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.

Tribe	Species	Amplicon size (bp)			Proportion of introns (%)		
		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).

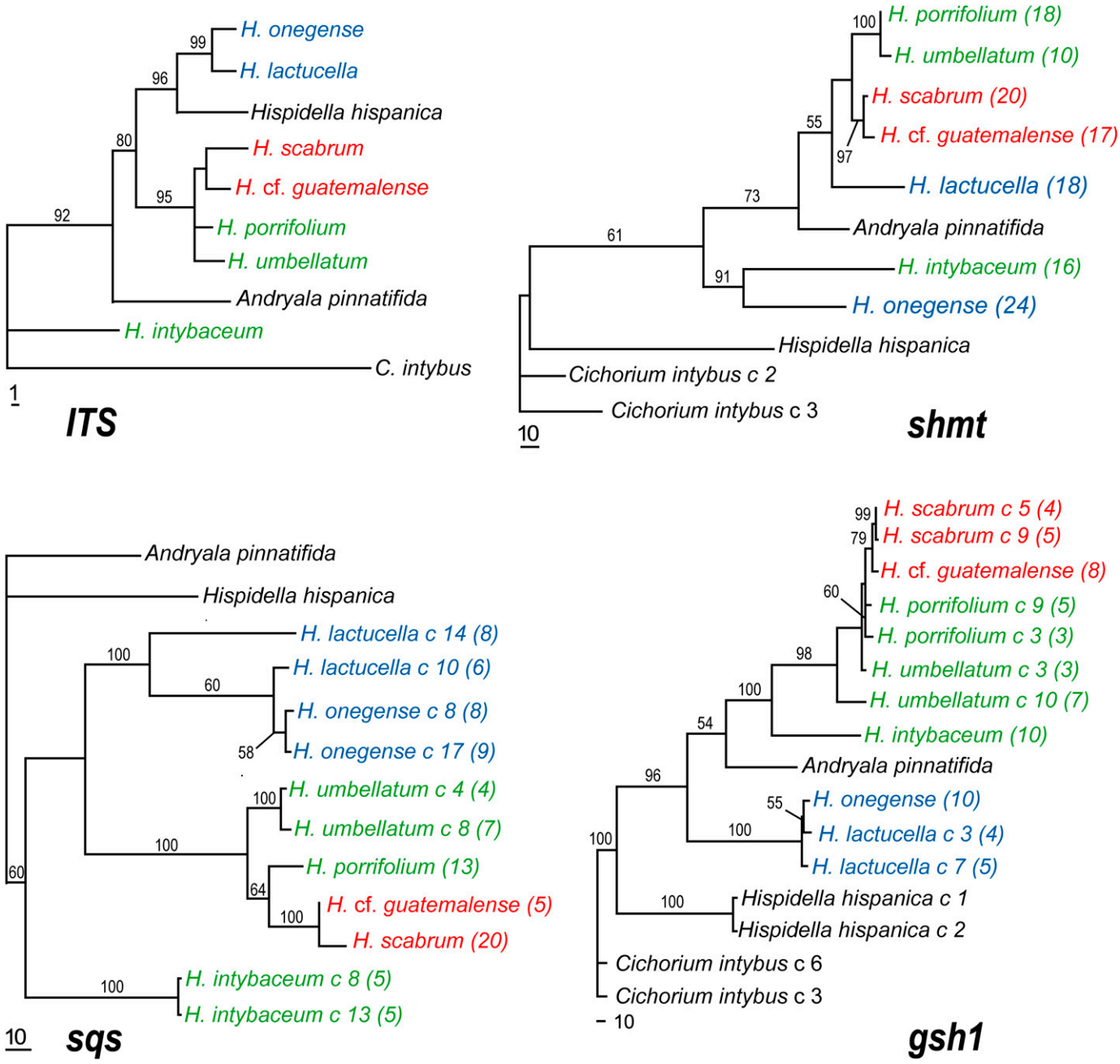

Fig. 1. Phylogenetic analyses of the three low-copy nuclear (LCN) markers and an ITS reference phylogeny for the Hieraciinae. For each of the three LCN markers, the single most parsimonious tree found in the analyses is presented. For ITS, one randomly selected tree from the four most parsimonious ones 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. Szelag. 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.