

RESEARCH ARTICLE

Open Access

# Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section *Petota*)

Danying Cai<sup>1,2</sup>, Flor Rodríguez<sup>2,3</sup>, Yuanwen Teng<sup>1</sup>, Cécile Ané<sup>4</sup>, Meredith Bonierbale<sup>5</sup>, Lukas A Mueller<sup>6</sup> and David M Spooner<sup>2\*</sup>

## Abstract

**Background:** Recent genomic studies have drastically altered our knowledge of polyploid evolution. Wild potatoes (*Solanum* section *Petota*) are a highly diverse and economically important group of about 100 species widely distributed throughout the Americas. Thirty-six percent of the species in section *Petota* are polyploid or with diploid and polyploid cytotypes. However, the group is poorly understood at the genomic level and the series is ideal to study polyploid evolution. Two separate studies using the nuclear orthologs GBSSI and nitrate reductase confirmed prior hypotheses of polyploid origins in potato and have shown new origins not proposed before. These studies have been limited, however, by the use of few accessions per polyploid species and by low taxonomic resolution, providing clade-specific, but not species-specific origins within clades. The purpose of the present study is to use six nuclear orthologs, within 54 accessions of 11 polyploid species, 34 accessions of 29 diploid species of section *Petota* representing their putative progenitors, and two outgroups, to see if phenomena typical of other polyploid groups occur within wild potatoes, to include multiple origins, loss of alleles, or gain of new alleles.

**Results:** Our results increase resolution within clades, giving better ideas of diploid progenitors, and show unexpected complexity of allele sharing within clades. While some species have little diversity among accessions and concur with the GBSSI and nitrate reductase results, such as *S. agrimonifolium*, *S. colombianum*, *S. hjertingii*, and *S. moscopanum*, the results give much better resolution of species-specific progenitors. Seven other species, however, show variant patterns of allele distributions suggesting multiple origins and allele loss. Complex three-genome origins are supported for *S. hougasii*, and *S. schenckii*, and one of the ten accessions of *S. stoloniferum*. A very unexpected shared presence of alleles occurs within one clade of *S. verrucosum* from Central America, and *S. berthaultii* from South America in six polyploid species *S. demissum*, *S. hjertingii*, *S. hougasii*, *S. iopetalum*, *S. schenckii*, and *S. stoloniferum*.

**Conclusions:** Our results document considerable genomic complexity of some wild potato polyploids. These can be explained by multiple hybrid origins and allele losses that provide a clear biological explanation for the taxonomic complexity in wild potato polyploids. These results are of theoretical and practical benefit to potato breeders, and add to a growing body of evidence showing considerable complexity in polyploid plants in general.

## Background

### Biology of polyploids

Over 70% of the monocots [1] and 70-80% of the dicots [2] were estimated to be polyploid. Recent data from genomics, however, suggest that almost all angiosperms, perhaps even all plant groups, have experienced one to several rounds of genome duplication [3,4], sometimes

followed by genome reorganization, homoeolog loss, and diploidization [3,5-12].

Polyploids are usually defined as autopolyploid (genome doubling from single species) or allopolyploid (genome doubling after hybridization or genome unreduced before hybridization) [13]. Others use the terms polysomic polyploid (instead of autopolyploid) and disomic polyploid (instead of allopolyploid) [14] to describe the genetic behavior of the plants rather than assuming anything about their origin or genome constitution.

A number of studies have focused on the single vs. multiple events of origins of polyploids [5,15-21]. Multiple

\* Correspondence: david.spooner@ars.usda.gov

<sup>2</sup>Department of Horticulture, USDA, Agricultural Research Service, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706-1590, USA  
Full list of author information is available at the end of the article

origins could generate more genetic diversity in the polyploid plant species than their diploid ancestors [22-24]. This study focuses on polyploid evolution in wild potatoes (*Solanum* L. section *Petota* Dumort.).

#### **Taxonomy and biology of *Solanum* section *Petota***

Section *Petota* contains four cultivated species [25,26], and about 100 wild species relatives [27]. About 70% of the wild species are diploid ( $2n = 2x = 24$ ), with the rest tetraploid ( $2n = 4x = 48$ ) and hexaploid ( $2n = 6x = 72$ ), with a few triploid or pentaploid populations [28]. The polyploids range from allopolyploids to autopolyploids. Using the classical taxonomic system of Hawkes [29] (Table 1), four taxonomic series of wild species that are wholly or predominately polyploid: series *Acaulia* (4x, 6x), *Conicibaccata* (2x, 4x, 6x), *Demissa* (6x), and *Longipedicellata* (4x). Other series with polyploids are predominately diploid: *Bulbocastana* (2x, 3x), *Commersoniana* (2x, 3x), *Maglia* (2x, 3x), *Pinnatisecta* (2x, 3x), *Piurana* (2x, 4x), and *Tuberosa* (2x, 4x, 6x) [28].

The origin of potato polyploids has been the subject of much debate, incorporating data from crossing studies, cytogenetics, morphology, and biogeography. Hawkes [29] speculated that section *Petota* arose in North and Central America, and possessed white stellate corollas, B genomes, and endosperm balance numbers (EBN) of 1. EBN is a strong biological isolating mechanism, empirically determined through artificial interspecific crosses and cytological examinations, and evidenced by endosperm death in EBN incompatible crosses. Entirely on the basis of empirical data, *Solanum* species have been assigned EBN based on their ability to hybridize with each other [30]. Barring other crossing barriers, successful hybridization is expected when male and female gametes have matching EBN, regardless of ploidy. Ploidy(EBN) combinations in potato include 2x(1EBN), 2x(2EBN), 4x(2EBN), 4x(4EBN), and 6x(4EBN). Hawkes [29] speculated that some of the North and Central American 2x(1EBN) species migrated to South America with A genome, rotate corollas and 2 or 4 EBN. A return migration of A genome back to Mexico and Central America took place around 3.5 MA, followed by polyploid formation of series *Conicibaccata*, *Demissa*, and *Longipedicellata* with rotate corollas.

Molecular phylogenetics suggests a classification of *Solanum* section *Petota*, and of polyploid origins in potato, often quite at odds with traditional classifications as outlined by section and series affiliations [29] (Table 1). The latest comprehensive taxonomic treatment of the section by Hawkes [29] recognized 232 tuber-bearing and non-tuber-bearing species divided into 21 taxonomic series. Plastid DNA restriction site studies [31], and nuclear DNA sequencing studies [32-37] have greatly changed our understanding of ingroup relationships in

section *Petota*. Relative to the last comprehensive taxonomic treatment by Hawkes [29], section *Petota* now excludes the non-tuber-bearing species, reclassified as sections *Etuberosum* (Bukasov and Kameraz) A. Child, *Juglandifolia* (Rydberg) A. Child, and *Lycopersicoides* (A. Child) Peralta [38]. Many of Hawkes 21 series are shown to be unnatural and the tuber-bearing species are divided into four clades (1 – 4) based on plastid DNA restriction site data or three clades based on nuclear DNA sequencing data, with both results similar except that the nuclear DNA sequencing data fail to distinguish clades 1 and 2. To maintain correspondence with the prior literature, we therefore refer to the nuclear clades here, as in prior studies, as clades 1 + 2, 3, 4. The allopolyploids combine sequences from different clades.

At lower taxonomic levels and important to this study, the Mexican hexaploid species *S. demissum* was shown to be related to the South American tetraploid species *S. acaule* and *S. albicans*, not to other members of series *Demissa* [39-41]. Spooner et al. [42] used these results to classify *S. acaule*, *S. albicans*, and *S. demissum* in an informal *Acaulia* group, and the other members of series *Demissa* (*S. hougasii*, *S. iopetalum*, and *S. schenckii*) in an informal *Iopetala* group. Because of complex hybrid origins and allopolyploid origins they also used the terms *Conicibaccata* group and *Longipedicellata* group instead of series. We use the terms *Acaulia*, *Conicibaccata*, *Iopetala*, and *Longipedicellata* groups in the text as they are putatively more natural, but show Hawkes's [29] traditional series classifications in Table 1.

The genome constitution of potato polyploids has been investigated by various workers [32, 33, 43, 44; Table 1]. Matsubayashi [43] speculated on the genome formation of diploid and polyploid species via insights from cytological analysis. *Solanum acaule* was thought to be a segmental allotetraploid with minor variants of a common A genome. *Solanum agrimonifolium*, *S. colombianum*; and *S. stoloniferum* were designated as strict allotetraploids, and shared the same A genome with *S. verrucosum*, the sole A genome species from Mexico. *Solanum demissum* was thought to be an allohexaploid with two similar genomes and a third different genome, which also had one common genome with many of diploid *Conicibaccata* group and series *Megistacroloba* and *Tuberosa* (Table 1). Pendinen et al. [44] supported a genome constitution of AABB for *S. hjertingii* and *S. stoloniferum* with genomic in situ hybridization (GISH) analysis, and they proposed *S. verrucosum* for the A genome donor, as well as at least one of three species in series *Pinnatisecta* (*S. cardiophyllum*, *S. ehrenbergii*, or *S. jamesii*) for the B genome donor, in concordance with the phylogenetic results using nuclear ortholog DNA sequences [32,33]. Nuclear DNA sequencing studies [32-36] supported clade 1+2 to contain B genome

**Table 1 Germplasm examined grouped by their classic series affiliations (Hawkes, 1990), country and state (or province or department) of collection, ploidy and endosperm balance numbers (EBN), and genome affiliations**

Section, series (ser. abbreviation)	Species	PI	Country, State	Ploidy	EBN	Genomes (Matsubayashi 1991)	Genomes (Hawkes 1990)				
<b>Polyploids</b>											
Acaulia Juz. (ACA)	<i>Solanum acaule</i> Bitter	310923	Bolivia, Cochabamba	4x	2	AAA <sup>a</sup> A <sup>a</sup>	A <sub>2</sub> A <sub>2</sub> A <sub>3</sub> A <sub>3</sub>				
		472648	Argentina, Jujuy	4x	2						
		473485	Peru, Lima	4x	2						
	<i>S. albicans</i> (Ochoa) Ochoa	230494	Peru, Cajamarca	6x	4	AAA <sup>a</sup> A <sup>a</sup> XX					
		365305	Peru, Apurímac	6x	4						
		365376	Peru, La Libertad	6x	4						
		561642	Ecuador, Chimborazo	6x	4						
		Demissa Buk. (DEM)	<i>S. demissum</i> Lindl.	161719	Mexico, Federal District			6x	4	AADD <sup>d</sup> D <sup>d</sup>	A <sub>1</sub> A <sub>1</sub> A <sub>4</sub> A <sub>4</sub> B <sub>1</sub> B <sub>1</sub>
				225711	Columbia, Boyacá			6x	4		
				275206	Mexico, Chihuahua			6x	4		
275211	Guatemala, Huehuetenango			6x	4						
498012	Mexico, Durango		6x	4							
545763	Mexico, Oaxaca		6x	4							
558482	Mexico, Mexico		6x	4							
<i>S. hougasii</i> Correll	161174		Mexico, Michoacán	6x			4	A <sub>1</sub> A <sub>1</sub> A <sub>4</sub> A <sub>4</sub> B <sub>2</sub> B <sub>2</sub>			
	161726		Mexico, Jalisco	6x			4				
239423	Mexico, Michoacán		6x	4							
558402	Mexico, Jalisco	6x	4								
558422	Mexico, Jalisco	6x	4								
<i>S. iopetalum</i> (Bitter) Hawkes	275181	Mexico, Puebla	6x	4	AADD <sup>b</sup> D <sup>b</sup> (as <i>S. brachycarpum</i> )	A <sub>1</sub> A <sub>1</sub> A <sub>4</sub> A <sub>4</sub> B <sub>3</sub> B <sub>3</sub>					
	275182	Mexico, Puebla	6x	4							
	498021	Veracruz	6x	4							
	498251	Mexico, Oaxaca	6x	4							
	558405	Mexico, Michoacán	6x	4							
	558409	Mexico, México	6x	4							
	607850	Mexico, Hidalgo	6x	4							

**Table 1 Germplasm examined grouped by their classic series affiliations (Hawkes, 1990), country and state (or province or department) of collection, ploidy and endosperm balance numbers (EBN), and genome affiliations (Continued)**

	<i>S. schenckii</i> Bitter	275261	Mexico, Oaxaca	6x	4	A <sub>1</sub> A <sub>1</sub> A <sub>4</sub> A <sub>4</sub> B <sub>4</sub> B <sub>4</sub>			
		498040	Mexico, Queretaro	6x	4				
		498250	Mexico, Oaxaca	6x	4				
		545733	Mexico, Puebla	6x	4				
		558456	Mexico, Oaxaca	6x	4				
		558458	Mexico, Puebla	6x	4				
		Longipedicellata Buk. (LON)	<i>S. hjertingii</i> Hawkes	186559	Mexico, Coahuila		4x	2	
				251067	Mexico, Nuevo León		4x	2	
				498019	Mexico, Coahuila		4x	2	
				498050	Mexico, San Luis Potosí		4x	2	
545713	Mexico, Coahuila			4x	2				
	<i>S. stoloniferum</i> Schlttdl. and Bouchet	570625	Mexico, San Luis Potosí	4x	2	AABB			
		275252	Mexico, Oaxaca	4x	2				
		283101	Mexico, Chihuahua	4x	2				
		497994	Mexico, Chihuahua	4x	2				
		498028	Mexico, Zacatecas	4x	2				
		545740	Mexico, Durango	4x	2				
		545787	Mexico, Zacatecas	4x	2				
		558395	Mexico, Baja California Sur	4x	2				
		558453	Mexico, Jalisco	4x	2				
		558454	Mexico, Queretaro	4x	2				
Conicibaccata Bitter (CON)	<i>S. agrimonifolium</i> Rydb.	558466	Mexico, Michoacán	4x	2	A <sup>c1</sup> A <sup>c1</sup> C <sup>a</sup> C <sup>a</sup>			
		243350	Guatemala, Huehuetenango	4x	2				
		558372	Mexico, Chiapas	4x	2				
		<i>S. colombianum</i> Bitter	561633	Ecuador, Pichincha	4x		2		
			583325	Venezuela, Táchira	4x		2		
<i>S. moscopanum</i> Hawkes	567812	Ecuador, Loja	6x	4					
	567843	Cauca Colombia	6x	4					

**Table 1 Germplasm examined grouped by their classic series affiliations (Hawkes, 1990), country and state (or province or department) of collection, ploidy and endosperm balance numbers (EBN), and genome affiliations (Continued)**

<b>Diploids</b>							
<i>Bulbocastana</i> Rydb. Hawkes (BUL)	<i>S. bulbocastanum</i> Dunal	347757	Mexico, Michoacán	2x	1	A <sup>b</sup> A <sup>b</sup>	
<i>Conicibaccata</i> Bitter (CON)	<i>S. chomatophilum</i> Bitter	310991	Peru, Amazonas	2x	2	A <sup>c2</sup> A <sup>c2</sup>	
		762575	Peru, Cajamarca.	2x	2		
	<i>S. laxissimum</i> Bitter	760350/473372	Peru, Cuzco	2x	2		
	<i>S. limbaniense</i> Ochoa	762336	Peru, Puno	2x	2		
		762848/607887	Peru, Cuzco	2x	2		
	<i>S. violaceimarmoratum</i> Bitter	473396	Bolivia, Cochabamba	2x	2		
		761797	Peru, Cuzco	2x	2		
760331		Bolivia, La Paz.	2x	2			
<i>Cuneolata</i> Hawkes (CUN)	<i>S. infundibuliforme</i> Phil.	472857	Argentina, Jujuy	2x	2	AA	
<i>Lignicaulia</i> Hawkes (LIG)	<i>S. lignicaule</i> Vargas	473351	Peru, Cuzco	2x	1		
<i>Megistacroloba</i> Cárdenas and Hawkes (MEG)	<i>S. boliviense</i> Dunal	597736	Bolivia, Potosí	2x	2	AA	
		265862	Peru, Cuzco	2x	2	AA	
<i>Pinnatisecta</i> (Rydb.) Hawkes (PIN)	<i>S. trifidum</i> Correll	255536	Mexico, Michoacán	2x	1		
	<i>S. stenophylidium</i> Bitter	255527	Mexico, Aguascalientes	2x	1		
<i>Piurana</i> Hawkes (PIU)	<i>S. albornozii</i> Correll	498206	Ecuador, Loja	2x	2		
	<i>S. cantense</i> Ochoa	762130	Peru, Lima	2x	2		
	<i>S. chilliasense</i> Ochoa	761590	Ecuador, El Oro	2x	2		
	<i>S. chiquidenum</i> Ochoa	762950	Peru, Julcán	2x	2		
	<i>S. hypacarthrum</i> Bitter	761259	Peru, Ancash	2x	1		
	<i>S. piurae</i> Bitter	761072	Peru, Piura	2x	2	A <sup>P</sup>	
<i>Polyadenia</i> Correll (POL)	<i>S. polyadenium</i> Greenm.	161728	Bolivia, Michoacán	2x	untested, likely 1	A <sup>PQ</sup> A <sup>PQ</sup>	
<i>Tuberosa</i> (Rydb.) Hawkes (TUB)	<i>S. andreanum</i> Baker	320345	Colombia, Cauca	2x	2		A <sub>1</sub> A <sub>1</sub>
	<i>S. brevicaule</i> Bitter	498091	Bolivia, Santa Cruz	2x	untested, likely 2		
	<i>S. berthaultii</i> Hawkes	265857	Bolivia, Cochabamba	2x	2		A <sub>1</sub> A <sub>1</sub>
	<i>S. brevicaule</i> Bitter	310957	Peru, Cuzco	2x	2	AA	A <sub>1</sub> A <sub>1</sub>
		498115	Cochabamba	2x	2		A <sub>1</sub> A <sub>1</sub>

**Table 1 Germplasm examined grouped by their classic series affiliations (Hawkes, 1990), country and state (or province or department) of collection, ploidy and endosperm balance numbers (EBN), and genome affiliations (Continued)**

	<i>S. candolleianum</i> P. Berthault	266385	Peru, Junin	2x	2	AA	A <sub>1</sub> A <sub>1</sub>
	<i>S. cajamarquense</i> Ochoa	310988	Peru, Cajamarca	2x	1		
	<i>S. gandarillasii</i> Cárdenas	265866	Bolivia, Cochabamba	2x	2		A <sub>1</sub> A <sub>1</sub>
	<i>S. microdontum</i> Bitter	500036	Argentina, Salta	2x	2	AA	A <sub>1</sub> A <sub>1</sub>
	<i>S. verrucosum</i> Schltld.	161173	Mexico, Michoacán	2x	2	AA	A <sub>1</sub> A <sub>1</sub>
Yungasensa Correll (YNG)	<i>S. chacoense</i> Bitter	275138	Argentina, Tucumán	2x	2		
	<i>S. berthaultii</i> Hawkes	442689	Argentina, Salta	2x	2		
Outgroup	<i>S. etuberosum</i> Lindl.	498311	Chile, Bio-Bio	2x	1	E <sup>e</sup> E <sup>e</sup>	
Outgroup	<i>S. dulcamara</i>	Spooner 2988	USA, Wisconsin	/	/		

species, clade 3 to contain P genome species, and clade 4 to contain A genome species.

Phylogenetic studies in section *Petota* have been hindered by the use of single genes with insufficient data to construct well resolved phylogenies. Recently, multiple nuclear orthologs have been shown to be phylogenetically useful across different angiosperm clades. Wu et al. [45] published a set of conserved orthologous nuclear markers that they termed conserved orthologous set II markers that provided superior phylogenetic resolution in *Solanum* [34-36,46]. We use these markers in our study but use the simple term nuclear orthologs because all low copy nuclear orthologous genes are similar, requiring about the same level of care concerning technical issues (e.g. PCR recombination) and are subjected to the same set of lineage-specific and hence variable evolutionary properties (variation in rates, degree of gene conversion, gene amplification or loss).

Cloning was the traditional approach for uncovering allelic variants in allopolyploids in these nuclear ortholog allopolyploid studies, but this technique is hindered by the formation of chimeric sequences combining the sequences of different alleles [47], high labor, and high cost [48]. Consequently, our study stimulated us to develop single strand conformation polymorphism (SSCP) that separates alleles by their different physical conformations, not by size, alleviating all three of these problems [49]. Asymmetric PCR single-strand conformation polymorphism is an efficient alternative technique for isolating allelic variants of highly heterozygous individuals that eliminates two common problems encountered in cloning: PCR recombination and heteroduplex fixation. It works by the electrophoretic separation of single-stranded nucleic acid, with differing tertiary structures formed by sequence differences as small as a single base pair, with visual detection using biological stains or radioactivity.

The present research is an outgrowth of our prior DNA phylogenetic [32,33,35,36], and GISH [44] studies of polyploidy in section *Petota*, and use of nuclear ortholog markers for phylogenetic studies as we used in diploids [34]. We pose the following questions: 1) Do individual species of wild polyploid potatoes have single or multiple origins? 2) Which one of several diploid parents is the possible progenitor of the polyploids? 3) How do the results of nuclear ortholog sequences compare to prior results using the nuclear orthologs GBSSI and nitrate reductase?

## Results

### Sequence alignment and variation

The aligned length of the individual six nuclear orthologs ranged from 461 characters for C2At1g32130 to 1473 for C2At1g20050. Introns were present in all of them, but these posed no particular alignment

difficulties. The total aligned length of all six nuclear orthologs was 4719 characters, although as described below we did not use a concatenated dataset. The sequence data are deposited in GenBank (Table 2) and the aligned matrix is available in TreeBASE ([www.treebase.org](http://www.treebase.org)) Study Accession URL: <http://purl.org/phylo/treebase/phyloids/study/TB2:S12288>.

### Phylogenetic analysis of the diploids

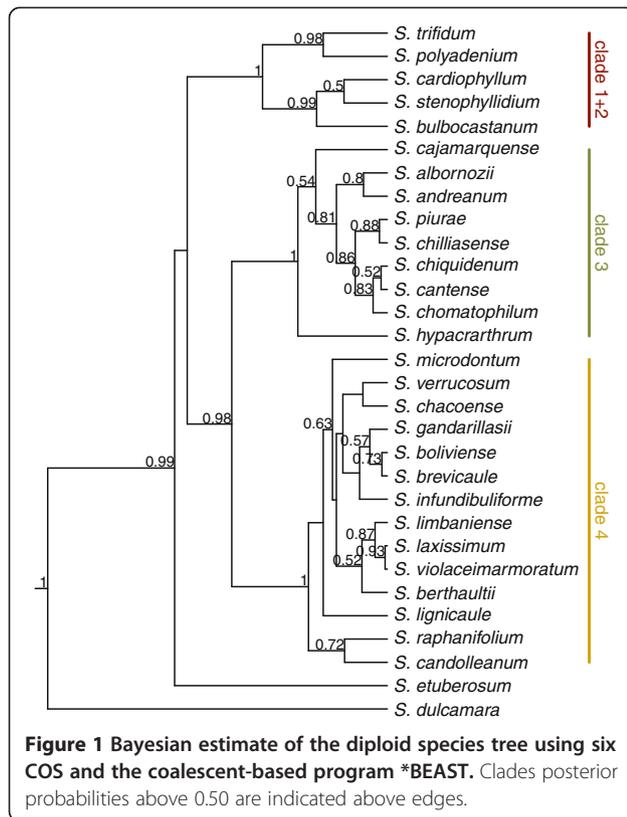
\*BEAST recovered three clades for the diploid species (Figure 1) with *S. dulcamara* recovered as outgroup and *S. etuberosum* as a close sister group of section *Petota* as in all prior nuclear DNA sequence phylogenies [32-36]. Our \*BEAST results using all six nuclear orthologs placed clades 3 and 4 as sister, with 1 + 2 sister to clade (3 + 4), as in two other nuclear ortholog phylogenies [34,45]. A GBSSI study [32] placed these three clades as polytomies, but another GBSSI study [37] and a nuclear ortholog study placed clades 1 + 2 and 4 as sister. These results clearly define these three clades, but the relationships among them are ambiguous. Because we used more sequence data here than in [32,33] and because of the concordance of our present results with [34,35], we consider the present cladistic structure to represent a dominant phylogeny. Not all prior analyses used the same species, but our study resolved all species at least in clade 4, although with low posterior probabilities for some relationships.

### Phylogenetic results of the polyploids

We analyzed all 54 of our polyploid accessions separately (Table 1, Figures 2, 3, 4 and 5). Four tetraploid species, *S. acaule* (three accessions examined), *S. agrimonifolium* (2), *S. colombianum* (2), and *S. hjertingii* (6) had invariant phylogenetic results among accessions within species, and *S. agrimonifolium* and *S. colombianum* had identical results to each other. The origin of some alleles is supported by high bootstrap values at the species level, which we report below and map in Figures 2, 3, 4 and 5 (Table 3). Other origin placements are ambiguous at the species level but well supported at the clade level.

**Table 2 GenBank numbers for the nuclear ortholog DNA sequence data used in this study**

Nuclear ortholog markers	GenBank numbers
C2At1g32130	FJ599275-FJ599284; HQ642429 - HQ642598
C2At5g14320	FJ599363-FJ599372; HQ642599 - HQ642764
C2At1g13380	FJ599242-FJ599251; HQ642096 - HQ642267
C2At1g20050	FJ599264-FJ599273; HQ642268 - HQ642428
C2At5g47390	HQ641924 - HQ642095
C2At4g10050	HQ641753 - HQ641923



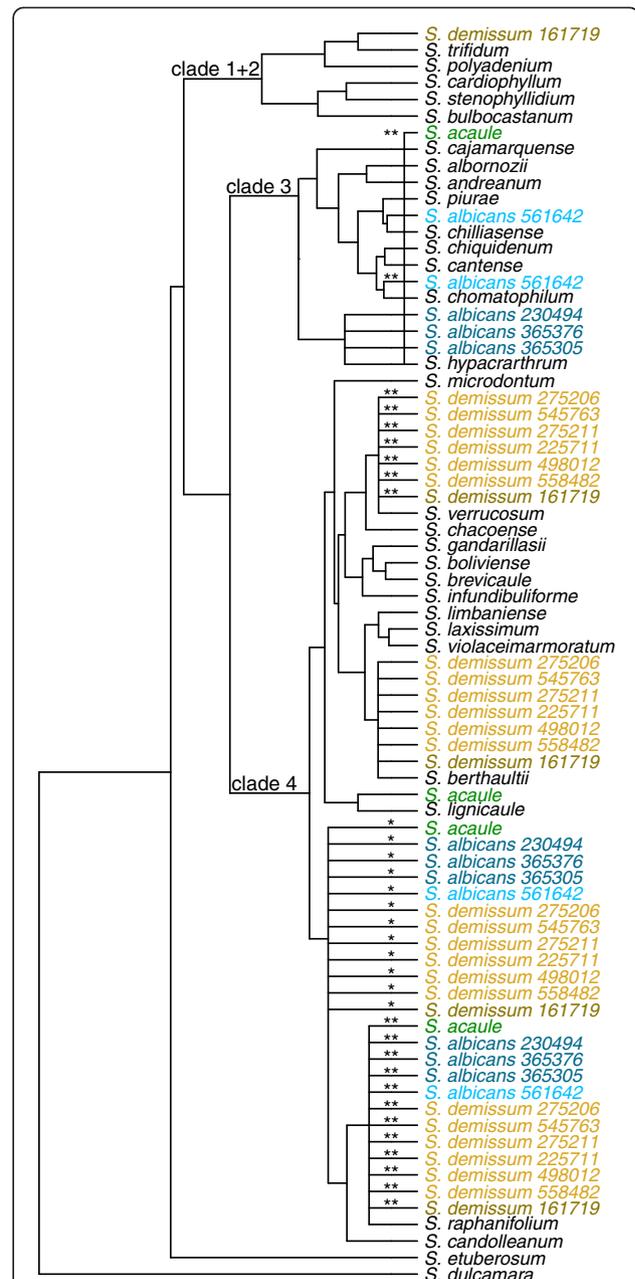
**Figure 1** Bayesian estimate of the diploid species tree using six COS and the coalescent-based program \*BEAST. Clades posterior probabilities above 0.50 are indicated above edges.

### Acaulia group (tetraploid and hexaploid)

*Solanum acaule* (tetraploid) shows distribution of most of its alleles very closely related within clade 4, with some of them placed confidently as sister to *S. lignicaule* and the *S. candolleum* and *S. raphanifolium* clade (Figure 2). Some of its alleles also fall within clade 3. The hexaploid species *S. albicans*, has three of the four accessions sharing alleles with the *S. raphanifolium* and *S. candolleum* clade (but not with *S. lignicaule*), and an additional allele with *S. hypacrarthrum* in clade 3. The remaining accession, PI 561642, shares alleles with the *S. raphanifolium* and *S. candolleum* clade, but shares clade 3 alleles with *S. chilliasense* and *S. chomatophilum*, not with *S. hypacrarthrum*. The hexaploid species *S. demissum* has six of its seven accessions sharing alleles with the *S. raphanifolium* and *S. candolleum* clade, and with *S. berthaultii* and *S. verrucosum* (and with alleles in clade 3 but with ambiguous species association). The remaining accession of *S. demissum*, PI 161719, shares these same alleles but an additional allele with *S. trifidum* in clade 1 + 2.

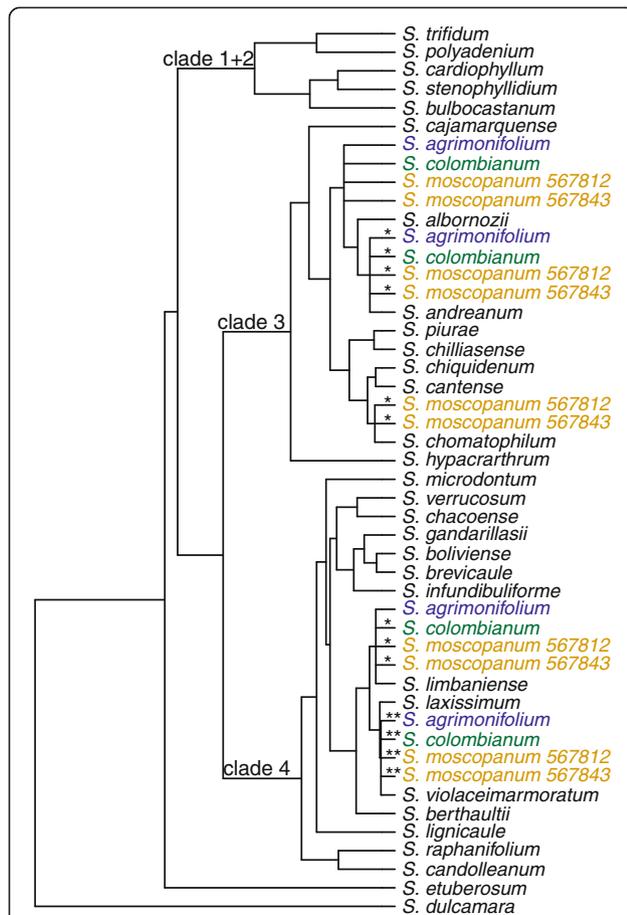
### Conicibaccata group (tetraploid and hexaploid)

All accessions of *S. agrimonifolium* and *S. colombianum* (tetraploid) are identical, grouping with all other examined diploid members of this series (*S. laxissimum*, *S. limbaniense*, *S. violaceimarmoratum*) in clade 4, and



**Figure 2** Tree summarizing the results of individual-allele analyses with the minimum evolution species tree (MEST) method, to place the polyploid alleles from the Acaulia group on the backbone diploid species tree from \*BEAST (Figure 1).

Colors correspond to accessions and rows in Table 3. The placement of parental origins which received a bootstrap support of 70% or higher for one of more alleles are indicated. Placements with lower bootstrap support are not indicated. Stronger bootstrap support (90%-95% or 95%-100%) for the placement of one or more alleles is indicated with stars (\* or \*\*). There was evidence of at least one allele of *S. acaule* originating from within clade 3 (bootstrap support of 96%) even though no single placement could be identified with strong bootstrap support. This origin from clade 3 is indicated with a vertical bar spanning the whole clade.



**Figure 3** Tree summarizing the placement of the polyploid allele origins from the Conicibaccata group on the backbone diploid species tree. Origin placements with bootstrap support over 70% for one or more alleles are displayed, as in Figure 2.

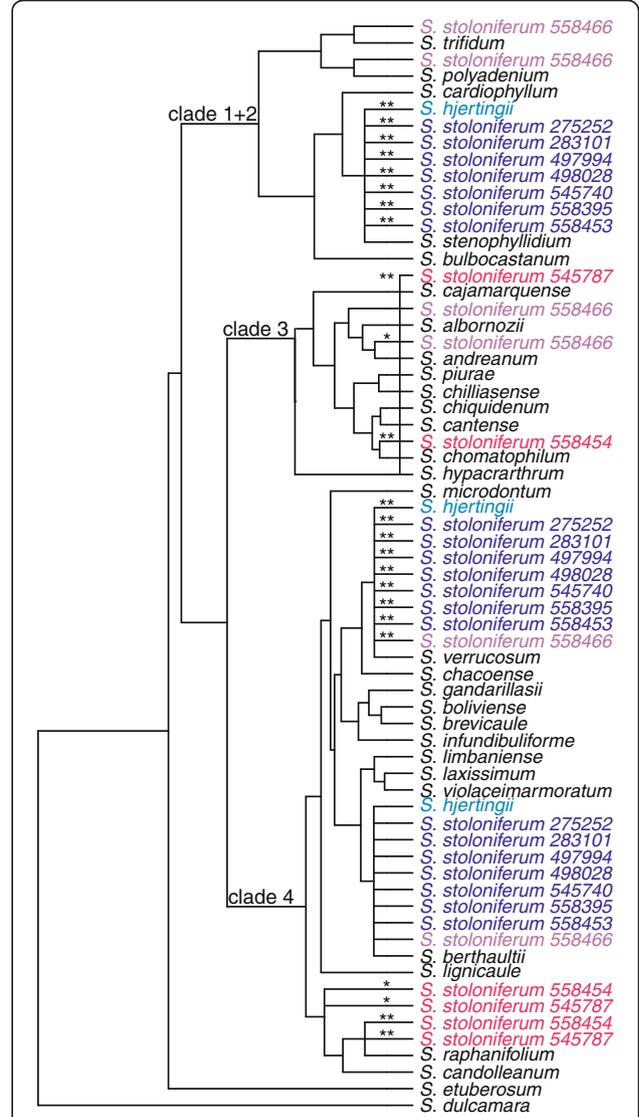
with the *S. albornozii* and *S. andreanum* clade in clade 3 (Figure 3). The two examined accessions of *S. moscopanum* (hexaploid) have identical clade associations to *S. agrimonifolium* and *S. colombianum*, but with an additional allele with *S. chomatophilum*.

**Longipedicellata group (all tetraploid)**

All six accessions of *S. hjertingii*, and seven of the ten accession of *S. stoloniferum* share alleles with *S. berthaultii* and *S. verrucosum* in clade 4, and an allele with *S. stenophyllidium* in clade 1 + 2 (Figure 4). The three remaining accessions of *S. stoloniferum* are all different from these accessions. One accession of *S. stoloniferum*, PI 558466, is identical to the above regarding clade 4, but shares alleles with the *S. polyadenium* and *S. trifidum* clade of clade 1 + 2, not with *S. stenophyllidium*, and alleles with the *S. andreanum* and *S. albornozii* clade of clade 3. The remaining accessions of *S. stoloniferum*, PI 558454 and PI 545787, share alleles with *S. candolleianum* and *S. raphanifolium* of clade 4 and alleles on clade 3.

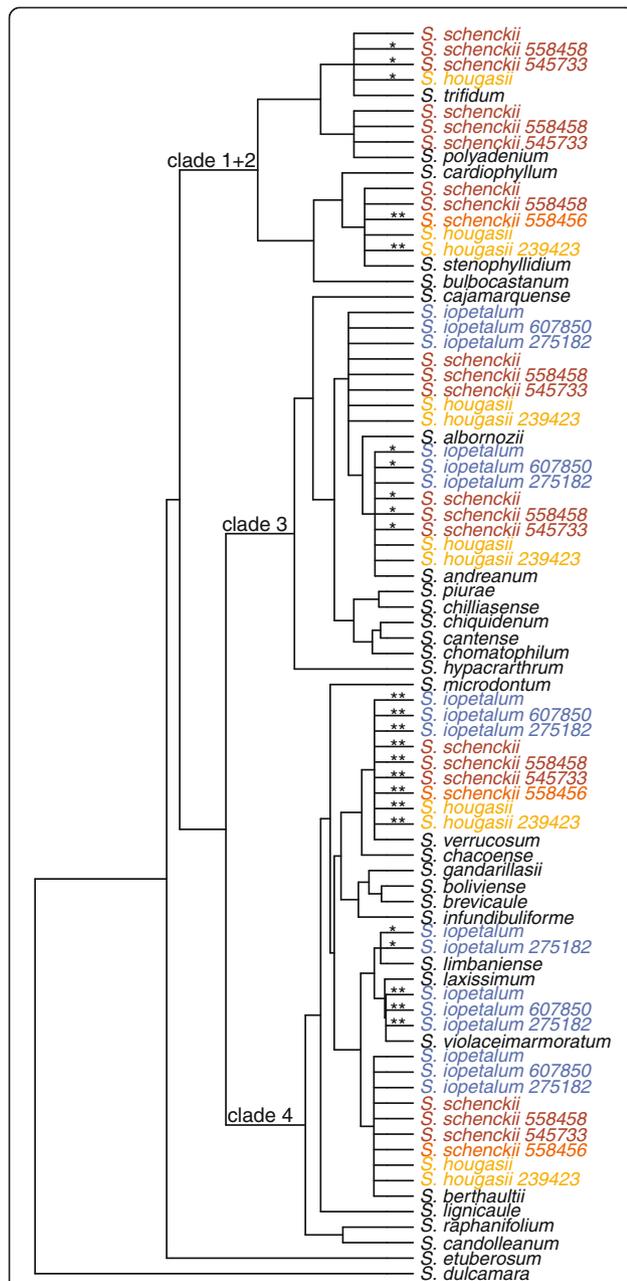
**lopetalata group (all hexaploid)**

All seven accessions of *S. iopetalum* share alleles with three clades of clade 4, *S. berthaultii*, *S. verrucosum*, and the three species of the diploid series Conicibaccata clade (*S. laxissimum*, *S. limbaniense*, and *S. violaceimarmoratum*). In addition, they share alleles with the *S. andreanum* and *S. albornozii* clade of clade 3 (Figure 5). Five of the six accessions of *S. schenckii* share alleles with *S. berthaultii* and separately *S. verrucosum* of clade 4, with the *S. andreanum* and *S. albornozii* clade of clade



**Figure 4** Tree summarizing the placement of the polyploid allele origins from the Longipedicellata group on the backbone diploid species tree. Origin placements with bootstrap support over 70% for one or more alleles are displayed, as in Figure 2.

There was evidence of at least one allele of *S. stoloniferum* 545787 originating from within clade 3 (bootstrap support of 96%) even though no single placement could be identified with strong bootstrap support. This origin from clade 3 is indicated with a vertical bar spanning the whole clade.



**Figure 5** Tree summarizing the placement of the polyploid allele origins from the lopetala group on the backbone diploid species tree. Origin placements with bootstrap support over 70% for one or more alleles are displayed, as in Figure 2. Tips named *S. hougasii* represent 4 of the *S. hougasii* accessions with similar results (161174, 161726, 558402, 558422). Similarly, tips marked *S. iopetalum* represent 5 of the *S. iopetalum* accessions (275181, 498021, 498251, 558405, 558409) and tips marked *S. schenckii* represent 3 of the *S. schenckii* accessions (275261, 498040, 498250).

3, and three separate clades in clade 1 + 2, *S. polyadenium*, *S. trifidum*, and *S. stenophyllidium* (although with low support for accession PI 545733). The remaining accession of *S. schenckii*, PI 558456, shares alleles with *S. berthaultii* and *S. verrucosum* of clade 4, and with

*S. stenophyllidium* of clade 1 + 2, but lacks alleles in clade 3.

All five accessions of *S. hougasii* are identical to four of six accessions of *S. schenckii* regarding alleles in clades 3 and 4, but *S. schenckii* shares alleles with three species (*S. polyadenium*, *S. stenophyllidium*, *S. trifidum*) in clade 1 + 2, and *S. hougasii* two species (*S. stenophyllidium*, *S. trifidum*).

## Discussion

### Phylogenetic results concordant with prior single-copy GBSSI and nitrate reductase results but with better resolution within clades

The diploid tree (Figure 1) recovers all three nuclear clades (1 + 2, 3, 4) concordant with prior results mentioned in Background. Forty-four of the 54 accessions in our present study place alleles in the major clades concordant with results from GSSSI [32] and nitrate reductase [33]. These include *S. acaule* (3 of 3 accessions), *S. agrimonifolium* (2 of 2), *S. colombianum* (2 of 2), *S. demissum* (6 of 7), *S. hjertingii* (6 of 6), *S. hougasii* (5 of 5), *S. iopetalum* (7 of 7), *S. moscopanum* (2 of 2), *S. stoloniferum* (7 of 10), and *S. schenckii* (4 of 6) (Table 3).

However, our results provide much greater resolution of species-specific associations of polyploid alleles within these clades. For example, one recurrent result is the distribution of polyploid alleles between the geographically separate clade 4 species *S. verrucosum* (Mexico) and *S. berthaultii* (central South America), as found in the North and Central American polyploids *S. demissum*, *S. hougasii*, *S. iopetalum*, *S. hjertingii*, *S. stoloniferum*, and *S. schenckii*. One possible explanation arises from the biogeographic hypothesis of Hawkes [29] who postulates that *S. verrucosum* evolved from a species that migrated from South America to Mexico, and was the A-genome (Table 1) contributor to these North and Central American polyploids. *Solanum berthaultii* (or its close relative) could have been that South American species, and *S. verrucosum* could retain some of its alleles. Another possible interpretation could be statistical error due to a violation of the coalescent model used by the species tree methods used here, such as gene flow across different species.

### Phylogenetic results incongruent with prior single-copy GBSSI and nitrate reductase results

Ten of the remaining 54 accessions show missing alleles or alleles in new clades 1 + 2, 3, 4, relative to the prior single-copy nuclear phylogenies [32,33] (Table 3). One example of new alleles relative to prior studies is found in *S. albicans* (hexaploid) that is morphologically very similar to one of its putative parents, *S. acaule*. Both species are cytological allopolyploids and the origin of the third genome in *S. albicans* relative to *S. acaule* has been the subject of long investigation [50]; note its unknown nature as

**Table 3 Summary of cladistic relationships of the polyploid alleles that are highly supported at the species level (mapped in Figures 2, 3, 4 and 5)**

Species	Number of accessions/ species examined	Clade 1 + 2	Clade 3	Clade 4
<i>Solanum acaule</i>	310923, 472648, 473485	No allele from clade 1 + 2 with BS = 0.85	Precise origin not identified, but at least 1 allele with BS = 0.96.	<i>S. lignicaule</i> ; <i>S. candolleanum</i> and <i>S. raphanifolium</i> clade
<i>S. albicans</i>	230494, 365305, 365376	No allele with BS $\geq$ 0.70	<i>S. hypacarthrum</i>	<i>S. candolleanum</i> and <i>S. raphanifolium</i> clade
	561642	No allele with BS = 0.84	<i>S. chilliasense</i> and <i>S. piurae</i> clade; <i>S. chomatophilum</i>	<i>S. candolleanum</i> and <i>S. raphanifolium</i> clade
<i>S. demissum</i>	225711, 275206, 275211, 498012, 545763, 558482	No allele with BS $\geq$ 0.82	Precise origin not identified, but at least 1 allele with BS > 0.96 for all accessions	<i>S. berthaultii</i> ; <i>S. verrucosum</i> ; <i>S. candolleanum</i> and <i>S. raphanifolium</i> clade
	161719	<i>S. trifidum</i>	Precise origin not identified, but at least 1 allele with BS = 0.95	<i>S. berthaultii</i> , <i>S. verrucosum</i> , <i>S. candolleanum</i> and <i>S. raphanifolium</i> clade
<i>S. agrimonifolium</i> , <i>S. colombianum</i>	243350, 558372, 561633, 583325	No allele with BS $\geq$ 0.91	<i>S. andreanum</i> and <i>S. alborozii</i> clade	<i>S. laxissimum</i> , <i>S. limbaniense</i> , and <i>S. violaceimarmoratum</i> clade
<i>S. moscopanum</i>	567812, 567843	No allele with BS $\geq$ 0.92	<i>S. andreanum</i> and <i>S. alborozii</i> clade; <i>S. chomatophilum</i>	<i>S. laxissimum</i> , <i>S. limbaniense</i> , and <i>S. violaceimarmoratum</i> clade
<i>S. hougasii</i>	161174, 161726, 558402, 558422, 239423	<i>S. stenophyllidium</i> ; <i>S. trifidum</i> (ambiguous for 239423)	<i>S. andreanum</i> and <i>S. alborozii</i> clade	<i>S. berthaultii</i> , <i>S. verrucosum</i>
<i>S. iopetalum</i>	275181, 275182, 498021, 498251, 558405, 558409, 607850	No allele with BS $\geq$ 0.90	<i>S. andreanum</i> and <i>S. alborozii</i> clade	<i>S. berthaultii</i> , <i>S. verrucosum</i> ; <i>S. laxissimum</i> , <i>S. limbaniense</i> , and <i>S. violaceimarmoratum</i> clade
<i>S. schenckii</i>	275261, 498040, 498250, 558458,	<i>S. polyadenium</i> ; <i>S. stenophyllidium</i> ; <i>S. trifidum</i>	<i>S. andreanum</i> and <i>S. alborozii</i> clade	<i>S. berthaultii</i> ; <i>S. verrucosum</i>
	545733	<i>S. polyadenium</i> ; <i>S. trifidum</i>	<i>S. andreanum</i> and <i>S. alborozii</i> clade	<i>S. berthaultii</i> ; <i>S. verrucosum</i>
	558456	<i>S. stenophyllidium</i>	No allele from clade 3 with BS = 0.99	<i>S. berthaultii</i> ; <i>S. verrucosum</i>
<i>S. hjertingii</i>	186559, 251067, 498019, 498050, 545713, 570625	<i>S. stenophyllidium</i>	No allele with BS $\geq$ 0.84	<i>S. berthaultii</i> ; <i>S. verrucosum</i>
<i>S. stoloniferum</i>	275252, 283101, 497994, 498028, 545740, 558395, 558453		No allele with BS $\geq$ 0.84	<i>S. berthaultii</i> ; <i>S. verrucosum</i>
<i>S. stoloniferum</i>	558454, 545787	No allele with BS $\geq$ 0.83	<i>S. chomatophilum</i> (species origin within the clade ambiguous for 545787)	<i>S. candolleanum</i> and <i>S. raphanifolium</i> clade
<i>S. stoloniferum</i>	558466	<i>S. polyadenium</i> , <i>S. trifidum</i>	<i>S. andreanum</i> and <i>S. alborozii</i> clade	<i>S. berthaultii</i> ; <i>S. verrucosum</i>

The origin placements that are ambiguous at the species level but well supported at the clade level are also indicated.

designated by XX by Matsubayashi [43] (Table 1). While the nuclear RFLP study of Hosaka and Spooner [50] could distinguish the genetic difference between *S. acaule* and *S. albicans*, no clade 3 species were used, and no clade 3 alleles were found in the single nuclear gene studies of [32,33]. This is the first study that documents clade 3 genomes in *S. albicans*. Of great interest is that of the four examined accessions of this species, one of them, PI 561642, is a northern disjunct in central Ecuador, the

others all distributed in central to northern Peru. Based on AFLP and morphological data, Kardolus [51] recognized a new subspecies of *S. acaule* subsp. *palmirensis* from the very accession we examined here. Although it has the hexaploid chromosome number and overall morphological similarity to *S. albicans*, AFLP data influenced him to classify it in *S. acaule* (typically tetraploid). We recognize this accession as *S. albicans*, but show a separate clade 3 genome origin for this species.

We found great variation in *S. stoloniferum*. While seven of the ten accessions showed identical origins to each other and to all six accessions of *S. hjertingii* (both species were the sole members of the Longipedicellata group), three accessions showed very different distributions of alleles. Accessions 558454 and 545787 lacked alleles from clade 1 + 2; and shared clade 4 alleles with clade *S. candolleianum* and *S. raphanifolium* but not with *S. berthaultii* and *S. verrucosum*. Accession 558466 shared clade 3 alleles with *S. andreanum* and *S. albornozi*, and clade 4 alleles with *S. berthaultii* and *S. verrucosum*. Accessions 558454 and 558466 are the only two accessions documented with alleles in clade 3.

*Solanum schenckii* PI 545733 has the same distribution of alleles as *S. stoloniferum* PI 558466. Unlike four of the six accessions of *S. schenckii*, it lacks alleles from *S. stenophyllidium*. *Solanum schenckii* PI 558456 is the only accession of this species that lacks alleles from clade 3. *Solanum demissum* PI 161719 has clade 4 alleles shared with all other 6 examined members of this species, but in addition possesses an allele from clade 1 + 2.

Various processes could explain the results we found above, to include multiple origins, introgressive hybridization subsequent to speciation, allele losses, or in the case of apparent allele losses because of procedural errors that failed to sequence “missing” alleles. We attempted to avoid the latter error, however, by our use of SSCP and cloning when expected alleles could not be located (Methods).

Multiple origins of polyploids appear to be a recurring and common pattern in plants. They have been documented in groups as diverse as the angiosperms in the Araliaceae [52], Asteraceae [24,53], Brassicaceae [54], Leguminosae [55,56], and Saxifragaceae [17]; in the bryophytes [57]; and ferns [22]. Regarding “missing” alleles, we consider missing alleles to be real, rather than an artifact of poor procedure, because of the procedures we outline in Methods. Allele loss in polyploids is appearing to be a common pattern in other groups [11]. Its cause could be stochastic, or caused by “genomic shock” during the early stages of polyploid formation [58]. Genomic changes are believed to be more common in allopolyploids than in autopolyploids, possibly correlated with greater genomic shock expected in genomically divergent parents of allopolyploids relative to diploids. The only study of genetic changes in section *Petota* was conducted in a synthetic autopolyploid [59], and this showed fewer expression differences than has been found in many allopolyploids [60]. However, this was conducted only with a first generation hybrid and was not subjected to selective forces allowing possible genomic rearrangements, so these data have little applicability here. The majority of the polyploids studied here are presumed allopolyploids or segmental autopolyploids (Table 1).

### Implications for the taxonomy of polyploids

The taxonomy of *Solanum* section *Petota* (including both the cultivated and wild potato species) is complicated by sexual compatibility among many species, introgression, interspecific hybridization, auto- and allopolyploidy, a mixture of sexual and asexual reproduction, possible recent species divergence, phenotypic plasticity, and consequent great morphological similarity and difficulty in defining and identifying species [27]. As this study and others [34-36] demonstrate, it is also complicated by phylogenetic results that are often incongruent among different phylogenetic markers (in this case nuclear ortholog markers).

Polyploids have long been recognized to be complex taxonomically and to complicate species coherence [61,62]. Thirty-six percent of the species in section *Petota* are polyploid or with diploid and polyploid cytotypes [28] and section *Petota* is notably difficult taxonomically. The present study documents considerable genomic complexity in polyploids in section *Petota*, helping to explain why taxonomists have traditionally had such difficulty in providing an easy taxonomic treatment of this group. Additional studies using more accessions and nuclear orthologs surely would expand such examples. Our results provide the very practical outcome in helping explain the cause of such taxonomic complexity, guiding taxonomists and genebank managers to rational classifications that do not search for clear differences. Clear differences will likely never be found in such systems. Our results also alert breeders to a storehouse of diversity within traditionally recognized polyploid species.

### Statistical analysis with polyploids and gene tree discordance

We faced here two major difficulties for tree inference. First, the discordance among nuclear gene trees was extensive even among diploid species. Second, the presence of multiple alleles for polyploids precluded a concatenated approach. More generally, alleles from a polyploid species cannot be paired up across different genes a priori. Much recent research has been devoted to address the first issue of species tree reconstruction from multiple conflicting gene trees [63]. In contrast, there is no standard statistical method to deal with the second issue and reconstruct the reticulate history of polyploid species from multiple gene trees. Our study illustrates a novel approach to dealing with these two issues. To account for a non-tree like history of polyploidy species, each polyploidy allele was placed separately within the diploid species tree, which was inferred using a gene tree/species tree approach to account for gene tree discordance. Finally, we summarized the results from all alleles and all accessions of a polyploid species by displaying the well supported parental origins. This workflow could be applicable to many other groups of

organisms for the inference of polyploidy origins, in the presence of extensive gene tree discordance.

### Future approaches

Buggs et al. [64] developed next generation sequencing approaches for investigating genomic changes in *Tragopogon* (Compositae) that highlight one possible next direction in this study, especially considering the availability of a genomic sequence for potato [65]. These investigators examined a wide range of genomic changes, including gene loss (quantifying the rapidity of such losses and examining parental biases in gene loss), gene silencing, subfunctionalization, and developing FISH markers for the study of genomic structural changes. Their general approach involves building an extensive genetic framework for the diploid parents via next-generation sequence data (using a combination of 454 and Illumina platforms), and then developing species-specific SNPs that are useful to investigate gene loss in the allopolyploids. They also used this approach to identify loci that exhibit apparent altered gene expression (silencing, or up- or down-regulation) in a selection of individuals of an allopolyploid relative to the parental alleles. A unique feature of their study was the use of Sequenom MassARRAY iPLEX genotyping to conduct a broad survey of homeolog loss across multiple allopolyploid populations. This method, which has been used in corn genomics, is especially suited for detecting homeologs that differ at only a few nucleotide positions. They make the point that next-generation sequencing technologies can be easily and inexpensively applied to many plant species, making any evolutionarily provocative system a potential new “model” system.

### Conclusions

Our results document considerable genomic complexity of some wild potato polyploids. These can be explained by multiple hybrid origins and allele losses that provide a clear biological explanation for the taxonomic complexity in wild potato polyploids. These results are of theoretical and practical benefit to potato breeders, and add to a growing body of evidence showing considerable complexity in polyploid plants in general.

### Methods

#### Plant materials, DNA isolation, amplification, and SSCP band sequencing

Fifty-four polyploid accessions, using 2–10 accessions per species from the *Acaulia* group (4x, 6x), *Conicibaccata* group (4x, 6x), *Longipedicellata* group, (4x), and *Iopetala* group (6x), were examined in our study (Table 1). We also examined 34 diploid accessions of 29 diploid species containing ingroup species of section *Petota* in series *Bulbocastana*, *Cuneoalata*, *Lignicaulia*, *Megistacroloba*, *Pinnatisecta*, *Piurana*, *Polyadenia*,

*Tuberosa* and *Yungasensa*, the *Conicibaccata* group, and two outgroups (*S. dulcamara*, *S. etuberosum*) (Table 1). These diploid species were chosen based on prior hypotheses of diploid progenitors of the polyploids [29,32,33,43,44], or results of phylogenetic studies within section *Petota*, including the polyploids [31-33,35,36]. DNA obtained from leaves of young plants grown from seeds in a greenhouse was extracted by the CTAB method [66] and qualified and quantified in 1% agarose gels with marker  $\lambda$  DNA digested with *Pst*I. All DNA amplification, and SSCP sequencing followed [49]. In brief, SSCP involved running SSCP, extracting the bands of interest, and sequencing them. In a few cases alleles could not be separated by SSCP because of smearing or poor amplification of the PCR products and these PCR products were then cloned and sequenced as in [34]. When we failed to obtain a DNA sequence found in prior results, or in the majority of the accessions examined here, we performed SSCP twice more. For example, in tetraploids, we expected two alleles in the PAGE or MDE gel while for hexaploids we expected three alleles. If the number of the alleles in one accession were less than these, we reran them in PAGE or MDE gels to make sure the allele number was right. Sometimes, the recovered bands of those accessions with missing alleles could not be amplified in the PCR for sequencing, and in such cases we then cloned our fragments as a final check of potential missing alleles.

#### Model selection

The molecular substitution models were evaluated with ModelTest [67] to select the preferred model among those that could be used in \*BEAST, separately for each locus. We used a likelihood ratio test to compare nested models, with a forward step-wise approach. For all loci, the selected model accounted for rate variation with a gamma-distributed rate variation among sites ( $\Gamma$ ). In five of the six nuclear orthologs the HKY +  $\Gamma$  model that includes five parameters was preferred. Only in nuclear ortholog C2At1g32130 did GTR +  $\Gamma$  best fit the data.

#### Analysis of diploid species

Our strategy was to use the diploid accessions as placeholders in a Bayesian framework to conduct further analyses to show relationships with the polyploids. In a few cases we encountered minor allelic variants from the same accession falling in the same clade in gene trees estimated with RAxML. In these cases we chose as representative the sequence that fell closest to the root of the clade, to limit the number of allele variants for future analysis.

Sequences were edited by Staden package 4.10 [68] and aligned in CLUSTALX 2.0.6 [69], with further manual alignments by MacClade 4.08 OS X [70]. The diploid

dataset was imported into BEAUTi (\*BEAST 1.6.1 package) to generate the XML format file for \*BEAST [71]. Models were selected for each COS on the basis of Akaike Information Criterion by using ModelTest 3.7. “Empirical” base frequencies were used and the Yule speciation process was selected as a prior on the species phylogeny. All MCMC chains were run for 100 million generations with subsampling every 10,000 generations and three independent runs. The three log files were then imported into Tracer 1.5 to get a combined tracer file and to check convergence to the stationary distribution and the effective sample size (ESS) of each parameter. The ESS values were exceeding 200 for all of the tree parameters except for the population size at one node, which had an ESS of 156. The sample files from the three independent runs were combined after discarding their first 10% as burn-in. They were summarized with a greedy consensus in TreeAnnotator 1.6.1 in the BEAST package. The resulting estimated trees (the diploid species tree and the six individuals COS trees) could be viewed in FigTree 1.3.1 in the BEAST package or using the package ape [72] in R [73].

#### Analysis of polyploid species

One difficulty with polyploid species is that the placement of allopolyploids or of hybrid species requires adding reticulation events in the species history, which cannot be represented by a bifurcating tree. A network is needed instead. An extra difficulty here is that we do not know which alleles come from the same side of a reticulation event. For a polyploid accession with two alleles at each locus, we can arbitrarily label each allele as “A” or “B” but the “A” alleles do not necessarily share the same parental origin. For instance, it is unknown which of allele “A” or “B” of C2At4g10050 comes from the same parental origin as allele “A” of C2At5g47390. The placement of individual alleles can still be represented in a bifurcating phylogenetic tree, which can provide some evidence about which alleles share the same placement and hence the same parental origin. In our situation however, the extent of conflict between gene trees, even across diploid species, made allele matching difficult and uncertain.

To avoid grouping alleles into putative common parental origins, we analyzed each polyploid allele separately. For each polyploid allele from each locus, we determined where the allele should be placed on the diploid species tree using the fast gene tree/species tree method “NJst,” described in [74]. This method was used instead of \*BEAST for two reasons. First, it is fast enough to be repeated once for each of the 823 alleles (and repeated 100 times to obtain bootstrap support values). Second, the NJst method could be modified to apply a subtree constraint. In our case, for each polyploid

allele, we constrained the subtree formed by the diploid species to the \*BEAST diploid species tree. By doing so, we were able to summarize the results from all polyploid alleles onto the same backbone diploid species tree.

The NJst method uses gene trees, which can include several individuals per species, and estimates a distance matrix between species. The distance between two species is defined as the average number of internodes between the two species, averaged across all gene trees and all pairs of individuals from the two species. Liu and Yu [74] showed that this distance provides a consistent estimate of the bifurcating tree topology under the coalescent model of gene tree discordance. While Saitou and Nei propose using Neighbor-Joining [75] to estimate the species trees based on the internode distance, we used instead the balanced Minimum Evolution (ME) criterion [76] for three reasons: (1) Neighbor Joining greedily aims to minimize the balanced ME criterion [77]. (2) The ME criterion can be evaluated on a set of trees for which a subtree constraint is enforced, whereas Neighbor Joining is an algorithm that builds a tree without any subtree constraint enforced. (3) The ME criterion as implemented in FastME [76] was shown to result in more accurate tree reconstruction than Neighbor Joining [78]. The modified NJst method, called the MEst method, was implemented in R [73] with an external call to fastME [76] for the calculation of the minimum evolution criterion.

For each polyploid allele of each COS, our MEst method was applied to the set of six gene trees as inferred with RAxML, where all polyploid alleles were pruned from the gene trees except for the one allele of interest. The set of candidate species trees consisted of all trees obtained from grafting the polyploidy species onto the \*BEAST diploid species tree. The candidate tree with the minimum evolution score was retained and the edge onto which the polyploid species branched off was recorded. If several candidate trees had the same best score, then all these best trees were retained and were given equal weights. For each edge in the diploid species tree, we recorded the number of polyploid alleles whose origin was estimated to be on that edge. To summarize the results at the clade level, we also counted the number of alleles for which the estimated parental origin was within the clade.

In order to account for uncertainty in gene tree estimation, this procedure was repeated 100 times, using gene trees estimated with RAxML from bootstrap sequence alignments. The input to a bootstrap replicate consisted of one bootstrap RAxML tree from each of the six COS. This bootstrap analysis resulted in a sample of size 100 for each edge and for each clade, giving the number of alleles for which the estimated parental origin was on the edge or from within the clade. We summarized each bootstrap sample for each edge and each clade

by: (1) the bootstrap support for at least one allele having a parental origin on the edge (or clade), calculated as the number of bootstrap replicates with 1 or more alleles supporting an origin on the edge, (2) the median number of alleles whose parental origin was placed on the edge (or within the clade), (3) a 90% bootstrap confidence interval for the number of alleles whose origin was placed on the edge, calculated by excluding the 5% lowest and 5% highest values in the bootstrap sample. In order to determine if accessions contributed equally to the various estimated parental origin placements, we repeated the procedure by separating out the different accessions for each polyploid species.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgments

This work was supported by the USDA and by NSF DEB (grant number 0316614; and USDA National Research Initiative grant number 2008-35300-18669) to DS and a grant from China Scholarship Council to DC, and by NSF DEB (grant number 0949121) to CA and DS. The use of trade, firm, or corporation names in this paper site is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

#### Author details

<sup>1</sup>Department of Horticulture, the State Agricultural Ministry Key Laboratory of Horticultural Plant Growth, Development and Quality Improvement, 866 Yuhangtang Road, Zhejiang University, Hangzhou, Zhejiang 310058, China. <sup>2</sup>Department of Horticulture, USDA, Agricultural Research Service, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706-1590, USA. <sup>3</sup>Centro Regional de Investigación Remehue, INIA, Xa Región de los Lagos, Km 8 Norte, Ruta 5 Sur, Casilla de Correos 24-O, Osorno, Chile. <sup>4</sup>Departments of Botany and of Statistics, 1300 University Ave., University of Wisconsin-Madison, Madison, WI 53706-1590, USA. <sup>5</sup>International Potato Center, P.O. Box 1558, Lima, 12, Peru. <sup>6</sup>Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14853, USA.

#### Authors' contributions

DS conceived the research and with LM, MB, and YT obtained funding. DC and FR helped design the research. DC performed all the laboratory work from isolating the DNA to generating the DNA sequences to DNA alignment, to data analyses with RAXML and \*BEAST. FR designed and screened the primers, developed the SSCP protocol, and advised us on laboratory techniques. CA designed the ME species tree method for polyploids and developed the associated computing pipeline in R. All authors revised several versions of the manuscript. LM generated nuclear orthologs and handled data archiving. All the authors reviewed and approved the final manuscript.

Received: 21 February 2012 Accepted: 24 May 2012

Published: 24 May 2012

#### References

- Goldblatt P: **Polyploidy in angiosperms: monocotyledons.** In *Polyploidy: biological relevance*. Edited by Lewis WH. New York: Plenum Press; 1980:219–239.
- Lewis WH: **Polyploidy in angiosperms: dicotyledons.** In *Polyploidy: biological relevance*. Edited by Lewis WH. New York: Plenum Press; 1980:241–268.
- Wendel JF: **Genome evolution in polyploids.** *Plant Mol Biol* 2000, **42**:225–249.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Patterson AH, Zheng C, Sankoff D, DePamphilis CW, Wall PK, Soltis PS: **Polyploidy and angiosperm diversification.** *Amer J Bot* 2009, **96**:336–348.
- Soltis PS, Soltis DE: **The role of genetic and genomic attributes in the success of polyploids.** *Proc Natl Acad Sci USA* 2000, **97**:7051–7057.
- Wolfe KH: **Yesterday's polyploidization and mystery of diploidization.** *Nat Rev Genet* 2001, **2**:333–341.
- Adams KL, Cronn R, Percifield R, Wendel JF: **Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing.** *Proc Natl Acad Sci USA* 2003, **100**:4649–4654.
- Comai L: **The advantages and disadvantages of being polyploid.** *Nature Rev Genet* 2005, **6**:836–846.
- Otto SP: **The evolutionary consequences of polyploidy.** *Cell* 2007, **131**:452–462.
- Sémon M, Wolfe KH: **Consequences of genome duplication.** *Curr Opin Genetics Dev* 2007, **17**:505–512.
- Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE: **On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae).** *BMC Plant Biol* 2009, **9**:80. doi:10.1186/1471-2229-9-80.
- Doyle JJ, Egan AN: **Dating the origins of polyploidy events.** *New Phytol* 2010, **186**:73–85.
- Chen ZJ: **Molecular mechanisms of polyploidy and hybrid vigor.** *Trends Plant Sci* 2010, **15**:57–71.
- Carputo D, Camadro EL, Peloquin SJ: **Terminology for polyploids based on their cytogenetic behavior: consequences in genetics and breeding.** *Plant Breed Rev* 2006, **26**:105–124.
- Soltis DE, Soltis PS: **Molecular data and the dynamic nature of polyploidy.** *Crit Rev Plant Sci* 1993, **12**:243–273.
- Soltis DE, Soltis PS: **Polyploidy: recurrent formation and genome evolution.** *Trends Ecol Evol* 1999, **14**:348–352.
- Segraves KA, Thompson JN, Soltis DE, Soltis PS: **Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*.** *Mol Ecol* 1999, **8**:253–262.
- Vanichanon A, Blake NK, Sherman JD, Talbert LE: **Multiple origins of allopolyploid *Aegilops triuncialis*.** *Theor Appl Genet* 2003, **106**:804–810.
- Albach DC: **Amplified fragment length polymorphisms and sequence data in the phylogenetic analysis of polyploids: multiple origins of *Veronica cymbalaria* (Plantaginaceae).** *New Phytol* 2007, **176**:481–498.
- Grubbs KC, Small RL, Schilling EE: **Evidence for multiple, autopolloid origins of agamosperous populations in *Eupatorium sessilifolium* (Asteraceae).** *Plant Syst Evol* 2009, **279**:151–161.
- Meimberg H, Njoku CC, McKay JK, Rice KJ, Milan NF: **Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae).** *Amer J Bot* 2009, **96**:1262–1273.
- Werth CR, Guttman SJ, Eshbaugh WH: **Recurring origins of allopolyploid species in *Asplenium*.** *Science* 1985, **228**:731–733.
- Soltis PS, Plunkett GM, Novak SJ, Soltis DE: **Genetic variation in *Tragopogon* species: additional origins of the allotetraploids *T. mirus* and *T. miscellus* (Compositae).** *Amer J Bot* 1995, **82**:1329–1341.
- Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate J, Mavrodiev E: **Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): genetics, genomic, and cytogenetic comparisons.** *Biol J Linn Soc* 2004, **82**:485–501.
- Spooner DM, Núñez J, Trujillo G, del Rosario Herrera M, Guzmán F, Ghislain M: **Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification.** *Proc Natl Acad Sci USA* 2007, **104**:19398–19403.
- Ovchinnikova A, Krylova E, Gavrilenko T, Smekalova T, Zhuk M, Knapp S, Spooner DM: **Taxonomy of cultivated potatoes (*Solanum* section *Petota*: *Solanaceae*).** *Bot J Linn Soc* 2011, **165**:107–155.
- Spooner DM: **DNA barcoding will frequently fail in complicated groups: an example in wild potatoes.** *Amer J Bot* 2009, **96**:1177–1189.
- Hijmans R, Gavrilenko T, Stephenson S, Bamberg J, Salas A, Spooner DM: **Geographic and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*).** *Global Ecol Biogeogr* 2007, **16**:485–495.
- Hawkes JG: *The potato: evolution, biodiversity, and genetic resources.* Washington, DC: Belhaven Press; 1990.
- Hanneman RE Jr: **Assignment of endosperm balance numbers to the tuber-bearing *Solanums* and their close non-tuber bearing relatives.** *Euphytica* 1994, **74**:19–25.
- Spooner DM, Castillo R: **Reexamination of series relationships of South American wild potatoes (*Solanaceae*: *Solanum* sect. *Petota*): evidence from chloroplast DNA restriction site variation.** *Amer J Bot* 1997, **84**:671–685.
- Spooner DM, Rodríguez F, Polgár Z, Ballard HE Jr, Jansky SH: **Genomic origins of potato polyploids: GBSSI gene sequencing data.** *The Plant Genome, a Suppl to Crop Sci* 2008, **48**(Suppl 1):S27–S36.

33. Rodríguez F, Spooner DM: Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. *Syst Bot* 2009, **34**:207–219.
34. Rodríguez F, Wu F, Ané C, Tanksley S, Spooner DM: Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? *BMC Evol Biol* 2009, **9**:191. doi:10.1186/1471-2148-9-191.
35. Ames M, Spooner DM: Phylogeny of *Solanum* series *Piurana* and related species in *Solanum* section *Petota* based on five conserved ortholog sequences. *Taxon* 2010, **59**:1091–1104 + 4pg. foldout.
36. Fajardo D, Spooner DM: Phylogenetic relationships of *Solanum* series *Conicibaccata* and related species in *Solanum* section *Petota* inferred from five conserved ortholog sequences. *Syst Bot* 2011, **36**:163–170.
37. Rodríguez F, Ghislain M, Clausen AM, Jansky SH, Spooner DM: Hybrid origins of cultivated potatoes. *Theor Appl Genet* 2010, **121**:1187–1198.
38. Peralta IE, Spooner DM, Knapp S: The taxonomy of tomatoes: a revision of wild tomatoes (*Solanum* section *Lycopersicon*) and their outgroup relatives in sections *Juglandifolium* and *Lycopersicoides*. *Syst Bot Monogr* 2008, **84**:1–186 + 3 plates.
39. Spooner DM, Van den Berg RG, Bamberg JB: Examination of species boundaries of *Solanum* series *Demissa* and potentially related species in series *Acaulia* and series *Tuberosa* (sect. *Petota*). *Syst Bot* 1995, **20**:295–314.
40. Kardolus JP: Morphological variation within series *Acaulia* Juz. (*Solanum* sect. *Petota*). In *Solanaceae IV: advances in biology and utilization*. Edited by Nee M, Symon DE, Lester RN, Jessop JP. Kew: Royal Botanic Gardens; 1999:257–274.
41. Nakagawa K, Hosaka K: Species relationships between a wild tetraploid potato species, *Solanum acaule* Bitter, and its related species as revealed by RFLPs of chloroplast and nuclear DNA. *Amer J Potato Res* 2002, **79**:85–98.
42. Spooner DM, Van den Berg RG, Rodríguez A, Bamberg J, Hijmans RJ, Lara-Cabrera SI: Wild potatoes (*Solanum* section *Petota*) of North and Central America. *Syst Bot Monogr* 2004, **68**:1–209.
43. Matsubayashi M: Phylogenetic relationships in the potato and its related species. In *Chromosome engineering in plants: genetics, breeding, evolution, part B*. Edited by Tsuchiya T, Gupta PK. Amsterdam: Elsevier Science BV; 1991:93–118.
44. Pendinen G, Gavrilenko T, Jiang J, Spooner DM: Allopolyploid speciation of the tetraploid Mexican potato species *Solanum stoloniferum* and *S. hjertingii* revealed by genomic in situ hybridization. *Genome* 2008, **51**:714–720.
45. Wu FN, Mueller LA, Crouzillat D, Petiard V, Tanksley SD: 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* 2006, **174**:1407–1420.
46. Levin RA, Whelan A, Miller JS: The utility of nuclear conserved ortholog set II (COSII) genomic regions for species-level phylogenetic inference in *Lycium* (*Solanaceae*). *Mol Phylogenet Evol* 2009, **53**:881–890.
47. Zhou L, Yang CF, Xiao LH: PCR-mediated recombination between *Cryptosporidium* spp. of lizards and snakes. *J Eukaryot Microbiol* 2003, **50**:563–565.
48. Ortí G, Hare MP, Avise JC: Detection and isolation of nuclear haplotypes by SSCP. *Mol Ecol* 1997, **6**:575–580.
49. Rodríguez F, Cai D, Teng Y, Spooner DM: Asymmetric single-strand conformation polymorphism: an accurate and cost-effective method to amplify and sequence allelic variants. *Amer J Bot* 2011, **98**:1061–1067.
50. Hosaka K, Spooner DM: RFLP analysis of the wild potato species, *Solanum acaule* Bitter (*Solanum* sect. *Petota*). *Theor Appl Genet* 1992, **84**:851–858.
51. Kardolus JP: A biosystematic analysis of *Solanum acaule*. In *Ph.D. Thesis*. Wageningen, The Netherlands: Wageningen Agricultural University; 1998.
52. Valcárcel V, Fiz O, Vargas P: Chloroplast and nuclear evidence for multiple origins of polyploids and diploids of *Hedera* (*Araliaceae*) in the Mediterranean basin. *Molec Phylo Evol* 2003, **27**:1–20.
53. Ashton PA, Abbott RJ: Multiple origins and genetic diversity in the newly arisen allopolyploid species, *Senecio cambrensis* Rosser (*Compositae*). *Heredity* 1992, **68**:25–32.
54. Brochmann C, Soltis PS, Soltis DE: Multiple origins of the octoploid Scandinavian endemic *Draba cacuminum*: electrophoretic and morphological evidence. *Nordic J Bot* 1992, **12**:257–272.
55. Doyle JJ, Doyle JL, Brown AHD, Palmer RG: Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (*Leguminosae*) polyploid complex: histone H3-D gene sequences. *Evolution* 2002, **56**:1388–1402.
56. Rauscher JT, Doyle JJ, Brown AHD: Multiple origins and nrDNA internal transcribed spacer homeologue evolution in the *Glycine tomentella* (*Leguminosae*) allopolyploid complex. *Genetics* 2004, **166**:987–998.
57. Wyatt R, Odrzykoski IJ, Stoneburner A, Bass HW, Galau GA: Allopolyploidy in bryophytes: multiple origins of *Plagiomnium medium*. *Proc Natl Acad Sci USA* 1988, **85**:5601–5604.
58. Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC: Genomic changes in resynthesized *Brassica napus* and their effects on gene expression and phenotype. *Plant Cell* 2007, **19**:3403–3417.
59. Stupar RM, Bhaskar PB, Yandell BS, Rensink WA, Hart AL, Ouyang S, Veilleux RE, Busse JS, Erhardt RJ, Buell CR, Jiang J: Phenotypic and transcriptional changes associated with potato autopolyploidization. *Genetics* 2007, **176**:2055–2067.
60. Doyle JJ, Flagel LE, Patterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF: Evolutionary genetics of genome merger in plants. *Ann Rev Genet* 2008, **42**:443–461.
61. Rieseberg L, Willis JH: Plant speciation. *Science* 2007, **317**:910–914.
62. Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS: Have we grossly underestimated the number of species? *Taxon* 2007, **56**:13–30.
63. Knowles LL, Kubatko LS S (Eds): *Estimating species trees: practical and theoretical aspects*. Hoboken, New Jersey: Wiley-Blackwell; 2010.
64. Buggs RJA, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, Leitch AR, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE: Next-generation sequencing and genome evolution in allopolyploids. *Amer J Bot* 2012, **99**:372–382.
65. The Potato Sequencing Consortium: Genome sequence and analysis of the tuber crop potato. *Nature* 2011, **475**:189–197.
66. Doyle J: DNA protocols for plants-CTAB total DNA isolation. In *Molecular techniques in taxonomy*. Edited by Hewitt GM, Johnston A. Berlin: Springer; 1991:283–293.
67. Posada D, Crandall KA: Modeltest: testing the model of DNA substitution. *Bioinformatics* 1998, **14**:817–818.
68. Staden R: The Staden sequence analysis package. *Mol Biotechnol* 1996, **5**:233–241.
69. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997, **24**:4876–4882.
70. Maddison DR, Maddison WP: *MacClade 4.03: Analysis of phylogeny and character evolution*. Sunderland: Sinauer Associates; 2001.
71. Heled J, Drummond AJ: Bayesian inference of species trees from multilocus data. *Mol Biol Evol* 2010, **27**:570–580.
72. Paradis E, Claude J, Strimmer K: APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 2004, **20**:289–290.
73. R Development Core Team: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2011. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
74. Liu L, Yu L: Estimating species trees from unrooted gene trees. *Syst Biol* 2011, **65**:661–667.
75. Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987, **4**:406–425.
76. Desper R, Gascuel O: Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol* 2002, **9**:687–705. software available: [www.atgc-montpellier.fr/fastme/](http://www.atgc-montpellier.fr/fastme/).
77. Gascuel O, Steel M: Neighbor-joining revealed. *Mol Biol Evol* 2006, **23**:1997–2000.
78. Desper R, Gascuel O: Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol Biol Evol* 2004, **21**:587–598.

doi:10.1186/1471-2148-12-70

Cite this article as: Cai et al.: Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section *Petota*). *BMC Evolutionary Biology* 2012 12:70.