Nuclear intron–targeting markers in genetic diversity analysis of black nightshade (Solanum sect. Solanum, Solanaceae) accessions

ARTICLE *in* GENETIC RESOURCES AND CROP EVOLUTION · JANUARY 2014

Impact Factor: 1.46 · DOI: 10.1007/s10722-013-0031-z

4 AUTHORS, INCLUDING:

Peter Poczai

University of Helsinki

32 PUBLICATIONS **203** CITATIONS

SEE PROFILE

István Cernák

University of Pannonia, Veszprém

30 PUBLICATIONS **56** CITATIONS

SEE PROFILE

Jaakko Hyvönen

University of Helsinki

47 PUBLICATIONS **859** CITATIONS

SEE PROFILE

Genet Resour Crop Evol (2014) 61:247–266
DOI 10.1007/s10722-013-0031-z

RESEARCH ARTICLE

Nuclear intron-targeting markers in genetic diversity analysis of black nightshade (*Solanum* sect. *Solanum*, Solanaceae) accessions

Péter Poczai · István Cernák · Ildikó Varga · Jaakko Hyvönen

Received: 26 February 2013 / Accepted: 29 July 2013 / Published online: 17 August 2013

Abstract Different molecular markers are routinely used in studies of potato (*Solanum tuberosum*) and the genus *Solanum* in general. Genome sequence databases provide potential to design new markers for various applications. Here we present the application of a recently developed core set of nuclear intron-targeting (indel) markers. These markers are based on the fact that in the plant genome introns are more variable than exons; therefore primers flanking exons can reveal polymorphisms related to introns. We detected such variation among accessions of the eight different species of black nightshades (*Solanum* sect. *Solanum*). Members of this group are important sources of food, mostly in Africa, while others are poisonous weeds with near global distribution. The tested 29 primers were designed previously for potato based on Solanaceae EST and other genomic databases and targeted 16 different genes. Our results showed that *Solanum* intron-targeting markers are not very polymorphic but identified considerable structure among accessions indicating fairly high interspecies differentiation. Further analyses showed that inbreeding is unlikely to be the major driving force in determining the genetic structure of the analyzed species. All phylogenetic analyses resolved the species included in our study as distinct clades with high support values, but provided weak information about their internal relationships. In summary, indel markers would be useful for the assignment of new *Solanum* germplasm to taxonomic groups or to identify certain taxa. They could also be used to address important question about genetic diversity and should yield results comparable to other markers covering the whole genome.

Keywords Cross-species amplification · Expressed sequence tags · Gene-targeted markers · Genetic data mining · Indels · Marker transferability

Péter Poczai and István Cernák have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10722-013-0031-z) contains supplementary material, which is available to authorized users.

P. Poczai (✉) · I. Varga · J. Hyvönen
Plant Biology, Department of Biosciences, University of Helsinki, PO Box 65, 00014 Helsinki, Finland
e-mail: peter.poczai@gmail.com

I. Cernák
Potato Research Centre, Centre of Agricultural Sciences, University of Pannonia, Festetics 7, 8360 Keszthely, Hungary

Introduction

Molecular markers have numerous applications in plant genetic research ranging from genetic variability and diversity studies to the construction of linkage maps, and tracking individuals or lines carrying

particular genes. Nowadays they are routinely used in breeding programs to identify genes with valuable traits that could be introduced to crop gene pools (Dwivedi et al. 2007). Besides basic PCR-based techniques like random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and inter-sample sequence repeats (ISSR) that are often termed collectively as arbitrarily amplified DNA markers (AADs) many new promising alternative techniques have been developed in plant genetics. This is largely due to rapid growth in genomic research initiating a trend away from AADs towards gene-targeted functional markers. Public genomic databases offer an enormous potential for the development of new markers in diverse plant species (Holland et al. 2001). Gene-targeted functional markers from the transcribed region of the genome offer great potential for various applications in plant genotyping as they reveal polymorphisms that might be directly related to gene functions (De Keyser et al. 2009). In contrast, AADs are generated from non-coding regions or randomly over the genome, and the loci obtained are generally far away from the genes of the targeted sites (Hu et al. 2010). The recent availability of the full plant genome sequences like those of *Arabidopsis thaliana* (L.) Heynh. (Arabidopsis Genome Initiative 2000) and potato (*Solanum tuberosum* L.; Potato Genome Sequencing Consortium 2011) makes it possible to expand the search for well-distributed markers targeting gene rich regions. Solanaceae genomes have undergone relatively few genome rearrangements and duplications (Mueller et al. 2005) and thus gene markers could be easily transferred from one species to another. The relative conservatism of the gene structures of the solanaceous plants makes it possible to utilize intron sequences as molecular markers, although they may require additional work compared to other techniques.

The gene-regions including introns already used in various phylogenetic analyses of the Solanaceae include COSII, GBSSI and nitrate reductase genes. Wu et al. (2006) developed almost 3000 primer pairs for the conserved orthologous set (COSII) and used these in a study of the euasterid clade of the Asteraceae. The advantage of these markers is the fact that they are found in multiple positions in the nuclear genome, include various amounts of exons and introns and thus allow use in phylogenetic analyses at different taxonomic levels (Tepe and Bohs 2010). These primers have now been successfully used in numerous phylogenetic analyses of various groups of Solanaceae as exemplified by Ames and Spooner (2010), Cai et al. (2012), Rodríguez et al. (2009) and Tepe and Bohs (2010). Spooner et al. (2008) used GBSSI sequences in their study of potato polyploids. In addition to GBSSI Levin et al. (2009) used nitrate reductase gene in their analysis of American species of tribe Lycieae and Rodríguez and Spooner (2009) in their study of the polyploidy species of the section *Petota* Dumort.

Recently we have developed a core set of intron-targeting (IT) primers in potato for genetic studies (Poczai et al. 2010). This method is based on the observation that introns are more polymorphic than exons, thus primers designed to anneal in conserved exon junctions could reveal length polymorphisms possessed by the introns. In addition, due to the high level of sequence conservation in exon sequences it is possible to design cross-species exon flanking primers, which are capable to reveal variation in closely related *Solanum* L. species like black nightshades. Beside major crops species, e.g., potato (*S. tuberosum*), tomato (*S. lycopersicum* L.) or eggplant (*S. melongena* L.) there are many minor food plants in the genus like species of section *Solanum* L. or black nightshades. Members of this group that belong to the Morelloid clade (Weese and Bohs 2007) serve as emerging food sources in Africa, but some varieties are cultivated elsewhere. In Europe they carry a negative stigma due to their similarity to *S. nigrum* L., a poisonous weed commonly found across the continent. Large number of traditional landraces and varieties exist in regions where they are utilized as food and/or medical plants (Edmonds and Chweya 1997). In addition to many local uses in Africa the species of sect. *Solanum* are potentially globally important for agriculture, human health, plant breeding and biotechnology. The leaves and berries are a potential source of coloring plant extracts, inks and dyes (Lehmann et al. 2007) and they are rich in proteins, fibers, vitamins and amino acids (Schippers 2000). The value of black nightshades as a genetic resource for *Solanum* breeding may also be considered since many species are resistant to late blight—economically one of the most important diseases of solanaceous plants—caused by an oomycete *Phytophthora infestans* (Mont.) de Bary (Lebecka 2008). Resistance genes from *S. nigrum* have been successfully transferred to potato (Colon et al. 1993;

Eijlander et al. 1994; Horsman et al. 1997) but the species has not proved to be a useful source for traditional breeding. However, there is growing interest to search for alternative sources of R genes in European *Solanum* species for example in *S. dulcamara* L. (Poczai et al. 2011; D'Agostino et al. 2013) of the Dulcamaroid group, the sister of the Morelloid clade. From this species two new resistance loci to late blight, *Rpi-dlc1* and *Rpi-dlc2*, have been reported from chromosome 9 and 10 (Golas et al. 2010 , 2013). Such European species may represent a promising new pool of possible R genes alternative to currently used American sources.

Section *Solanum* often called as the *S. nigrum* complex is one of the largest and most variable species group of the genus (Edmonds and Chweya 1997) and this is the reason why we chose these plants to demonstrate the utility of intron-targeting (IT) in genetic diversity assessment. We aim to investigate the efficiency and transferability of IT by cross-species amplification of a core primer set developed for potato in chosen accessions of black nightshades. Although, sect. *Solanum* is under intensive taxonomic studies we concentrate here to overview the advantages and drawbacks of the IT method rather than revising the systematics and species relationships of the section.

Materials and methods

Plant material, in vitro culture and DNA extraction

A total of 61 accessions of black nightshades related to *S. nigrum* (Fig. 1) of diploid (*S. americanum* Mill., *S. physalifolium* Rusby, *S. chenopodioides* Lam.), tetraploid (*S. retroflexum* Dunal, *S. villosum* Mill.) and hexaploid (*S. nigrum*, *S. opacum* A. Br. et Bouché, *S. scabrum* Mill.) taxa were used in this study (Table 1). The Hungarian potato cultivar white lady was used as the reference genotype and as the outgroup together with a bittersweet accession (*S. dulcamara*). Plants were grown in vitro to obtain fresh tissues for DNA extractions. Seeds were surface-sterilized by a brief ethanol (70 %) immersion followed by surface-sterilization treatments with a 7 % (v/v) aqueous solution of sodium hypochlorite for 15 min and four washes with sterile distilled water. The sterilization step was repeated with aqueous H_2O_2 (20 % v/v) and $HgCl_2$ (1 % m/v) solutions followed by four washing steps. Finally, seeds were placed on 30 ml solid MS (Murashige and Skoog 1962) medium supplemented with Murashige-Skoog vitamins, 0.8 % agar and 3 % sucrose, with 5–6 seeds per vessel. Cultures were grown at 23 ± 2 °C, with a day length of 16 h and a light intensity of 106 $\mu mol^{-2}s^{-1}$. Adult plants were further maintained in a growth chamber under optimal physiological conditions for *Solanum*: 25 °C day (11 h) and 15 °C night (13 h), 455 $\mu mol\ m^{-2}\ s^{-1}$ illumination intensity, and irrigated as required. DNA was isolated from 50 mg leaves using the technique of Walbot and Warren (1988).

Intron-targeting (IT) amplifications

PCR reactions targeting 29 nuclear intron loci (Poczai et al. 2010; Table 2) were performed in 10 µl volumes containing: 5 µl Nuclease Free Water, ca. 20 ng template DNA, 0.5 µM of each primer, 0.2 mM dNTP (Fermentas, Lithuania), 1 µl 10 × PCR buffer (1 mM Tris–HCl, pH 8.8 at 25 °C, 1.5 mM $MgCl_2$, 50 mM KCl and 0.1 % Triton X-100) and 0.5 U of DyNazyme II (Finnzymes, Finland) polymerase. All reactions were performed with the following settings with a MasterCycler ep96 (Eppendorf, Germany): 2 min at 94 °C for initial denaturation, 35 cycles of 30 s denaturation at 94 °C, 1 min annealing at optimal temperature, and 2 min extension at 72 °C, followed by a final extension for 5 min at 72 °C. Amplification products were separated on a 1.5 % agarose gel (GE Healthcare, UK) in 0.5 × TBE buffer (220 V, 0.5 h) and stained with ethidium-bromide. Each amplification contained one negative and one positive control, to check the reliability of the primers and the patterns produced. In many cases where the polymorphisms manifested as small size difference, the PCR products were separated on 2.5 % Metaphor agarose gels (Cambrex, East Rutherford, NJ).

Genetic diversity analysis

Resulting banding patterns were scored as binary characters and used for further analyses. Genetic variability and genetic differentiation for each species was assessed by calculating the number of polymorphic bands (*PB*) and the percentage of polymorphic bands (*PP*) for the IT data set. A locus was considered polymorphic when the band was present at a frequency

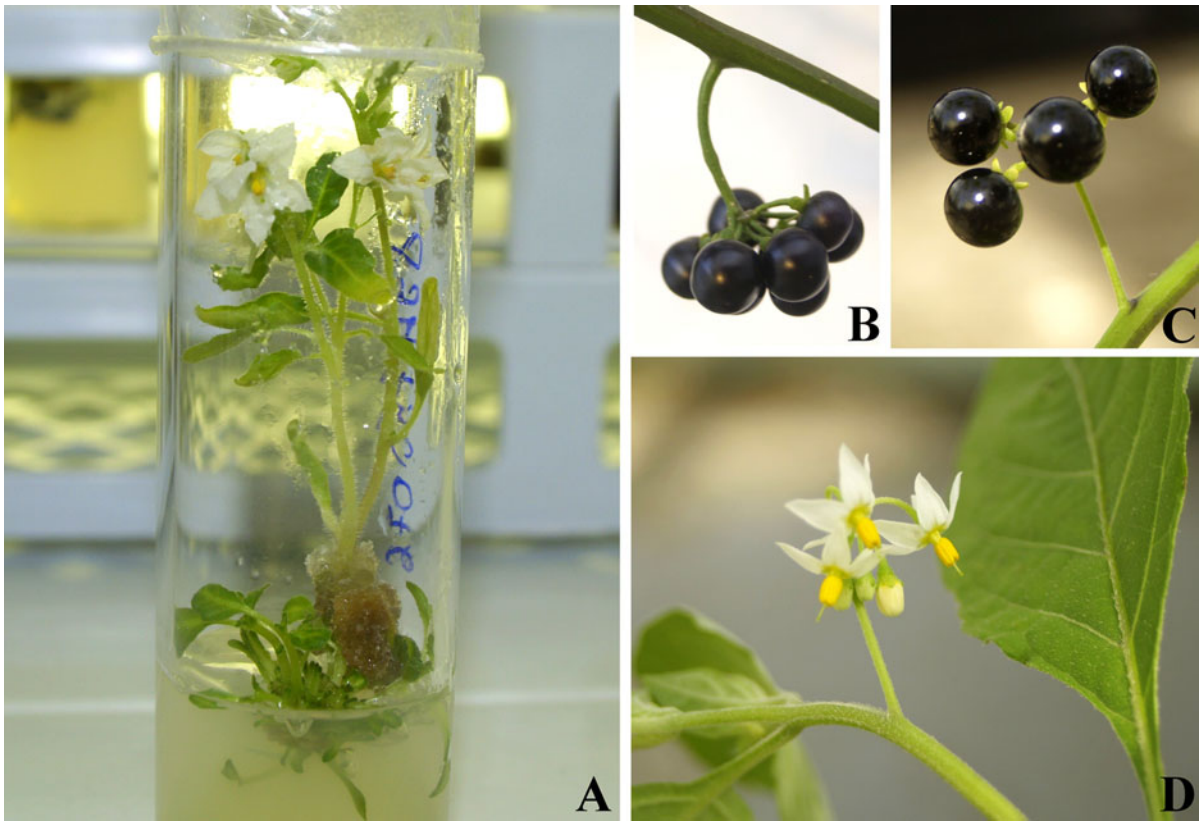

Fig. 1 **A** In vitro culture of *S. physalifolium*. The approximately equal length and width of the petals is apparent, with the distinct *brown–yellow* basal star. **B** Deep purple and broadly ovoid berries (15–17 mm) of *S. scabrum* with opaque cuticles; **C** the globose black berries of *S. americanum* with shiny opaque cuticles. The sepals are reflexed away from the mature berries found on ca. 14 mm long erect and splayed pedicels; **D** simple umbellate 6-flowered cyme of *S. americanum*. Corollas have a *yellow–green* translucent basal star, while the styles are usually exserted beyond the anther cone ca. 2.5 mm

between 5 and 95 %. Further matrices such as the observed number of alleles (n_a) and the effective number of alleles (n_e) were also calculated according to Kimura and Crow (1964) for each species. The lower bound for the actual number of alleles in the each species was defined as the effective number of alleles (n_e), which is equivalent to the inverse of the homozygosity. Shannon's information index (I) of phenotypic diversity (Lewontin 1972) was also calculated from the data set together with expected heterozygosity (H_E) or Nei's gene diversity (Nei 1973). The latter was calculated from allele frequencies based on the square root of the frequency of the null (recessive) allele. All calculations were done within all analyzed species and among all taxa representing an overall statistic measure with the program POPGENE v1.32 (Yeh et al. 1997).

Bayesian estimates of genetic diversity

We obtained genetic diversity estimates by using a Bayesian approach developed by Holsinger (1999) and implemented in the program HICKORY v.1.0.4 (Holsinger and Lewis 2003). This method does not require the assumption of Hardy–Weinberg equilibrium (HWE) for the genotypes and it takes full advantage of the information provided by molecular markers. The posteriori distribution of θ^B (analogous to Wright's F_{st}) was approximated through a Markov Chain Monte Carlo (MCMC) simulations using default parameters (burn-in = 5,000; sample = 25,000; thin = 5). We also attempted to estimate the panmictic heterozygosity, or total heterozygosity (H_T), based on mean allele frequencies and the average panmictic heterozygosity (H_S). The Bayesian

Table 1 Information about black nightshade accessions (*Solanum* section *Solanum*) used for intron-targeting analysis

Nr.	Accession number	Accession code	Taxon	Received	Origin	Voucher[a]
	Ingroup					
1	HUGEO09022	Ame01	*S. americanum*	G	Australia	UPG0001
2	HUGEO09020	Ame02	*S. americanum*	G	Hungary	UPG0002
3	HUGEO09023	Ame03	*S. americanum*	G	Australia	UPG0003
4	904750023	Ame04	*S. americanum*	N	Unknown	JM Edmonds
5	HUGEO09024	Ame05	*S. americanum*	G	Australia	UPG0005
6	954750174	Ame06	*S. americanum*	N	Unknown	JM Edmonds
7	954750178	Ame07	*S. americanum*	N	Unknown	P Poczai
8	954750184	Ame08	*S. americanum*	N	Unknown	P Poczai
9	HUGEO09021	Ame09	*S. americanum*	G	Hungary	UPG0006
10	884750042	Che01	*S. chenopodioides*	N	Switzerland	UPG0007
11	HUGEO09025	Che02	*S. chenopodioides*	G	France	UPG0008
12	UHBG211-1470	Che03	*S. chenopodioides*	H	Germany	UPG0009
13	HUGEO09005	Che04	*S. chenopodioides*	G	Hungary	UPG0010
14	HUGEO09026	Che05	*S. chenopodioides*	G	Romania	UPG0011
15	HUGEO09027	Che06	*S. chenopodioides*	G	Hungary	UPG0012
16	954750185	Che07	*S. chenopodioides*	N	Australia	DE Symon
17	A14750051	Che08	*S. chenopodioides*	N	Australia	DE Symon
18	904750124	Che09	*S. chenopodioides*	N	Australia	DE Symon
19	HUGEO06004	Nig01	*S. nigrum*[b]	G	Hungary	UPG0013
20	HUGEO06005	Nig02	*S. nigrum*[c]	G	Hungary	UPG0014
21	HUGEO06006	Nig03	*S. nigrum*[c]	G	Hungary	UPG0015
22	HUGEO09028	Nig04	*S. nigrum*[c]	G	Croatia	UPG0016
23	HUGEO09029	Nig05	*S. nigrum*[c]	G	Romania	UPG0017
24	HUGEO09030	Nig06	*S. nigrum*[b]	G	Italy	UPG0018
25	824750016	Nig07	*S. nigrum*[c]	N	Italy	GM van der Weerden
26	824750029	Nig08	*S. nigrum*[b]	N	Germany	P Poczai
27	834750011	Nig09	*S. nigrum*[b]	N	Netherlands	P Poczai
28	884750070	Nig10	*S. nigrum*[c]	N	Belgium	P Poczai
29	884750223	Opa01	*S. opacum*	N	Unknown	UPG0019
30	HUGEO09015	Opa02	*S. opacum*	G	Australia	UPG0020
31	HUGEO09016	Opa03	*S. opacum*	G	Australia	UPG0021
32	HUGEO09017	Opa04	*S. opacum*	G	Australia	UPG0022
33	HUGEO09018	Opa05	*S. opacum*	G	Australia	UPG0023
34	HUGEO09019	Opa06	*S. opacum*	G	Australia	UPG0024
35	894750076	Phy01	*S. physalifolium*	N	Germany	UPG0025
36	HUGEO09010	Phy02	*S. physalifolium*	G	Hungary	UPG0026
37	HUGEO09011	Phy03	*S. physalifolium*	G	Hungary	UPG0027
38	HUGEO09012	Phy04	*S. physalifolium*	G	Romania	UPG0028
39	HUGEO09013	Phy05	*S. physalifolium*	G	Romania	UPG0029
40	HUGEO09014	Phy06	*S. physalifolium*	G	Slovakia	UPG0030
41	PI63475502SD	Ret01	*S. retroflexum*	U	USA	UPG0031
42	HUGEO09006	Ret02	*S. retroflexum*	G	USA	UPG0032
43	HUGEO09007	Ret03	*S. retroflexum*	G	Australia	UPG0033
44	HUGEO09009	Ret04	*S. retroflexum*	G	South Africa	UPG0034

Table 1 continued

Nr.	Accession number	Accession code	Taxon	Received	Origin	Voucher[a]
45	904750228	Ret05	*S. retroflexum*	N	Australia	UPG0035
46	HUGEO09008	Ret06	*S. retroflexum*	G	Australia	UPG0036
47	824750011	Sca01	*S. scabrum*	N	Unkonwn	UPG0037
48	Grif1419801SD	Sca02	*S. scabrum*	U	USA	UPG0038
49	HUGEO09004	Sca03	*S. scabrum*	G	Uganda	UPG0039
50	UHBG211-1465	Sca04	*S. scabrum*	H	Germany	UPG0040
51	HUGEO09002	Sca05	*S. scabrum*	G	South Africa	UPG0041
52	HUGEO09003	Sca06	*S. scabrum*	G	South Africa	UPG0042
53	HUGEO09031	Vil01	*S. villosum*[d]	G	Hungary	UPG0043
54	HUGEO09032	Vil02	*S. villosum*[d]	G	Romania	UPG0044
55	HUGEO09033	Vil03	*S. villosum*[e]	G	Hungary	UPG0045
56	804750186	Vil04	*S. villosum*[e]	N	Unknown	UPG0046
57	HUGEO09034	Vil05	*S. villosum*[d]	G	Unknown	UPG0047
58	HUGEO09035	Vil06	*S. villosum*[d]	G	Unknown	UPG0048
59	814750090	Vil07	*S. villosum*[e]	N	France	MLK Manoko
60	954750158	Vil08	*S. villosum*[d]	N	Unknown	P Poczai
61	954750176	Vil09	*S. villosum*[d]	N	Unknown	P Poczai
	Outgroup					
62	S001/2009	Dul01	*S. dulcamara*	V	Hungary	UPG0049
63	White Lady	WL1	*S. tuberosum*	C	Hungary	I Cernák

C Potato Research Centre, University of Pannonia

G Georgikon Botanical Garden, University of Pannonia, Keszthely, Hungary

H Botanical Garden of the University of Hohenheim, Stuttgart, Germany

N Botanical and Experimental Garden of the Radboud University Nijmegen, The Netherlands

U National Plant Germplasm System (NPGS), Plant Genetic Resources Conservation Unit, Griffin, Georgia, United States

V Botanical Gardens, Institute of Ecology and Botany, Hungarian Academy of Sciences, Vácrátót, Hungary

[a] University of Pannonia, Georgikon Faculty Voucher System of the Dept. of Plant Science and Biotechnology

Interspecific taxon marks:

[b] subsp. *shultesii*

[c] subsp. *nigrum*

[d] subsp. *villosum*

[e] subsp. *miniatum*

estimates of genetic diversity were calculated using four models: (1) *Full* model (with non-informative priors for f and θ^{B}); (2) $f = 0$ (assuming no inbreeding); (3) $\theta^{B} = 0$ (assuming no population structure); (4) f free (allowing the incorporation of uncertainty about f into the analysis). The model choice was based on the deviance information criterion (*DIC*), which combines a measure of model fit (*Dbar*) with the measure of model complexity (*pD*).

Selection of a subset of intron-targeting markers for quick genotyping

We calculated a number of parameters to select a small number of markers (core set) that can effectively differentiate the analyzed accessions. We calculated the polymorphism information content (PIC) according to Botstein et al. (1980) and the heterozygosity level (H; Liu 1998) for each primer using PICcalc (Nagy et al. 2012). To further assess the capacity of a

Table 2 Details of intron-targeting markers used in this study

Primer	Primer sequence (5′–3′)	Size	T_a	Putative function	No. of bands and size ranges
Adk-242	F: TGCTTTTAAAGTCGCACCA R: TTATATCCGGAGCATGTCCAC	242	55	Adenylate kinase gene	4 (112–242)
Adk-795	F: GCATGGTTCTTTCCTTCCTG R: TGGGCCAGGAATTTTGCTATC	795	54	Adenylate kinase gene	4 (385–805)
Cat-232	F: AGGAGGCGGATCTAGCCTTA R: TGTCAAGAAAGGGGTGTCGT	232	55	Potato catalase gene	2 (210–232)
Cat-260	F: TGACAACAAATGCTGGTGGT R: AAGGTGGCAAGCTTCTCAAT	260	53	Potato catalase gene	3 (260–310)
GPSS-275	F: CTTTTGATGGGGGCAGATTA R: CAGCTTCCTGTCAGCATCAG	275	53	ADP-glucose pyrophosphorylase small-subunit gene	3 (125–275)
GPSS-943	F: TCATTGGTGAAGGTTGTGTGA R. ACCACGGAATGGTGAATCTT	943	53	ADP-glucose pyrophosphorylase small-subunit gene	7 (150–1043)
INHWI-509	F: TGAAACTCTCTTGGCACGAA R: TTCTGGCCACCTTTGTTTTC	509	54	Wound-inducible proteinase inhibitor I gene	4 (205–509)
INHWI-545	F: TCAAGTTTGCTCACATCTTGT R: TCGTGCCAAGAGAGTTTCAA	545	54	Wound-inducible proteinase inhibitor I gene	4 (425–685)
InvG-220	F: ACAGGAATCACACCTGCACA R: TCTGCACCCTTAAGTCCACA	220	54	Invertase gene	3 (150–280)
InvG-262	F: TTCTCATGTGCTCAGATGCT R: GAGGGCTTGTACATTGACTTCA	262	53	Invertase gene	3 (182–262)
InvG-393	F: TGGTTACCATTTTCAGCCAGA R: CCATTGAAATACATTGGTGCTG	393	55	Invertase gene	3 (320–393)
LBr-G9	F: TGGATCTGAAGATGGCACTG R: TTGCTCTCAAATCCCACACA	652	55	Transducin family protein	3 (652–1100)
LBr-4D6	F: GAGTATTCATTCGGGCTTGG R: CTCTACCGACCCGTAGCAAG	196	54	Plastidic ATP/ADP-transporter protein	4 (196–597)
PatI-433	F: TCAAGCTCGTCATTCACAAAA R: TCAGACGCATCATCCATTTC	433	54	Potato patatin class I. gene	3 (285–510)
PatI-838	F: CGAACATGGCCCTCATATTT R: TGCACACGAGTTTCTCCAAG	838	53	Potato patatin class I. gene	5 (145–938)
Pat-In3	F: CAGAAAGTTGCCATCTCAAGC R: GCTGCTGCTGTGGAATAACA	581	53	Potato patatin gene Intron 3	4 (220–595)
Poni1a-718	F: GGTGGTGGTGGTAGCTCAGT R: CCCARRGGCATTAACACTCC	718	55	Potato membrane protein	3 (718–850)
Poni1a-442	F: TTTGCCTCGGAACTCTTCAG R: GCCTCAGAAGCAAAGCAAAT	442	54	Potato membrane protein	3 (296–510)
Ry1-In3	F: AATGCAGAAGGTGCAACGAT R: TGGGCGAAATTTCATTAACA	199	54	Ry1 resistance gene-like function protein intron 3	2 (199–450)
Ry1-In4	F: TCGAAAAATTCTCAAATGCAAA R: GATTGCTTCGATAGCCTTGG	477	55	Ry1 resistance gene-like function protein intron 4	6 (125–590)
Ry1-In5	F: CCAGCAGAGTTCACTGTTTCA R: GTTGCACAGCTGCTGAGAT	481	54	Ry1 resistance gene-like function protein intron 5	2 (481–501)
Ry1-In6	F: GCTCTCGTCTCCACTTCTGC R: AACTCCTCAGCAACTGCACA	741	56	Ry1 resistance gene-like function protein intron 5	7 (110–751)
S2-317	F: CGGCCAGTTACAATTCTGC R: AATCCAGTGGTGGTCCAGAG	202	55	Self-incompatibility locus linked stylar Rnase gene	6 (202-455)
STAC1-226	F: GTCTTCCCGCTTTCAAAGAT R: TCAGCAAGGCAAAACATGAG	226	53	1-Aminocyclo- propane-1-carboxylate synthase gene	5 (135–295)

Table 2 continued

Primer	Primer sequence (5′–3′)	Size	T_a	Putative function	No. of bands and size ranges
Suc16-321	F: TCACCGCAATGAGATACTGC R: TATCCCTTTTCCGTGGCTTT	321	54	Sucrose synthase gene	4 (270–385)
Suc16-349	F: TGACGTTGAGAATGACGAACA R: CCAACCTTGCCATTGTGAAT	349	55	Sucrose synthase gene	4 (275–605)
UBQ-627	F: TCTCAATTGCCTTCAATTTCTC R: TCCGGTGAGAGTTTTCACAA	627	53	Polyubiquitin gene	6 (520–715)
Ure-242	F: TGCTTTTAAAGGTCGCACCA R: TTATATCCGGAGCATGTCCAC	242	56	Ure gene for Urease	4 (242–257)
Ure-271	F: GAGCAGCCACGAGATTTGA R: CACAAATCAATGCCCAAGC	271	55	Ure gene for Urease	4 (271–451)

In the table forward and reverse primer sequences are shown with the applied annealing temperatures (T_a), size of the expected band in potato (*Solanum tuberosum*) and its predicted putative function together with the number of scored bands and their size range.Polymorphism information content (PIC) is calculated for each primer across all samples

given primer to distinguish among various genotypes we also calculated the resolving power (Rp) described by Prevost and Wilkinson (1999). The PIC and H values were used to determine the information of a primer to detect polymorphism within the analyzed accessions, while Rp was used as a quantification tool to assess the probability that two randomly chosen individuals have different patterns.

Phylogenetic analyses

Parsimony analysis

We performed phylogenetic analyses with parsimony as an optimality criterion using the program Nona (Goloboff 1994) within a winclada (Nixon 2002) shell. We performed six separate analyses (using processor time as a seed to randomize the order of the terminals) with the following settings: hold 100 (holding defined number of trees), 50 replications (search performed with multiple tree-bisection-reconnection algorithm mult*max*), hold/1 (keeping 1 starting tree for each replication). In addition, we performed also larger analyses by holding up to 3,000, 30,000 and 300,000 trees (hold 3,000, 30,000 and 300,000), 100, 1,000 and 10,000 replications respectively, and keeping 20 or 30 starting trees for each replication (hold/20 or 30). In addition we performed analyses using also parsimony ratchet (Nixon 1999) in two additional analyses with the following settings: 1000 replicates, two trees hold per iteration, 23 characters (ca. 20 %) reweighted, and with amb-poly = (default setting of Nona: if any of the reconstructed states are shared between ancestor and descendant node, the branch is collapsed). Another ratchet analysis was performed with the same settings but with an increased number of trees held per iteration (hold/10), and with 35 characters (ca. 30 %) reweighted. Jackknife (Farris et al. 1996) support values were calculated using 1,000 replications, with 10 search replications (multi*10) and with 1 starting tree per replication (hold/1).

Bayesian analysis

Bayesian analysis was performed with MrBayes v3.2 (Huelsenbeck and Ronquist 2001) applying the binary model with gamma-shaped distribution of across-site rate variation (*lset rates = gamma*). Posterior probability (PP) distributions were created using the Metropolis-coupled Markov chain Monte Carlo $(MC)^3$ method. The binary (restriction) site model implemented in MrBayes employs a simple F81-like model that offers a number of options to correct for coding bias. We used the option variable (*lset coding = variable*) that may closely fit the intron-targeting (IT) dataset. The detection of indels revealed by IT is typically depending on the detection of some sequence length variation in the targeted site. Therefore, neither all the absence (0) nor all the presence (1) characters can be observed. The character frequencies π_0 and π_1 in the binary model would represent the rate at which insertions and deletions occur. We attempted

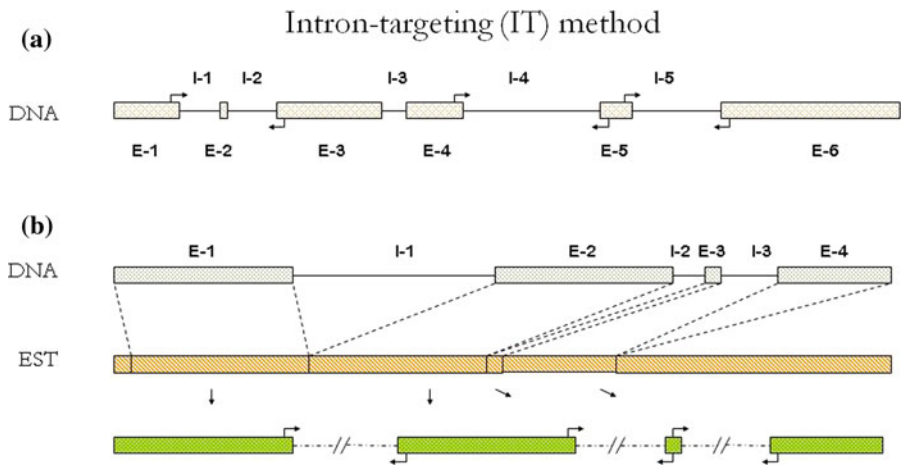

Fig. 2 **a** Outline of intron-targeting (IT) markers with primers flanking the 5′ and 3′ splice site of the exons. **b** *Schematic* representation of IT markers designed from ESTs. ESTs are compared with known genes from genome databases to find putative exon and intron sequences. After locating the precise positions of the introns, the aligned sequences of the flanking exons are used for primer deign. *Colored boxes* indicate exons, while (*dashed*) *lines* introns; *arrows* attached to the *colored boxes* indicate the primers

Table 3 Genetic diversity statistics calculated under Hardy–Weinberg equilibrium for black nightshade (sect. *Solanum*) accessions based on intron-targeting (IT) markers

Species	*N*	*PB*	*PP* (%)	n_a	n_e	*I*	H_E
S. americanum	9	24	20.87	1.21 (0.41)	1.11 (0.25)	0.10 (0.21)	0.06 (0.14)
S. chenopodioides	9	37	32.17	1.32 (0.46)	1.17 (0.30)	0.15 (0.25)	0.10 (0.17)
S. nigrum	10	32	27.83	1.28 (0.45)	1.13 (0.27)	0.13 (0.22)	0.08 (0.15)
S. opacum	6	30	26.09	1.27 (0.44)	1.13 (0.27)	0.12 (0.22)	0.07 (0.15)
S. physalifolium	6	20	17.39	1.17 (0.38)	1.11 (0.25)	0.09 (0.22)	0.06 (0.15)
S. retroflexum	6	12	10.43	1.10 (0.30)	1.07 (0.21)	0.06 (0.19)	0.04 (0.12)
S. scabrum	6	34	29.57	1.30 (0.45)	0.16 (0.30)	0.14 (0.24)	0.09 (0.16)
S. villosum	9	31	26.96	1.26 (0.44)	0.14 (0.30)	0.12 (0.23)	0.08 (0.16)
Total	61	105	91.3	1.91 (0.28)	0.36 (0.30)	0.37 (0.20)	0.24 (0.15)

Standard deviations (SD) are indicated in parentheses

ID sample identification code used in the study, *N* number of samples taken from the species, *PB* the number of polymorphic bands, *PP* (%) the percentage of polymorphism, n_a the observed number of alleles, n_e the effective number of alleles, *I* Shannon's information index of phenotypic diversity, H_E expected heterozygosity or Nei's gene diversity

to sample all trees that have a reasonable probability given the assembled datasets. For this reason we kept all sequences in the alignments even if they had identical sequences. This was based on population genetic reasons not to mislead the BI assuming a larger population. Analyses were initiated with four runs and eight chains (2×10^6 generations each) sampling every 1,000th generation. Simulations were run until it was necessary to reach stationarity assessed from the average standard deviation of split frequencies <0.01. MC^3 convergence was explored by examining the Potential Scale Reduction Factor (PSRF) for all parameters in the model and plots of log-likelihoods over time together with other plots for all parameters allowed by Tracer v1.4 (Rambaut and Drummond 2007). Additional tests of convergence were conducted with the online program AWTY (Nylander et al. 2008) using the 'cumulative' and 'compare' functions. The states of the chains sampled before stationarity (split freq> 0.01) were discarded as burn-

Table 4 DIC statistics for intron-targeting data of black nightshade accessions under four models

Model	*Dbar*	*Dhat*	*pD*	DIC
Full	930.65	684.725	245.93	1176.59
$f = 0$	924.79	650.87	273.913	1198.70
$\theta^B = 0$	5174.77	5076.65	98.12	5272.89
f free	957.22	704.692	252.52	1209.75

DIC is the Deviance Information Criterion is similar to Akaike's Information criterion and shows how a particular model fits to the dataset. *Dbar*, is the posterior mean of the deviance; *Dhat*, is the deviance at the posterior mean of the parameters and *pD*, is the effective number of parameters. (See text for further explanation)

in (25 %). Trees from BI analyses were summarized as majority-rule consensus trees and edited with TreeGraph2 (Stöver and Müller 2010).

Maximum likelihood analysis

Phylogenetic analyses were performed with RAxML 7.2.6 (Stamatakis 2006) using maximum likelihood (ML) as optimality criteria via raxmlGUI 0.93 (Silvestro and Michalak 2012). An ML + thorough bootstrap ($-b$) search was performed with 1,000 replicates and 100 runs using a BINGAMMA evolutionary model implemented for binary data. Bootstrap support values were drawn on the most likely ($-f\,b$) majority rule consensus tree. Throughout this paper, 70–84 % bootstrap support is considered moderate and 85–100 % as strong support.

Results

Intron-targeting amplifications and genetic diversity

Insertion-deletion derived markers are considered to reveal polymorphism in genetic studies. Despite of the increasing molecular studies aiming to reveal the genetic diversity and phylogeny of sect. *Solanum* only few attempts have been made to analyze the value of nuclear indels as genetic markers in this section. In our study we revealed indel polymorphism in eight species of black nightshades (sect. *Solanum*) with primers previously designed based on Solanaceae EST and genomic databases. These primers were designed to amplify fragments of 110–1,100 bp flanking the intron–exon splice sites (Fig. 2) of 16 genes and size variation of the amplicons was detected by simple agarose gel electrophoresis on samples of 61 accessions of a core germplasm collection. One hundred and five loci (91.3 %) were polymorphic among the total accessions. The number of polymorphic bands (PB) varied between 12 and 37 with the lowest value (10.43 %) detected in *S. retroflexum* and the highest (32.17 %) in *S. chenopodioides*. Population genetic parameters, namely, the observed number of alleles, number of effective alleles, Shannon's information index, and expected heterozygosity are shown in Table 3. All of these calculated indices of genetic diversity confirm that variability was low within populations, but the species were well differentiated.

Table 5 Genetic diversity statistics from Bayesian inference not requiring Hardy–Weinberg proportions based on the $f = 0$ model

Species	H_T			H_S			θ^B		
	Mean	SD	CI	Mean	SD	CI	Mean	SD	CI
Solanum americanum	0.3469	0.0196	0.3088, 0.3855	0.1116	0.0088	0.0950, 0.1299	0.0931	0.0664	0.0028, 0.2378
S. chenopodioides	0.3334	0.0161	0.3010, 0.3654	0.1473	0.0095	0.1289, 0.1658	0.1107	0.0729	0.0038, 0.2562
S. nigrum	0.3180	0.0165	0.2857, 0.3511	0.1177	0.0081	0.1020, 0.1338	0.1198	0.0847	0.0034, 0.2851
S. opacum	0.3340	0.0220	0.2912, 0.3775	0.1399	0.0109	0.1186, 0.1616	0.0642	0.0480	0.0020, 0.1769
S. physalifolium	0.3897	0.0268	0.3367, 0.4402	0.1235	0.0101	0.1036, 0.1430	0.0160	0.0166	0.0003, 0.0609
S. retroflexum	0.4087	0.0307	0.3440, 0.4624	0.1057	0.0103	0.0852, 0.1264	0.0269	0.0292	0.0005, 0.1080
S. scabrum	0.3534	0.0208	0.3121, 0.3943	0.1392	0.0094	0.1214, 0.1582	0.0436	0.0345	0.0015, 0.1303
S. villosum	0.3234	0.0169	0.2905, 0.3570	0.1221	0.0084	0.1220, 0.1390	0.0871	0.0576	0.0043, 0.2134
Total	0.3077	0.0032	0.3012, 0.3140	0.1259	0.0042	0.1257, 0.1344	0.6065	0.0204	0.5654, 0.6466

Posterior means, standard deviation (SD), and upper and lower bounds of 97.5 % credible intervals (CI) of panmictic or total heterozygosity based on mean allele frequencies (H_T), average panmictic heterozygosity (H_S), and fixation index (θ^B), respectively. Values are shown in four decimal formats because of the observed small differences in some cases

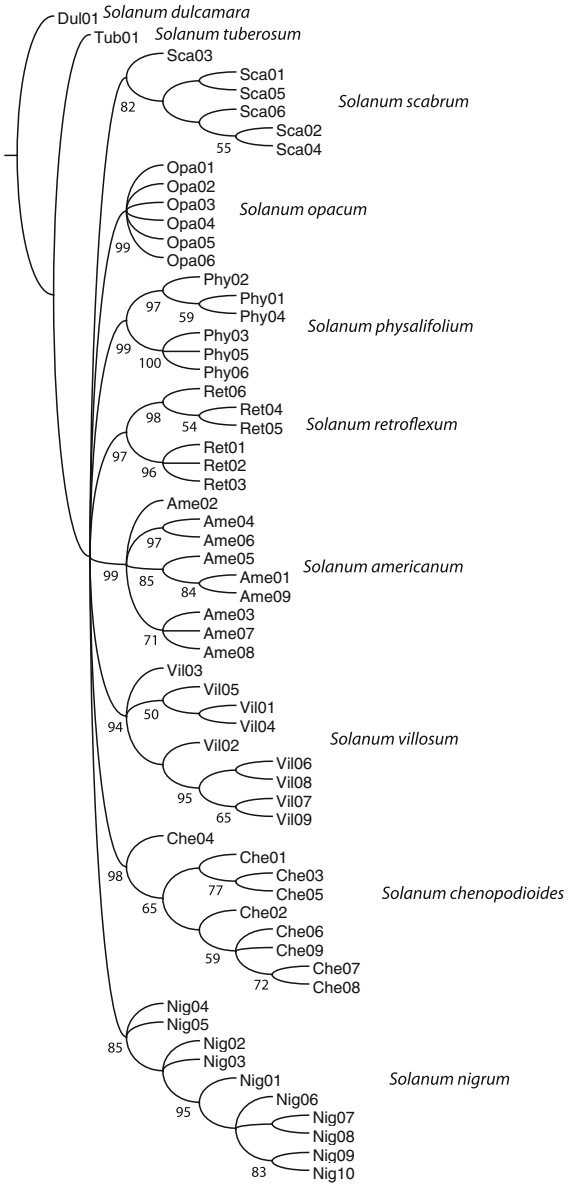

Fig. 3 The resulting strict consensus tree of the parsimony analysis. Numbers below branches represent jackknife support values

Within the species effective number of alleles (n_e) varied between 0.14 and 1.17, while the observed number of alleles (n_a) ranged between 1.10 and 1.32. The values of Shannon's information index (I) were similar to expected heterozygosity values, ranging from 0.06 to 0.15, while expected heterozygosity (H_E) averaged across all species ranged from 0.04 to 0.10 with an average value of 0.24.

The Bayesian analysis carried out with HICKORY showed that inbreeding is unlikely to be the major driving force in determining the genetic structure of the analyzed accessions. All inferences strongly rejected the $\theta^B = 0$ model, confirming that there is considerable structure in the dataset, indicating that the analyzed species are well differentiated. The cross-species amplification of IT markers was successful in terms of species identification. The DIC parameter was lowest in the *Full* model (1176.59) than in case of others f free (difference in DIC value = 33.16) or $f = 0$ (difference in DIC value = 22.11). The resulting DIC values indicated that the best model that fits our dataset is the *Full* model. Comparing other components of the DIC statistics led us to conclude that a simpler model with $f = 0$ should be preferred. As DIC = *Dbar* + *pD* and it combines the goodness-of-fit with the model complexity where *pD* = *Dbar*−*Dhat*. *Dbar* is −2 times the mean posterior likelihood and gives a measure how well the model fits to the data. The model with the lowest *Dbar* is that which best fits the data, this holds for the $f = 0$ model (inbreeding equals zero). The *Dhat* value, which is−2 times the log likelihood evaluated at the posterior mean, and gives a measure about how well the best point estimate fits the data was also the lowest for $f = 0$. This suggested caution with the *Full* model as the lowest DIC was entirely due to the smaller number parameters estimated (*pD*). As indicated by *Dbar* and *Dhat* values $f = 0$ fitted our dataset better (see Table 4) and was used in our estimations instead of the *Full* model that had the best DIC score. For the results derived from the $f = 0$ run with default settings, illustrating the accurate and precise estimates of θ^B see Electronic Supplementary Material 1. The lower panel shows that the sampler thoroughly explored the relevant parts of parameter space.

The Bayesian estimate of fixation index (θ^B) was slightly higher compared to heterozygosity calculations assuming Hardy–Weinberg proportions, ranging from 0.016 to 0.1198 with an total value of 0.6065 calculated across all samples (Table 5). The estimate of total heterozygosity (H_T) with the Bayesian approach (not assuming Hardy–Weinberg proportions) returned similar posterior distributions for all species ranging from 0.3180 to 0.4087 with an a value of 0.3077 across all accessions. These estimates for each species are much higher than those obtained under Hardy–Weinberg equilibrium, while the

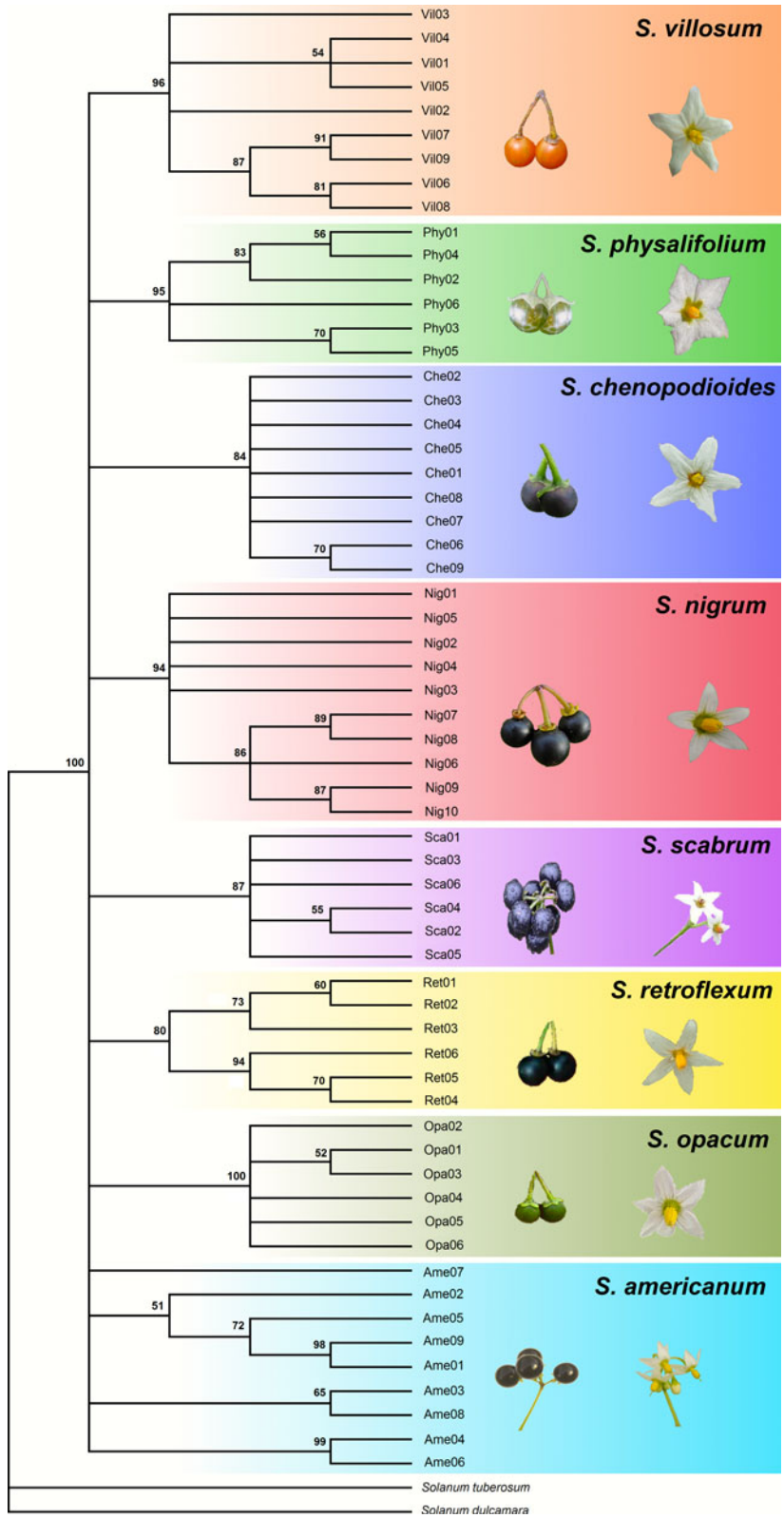

Fig. 4 The best scoring ($lnL = -1626.62$) maximum likelihood (ML) majority rule consensus tree for the analyzed *black nightshade* (sect. *Solanum*) accessions and outgroup taxa obtained from ITs. *Numbers* above branches are bootstrap values from 1,000 replicates. *Images* of *representative flowers and berries* of the analyzed species are shown

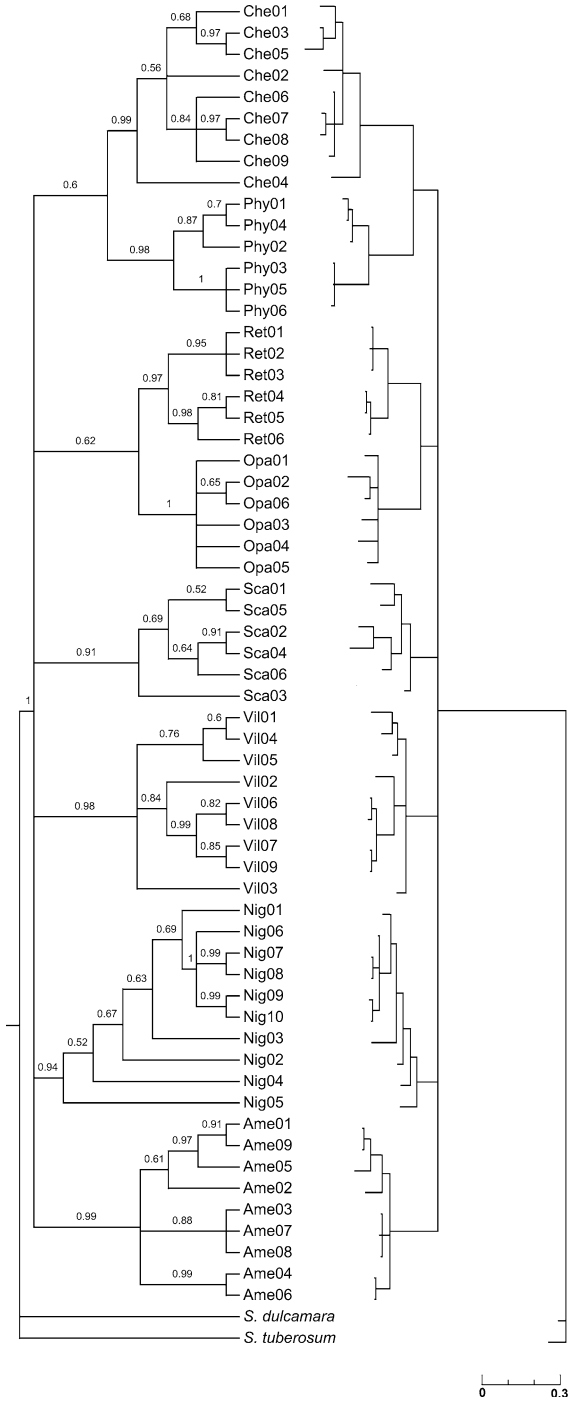

Fig. 5 Bayesian majority rule (50 %) consensus tree (*left*) and phylogram (*right*) from the analysis of IT dataset with MrBayes v3.2. Numbers above branches are posterior probabilities (PP)

Bayesian estimates among the whole population matched those obtained under the above mentioned criteria.

Table 6 Characteristics of intron-targeting markers

Primer	NSB	H	PIC	Rp	Order of Rp
Adk-242	1	0.82	0.79	0.99	16
Adk-795	1	0.86	0.84	0.57	20
Cat-232	1	0.68	0.62	1.76	5
Cat-260	1	0.80	0.78	1.49	11
GPSS-275	0	0.87	0.86	1.73	6
GPSS-943	1	0.92	0.91	3.32	1
INHWI-509	1	0.85	0.83	1.89	4
INHWI-545	4	0.80	0.77	1.27	12
InvG-220	2	0.71	0.66	0.27	29
InvG-262	0	0.75	0.71	0.57	21
InvG-393	1	0.80	0.77	1.22	13
LBr-G9	3	0.73	0.68	0.48	24
LBr-4D6	1	0.82	0.80	0.85	18
PatI-433	0	0.79	0.77	1.21	14
PatI-838	0	0.87	0.85	1.51	10
Pat-In3	3	0.81	0.79	0.54	22
Poni1a-718	2	0.74	0.70	0.45	26
Poni1a-442	1	0.73	0.69	0.44	27
Ry1-In3	3	0.62	0.54	1.19	15
Ry1-In4	1	0.88	0.87	0.46	25
Ry1-In5	2	0.69	0.64	2.06	3
Ry1-In6	3	0.92	0.91	1.69	7
S2-317	2	0.86	0.85	0.62	19
STAC1-226	2	0.87	0.83	0.86	17
Suc16-321	0	0.85	0.86	1.68	9
Suc16-349	1	0.85	0.83	2.83	2
UBQ-627	2	0.90	0.89	1.69	8
Ure-242	4	0.80	0.77	0.53	23
Ure-271	4	0.79	0.76	0.40	28

NSB number of synapomorphic bands, *H* heterozygosity, *PIC* polymorphism information content, *Rp* resolving power

Phylogenetic analyses

Comparison of the results of the parsimony analyses (Fig. 3) reveals that already the first analysis with only 100 repetitions found 100 equally parsimonious trees (EPTs) with the length of 435 steps, CI (Kluge and Farris 1969) 0.25 and RI (Farris 1989) 0.78. The consensus of the 100 optimal trees found in this analysis, as compared with the consensus of the trees of all analyses, differs only in the presence of one basal clade within *Solanum nigrum*. As noted by Farris et al. (1996) it is not necessary to find *all* parsimonious trees. The tree "space" can be considered to be

sufficiently sampled if no change in the consensus tree can be detected. In our case already the second analysis with 1,000 repetitions and 353 EPTs resulted in a consensus that was identical with the consensus based on the trees of all analyses.

The most likely tree (Fig. 4) of the maximum likelihood (ML) analysis reached a log likelihood score—1,626.62. This tree was consistent with the tree obtained with parsimony and Bayesian analyses. The results of the four BI runs conducted with MrBayes were highly congruent with each other. PSRF values averaged 1.0000–1.0001 strongly suggesting that stationarity had been reached. The comparison of the topologies and associated posterior probability values obtained across the independent runs for the IT dataset with AWTY verified convergence. The inspection of the log likelihood trace plot and the comparison of each run also showed that the runs reached stationarity. Example trace plots comparing four runs with each other and bivariate plots created with the compare function of AWTY are shown in Electronic Supplementary Material 2. Runs with the IT dataset reached stationarity after 1.2×10^6 generations. The Bayesian majority rule consensus tree is illustrated in Fig. 5.

All analyses resolved the species included in our study in separate well supported groups. The exception was the *S. americanum* clade which collapsed in the ML analysis. This might be due to the taxonomic dispute surrounding this species and its relationship with *S. nodiflorum* Jacq (see Manoko et al. 2007). The Bayesian analysis, however, supported a separate *S. americanum* clade with strong posterior evidence (PP 0.99). The phylogenetic analyses provided only weak support for the genetic relationships of the black nightshade accessions analyzed. *Solanum chenopodioides* and *S. physalifolium* were united in the BI tree in a very weakly supported clade and the same is true for *S. retroflexum* and *S. opacum*.

Selection of a core set of intron-targeting markers

The identification of a subset of IT markers, which can quickly separate black nightshade accessions, would be useful for reducing the costs of genotyping. We calculated various indices, which can help to select the most suitable primers for such experiments. The resolving power of the primers ranged from 0.27 to 3.32, while PIC and H values varied from 0.54 to 0.91 and 0.62 to 0.92, respectively (Table 6). Theoretically even one primer with the highest Rp value could differentiate all analyzed species (10.2), based on the equation $0.15x + 1.78 = Rp$ described by Prevost and Wilkinson (1999), where x in our case would define the number of analyzed species. Nevertheless, there might be primers with high ranking Rp values that are unable to distinguish all species when used solely; this was also described by Prevost and Wilkinson (1999). Therefore we suggest the use of at least three to five intron-targeting markers with high Rp values to separate sect. *Solanum* species if the analysis is aimed at intraspecific level. However, at least twelve primers, with the highest Rp values, should be used to distinguish all accessions. However, even when the entire indel primer set was used we failed to distinguish accessions as for example with *S. opacum*. On the parsimony consensus tree (Fig. 3) also clades of *S. americanum* or *S. physalifolium* were totally unresolved.

Discussion

Intron-targeting markers are less polymorphic but reveal high interspecies differentiation

Introns have long been considered as a source of polymorphism due to their moderate sequence evolution, which is presumed to take place under minimal constraints in a way consistent with the neutral theory of sequence evolution (Kimura 1983). Recent reports have shown intron length polymorphism to be a convenient and reliable source of information with high interspecies transferability. The intron-targeting markers can originate from either multiple or single loci depending on the features of the targeted regions. Another important feature of IT is that primers can be designed from both genomic and EST databases. These primers may generate amplicons from any gene or gene family corresponding to conserved exon sequences flanking the introns. The close proximity of introns to exons makes them well suited for the detection of length polymorphism in their structure that can be utilized for various purposes. However, development of such markers depends on the availability of genomic databases with several target sequences for IT markers. Functional gene characterization might be a limiting factor, since it is not

possible to establish functions for all genes. The crucial question is whether useful allelic variation can be identified for all genes of, for example, ecological relevance in the targeted organism.

The present primer set of IT markers differentiated all the eight analyzed species of sect. *Solanum*, but in some cases they failed to distinguish accessions from the same species. A better differentiation may be obtained with more IT markers, but the low mean number of alleles per locus may characterize the limitation of this method compared to other techniques based on multi-locus approaches, e.g., targeting fingerprinting markers (see Poczai et al. 2013). However, intron-targeting markers may be less polymorphic at intraspecific level (see Table 3 and 5) but they proved to be useful at interspecies differentiation. This finding is confirmed by both phylogenetic analyses and the population genetic diversity analyses presented here. Intron-targeting markers may fail to identify variation in highly inbred species, or in cases where the species have experienced severe genetic bottlenecks followed by range expansions and rapid dispersal. This may be the case with cultivar groups that are based on limited genetic diversity. The recent study of García-Lor et al. (2012) comparing indels markers and SSRs in the genus *Citrus* L. also showed similar results. They concluded that indel markers are better phylogenetic markers than SSRs with lower level of homoplasy and are suitable to trace the genomic contribution of three ancestral *Citrus* species, while SSRs are better for intraspecific diversity analysis. Indel markers could be useful for the assignment of new solanaceous germplasm to taxonomic groups or to identify species. The results obtained with the IT method are comparable with those reached with other molecular markers like RAPDs that have been used to study genetic variation among accessions of sect. *Solanum*, or with others using similar DNA fingerprinting markers like AFLP or SSR (Dehmer and Stracke 1999; Dehmer 2001; Jacoby et al. 2003; Dehmer and Hammer 2004; Olet et al. 2005; Manoko et al. 2008). The successful transferability and cross-species amplification capacity of IT markers, however, may depend on the conservation of exon–intron junctions and gene structures across related genomes in different species. It seems that the shared syntenies of the presently targeted 16 different genes as well as their sequence features are relatively conserved. This allowed easy transfer of primers between species developed for potato (*S. tuberosum*) of the sect. *Petota* to black nightshades of the sect. *Solanum*. This phenomenon may be valuable for generating functional markers directly related to gene regions and to facilitate the discovery of specific markers linked to a given phenotype. Indels play a major role in sequence divergence between closely related DNA sequences in plants (Xu et al. 2009). Studies of human genome suggest that they are also responsible for other genetic changes, gene defects or even diseases (Britten et al. 2003). In plants this could be useful in studies of resistance to pests or different diseases as indels in coding regions may have functional role also in plant genomes, or they can contribute to evolutionary changes (García-Lor et al. 2012). Intron-targeting markers showing functional diversity among the amplified alleles could be useful for genetic or QTL mapping as shown by Gorji et al. (2012). It is also possible to tag specific genes related to environmental factors that could have useful applications, for example in molecular ecology. This is because IT uses primers based on allele sequences of functionally characterized genes, and thus specific banding patterns corresponding to plant phenotypes can be identified.

The polymorphism and genome coverage of intron-targeting markers

With the present primer set we surveyed over 12,672 bp for insertion-deletion polymorphism regarding the size of the expected amplicons based on GenBank data. According to this data the frequency of indels per kb in black nightshade accessions was 0.11. This is considerably lower than frequencies reported for other species, 0.43 in *Zea* L. (Ching et al. 2002); 0.45 in *Brassica* L. (Park et al. 2010) or 0.60/kb in *Cucumis* L. (Morales et al. 2004). The mean alleles (bands) per locus were 3.96 with a maximum of seven for the GPSS-943 and Ry1-In6 locus, this being higher than reported by the previously mentioned studies. This may be attributed to the fact that genomes in Solanaceae have undergone fairly few rearrangements and duplications and gene structures are relatively conserved (Mueller et al. 2005). However, in a similar study with indel markers in tomato (*S. lycopersicum*) and related species revealed 0.74 indels per kb (Wang et al. 2010) indicating that our findings may be restricted to sect. *Solanum*. The genomes of analyzed

species may be highly similar or very closely related and it is very likely that intron variability may increase with taxonomic distance. However, the size of the conserved intervening sequences amplified by intron-targeting can be highly variable. Plant specific gene families, e.g., homeobox genes that function as transcription factors with a unique homeodomain (KNOX) could reveal more polymorphism than those which are more conserved. Intron-targeting markers could combine reliability and reproducibility with easy access to the generated raw data. Banding patterns are based on length polymorphism, which requires no further laboratory treatment. However, novel primer design may become problematic if genomic annotations of conserved sequences are missing.

Once they are developed they might be used for further analysis besides indel screening. Although our study did not address the utility of IT amplicons as a source of SNP markers further downstream applications may reveal potential of the marker system as a source of nucleotide polymorphism. Many loci investigated have proved this potential in other studies. A good example could be the sucrose synthase gene also targeted by our primers. This gene is coding the main enzyme that degrades sucrose in potato tubers providing substrate for starch synthesis in this storage organ (Sun et al. 1992; Wang et al. 1994). Boris et al. (2011) detected both indels and SNPs in the sucrose synthase gene and identified four alleles in potato depending on the presence of a mononucleotide $(T)_8$ repeat in intron4 and 5. This is consistent with our findings and the number of detected alleles (also four) for the Suc16-321 and Suc16-349 loci. Sequencing of the amplicons revealed further cultivar-specific nucleotide and amino acid substitutions, which were identified in potato genotypes (Boris et al. 2011). Similar results were obtained by (Draffehn et al. 2010) for the invertase gene, coding an enzyme that irreversibly cleaves sucrose into fructose and glucose. This gene was also included in our study. Substantial allelic diversity was found for potato invertases both in indels and SNPs. This suggests that sequencing of IT bands may provide further information for genetic diversity analysis.

Reproducibility, specificity and amplification artifacts

Using the same primer combination, we repeated the PCR reactions to test the reproducibility of the marker system. During the replicate experiments the same banding patterns were detected. The polymorphic bands were fully reproducible when the same DNA samples were run in independent experiments, under the same reaction conditions in different PCR instruments. Some results using IT have shown that amplicons generated from different samples may not only amplify the targeted specific genes, but also multiple analogues or alternative copies of targeted gene. These can result from pseudogene loci, or from PCR errors generating artifacts. Therefore, we suggest that PCR conditions for all reactions should be carefully optimized, because unspecific amplification products, like heteroduplexes, may occur in the reaction. In the IT method, polymorphism between individuals ultimately results from insertions/deletions (or from single nucleotide changes) occurring in the amplified fragments, representing different alleles of the targeted gene. Therefore, amplicons could form heteroduplex artefacts, where a double-stranded product is generated, from the single complementary strand derived from the alternate allelic sequences of the targeted gene. We observed a heteroduplex formed by different DNA strands having homologue sites, originating from different alleles of the *Cat-In*2 locus, linked to the Ry_{sto} gene from *Solanum stoloniferum* Schltdl. during our ongoing experiments related to potato breeding (data not shown). These products, however, can be eliminated with precise reaction condition optimization, or with the inclusion of adjuvants, such as BSA, in the reaction mix. Our observation is that non-specific fragments tend to appear when larger gene-families are targeted, but remain insignificant, or unnoticed in conserved genes with few copies. This phenomenon may appear to be problematic in studies where homology of the bands is essential. However, other fragments related to the targeted or expected gene loci might still be useful for functional marker studies, as well as for genetic diversity assessment. This is why we recommend choosing the targeted genes according to the aims of the study.

Phylogenetic utility

We mapped the characters on the parsimony consensus tree (see Electronic Supplementary Material 3) using NONA and counted synapomorphic and homoplastic bands produced for each primer (Table 6). We

observed that primers with the highest PIC and Rp values may not be the ideal choice for phylogenetic reconstructions, but they proved to be the most useful ones for genotyping. For example the GPSS-943 primer targeting the small-subunit of the ADP-glucose pyrophosphorylase gene was ranked first according to its resolving power (3.32) but produced only one synapomorphic band. Contrary, Ure-242 and Ure-271 targeting the Ure gene for urease enzyme were ranked only 23rd and 28th based on Rp values but they produced four synapomorphic bands. We suggest choosing primers for phylogenetic purposes, which are less polymorphic within each species, have lower Rp values but produce banding patterns characteristic for the particular species. Indel markers tend to produce more synapomorphies between species, and homoplasy increase at infraspecific level. There is a slight correlation between taxonomic distance and the utility of indel markers as synapomorphies are more common at infrageneric level. Insertion and deletions of single base pairs and monomeric base pair extensions with variable length are commonly found among indels (Vasemägi et al. 2010), but other types due to transposon insertion, slippage in simple sequence replication or unequal crossing over events also occur (Xiong et al. 2010; Haile et al. 2013; Hachiken et al. 2013). It seems that such events have taken place multiple times in different lineages limiting the use of techniques at small scale phylogenetic studies.

Genetic relationships based on intron-targeting markers

Intron-targeting markers were used to fingerprint a small diverse set of black nightshade accessions. The fingerprinting analysis reported here supports previous taxonomy of sect. *Solanum* as presented by Dehmer and Hammer (2004) and Manoko et al. (2008). However, the topologies presented here are only tentative, because only small subset of accessions was analyzed to investigate the transferability and utility of the described technique. In this context the presented results should be regarded only as a mere experiment with the developed markers. For further analyses and clarification, it would be possible to use the newly developed intron-targeting marker system simultaneously with others on a broader range of samples from different continents to confirm or reject results obtained here. Our results also support the differentiation of the two diploid species *S. americanum* and *S. chenopodioides* from the tetraploid *S. retroflexum, S. villosum* and hexaploid species *S. nigrum*, *S. opacum* and *S. scabrum*. Phylogenetic analyses with cladistic methods might not be appropriate when polyploid species, possibly resulting from reticulate evolution, are included. These events are complex and will not be investigated in this study [for further details see Poczai and Hyvönen (2011)], but it should be emphasized that intron-targeting might be useful also for such studies besides and/or together with other marker systems. Previous studies of black nightshades germplasm reported difficulties due to the taxonomic complexity of this group. Various classical experimental and numerical studies have demonstrated that this complexity can be attributed to a number of causes such as phenotypic plasticity and genetic variation (Edmonds 1979). Many diploid taxa hybridize readily and show both pre- and post-zygotic isolating mechanisms resulting in a polyploid complex (Edmonds 1979). It seems clear, that for the complete investigation of species relationships in this section could only be achieved by a complex analysis combining marker systems as suggested by Poczai and Hyvönen (2011).

Concluding remarks

Our interpretations and conclusions are necessarily tentative at this stage, but this study could serve as a useful basis to develop more comprehensive and detailed work in the future. Given the preliminary nature of this study, we chose to use intron-targeting markers as an efficient technique for obtaining a picture of interspecies diversity in the selected black nightshade accessions. The markers applied here, may have wider application in other species of the genus *Solanum*. These primers might be potentially useful in other fields of applied research, e.g., crop breeding programs of the family as further new primers can be easily designed based on the methods described here. Since the targeted intron sequences are generally less conserved than the exons, the amplified products may display polymorphism due to length/nucleotide variation among introns in the alleles of the gene. This fact makes the intron-targeting method an effective tool in the identification of several markers for plant molecular genetic studies. Using next generation sequencing

technologies, e.g., 454-Roche and iLLUMINA, large amounts of new sequences could be produced in a relatively easy and cost effective way providing new basis for IT marker development. This could be very useful for orphan crops, or poorly known taxonomic groups of *Solanum* where further sequence information is needed.

Acknowledgments PP gratefully acknowledges support from a Marie Curie Fellowship Grant (PIEF-GA-2011-300186) under the seventh framework program of the European Union. This study was partially supported by a Hungarian Eötvös Research Grant and a Campus Hungary Grant provided to IV. IC is supported by the János Bolyai Research Fellowship of the Hungarian Academy of Sciences.

References

Ames M, Spooner DM (2010) Phylogeny of *Solanum* series *Piurana* and related species in *Solanum* section *Petota* based on five conserved ortholog sequences. Taxon 59:1091–1101

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Boris KV, Ryzhova NN, Kochieva EZ (2011) Identification and characterization of intraspecific variability of the sucrose synthase gene Sus4 of potato (*Solanum tuberosum*). Russ J Genet 47:168–175

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Britten RJ, Rowen L, Williams J, Cameron RA (2003) Majority of divergence between closely related DNA samples is due to indels. PNAS 100:661–4665

Cai D, Rodríguez F, Teng Y, Ané C, Bonierbale M, Mueller LA, Spooner DM (2012) Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section *Petota*). BMC Evol Biol 12:70

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Colon LT, Eijlander R, Budding DJ, Pieters MMJ, Hoogendoorn J, Van-Ijzendoorn MT (1993) Resistance to potato late blight (*Phytophthora infestans* (Mont.) de Bary) in *Solanum nigrum*, *S. villosum* and their sexual hybrids with *S. tuberosum* and *S. demissum*. Euphytica 66:55–64

D'Agostino N, Golas T, van de Geest H, Bombarely A, Dawood T, Zethof J, Driedonks N, Wijnker E, Bargsten J, Nap JP, Mariani C, Rieu I (2013) Genomic analysis of the native European *Solanum* species S. dulcamara. BMC Genomics 14:356

De Keyser E, De Riek J, Van Bockstaele E (2009) Discovery of species-wide EST-derived markers in *Rhododendron* by intron-flanking primer design. Mol Breeding 23:171–178

Dehmer KJ (2001) Conclusions on the taxonomy of the *Solanum nigrum* complex by molecular analysis of IPK germplasm accessions. In: van den Berg RG, Barendse GWM, van der Weerden GM, Mariani C (eds) Solanaceae V: advances in taxonomy and utilization. Nijmegen University Press, The Netherlands, pp 85–96

Dehmer KJ, Hammer K (2004) Taxonomic status and geographic provenance of germplasm accessions in the *Solanum nigrum* L complex: AFLP data. Genet Resour Crop Evol 51:551–558

Dehmer KJ, Stracke S (1999) Molecular analyses of genebank accessions of the *Solanum nigrum* complex. In: Andrews S, Leslie AC, Alexander C (eds) Third international symposium on the taxonomy of cultivated plants. Royal Botanic Gardens, Kew, pp 343–345

Draffehn AM, Meller S, Li L, Gebhardt C (2010) Natural diversity of potato (*Solanum tuberosum*) invertases. BMC Plant Biol 10:271

Dwivedi SL, Crouch JH, Mackill DJ, Xu Y, Blair MW, Ragot M, Upadhyaya HD, Ortiz R (2007) The molecularization of public sector crop breeding: progress, problems, and prospects. Adv Agron 95:163–318

Edmonds JM (1979) Biosystematics of *Solanum* L. section *Solanum* (*Maurella*). In: Hawkes JG, Lester RN, Skelding AD (eds) The biology and taxonomy of the Solanaceae. Academic Press, London, pp 529–548

Edmonds JM, Chweya JA (1997) Black nightshades. *Solanum nigrum* L. and related species. Promoting the conservation and use of underutilized and neglected crops, vol 15. International Plant Genetic Resources Institute, Rome, Italy

Eijlander R, Stiekema WJ (1994) Biological containment of potato (*Solanum tuberosum*): outcrossing to the related wild species black nightshade (*Solanum nigrum*) and bittersweet (*Solanum dulcamara*). Sex Plant Reprod 7:29–40

Farris JS (1989) The retention index and homoplasy excess. Syst Zool 38:406–407

Farris JS, Albert V, Källersjö M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. Cladistics 12:99–124

García-Lor A, Luro F, Navarro L, Ollitrault P (2012) Comparative use of indel and SSR marker sin deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. Mol Genet Genomics 287:77–94

Golas TM, Sikkema A, Gros J, Feron RM, van den Berg RG, van der Weerden GM, Mariani C, Allefs JJ (2010) Identification of a resistance gene *Rpi-dlc1* to *Phytophthora infestans* in European accessions of *Solanum dulcamara*. Theor Appl Genet 120:797–808

Golas TM, van de Geest H, Gros J, Sikkema A, D'Agostino N, Nap JP, Mariani C, Allefs JJ, Rieu I (2013) Comparative next-generation mapping of the *Phytophthora infestans* gene *Rpi-dlc2* in the European species *Solanum dulcamara*. Theor Appl Genet 126:59–68

Goloboff PA (1994) NONA: a tree searching program. Program and documentation. www.cladistics.com/aboutNona.htm

Gorji AM, Matyas KK, Dublecz Z, Decsi K, Cernak I, Hoffmann B, Taller J, Polgar Z (2012) In vitro osmotic stress tolerance in potato and identification of major QTLs. Am J Pot Res 89:453–464

Hachiken T, Sato K, Hasegawa T, Ichitani K, Kawase M, Fukunaga K (2013) Geographical distribution of waxy gene SNPs and indels in foxtail millet, Setaria italic (L.) P. Beauv. Genet Resour Crop Evol 60:1559–1570

Haile JK, Hammer K, Badebo A, Nachit MM, Röder MS (2013) Genetic diversity assessment of Ethiopian tetraploid wheat landraces and improved durum wheat varieties using microsatellites and markers linked with stem rust resistance. Genet Resour Crop Evol 60:513–527

Holland JB, Helland SJ, Sharapova N, Rhyne DC (2001) Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. Genome 44:1065–1076

Holsinger KE (1999) Analysis of genetic diversity in geographically structured populations: a Bayesian perspective. Hereditas 130:245–255

Holsinger KE, Lewis PO (2003) HICKORY: a package for the analysis of population genetic data, version 1.0.4. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA. Distributed by the authors

Horsman K, Bergervoet JEM, Jacobsen E (1997) Somatic hybridization between *Solanum tuberosum* and species of the *S. nigrum* complex: selection of vigorously growing and flowering plants. Euphytica 96:345–352

Hu J, Seiler G, Kole C (2010) Genetics, genomics and breeding of sunflower. CRC Press, Routledge

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Jacoby A, Labuschagne MT, Viljoen CD (2003) Genetic relationships between Southern African *Solanum retroflexum* Dun and other related species measured by morphological and DNA markers. Euphytica 132:109–113

Kimura M (1983) Rare variant alleles in the light of the neutral theory. Mol Biol Evol 1:84–93

Kimura M, Crow J (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of Anurans. Syst Zool 18:1–32

Lebecka R (2008) Host–pathogen interaction between *Phytophthora infestans* and *Solanum nigrum*, *S. villosum*, and *S. scabrum*. Eur J Plant Pathol 120:233–240

Lehmann C, Biela C, Töpfl S, Jansen G, Vögel R (2007) *Solanum scabrum*—a potential source of a coloring plant extract. Euphytica 158:189–199

Levin RA, Blanton J, Miller JS (2009) Phylogenetic utility of nuclear nitrate reductase: a multi-locus comparison of nuclear and chloroplast sequence data for inference of relationships among American Lycieae (Solanaceae). Mol Phylogenet Evol 50:608–617

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:391–398

Liu BH (1998) Statistical genomics: linkage, mapping and QTL analysis. CRC Press, Boca Raton

Manoko MLK, van den Berg RG, Feron RMC, van der Weerden GM, Mariani C (2007) AFLP markers support separation of *Solanum nodiflorum* from *Solanum americanum* sensu stricto (Solanaceae). Plant Syst Evol 267:1–11

Manoko MLK, van den Berg RG, Feron RMC, van der Weerden GM, Mariani C (2008) Genetic diversity of the African hexaploid species of *Solanum scabrum* Mill. and *Solanum nigrum* L. (Solanaceae). Genetic Res Crop Evol 55: 409–418

Morales M, Roig E, Monforte AJ, Arús P, Garcia-Mas J (2004) Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.). Genome 47:352–360

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y et al (2005) The SOL genomics network: a comparative resource for solanaceae biology and beyond. Plant Physiol 138:1310–1317

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473–497

Nagy S, Poczai P, Cernák I, Gorji AM, Hegedűs G, Taller J (2012) PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. Biochem Genet 50:9–10

Nei M (1973) Analysis of gene diversity in subdivided populations. PNAS USA 70:3321–3323

Nixon KC (1999) The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15:407–414

Nixon KC (2002) Winclada. Version 1.00.08. http://www.cladistics.com/aboutWinc.htm. Ithaca, NY

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL (2008) AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583

Olet EA, Huen M, Lye KA (2005) African crop or poisonous nightshade; the enigma of poisonous or edible black nightshades solved. Afr J Ecol 43:158–161

Park S, Yu HJ, Mun JH, Lee SC (2010) Genome-wide discovery of DNA polymorphism in *Brassica rapa*. Mol Genet Genomics 283:135–145

Poczai P, Hyvönen J (2011) On the origin of *Solanum nigrum*: can networks help? Mol Biol Rep 38:1171–1185

Poczai P, Cernák I, Gorji AM, Nagy S, Taller J, Polgár J (2010) Development of intron-targeting (IT) markers for potato and cross-species amplification in *Solanum nigrum* (Solanaceae). Amer J Bot 97:e142–e145

Poczai P, Varga I, Bell NE, Hyvönen J (2011) Genetic diversity assessment of bittersweet (*Solanum dulcamara*, Solanaceae) germplasm using conserved DNA-derived polymorphism and intron-targeting markers. Ann Appl Biol 159:141–153

Poczai P, Varga I, Laos M, Cseh A, Bell N, Valkonen JPT, Hyvönen J (2013) Advances in plant gene-targeted and functional markers: a review. Plant Methods 9:6

Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivar. Theor Appl Genet 98:107–112

Rambaut A, Drummond AJ (2007) Tracer v1.4. http://tree.bio.ed.ac.uk/software/tracer/. Accessed 13 Feb 2013

Rodríguez F, Spooner DM (2009) Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota)* genomes with emphasis on the origins of the polyploid species. Syst Bot 34: 207–219

Rodríguez F, Wu F, Ané C, Tanksley S, Spooner DM (2009) Do potatoes and tomatoes have a single evolutionary history,

and what proportion of the genome supports this history? BMC Evol Biol 9:191

Schippers RR (2000) African indigenous vegetables. An overview of the cultivated species. Natural Resources Institute/ ACPEU Technical Centre of Agriculture and Rural Cooperation, Catham, pp 147–176

Silvestro D, Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. Org Divers Evol 12:335–337

Spooner DM, Rodríguez F, Polgár Z, Ballerd HE, Jansky SH (2008) Genomic origins of potato polyploids: GBSSI gene sequencing data. Crop Sci 48:S27–S36

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

Stöver BC, Müller KF (2010) TreeGraph2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11:7

Sun J, Loboda T, Sung SS, Black CC (1992) Sucrose synthase in wild tomato, *Lycopersicon chmielewskii*, and tomato fruit skin strength. Plant Physiol 98:1163–1169

Tepe EJ, Bohs L (2010) A molecular phylogeny of *Solanum* sect. *Pteroidea* (Solanaceae) and the utility of COSII markers in resolving relationships among closely related species. Taxon 59:733–743

Vasemägi A, Gross R, Plam D, Paaver T, Primmer CR (2010) Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. BMC Genomics 11:156

Walbot V, Warren C (1988) Regulation of *Mu* element copy number in maize lines with an active or inactive Mutator transposable element system. Mol Gen Genet 211:27–34

Wang F, Smith A, Brenner ML (1994) Temporal and spatial expression pattern of sucrose synthase during tomato fruit development. Plant Physiol 104:535–540

Wang Y, Chen J, Francis DM, Shen H, Wu T, Yang W (2010) Discovery of intron polymorphisms in cultivated tomato using both tomato and Arabidopsis genomic information. Theor Appl Genet 121:1199–1207

Weese TL, Bohs L (2007) A three-gene phylogeny of the genus *Solanum* (Solanaceae). Syst Bot 32:445–463

Wu F, Mueller LA, Crouzillat D, Pétiard 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 174:1407–1420

Xiong Z, Zhang S, Wang Y, Ford-Lloyd BV, Tu M, Jin X, Wu Y, Yan H, Yang X, Liu P et al (2010) Differentiation and distribution of *indica* and *japonica* rice varieties along the altitude gradients in Yunnan Province of China as revealed by InDel molecular markers. Genet Resour Crop Evol 57:891–902

Xu G, Ma H, Nei M, Kong H (2009) Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. PNAS 106: 835–840

Yeh FC, Yang R-C, Boyle TJB, Ye Z-H, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, MN