taxa to be inhomogeneous entities – which is interesting in itself – and this may apply to other species as well if a larger number of accessions will be investigated. Nevertheless, a preliminary assessment of accessions in relation to the taxon as a whole provides useful insights and generates working hypotheses that can guide future research strategies and sampling schemes. Source information for all accessions and outgroup taxa is given in a table at the end of this file.

***H. humile* Jacq.**
The species is morphologically more or less uniform and has its major distribution in the Western Alps. Two accessions, a triploid and a tetraploid one, were analyzed. *ETS* placed both accessions into the 'Western' clade. Together with one accession of *H. pilosum*, *H. humile* was the species with the lowest number of polymorphic sites (1–2) despite the polyploidy of the analyzed accessions. It is possibly an old species; morphologically, it cannot be derived from any recent species. The chloroplast DNA of both accessions was identical to that of other 'Western' clade species. The low amount of intra-individual polymorphism suggests autopolyploid origin from extinct diploid accessions of the same taxon, because diploid populations are unknown.

***H. tomentosum* L.**
This species occurs in the Western Alps and also belonged to the 'Western' clade. A diploid accession shared one polymorphic site with the majority of 'Western' species, reflecting a substitution characteristic for the 'Pyrenean' lineage. A similar structure of the indumentum shared between *H. tomentosum* and some Pyrenean species could also be indicative of a loose relation. Most other polymorphisms were unique. Its cpDNA haplotype was 'Western'.

***H. murorum* L.**
This species is distributed all over Europe. Despite its being one of the morphologically most polymorphic taxa in *Hieracium*, intraspecific differences are still much smaller than, for example, in *H. prenanthoides* (see below). Although this is a polyploid taxon, the analyzed triploid accession showed a relatively low number of *ETS* polymorphisms suggesting autopolyploid origin. According to *ETS* and cpDNA, it belonged to the 'Western' clade. Analysis of additional accessions might reveal genetic heterogeneity of this species.

***H. bifidum* Kit.**
This is a Central and North European, predominately mountainous calciphilous species with morphological similarities to *H. murorum* and *H. stelligerum*. The analyzed triploid accession from the Slovakian mountains had a moderate amount of polymorphic sites and was of western origin according to nuclear and chloroplast DNA.

***H. schmidtii* Tausch**
The distribution of this Mediterranean-Atlantic species extends in Western Europe far to the north; the further to the east, the more southern the distribution becomes, reaching Lebanon. The analyzed Czech accession (triploid) belonged to the 'Western' clade according to *ETS*, cpDNA, and DNA content. Intra-individual polymorphisms were either unique or homoplasious, i.e., shared with different species without any apparent pattern. The species has special ecological demands (canyon-like valleys, rock-exposed if local humidity is high) and a fragmented distribution area.

***H. pictum* Pers.**
This polyploid species occurs in the Western Alps (Southwestern Switzerland, France, Northwestern Italy), rarely also in Central Italy. It is morphologically more or less uniform, and resembles *H. bifidum* and an intermediate species between *H. bifidum* and *H. schmidtii* (*H. hypochoeroides*, syn. *H. wiesbaurianum*). *ETS*

and cpDNA of both triploid accessions corresponded to a 'Western' origin. Despite being polyploid, only a small number of polymorphic sites (1 and 4, respectively) occurred which differed between the accessions. Autopolyploid origin from extinct diploids of the same taxon is probable.

***H. stelligerum* Froel.**

It is morphologically rather uniform, only indumentum and leaf shape are highly variable. Only ca 15 populations in a small area of Southern France are known, highlighting the relict character of this diploid species. It co-occurs with *H. schmidtii*, *H. murorum*, and *H. bifidum* whose analyzed accessions also belonged to the 'Western' clade. Its ecology is similar to *H. bifidum* with which it shared two out of three polymorphic sites in *ETS* and a 'Western' cpDNA haplotype. The taxon could be ancestral to polyploid *H. bifidum*.

***H. transylvanicum* Heuff.**

This is a frequent species in the Eastern and Southern Carpathians and also occurs in the Eastern Alps and the Northern Balkans. It is more or less uniform morphologically, some plasticity concerns only the growth form. Despite its eastern distribution, *ETS* sequences placed both diploid accessions into the 'Western' clade. The cpDNA was unique; the DNA content was far above the usual values for 'Western' species. Accessions 1077 and tra.Boa differed in their polymorphic sites; tra.Boa shared a single one with several hybrids at a homoplasious position, 1077 shared one with some species from the 'Western' clade, one with Pyrenean taxa, and two reflected substitutions of accession tra.Boa. Two alternative scenarios for the origin of this species are proposed: 1) The species has an originally eastern origin as suggested by its current distribution and DNA content and was introgressed a long time ago by species from the 'Western' clade some of which are or were widespread. *ETS* sequences then became completely homogenized towards the 'Western' type. In this case, the introgression by 'Western' species should be rather ancient given complete homogenization of *ETS* towards this type plus additional unique substitutions. 2) The species originated in Western Europe, spread towards the east, the original populations became extinct (probably during the Ice Ages), and only the eastern populations survived in eastern glacial refugia like the Carpathian basin. In this case, the high DNA content must have other reasons than reflecting phylogenetic signal. Shared polymorphisms with different 'Western' clade species in accession 1077 could then be interpreted as retained ancestral polymorphism and, together with a unique cpDNA, imply an old, truely western origin. The first scenario may be more probable, because *H. transylvanicum* is rather frequent in Eastern Europe, does not have any relict populations in Western Europe, and apparently did not leave any traces in intermediate apomictic taxa.

***H. lucidum* Guss.**

This critically endangered species (*sensu stricto*) is only known from a single relict population in Northwestern Sicily. The diploid accession belonged to the 'Western' clade and showed a high level of polymorphic sites; more than half of them were unique. Four intra-individual polymorphisms were indicative of the 'unknown Western 1' ribotype which was present in low amounts. The chloroplast haplotype was identical to that of two accessions of *H. prenanthoides* (1252 and 1187). Accession 1252 was diploid and restricted to the Southwestern Alps supporting a western origin also for *H. lucidum*. The data are concordant with ancient hybridization between a 'Western' species (but not extant *H. prenanthoides*) and an unsampled or extinct species of western origin. Morphologically, *H. lucidum* has some similarity with *H. racemosum* which could be caused by their sharing an 'unknown Western 1' genome. Also, *H. crinitum*, a polyploid subspecies of *H. racemosum*, is rather frequent in Sicily and may have introgressed *H. lucidum* in former times. Nowadays, the only remaining population of *H. lucidum* is geographically isolated from other *Hieracium* species.

***H. lawsonii* Vill.**

This species is from the Pyrenees and the Western Alps. A triploid plant shared most of its few polymorphic sites with other Pyrenean species and, according to *ETS* and cpDNA, belonged to the 'Pyrenean' subclade.

***H. ramondii* Griseb.**

A triploid of this Pyrenean taxon shared two out of three polymorphisms with other Pyrenean species. It also belonged to the 'Pyrenean' subclade, but had another 'Pyrenean' cpDNA haplotype than *H. lawsonii*.

***H. recoderi* De Retz**
This Pyrenean species shared four polymorphic sites with other species from the same clade, three of them mostly with other Pyrenean taxa. *ETS* attributed it to the 'Pyrenean' clade of which it was the only diploid representative; its cpDNA haplotype was the same as in *H. ramondii*.

***H. cordifolium* Lapeyr.**
This diploid Pyrenean species shared a few polymorphic sites with other 'Pyrenean' and 'Western' clade taxa, other polymorphisms were unique. It is probably a hybrid between 'Pyrenean' and undifferentiated 'Western' taxa. Its cpDNA haplotype was derived from one of the 'Pyrenean' haplotypes.

***H. gymnocerinthe* Arv.-Touv. & G. Gaut.**
Like most other Pyrenean species, the analyzed triploid accession shared polymorphic sites with 'Pyrenean' and some other 'Western' clade taxa, one site with all accessions of *H. prenanthoides*. Two polymorphisms at homoplasious positions occurred also in different 'Eastern' species. The cpDNA haplotype was one of the 'Pyrenean' variants. We consider it as a hybrid between the 'Pyrenean' lineage and other 'Western' taxa.

***H. candidum* Scheele**
Like the previous species, the triploid accession shared polymorphisms with other 'Pyrenean' taxa, some with basal 'Western' clade species, a few probably homoplasious ones also with different 'Eastern' taxa. One polymorphic site was exclusively shared with *H. amplexicaule*. Its cpDNA haplotype was one of the two 'Pyrenean' variants. Similar hybrid origin as for the preceding species is suggested.

***H. cerinthoides* L.**
The degree of morphological polymorphism of this Pyrenean species is comparable with that of other Pyrenean taxa. All six polymorphic sites in the *ETS* of the triploid accession were shared with 'Western' clade taxa, predominantly with other Pyrenean species. Its cpDNA haplotype corresponded to one of the 'Pyrenean' variants. The accession had most likely hybrid origin involving a 'Western' clade taxon and one belonging to the 'Pyrenean' lineage.

***H. gouani* Arv.-Touv.**
This species occurs at the foothills of the Pyrenees in refugial localities and has relict character. It belongs to the *H. cerinthoides* group. The only co-occurring *Hieracium* species is *H. sabaudum*. The analyzed diploid accession was a hybrid between the 'Eastern' and 'Western' clade as it showed additive sites at all positions differing between the clades. In addition, it shared six polymorphisms with different Pyrenean species indicating that a taxon from the 'Pyrenean' subclade hybridized with an 'Eastern' clade species (but probably not with *H. sabaudum*, see below). No particular 'Eastern' clade taxon could be identified. The 'Western' sequence type was by far dominating (about 90% across the sites) which fits to its low genome size and its chloroplast DNA of Pyrenean origin.

***H. pilosum* Schleich. ex Froel.**
This is a morphologically variable Central European high mountain taxon. In morphology and distribution, it resembles *H. villosum*. Of two triploid accessions from the same locality in the Slovenian Alps, one had 'Eastern' origin, the other was a hybrid between the two major clades. The 'pure' species had only a single polymorphic site which it shared with other species of the same 'Eastern' ('*H. porrifolium*') subclade. The hybrid accession showed strong additivity of positions differing between the major clades. In addition, six polymorphisms were shared with different species of the '*H. porrifolium*' subclade, mostly with *H. villosum*. Four further polymorphisms – consistently shared with several other hybrids – reflected contribution of the 'unknown Western 2' lineage. The non-hybrid accession had a unique substitution (A, consensus G) reflected by a polymorphism (r) in the hybrid accession. Thus, the hybrid most likely was a cross between 'true' *H. pilosum* and an unknown or extinct ('unknown Western 2') species. Both accessions had the same cpDNA which was shared with other members of the '*H. porrifolium*' subclade.

***H. villosum* Jacq.**

In morphology and distribution, it is very similar to *H. pilosum*, but has a less dense involucrum with distinctly broader outer bracts. The species is morphologically variable. A tetraploid accession from Slovakia had few polymorphic sites most of which were shared with other species of the '*H. porrifolium*' subclade to which also its cpDNA corresponded. It could be an autotetraploid derived from extinct diploid or triploid ancestors. A triploid accession from France showed all additive sites typical for interclade hybrids with the 'Eastern' sequence variant strongly dominating. Correspondingly, its DNA content matched that of 'Eastern' species which may indicate backcrossing or unequal proportions of genomes obtained from the parental accessions. Other polymorphisms and cpDNA were shared with species from the '*H. porrifolium*' subclade. This triploid accession had hybrid origin involving a parent from that subclade and an 'unknown Western 2' taxon as pollen donor (see also *H. pilosum*). The species had multiple origin which might be suspected already from its morphological variability.

***H. bupleuroides* C.C. Gmel.**

Like the previous two taxa, this is also a Central European mountain species. Morphologically, it is most similar to *H. porrifolium*. Out of two triploid accessions from Slovakia, one (1212) had few polymorphisms all of which were shared with species of the '*H. porrifolium*' subclade, two of them with *H. porrifolium* itself. Either the species was derived from that diploid taxon or they share a common ancestor. Its cpDNA corresponded to other species of the '*H. porrifolium*' subclade. The second accession (1033) somewhat resembling *H. umbellatum* morphologically was actually introgressed by this species according to seven additive polymorphisms and cpDNA. It had also a higher DNA content (accessions of the '*H. umbellatum*'clade had the highest values among 'Eastern' taxa) than the 'pure' accession (1212) with which it shared five prominent polymorphisms. Additionally, two adjacent polymorphic sites were triple peaks that were additive between polymorphic sites of 1212 and different character states specific for all '*H. umbellatum*' clade taxa (A/T plus G, and T/G plus A). Thus, this accession of *H. bupleuroides* was introgressed by *H. umbellatum*, the only widespread diploid species, which acted as the maternal parent. With respect to these patterns and to its morphology, accession 1033 was most likely misidentified and probably corresponds to the intermediate species *H. virgicaule* Nägeli & Peter, which is supposed to have the inferred parentage.

***H. porrifolium* L.**

This diploid taxon is morphologically rather uniform and occurs predominantly in the Southeastern Alps which are considered as a glacial refuge area. Morphologically, it mostly resembles *H. bupleuroides*. The number of polymorphic sites was small. Most were homoplasious, but two were shared with both accessions of *H. bupleuroides*. The species clustered with 'pure' accessions of the previous three taxa ('*H. porrifolium*' clade) with which it shared morphological, distributional and ecological features, and a particular cpDNA haplotype.

***H. umbellatum* L.**

This is the *Hieracium* species with the largest distribution (Eurasia and North America). It is a tall-growing perennial species, and is predominantly diploid (triploid chromosome counts occur, but are much less frequent). Two diploid accessions had few polymorphic sites most of which were unique, one was shared among the two accessions, with *H. eriophorum*, *H. canadense*, and some hybrids involving the '*H. umbellatum*' clade. Accession um.AM.1 shared two further polymorphisms with *H. eriophorum* and *H. pojoritense*. Together with the following three species, *H. umbellatum* belonged to a well-supported subclade of eastern origin ('*H. umbellatum*' clade) which was also characterized by a particular chloroplast haplotype.

***H. eriophorum* St.-Amans**

This diploid endangered species occurs only along the Atlantic coast in Southwest France. Two accessions shared three prominent species-specific polymorphisms and additionally the above-mentioned ones with *H. umbellatum* to which also the cpDNA corresponded. The species is probably a young offspring of *H. umbellatum*. The particular morphology of *H. eriophorum* could be interpreted as a local adaptation to sand dunes along the sea coast.

***H. canadense* Michx.**
This species of subgenus *Hieracium* exceptionally occurs in North America. Some authors consider it as a subspecies of *H. umbellatum*, a view that does not contradict our data. The triploid accession shared one polymorphic site and its cpDNA haplotype with *H. umbellatum* (and *H. eriophorum*).

***H. virosum* Pall.**
This is a polyploid tall-growing perennial with Eastern European and Siberian distribution and some morphological similarities to *H. sabaudum*. Two triploid accessions had five polymorphic sites of which they shared only one. At that site, the '*H. umbellatum*' clade differed from all non-hybrid taxa which indicates that either *H. virosum* was introgressed by another 'Eastern' species (but no other evidence was found to support this possibility) or that it – or more likely its putative diploid predecessor – could be the oldest species of the subclade in which the polymorphism arose of which the derived character state became fixed in the whole subclade. The Siberian accession (1238) had '*H. umbellatum*' type cpDNA while the more Western Russian accession (vir.1) had a unique variant differing by two mutations.

***H. alpinum* L.**
This is a locally widespread mountain species with unique morphology of which diploid populations occur nowadays only in the Eastern and Southern Carpathians. Two diploid accessions had prominent, but few and mostly unique polymorphisms. The species formed a lineage of its own within the 'Eastern' clade. The chloroplast DNA was also unique and shared only with hybrid taxa with a captured 'original' *H. alpinum* chloroplast.

***H. pojoritense* Woł.**
This is a well circumscribed diploid taxon growing in calcareous crevices in the Eastern Carpathians. Morphology suggests some influence of *H. umbellatum*. Zahn [3] considered it as an 'intermediate' species and placed it between *H. sparsum* and *H. racemosum*. The analyzed accession had a very high number of polymorphic sites, many of which were unique, few were shared with species of the '*H. umbellatum*' clade and some other taxa, but seven reflected synapomorphic substitutions of that clade, and four mirrored *H. alpinum*-specific synapomorphies. *Hieracium umbellatum* (or one of its hybrids, for example *H. racemosum*) is the most likely paternal parent of this hybrid. The *H. alpinum ETS* ribotype was dominating across all positions (about 80% of the total signal), and the chloroplast DNA also corresponded to that species. However, *H. pojoritense* does not show any morphological influence of *H. alpinum*. Also, experimental hybrids between *H. alpinum* and *H. umbellatum* have a different phenotype (P. Mráz, pers. comm.). Taken together with the diploid condition of *H. pojoritense*, the relatively large number of unique polymorphic sites and the species' distribution in a well-known glacial refuge area, it might be concluded that it is an old taxon that either for some reason does not resemble *H. alpinum* morphologically, or that it has originated from an extinct species closely related to, but morphologically different from recent *H. alpinum*. In this context, the captured '*H. alpinum*' cpDNA of *H. sparsum* could be a hint that Zahn's opinion was quite accurate.

***H. petrovae* Vladimirov & Szeląg**
This is a rare, recently described diploid species of section *Pannosa* from the Balkans. Expectedly, it fell into the 'Eastern' clade. It had few polymorphisms; the two most prominent ones were shared only with polyploid *H. pannosum*. Its chloroplast haplotype was the same as that of *H. kittanae* and *H. pannosum*, other Balkan species.

***H. pannosum* Boiss.**
This is a widespread polyploid taxon occurring from the Balkans to Anatolia. It belonged to the 'Eastern' clade and had a rather high number of polymorphisms. Nine out of twelve shared polymorphisms occurred also in other Balkan species, two of these were exclusively shared with *H. petrovae*. An additional polymorphism reflected an autapomorphic substitution of *H. petrovae* with which it also shared its cpDNA haplotype. Thus, the taxon could be a polyploid derivative of *H. petrovae*.

***H. kittanae* Vladimirov**

This is another recently described diploid species. Its has unique morphology, is restricted to a few localities in Bulgaria, and has relict character. Nowadays, few *Hieracia* co-occur with this taxon, but in former times there may have been more species in this area. It formed a separate lineage within the 'Eastern' clade and had the highest number of polymorphic sites of any non-hybrid species. Most were shared with other species from the Balkans. They had apparently accumulated on various different rDNA copies according to cloned sequences (Additional file 2). Its cpDNA corresponded to that of *H. pannosum* and *H. petrovae*.

***H. naegelianum* Pančić**

The species occurs in the Balkan Peninsula and in the Abruzzy Mountains in Central Italy, mostly in refugial areas. It has a rather unique morphology and forms elongated, stolon-like underground rhizoms, an unusual trait in *Hieracium* s.str. The analyzed triploid accession fell into the 'Eastern' clade and had a rather high number of polymorphic sites, many of them unique. Half of the shared polymorphisms were inconclusive as they matched different (unrelated) taxa and hybrids; the other half were shared with a few species and hybrids from the Balkans. Its chloroplast haplotype was also unique. Its morphology, *ETS* polymorphisms, and cpDNA support a rather isolated position within the 'Eastern' clade. While no evidence of introgression from a 'Western' species is apparent from the molecular data, its occurrence in Italian glacial refuges could be indicative of past contacts and introgression from which only an unusually small genome size is left.

***H. sparsum* Friv.**

Hieracium sparsum s.l. occurs in Southeastern Europe, Northern Anatolia and the Caucasus, and comprises morphologically rather different growth forms. In contrast, *H. sparsum* s.str. is diploid, morphologically quite uniform, and is missing in the Caucasus. Two analyzed diploid accessions fell into the 'Eastern' clade and shared two otherwise unique polymorphisms. Apart from these, the *ETS* sequences lacked aut- or syn-apomorphic character states which might indicate that it is a rather old taxon which is also reflected by its position at the base of the 'Eastern' polytomy in the phylogenetic tree. Eight polymorphic sites were shared with other species and hybrids from the Balkans, one with a Romanian accession of *H. alpinum* (alp.Boa.2). Surprisingly, its chloroplast DNA was apparently derived from an '*H. alpinum*' haplotype. The diploid's recent distribution area does not overlap with that of *H. alpinum*. However, polyploid *H. sparsum* (s.l.) is subalpine and known to form hybrid taxa with *H. alpinum*. As morphology of the analyzed accessions does not show any evidence for an introgression of *H. alpinum*, we assume either a chloroplast capture event that has happened very early in the history of the species or an origin from an unknown or extinct species with *H. alpinum* cpDNA, but different morphology (maybe analogous to the situation in *H. pojoritense*).

***H. amplexicaule* L.**

This mountain species is rather widespread in the Alps, in South and Southwest Europe, and North Africa (with numerous secondary occurrences outside its native range). It is morphologically variable with characters connecting it to Pyrenean taxa, but with a much wider ecological amplitude. It forms many intermediates with other taxa. The analyzed triploid accession from Austria showed additivity at positions differing between 'Eastern' and 'Western' clade species. The 'Western' variant was predominant and constituted about 60–70% of the total signal. In addition, it shared six polymorphic sites with Pyrenean taxa indicating that it is a product of a hybridization between a taxon of the 'Pyrenean' lineage and an 'Eastern' clade species. Several polymorphisms were shared with different 'Eastern' species so that its exact origin could not be inferred. Its chloroplast haplotype was derived from one of the 'Pyrenean' variants.

***H. caesium* (Fr.) Fr.**

This is a locally rather widespread, predominantly calciphilous mountain species with mainly North and Central European distribution. The analyzed tetraploid accession from Sweden (microspecies *H. plumbeum* Blytt et Fr.) had hybrid origin and involved a member of the '*H. umbellatum*' group and a 'Western' clade species. The 'Western' ribotype was dominating (about 80–90% relative contribution according to direct sequencing; only one out of six clones represented the '*H. umbellatum*' variant). One polymorphim was exclusively shared with *H. umbellatum* um.AM.1 which is a common genotype according to its inferred

contribution to species and hybrids across a large geographic area. A rare 1 bp-deletion was shared with *H. villosum* 1029. As the only clone showing this feature belonged to the 'Western' ribotype while *H. villosum* was an 'Eastern' clade taxon, this was probably a parallel mutation. The 'Western' parent could not be resolved: only one polymorphic position was shared with many 'Western' species. Morphologically, *H. caesium* resembles *H. bifidum* which might have been involved in its origin. It had a chloroplast of 'Western' origin. Its low 1Cx genome size is indicative of larger amounts of a 'Western' genome.

***H. mixtum* Froel.**

This Pyrenean-Cantabrian mountainous-alpine species is occasionally used as an ornamental plant and has a neophytic occurrence in Germany from which we analyzed one accession. According to published chromosome counts, this species is triploid. Its *ETS* had a relatively high number of small polymorphisms. Alternative character states showed shared polymorphisms for interclade hybrids at all 15 positions. However, neither 'Eastern' nor 'Western' clade-specific character states were dominating, but a mixture of these. This indicates that subsequent to hybridization, the *ETS* arrays became to a large degree homogenized towards a novel unique hybrid sequence. No particular parental taxa could be identified. Its cpDNA was also unique.

***H. prenanthoides* Vill.**

Diploid *H. prenanthoides* is a morphologically rather uniform mountaneous taxon of the Southwestern Alps. As diploids are only known from this area, they can be considered as relict populations that may have survived in this glacial refuge area. Their distribution coincides with a western origin of this species as suggested by the *ETS* data. Polyploids of this species are morphologically variable and occur in subalpine habitats throughout the European montains, the Caucasus and neighboring areas, Central Asia and Siberia; in Northeastern Europe and Siberia also in the lowlands. Their distribution area is large and fragmented. Polyploids form many intermediate types with other species. Zahn [3] supposed morphological influences of section *Cerinthoidea* (mostly Pyrenean taxa) and also similarities to *H. umbellatum*. The three accessions analyzed were similar in sharing two unique polymorphic sites and a strong interclade hybrid signature at the 3'-end of the *ETS* region, and in containing predominantly the 'Western' ribotype. Their cpDNA also belonged to the 'Western' variant or was derived from it. The dominating 'Western' *ETS* ribotype and the geographic distribution of the diploid could hint at an early introgression by a diploid species of eastern origin (or a hybrid contributing an 'Eastern' genome) occurring in the Southwestern Alps. Backcrossing towards *H. prenanthoides* (because of low 'Western' genome size) and a recombination event may have left a trace of the 'Eastern' introgressant's genome in a predominantly 'Western' *ETS* background. This signature was still present in the triploid accessions that arose later from the diploids of this species by further hybridization: The triploid accession 1161 showed additionally the 'unknown Western 1' *ETS* variant, suggesting introgression by a further, potentially extinct taxon. The triploid 1187 lacked this variant, but instead showed subsequent introgression by an 'Eastern' clade taxon. More specifically, part of the additional polymorphisms were shared with the '*H. umbellatum*' group, and cloned recombinant sequences suggested that the other part was lost from the genome (Additional file 2: Patterns of *ETS* recombination). Thus, *H. prenanthoides* seems to have a cryptic history of reticulation between different clades which in case of the triploids involved at least two subsequent hybridization events. The morphological variability of polyploids of this species may be due to recurrent origin from diploid *H. prenanthoides* hybridizing with different other taxa. Whether the initial 'Eastern' introgressant also belonged to the '*H. umbellatum*' clade cannot be decided, because the 3'-end bearing the 'Eastern' hybrid signature contains no substitutions distinguishing between the '*H. umbellatum*' subclade and other 'Eastern' taxa.

***H. lachenalii* Suter**

The species is widespread across all of Europe and Western Asia, morphologically very polymorphic and mostly lacking in specific characters. It fills the morphological space between *H. murorum* and *H. laevigatum*. According to *ETS* sequence, the triploid accession belonged to the 'Western' clade, but its cpDNA matched that of the '*H. umbellatum*' group to which the species has some morphological similarity. However, not a single polymorphic site specific for that group was present in its *ETS* sequence. This suggests

ancient introgression by *H. umbellatum* (or a sexual hybrid carrying an '*H. umbellatum*' genome) with either almost complete homogenization of *ETS* towards the 'Western' clade (only two polymorphic positions showed a small amount of general 'Eastern' clade character states) or repeated backcrossing to a 'Western' species prior to polyploidization. Its higher DNA content – about 5% above the usual values for 'Western' clade species – may be a consequence of that introgression. With respect to its morphological variability and poor taxonomic circumscription, it is likely that *H. lachenalii* is composed of types with multiple origin from similar species combinations.

***H. laevigatum* Willd.**

Like *H. lachenalii*, this is a widespread and highly polymorphic species, but with prominent morphological similarity to *H. umbellatum*. All 15 sites differing between clades were additive in the analyzed triploid accession. In addition, it shared seven polymorphisms or substitutions with the '*H. umbellatum*' group. Its chloroplast DNA also belonged to that group. Thus, the accession had hybrid origin involving a genetically undifferentiated 'Western' clade species and a member of the '*H. umbellatum*' group.

***H. racemosum* Waldst. et Kit. ex Willd.**

This is a very polymorphic species, some accessions resembling *H. lucidum*, some *H. sabaudum*. With the latter species, it forms a morphological continuum, but has a more southern distribution (East-Submediterranean to Southern Silesia, further east in northern areas, overlapping with *H. virosum*). The analyzed triploid accession showed additive patterns between 'Eastern' and 'Western' clade taxa with the 'Eastern' variant slightly dominating. In addition, eight character states or polymorphisms as well as its chloroplast haplotype were shared with the '*H. umbellatum*' group. There was also a strong contribution of the 'unknown Western 1' sequence variant at all four positions diagnostic for that type. Thus, this unsampled or extinct variant reflected the 'Western' clade parent, and a species of the '*H. umbellatum*' group or a hybrid involving that group provided the 'Eastern' *ETS* ribotype. Its morphological similarity with *H. lucidum* or *H. sabaudum* could be due to the shared 'unknown Western 1' genome; its high variability might be caused by multiple origins involving different additional taxa.

***H. sabaudum* L.**

It has a wider geographic distribution than *H. racemosum* in northern, eastern and western directions. The involucral bracts of both species are similar. Also, *ETS* character additivity between both major clades, contribution of the 'unknown Western 1' sequence variant, and sharing of character states, polymorphisms and cpDNA with the '*H. umbellatum*' group of the triploid accession analyzed corresponded to *H. racemosum*. Additional polymorphic sites narrowed the parentage of the '*H. umbellatum*' group down to *H. umbellatum* um.AM.1 sampled from the same locality. In accessions of both *H. sabaudum* and *H. racemosum*, the 'Eastern' sequence variant slightly predominated according to relative peak heights, more strongly so in *H. sabaudum*. Correspondingly, two out of three clones in *H. sabaudum* and three out of six clones in *H. racemosum* reflected the '*H. umbellatum*' ribotype.

***H. bracteolatum* Sibth. & Sm.**

This species has a unique morphology, belongs to a monotypic section, and occurs only in Greece. Nevertheless, the analyzed triploid accession had the same inferred parentage as the previous two taxa (but lacking the additional polymorphisms of *H. sabaudum*). 'Western' and 'Eastern' sequence variants were present in about equal amounts. Either the morphological differences despite similar parentage reflect ecological adaptations that did not affect nrDNA variation, or the interspecific genetic variation was too low to distinguish between morphologically different taxa of the same subclade (as in the case of *H. umbellatum* and *H. eriophorum*).

***H. glaucum* All.**

Morphologically, this species shows influences of *H. bifidum* and either *H. porrifolium* or *H. bupleuroides*. The analyzed triploid accession was indeed an interclade hybrid. Two additional additive sites showed contribution of the '*H. porrifolium*' subclade. Another polymorphism was shared with *H. porrifolium* and *H. bupleuroides*. In addition, a three-character state additive site composed of a further polymorphism shared by

these two species and a 'Western' clade substitution occurred (W + C). CpDNA also corresponded to the '*H. porrifolium*' haplotype. The 'Western' clade was less differentiated, and there were only few sites available for tracing the second parent. However, at a single site, *H. glaucum* shared a polymorphism with *H. bifidum* and *H. humile* 1064. Thus, in this case, the *ETS* data reflected the morphological similarities.

***H. olympicum* Boiss.**

This is a species from the Southern Balkans and Northern Anatolia. The triploid accession showed character additivity at all 15 positions distinguishing 'Eastern' and 'Western' ribotypes. The 'Eastern' variant was somewhat dominating, especially towards the 3'-end, which was reflected by the recombinant sequences (Additional file 2: Patterns of *ETS* recombination). The 'Western' clade was represented by the 'unknown Western 1' ribotype. Eight additional polymorphisms were shared with different species from the Balkans. The most similar pattern belonged to *H. sparsum*. This combination of 'Eastern' and 'Western' subtypes was unique among all analyzed accessions. Its cpDNA was also unique, most similar to that of *H. gymnocephalum*, another interclade hybrid. However, none of the identified ribotypes were shared among these two taxa suggesting the involvement of an unknown maternal parent not reflected by *ETS* for at least one of them.

***H. gymnocephalum* Griseb. ex Pant.**

The species occurs in the Western Balkans. In Macedonia, its distribution area overlaps with *H. olympicum*. The analyzed accessions – one of them diploid, one triploid – indicated this to be also a hybrid taxon between the 'Eastern' and 'Western' clade. Apart from the interclade additive sites, a rather large number of additional intra-individual polymorphisms occurred. Eight (4 + 4) represented the 'unknown Western 2' and the 'unknown Eastern' ribotypes (Additional file 2). Only 1–2 polymorphisms were shared with 'pure' species from the Balkans (Figure 4). A unique cpDNA haplotype and a unique substitution in the *ETS* (note that substitutions were generally rare and most of the variation within clades was based on shared polymorphisms) suggested that this might be a rather old taxon and, in addition, that perhaps no extant species can be considered as a direct predecessor. It is also remarkable that despite different ploidy levels, the two accessions had a nearly identical pattern of intra-individual polymorphisms and equal ratios of 'Eastern' and 'Western' ribotypes. Nearly identical, but rather derived cpDNA haplotypes of *H. olympicum* and *H. gymnocephalum* indicate a common maternal origin of these species from an unknown donor. It may have been a taxon belonging to the 'unknown Eastern' lineage according to geographic distribution. In any case, either *H. olympicum* or *H. gymnocephalum* (or both) must have had an additional genome donor that could not be traced as their parentages inferred from *ETS* ribotypes showed no overlap while their cpDNAs matched.

***H. heterogynum* (Froel.) Guterm.**

This species has a similar distribution to *H. gymnocephalum*, mostly occurring in former Yugoslavia. Morphologically, it is characterized by a particular kind of indumentum. The analyzed accession was also an interclade hybrid. A high number of additional shared polymorphisms occurred in the *ETS*. Many were accession-specific or occurred at homoplasious positions. Four reflected the 'unknown Eastern' ribotype and were shared with the interclade hybrids *H. gymnocephalum* and *H. plumulosum*. Two polymorphisms showed additivity with species of the '*H. umbellatum*' group to which also the cpDNA corresponded. Apparently, an 'Eastern' species from the '*H. umbellatum*' group was involved in its origin, but only two out of seven sites in the *ETS* (plus the cpDNA) reflected this. Both remaining character states of the '*H. umbellatum*' ribotype were situated on the 'wrong' strands suggesting that their character states were probably maintained by gene conversion while the other five were lost from the genome (Additional file 2), maybe together with a particular rDNA locus. An '*H. umbellatum*' chloroplast haplotype in combination with an almost erased '*H. umbellatum*' *ETS* pattern also suggests that this clade made a rather ancient genome contribution to *H. heterogynum* and that several subsequent hybridization events occurred through pollination by different donors. A clone representing the 'unknown Western 2' ribotype was also found (in addition to 'ordinary' 'Western' sequences) although exceptionally direct sequencing did not show any trace of this, i.e., this ribotype must be present in less than 5% of all copies in the genome. Further accession-specific polymorphisms were retrieved by cloned sequences that were not present in the direct sequence, but the clones

also did not retrieve all of the accession-specific substitutions inferred from direct sequencing. It seems that this accession has particularly complex rDNA arrays and, in addition to the two '*H. umbellatum*'-specific characters, more examples of potential gene conversion were found (Additional file 2). In the triploid accession of *H. heterogynum*, three different ribotypes could be clearly identified, plus remnants of a fourth whose genomic contribution was also apparent from cpDNA.

***H. plumulosum* A. Kern.**

This species occurs in the Western Balkans. Species with a similar kind of indumentum are known from the Southwestern Alps and from the Appennin. But as this is a rather widespread feature of *Hieracia* in Submediterranean regions, it could be an adaptation to dry climate instead of reflecting species relationships. Morphologically, *H. plumulosum* stands between *H. pannosum*-like species and possibly *H. tomentosum*. It also had a genome composed of the two major clades, however, the 'Eastern' ribotype was strongly overrepresented (about 80% of the total signal). The analyzed diploid from a deep river valley with relict character had by far the highest number of polymorphic sites among all investigated accessions (37). Many of them were unique, but it also shared more polymorphic sites with other taxa than most samples. In addition to 'ordinary' 'Western' and 'Eastern' variants, the 'unknown Western 2' and the 'unknown Eastern' ribotypes were found (Additional file 2). The latter was shared with the interclade hybrids *H. heterogynum* and *H. gymnocephalum* and may indicate a genome donor geographically restricted to the Balkans. This is supported by 1–2 'Balkan' polymorphisms occurring on the 'unknown Eastern' strands suggestive of the origin of this ribotype from a Balkan species. Its chloroplast haplotype was unique and comparably divergent from all other species. This taxon (or at least this accession) had a highly reticulate history with four *ETS* ribotypes and a chloroplast haplotype not corresponding to any other taxon sampled. This example points out that at least two major rDNA loci per haploid genome should exist.

Details of accession origin (including outgroup taxa)

Species	Accession	Origin of samples
H. alpinum L.	alp.Ukr	Ukraine: Chornohora Mts., Ukrainian Carpathians, Polonina Breskulska ridge, saddle between Mt. Hoverla and Mt. Breskul, 1800 m a.s.l., 48°09'09.8''N, 24°30'14.6'', 23 July 2003, leg. P. Mráz & J. Chrtek
	alp.Boa.2	Romania: Munţii Rodnei Mts, glacial cirque on the NE slopes of Mt. Pietrosul Mare, ca 0.3 km SE from Staţia Meteo, ca 1900 m a.s.l., 5 July 2001, leg. P. Mráz
H. amplexicaule L.	1050	Austria: Carinthia, Hohe Tauern, Goldberggruppe: Innerfragant, near the old (unmarked) path to the Fraganter Hütte, ca 1 km SW of the village, rocks above the brook, 1233 m, 13º 04´20´´ E, 46º 57´56´´ N, 28 July 2005, leg. J. Chrtek & P. Mráz
H. bifidum Kit.	1213	Slovakia: Orava, distr. Tvrdošín, Oravice: Juráňova dolina-Tiesňavy, 3 km SE of the village, limestone gorge, 930 m, 19º 46´19´´ E, 49º 16´31´´ N, 15 August 2006, leg. J. Chrtek
H. bracteolatum Sibth. & Sm.	1240	Greece: Thessalia, Pilion Mts., Agriolefkes, 1400 m, leg. Binder et al. (S, BGBM Berlin-Dahlem)
H. bupleuroides C.C. Gmel.	1033	Slovakia: distr. Ilava, Biele Karpaty Mts.: Vršatské Podhradie, castle ruine of Vršatec, 770 m, 18º 09´00´´ E, 49º 03´57´´ N, 15 June 2005, leg. J. Chrtek
	1212	Slovakia: Orava, distr. Tvrdošín, Oravice: Juráňova dolina-Tiesňavy, 3 km SE of the village, limestone gorge, 930 m, 19º 46´19´´ E, 49º 16´31´´ N, 15 August 2006, leg. J. Chrtek
H. caesium (Fr.) Fr.	1231	Sweden: prov. Gotland, par. Hall, open limestone scree by the sea 1.3 km SE of Hallshuk (close to the NW point of the island of Gotland ca 40 km NNE of Visby), 20 m, 18º 43´ E, 57º 55´ N, July 2006, leg. et det. T. Tyler & A. Sennikov
H. canadense Michx.	canad	Canada: B&T World Seed, http://www.b-and-t-world-seeds.com/index.html, herbarium GLM 157756
H. candidum Scheele	1197	Spain: Catalunya, prov. Lérida, distr. La Seu d´Urgell: Adraén, Serra del Cadí mountain ridge, NW slopes, 3 km SE of the village, 1750 m, road margin in a pine forest, 23 July 2006, leg. J. Chrtek, G. Mateo & J. A. Rosselló, det. G. Mateo
H. cerinthoides L.	1176	Spain: Catalunya, prov. Lérida, Pirineus Mts: Os de Civís, 1 km WSW of the village, margin of a pasture, 1720 m, 1º 25´46´´ E, 42º 26´47´´ N, 21 July 2006, leg. J. Chrtek, G. Mateo & J. A. Rosselló, det. G. Mateo
H. cordifolium Lapeyr.	1177	Andorra: Pirineus Mts, Bixessarri (NW of Sant Julià de Lòria), valley of Torrent dels Llimois, rocks and margins of a path ca 100 m from the street, 1.5 km NW of the village, 1305 m, 1º 26´44´´ E, 42º 29´33´´ N, 21 July 2006, leg. J. Chrtek, G. Mateo & J. A. Rosselló, det. G. Mateo
H. eriophorum St.-Amans	1221	France: dépt. Landes, Labenne: plage de Labenne Océan Sud, 10 m, 1º 27´20´´ W, 43º 36´17´´ N, 27 September 2006, leg. E. Forey
	1222	France: dépt. Landes, Seignosse-le-Penon, plage de Estagnols Seignosse, 10 m, 1º 25´51´´ W, 43º 41´40´´ N, 27 September 2006, leg. E. Forey
H. glaucum All.	1230	Slovenia: Julian Alps, village of Trenta: Zadnja Trenta valley, 0.5 km W of the chalet 'Koča pro izviru Soče', margin of gravel alluvium, 910 m, 13º 43´41´´ E, 46º 24´22´´ N, September 2006, leg. J. Chrtek
H. gouani Arv.-Touv.	1171	Spain: Catalunya, prov. Gerona: rocks at the road between Ripoll and Ribes de Fresser, 1290 m, 2º 10´02´´ E, 42º 15´27´´ N, 24 July 2006, leg. J. Chrtek
H. gymnocephalum Griseb. ex Pant.	1215	Albania: NW part, Jezercës, 21 km NW of Bajram Curri, 1800 m, 19º 50´23´´ E, 42º 25´46´´ N, August 2006, leg. J. Zahradníček, det. Z. Szeląg & V. Vladimirov
	1207	Montenegro: Durmitor Mts, August 2006, leg. and det. Z. Szeląg
H. gymnocerinthe Arv.-Touv. & G. Gaut.	1172	Spain: Catalunya, prov. Lérida, distr. La Seu d'Urgell: Adraén, Serra del Cadí mountain ridge, NW slopes, 1 km SE of the village, 1600 m, road margin in a pine forest with dominating *Arctostaphylos uva-ursi*, 1º 30´34´´ E, 42º 16´15´´ N, 23 July 2006, leg. J. Chrtek, G. Mateo & J. A. Rosselló, det. G. Mateo
H. heterogynum (Froel.) Guterm.	1250	Montenegro: Kotor, Mt. Lovčen, 1350 m, 18º 47´ E, 42º 23´ N, 20 August 2006, leg. M. Niketić
H. humile Jacq.	1064	Austria: Oberösterreich, Dachstein massif: Vorderer Gosausee (mountain lake), rocks on NW bank, 6 km SSW of the village of Gosau, 940 m, 13º 29´55´´ E, 47º 31´53´´ N, 13 August 2005, leg. J. Chrtek
	1188	France: Le Midi, dépt. Aude, Corbières Mts: Bugarach, Mt. Pech de Bugarach, 1130 m, 2º 22´47´´ E, 42º 52´06´´ N, 27 July 2006, leg. J. Chrtek
H. kittanae Vladimirov	1228	Bulgaria: Central Rhodope Mts.: Trigrad gorge, limestone rocks near the natural entrance to Dyavolskoto garlo cave, 750–800 m, September 2005, leg. P. Ignatova
H. lachenalii Suter	1160	Czech Republic: Moravia, distr. Znojmo: Lukov, forest 1.3 km SSW of the village, 410 m, 15º 54´27´´ E, 48º 51´04´´ N, June 2006, leg. J. Zahradníček
H. laevigatum Willd.	1031	Czech Republic: Bohemia, distr. Rokycany: Strašice, N part of the village, margin of a forest, 550 m, 13º 45´07´´ E, 49º 44´51´´ N, 29 June 2005, leg. J. Chrtek
H. lawsonii Vill.	1175	France: Le Midi, dépt. Aude, Corbières Mts: Bugarach, Mt. Pech de Bugarach, 1130 m, 2º 22´47´´ E, 42º 52´06´´ N, 27 July 2006, leg. J. Chrtek
H. lucidum Guss.	H.lucidum	Italy: Sicily, distr. Palermo: Sferracavalio, limestone rocks between the village and Capo Gallo, 40 m, 13º 17´56´´ E, 38º 12´56´´ N, 11 April 2007, leg. J. Chrtek et al.
H. mixtum Froel.	H.mixtum	Germany: Lower Saxony, 3723/33, southern part of the Deister hill near Springe, quarry at the Ebersberg, 345 m, herbarium GLM
H. murorum L.	875	Czech Republic: Bohemia, distr. Plzeň: Plzeň, village of Koterov, street "V závrtku", 0.6 km SSW of the railway station "Plzeň-Koterov", slopes along the street, ca 320 m, 13º 25´07´´ E, 49º 43´02´´ N, 12 August 2003, leg. M. Král
H. naegelianum Pančić	1208	Montenegro: Durmitor Mts.: Mt. Veliki Međed, alpine grassland on limestone, 2050 m, 43º 03´31´´ N, 19º 04´13´´ E, 1 August 2006, leg. and det. Z. Szeląg
H. olympicum Boiss.	1206	Bulgaria: Stara Planina Mts., Kaloferska Planina Mts.: Valley of Vidima riiver, 2 km NE of the Kaloferski Monastyr, eroded slope in *Carpinus orientalis* forest, 870 m, 24º 58´42´´ E, 42º 40´36´´ N, 9 August 2006, leg. and det. Z. Szeląg
H. pannosum Boiss.	1205	Bulgaria: Stara Planina Mts., Trojanska Planina Mts.: Mt. Kozja stena, grassy slope on limestone, 1570 m,

		42°47′27′′ N, 24°34′06′′ E, 8 August 2006, leg. and det. Z. Szeląg
H. petrovae Vladimirov & Szeląg	1229	Bulgaria: Central Rhodope Mts.: Trigrad gorge, crevices of limestone rock (*locus classicus*), 750–800 m, 41°39′ 55′′N, 24°21′ 50′′E, 15 October 2005, leg. V. Vladimirov
H. pictum Pers.	1067	France: dépt. Alpes Maritimes, valley of la Roya: Tende, along the old road to the Col de Tende, ca 0.5 km above the tunnel, 6 km NNW of the village, 1331 m, 07º 33′ 57′′ E, 44º 08′ 19′′ N, 28 August 2005, leg. J. Chrtek & P. Mráz
	1307	France: dépt. Hautes Alpes, Briançon, near a path to Fort des Sallettes, 1360 m, 6º 38′ 52′′ E, 44º 54′ 33′′ N, 2 August 2007, leg. J. Chrtek et al.
H. pilosum Schleich. ex Froel.	1226 (2)	Slovenia: Julijske Alpe Mts, Log pod Mangartom: Mt. Travnik (2200 m), SW slopes, limestone, 2120 m, 13º 38′40′′ E, 46º 26′22′′ N, 7 September 2006, leg. J. Chrtek, J. Fehrer et al.
H. plumulosum A. Kern.	1218	Montenegro: Canyon of the Mrtvica river, 35 km SW of Kolasin, halfway through the canyon, near the bridge, 1000 m, 19° 48′ 59′′ E, 42° 28′ 40′′ N, August 2006, leg J. Zahradníček, det Z. Szeląg
H. pojoritense Woł	poi.Rom	Romania: Pojorita, Cimpalung Moldovec, herbarium P. Mráz
H. porrifolium L.	1052	Austria: Carinthia, the Karawanken Mts.: Bad Eisenkappel, limestone rocks and pine forests (alliance *Erico-Pinion*) near the road to Bad Vellach, 4.5 km SSW of the town, 658 m, 14º 34′20.5′′ E, 46º 27′07.1′′ N, 26 July 2005, leg. J. Chrtek & P. Mráz
H. prenanthoides Vill.	1161	Poland: Województwo dolnośląskie, Karkonosze Mts., Jagniątków: Mały Kocioł Śnieżny glacial cirque, along the path, 1250 m, 15° 33′ 28′′ E, 50° 46′ 27′′ N, July 2006, leg. J. Zahradníček
	1252	France: dépt. Hautes Alpes, La Grave, below the village, ca 1500 m, 06º 18′21′′ E, 45º 02′37′′ N, June 2003, leg. P. Mráz
	1187	Andorra: Pirineus Mts, Canillo, SE margin of the village, 1530 m, 1º 36′11′′ E, 42º 33′56′′ N, 22 July 2006, leg. J. Chrtek, G. Mateo & J. A. Rosselló
H. racemosum Waldst. et Kit. ex Willd.	874	Czech Republic: SW Moravia, distr. Znojmo, Vranov nad Dyjí, forest 0.6 km SE of the village, 370 m, 15º 49′09′′ E, 48º 53′32′′ N, 27 September 2003, leg. J. Chrtek
H. ramondii Griseb.	1173	Andorra: Pirineus Mts., Encamp, valley of Riu de les Deveses, NW slopes of Mt. Alt del Griu, 3.8 km E of the town, rocky outcrops in a light mountain forest, 2040 m alt., 1º 37′52.1′′ E, 42º 32′07.1′′ N, 22 July 2006, leg. J. Chrtek, G. Mateo & J. A. Rosselló, det. G. Mateo
H. recoderi De Retz	1174	Spain: Catalunya, prov. Barcelona: Berga, monastery of Queralt, rocks ca 200 m below the parking place, 805 m, 1º 49′24′′ E, 42º 06′54′′ N, 24 July 2006, leg. J. Chrtek
H. sabaudum L.	1098	Germany: Oberlausitz , distr. Löbau-Zittau: Schönau-Berzdorf, 220 m, 14º 53′48′′ E, 51º 04′01′′ N, July 2004, leg. S. Bräutigam
H. schmidtii Tausch	1025	Czech Republic: Bohemia, distr. Litoměřice: Boreč, the Boreč hill, E slope, 500 m N of the village, 367 m, 13º 59′25′′ E, 50º 30′53.4′′ N, 15 May 2005, leg. J. Chrtek
H. sparsum Friv.	1251	Bulgaria: Sofia, Vitoša Mts.: NE slope of Mt. Vitoša, Bistriško Branište biosphere reserve, 2000 m, 23º 17′56′′ E, 42º 34′07′′ N, 23 June 2006, leg. F. Krahulec & A. Krahulcová
	spa.sst.2	Bulgaria: Pirin Mts., Vihren Mt., garden culture Z. Szeląg, herbarium PRA
H. stelligerum Froel.	1233	France: dépt. Ardèche, Vallon Pont d`Arc: crevices of calcareous rocks along the road D 390, opposite of 'le Pont d'Arc', ca 3.5 km SE of the village, 500 m, 04º 24′ 10′′ E, 44º 24′ 25′′ N, October 2006, leg. P. Mráz
H. tomentosum L.	1066	France: dépt. Alpes Maritimes, valley of la Roya: Tende, along the old road to the Col de Tende, ca 0.5 km above the tunnel, 6 km NNW of the village, 1331 m, 07º 33′ 57′′ E, 44º 08′ 19′′ N, 28 August 2005, leg. J. Chrtek & P. Mráz
H. transylvanicum Heuff.	1077	Ukraine: Oblast′ Zakarpatska, Marmaros′ki Al′py Mts.: Mt. Berlebashka (1480 m), NW slope along the trail (red marked), E of the village of Dilove, 1200 m, 24º 21′31′′ E, 47º 56′13′′ N, 19 September 2005, leg. J. Zahradníček
	tra.Boa	Romania: Munții Rodnei Mts, border of the tourist path from the village of Borșa to Mt. Pietrosul Mare, spruce forest, 1300–1400 m a.s.l., 47°39' N, 24°39' E, 5 July 2001, leg. P. Mráz
H. umbellatum L.	1021	Poland: Województwo pomorskie, Baltic coast, Jantar, 5 m, 19º 02′27′′ E, 54º 20′00′′ N, 27 June 2002, leg. et det. Z. Szeląg
	um.AM.1	Germany: Upper Lusatia, SE Schönau-Berzdorf, herbarium GLM 46889
H. villosum Jacq.	1029	Slovakia: Strážovské vrchy Mts, distr. Ilava: Mt. Strážov, summit region, calcareous rocks, 1310 m, 18º 27′44′′ E, 48º 57′19′′ N, 18 June 2005, leg. J. Chrtek
	1305	France: dépt. Savoie, slopes below (N of) Col du Galibier, the mountain lake 10 km S of the village of Valloire, 2380 m, 6º 24′43′′ E, 45º 04′21′′ N, 3 August 2007, leg. J. Chrtek et al.
H. virosum Pall.	vir.1	Russia: Rostov-na-Donu, vicinity of the town, 150 m, 39º 49′41′′ E, 47º 17′05′′ (Botanical Garden Rostov)
	1238	Russia: Siberia, Altajskij kraj, W of Barnaul, 220 m, leg. Ristow/Seitz (BGBM Berlin-Dahlem)
Pilosella lactucella (Wallr.) P. D. Sell & C. West	lac.Jon.1	Germany: Erzgebirge, culture Görlitz, herbarium GLM 140613
Hispidella hispanica Barnades ex Lam.	His.his.2	Spain: Sierra de Guadarrama, leg. J. Pizarro et C. Navarro no CN 2460, herbarium M
Andryala integrifolia L.	Ant.int.2/2	Spain: Andalusia, Malaga, herbarium GLM 141138
Andryala levitomentosa (Nyár.) P.D. Sell	A.lev.maj.1	Romania: Pietrosul Bogolini, herbarium GLM 156367
'*Hieracium*' *intybaceum* All.	H.intybac	B&T World Seed, http://www.b-and-t-world-seeds.com/index.html, herbarium PRA
	inb.Kaer	Austria: Kärnten, S. Jagalski 4, herbarium M

Additional file 5: Species/accessions, their origin, cytotype, *ETS* and cpDNA features

Species	Access.	Origin	Ploidy[1]	DNA content 2C [pg][1]	Inferred origin[2] (*ETS*)	cpDNA[3] (*trnT-L*)	Remarks
H. alpinum	alp.Ukr	Ukraine	2x	7.9	EA	EA	
	alp.Boa.2	Romania	2x	n.d.	EA	EA	
H. amplexicaule	1050/1	Austria	3x	10.8	WP-E	dWP1	'Western' *ETS* dominating
H. bifidum	1213/2	Slovakia	3x	10.7	W	W	
H. bracteolatum	1240/2	N Greece	3x	12.4	Wx-EU	EU	
H. bupleuroides	1212/2	Slovakia	3x	11.7	Epo	Epo	
	1033/3	Slovakia	3x	12.0	EU-Epo	EU	
H. caesium	**1231**	Sweden	4x	14.6	W-EU	W	'Western' *ETS* dominating
H. canadense	canad	Canada	3x	12.3	EU	EU	
H. candidum	1197/3	Spain	3x	n.d.	WP-W	WP2	
H. cerinthoides	1176/2	Spain	3x	10.8	WP-W	WP2	
H. cordifolium	1177/5	Andorra	2x	7.2	WP-W	dWP2	
H. eriophorum	1221/1	France	2x	8.5	EU	EU	
	1222/2	France	2x*	n.d.	EU	EU	
H. glaucum	1230/3	Slovenia	3x	11.4	W-Epo	Epo	
H. gouani	1171/4	Spain	2x	7.1	WP-E	WP1	'Western' *ETS* dominating
H. gymnocephalum	**1215**/1	Albania	2x	8.4	Wy-Ex	dO	
	1207/2	Montenegro	3x	n.d.	Wy-Ex	dO	
H. gymnocerinthe	1172/4	Spain	3x	10.7	WP-W	WP1	
H. heterogynum	**1250**/2	Montenegro	3x	12.5	W-Wy-Ex-EU	EU	partial *ETS* additivity for EU
H. humile	1064/2	Austria	4x	14.4	W	W	
	1188/2	France	3x	10.6	W	W	
H. kittanae	**1228**/2	Bulgaria	2x	8.4	EB	EB	
H. lachenalii	1160/2	Czechia	3x	11.3	W	EU	'Western' *ETS*, cpDNA captured from EU
H. laevigatum	1031/11	Czechia	3x	12.2	W-EU	EU	
H. lawsonii	1175/1	France	3x	10.9	WP	WP2	
H. lucidum	H. lucidum	Sicily	2x*	n.d.	W-Wx	W	
H. mixtum	**H. mixtum**	Germany	3x*	n.d.	W-E	unique	hybrid *ETS* dominating (Table 1)
H. murorum	875/1	Czechia	3x	10.7	W	W	
H. naegelianum	1208/2	Montenegro	3x	10.9	EB	unique	
H. olympicum	**1206**/3	Bulgaria	3x	12.1	Wx-EB	O	
H. pannosum	1205/1	Bulgaria	3x	11.7	EB	EB	
H. petrovae	1229	Bulgaria	2x	7.9	EB	EB	
H. pictum	1067/4	France	3x	10.6	W	W	
	1307/5	France	3x	10.7	W	W	
H. pilosum	1226/1	Slovenia	3x	11.6	Epo	Epo	
	1226/2	Slovenia	3x	11.8	Wy-Epo	Epo	
H. plumulosum	**1218**/2	Montenegro	2x	8.6	W-Wy-Ex-E	unique	'Eastern' *ETS* dominating (Table 1), multiple introgressions probable
H. pojoritense	poi.Rom.1	Romania	2x	n.d.	EA-EU	EA	'*H. alpinum*' *ETS* dominating
H. porrifolium	1052/9	Austria	2x	7.8	Epo	Epo	
H. prenanthoides	1252	France	2x	7.1	W-E	W	'Western' *ETS*, 3'-end additive (Table 1)
	1161/2	Poland	3x	10.9	W-E-Wx	dW	'unknown Western 1' *ETS*, 3'-end additive (Table 1)
	1187/1	Andorra	3x	11.5	W-E-EU	W	'Western' *ETS* dominating, 3'-end additive (Table 1), partial *ETS* additivity for EU
H. racemosum	874	Czechia	3x	12.5	Wx-EU	EU	
H. ramondii	1173/3	Andorra	3x	10.7	WP	WP1	
H. recoderi	1174/4	Spain	2x	7.1	WP	WP1	
H. sabaudum	1098/2	Germany	3x	12.6	Wx-EU	EU	
H. schmidtii	1025/3	Czechia	3x	10.7	W	W	
H. sparsum	1251/1	Bulgaria	2x	8.1	EB	dEA	cpDNA captured from EA
	spa.sst.2	Bulgaria	2x*	n.d.	EB	dEA	cpDNA captured from EA
H. stelligerum	1233/1	S France	2x	7.0	W	W	
H. tomentosum	1066/8	France	2x	7.5	W	W	
H. transylvanicum	tra.Boa	Romania	2x	8.3	W	unique	genome size and distribution may suggest cryptic 'Eastern' origin
	1077/7	Ukraine	2x	8.5	W	unique	genome size and distribution may suggest cryptic 'Eastern' origin
H. umbellatum	1021/1	Poland	2x	8.3	EU	EU	
	um.AM.1	Germany	2x	n.d.	EU	EU	
H. villosum	1029/1	Slovakia	4x	15.7	Epo	dEpo	
	1305/3	France	3x	11.6	Wy-Epo	Epo	'Eastern' *ETS* dominating
H. virosum	1238/1	Russia	3x	13.0	EU	EU	
	vir.1	Russia	3x	13.0	EU	unique	

[1] Chromosome counts and DNA content determined by flow cytometry, data adopted from [46].

[2] Apart from 'Eastern' (E) and 'Western' (W) ribotypes that could not be further differentiated, the following subgroups are indicated (corresponding to Figures 2 and 5): 'Pyrenean' (WP), 'unknown Western 1' (Wx), 'unknown Western 2' (Wy), '*H. alpinum*' (EA), '*H. umbellatum*' (EU), '*H. porrifolium*' (Epo), 'Balkan' (EB), 'unknown Eastern' (Ex).

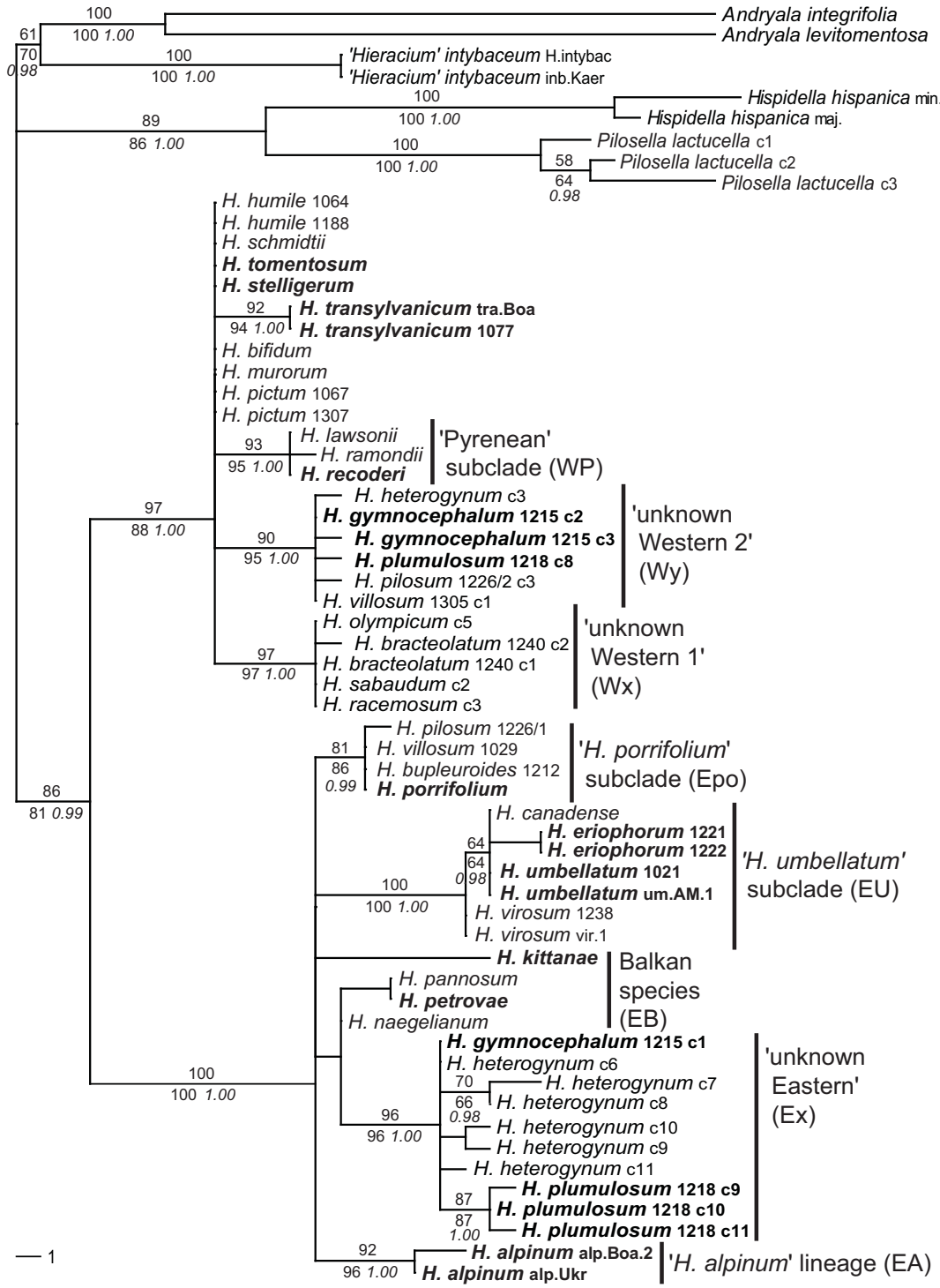
[3] Chloroplast haplotypes (see Figure 3): WP1 and WP2 are different haplotypes of Pyrenean species; the '*H. olympicum*' haplotype is specified as "O". Haplotypes derived from other haplotypes / haplotype groups are given as, e.g., dWP1.

Asterisks after ploidy levels indicate chromosome counts reported for that species (in the strict sense), but not available for the particular accession.

The German accession of the Pyrenean species *H. mixtum* is from a recently described neophytic population (Bräutigam S, Gottschlich G, Hänel K. 2007. *Hieracium mixtum* FROEL. – ein für Deutschland neuer Neophyt. Kochia 2: 25-30).

For accessions marked in boldface, cloned sequences are shown in Additional file 2: Patterns of *ETS* recombination.

Additional file 6 - *ETS* phylogeny with ribotypes present only in hybrids



Fehrer J. et al. (2009) BMC Evol. Biol.

Supplementary files to Paper 4, Krak et al.: **Reticulate evolution and lineage sorting in *Hieracium* s. str. (Asteraceae): evidence from a low-copy nuclear gene and cpDNA**

Supplementary File 1. (provided only in the electronic version of the PhD thesis in .fasta format).

Sequence alignment of the combined cpDNA dataset. The alignment already contains the matrix of indels coded according to Fehrer et al. (2007).

Positions 1-1133: sequences of the *trn*V-*ndh*C intergenic spacer
Positions 1135-1146: matrix representing the indels of the *trn*V-*ndh*C region
Positions 1149-1724: sequences of the *trn*T-*trn*L intergenic spacer
Positions 1726-1734: matrix representing the indels of the *trn*T-*trn*L region

Supplementary File 2. (provided only in the electronic version of the PhD thesis in .fasta format).

Sequence alignment of the *sqs* dataset used for the initial phylogenetic analyses.

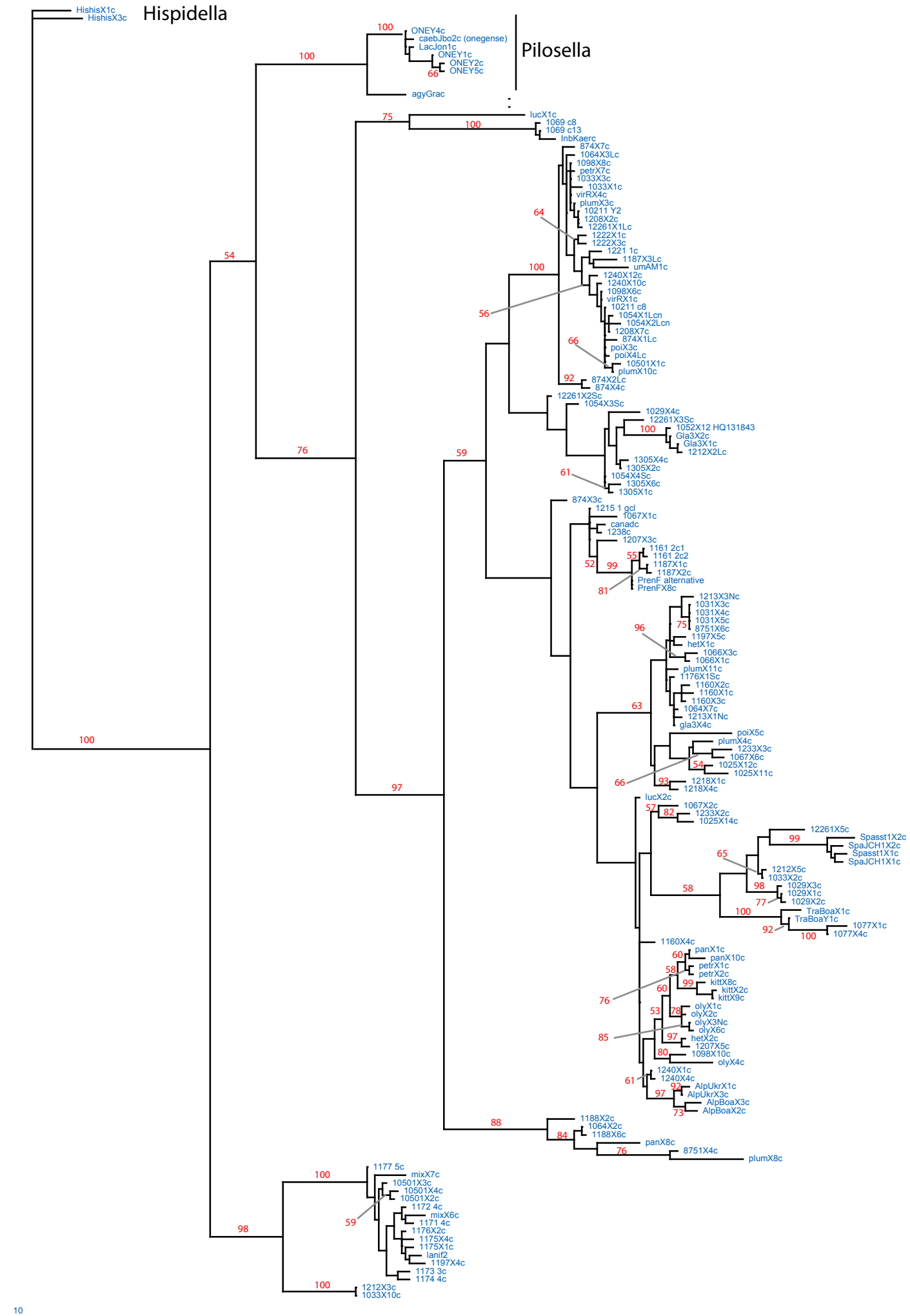
Supplementary File 3. (provided only in the electronic version of the PhD thesis in .fasta format).

Reduced sequence alignment of the *sqs* dataset used for the second round of phylogenetic analyses.

Supplementary files to Paper 4, Krak et al.: **Reticulate evolution and lineage sorting in *Hieracium* s. str. (Asteraceae): evidence from a low-copy nuclear gene and cpDNA**

Supplementary File 4.

Results of the initial phylogenetic analysis, based on the complete *sqs* dataset. One of the 100 most parsimonious trees (CI=0.492, RI=0.910) is presented. Bootstrap supports are indicated by the numbers above or below the branches.


Hispidella
Pilosella
HishisX1c
HishisX3c
ONEY4c
caebJbo2c (onegense)
LacJon1c
ONEY1c
ONEY2c
ONEY5c
agyGrac
lucX1c
1069 c8
1069 c13
InbKaerc
874X7c
1064X3Lc
1098X8c
petrX7c
1033X3c
1033X1c
virRX4c
plumX3c
10211 Y2
1208X2c
12261X1Lc
1222X1c
1222X3c
1221 1c
1187X3Lc
umAM1c
1240X12c
1240X10c
1098X6c
virRX1c
10211 c8
1054X1Lcn
1054X2Lcn
1208X7c
874X1Lc
poiX3c
poiX4Lc
10501X1c
plumX10c
874X2Lc
874X4c
12261X2Sc
1054X3Sc
1029X4c
12261X3Sc
1052X12 HQ131843
Gla3X2c
Gla3X1c
1212X2Lc
1305X4c
1305X2c
1054X4Sc
1305X6c
1305X1c
874X3c
1215 1 gcl
1067X1c
canadc
1238c
1207X3c
1161 2c1
1161 2c2
1187X1c
1187X2c
PrenF alternative
PrenFX8c
1213X3Nc
1031X3c
1031X4c
1031X5c
8751X6c
1197X5c
hetX1c
1066X3c
1066X1c
plumX11c
1176X1Sc
1160X2c
1160X1c
1160X3c
1064X7c
1213X1Nc
gla3X4c
poiX5c
plumX4c
1233X3c
1067X6c
1025X12c
1025X11c
1218X1c
1218X4c
lucX2c
1067X2c
1233X2c
1025X14c
12261X5c
Spasst1X2c
SpaJCH1X2c
Spasst1X1c
SpaJCH1X1c
1212X5c
1033X2c
1029X3c
1029X1c
1029X2c
TraBoaX1c
TraBoaY1c
1077X1c
1077X4c
1160X4c
panX1c
panX10c
petrX1c
petrX2c
kittX8c
kittX2c
kittX9c
olyX1c
olyX2c
olyX3Nc
olyX6c
hetX2c
1207X5c
1098X10c
olyX4c
1240X1c
1240X4c
AlpUkrX1c
AlpUkrX3c
AlpBoaX3c
AlpBoaX2c
1188X2c
1064X2c
1188X6c
panX8c
8751X4c
plumX8c
1177 5c
mixX7c
10501X3c
10501X4c
10501X2c
1172 4c
mixX6c
1171 4c
1176X2c
1175X4c
1175X1c
lanif2
1197X4c
1173 3c
1174 4c
1212X3c
1033X10c
10