# Searching for species, relationships and resistance in *Solanum* section *Petota*

Mirjam M.J. Jacobs

Promotoren:

**Prof. dr. M.S.M. Sosef**
Hoogleraar Biosystematiek, leerstoelgroep Biosystematiek,
Wageningen Universiteit

**Prof. dr. R.G.F. Visser**
Hoogleraar Plantenveredeling, leerstoelgroep Plantenveredeling,
Wageningen Universiteit

Co-promotoren:

**Dr. B.J. Vosman**
Senior onderzoeker, Plant Research International,
Wageningen UR

**Dr. R.G. van den Berg**
Universitair hoofddocent Biosystematiek, leerstoelgroep Biosystematiek,
Wageningen Universiteit

Promotiecommissie:

**Prof. dr. E.F. Smets**
Universiteit Leiden

**Prof. dr. W.J. Stiekema**
Wageningen Universiteit

**Dr. ir. L. Visser**
Centrum voor Genetische Bronnen Nederland

**Dr. ir. W.J.M. Koopman**
Koopman Scientific Services, Enkhuizen

Dit onderzoek is uitgevoerd binnen de onderzoeksschool EPS (Experimental Plant Sciences)

# Searching for species, relationships and resistance in *Solanum* section *Petota*

Mirjam M.J. Jacobs

**Proefschrift**

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. M.J. Kropff
in het openbaar te verdedigen
op vrijdag 31 oktober 2008
des namiddags half twee in de Aula

# CHAPTER 1

## General introduction

Mirjam M.J. Jacobs

**The potato**

The year 2008 has been declared the United Nations International Year of the Potato, intended to "raise awareness of the key role played by the 'humble tuber' in agriculture, economy and world food security" (http://www.potato2008.org/). Ironically, in May 2008 a conflict between Peru and Chile arose about the origin of the cultivated potato, Chile claiming that 99% of the world's potatoes derive from material native to its territory, while Peru points to the area near Lake Titicaca as the site of origin of the crop. The origin of the potato has thus become a 'hot potato' (http://www.usatoday.com/ news/ world/2008-05-27-peru-chile-potato_N.htm). Regardless of who is right in this case, the discussion demonstrates the importance given to the issue of genetic resources of the cultivated potato. Even in these modern times it is worth competing for the honour of being the country of origin of the potato.

Potato is a crop with a long history. It was cultivated for many millennia in the Andes region in South America. Plant remains from archaeological sites, dating back as far as 2500 BC and 5000 BC, have provided evidence for ancient potato cultivation (Hawkes, 1990). After the conquest of the Americas in the 16th century by the Spanish conquistadores they were successfully introduced in Europe and were distributed from there to other continents. In most countries in Europe and in the Americas the potato is still one of the largest sources of starch in the daily diet and the potato production still continues to increase, mainly in the developing countries (http://www.cipotato.org/ potato/facts/growth.asp).

**Potato systematics**

Because potato is such an economically important crop, it is not surprising that its botany and taxonomy has been the subject of intensive study for many years. The crop has been classified in a number of cultivated taxa (the species *Solanum tuberosum*, *S. ajanhuiri*, *S. chaucha*, *S. curtilobum*, *S. juzepczukii*, *S. phureja* and *S. stenotomum*; Hawkes 1990) but these taxa have recently (Huaman, Spooner, 2002) been considered as formal cultivar-groups within one cultigenic species *S. tuberosum*, following the rules of the International Code of Nomenclature for Cultivated Plants (Brickell *et al.*, 2004). Besides *Solanum tuberosum* there are circa 200 wild tuber-bearing species, belonging to section *Petota* Dumort. within subgenus *Potatoe* (G. Don) D'Arcy of the large genus *Solanum*.

It was recognized early that these wild relatives of the cultivated potato could provide crossing material to improve the cultivated material, so they have been the subject of study already since the 19th century. At the end of the 19th century and the beginning of the 20th century, taxonomists like G. Don, Bitter and Dunal produced the first classifications of the group of tuber-bearing *Solanum* species. Later, especially Russian scientists extended the knowledge on the taxonomy of wild potatoes, Vavilov and other scientists carried out many expeditions in South America. After the Second World War, J.G. Hawkes, C. Ochoa and many others described substantial numbers of new taxa.

The taxonomy of the tuber-bearing species is complicated because of the occurrence of phenomena like polyploidization, hybridization and morphological plasticity. Furthermore, crossing barriers between certain species are presumably influenced by an unknown mechanism called EBN (Embryo Balance Number) (Hawkes & Jackson, 1992; Johnston *et al.*, 1980) which adds to the confusion. Despite many extensive studies from various taxonomists several taxonomic problems still remain. These are mainly:

1) The difficulty of correct identification using morphological keys;

2) Over-classification of parts of section *Petota* (overclassification means in this case that (too) many taxa like species, subspecies, varieties, etc., have been assigned to explain the variation visible, but that less species exist in reality);

3) Problematic classification of the series (Spooner & Salas, 2006; Spooner & van den Berg, 1992).

Nonetheless, the increasing application of molecular methods provides hope for a more suitable and durable potato taxonomy. Until now several molecular studies using various methods like RAPDs, AFLP and RFLP have been conducted. However, most molecular studies have only focused on a small part of the variation present within section *Petota* (Spooner & Castillo, 1997; Spooner *et al.*, 1996); (Sukhotu & Hosaka, 2006; van den Berg *et al.*, 2002).

**Phytophthora infestans and its impact on potato cultivation**

The name *Phytophthora infestans* (Mont.) de Bary (de Bary, 1876) suits the disease this pathogen causes. "Phytophthora" means in classic Greek "plants destroyer" and "infestans" refers to its infectious abilities. *P. infestans* causes late blight, the most important disease in potato cultivation. Late blight has the ability to destroy entire fields of potato in a few weeks or even days. It affects foliage and stems and additionally it can also infect fruits and tubers (Fry, 2008). In 1845 and 1846 severe late blight epidemics destroyed potato crops in the whole of Europe, causing the infamous Irish potato famine and the following mass emigration of Irish people to the U.S.A. Nowadays, late blight remains a major problem in potato production. The costs of control efforts and lost production are estimated at more than $3 billion dollar each year (CIP, 1996). The control of the disease heavily dependents on the use of fungicides, but despite the frequent use late blight still proofs increasingly difficult to control (Erwin & Ribeiro, 1996; Fry, 2008; Hijmans *et al.*, 2000).

*P. infestans* belongs to the oomycetes, organisms that resemble fungi but are more closely related to the algae. One of the many features that distinguish oomycetes from true fungi is that oomycetes are diploid and lack a free haploid life stage (Judelson & Blanco, 2005). In the normal, asexual, life cycle of *P. infestans* the pathogen forms a mycelium in the host plant. It then produces zoospores, which detach easily from the mycelium thus spreading the disease rapidly (Grunwald & Flier, 2005).

Until the 20th century, *P. infestans* was known as an asexual organism worldwide existing only as mating type A1, except for central Mexico where A1 and A2 mating types were found. Recently, since the 80ths the situation has changed, both mating types are now present in Europe and other parts of the world, and sexual reproduction can occur. Through sexual reproduction the pathogen produces oospores that stay alive longer than zoospores and can hibernate in the soil. The sexual life-cycle complicates the fight against *P. infestans* because it gives the pathogen the possibility of enhancing its genetic variation while it also produces longer-living spores (Fry, 2007).

Already in the first half of the 20th century it was recognized that potato clones reported to be resistant elsewhere became severely diseased when exposed to populations of *P. infestans* in the Toluca Valley in Central Mexico (Galindo & Gallegly, 1960; Gallegly & Galindo, 1958; Niederhauser & Millis, 1953). It is therefore not surprising that the A2 mating type was first reported from the Toluca valley (Galindo & Gallegly, 1960; Gallegly & Galindo, 1958). The region of the Toluca valley is considered to be the centre of origin and genetic diversity for *P. infestans* (Flier *et al.*, 2003; Grunwald & Flier, 2005) and it stills plays an important role in the study of the biology of *P. infestans*.

**Late blight resistance genes in potato**

Plants are attacked by a wide range of organisms including viruses, bacteria, oomycetes, fungi, nematodes and insects. They have evolved passive and active ways to defend themselves against these attackers. One of the active defence systems is a type of immunity that is described by the "gene for gene" resistance theory, which was developed by Flor in the 1940's. It considers the gene causing resistance, the R gene in the host, to be complementary to an *avr* (avirulence) gene in the pathogen (Flor, 1942). The pathogen infection in the plant leads to the recognition of an avirulence gene product, a so-called elicitor, by a corresponding R gene product in the host plant. The initial recognition sets of a complex cascade of defence responses that eventually all lead to the restriction or the further development of the pathogen (Keen, 1990). If the plant lacks the appropriate R gene or the pathogen lacks the *avr* gene, activation of plant defence responses may be delayed or ineffective and the disease can develop (Thatcher, 2005).

To date, more than 90 resistance genes have been cloned from plants, by a wide variety of methods including map-based cloning, transposon tagging, and homology based DNA library screening (Ingvardsen *et al.*, 2008). In the genus *Solanum*, many R genes have been mapped and cloned (sequenced) in the last two decades (van Ooijen *et al.*, 2007). Some of these R genes confer resistance to *P. infestans*, some to potato virus X, and others to potato cyst nematodes. The R genes are sometimes found in cultivated potato germplasm but mainly originate from wild potato germplasm. Many of the R genes in *Solanum* seem to be positioned in relatively few DNA clusters (Bakker *et al.*, 2003; Wang *et al.*, 2008). Within these clusters, repeats of similar genes, as well as several different genes can be recognized. For the clusters in the *Solanum* genome, as known so far, we refer to the 'SOLanaceae Function Map for Pathogen Resistance', which is compiled by Gebhardt and co-workers and is a representation of published literature in the form of a genetic map (Meyer *et al.*, 2005) (http://gabi.rzpd.de/projects/Pomamo/).

Over the last century, 11 late blight resistance genes were introduced into cultivated potato from the wild potato species *S. demissum* (Gebhardt & Valkonen, 2001). As the resistances conferred by these R genes were quickly broken by the pathogen (Wastie, 1991), the presence of R genes in other relatives of the cultivated potato was investigated as well. In the following species late blight R-genes or QTLs have been identified and mapped: *S. microdontum*, *S. mochiquense*, *S. paucissectum*, *S. spegazinni*, *S. pinnatisectum*, *S. berthaultii* and *S. bulbocastanum* and *S. stoloniferum* (Bisognin *et al.*, 2005; Ewing *et al.*, 2000; Ghislain *et al.*, 2001; Kuhl *et al.*, 2001; Naess *et al.*, 2000; Oberhagemann *et al.*, 1999; Park *et al.*, 2005; Rauscher *et al.*, 2006; Sandbrink *et al.*, 2000; Sliwka *et al.*, 2006; Smilde *et al.*, 2005; van der Vossen *et al.*, 2003; Villamon *et al.*, 2005; Wang *et al.*, 2008).
Most R genes can be assigned to one of the five major classes of R genes (Dangl & Jones, 2001). The largest of these classes contains genes that encode proteins with a nucleotide binding site and a leucine-rich repeat region (the so called NBS-LRR genes). NBS-LRR resistance genes and the resistance gene analogs (RGA's) are numerous in plant genomes and are often organized in clusters (AGI, 2000; Michelmore & Meyers, 1998). RGA's are parts of the genome that are presumed genes and share conserved common motifs with known R genes. They may possibly also code for proteins involved in resistance physiology, but this has yet to be described. The NBS region of R genes and RGA's contain highly conserved common motifs like the P-loop, the kinase-2 motif and the GLPL motif (Meyers *et al.*, 1999; Meyers *et al.*, 2003; Monosi *et al.*, 2004). The conserved motifs within the NBS-LRR genes have been used successfully to sequence (parts of) NBS regions from various plant species (Collins *et al.*, 1998; Pflieger *et al.*, 1999; van der Linden *et al.*, 2004; Zhang *et al.*, 2007). Van der Linden et al. (2004) published on a method called Nucleotide Binding Site profiling (NBS profiling). NBS profiling is a PCR based method that uses primers that target different conserved motifs in the NBS domain. It produces a DNA profile that is highly enriched for R genes and RGA's. Studies in apple (Calenge *et al.*, 2005) and in potato, tomato, barley and lettuce (van der Linden *et al.*, 2004) show that NBS profiling produces markers that are tightly linked to R genes and R gene clusters. The major advantage of this method is that it can be applied to study resistance in plants, even if there is no information available on the resistance gene present in the plant (Wang *et al.*, 2008).

**Aims and scope of this study**

The present PhD project is part of the potato programme carried out within the Centre for Biosystems Genomics (CBSG). The CBSG is a network of Dutch scientists in the field of plant genomics, as well as Dutch companies involved in plant genomics, breeding, cultivation and processing. The aim of the CBSG is to contribute to sustainable improvement of important world food and non-food crops. The potato programme is composed of a resistance and a quality part, each with a number of subprojects that are strongly interconnected (http://www.cbsg.nl/). In the potato programme, one project analyses over 1000 potato accessions for *P. infestans* resistance, using the same plants that were used in this PhD project to analyse the biosystematics, while another project delivered information on the locations of the Resistance Gene Homologs on the different chromosomes. The present project (P4) has close relationships with these subprojects P1 and P3 (Figure 1).

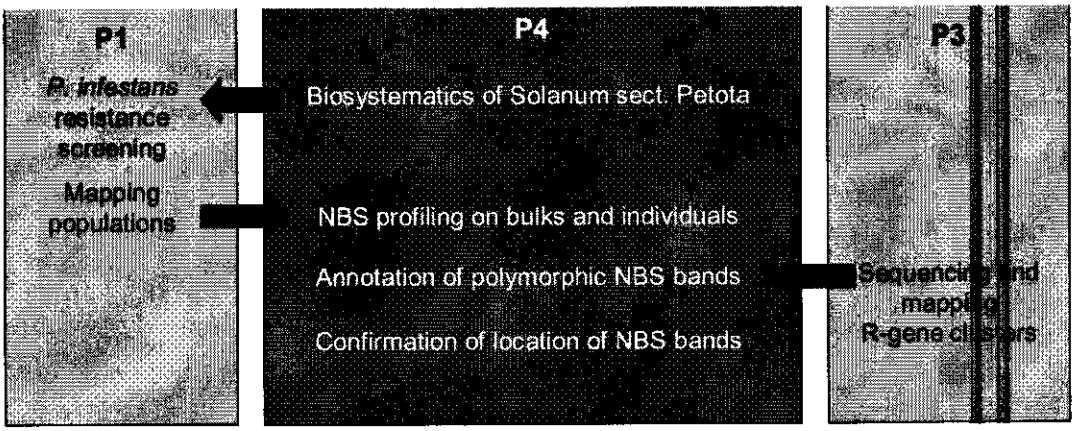



*Figure 1. The relationships of project P4 (this PhD thesis) and other resistance projects within CBSG.*

The evolution of new races of *P. infestans* is rapid and the spread of races that have overcome resistance genes is a serious problem. As a consequence, there is a continuous and growing need to find new R genes that can be deployed in breeding (Fry, 2008). The wild gene pool of *S. tuberosum* can offer new resources of R genes. There are many wild species that have not been tested yet for the presence of R genes against *P. infestans*. For an efficient study of valuable traits in general and *P. infestans* resistance in particular, it is important to have information on the phylogenetic relationships of the wild tuber-bearing species. This will ensure that the search for new resistances efficiently utilizes all of the possible sources and will not be restricted to just a part of the group.

The present study aimed at elucidating the taxonomy of wild potatoes and searching for new resistance genes against *P. infestans*. To elucidate the systematic relationships of the wild *Solanum* section *Petota* taxa, we analysed a large AFLP dataset. No other study has been based on such extensive sampling. The taxonomic information generated will allow the easy selection of genebank accessions for a range of purposes, amongst others for finding new sources of resistance.

**Thesis contents**

Chapter 2 provides a review of the results of molecular analyses of datasets of tuber-bearing *Solanum* species. A short introduction on the classical taxonomy of section *Petota* is given, followed by an overview of recent new insights. Remaining taxonomic problems and remaining needs for research are discussed. The purpose of Chapter 3 is to gain insight in the internal taxonomic structure of section *Petota* by using both cpDNA sequences and AFLP patterns for the same individuals. This allows a direct comparison of the results of both methods. This reveals major incongruencies between the AFLP data and the chloroplast data which might point at the occurrence of former hybridization events. In Chapter 4 the taxonomic structure present in section *Petota* is analyzed by using AFLP. Phylogenetic and a phenetic analysis were performed on the largest dataset ever constructed for *Solanum* section *Petota.* In total, around 1000 accessions were sampled, and approximately 5000 individual plants were genotyped using over 200 AFLP markers. The data obtained were used to evaluate the 21 series hypothesis put forward by Hawkes and the 4 clade hypothesis of Spooner and co-workers. From the results in Chapters 3 and 4 we learnt that most of the wild *Solanum* species from South America are closely related and that species boundaries between many species are unclear. In Chapter 5, using a population genetics approach, the status of species names (species labels) in *Solanum* section *Petota* was investigated with a strong focus on the South American species. Herewith, we attempt to elucidate the inner systematic structure of the South American part of *Solanum* section *Petota*. In Chapter 6, a novel approach to map the position of new resistance genes is presented and tested. It aims at quick identification of the gene cluster and obtaining markers that can be used for introgression breeding. Our approach consisted of combining NBS profiling on small segregating populations, followed by sequencing and annotating of polymorphic NBS bands and confirming map positions by using PCR based markers. The thesis is completed with a general discussion in Chapter 7, in which the results from all the chapters are evaluated and concluding remarks are made.

# CHAPTER 2

## Molecular Taxonomy

Ronald G. van den Berg and Mirjam M.J. Jacobs

(Review, adapted from a chapter in Potato Biology and Biotechnology; Advances and Perspectives, 2007, edited by D. Vreugdehil)

## Introduction

In this chapter, we look at the results of analyses of molecular data sets that have been used to answer questions on the taxonomy of the group of tuber-bearing *Solanum* spp., *Solanum* sect. *Petota*, the cultivated potato and its wild relatives.

## Taxonomic background

### Wild and cultivated potatoes

The cultivated potato is an unusual crop in that it has an extremely large secondary genepool consisting of related wild species that are tuber-bearing, albeit with small inedible tubers. The taxonomy of the cultivated potato and its wild relatives has been the subject of study for many years. Most of these studies relied on morphological observations and, on a limited scale, experimental methods like cytogenetics and hybridization experiments. More than 200 species have been described and many infraspecific taxa. These taxa have been classified in series, with different authors recognizing different numbers of series, often with different circumscriptions. Two authorative treatments (Correll, 1962; Hawkes, 1990) recognized 26 and 21 series, respectively (Table 1). Hawkes (1989) suggested a division of the series into two superseries, *Stellata* and *Rotata*, emphasizing the outline of the corolla as a major distinctive character. Some of the series contain only one or just a few species, indicating that their relationship to the other species is not clear. On the contrary, series such as *Piurana* and, especially, *Tuberosa*, are large groups of species that may not be closely related to each other. Hijmans and Spooner (2001) and Hijmans et al. (2002) documented the geographic distribution of wild potato species, with the majority occurring in Argentina, Bolivia, Mexico and Peru, many with only restricted distribution areas.

There are polyploid series present with diploids, triploids, tetraploids, pentaploids and hexaploids. The polyploids are considered to be allopolyploids derived from hybdridization events involving 2n gametes. The odd numbered polyploids, while mostly sterile, are able to maintain themselves vegetatively through the tubers. The cultivated potato, *Solanum tuberosum* L., is accommodated in series *Tuberosa*, a rather large and variable group without clear diagnostic characters. The link between wild and cultivated potatoes, the direct ancestors of the crop, must be looked for in the so-called brevicaule complex, a group of morphologically variable, diploid species within series *Tuberosa*. Within this complex, about 20 species have been distinguished, but Ugent (1966) suggested that these could be drastically reduced to one species (*Solanum brevicaule*) and Van den Berg et al. (1998) by and large confirmed that conclusion. Morphologically, many of the wild species in the brevicaule complex are similar to some of the cultivated potatoes, the main differences being found in leaf dissection, in corolla colour and – obviously – in the tuber.

*Table 1. series according to Hawkes (1990)*

| **Series according to Hawkes (1990)** |
|---|
| Subsection *Estolonifera* |
| Series *Etuberosa* |
| Series *Juglandifolia* |
| Subsection *Potatoe* |
| Superseries *Stellata* |
| Series *Morelliforme* |
| Series *Bulbocastana* |
| Series *Pinnatisecta* |
| Series *Polyadenia* |
| Series *Commersoniana* |
| Series *Circaeifolia* |
| Series *Lignicaulia* |
| Series *Olmosiana* |
| Series *Yungasensa* |
| Superseries *Rotata* |
| Series *Megistacroloba* |
| Series *Cuneoalata* |
| Series *Conicibaccata* |
| Series *Piurana* |
| Series *Ingifolia* |
| Series *Maglia* |
| Series *Tuberosa* |
| Series *Acaulia* |
| Series *Longipedicellata* |
| Series *Demissa* |

The origin of the cultivated potatoes has been described as the result of successive hybridizations between diploid members of the brevicaule complex, accompanied by chromosome doubling leading to the tetraploid forms. The crop itself has been classified into seven cultivated species (*Solanum ajanhuiri, Solanum chaucha, Solanum curtilobum, Solanum juzepczukii, Solanum phureja, Solanum stenotomum* and *S. tuberosum* with two subspecies, *tuberosum* and *andigena*), showing several ploidy levels. The discussion about the taxonomic status of cultivated plant material (Hetterscheid & Brandenburg, 1995) suggests that the taxon 'species' (with its connotation of a product resulting from evolutionary processes) is not suitable for the classification of cultivated plants as the influence of humans seriously disturbs the patterns of variation used to classify species. Rather, cultivated material should be treated as artificial entities such as landraces or cultivars and classified into cultivar groups as advocated in the International Code of Nomenclature of Cultivated Plants (ICNCP, 2004).

This was anticipated by Dodds (1962), who, in an appendix to Correll's book, suggested the informal groups Stenotomum, Phureja, Chaucha, Andigena and Tuberosum within the species *S. tuberosum* to accommodate the cultivated potatoes. Huaman and Spooner (2002) suggested a similar solution with eight groups (Ajanhuiri, Juzepczukii, Curtilobum, Chilotanum, Andigenum, Chaucha, Phureja and Stenotomum). The crop 'potato', making up the total of these groups, can still be assigned to the 'species' *S. tuberosum*, if so desired. This species name should then be considered as a cultigen (a species consisting of cultivated plants only, and as such without wild representatives, without a natural geographic distribution area and without a natural population structure). If the six other species names are used, these too are to be considered as cultigens (Table 2).

*Table 2. Alternative classifications of the cultivated potatoes*

| Cultigen | Groups Dodds (1962) | Groups Huaman & Spooner (2002) |
|---|---|---|
| *Solanum tuberosum* | Tuberosum group | Modern varieties, cultivar-group name(s) yet to be proposed |
| *Solanum stenotomum* | Stenotomum Group | Stenotomum group |
| *Solanum phureja* | Phureja | Phureja Group |
| *Solanum chaucha* | Chaucha | Chaucha Group |
| *Solanum andigena* | Andigena | Andigenum Group |
| *Solanum curtilobum* | *S. x curtilobum* | Curtilobum Group |
| *Solanum juzepczukii* | *S. x juzepczukii* | Juzepczukii Group |
| *Solanum ajanhuiri* | | Ajanhuiri Group |
| | | Chilotanum Group |

**The evolutionary framework**

The place of origin of the group of tuber-bearing potato species has been suggested to be the Mexican/ Central American area, where those species are found that are considered to be phylogenetically primitive. These species are diploids, with stellate corollas and an endosperm balance number (EBN) of 1 (EBN refers to a genetic isolating mechanism that allows crosses between species with the same EBN and prevents crosses between different EBN groups; there are five combinations of ploidy level and EBN that determine crossability groups: 2x/EBN1, 2x/EBN2, 4x/EBN2, 4x/EBN4 and 6x/EBN4; Hawkes, 1990). The further history of the group has been principally determined by two migrations across the landbridge between North and South America. A first migration southward from the Mexican/Central American area introduced the diploid tuber-bearing species to the variety of niches available in the South American continent, especially those in the mountain range of the Andes. This provoked a rapid speciation, producing the numerous species now occurring in Ecuador, Peru, Bolivia and Argentina. This speciation was accompanied by an increase in EBN and chromosome doubling. Morphologically, the corolla shape developed from stellate to rotate. A northward migration led to the establishment of polyploid species of the series Conicibaccata in Central America and, finally, to the derived polyploids nowadays found in Mexico, which include the well-known hexaploid species *Solanum demissum*. The cultivated forms originated in the area around lake Titicaca, on the border of Peru and Bolivia, where several members of the brevicaule complex still occur.

**Remaining taxonomic problems**

Much of what is known about the taxonomy of potato is due to the work of two formidable taxonomists, Jack Hawkes and Carlos Ochoa. They described numerous species, classified them in series and provided keys based on morphological characters. However, these keys are generally difficult to apply because of the extensive variability in characters such as leaf dissection, pubescence and corolla colour. Many described species are extremely similar to each other, and the group of tuber-bearing *Solanum* spp. seems to be somewhat over-classified, with the application of a rather narrow typological species concept, where all deviations from the 'typical' habit are considered to be due to hybridization. Although hybridization within EBN groups is certainly taking place, another approach would be the recognition of a smaller number of broader circumscribed species, applying a polythetic species concept that allows overlap of character states among species. In certain groups, there is a lack of distinctive characters and species boundaries are difficult to trace. Especially, the interaction between wild and cultivated forms and the influence of human selection have obscured species boundaries, and in some cases, described species might be weedy relatives of cultivated plants or escapes from cultivation. Also, the series classification is problematic, with some series difficult to distinguish from each other and others containing subgroups that could be distinguished as separate series. Consistent with work in other groups within the genus *Solanum* (Knapp, 1991, 2000), it would be advisable to apply the informal concept of 'species groups' instead of the formal taxon 'series'. The advance of molecular methods has offered the hope to arrive at solutions of the aforementioned problems and improve our understanding of the taxonomy of the potato.

## Molecular data

**Molecular markers applied to tuber-bearing *Solanum* spp.**

The available morphological data on potato species have been supplemented with data from cytology, serology, isozymes and several types of DNA data. The cytological data have helped in acquiring more insight into the origin and distribution of polyploids in the group (Swaminathan & Howard, 1953), the serological data gave indications of interrelationships among groups of species but were difficult to interpret (Lester, 1965) and the isozyme data provided valuable information mainly on the diversity of the cultivated forms (Quiros & McHale, 1985; Douches & Quiros, 1988). DNA data are basically in one of two types: restriction site and primer-based data, like restriction fragment length polymorphism (RFLP), RAPD, amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR); giving genome-wide, multilocus information and sequence data providing detailed, single locus information on only a small part of the genome. There is a strong relationship between the level of variability of a molecular marker and its suitability at a given taxonomic level.

**Methods of analysis of molecular data sets – phenetic versus cladistic approaches**

The preferred method to visualize taxonomic interrelationship is to construct bifurcating trees (although scatter plots from ordination techniques have also been found useful). It is important to distinguish the two fundamentally different approaches to tree building, i.e. phenetic versus cladistic approaches, which are distance-based versus character-based, respectively. The distance-based approach calculates the pairwise distances between all combinations of the investigated entities [often called operational taxonomic units (OTUs)], resulting in a triangular distance (or similarity) matrix. Starting from this matrix, different algorithms are used to produce distance-based trees or dendrograms. Consecutive clustering of OTUs with the smallest distances results in an Unweighted Pair Group Method using arithmic averages (UPGMA) tree, whereas for Neighbour Joining trees at each clustering step the effect on the total tree length is taken into account. Both approaches make use of overall similarity based on all the characters simultaneously. Character-based approaches try to construct a tree topology where the character states of each character can be placed in a consistent way (e.g. such that a character state changes into another state just once on the tree). The branches in the tree are considered to be natural groups, called clades (hence cladistic approach). Usually, an analysis generates many possible character-based trees (and just one or very few distance-based trees), making it necessary to adhere to an optimality criterion to choose the 'best' tree. In the character-based approach, this is often the parsimony criterion where the shortest tree (with the minimum number of steps between character states) is considered best.

**Application of molecular data to the taxonomy of the tuber-bearing *Solanum* spp.**

*Delimitation of the group*

The genus *Solanum* has been subdivided into seven subgenera. The group of wild relatives of the potato is classified within the subgenus *Potatoe* in section *Petota*. Hawkes (1989) subdivided this section into two subsections, *Potatoe* and *Estolonifera*, accommodating two non-tuber-bearing series (*Etuberosa* and *Juglandifolia*) in the latter. Using chloroplast DNA RFLPs, Spooner et al. (1993) showed that these two non-tuber-bearing series were in fact less closely related to the tuber-bearing series than to the tomato and should be excluded from section *Petota*. This article also presented conclusive evidence for the inclusion of the genus *Lycopersicon* in the genus *Solanum*, as a section closely related to, but separate from, the potatoes. The nomenclatural consequences of this were published by Child (1990). Kardolus (1998) showed a scatterplot of the first two multidimensional scaling axes calculated from an AFLP data set, where most of the investigated tuber-bearing species form a dense cluster with only a few species outside this core group. Besides three tomato species and members of series *Etuberosa*, also representatives of the Mexican diploid species, series *Circaeifolia* and two accessions of *Solanum mochicense* were plotted away from the dense central cluster. This would indicate that the Mexican diploid species and the South American species *Solanum circaeifolium* and *S. mochicense* are relatively distantly related to the South American species and the Mexican polyploids.

*Overall phylogeny of the group based on molecular markers*

Many studies have focused on the elucidation of the structure within the group of tuberbearing *Solanum* spp., using several molecular markers. Most of these have used RFLPs of the chloroplast (Hosaka *et al.*, 1984; Spooner *et al.*, 1991a; Spooner & Sytsma, 1992; Spooner & Castillo, 1997) or the nuclear genome (Debener *et al.*, 1990). Later, the so-called AFLP reactions were applied (Kardolus, 1998; Kardolus *et al.*, 1998), and recently, SSR (Bryan *et al.*, 1999) and sequence data (Volkov *et al.*, 2001, 2003) became available.

Hosaka et al. (1984) used cpDNA digest patterns to study the interrelationships of 26 species of section *Petota*, supplemented with four outgroup species of the genus *Lycopersicon* and the series *Juglandifolia* and *Etuberosa*. They found clear differences between the outgroup species and the wild potato species, but within section *Petota* were only able to distinguish the Mexican diploid species from the rest (comprising the Mexican polyploids and the South American species). Debener et al. (1990) used nuclear RFLPs, studying 14 wild and 2 cultivated potato species, with *Solanum etuberosum* as an outgroup. They could distinguish clearly separated groups, with all the species of series *Tuberosa* in two related groups, one with wild representatives and one with the cultivated potato, *S. tuberosum*, clustering with *S. stenotomum* and *Solanum canasense*. *Solanum acaule* and *S. demissum* together formed a well-separated branch, and *S. etuberosum* was most distant. Spooner and collaborators (Spooner *et al.*, 1991a; Spooner & Sytsma, 1992; Spooner & Castillo, 1997) used probes rather than directly observed cpDNA digest patterns and greatly extended the number of species studied. They provided evidence for four clades: (1) The Mexican diploids, but excluding *Solanum bulbocastanum*, *Solanum cardiophyllum* and *Solanum verrucosum*. (2) *S. bulbocastanum* and *S. cardiophyllum*. (3) Members of series *Piurana*, with a number of species from other series. (4) *Solanum verrucosum*, all remaining South American species and the polyploid species from Mexico and Central America.

Kardolus et al. (1998) and Kardolus (1998) applied AFLP. In the latter study, 53 species were investigated. The method proved to be highly efficient in producing 997 markers with three primer combinations. Three tomato species and two species from series *Etuberosa* constituted the outgroups. Representatives of the Mexican diploids were placed as the sistergroup of the rest of the tuber-bearing species. The species of series *Tuberosa* were subdivided into geographical groups, with *S. tuberosum* in the Peruvian group associated with species such as *S. canasense*, *Solanum bukasovii* and *Solanum multidissectum* and other members of the brevicaule complex from Bolivia and Argentina grouping together. *Solanum demissum* was united with species of series *Acaulia*, recalling the results of Debener et al. (1990). Series *Circaeifolia* was placed as the most primitive group of the South American species. Bryan et al. (1999) used polymorphic SSRs from the chloroplast genome (cpSSRs), studying 24 species and 30 cultivars. This marker system detected high levels of interspecific cpDNA variation, and the authors suggest its utility in population genetics, germplasm management and phylogenetic studies. The resulting UPGMA tree, however, does not provide much resolution, with cultivated accessions clustering among the wild species (indicating the introgression from wild species into cultivated material) and a tree topology that does not enable the recognition of clear subgroups.

Volkov et al. (2001) used nucleotide sequences of 5S ribosomal DNA genes of 26 wild species, 4 *S. tuberosum*-breeding lines and a tomato accession, with *Solanum dulcamara* as outgroup. This first sequence data set proved difficult to analyse because of the high abundance of indels in comparison with base substitutions, and the dendrograms resulting from different clustering algorithms differed essentially from each other. Because the dendrogram topology was extremely unstable, the authors evaluated the indels 'manually', producing a schematic representation of the molecular evolution. This shows the Mexican diploids (series *Polyadenia, Pinnatisecta* and *Bulbocastana*) as basal and a group with rather conserved 5S rDNA organization comprising *Solanum brevidens, Solanum commersonii, S. circaeifolium* and – surprisingly – *S. bukasovii*, which is far removed from the remaining cluster of species belonging to superseries *Rotata*. Within the latter group, the species of series *Tuberosa* are divided into several subgroups, and *S. acaule* and *S. demissum* are grouped together. Although the overall picture conforms with earlier results, the sequenced region does not seem to be optimal for phylogenetic reconstruction. Volkov et al. (2003) turned to the 5_ external transcribed spacer (ETS) region of rDNA, comparing 30 species of *Solanum* sect. *Petota*, with *S. dulcamara* as outgroup. Three structural variants of ETS (variants A–C) could be recognized . Variant A is present in the outgroup *S. dulcamara*, in the non-tuber-bearing species of series *Etuberosa* and in the representatives of the Mexican diploid series *Bulbocastana, Pinnatisecta* and *Polyadenia*. Species of the series *Commersoniana* and *Circaeifolia* possess variant B, and variant C is present in all other investigated species. Variant C can be subdivided into two subgroups, C1 and C2. Group C1 contains species from the series *Megistacroloba, Conicibaccata* and *Acaulia*, whereas group C2 consists of all diploids of series *Tuberosa*. The dendrograms presented show many polytomies (multiple branching instead of dichotomously branching), indicating that resolution within the groups is mostly lacking. Also, representatives of a species like *S. demissum* are widely separated in all trees, indicative of intraspecific variation. According to the authors, the groups are defined by large rearrangements, while base substitutions allow additional discrimination of closely related species, and this broad range of resolving power is taken to suggest the utility of this marker system for phylogeny reconstruction. The authors further suggest – in contrast with the evolutionary scenario in Hawkes (1990) – an origin of primitive *Petota* spp. in South America, followed by a migration of primitive *Stellata* spp. to Mexico, and a development in South America from other primitive *Stellata* towards more advanced *Stellata* and *Rotata* spp. Summarizing the data from the various studies mentioned above, it seems clear that our insight into the phylogenetic structure of the group of tuber-bearing *Solanum* spp. has been improved by molecular studies. The phylogenetic position of certain species, like e.g. the Mexican diploids, and the series *Circaeifolia*, and the reality of a *S. acaule/S. demissum* assemblage, are supported by several sources. However, the lack of resolution within section *Petota* (4 clades instead of 20 series based on chloroplast RFLP data) seems to be a real phenomenon. Except for the rather distinctive groups in Mexico, the differentiation among the other South American groups is not large, and it remains difficult to subdivide the group into natural units.

The studies discussed above have one serious problem in common: most of them were not able to sample the complete width of the variation of the group of tuberbearing *Solanum* spp., and undersampling can influence the results of (especially cladistic) analyses. The most complete effort has been the studies by Spooner and collaborators.

If one combines these three studies, 86 species from most of the series are considered, but this still is only less than half of the total number of species. The most promising ways forward could be the extension of a molecular data set to encompass all the relevant taxa of the group (Jacobs *et al.*, 2005) and the search for suitable nuclear sequences as undertaken by Spooner and collaborators [nitrate reductase (NIA), Rodriguez and Spooner, 2004; single-copy waxy gene (GBSSI), Spooner et al., 2004; internal transcribed spacer (ITS), Stephenson et al., 2004]

*Detailed studies of parts of the group*

Molecular data have been utilized to study certain groups of species in detail. For convenience sake, these studies will be discussed according to the series names applied, even though not all series received support in the phylogenetic studies mentioned in the previous paragraph.

Series *Acaulia*

Series *Acaulia* has attracted many workers, due to the extreme frost tolerance present in the species *S. acaule*. The pentaploid cultivated potato *S. curtilobum*, which is used to produce the freeze-dried 'chuno', resulted from crosses between Andigenum-type potatoes and the triploid cultigen *S. juzepczukii*, itself a cross between *S. acaule* and diploid Stenotomum potatoes. The taxa within the series comprise tetraploids and hexaploids, which have been recognized at different taxonomic levels by different authors. Hosaka and Spooner (1992), using RFLPs of genomic DNA, studied 105 accessions of *S. acaule* including all four subspecies (*acaule, aemulans, albicans* and *punae*) that were recognized at that time. The results placed subspecies *albicans* as most distant (this hexaploid taxon was later raised to the species level), could not distinguish subspecies *acaule* and *punae* and divided subspecies *aemulans* into two groups, from the provinces La Rioja and Jujuy (Argentina), respectively. Kardolus (1998), studying this group with AFLP reactions, also could not consistently distinguish the subspecies *acaule* and *punae*, but recognized a new, hexaploid subspecies, subspecies *palmirense*. The occurrence of this hexaploid cytotype within the species *S. acaule* may indicate the need to re-evaluate the recognition of *Solanum albicans* on the species level. McGregor et al. (2002) investigated 314 accessions of the Centre for Genetic Resources, The Netherlands (CGN) germplasm collection with AFLP reactions and concluded that most plants were grouped in an UPGMA tree according to the species and subspecies designations in the passport data. The subspecies *acaule* and *punae* were distinguishable, although only separated by a small genetic distance. The classification of the hexaploid *palmirense* taxon in *S. acaule*, separate from *S. albicans*, was confirmed. Nakagawa and Hosaka (2002) combined RFLP data from chloroplast and nuclear DNA to study the relationships between *S. acaule, S. albicans* and 27 morphologically closely related species. They found high similarity between *S. acaule, S. albicans* and *S. demissum*, and suggested *Solanum megistacrolobum* and *Solanum sanctae-rosae* as the closest relatives, and possibly involved in the origin of the series *Acaulia* spp.

Seven cultigenic species *S. stenotomum, S. ajanhuiri, S. chaucha, S. phureja, S. curtilobum, S. juzepczukii* and *S. tuberosum* (with two subspecies: *andigena* and *tuberosum*) are currently recognized as cultivated species (Hawkes, 1990). All tetraploid South-American landraces are classified in *S. tuberosum* ssp. *andigena*. All modern cultivars known to us as the common potato can be accommodated in *S. tuberosum* ssp. *tuberosum*. The transition from subspecies *andigena* to subspecies *tuberosum* apparently resulted from transporting material from the short-day environment of the Peruvian/Bolivian Andes to long-day circumstances. This transport accompanied by adaptation is believed by Hawkes (1990) to have occurred twice: the first event would have taken place in Chile where original subspecies *andigena* material, brought here by migrating Indian tribes from the Andes, underwent adaptation to long-day length and cool climatic conditions, and the second time, this development took place was in Europe after the Spaniards introduced the potato there. Hawkes (1990) regarded the cultigen *S. stenotomum* as being the most primitive of the cultivated material and as the progenitor of the other cultivated 'taxa'. A wild diploid species like *Solanum leptophyes* would have been the progenitor of *S. stenotomum*. Solanum tuberosum ssp. *andigena* originated from a hybridization event between *S. stenotomum* and the wild species *Solanum sparsipilum. Solanum tuberosum* ssp. *tuberosum* later developed from subsp. *andigena*. Grun (1990) described a similar origin of the cultivated potato, with the primitive diploid cultigen *S. stenotomum* arising from a wild progenitor from the brevicaule complex.

The most extensive study using molecular data on the origin of *S. tuberosum*, the relationships among the cultivated species and the relationships between wild and cultivated species, has been conducted by Hosaka and co-workers. In a series of publications ranging from 1986 to 2004, they focused on restriction data of cpDNA of wild and cultivated potatoes. Hosaka (1986) distinguished seven different chloroplast haplotypes in a selection of wild and cultivated species:
(1) type T was restricted to *S. tuberosum* ssp. *tuberosum*;
(2) type A was characteristic for *S. tuberosum* ssp. *andigena* and *Solanum maglia*;
(3) type S was found in *Solanum goniocalyx, S. phureja, S. stenotomum, S. chaucha* and one accession of subspecies *andigena*;s
(4) type C was found in *S. acaule, S. bukasovii, S. canasense, S. multidissectum* and *S. juzepczukii*;
(5) type W was found in wild species and was considered as the most primitive type;
(6) type W' was found in *S. chacoense* f. *gibberulosum*;
(7) type W" was found in *Solanum tarijense*.
The author concluded that, indeed, the cultivated potatoes derived from *S. stenotomum*, which itself might have developed from *S. canasense*, and, furthermore, that the chloroplast genome of the European potato derived from Chilean material, which itself was the result of the combination of the nuclear genome of subspecies *andigena* with cytoplasm from an unknown species.

In 1988, a series of three articles on cpDNA data of potato were published by Hosaka and co-workers. Hosaka et al. (1988) showed that the differences between types T and W found with five different restriction enzymes in the earlier study (Hosaka, 1986) were in fact all caused by one physical deletion in the chloroplast genome of the T-type chloroplast.

The authors concluded that the T-type chloroplast could easily have evolved from the primitive W-type, whereas in the former publication this had not seemed probable. Hosaka and Hanneman (1988a) found a geographical cline from the Andean region to coastal Chile, supporting the Andean origin of Chilean subspecies *tuberosum*. Material considered as a relic of the first European potato (a hybrid of the cultivar 'Myatt's Ashleaf') showed the A-type chloroplast, confirming Hawkes' opinion that the first European potatoes were subspecies *andigena*, later replaced by subspecies *tuberosum* from Chile.

Hosaka and Hanneman (1988b) noted extensive cpDNA variation in cultivated potatoes as well as in wild potato species. They hypothesized that the Andean cultivated tetraploid potato, subspecies *andigena*, could have arisen many times from the cultivated diploids. Hosaka (1995) determined the chloroplast types of 35 accessions of *S. stenotomum* and 97 accessions of putative ancestral wild species, including *S. brevicaule, S. bukasovii, Solanum candolleanum, S. canasense, S. leptophyes* and *S. multidissectum*. Except for *S. brevicaule*, which had only the W type, the wild species proved polymorphic for cpDNA types. Sexual polyploidization formed a wide cpDNA diversity among the Andean tetraploid potatoes and selection caused the limited diversity found in Chilean tetraploid potatoes.

Hosaka (2002) explored the maternal ancestry of the common potato by determining the presence/ absence of a 241-bp deletion characteristic for the T-type cpDNA. Sixteen of 80 accessions of *S. tarijense, S. berthaultii* and *S. neorossii* showed the same deletion at the same position. Hosaka (2003) found that all the T-type accessions of cultivated potatoes shared this haplotype only with some accessions of *S. tarijense*. The author concluded that some populations of *S. tarijense* acted as the maternal ancestor of potato. Hosaka (2004), investigating 215 accessions of *S. stenotomum* and 286 accessions of *S. tuberosum* subsp. *andigena*, noted the absence of T-type chloroplast in *S. stenotomum* while this type was present in nine accessions of subsp. *andigena* and concluded that *S. stenotomum* did not play a role in the formation of the tetraploid potatoes. All the data presented above are based on cpDNA and therefore only show maternal inheritance. Furthermore, the cpDNA types do not seem to be monomorphic within species, which makes it difficult to discuss ancestor/ derivative relationships between species. There is a need for suitable nuclear markers (such as AFLP or a suitable nuclear sequence) to complement the work done on the chloroplast genome. Many studies have taken this approach.

Debener et al. (1991) showed with nuclear RFLPs that *S. andigena, S. stenotomum* and *S. canasense* were very closely related to each other and could in fact not be distinguished with the single locus information. Miller and Spooner (1999) used single to low-copy nuclear RFLPs and RAPDs to investigate the species boundaries and relationships among the members of the brevicaule complex. They confirmed the separation of populations from Peru and immediately adjacent northwestern Bolivia, including most cultivated accessions, and of populations from northwestern Bolivia and Argentina. This had been found by Van den Berg et al. (1998) using morphological data and Kardolus et al. (1998) using AFLP, which also showed *S. tuberosum* clustering together with the wild Peruvian species *S. canasense* and *S. multidissectum*. Miller and Spooner (1999) indicated the paraphyletic nature of the brevicaule complex and the need to reduce the number of species names in this group.

Raker and Spooner (2002) tested the genetic differences between accessions of *S. tuberosum* ssp. *andigena* and *S. tuberosum* ssp. *tuberosum* using nuclear DNA microsatellites. The two subspecies could be separated from each other although the separation is not very firm. Other cultivated species (*S. stenotomum* and *S. phureja*) and wild species (*S. bukasovii, S. multidissectum* and *S. canasense*) used in this study were mixed with *S. tuberosum* ssp. *andigena*.

Sukhotu et al. (2004) combined data on the cpDNA types of Hosaka with chloroplast microsatellite markers and nuclear RFLPs. The differences among cpDNA types were highly correlated with the microsatellite markers. The nuclear RFLPs supported the differentiation between the W type versus the C, S and A types, but not the differentiation among the three latter types, suggesting frequent genetic exchange among them. In a UPGMA dendrogram of the nuclear DNA restriction data, three clusters could be identified, with both *S. tuberosum* ssp. *andigena* and *S. tuberosum* ssp. *tuberosum* accessions placed together with most other cultivated *Andean* spp. and members of the brevicaule complex.

In a recent study of the brevicaule complex, Spooner et al. (2005a), using AFLP data, reconfirmed the distinction of the northern and southern subgroups within the complex and argued that cultivated potatoes have had a monophyletic origin in the northern part of the distribution area of the brevicaule complex, as all the landrace populations form a clade in the parsimony cladogram. The progenitor of the cultivated potato should thus be sought in the members of the brevicaule complex occurring in southern Peru. The authors note that these species are poorly defined and may have to be reduced to a single species, the earliest valid name being *S. bukasovii*. The brevicaule complex itself is designated to be polyphyletic.

The origin of our modern cultivated potato varieties and the manner of introduction of the cultivated potato in Europe have been the subject of controversy. According to many authors, the first potato material to be introduced in Europe belonged to *S. tuberosum* ssp. *andigena*. Most of the potato stock derived from this original material was believed to have been wiped out during the late-blight outbreak in Europe in the 1840s. After this, the breeding stock would have been replaced with introductions from Chile of *S. tuberosum* ssp. *tuberosum* material (Grun, 1990). Juzepczuk and Bukasov (1929) had, however, suggested that the early European introductions already consisted of subspecies *tuberosum* germplasm from Chile, because of the similarity in morphology and growing conditions. Spooner et al. (2005b) published results from a nuclear microsatellite analysis of mainly Indian potato cultivars. The analysis included several accessions that were considered to be derived from *S. tuberosum* ssp. *andigena* to test the idea that the first potato introductions in the old world were actually subspecies *andigena*. Late blight was not recorded in India until 1870, so only after the late blight disaster in Europe. The *andigena* germplasm in India would therefore not have been eliminated by the epidemic. The microsatellite results showed, however, that all Indian cultivars, including those that were thought to be derived from subspecies *andigena*, clustered together with the subspecies *tuberosum* landraces and European cultivars. All 12 tested subspecies *andigena* landraces from Central and South America clustered together separately from this group. The andigena introduction theory was thus not supported.

Spooner et al. (2005b) concluded that no remnant landraces of subspecies *andigena* were involved in the development of the Indian germplasm. Considering this evidence and other historical and cytological information, they suggested that the early introductions of cultivated potatoes of India (and Europe) came from both the Chile and the Andes. The Chilean landraces became the predominant breeding germplasm before the outbreak of late blight, likely because of their pre-adaptation to long-day/cool climate conditions. Remarkably, five Indian cultivars that, based on the nuclear microsatellite data were linked to subspecies *tuberosum*, lacked the typical 241 bp deletion. This could have been caused by either a subspecies *tuberosum* progenitor lacking the typical deletion or the incorporation of other non-tuberosum accessions as maternal material.

Huaman and Spooner (2002) examined morphological support for the classification of landrace populations of cultivated potatoes. They recognized all landrace populations as a single species, *S. tuberosum*, with eight cultivar groups. Following the philosophy of cultivar-group classification, the remaining cultivated materials, e.g. the modern varieties, were not automatically classified as a ninth 'Tuberosum' group. Many authors have suggested that molecular markers are appropriate to identify cultivars and reveal infraspecific variation (e.g. Debener *et al.*, 1990; Hosaka *et al.*, 1994; Bryan *et al.*, 1999; Bornet *et al.*, 2002), but these methods have not been used to produce an overall classification of cultivars. Most studies are restricted to the assessment of genetic diversity of cultivars or their discrimination with fingerprinting techniques (Görg *et al.*, 1992).

Provan et al. (1999) used polymorphic chloroplast and nuclear SSRs to study the diversity in most modern potato cultivars grown in the UK. In total, 151 of 178 accessions tested showed the same chloroplast haplotype, named haplotype A, which corresponds with the T type of Hosaka (1986). A much higher diversity was found in the remaining accessions outside the T-type group, which were assigned to 25 different haplotypes. The diversity of the nuclear SSR loci did not show this difference between the T-type group and the rest. The authors suggested that the dominance of the T-type cytoplasm was caused by the use of only a limited number of maternal lineages in breeding programmes. Bryan et al. (1999) using cpSSRs demonstrated that among a set of 30 tetraploid potato cultivars, a single chloroplast haplotype was prevalent and they attributed this to the widespread use as a female parent of the imported US cultivar 'Purple Chili' in the latter half of the 19th century. The chloroplast diversity that is present has arisen through introgression from wild and primitive cultivated material. The low level of genetic diversity of European cultivated potatoes was confirmed in an analysis using ISSRs by Bornet et al. (2002). Their results showed that European potatoes are quite homogenous, and the genetic diversity was very low compared with Argentinian cultivars.

Molecular data have been used to address three main issues about the cultivated potato:
(1) the mode of origin of the crop and the relationships with its wild relatives;
(2) the relationship between the andigena and tuberosum groups and the introduction of the cultivated potato from South America to Europe and the rest of the world;
(3) the genetic diversity of the crop.

The conclusions about these issues are not unequivocal. Results from the chloroplast and nuclear genome conflict as to the role taxa like *S. tarijense*, *S. stenotomum* and the brevicaule complex have played in the origin of the crop. Different data sets give rise to different hypotheses on the multiple or single domestication event(s) that occurred, most probably, in southern Peru. The role that Chilean material played in the introduction of the cultivated potato in Europe has been clarified. The genetic diversity of the crop has been shown to have suffered a severe maternal bottleneck during the development of the modern cultivated potato. Finally, the classification of the modern cultivars of potato in subgroups has not really been addressed yet with molecular markers.

## Conclusions

Molecular data have been used to establish the phylogeny of the group of tuber-bearing *Solanum* spp., to evaluate hybridization hypotheses, to evaluate infraspecific classifications, to establish the ancestry of the cultivated potato, to trace introgression from wild species and to assess genetic diversity within species and cultivated material. In the context of genebank management, the effect of seed increases on the diversity of genebank accessions (Del Rio & Bamberg, 2003), and the extent of redundancy (McGregor *et al.*, 2002) has been studied. Furthermore, molecular data allow checking for misidentifications and can be utilized in risk-assessment studies. Although the search for the phylogenetic structure of the group has suffered from a lack of resolution, at the species level, the utility of AFLP is evident, as long as closely related taxa are compared. There remains a need for a suitable nuclear marker to fill the gap between the high level chloroplast derived data and the fingerprinting data like SSRs, but this will most probably be forthcoming in the near future.

# CHAPTER 3

## Comparison of Chloroplast DNA and AFLP data from *Solanum* section *Petota* reveals incongruencies

Mirjam M.J. Jacobs, Ronald R.G. van den Berg, Marc S.M. Sosef and Ben Vosman

## Abstract

Chloroplast (cp) DNA sequence data and nuclear AFLP data were used for phylogeny reconstruction in *Solanum* section *Petota*. A comprehensive set of accessions (199 accessions from 174 taxa), covering the section as widely as possible, was chosen from gene banks worldwide. The chloroplast regions *trn*TLF (1907 nucleotides) and *psb*A/*trn*H (464 nucleotides) were sequenced. AFLP data were obtained for two primer combinations. The AFLP tree showed much more phylogenetic resolution than the tree based on Chloroplast DNA. Neither the chloroplast results nor the AFLP results provide support for maintaining the classification of section *Petota* in 21 series, as proposed by Hawkes. The majority of the series proposed by Hawkes could not be identified as separate clades. Comparison of the cladograms obtained from the cpDNA and AFLP data, showed several incongruencies. These differences are most likely related to the mode of inheritance of the different genomes targeted, in combination with extensive hybridization between species. The low resolution found in large sections of the trees suggests that many species within the section *Petota* have not diverged substantially.

## Introduction

The tuber-bearing *Solanum* species, including the cultivated potato and its wild relatives, are accommodated in *Solanum* section *Petota*. Based on morphological characteristics, crossability and cytology, Hawkes (1990) divided the species of section *Petota* in 21 series, 19 of which contain tuber-bearing species plus two series (*Etuberosa* and *Juglandifolia*) containing related non-tuber-bearing species. The series were put in an order that reflected Hawkes ideas on their evolutionary relationships. Since Hawkes (1990), besides further morphological studies, many molecular studies have been carried out to establish the taxonomic structure within section *Petota*, focusing on the nuclear and chloroplast genome (for a review see Van den Berg & Jacobs, 2007).

Both sequences and restriction fragments of Chloroplast DNA (cpDNA) have been used to solve taxonomic problems at different taxonomic levels (Olmstead & Palmer, 1994), using coding regions like *rbcL* for revealing family level taxonomy and non-coding regions for lower taxonomic levels. Mutation rates in cpDNA are low, which makes cpDNA valuable for inferring relationships at the interspecies level and above (Palmer, 1987). Mutation rates in non-coding chloroplast sequences are higher than coding cpDNA regions (Gielly & Taberlet, 1994). For section *Petota*, cpDNA restriction fragments length polymorphisms (RFLPs) have been used but most studies have treated only a small part of the section, except for Spooner and Castillo (1997), Spooner and Sytsma (1992), and Spooner et al. (1991). These three studies together treated 90 accessions from 86 species representing 17 series. The results from these cpDNA RFLPs did not yield support for a classification of potato species into 19 series. Only four clades of different size were found (Spooner and Castillo, 1997).

For closely related taxa, especially at the species level, the AFLP method has the potential to solve phylogenies when other markers fail due to lack of genetic variability (Despres *et al.*, 2003).The AFLP method generates a large number of polymorphisms (Wolfe & Liston, 1998).

Several authors report that the AFLP method could generate non-homologous fragments, especially at higher taxonomic levels (Despres *et al.*, 2003; Koopman & Gort, 2004). Still, in spite of possible homoplasy, AFLP datasets generally contain sufficient phylogenetic signal (Koopman, 2005). Again, only a few AFLP studies covered a substantial part of section *Petota*. Kardolus (1998) studied the AFLP patterns of 53 taxa, representing 17 series in total. He also found that the phylogenetic relations did not reflect the series classification by Hawkes (1990). Lara-Cabrera and Spooner (2004) used AFLP data to infer the phylogeny of the North and Central American diploid potato species. Their results support sister taxon relationships for a number of species: *S. cardiophyllum* subsp. *ehrenbergii* and *S. stenophyllidium*; *S. tarnii* and *S. trifidum*; *S. jamesii* and *S. pinnatisectum*; *S. lesteri* and *S. polyadenium* and *S. clarum* and *S. morelliforme*.

The purpose of the present study is to gain insight in the internal taxonomic structure of section *Petota* by using both cpDNA sequences and AFLP patterns for the same individuals, which will allow a direct comparison of the results of both methods. We have analyzed as many species and as many series from section *Petota* as were available. The only two series not represented here are series *Ingifolia* and series *Olmosiana*. The results are used to evaluate the classification of Hawkes (1990) and the four clade hypothesis of Spooner and Castillo (1997).

## Materials and methods

### Plant material

The material used was chosen to sample the *Solanum* section *Petota* material available in the genebanks worldwide as widely as possible. In total, 199 accessions from 174 taxa were used. Table 1 lists the accessions used and gives species names according to the passport information from the genebanks and notes on synonymy based on several recent publications. Seeds were surface-sterilized and sown *in vitro* at 25°C. The collection of individual *Solanum* clones was grown *in vitro* for at least 6 weeks on MS medium supplemented with 20% sucrose (Murashige & Skoog, 1962) at 18°C. DNA was extracted from leafs according to the method described by Stewart and Via (1993).

### Nomenclature

The labels used in this paper are taken from the original passport data belonging to the genebank accessions. They are not corrected according to the synonymy in recent taxonomic revisions because we do not want to change an original label of an accession without actually checking the identity of that accession. Moreover, by displaying the original genebank species names, the discrepancies of our results with earlier taxonomic treatments become apparent. We have included information on revised taxonomy from Hawkes (1990), Ochoa (1990), Ochoa (1999), Spooner and Hijmans (2001), Spooner et al. (2004), Spooner and Salas (2006), Van den Berg and Spooner (1992) and Huaman and Spooner (2002) in Table 1.

**Chloroplast DNA sequencing**

To amplify the chloroplast *trn*TLF region, primers a (CATTACAAATGCGATGCTCT) / d (GGGGATAGAGGGACTTGAAC) and c (CGAAATCGGTAGACGCTACG) / f (ATTTGAACTGGTGACACGAG) described by Taberlet et al. (1991), were used. Because some of the samples did not amplify very well with the primer combination a/d, an extra primer, which is here described as primer b4 (CGGATTCGGGTCGTCAT), was used for some of the samples. Amplification of the intergenic spacer between *psb*A and *trn*H was done by using primers *psb*AH (CGAAGCTCCATCTACAAATGG) and *trn*HH (ACTGCCTTGATCCACTTGGC) described by Hamilton (1999). All samples were sequenced in both directions at least once. The expected length for the chloroplast regions *trn*TLF was 1907 nucleotides and that of the *psbA*/*trn*H region was 464 nucleotides. To amplify the samples, approximately 10 ng of genomic DNA was mixed in a total volume of 20 µl containing (endconcentration per reaction) 1x PCR buffer, 2 mM MgCl2, 0.1 mM mixture of all dNTPs, 0.2 pM each primer and 0.4 unit Goldstar ® DNA polymerase (Eurogentec). The following PCR protocol was used: a first step of 3 minutes at 94 °C followed by 30 cycli of 0.5 minutes at 94 ° C, 0.5 minutes at 50 ° C, 2 minutes at 72 ° C, concluding with 10 ° C. PCR products were checked for presence and correct length on a 1.0% agarose gel. All the PCR products were purified by filtration in Sephadex G-50 columns. Around 50 ng of the PCR product was added to a total volume of 10 µl containing 0.5 pM primer, 2 µl AmerDye and 2 µl Amerdye Buffer (Amersham). The PCR programme included 25 cycles of 20 seconds at 94 ° C, 15 seconds at 50° C, and 1 minute at 60° C, ending with 10° C. The sequence products were purified in Sephadex G-50 columns. The samples were run on a ABI 3000 sequence machine. Sequences were aligned with the software program Seqman DNAstar v6.

**AFLP analysis**

AFLP analysis was carried out as described by Vos et al. (1995) using the enzyme combination *Eco*RI/*Mse*I. The two primer combinations (PCs), E32/M49 and E35/M48, used for the analysis were selected based on previous results obtained with potato material.

The gel analysis on a capillary electrophoresis system (MegaBACE™) was performed according to Van Eijk et al. (2004). MegaBACE allows multiloading of two AFLP reactions in parallel, each reaction is labelled with a specific fluorphorescent. Only the *Eco*RI-primers were end-labelled using fluorescent label (FAM and JOE). Pseudo gel images were generated and all AFLP markers were scored dominantly using proprietary software developed specifically for AFLP analysis at Keygene N.V. This software allows the display, and analysis of pseudo gel images. For the analysis of pixel images, the software provides tools to navigate through the image to size and quantify the AFLP bands with great precision. Each band of a specific marker is classified with respect to its intensity using a mixture model of normal distributions, as described by Jansen et al. (2001). A MegaBACE ET900-R size standard from Amersham Biosciences was used in each capillary to estimate the molecular weight of the fragments. AFLP® is a registered trademark and the AFLP technology is covered by patents and patent applications of Keygene N.V.

**Data Analysis**

The data were analyzed as two separate data sets (a cpDNA data set and an AFLP data set).
Both phenetic and cladistic analyses were conducted using PAUP 4.0 Altivec (Swofford, 2002). To calculate the distance matrix in the phenetic analyses we used the NeiLi distance ( Nei & Li, 1979) for the AFLP data and the Jukes-Cantor distance measure for the cpDNA data. Neighbor Joining clustering was carried out on both data sets. In the cladistic analysis heuristic searches were run by using a 2-step-procedure modified from Maddison (1991). In step 1 of this procedure a set of 10.000 starting trees is created using the options TBR, MULPARS, SAVEREPS. One tree for each replicate was saved. The resulting trees of step 1 were then used as starting trees in step 2 of the procedure. In step 2, the options TBR, MULPARS, SWAPALL and Nbest=10.000 were used. Probably more trees with this length could have been found but because of memory capacity the search was restricted to find no more than 10.000 trees. Jackknife analysis was performed to obtain statistical support for the branches of both phenetic and cladistic trees. In the heuristic search the jackknife analysis was performed using 1000 replicates, TBR swapping, multrees=yes, saving no more than five shortest trees per starting tree.

**Testing significance of congruence between datasets**

As visual comparison suggested incongruence between the AFLP and the cpDNA trees, the significance of the incongruence in the phylogenetic trees was tested by using the Incongruence Length Difference test, implemented in PAUP as the so-called partition homogeneity test (Farris *et al.*, 1995). The Neighbor Joining trees were tested by using the Mantel test as implemented in Ntsys 2.1 and described by Lapointe and Legendre (1992).

## Results

Because of the similar topology of the trees resulting from the phenetic and cladistic analyses we only present the results of the heuristic searches. Figure 1 and 2 show the strict consensus trees based on the 10.000 most parsimonious trees derived from the cpDNA and AFLP data, respectively. The most important results of the Neighbor Joining (NJ) and Maximum Parsimony (MP) Jackknife analyses are summarized in Table 2. The clades shown in the strict consensus tree of the 10.000 most parsimonious trees coincide with the groups found in the jackknife tree of the NJ analysis. The jackknife support for the groups found in the NJ jackknife analysis is also very similar to that of the MP analysis, as shown in Table 2. The results are described using the series names of Hawkes (1990).

**The Chloroplast DNA results**

The *trn*TLF sequences yielded 15 informative indels and 91 snp's, the *psbA/trn*H sequences gave 3 indels and 42 snp's. Indels were (except for a microsatellite region in *psba/trn*H that was excluded because of instability) included in the dataset. They were coded as absent/present (T/A) irrespective of their length. This resulted in a combined chloroplast sequence dataset with 2421 nucleotides and 151 parsimony informative markers and 109 non-informative markers.

The 10.000 most parsimonious trees found all had a length of 282 steps (with CI=0.691 and RI= 0.890). The strict consensus tree of the cpDNA data with jackknife support values is shown in figure 1. In this tree clades are labeled from A to Q.

Clade Q represents the outgroup. The ingroup consists of two successive polytomies. The first polytomy contains the second polytomy plus four branches each representing individual accessions. The second polytomy consists of Clade A to P plus a large number of branches representing individual accessions.

Clade A contains accessions of the Central and North American polyploid species of series *Longipedicellata* and *Demissa*, plus *S. verrucosum* and *S. andreanum* from series *Tuberosa* group i (Mexico, Venezuela, Colombia and Equador). Between Clade A and the clades B to P, the tree continues with a large block of unsupported branches each consisting of individual accessions. These accessions mainly represent diploid and polyploid species of series *Tuberosa* group ii (Peru), *Tuberosa* group iii (Bolivia, Argentina and Chile) and *Tuberosa* group iv (cultivated species), some polyploid species from series *Conicibaccata* and species from the series *Piurana*, *Yungasensa*, *Megistacroloba*, *Cuneoalata*, *Lignicaulia*, *Commersoniana*, *Maglia* and *Acaulia*.

Clades B to G each unite only two to four species. Clade H contains mainly species of series *Tuberosa* from Bolivia, Argentina and Chile. The accessions of species from series *Circaeifolia* are split up in 3 separate clades (J, K, and L). Clade M contains accessions mainly from series *Piurana*, although there are representatives of other series in this clade: *S. chomatophilum*, *S. irosinum* and *S. paucijugum* from series *Conicibaccata*, *S. immite*, *S. augustii*, *S. acroscopicum* from series *Tuberosa* (from Bolivia Argentina and Chile), and *S. sogarandinum* from series *Megistacroloba*. The clades N, O and P contain all the Mexican and Northern American diploids species. Clade N consists of all the accessions of *S. bulbocastanum* but not *S. clarum* that Hawkes (1990) also considered to be a member of series *Bulbocastana*. Clade O contains three accessions of *S. cardiophyllum*. Two other accessions of *S. cardiophyllum* and the *S. cardiophyllum* subsp. *ehrenbergii* accessions are placed in clade P. This clade includes almost all the North/Central American diploid species belonging to the series *Pinnatisecta*, *Polyadenia* and *Morelliformia*, plus *S. clarum* of series *Bulbocastana*. Clade Q contains the outgroup accessions of *S. etuberosum*, *S. palustre* and *S. fernandezianum*.

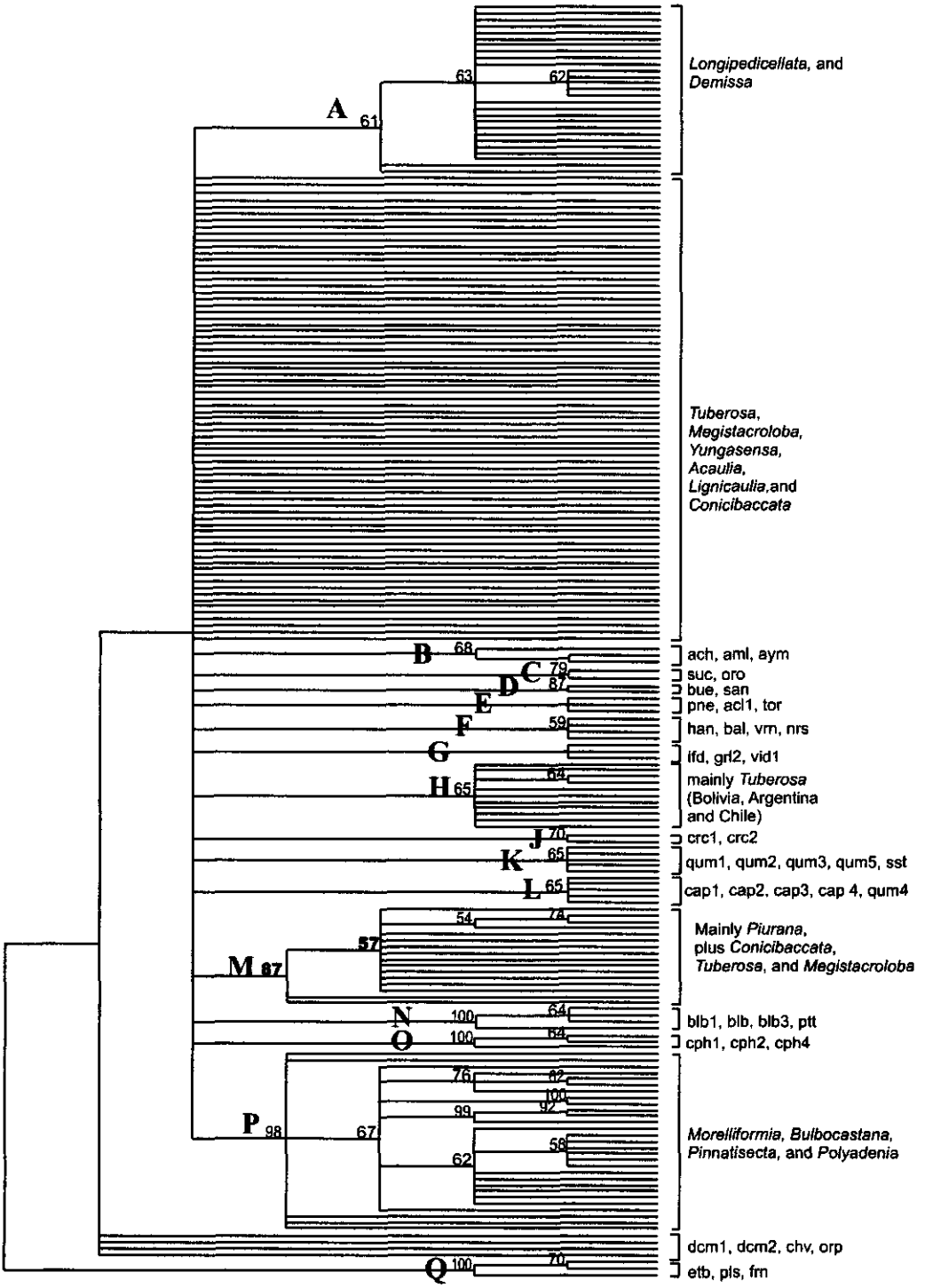



*Figure 1. Strict consensus tree of 10.000 most parsimonious trees based on chloroplast DNA data. Jackknife values >50 are indicated above the branches.*

**The AFLP results**

Only markers which could be scored dominantly in an unambiguous manner were scored (presence/ absence polymorphisms). In total 224 AFLP markers were scored. The 10.000 most parsimonious trees had a length of 3088 steps (with CI=0.073 and RI=0.664). The strict consensus tree of these 10.000 most parsimonious trees (figure 2) shows more structure than the strict consensus of the cpDNA data, but the jackknife support of several of these groups is low.

The clades in figure 2 are coded from I to XIV.

Clade I, VII, VIII, IX, and XI do not show jackknife supports above 50. They mainly contain accessions of species belonging to series *Tuberosa* plus some species belonging to series *Megistacroloba*, *Yungasensa* and *Commersoniana*. Clade II is not supported itself but contains a highly supported clade with only accessions representing taxa from series *Circaeifolia*: *S.circaeifolium*, *S. circaeifolium* subsp. *quimense*, *S. capsicumbaccatum* and *S. soestii*. Clade III is a highly supported clade with the *Conicibaccata* species *S. sucubunense*, *S. orocense*, *S. agrimonifolium*, *S. colombianum*, *S. longiconicum*, *S. flahaultii*, *S. subpanduratum*, *S. otites*, and *S. garcia-barrigae*. Clade IV contains representatives of series *Demissa* : *S. iopetalum*, *S. brachycarpum*, *S. guerreroense* and *S. schenckii*. However, it does not contain *S. demissum* which is placed in clade X. Clade V contains accessions of the Mexican and North American polyploid series *Longipedicellata* : *S. hjertingii*, *S. matehuale*, *S. fendleri* subsp. *arizonicum*, *S. papita*, *S. polytrichon* and *S. leptosepalum*. Clade VI only consists of the species *S. verrucosum* and *S. macropilosum*. Clade X is not supported but contains a highly supported clade of the species *S. acaule* and *S. demissum* and their closest relatives. The jackknife support for their common branch is high (99) but for the branch connecting them with *S. sanctae-rosae*, *S. megistacrolobum*, and *S. megistacrolobum* subsp. *toralapanum* the jackknife value is lower (64). The accessions of *S. demissum*, *S. semidemissum* and *S. edinense* form a strongly supported (jackknife value 87) clade. The clade of *S. acaule* and species related to *S. acaule* does not appear in the strict consensus while in the jackknife tree this clade has also high support (Table 2).

Clade XII consists of the *Piurana* accessions *S. paucissectum*, *S. chomatophilum*, *S. solisii*, *S. plurae*, *S. paucijugum*, *S. tuquerrense* and *S. irosinum*. Clade XIII consists of a group of accessions from Mexican diploid series *Morelliformia*, *Polyadenia*, *Pinnatisecta* and *Bulbocastana*. Within this clade several subclades can be distinguished. Clade XIV contains the outgroup accessions of species *S. etuberosum*, *S. palustre*, and *S. fernandezianum*.

**Statistical tests**

Many differences between the cpDNA tree and the AFLP tree are apparent. Two tests, the ILD test for the cladistic trees and the Mantel test for the phenetic trees, were carried out to evaluate the significance of the observed differences. The Mantel test showed a correllation of r=0.56 (poor fit) between the tree structure of the cpDNA and the tree structure of the AFLP data, with a p value of 0,001. The ILD test resulted in a value of 124 but with a p value of 0.095: the observed ILD is not significantly greater than can be expected from chance. This means that the null hypothesis, the datasets are congruent with each other, cannot be rejected.

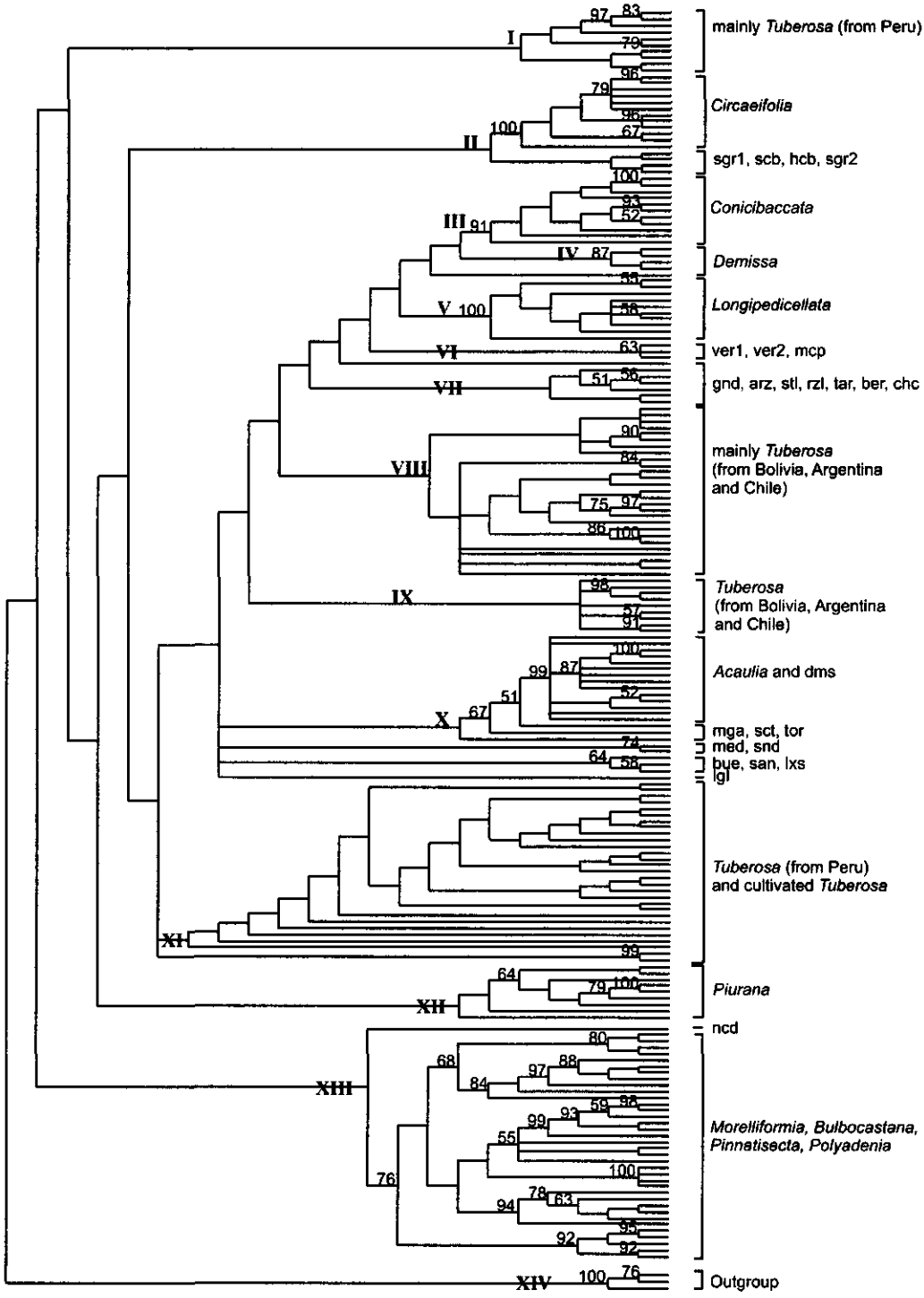



*Figure 2. Strict consensus tree of 10.000 most parsimonious trees based on AFLP data. Jackknife values > 50 are indicated above the branches.*

## Discussion

### Comparing the Chloroplast DNA results to previous studies

The groups found in our Chloroplast DNA results correspond largely with the results from three cpDNA restriction site studies (Spooner *et al.*, 1991; Spooner & Sytsma, 1992; Spooner & Castillo, 1997). They found four clades:

Clade 1, consisting of Mexican diploid species, is similar to our group of Mexican diploid species. Clade 2 consists of *S. cardiophyllum* and *S. bulbocastanum* accessions. In our results *S. cardiophyllum* and *S. bulbocastanum* are also not included in the group of Mexican diploids, but they form two separate groups that are not connected to each other. In the results of Spooner and Castillo (1997) and Rodriguez and Spooner (1997) *S. cardiophyllum* and *S. bulbocastanum* form a clade with a high bootstrap value. The absence of a connection between the *S. cardiophyllum* and the S. *bulbocastanum* in the present study might be due to the different markers used. However, in the present study four accessions of *S. cardiophyllum* subsp. *ehrenbergii* are included in the Mexican diploid group. These results correspond to the results of Rodriguez and Spooner (1997). Clade 3 consists of *Piurana* species including some accessions from other series. This clade contains almost all the species found in the *Piurana* clade in the present Chloroplast DNA analysis. Clade 4 corresponds partly with the species directly attached to the large polytomy in our cpDNA analysis, containing species from series *Tuberosa*, *Megistacroloba* and *Conicibaccata*. A difference between these studies and our results is that, while in the cpDNA RFLP results the polyploid Mexican and Central American species are placed together with South American polyploids and diploids in clade 4 and there seems to be no resolution within clade 4, in our results the polyploid Mexican and Central American species are placed in a separate clade. On the other hand, Spooner and Castillo (1997) found sistergroup relationships between clade 3 and clade 4, and between clade 2 and the combination of clade 3 and 4. Our results did not show resolution on this level of the connections of the different clades.

### Comparing the AFLP results to other nuclear analyses

Bonierbale et al. (1990) used nuclear RFLPs to study 90 *Solanum* accessions representing 18 species of *Solanum* section *Petota*. Their tree based on the calculated genetic distances shows that the diploid species *S. capsicumbaccatum* and *S. bulbocastanum* are most different from all the other *Petota* species. The North/Central polyploid American species *S. stoloniferum* is most similar to *S. verrucosum* and together they cluster with species from series *Tuberosa* group iii. They show that *S. demissum* is more similar to *S. acaule* than to the other polyploid Mexican species *S. stoloniferum*. Furthermore, the cultivated species *S. tuberosum*, *S. phureja*, and *S. stenotomum* are more similar to each other than to the other species. Their results roughly correspond to our present AFLP results. In our results, *S. verrucosum* is linked to the *Longipedicellata* and *Demissa* species. Furthermore, the *S. demissum* accessions in our study are also more related to the *S. acaule* accessions than they are to accessions belonging to *Demissa*. The cultivated species of series *Tuberosa* are mixed among other representatives of series *Tuberosa* in clade XI in the present AFLP results.

Our AFLP results are concordant with results found in the first AFLP analysis by Kardolus et al. (1998) and his more extended AFLP analysis (Kardolus, 1998). There it was also found that *S. demissum* appeared to be closely related to series *Acaulia*. Accessions from this series and *S. demissum* shared many AFLP bands. *S. juzepcukii*, a triploid hybrid between *S. acaule* and *S. stenotomum*, also shared bands with *S. demissum* and accessions from *Acaulia*. Furthermore, Kardolus (1998) found two large clusters of *Tuberosa* accessions, reflecting their geographic origin. One of these clusters consists of species from series *Tuberosa* group iii (from Bolivia, Argentina and Chile) and the other cluster consists of species from *Tuberosa* group ii (from Peru) plus the cultivated potato species from series *Tuberosa*. Our study also shows some separation between groups of *Tuberosa* species from different geographic regions.

The present study shows many similarities with an earlier AFLP study of Lara-Cabrera and Spooner (2004) concerning the inner structure of the North and Central American diploid species group. In the present study, eight clades with jackknife support of 60 or higher were found within the clade of North and Central American diploid species. At least four clades that were present in the earlier study could also be recognised in the present AFLP tree. A difference is the position of the clade with *S. cardiophyllum* accessions. In our study the *S. cardiophyllum* accessions form a strong subgroup (with a jackknife support value of 100) but they are not a sistergroup to all the rest of the diploid North and Central American species together, like in the earlier study.

**Comparing the cpDNA and AFLP trees**

The results of the cpDNA and the AFLP analyses were visually compared. Table 2 lists similarities and incongruencies between them. A few examples of incongruencies will be discussed here. First, the accessions of *S. demissum* and *S. acaule* and their closest relatives are strongly connected in the AFLP tree while they are not linked at all in the cpDNA tree. In the cpDNA tree, *S. demissum* is placed amidst the *Demissa / Longipedicellata* group. Both Spooner et al. (1995) and Nakagawa and Hosaka (2002) hypothesize that *S. demissum* could be derived from *S. acaule* and an unknown female parent. Although Nakagawa and Hosaka (2002) suggest this unknown maternal parent to be a diploid South American species having W type chloroplast, based on the present results it would be more logical to assume that an unknown species from series *Longipedicellata* or *Demissa* has acted as a maternal parent.

Secondly, in the cpDNA strict consensus tree, the different taxa of series *Circaeifolia* do not form one group but end up in three separate groups. Each small group contains representatives of one taxon, except for accession qum4, a *S. circaeifolium* subsp. *quimense* accession that clusters with *S. capsicibaccatum*. The only *S. soestii* accession appears together with *S. capsicibaccatum* in one clade. In contrast to these cpDNA data all the series *Circaeifolia* species come together in one single clade in the AFLP strict consensus tree. The Chloroplast DNA results from our study are surprising because according to earlier studies (Van den Berg & Groendijk-Wilders, 1999; Van den Berg *et al.*, 2001) *S. capsicibaccatum*, *S. circaeifolium* subsp. *quimense*, *S. circaeifolium* and *S. soestii* can be regarded as subspecies from the same species. The Chloroplast DNA results would therefore suggest that different Chloroplast DNA types are found within one species.

This phenomenon could be caused by introgression, lineage sorting or Chloroplast capture (Wendel & Doyle, 1999). In a review on cytoplasmic gene flow in plants, Rieseberg and Soltis (1991) reviewed many of these kind of incongruencies between Chloroplast DNA trees and organismal trees. The genera *Helianthus*, *Heuchera* and *Populus* are mentioned amongst many others as being known for numerous examples. The review shows that five species of *Helianthus* possessed more than one cpDNA genotype, which would be indicative of recent cytoplasmic introgression. This could also be the case in our *Solanum* species.

A last example refers to the boundaries and status of the series *Piurana* and *Conicibaccata*. The cpDNA tree only shows a strongly supported clade with species belonging to series *Piurana* but it lacks a clade with species from series *Conicibaccata*. In the cpDNA strict consensus tree almost all species of the series *Conicibaccata* are attached directly to the polytomy, without any structure, like so many other species from series *Tuberosa* group (iii) and series *Megistacroloba*. The clade of series *Piurana* species in the cpDNA tree contains also many non-*Piurana* species: *S. sogarandinum*, *S. huancabambense*, *S. immite*, *S. augustii*, *S. acroscopicum* and *S. mochiquense*. The AFLP tree shows a strongly supported clade formed by species belonging to series *Conicibaccata*. Additionally, a *Piurana* clade with moderate support can be found but it is smaller and the species that are included differ from those in the *Piurana* clade found with the cpDNA results. The *Piurana* clade in the AFLP results contains four species that Hawkes (1990) classified in series *Piurana* and three *Conicibaccata* species that Castillo and Spooner (1997) recognized as belonging to series *Piurana* (*S. chomatophilum*, *S. paucijugum* and *S. irosinum*). In summary, the boundaries of both series *Piurana* and *Conicibaccata* seem to be blurred and unclear and the Chloroplast DNA results do not reflect the AFLP results nor the species classification based on morphological features. This discrepancy might be explained by assuming that (ongoing) gene flow between species causes confusion. The series *Conicibaccata* might be closer related to series *Longipedicellata* as shown in the strict consensus tree in figure 2.

The results from the statistical test are not very consistent. The low value of the outcome of the Mantel test would suggest that the datasets are significantly different from each other, but the outcome of the ILD test is that the null hypothesis (the datasets are not significantly incongruent) cannot be rejected. These discrepancies between the outcomes of the statistical tests might be caused by the differences in level of resolution between the two datasets.

**Resolving power of cpDNA and AFLP markers used**

The observation that the data presented show a lack of resolution raises questions on the markers used. First, one could argue on the number of AFLP markers that was used to reconstruct the phylogeny. Although the present number of 224 markers seems low, other studies in which many more AFLP markers were used, also point to a lack of structure. Kardolus (1998), using 3 AFLP primer combinations, produced in total 997 markers in 171 genebank accessions of *Solanum* section *Petota* species, and found no more structure than in the present study. A recent study (Spooner et al. 2005) used 438 AFLP markers from 6 AFLP primer combinations to produce a phylogenetic tree of 261 wild (mainly brevicaule complex members) and 98 landrace members of section *Petota*.

Their strict consensus tree shows very few supported clades. Furthermore, we also experimented in our study with scoring extra markers for each AFLP primer combination in certain groups of the dataset (results not shown). The number of AFLP markers increased with up to 30%, but the resolution within the groups did not improve. All these results together suggest that increasing the number of AFLP primer combinations or the number of markers scored does not improve resolution. Regarding the two chloroplast regions (*trnLF* and *psbA/trnH*) one can ask the question whether the chosen regions are the most variable in the chloroplast genome. In a recent publication, Shaw et al. (2005) have compared 21 non coding cpDNA regions on 3 species from each of 10 groups representing eight major phylogenetic lineages. Although in that study no representatives of *Solanum* section *Petota* were included we can conclude for *Solanum* as a whole that the overall level of variability in chloroplast regions is very low. The number of potentially informative characters (PICs) ranges from 0 to 8 as compared to other genera, for example *Prunus* with a PIC value ranging from 0 to 27 or *Gratiola* with a PIC value ranging from 10 to 82. Other studies that used RFLP on Chloroplast DNA to study the taxonomic structure of *Solanum* section *Petota* also do not show any higher levels of resolution than the results shown in this study, but present the data in a different way. A good example is the strict consensus tree on page 682 from Spooner and Castillo (1997) From the 4 clades they found, using cpDNA restriction enzyme site analysis, only clade 2 and clade 1 show acceptable bootstrap values. The other 2 main clades have bootstrap levels that do not reach the 70 bootstrap level as recommended in Hillis and Bull (1993). So the low variability of the Chloroplast DNA might be an intrinsic characteristic of *Solanum*, Adding several other regions might improve the resolution slightly but is probably not worth the effort.

**Structure inside section Petota**

Neither the Chloroplast DNA results nor the AFLP results provide support for maintaining the classification of section *Petota* in 21 series. Although some of these series, like *Circaeifolia* and *Longipedicellata*, can be recognized as clades in the AFLP tree (but with jackknife support varying from high to almost zero), the majority of the series accepted by Hawkes (1990) are not retrieved. The four clades found in the cpDNA RFLP results are largely supported by similar results from the present cpDNA analysis. Our Chloroplast DNA results are concordant with earlier Chloroplast DNA results and the present AFLP results correspond with results from earlier AFLP studies and results from other nuclear data.

In contrast, the four clades of the earlier cpDNA RFLP studies and the groups found in the present Chloroplast DNA analysis do not correspond to the groups found in the AFLP analysis. The cause of these differences can be attributed to the different evolutionary histories that underlie the results of the two marker systems. The evolutionary history of the chloroplast genome is only determined by maternal inheritance. The AFLP results, on the other hand, show the evolutionary history of the nuclear genome to which both male and female parents have contributed equally. It is therefore not surprising that in cases of hybridization, trees derived from such differently inherited genomes will be different. The incongruencies between cpDNA and AFLP make it also difficult to construct 'backbone phylogenies' at the higher taxonomic level using cpDNA and filling in the lower level systematics using another marker like AFLP.

The hybridization events that cause the conflicts between the Chloroplast DNA and nuclear results may also contribute to the difficulties in the taxonomy of section *Petota*. Even with AFLP, a method that generally would produce detailed structure within closely related groups, the relationships among many of the species in section *Petota* are unresolved. Nevertheless, the phylogenetic resolution of the AFLP analysis surpasses by far the structure found with Chloroplast DNA. The AFLP results show more resolution, but the degree of resolution found depends on which part of section *Petota* is studied. Most of the South American diploid species belong to a group of species which shows a lack of supported structure, whereas higher resolution and support is found for the Mexican diploid species. Altogether these results suggest that many of the species within the section *Petota* are genetically very closely related.

## Acknowledgements

We would like to thank Roel Hoekstra from the Centre for Genetic Resources, The Netherlands (CGN) for acquiring the seeds, Vivianne Vleeshouwers from the department of Plant Breeding, Wageningen University and Research centre for growing the plants, and Rolf Mank and Mariëlle Sengers from Keygene B.V. for producing and scoring the AFLP patterns.
We also would like to thank the two anonymous reviewers that provided us with their useful comments.

This project was (co)financed by the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research.

## Supplementary material

The more detailed versions of Figure 1 and Figure 2 containing all the individual labels of the accessions are available upon request.

***Additional information Table 1. (next page)***

*a) Genebank abbreviations: GLKS, Groß Lüsewitz Potato Collection Germany; CGN Centre for Genetic Resources, The Netherlands; PI, Plant Introduction number, The US Potato Genebank (NRSP-6); CPC, Commonwealth Potato Collection, Schotland; CIP, The Potato Collection of the International Potato Centre (CIP), Peru; BGRC, Braunschweig Genetic Resources Centre.*

*b) We list information on synonymy and status changes based on the following treatments: Hawkes(1990), Ochoa (1990), Ochoa (1999), Spooner and Hijmans (2001), Spooner et al. (2004), Spooner and Salas (2006), van den Berg and Spooner (1992), Huaman and Spooner (2002).*

*Table1. Accession codes with information on species, gene bank number, series classification and species status.*

| Label | Species | Source [a] | Series classification | Remarks [b] |
|---|---|---|---|---|
| abz | *S. albornozii* Correll | GLKS5298 | *Piurana* | |
| acg | *S. acroglossum* Juzepczuk | PI498204 | *Piurana* | |
| ach | *S. achacachense* Cárdenas | GLKS2830 | *Tuberosa* (Bol, Arg, Chi) | |
| acl1 | *S. acaule* Bitter | CPC3768 | *Acaulia* | |
| acl2 | *S. acaule* Bitter | CGN20620 | *Acaulia* | |
| acl3 | *S. acaule* Bitter | CGN17930 | *Acaulia* | |
| acp | *S. ancophilum* (Correll) Ochoa | CIP761448 | *Tuberosa* (Per) | |
| acs | *S. acroscopicum* Ochoa | GLKS760030 | *Tuberosa* (Per) | |
| adg | *S. tuberosum* L. subsp. *andigena* Hawkes | CGN20614 | *Tuberosa* (cult) | classified as cultivar-group: Andigenum Group within *S. tuberosum* L. |
| aem | *S. acaule subsp. aemulans* (Bitter and Wittmack) Hawkes and Hjerting | CGN21331 | *Acaulia* | |
| agf | *S. agrimonifolium* Rydberg | CGN18285 | *Conicibaccata* | |
| agu | *S. augustii* Ochoa | CIP762631 | *Tuberosa* (Per) | |
| ajh | *S. ajanhuiri* Juzepczuk and Bukasov | CGN18239 | *Tuberosa* (cult) | classified as cultivar-group: Ajanhuiri Group within *S. tuberosum* L. |
| alb | *S. albicans* (Ochoa) Ochoa | CIP761605 | *Acaulia* | |
| aln | *S. alandiae* Cárdenas | CGN18245 | *Tuberosa* (Bol, Arg, Chi) | |
| aml | *S. amabile* Vargas | PI365356 | *Tuberosa* (Per) | synonym of *S. bukasovii* Juzepczuk |
| amy | *S. amayanum* Ochoa | CIP763005 | *Tuberosa* (Per) | |
| arp | *S. aracc-papa* Juzepczuk | GLKS81 | *Tuberosa* (Per) | nomen dubium |
| arz | *S. arnezii* Cárdenas | GLKS2831 | *Yungasensa* | |
| ast | *S. astleyi* Hawkes and Hjerting | CGN18210 | *Megistacroloba* | subspecies of *S. boliviense* Dunal |
| avl | *S. avilesii* Hawkes and Hjerting | CGN18256 | *Tuberosa (Bol, Arg, Chi)* | |
| aym | *S. aymaraesense* Ochoa | PI607896 | *Tuberosa* (Per) | |
| azn | *S. fendleri* subsp. *arizonicum* Hawkes | GLKS5351 | *Longipedicellata* | synonym of *S. stoloniferum* Schlechtendal and |
| bal | *S. vernei* subsp. *ballsii* (Hawkes) Hawkes and | CGN17992 | *Tuberosa* (Bol, Arg, Chi) | |
| bcp | *S. brachycarpum* Correll | CGN17721 | *Demissa* | synonym of *S. iopetalum* (Bitter) Hawkes |
| ber | *S. berthaultii* Hawkes | CGN20650 | *Tuberosa* (Bol, Arg, Chi) | |
| blb1 | *S. bulbocastanum* Dunal | CGN21306 | *Bulbocastana* | |
| blb2 | *S. bulbocastanum* Dunal | CGN17693 | *Bulbocastana* | |
| blb3 | *S. bulbocastanum* Dunal | CGN21364 | *Bulbocastana* | |
| blg | *S. blanco-galdosii* Ochoa | PI498214 | *Piurana* | |
| blv | *S. boliviense* Dunal | PI472736 | *Megistacroloba* | |
| brc | *S. brevicaule* Bitter | CGN18231 | *Tuberosa* (Bol, Arg, Chi) | |
| bst | *S. brachistotrichium* (Bitter) Rydberg | GLKS2801 | *Pinnatisecta* | synonym of *S. stenophyllidium* Bitter |
| bue | *S. buesii* Vargas | PI568922 | *Conicibaccata* | |
| buk | S. bukasovii Juzepczuk | CGN17737 | *Tuberosa* (Per) | |

*Table1. (Continued) Accession codes with information on species, gene bank number, series classification and species status.*

| Label | Species | Source [a] | Series classification | Remarks [b] |
|---|---|---|---|---|
| can | *S. canasense* Hawkes | CGN17672 | *Tuberosa* (Per) | synonym of *S. bukasovii* Juzepczuk |
| cap1 | *S. capsicibaccatum* Cárdenas | CPC3554 | *Circaeifolia* | variety of *S. circaeifolium* Bitter |
| cap2 | *S. capsicibaccatum* Cárdenas | CGN18297 | *Circaeifolia* | |
| cap3 | *S. capsicibaccatum* Cárdenas | CGN18268 | *Circaeifolia* | |
| cap4 | *S. capsicibaccatum* Cárdenas | CGN22388 | *Cicaeifolia* | |
| cha | *S. chaucha* Juzepczuk and Bukasov | GLKS118 | *Tuberosa* (cult) | classified as cultivar-group: Chaucha Group within *S. tuberosum* L. |
| chc | *S. chacoense* Bitter | CGN18248 | *Yungasensa* | |
| chi | *S. chillonanum* Ochoa | PI607890 | *Tuberosa* (Per) | |
| chm | *S. chomatophilum* Bitter | CGN17713 | *Conicibaccata* | |
| chv | *S. chavinense* Correll | PI498235 | *Megistacroloba* | synonym of *S. dolichocremastrum* Bitter |
| clr1 | *S. clarum* Correll | PI275202 | *Bulbocastana* | |
| clr2 | *S. clarum* Correll | PI604052 | *Bulbocastana* | |
| cmm | *S. commersonii* Dunal | CGN22351 | *Commersoniana* | |
| col | *S. colombianum* Bitter | CGN18289 | *Conicibaccata* | |
| cop | *S. coelestipetalum* Vargas | CIP761755 | *Tuberosa* (Per) | |
| cph1 | *S. cardiophyllum* Lindley | CGN18325 | *Pinnatisecta* | |
| cph2 | *S. cardiophyllum* Lindley | CGN18326 | *Pinnatisecta* | |
| cph3 | *S. cardiophyllum* Lindley | bgrc55227 | *Pinnatisecta* | |
| cph4 | *S. cardiophyllum* Lindley | CGN22387 | *Pinnatisecta* | |
| cph5 | *S. cardiophyllum* Lindley | CGN17697 | *Pinnatisecta* | |
| crc1 | *S. circaeifolium* Bitter supsp. *circaeifolium* | BGRC27058 | *Circaeifolia* | |
| crc2 | *S. circaeifolium* Bitter supsp. *circaeifolium* | CGN18133 | *Circaeifolia* | |
| cur | *S. curtilobum* Juzepczuk and Bukasov | GLKS1620 | *Tuberosa* (cult) | classified as cultivar-group: Curtilobum Group within *S. tuberosum* L. |
| dcm1 | *S. dolichocremastrum* Bitter | GLKS5348 | *Tuberosa* (Per) | |
| dcm2 | *S. dolichocremastrum* Bitter | CIP762533 | *Tuberosa* (Per) | |
| dds | *S. x doddsii* Correll | CGN18359 | *Tuberosa* (Bol, Arg, Chi) | |
| dms1 | *S. demissum* Lindley | CGN18313 | *Demissa* | |
| dms2 | *S. demissum* Lindley | CGN17820 | *Demissa* | |
| dms3 | *S. demissum* Lindley | CGN20571 | *Demissa* | |
| dms4 | *S. demissum* Lindley | CGN17810 | *Demissa* | |
| edn1 | *S. x edinense* P. Berthault | GLKS5493 | *Demissa* | |
| edn2 | *S. x edinense* P. Berthault | GLKS5492 | *Demissa* | |
| ehr1 | *S. cardiophyllum* subsp. *ehrenbergii* Bitter | GLKS5332 | *Pinnatisecta* | recognised as a separate species *S. ehrenbergii* (Bitter) Rydberg |
| ehr2 | *S. cardiophyllum* subsp. *ehrenbergii* Bitter | GLKS2185 | *Pinnatisecta* | |
| etb | *S. etuberosum Lindley* | CGN17714 | *Etuberosa* | |

*Table1. (Continued) Accession codes with information on species, gene bank number, series classification and species status.*

| Label | Species | Source [a)] | Series classification | Remarks [b)] |
|---|---|---|---|---|
| flh | *S. flahaultii* Bitter | PI570620 | *Conicibaccata* | |
| fm | *S. fernandezianum* Phil. | CGN18360 | *Etuberosa* | |
| gab | *S. garcia-barrigae* Ochoa | PI498158 | *Conicibaccata* | |
| gig1 | *S. microdontum* subsp. *gigantophyllum* (Bitter) Hawkes and Hjerting | CGN18046 | *Tuberosa* (Bol, Arg, Chi) | synonym of *S. microdontum* Bitter |
| gig2 | *S. microdontum* subsp. *gigantophyllum* (Bitter) Hawkes and Hjerting | CGN18200 | *Tuberosa* (Bol, Arg, Chi) | |
| gnd | *S. gandarillasii* Cárdenas | CGN20560 | *Tuberosa* (Bol, Arg, Chi) | |
| gon | *S. stenotomum* subsp. *goniocalyx* (Juzepczuk and Bukasov) Hawkes | CGN18314 | *Tuberosa* (cult) | classified as cultivar-group: Stenotomum Group within S. tuberosum L. |
| grl1 | *S. gourlayi* Hawkes | CGN17851 | *Tuberosa* (Bol, Arg, Chi) | synonym of *S. leptophyes* Bitter |
| grl2 | *S. gourlayi* Hawkes | CGN22705 | *Tuberosa* (Bol, Arg, Chi) | |
| grr | *S. guerreroense* Correll | CGN18290 | *Demissa* | |
| han | *S. hannemanii* | CGN17856 | | provisional name |
| haw | *S. hawkesii* Cárdenas | GLKS2762 | *Megistacroloba* | |
| hcb | *S. huancabambense* Ochoa | CGN17719 | *Yungasensa* | |
| hdm | *S. hondelmannii* Hawkes and Hjerting | CGN18106 | *Tuberosa* (Bol, Arg, Chi) | synonym of *S. oplocense* Hawkes |
| hjt1 | *S. hjertingii* Hawkes | CGN17717 | *Longipedicellata* | |
| hjt2 | *S. hjertingii* Hawkes | CGN17718 | *Longipedicellata* | |
| hmp | *S. humectophilum* Ochoa | GLKS2829 | *Tuberosa* (Per) | |
| hps | *S. hoopesii* Hawkes and Okada | CGN18363 | *Tuberosa* (Bol, Arg, Chi) | |
| hro | *S. huarochiriense* Ochoa | CIP761224 | *Tuberosa* (Per) | |
| ifd | *S. infundibuliforme* Phil. | GLKS1700 | *Cuneoalata* | |
| imt | *S. immite* Dunal | GLKS2819 | *Tuberosa* (Per) | |
| inm | *S. incamayoense* Okada and Clausen | CGN22335 | *Tuberosa* (Bol, Arg, Chi) | |
| iop1 | *S. iopetalum* (Bitter) Hawkes | CPC2922 | *Demissa* | |
| iop2 | *S. iopetalum* (Bitter) Hawkes | CGN20561 | *Demissa* | |
| irs | *S. irosinum* Ochoa | CIP762259 | *Conicibaccata* | |
| jam1 | *S. jamesii* Torres | CGN18349 | *Pinnatisecta* | |
| jam2 | *S. jamesii* Torres | BGRC53630 | *Pinnatisecta* | |
| juz | *S. juzepczukii* Bukasov | GLKS5467 | *Tuberosa* (cult) | classified as cultivar-group: Juzepczukii Group within *S. tuberosum* L. |
| ktz | *S. kurtzianum Bitter and Wittmack* | CGN22338 | *Tuberosa* (Bol, Arg, Chi) | |
| les1 | *S. lesteri* Hawkes and Hjerting | PI442679 | *Polyadenia* | |
| les2 | *S. lesteri* Hawkes and Hjerting | PI558435 | *Polyadenia* | |
| lgc1 | *S. longiconicum* Bitter | GLKS5363 | *Conicibaccata* | |
| lgc2 | *S. longiconicum* Bitter | PI604094 | *Conicibaccata* | |
| lgl | *S. lignicaule* Vargas | CGN17723 | *Lignicaulia* | |

*Table1. (Continued) Accession codes with information on species, gene bank number, series classification and species status.*

| Label | Species | Source [a] | Series classification | Remarks [b] |
|---|---|---|---|---|
| lmb | *S. limbaniense* Ochoa | CGN22720 | *Conicibaccata* | |
| lph | *S. leptophyes* Bitter | CGN18140 | *Tuberosa* (Per) | |
| lps | *S. leptosepalum* Correll | PI607843 | *Tuberosa* (Mex, Ven, Col, | synonym of *S. stoloniferum* Schlechtendal and |
| lxs | *S. laxissimum* Bitter | CGN22721 | *Conicibaccata* | |
| mag | *S. maglia Schlechtendal* | CGN18064 | *Maglia* | |
| mat | S. matehualae Hjerting et Tarn | GLKS5364 | *Longipedicellata* | synonym of *S. hjertingii* Hawkes |
| mcd | *S. microdontum* Bitter | CGN17596 | *Tuberosa* (Bol, Arg, Chi) | |
| mch | *S. x michoacanum* (Bitter) Rydberg | GLKS2346 | *Pinnatisecta* | |
| mcp | *S. macropilosum* Correll | PI607844 | *Tuberosa* (Mex, Ven, Col, | synonym of *S. verrucosum* Schlechtendal |
| mcq | *S. mochiquense* Ochoa | CGN17731 | *Tuberosa* (Per) | |
| med | *S. medians* Bitter | CGN18043 | *Tuberosa* (Per) | |
| mga | *S. megistacrolobum* Bitter | CGN17828 | *Megistacroloba* | |
| min | *S. minutifoliolum* Correll | PI583298 | *Tuberosa* (Mex, Ven, Col, | |
| mlm | *S. commersonii* Dunal subsp. *malmeanum* | CPC7520 | *Commersoniana* | |
| mlt1 | *S. multidissectum* Hawkes | CGN17824 | *Tuberosa* (Per) | |
| mlt2 | *S. multidissectum* Hawkes | CGN21344 | *Tuberosa* (Per) | |
| mrl1 | *S. morelliforme* Bitter and Muench | GLKS2245 | *Morelliformia* | |
| mrl2 | *S. morelliforme* Bitter and Muench | PI619119 | *Morelliformia* | |
| mrl3 | *S. morelliforme* Bitter and Muench | PI545720 | *Morelliformia* | |
| mrn | *S. marinasense* Vargas | GLKS2281 | *Tuberosa* (Per) | |
| mtp | *S. multiinterruptum* Bitter | GLKS2431 | *Tuberosa* (Per) | |
| ncd | *S. neocardenasii* Hawkes and Hjerting | CGN18217 | *Tuberosa* (Bol, Arg, Chi) | |
| nrs | *S. neorossii* Hawkes and Hjerting | CGN18000 | *Tuberosa* (Bol, Arg, Chi) | |
| nyr1 | *S. nayaritense (Bitter)* Rydberg | PI545825 | *Pinnatisecta* | synonym of *S. stenophyllidium* Bitter |
| nyr2 | *S. nayaritense (Bitter)* Rydberg | PI545820 | *Pinnatisecta* | |
| oka | *S. okadae* Hawkes and Hjerting | CGN18108 | *Tuberosa* (Bol, Arg, Chi) | |
| opl | *S. oplocense* Hawkes | CGN23049 | *Tuberosa* (Bol, Arg, Chi) | |
| oro | *S. orocense* Ochoa | PI583307 | *Conicibaccata* | |
| orp | *S. orophilum* Correll | PI498213 | *Tuberosa* (Per) | |
| oti | *S. otites* Dunal | PI570618 | *Conicibaccata* | |
| pam | *S. pampasense* Hawkes | CGN20575 | *Tuberosa* (Per) | |
| pcj | *S. paucijugum* Bitter | PI561650 | *Conicibaccata* | |
| pcs | *S. paucissectum* Ochoa | PI590922 | *Piurana* | |
| phu | *S. phureja* Juzepczuk and Bukasov | CGN17667 | *Tuberosa* (cult) | classified as cultivar-group: Phureja Group within *S. tuberosum* L. |
| pld1 | *S. polyadenium* Greenman | CGN17749 | *Polyadenia* | |
| pld2 | *S. polyadenium* Greenman | CGN17746 | *Polyadenia* | |
| pld3 | *S. polyadenium* Greenman | CGN23013 | *Polyadenia* | |

*Table1. (Continued) Accession codes with information on species, gene bank number, series classification and species status.*

| Label | Species | Source [a)] | Series classification | Remarks [b)] |
|---|---|---|---|---|
| pld4 | *S. polyadenium* Greenman | CGN23014 | *Polyadenia* | |
| pls | *S. palustre* Poeppig | CPC7034 | *Etuberosa* | |
| plt | *S. polytrichon* Rydberg | CGN17750 | *Longipedicellata* | synonym of *S. stoloniferum* Schlechtendal and |
| pne | *S. acaule Bitter subsp. punae (Juzepczuk)* Hawkes and Hjerting | CGN20665 | *Acaulia* | |
| pnt1 | *S. pinnatisectum* Dunal | GLKS1586 | *Pinnatisecta* | |
| pnt2 | *S. pinnatisectum* Dunal | CGN17745 | *Pinnatisecta* | |
| pnt3 | *S. pinnatisectum* Dunal | CGN17743 | *Pinnatisecta* | |
| pnt4 | *S. pinnatisectum* Dunal | CGN23011 | *Pinnatisecta* | |
| prm | *S. paramoense* Bitter | Pi604202 | *Tuberosa* (Mex, Ven, Col, | synonym of *S. tuberosum* L. |
| pta1 | *S. papita* Rydberg | CGN17830 | *Longipedicellata* | synonym of *S. stoloniferum* Schlechtendal and |
| pta2 | *S. papita* Rydberg | CGN17832 | *Longipedicellata* | |
| ptr1 | *S. gourlayi* subsp. *pachytrichum* (Hawkes) Hawkes and Hjerting | CGN18102 | *Tuberosa* (Bol, Arg, Chi) | synonym of *S. leptophyes* Bitter |
| ptr2 | *S. gourlayi* subsp. *pachytrichum* (Hawkes) Hawkes and Hjerting | CGN18176 | *Tuberosa* (Bol, Arg, Chi) | |
| ptt | *S. bulbocastanum* subsp. *partitum* (Correll) | GLKS5322 | *Bulbocastana* | synonym of *S. bulbocastanum* Dunal |
| pur | *S. piurae* Bitter | PI365365 | *Piurana* | |
| qum1 | *S. circaeifolium* subsp. *quimense* Hawkes and | CGN18127 | *Circaeifolia* | synonym of *S. circaeifolium* var. *capsicibaccatum* |
| qum2 | *S. circaeifolium* subsp. *quimense* Hawkes and | CGN18128 | *Circaeifolia* | |
| qum3 | *S. circaeifolium* subsp. *quimense* Hawkes and | CGN20643 | *Circaeifolia* | |
| qum4 | *S. circaeifolium* subsp. *quimense* Hawkes and | CGN22767 | *Circaeifolia* | |
| qum5 | *S. circaeifolium* subsp. *quimense* Hawkes and | CGN18158 | *Circaeifolia* | |
| rzl | *S. x ruiz-lealii* (Brücher) | CGN18117 | *Tuberosa* (Bol, Arg, Chi) | |
| san | *S. santolallae* Vargas | CGN18293 | *Conicibaccata* | |
| scb | *S. scabrifolium* Ochoa | PI365363 | *Tuberosa* (Per) | |
| sct | *S. sanctae-rosae* Hawkes | CGN17837 | *Megistacroloba* | |
| sem | *S. x semidemissum* Juzepczuk | CPC7331 | *Demissa* | synonym of S. *demissum* Lindley |
| sgr1 | *S. sogarandinum* Ochoa | GLKS5382 | *Megistacroloba* | |
| sgr2 | *S. sogarandinum* Ochoa | CGN17601 | *Megistacroloba* | |
| snd | *S. sandemanii* Hawkes | CGN17600 | *Tuberosa* (Per) | |
| snk | *S. schenckii* Bitter | CPC7164 | *Demissa* | |
| sol | *S. solisii* Hawkes | GLKS5383 | *Piurana* | |
| sou | *S. soukupii* Hawkes | CGN18061 | *Tuberosa* (Per) | synonym of *S. bukasovii* Juzepczuk |
| spg | *S. spegazzinii* Bitter | CGN17839 | *Tuberosa* (Bol, Arg, Chi) | |
| spl | *S. sparsipilum* ( Bitter) Juzepczuk and Bukasov | CGN20653 | *Tuberosa* (Per) | |
| sst | *S. soestii* Hawkes and Hjerting | BGRC63075 | *Circaeifolia* | |
| stl | *S. x setulosistylum* Bitter | CGN20655 | *Tuberosa* (Bol, Arg, Chi) | |

*Table1. (Continued) Accession codes with information on species, gene bank number, series classification and species status.*

| Label | Species | Source [a)] | Series classification | Remarks [b)] |
|---|---|---|---|---|
| stn | *S. stenotomum* Juzepczuk and Bukasov | CGN20616 | *Tuberosa* (cult) | classified as cultivar-group: Stenotomum Group within *S. tuberosum* L. |
| sto1 | *S. stoloniferum* Schlechtendal and Bouchet | CGN17605 | *Longipedicellata* | |
| sto2 | *S. stoloniferum* Schlechtendal and Bouchet | CGN18332 | *Longipedicellata* | |
| sub | *S. x subandigena* Hawkes | GLKS722 | *Tuberosa* (Bol, Arg, Chi) | synonym of *S. oplocense* Hawkes |
| suc | *S. sucubunense* Ochoa | PI583320 | *Conicibaccata* | |
| sup | *S. subpanduratum* Ochoa | GLKS2873 | *Conicibaccata* | |
| tar | *S. tarijense* Hawkes | CGN18107 | *Yungasensa* | |
| tbr | *S. tuberosum* L. | BGRC28469 | *Tuberosa* (cult) | |
| tor | S. toralapanum Cárdenas and Hawkes | CGN18147 | *Megistacroloba* | subspecies of *S. megistacrolobum* Bitter |
| trf1 | *S. trifidum* Correll | CGN18335 | *Pinnatisecta* | |
| trf2 | *S. trifidum* Correll | CGN22722 | *Pinnatisecta* | |
| trn1 | *S. tarnii* Hawkes and Hjerting | GLKS5385 | *Pinnatisecta* | |
| trn2 | *S. tarnii* Hawkes and Hjerting | PI570642 | *Pinnatisecta* | |
| tuq | *S. tuquerrense* Hawkes | CGN18353 | *Piurana* | |
| ugt | *S. ugentii* Hawkes and Okada | GLKS2887 | *Tuberosa* (Bol, Arg, Chi) | |
| ver1 | *S. verrucosum* Schlechtendal | CGN17768 | *Tuberosa* (Mex, Ven, Col, | |
| ver2 | *S. verrucosum* Schlechtendal | CGN17769 | *Tuberosa* (Mex, Ven, Col, | |
| vid1 | *S. gourlayi* subsp. *vidaurrei* (Cárdenas) Hawkes and Hjerting | CGN18040 | *Tuberosa* (Bol, Arg, Chi) | |
| vid2 | *S. gourlayi* subsp. *vidaurrei* (Cárdenas) Hawkes and Hjerting | CGN18038 | *Tuberosa* (Bol, Arg, Chi) | |
| vlr | *S. velardei* Ochoa | CGN18324 | *Tuberosa* (Per) | |
| vnt | *S. venturii* Hawkes and Hjerting | CGN17755 | *Tuberosa* (Bol, Arg, Chi) | |
| vrg | *S. virgultorum* (Bitter) Cárdenas and Hawkes | BGRC31203 | *Tuberosa* (Bol, Arg, Chi) | |
| vrn | *S. vernei* Bitter and Wittmack | CGN18278 | *Tuberosa* (Bol, Arg, Chi) | |

*Table 2. Summary of the results found with Neighbour Joining (NJ) and maximum Parsimony (MP) jackknife anlyses for the cpDNA and AFLP data sets*

| Groups found in analyses | AFLP support in NJ jackknife | AFLP support in MP jackknife | cpDNA support in NJ jackknife | cpDNA support in MP jackknife |
|---|---|---|---|---|
| ***Morelliformia / Bulbocastana / Pinnatisecta / Polyadenia (summarized as Mexican diploids)*** | 93<br>Including cph1, cph2, cph4, blb1, blb2, blb3, and ptt. | 76<br>Including cph1, cph2, cph4, blb1, blb2, blb3, and ptt. | 99<br>Excluding cph1, cph2, cph4, blb1, blb2, blb3, and ptt. | 98<br>Excluding cph1, cph2, cph4, blb1, blb2, blb3, and ptt. |
| ***cph1, cph2, and cph4*** | 80<br>cph1, cph2, and cph4 together with cph3 embedded in cluster of Mexican diploids. | 100<br>cph1, cph2, and cph4 together with cph3 and cph5 as a clade embedded within clade of Mexican diploids. | 100<br>cph1, cph2, and cph4 form a separate cluster outside cluster of Mexican diploids. | 100<br>cph1, cph2, and cph4 form a separate clade outside clade of Mexican diploids. |
| ***blb1, blb2, blb3, and ptt*** | 99<br>blb1, blb2, blb3, and ptt embedded in cluster of Mexican diploids. | 83<br>blb1, blb2, blb3 and ptt embedded as a clade within clade of Mexican diploids. | 100<br>blb1, blb2, blb3, and ptt as separate cluster outside cluster of Mexican diploids. | 100<br>blb1, blb2, blb3, and ptt as a separate clade outside clade of Mexican diploids. |
| ***Circaeifoliae*** | 100<br>Present as a separate cluster. | 100<br>Present as a separate clade. | -<br>Not present as a cluster.<br>Represented by 3 separate clusters:<br>cap1, cap2, cap3, cap4 and qum4: 71;.<br>crc1, crc2: 86;<br>qum1, qum2, qum3, qum5, sst: 61. | -<br>Not present as a clade.<br>Represented by 3 separate clades:<br>cap1, cap2, cap3, cap4 and qum4: 65;<br>crc1, crc2: 70;<br>qum1, qum2, qum3, qum5, sst: 65. |
| ***Acaulia*** | 96<br>Present as a separate cluster. | 99<br>Present as a separate clade. | -<br>Not present as a cluster. | -<br>Not present as a clade. |
| ***edn1, edn2, sem, dms1, dms4, dms2, and dms3.*** | 100<br>Present as a separate cluster. | 87<br>Present as a separate clade. | -<br>Not present as a cluster, genotypes embedded within cluster of Mexican polyploids. | -<br>Not present as clade, genotypes embedded within clade of Mexican polyploids. |
| ***Acaulia / edn1, edn2, sem, dms1, dms4, dms2, and dms3.*** | 96<br>Present as a separate cluster. | 99<br>Present as a separate clade. | -<br>Not present as a cluster. | -<br>Not present as a clade. |
| ***Longipedicellata*** | 100<br>Present as a separate cluster. | 100<br>Present as a separate clade. | -<br>Not present as a separate cluster but genotypes are mixed with *Demissa* genotypes within cluster of Mexican polyploids. | -<br>Not present as separate clade but genotypes are mixed with *Demissa* genotypes within cluster of Mexican polyploids. |

*Table 2. (Continued) Summary of the results found with Neighbour Joining (NJ) and maximum Parsimony (MP) jackknife anlyses for the cpDNA and AFLP data*

| Groups found in analyses | AFLP support in NJ jackknife | AFLP support in MP jackknife | cpDNA support in NJ jackknife | cpDNA support in MP jackknife |
|---|---|---|---|---|
| ***ver1, ver2, mcp*** | 100<br>Present as a separate cluster. | 63<br>Present as a separate clade. | -<br>Not present as a cluster, genotypes embedded within cluster of Mexican polyploids. | -<br>Not present as a clade, genotypes embedded within clade of Mexican polyploids. |
| ***Demissa*** | 99<br>Present as a separate cluster, but excluding<br>edn1, edn2, sem, dms1, dms4, dms2, and dms3. | 68<br>Present as a separate clade, but excluding edn1, edn2, sem, dms1, dms4, dms2, and dms3 (in the strict consensus a clade excluding snk has support of 87). | -<br>Not present as a separate cluster but genotypes are mixed with *Longipedicellata* genotypes within cluster of Mexican polyploids. | -<br>Not present as separate clade but genotypes are mixed with *Longipedicellata* genotypes within clade of Mexican polyploids. |
| ***Longipedicellata / Demissa / ver1, ver2, mcp (summarized as Mexican polyploids)*** | -<br>Not present as a cluster. | 45<br>Present as a clade. | 65<br>Present as a separate cluster. | 61<br>Present as a separate clade. |
| ***Demissa / Conicibaccata*** | 69<br>Present as a cluster. | -<br>Not present as a clade. | -<br>Not present as cluster. | -<br>Not present as a clade. |
| ***Longipedicellatan / ver1, ver2, and mcp*** | 58<br>Present as a cluster. | -<br>Not present as a clade. | -<br>Not present as a separate cluster. | -<br>Not present as a clade. |
| ***Conicibaccata*** | 85<br>Present as a cluster of suc, oro, agf, col, flh, lgc2, lgc1, gab, oti, sup, and prm. | 91<br>Present as a clade of suc, oro, agf, col, lgc2, lgc1, flh, sup, oti, gab, and prm. | -<br>Not present as a cluster, *Conicibaccata* genotypes mixed with *Tuberosa* genotypes. | -<br>Not present as a clade, *Conicibaccata* genotypes mixed with *Tuberosa* genotypes. |
| ***Piurana*** | 61<br>Small cluster of pcs, pur, chm and irs (plus other separate cluster of sol, pcj and tuq with jackknife value of 58). | 64<br>Small clade of pcs, pur, irs, chm, sol, pcj and tuq. | 80<br>Cluster of pcs, imt, sgr2, sgr1, hcb, blg, acg, irs, chm, agu, acs, sol, pcj, tuq, pur, and mcq. | 78<br>Clade of pcs, sgr1, hcb, imt, sgr2, blg, acg, irs, chm, agu, acs, sol, pcj, tuq, pur, and mcq. |
| ***Tuberosa cultivated*** | -<br>Not present as cluster but genotypes mixed within *Tuberosa* (Peru) genotypes and one small cluster of cur, can, phu, cha, sub and gon with jackknife value of 79. | -<br>Not present as clade but genotypes mixed within *Tuberosa* (Peru) genotypes and one small cluster of cur, can, phu, cha, sub and gon with jackknife value of 44. | -<br>Not present as cluster but genotypes are embedded in large grade with other *Tuberosa*. | -<br>Not present as clade but genotypes are embedded in large grade with other *Tuberosa*. |
| ***Tuberosa (Peru)*** | -<br>Not present as supported cluster but several small/very small clusters of genotypes with support. | -<br>Not present as supported clade but several small/very small clades of genotypes with support. | -<br>Embedded in large grade with other *Tuberosa*. | -<br>Embedded in large grade with other *Tuberosa*. |
| ***Tuberosa (Bolivia, Argentina, Chile)*** | -<br>Not present as supported cluster but several small/very small clusters of genotypes with support. | -<br>Not present as supported clade but several small/ very small clades of genotypes with support. | 61<br>Cluster of dds, ast, tbr, hdm, ber, opl, blv, lxs, aln, gig1, and gig2, other genotyopes are embedded in grade with other *Tuberosa*. | 65<br>Clade of dds, hdm, aln, tbr, ast, blv, ber, opl, lxs, gig1, and gig2, other genotypes are embedded in grade with other *Tuberosa*. |

# CHAPTER 4

## AFLP analysis reveals a lack of phylogenetic structure within *Solanum* section *Petota*

Mirjam M.J. Jacobs, Ronald R.G. van den Berg, Vivianne G.A.A. Vleeshouwers, Marcel Visser, Rolf Mank, Mariëlle Sengers, Roel Hoekstra and Ben Vosman

(Published in BMC Evolutionary Biology, 2008, 8:145)

## Abstract

The secondary genepool of our modern cultivated potato (*Solanum tuberosum* L.) consists of a large number of tuber-bearing wild *Solanum* species under *Solanum* section *Petota*. One of the major taxonomic problems in section *Petota* is that the series classification (as put forward by Hawkes) is problematic and the boundaries of some series are unclear. In addition, the classification has received only partial cladistic support in all molecular studies carried out to date. The aim of the present study is to describe the structure present in section *Petota*. When possible, at least 5 accessions from each available species and 5 individual plants per accession (totally approx. 5000 plants) were genotyped using over 200 AFLP markers. This resulted in the largest dataset ever constructed for *Solanum* section *Petota*. The data obtained are used to evaluate the 21 series hypothesis put forward by Hawkes and the 4 clade hypothesis of Spooner and co-workers.

We constructed a NJ tree for 4929 genotypes. For the other analyses, due to practical reasons, a condensed dataset was created consisting of one representative genotype from each available accession. We show a NJ jackknife and a MP jackknife tree. A large part of both trees consists of a polytomy. Some structure is still visible in both trees, supported by jackknife values above 69. We use these branches with >69 jackknife support in the NJ jackknife tree as a basis for informal species groups. The informal species groups recognized are: Mexican diploids, Acaulia, Iopetala, Longipedicellata, polyploid Conicibaccata, diploid Conicibaccata, Circaeifolia, diploid Piurana and tetraploid Piurana. Most of the series that Hawkes and his predecessors designated can not be accepted as natural groups, based on our study. Neither do we find proof for the 4 clades proposed by Spooner and co-workers. A few species groups have high support and their inner structure displays also supported subdivisions, while a large part of the species cannot be structured at all. We believe that the lack of structure is not due to any methodological problem but represents the real biological situation within section *Petota*.

## Background

The secondary genepool of our modern cultivated potato (*Solanum tuberosum* L.) consists of a large number of tuber-bearing wild *Solanum* species which grow in various habitats from the southern states of the USA to the most southern parts of Chile and Argentina. These wild species are important as a resource for valuable traits that can be used to improve the quality of the cultivars, including resistance against important diseases like *Phytophthora infestans* and potato cyst nematodes (*Globodera* spp.). Therefore it is no surprise that the wild relatives of the cultivated potato have since long drawn the attention of many plant breeders and botanists. To benefit most from the possibilities that the secondary genepool has to offer, it is necessary to have a good insight in the taxonomy. The classical treatments of potato taxonomy are from Correll (1962), and Hawkes (1990), later followed by reviews from Spooner and Hijmans (2001), Spooner and Salas (2006), and van den Berg and Jacobs (2007).

There are two major taxonomic problems in the section *Petota*. First, many described species are extremely similar to each other and section *Petota* seems to be overclassified (Van den Berg & Jacobs 2007). In many cases, potato species can only be distinguished by means of multivariate analysis of quantitative characters and/or on the basis of geographic origin (Giannattasio & Spooner 1994; Van den Berg *et al.* 1998; Van den Berg & Groendijk-Wilders 1999; Kardolus 1998).
The main cause for these difficulties is the ability of many species in section *Petota* to hybridize easily with other species (Spooner & Salas 2006). Many species have been suspected to arise from hybrid speciation. Other causes are high morphological similarity among species, and phenotypic plasticity in different environments (Spooner & Hijmans 2001). In recent reviews the number of species is reduced due to increased insights in potato taxonomy. Hawkes (1990) recognized 227 tuber bearing species (7 cultivated species included) and 9 non-tuber-bearing species within section *Petota*. Spooner and Hijmans (2001) recognized 203 tuber-bearing species including 7 cultivated species. Finally, Spooner and Salas (2006) reduced the number further to 189 species (including 1 cultivated species) in section *Petota*.

The second taxonomic problem is the series classification. Hawkes (1990) classified section *Petota* into 19 tuber bearing series plus two non-tuber bearing series that vary considerably in the number of species included. The boundaries between some series are unclear. As outlined earlier by Spooner et al. (2004) the series classification of Hawkes and previous authors has received only partial cladistic support in any molecular study to date. The cpDNA RFLP data from Spooner and Sytsma (1992), Castillo and Spooner (1997), Rodriguez and Spooner (1997), and Spooner and Castillo (1997) could only find support for a classification in 4 clades.

The aim of the present study is to focus on the second problem and to describe the structure within section *Petota*. In the present study the largest number of species and accessions to date are examined in one simultaneous AFLP analysis. The obtained data are used for evaluation of the hypothesis put forward by Hawkes (1990) that section *Petota* can be divided in 21 series and the hypothesis of Spooner and Castillo (1997), that the section consists of 4 clades only.

AFLP has proven to be a useful method to solve phylogenetic relationships at a low taxonomic level (Despres et al. 2003; Koopman 2005; Meudt & Clarke 2006). The application of AFLP has many advantages. It produces highly reproducible data (Jones *et al.* 1997), it does not need a priori sequence information and it has the ability of high resolution (Meudt & Clarke 2007). Because AFLP generates fragments at random over the whole genome it avoids the problem that many sequence data based phylogeny reconstructions have, e.g. the generation of a gene tree instead of a species tree (Despres et al. 2003).

## Methods

### Plant Material

In total 951 accessions representing 196 different taxa, species, 15 subspecies and 17 hybrids were sampled. We tried to include as many species as possible from various gene banks. In principle, at least 5 accessions from each available species and 5 individual plants per species (totally approx. 5000 genotypes) were included. Seeds were surface-sterilized and sown in vitro at 25°C. The collection of individual *Solanum* clones was grown in vitro for at least 6 weeks on MS medium supplemented with 20% sucrose (Murashigi & Skoog 1962) at 18°C. DNA was extracted from leafs according to the method described by Stewart and Via (1993).

### Nomenclature

Additional file 1 lists the species used and the accessions representing the species names according to the passport information from the gene bank. The labels used are not corrected according to the synonymy in recent taxonomic revisions for two reasons. First, we do not want to change an original label of an accession without actually checking the identity of that accession. Furthermore, by retaining the original labels it is possible to check many hypotheses on the taxonomy of species. However, we have included some remarks about recent taxonomy changes in additional file 1. In some cases names/labels were corrected by us after preliminary AFLP results and visual inspection of the plant material in the greenhouse or on the field. If an accession could be assigned to another species according to AFLP pattern and morphology, it was given the name of this species, If there were any doubts on the identification the species was given the label *S. spec.* The accessions which labels were changed are indicated in additional file 1.

### AFLP

The samples were fingerprinted with two *Eco*RI/*Mse*I AFLP primer combinations: E32/M49 and E35/M48. The protocol of Vos et al. (1995) was used to generate AFLP fragments. Primer combination E32/M49 yielded 91 polymorphic bands and primer combination E35/M48 yielded 131 bands. Keygene carried out the AFLP analysis on a MegaBACE 2.1 and scored the bands using their proprietary software. Bands were scored as dominant markers, so only the presence or the absence of a band was scored.

### Datasets

The dataset in this study originally contained 4929 genotypes. This large dataset was analyzed with NJ and UPGMA. Because of the size of the dataset, it proved impossible to analyze it with cladistic methods nor to analyze it for statistical support, even using the SARA supercomputer (see below). It was sheer impossible for a personal computer to do any further analyses apart from the NJ and UPGMA, and for the SARA computer cluster it would have taken many months/years of computing time. For further analysis a condensed dataset was created by carefully choosing a representative genotype from all the available accessions. This condensed dataset consisted of 916 genotypes. The condensed dataset was used in both phenetic and cladistic analyses and in the resampling methods.

Besides choosing only one genotype per accession to represent the accession in the condensed dataset, other adjustments were made to create this dataset. All the 22 known interspecific hybrid accessions were removed, 23 other accessions were completely removed because of the extreme heterogeneity of the accession (possibly resulting from a mixture of species) in both the NJ and the UPGMA trees. Species labels of 49 accessions were changed based on their position in the NJ and/ or UPGMA tree (not shown) and visual inspection of the plants in the experimental field or greenhouse in 2005 and 2006. In total 11 outgroup accessions were removed because preliminary AFLP results showed these outgroups to be too distant (*S. sitiens, S. nigrum, S. chaparense, S. lycopersicoides, S. canense, S. fraxinifolium*). The outgroup species *S. etuberosum*, *S. palustre* and *S. fernandezianum* were retained in the dataset.

**Data analysis**

Both the phenetic and the cladistic analyses were conducted using PAUP 4.0 Altivec (Swofford, 2002) on the TERAS computing cluster of SARA computing facilities in Amsterdam. For the 4929 phenetic analysis we used the total character distance, for the 916 data set we used the NeiLi distance (Nei & Li, 1979) to calculate the distance matrix. A Neighbor Joining Jackknife tree was calculated using 10.000 replicates. The cladistic analysis heuristic searches were done by using PRAP, Parsimony Ratchet Analyses using PAUP, a program that writes commands for PAUP. The commands in PRAP describe how PAUP should carry out parsimony ratchet searches (http://www.nees.uni-bonn.de/ downloads/PRAP). By using parsimony ratchet, as described by Nixon (1999), many tree islands are searched instead of thoroughly searching through each island. For the MP jackknife analysis, we followed the conclusions drawn by Muller (http://www.nees.uni-bonn.de/downloads/PRAP) that using random addition sequence instead of simple addition sequence has no beneficial effect on bootstrap or jackknife support. Also, a jackknife or bootstrap analysis using one heuristic search saving one tree per jackknife replicate and simple addition sequence, performed as good as or even better than an analysis using 10 parsimony ratchet iterations using the shortest tree only or using a strict consensus tree of all shortest trees (http://www.nees.uni-bonn.de/downloads/PRAP). Therefore, we conducted a jackknife MP analysis by performing 10.000 replicates using simple addition, and saving one shortest tree per replicate.

## Results

**The large dataset (4929 genotypes)**

Figure 1 shows the Neighbor Joining (NJ) tree of the 4929 genotypes dataset. To describe the structure found in this NJ tree, we differentiate between 3 levels of structure: the accession level, the species level and the interspecies level. At the accession level, the genotypes of the majority of the accessions cluster together. Of those accessions that do not form complete clusters, in most cases only one genotype deviates from the other 4 genotypes. In other cases, the accession was apparently so closely related with one or more other accessions that their genotypes formed a mixed group. At the species level, 58 species or subspecies show consistency in their clustering, e.g. all accessions of a species cluster together. Nevertheless there are also many species (38 in total) whose accessions did not cluster all together and 48 species whose accessions were mixed with accessions of other species. The latter was often the case with species that occur in South America, the borders of many of these species are not clearly recognizable from the NJ tree. Above the species level, a few clusters of species groups can be distinguished in the large NJ tree (but there is no indication on the statistical strength of the structure observed). Roughly, the following groups can be found in the NJ tree of the large dataset: 1) an outgroup with *S. nigrum, S. chaparense, S. sitiens, and S. fraxinifolium* 2) North and Central American diploid series *Polyadenia*, *Pinnatisecta*, *Bulbocastana* and *Morelliformia*, 3) *Circaeifolia* and *Piurana* accessions, 4) *Longipedicellata* accessions, 5) *Demissa* and *Conicibaccata* accessions but without *S. demissum* and *S. semidemissum*, 6) *S. verrucosum* accessions, 7) *Tuberosa* from Bolivia, Argentina and Chile plus some accessions from other series such as *Yungasensa*, 8) accessions from cultivated *Tuberosa* species and wild *Tuberosa* from Peru, 9) accessions from *Tuberosa* and *Megistacroloba*, 10) accessions from *S. acaule* (and its subspecies), *S. albicans, S. demissum, S x semidemissum* and *S. edinense*.

**The condensed dataset (916 genotypes)**

Because of the size of the dataset, it proved impossible to analyze it with cladistic methods nor to analyze it for statistical support. A condensed dataset was created by choosing a representative genotype from all the available accessions (see methods section for exact details). This condensed dataset consisted of 916 genotypes. A single ratchet parsimony search consisting of 200 iterations yielded a Maximum Parsimony (MP) tree of 9669 steps. Furthermore, 20 individual independent ratchet searches each consisting of 50 iterations also yielded a MP tree of 9669 steps. Figure 2 shows the schematised majority rule consensus NJ jackknife tree and Figure 3 shows the schematised majority rule consensus MP jackknife tree of the condensed dataset. The strict consensus trees were manipulated in such a manner that not all the separate branches were represented but some were summarised. The schematised trees only show branches with more than 69 jackknife support. The original majority rule consensus NJ jackknife tree and majority rule consensus MP jackknife tree are available from the authors as supplemental data.

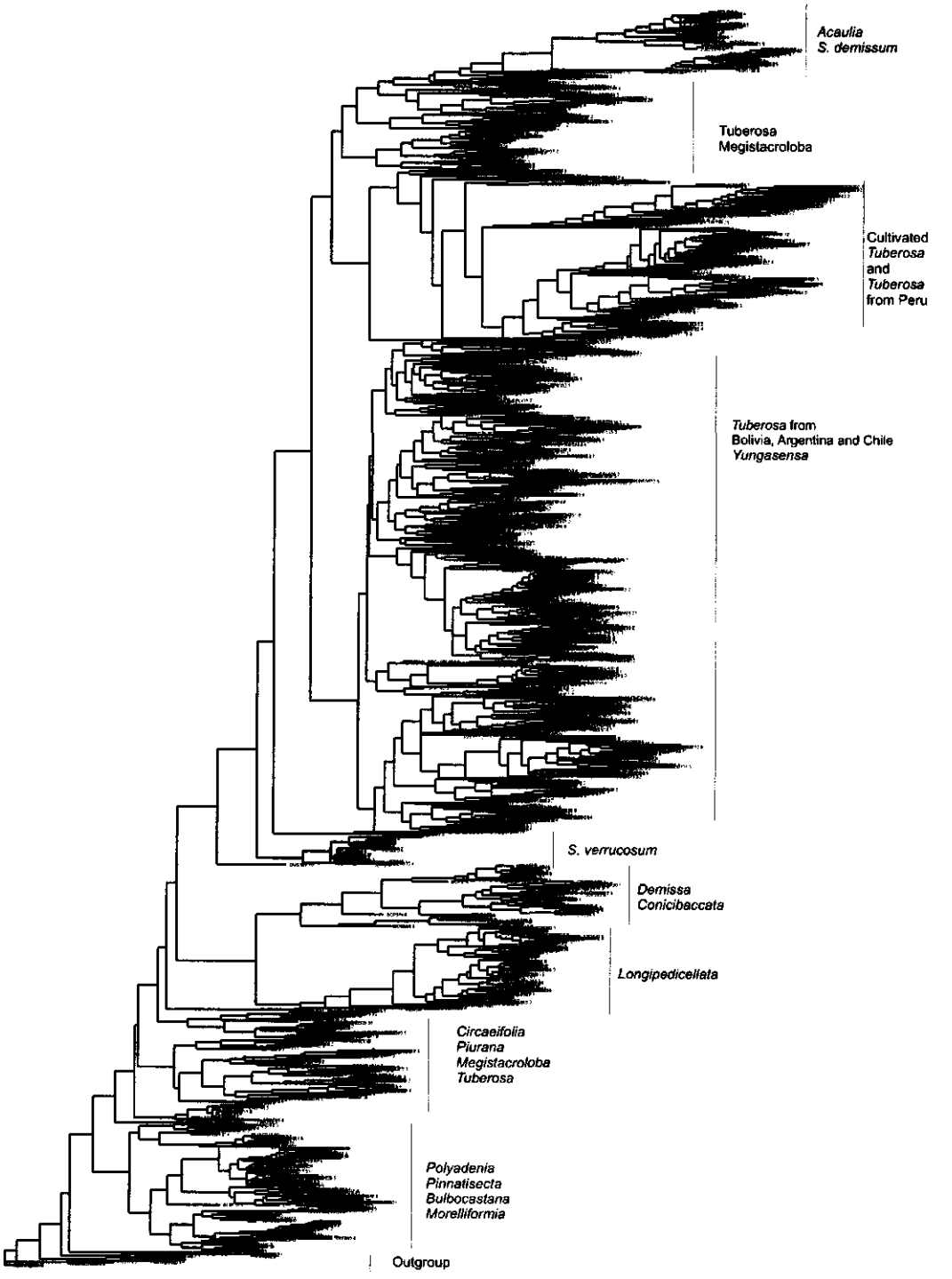



*Figure 1. Neighbour Joining tree, complete dataset*

When comparing the NJ and the MP jackknife trees it is apparent that a large part of both trees consists of a polytomy. However, some structure is still visible in both trees, supported by jackknife values above 69. The following groups can be recognized in both the NJ jackknife tree and the MP jackknife tree:

1) Mexican diploid species, with a jackknife support of 73 for the MP tree and 99 for the NJ tree; the substructure found within the Mexican and Northern American diploids is almost the same for both trees.

2) A group of tetraploid Mexican/ North and Central American species belonging to series *Longipedicellata*, with a jackknife support of 100 in both trees.

3) A group consisting of accessions of *S. acaule*, *S. demissum*, and closely related species with a jackknife support of 100 in the MP tree and 99 in the NJ tree.

4) A group consisting of the species belonging to series *Circaeifolia*, with a jackknife support of 100 in both trees.

5) A small group of accessions belonging to *S. paucijugum*, *S. tuquerrense*, and *S. solisii,* tetraploid species belonging to the series *Piurana*, with a jackknife support of 96 in the NJ tree and 92 in the MP tree.

There are also differences in group structure between the two trees. There are a number of groups that have good jackknife support in the NJ tree but are not supported in the MP jackknife tree:

1) A group of hexaploid Mexican species belonging to series *Demissa* with a jackknife support of 79. In the MP tree only 2 species that are part of this group were found in one small clade: *S. schenckii* and *S. hougasii.*

2) A group of accessions from species belonging to series *Conicibaccata* has a jackknife support of 82 in the NJ jackknife tree. In the MP jackknife tree the same accessions are part of the polytomy. These clades represent the subgroups found within the *Conicibaccata* group in the NJ tree. Only one subgroup is not represented by a similar clade in the MP jackknife tree.

3) A group of species belonging to series *Piurana* has a jackknife support of 69 in the NJ tree. In the MP tree, the jackknife support was low, so this group collapsed and 4 out of 5 supported subgroups found in the NJ jackknife tree are visible as supported separate small groups in the MP jackknife tree.

4) A group consisting of accessions from diploid species of series *Conicibaccata*, *S. buesii*, *S. sandemannii* and *S. laxissimum* with jackknife support of 92.

5) A group which contains accessions of *S. medians*, *S. sandemanii*, *S. weberbauerii* and a unknown species with a jackknife support of 85.

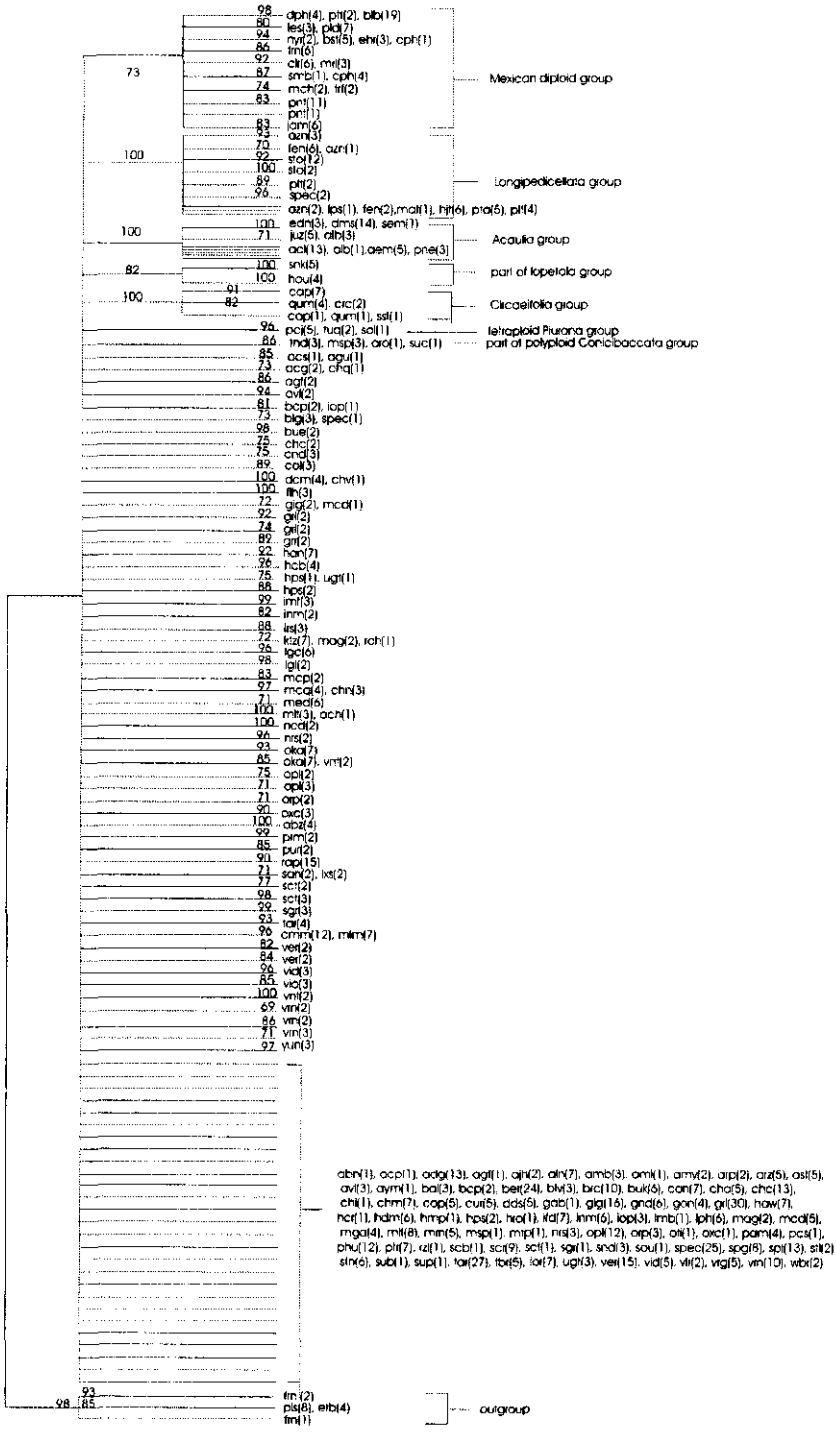



*Figure 2. Maximum Parsimony majority rule consensus tree, condensed dataset, the numbers in the parentheses indicate the number of accessions. The numbers above the branches are jackknife support values.*

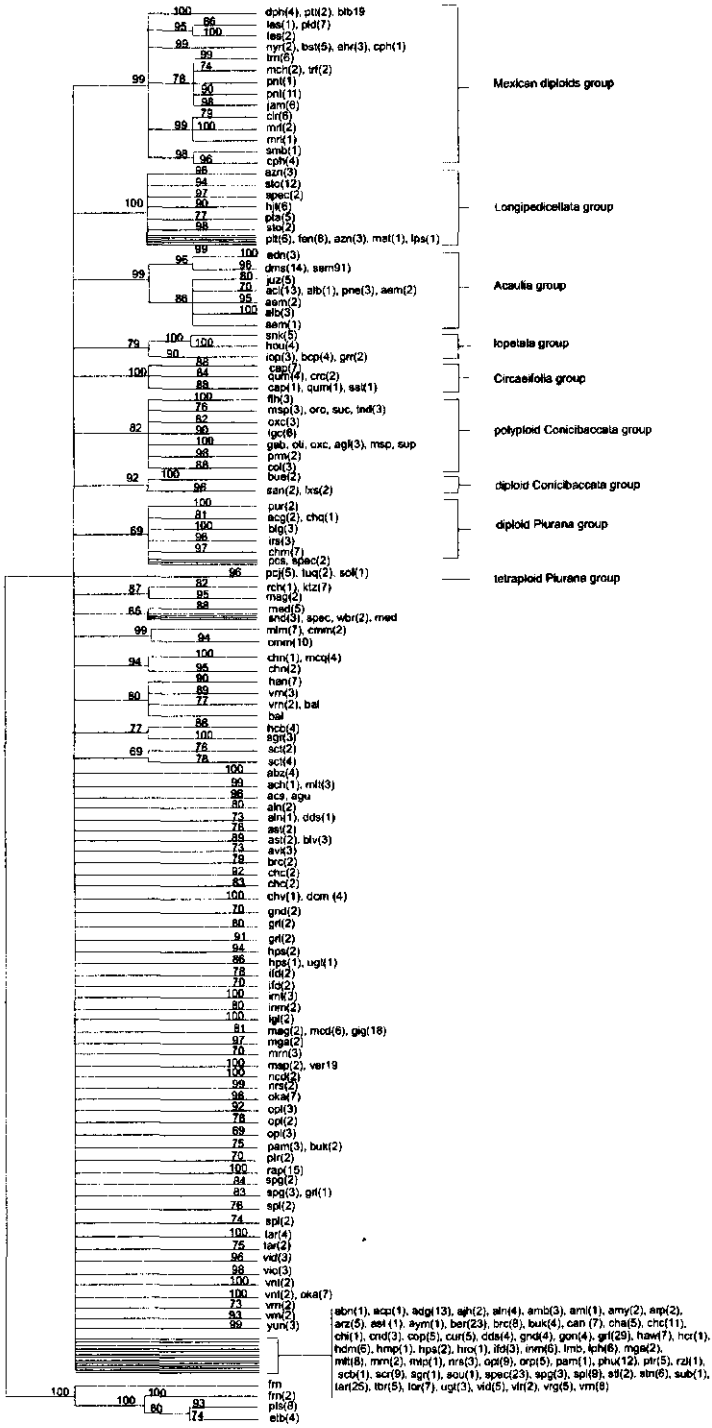



*Figure 3. Neighbour Joining majority rule consensus tree, condensed dataset, the numbers in parentheses indicate the number of accessions. The numbers above the branches are jackknife support values.*

## Discussion

### The value of AFLP

One of the arguments against the use of AFLP is the possible bias caused by homoplasy (Meudt & Clarke 2007; Kardolus *et al* 1998; Koopman & Gort 2004). Non-identical co-migrating bands in the AFLP fingerprints can contribute noise instead of signal to the dataset without being detected. However, it is not likely that in the tuber-bearing wild potatoes homoplasy will cause many problems because the species are all very closely related and homoplasy becomes a problem when distantly related species are involved. Koopman (2005) showed that in a set of closely related *Lactuca* species, sufficient phylogenetic signal was present and concluded that in practice the influence of possible limitations of AFLP, such as co-migration of nonhomologous fragments is limited. However, he stresses that the conclusion only applies to datasets with closely related species. Moreover, Kardolus et al. (1998) concludes from his AFLP results that in *Solanum* section *Petota* the AFLP technique is suitable up to the species level. The AFLP method has since then successfully been used in more studies on potato taxonomy (Van den Berg *et al.* 2002; Lara-Cabrera & Spooner 2004; Mc Gregor *et al.* 2002; Spooner et al. 1992).

### Status of groups within section *Petota*

Not all the groups found in this study have the same level of cohesion or have the same level of demarcation. Some groups have clear borders, while from others we can only vaguely recognize the contours. First, there is a number of groups that are always well supported, whether the analysis is done in a phenetic or phylogenetic way, see Figure 2 and 3. This is the case for the group of Mexican diploid species, the group of Mexican tetraploids, the group of *S. demissum* and *S. acaule*, the group of *S. circaeifolium,* the group of *S. commersonii* and the group of *S. schenckii* and *S. hougasii*. Then there are groups that are not supported in the MP jackknife tree (Figure 2) but that can be found in both the original MP trees and NJ trees (not shown) and are supported in the NJ jackknife tree (Figure 3). This applies to the group with Mexican hexaploid species, the group containing polyploid species belonging to series *Conicibaccata*, the group containing diploid *Piurana* species, and the small groups of *S. huancabambense*, *S. kurtzianum*, *S. medians*, *S. mochiquense*, *S. hannemanii*, *S. buesii*, and *S. paucijugum*.

The largest part of the jackknife trees consists of a polytomy of species that does not seem to contain structure at all. If one was only to consider the structure shown in the jackknife trees, the conclusion would have to be that according to the results of the present AFLP analyses the largest part of section *Petota* is without any taxonomic structure. However, it is possible to identify additional groups that are present in many of the original NJ and MP trees, but do not have enough support to be shown in the jackknife trees. For example, in the 4929 dataset NJ tree a cluster represents the group of cultivated potatoes together with species of series *Tuberosa* from Peru. The groups that are found in both the phenetic and phylogenetic analysis are strong groups with clear borders. The exchange of genetic material is most likely restricted to the members of the group.

The groups with only low support in the MP alone or in both trees are groups that probably share a considerable amount of genetic material with genotypes outside the group. In a study of Jacobs, van den Berg and Vosman (unpublished, but to be submitted) comparison of Chloroplast DNA and AFLP data from *Solanum* section *Petota* reveals incongruencies between the datasets, submitted, the incongruencies found between the chloroplast data and the AFLP data suggest that hybridization occurs between species of different series in section *Petota*. For example, the composition of species of the clade representing the series *Piurana* in the chloroplast tree is different from that of the clade representing the *Piurana* series in the AFLP tree.

The resulting groups also have implications for the theory on EBN of Hawkes and Jackson (1992). EBN stands for Endosperm Balance Number and refers to a hypothetical genetic factor that would explain the success or failure of crosses due to the functioning or breakdown of the endosperm after fertilization. Crosses between species with the same EBN are generally successful and crosses between species with different EBN generally are not, independent of ploidy levels. Hawkes and Jackson (1992) claim that there is a correlation between the EBN hypothesis and the evolution of the group of tuber-bearing *Solanum* species. EBN 1 is found mainly in species that are considered to be close to the ancestors of the group: Mexican series *Morelliformia*, *Bulbocastana*, *Pinnatisecta*, and *Polyadenia*. The EBN 2 condition would have arisen as an isolating mechanism when potato species moved southwards. The EBN 4 condition occurs in hexaploids which are allopolyploids. From the present results it is clear that there is no absolute relationship between EBNs and the groups found. In the group which contains *S. acaule*, *S. demissum*, *S. semidemissum* and *S. edinense*, different ploidy levels and different EBNs occur. This mixture of ploidy and EBN levels also occurs in the group with representatives of series *Conicibaccata*. The species *S. moscopanum* and *S. tundalomense* both are hexaploid and have EBN 4 and they form a group or cluster together with other series *Conicibaccata* species which are known to be tetraploid and have EBN 2. Although these tetraploid and hexaploid species from series *Conicibaccata* are mixed, the diploid series *Conicibaccata* (EBN 2) species do form a separate cluster.

With regard to the overall structure of the section as found in this study two main observations can be made. There seems to be a lack of supported structure, especially in the South American part of section *Petota*. Furthermore, there is a lack of support for the relationships between the different groups that were found in the NJ and MP trees. It is important to differentiate between these two phenomena because the causes underlying both cases could be different.

**Lack of structure in South American part of section Petota**

The AFLP jackknife NJ tree and the jackknife MP tree in this study shows a lack of structure or rather, an unresolved structure for the part of the tree which contains South American species while the other part of the tree shows several well supported groups. Kardolus et al. (1998) mentioned that within series *Tuberosa* different genotypes of the same species are not always grouped together and are scattered among genotypes from other species. He claims that the cause of this phenomenon is not the lack of resolution of AFLP but the overclassification of a group of species, the so-called brevicaule-complex. The cpDNA RFLP studies of Spooner and Sytsma (1992) and Spooner and Castillo (1997) also showed a lack of support for a resolved structure within the group of South American species, and the branch uniting all these species had a bootstrap support value of only 67. Volkov et al. (2003) compared the ETS region of rDNA for 30 species of *Solanum* section *Petota* and found high bootstrap values for the branch uniting all the South American species in three different types of dendrogram (Maximum parsimony, Bayesian statistics and Neigbour Joining). However, the two subgroups within the South American clade that they distinghuished (variants C1 and C2) often show polytomies and resolution within the groups is mostly lacking.

Outside the field of potato taxonomy, researchers have reported similar patterns. Hughes and Eastwood (2006) report a low sequence divergence and lack of resolution in the large Andean clade of the genus *Lupinus*. This would point at a rapid and recent diversification in the Andes. The authors also suggest that *Lupinus* is probably only one example of many plant radiations that followed the final uplift of the Andes. They assume that many of these plant radiations are yet unknown. It is possible that the factors underlying the *Lupinus* diversification are also responsible for the *Solanum* section *Petota* diversification. According to Hughes and Eastwood (2006) these factors would be the large scale of the area over which the radiation extends, repeated fragmentation of high altitude habitats due to quaternary climate fluctuations, the extremely dissected topography, and the habitat heterogeneity.

**Lack of support for relationships between different groups**

Except for the outgroup consisting of *S. etuberosum*, *S. palustre* and *S. fernandezianum* which connects to the main branch of the NJ jackknife and MP jackknife tree with respectively 100 or 98 support value, none of the branches connecting two or more groups have jackknife support of 69 or higher. That is the reason why in the schematized jackknife NJ and jackknife MP trees these branches collapse in a polytomy. Contrastingly, the branches of the groups that can be recognised within the polytomy do have jackknife support, although not all species can be put in groups as discussed previously.

In the first study on the use of AFLP in *Petota* taxonomy by Kardolus et al. (1998), it proved also difficult to find bootstrap support for branches connecting the different groups in section *Petota*. Bootstrap support above 70 were given for a NJ tree branch connecting the outgroup of *S. etuberosum* and *S. brevidens*, for a branch connecting the outgroups, and for the Mexican diploids and *S. circaeifolium* and *S. circaeifolium* subspecies *quimense* with the other part of the tree. In the cpDNA RFLP studies on the South American part of section *Petota (*Spooner & Castillo, 1997*)* only a few branches connecting the larger groups showed bootstrap support above 70. Clade 1, consisting of Mexican diploids (except *S. cardiophyllum* and *S. bulbocastanum*) is connected to the other clades with a bootstrap value of 87, and Clade 3 (mainly accessions belonging to series *Piurana*) and Clade 4 (the rest of section *Petota*) are connected to each other with a branch with 96 bootstrap support.

We can conclude from these previous results that it is indeed difficult to find good support for the backbone structure of section *Petota* in general. This indicates that our and previous results represent the real biological situation in *Solanum* section *Petota*. Since the phylogenetic signal is clearly present in our data as shown in the well-supported groups in the present study, the lack of structure in parts of the tree is not caused by the lack of phylogenetic signal in AFLP markers.

**New informal species groups for Solanum section Petota**

As outlined in this paper and in other earlier studies, there are no results that support the classification of section *Petota* in 21 series. Although a few of the series seem to form natural groups, the majority of the series as proposed by Hawkes (1990) could not be found as separate clusters or clades. Our goal is to use the found structure in the present study at maximum for classifying the section *Petota*.
We propose to divide section *Petota* in informal species groups, following the approach of Spooner et al. (2004) who constructed 11 informal species groups for the North and Central American species. They followed the approach of Whalen (1984) and Knapp (1991; 2000) who applied a similar informal species group classification. We will use the names already used by Spooner et al. (2004) if applicable, and add new groups that were not treated in their study. We chose to base the informal group classification on the groups that are supported in the NJ jackknife tree. The NJ jackknife tree shows more resolution relative to the MP. However, it would not be useful to consider every small group that appears in the schematized tree as a biologically meaningful group. Therefore, the choice for species groups is restricted to groups of species that make sense in the light of former studies and contain at least 3 species. We maintain the species group Verrucosa which contains only one species, because this species group is already designated by Spooner et al (2004).

In total, the NJ jackknife tree can be partitioned into 10 species groups. It would be possible to construct more species groups based on the structure shown in the various trees made in the present study, but these groups would then not be supported by bootstrap or jackknife supports. Although a closed classification following the rules of the Botanical Code is desirable, it seems in this case difficult to apply. In the present study, many species cannot be accommodated in groups. These species do not automatically form a group themselves, but are intentionally left unclassified. We suggest recognizing the following informal species groups as shown in the NJ jackknife tree (Figure 3):

Diploid Mexican group

This group contains the species groups of Spooner et al. (2004): Pinnatisecta, Stenophyllidia, Trifida, Polyadenia , Morelliforme, and Bulbocastana. These species groups can be recognized in the present study as separate branches within the NJ cluster which represents this species group. In the present study we recognize a higher level of group structure which contains all the mentioned species groups, because the detailed contents of each subgroup in our study (Figure 3) differs from the contents from the species groups from Spooner et al. (2004).

Acaulia group

In our study this group contains 2 supported subgroups, one branch with jackknife support of 96 containing the species *S. semidemissum*, *S. demissum* and *S. x edinense*. The other group shows a jackknife support of 98 and contains *S. juzepczukii*, *S. albicans* and the three subspecies *S. acaule* subsp. *acaule* , *S. acaule* subsp. *aemulans*, *S. acaule* subsp. *punae*.

Iopetala group

This group contains the species *S. schenckii*, *S. hougasii*, that form a strongly supported cluster together (jackknife support 100) and a cluster containing the species *S. iopetalum*, *S. brachycarpum*, *S. guerreroense* (jackknife support 90). All species were formerly designated by Hawkes (1990) to series *Demissa* which also included the species *S. demissum* and closely related species. The species in our group are the same as in the species group Iopetala designated by Spooner et al. (2004). They reduced the species *S. brachycarpum* as a synonym of *S. iopetalum*.

Longipedicellata group

As the name does suggest, this group contains species that were formerly placed by Hawkes (1990) in the series of *Longipedicellata*. The species included in this group are *S. fendleri* including *S. fendleri* subsp. *arizonicum*, *S. stoloniferum*, *S. hjertingii*, *S. papita*, *S. polytrichon*, *S. leptosepalum*, *S. matehualae*. The species *S. leptosepalum*, *S. fendleri*, *S. papita*, and *S. polytrichon* have been reduced as synonyms of *S. stoloniferum* (Spooner *et al.* 2004). The species *S. matehualae* is reduced as synonym of *S. hjertingii* (Spooner *et al.* 2004).

Polyploid Conicibaccata group

This group contains species placed there by Spooner et al. (2004), complemented with South American species. The species in this species group are mainly the same as Hawkes (1990) placed in series *Conicibaccata*. According to the present study the group consists of *S. flahaultii*, *S. moscopanum*, *S. orocense*, *S. sucubunense*, *S. tundalomense*, *S. oxycarpum*, *S. longiconicum*, *S. garcia-barrigae*, *S. otites*, *S. oxycarpum*, *S. agrimonifolium*, *S. moscopanum*, *S. subspanduratum*, *S. paramoense*, and *S. colombianum*.

Diploid Conicibaccata group

Although most of the series *Conicibaccata* can be put in the species group Conicibaccata there are a few species that form a separate group. This group consists of the diploid species *S. buesii*, *S. sandemanii*, and *S. laxissimum*.

Diploid Piurana group

This species group was not designated by Spooner et al. (2004). The name refers to the former series *Piurana* as the contents of the group are roughly similar: *S. piurae*, *S. acroglossum*, *S. blanco-galdosii*, *S. irosinum*, *S. chomatophilum*, and *S. paucissectum* from series *Piurana* and *S. chiquidenum* from series *Tuberosa*.

Tetraploid Piurana group

The situation as described before for the Conicibaccata group also applies partly for the Piurana group. There are a few species from the formerly designated *Piurana* series (1990) that form their own species group. This species group contains the tetraploid species *S. paucijugum*, *S. tuquerrense*, and *S. solisii*.

Circaeifolia group

This group consists of *S. circaeifolium*, *S. soestii*, *S. capsicumbaccatum* and *S. circaeifolium* subsp. *quimense*. The contents is conform Hawkes' series *Circaeifolia*.

Verrucosa group

This group contains only 2 species; *S. macropilosum* and *S. verrucosum*. The species *S. macropilosum* was reduced to a synonym of *S. verrucosum* by Spooner et al. (2004)

## Conclusions

As far as we know, this paper treats the largest collection of *Solanum* section *Petota* accessions ever analysed simultaneously. All other previous studies used datasets that included less variation and fewer species. Because of the thorough sampling, it is possible to propose species groups without too many reservations. A number of species groups coincide with certain series recognized by Hawkes (1990). However, most of the series that Hawkes and his predecessors recognized, cannot be supported any longer as natural groups, based on our current knowledge. The present study shows that the taxonomic structure of *Solanum* section *Petota* is highly unbalanced. A few species groups have high support and their inner structure displays also supported subdivisions, while a large part of the species cannot be structured and they seem to be all equally related to each other and to the supported groups.

It might be difficult to accept that a part of genus *Solanum* section *Petota* cannot be structured or subdivided. We even doubt that it would be possible to find more resolution with other methods or more markers, and we consider it likely that the polytomy is indicative of the real situation in section *Petota*. A relatively fast spread of tuber-bearing *Solanum* species over South America, due to the geographic conditions in the Andes (Hughes & Eastwood, 2006), combined with high levels of hybridisation may explain why the phylogenetic links between species are so difficult to establish.

## Acknowledgements

This project was (co)financed by the research programme of the Centre of Biosystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research. We thank SARA Computing and Networking Services in Amsterdam for the use and help with running the analyses on the TERAS computing cluster. The computing time and services at SARA were financed by a grant for pilot projects of the programme Scientific use Supercomputers of the Netherlands Organization for Scientific Research, (NWO). The acronym AFLP is a registered trademark (AFLP®) of Keygene N.V. and the AFLP® technology is covered by patents and patent applications of KeyGene N.V.

*Additional file 1. List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa II* | *S. abancayense* Ochoa | abn | 423 | CGN 18357 | 1 (1) | synonym of *S. bukasovii* (Ochoa, 1999) |
| *Acaulia* | *S. acaule* Bitter | acl | 256, 317, 318, 424, 425, 426, 427, 428, 429, 430, 431, 432, 513*, 681*, 861* | CPC 3768, CGN 17843, CGN 20620, CGN 18179, CGN 18203, cgn962297, CGN 22768, CGN 18068, CGN 22332, CGN 21328, CGN 17930, CGN 22359, CGN 21366*, CGN 20623*, CGN 17924* | 13 (13) | |
| *Acaulia* | *S. acaule subsp. aemulans* (Bitter and Wittm.) Hawkes and Hjert. | aem | 434, 435, 436, 437, 669* | CGN 21303, CGN 21330, CGN 21331, CGN 23789, CGN 20562* | 5 (4) | |
| *Acaulia* | *S. acaule subsp. punae* (Juz.) Hawkes and Hjert. | pne | 441, 442, 443 | CGN 20672, CGN 20665, CGN 20669 | 3 (3) | |
| *Tuberosa III* | *S. achacachense* Cárdenas | ach | 99 | GLKS 32830 | 1(1) | |
| *Piurana* | *S. acroglossum* Juz. | acg | 45, 448 | PI 498204, PI 365313 | 2( 2) | |
| *Tuberosa II* | *S. acroscopicum* Ochoa | acs | 100 | GLKS 32436 | 1 (1) | |
| *Conicibaccata* | *S. agrimonifolium* Rydb. | agf | 101, 319, 450 | GLKS 32292, CGN 18285, CGN 22356 | 3 (3) | |
| *Tuberosa III* | *S. ajanhuiri* Juz. and Bukasov | ajh | 47$, 453, 454 | PI 611096$, CGN 18239, CGN 22389 | 3 (2) | classified as cultivar-group: Ajanhuiri Group within *S. tuberosum* L. (Huaman and Spooner, 2002) |
| *Tuberosa III* | *S. alandiae* Cárdenas | aln | 257, 320, 455, 457, 458, 459, 460 | CPC 7212, CGN 18245, CGN 22349, cgn962384, CGN 20651, CGN 18260, CGN 18264 | 7 (7) | |
| *Acaulia* | *S. albicans* (Ochoa) Ochoa | alb | 301, 461, 462, 464, | CIP 761605, CGN 20667, CGN 20670, CGN 20674 | 4 (4) | |
| *Piurana* | *S. albornozii* Correll | abz | 2, 102, 103, 466 | PI 561637, GLKS 35297, GLKS 35298, CGN 22731 | 4 (4) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa II* | *S. amabile* Vargas | aml | 3 | PI 365356 | 1 (1) | synonym of *S. canasense* (Hawkes, 1990) |
| *Tuberosa II* | *S. amayanum* Ochoa | amy | 302, 303 | CIP 763004, CIP 763005 | 2 (2) | |
| *Tuberosa II* | *S. ambosinum* Ochoa | amb | 104, 105, 467, | GLKS 32282, GLKS 35299, CGN 18358 | 3 (3) | |
| *Tuberosa II* | *S. ancophilum* (Correll) Ochoa | acp | 304 | CIP 761448 | 1 (1) | synonym of *S. rhomboideilanceolatum* Ochoa (Hawkes, 1990) |
| *Megistacroloba* | *S. aracc-papa* Juz. | arp | 109, 110 | GLKS 30082, GLKS 30081 | 2 (2) | nomen dubium (Hawkes, 1990) |
| *Yungasensa* | *S. arnezii* Cárdenas | arz | 4, 111, 112, 113, 471 | PI 545880, GLKS 32832, GLKS 32833, GLKS 32834, GLKS 32831 | 5 (5) | |
| *Yungasensa / Tuberosa* | *S. arnezii* x *S. hondelmanii* | arz x hdm | 401# | CGN 18191# | 1 (0) | |
| *Megistacroloba* | *S. astleyi* Hawkes and Hjert. | ast | 114, 472, 474, 475, 476 | GLKS 32836, CGN 18207, CGN 18210, CGN 18211, CGN 18212 | 5 (5) | |
| *Tuberosa II* | *S. augustii* Ochoa | agu | 305 | CIP 762631 | 1 (1) | |
| *Tuberosa III* | *S. avilesii* Hawkes and Hjert. | avl | 477, 478, 479, | CGN 18255, CGN 18256, CGN 18257 | 3 (3) | |
| *Tuberosa II* | *S. aymaraesense* Ochoa | aym | 5 | PI 607896 | 1 (1) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa III* | *S. berthaultii* Hawkes | ber | 322, 323, 324, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 561*, 939, 940, 941, 943, 944 | CGN 20644, CGN 20650, CGN 18042, CGN 18074, CGN 18190, CGN 20635, CGN 20636, CGN 22715, CGN18216, CGN 22716, CGN 20645, CGN 18246, CGN 23804, CGN 18228, CGN 22727, BGRC 15479, CGN 17823, CGN 18118, GLKS 31670*, CGN 18189, CGN 23508, CGN 18267, CGN 17716, CGN 23477 | 23 (24) | |
| *Tuberosa III / Tuberosa II* | *S. berthaultii x S. sparsipilum* | ber x spl | 402# | CGN 18229# | 1 (0) | |
| *Piurana* | *S. x blanco-galdosii* Ochoa | blg | 48, 115, 996 | PI 442701, GLKS 35309, CIP 761051 | 3 (3) | |
| *Megistacroloba* | *S. boliviense* Dunal | blv | 496, 498, 499 | CGN 18196, CGN 18070, INTA 73228B | 3 (3) | |
| *Pinnatisecta* | *S. brachistotrichium* Bitter (Rydb.) | bst | 116$, 117, 118, 258, 325, 500, | GLKS 32717$, GLKS 32801, GLKS 32714, CPC 3822, CGN 17681, CGN 17603 | 6 (5) | synonym of *S. stenophyllidium* (Spooner et al., 2004) |
| *Demissa* | *S. brachycarpum* Correll | bcp | 259, 321*, 326, 501, 504$ | CPC 7028, CGN 18344*, CGN 17721, GLKS 31686, CGN 18347$ | 4 (4) | synonym of *S. iopetalum* (Spooner et al., 2004) |
| *Tuberosa III* | *S. brevicaule* Bitter | brc | 327, 505, 506, 507, 509, 1020, 1025, 1026, 1040, 1047, | CGN 18231, CGN 17841, CGN 18226, CGN 18232, CGN 22321, CGN 18030, CGN 18223, CGN18247, CGN22322, CGN22717 | 10 (10) | |
| *Tuberosa III / Tuberosa II* | *S. brevicaule x S. sparsipilum* | brc x spl | 403#, 404# | cgn962403#, cgn962443# | 2 (0) | |
| *Conicibaccata* | *S. buesii* Vargas | bue | 7, 8 | PI 568922, PI 607889 | 2 (2) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa II* | *S. bukasovii* Juz. | buk | 328, 511, 512, 514, 955, 971, | CGN 17683, CGN 17684, CGN 17737, CGN 17821, CGN 21305, CGN 17738 | 6 (6) | |
| *Bulbocastana* | *S. bulbocastanum* Dunal | blb | 330, 331, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 945, 946, 947, 948, 949, 950 | CGN 21306, CGN 17693, CGN 17690, CGN 17691, CGN 22698, CGN 17692, cgn960631, CGN18310, CGN17687, GLKS 31741, CGN 21363, CGN 21364, CGN 23075, CGN 17689, CGN 22367, pi275199, CGN 23010, CGN 23074, CGN 22732 | 19 (19) | |
| *Bulbocastana* | *S. bulbocastanum* subsp. *dolichophyllum* (Bitter) Hawkes | dph | 9, 10, 119, 937 | PI 255516, PI 545752, GLKS 35399, CGN 17688 | 4 (4) | synonym of *S. bulbocastanum* (Spooner et al. 2004) |
| *Bulbocastana* | *S. bulbocastanum* subsp. *partitum* (Correll) Hawkes | ptt | 120, 121 | GLKS 35322, GLKS 35323 | 2 (2) | synonym of *S. bulbocastanum* (Spooner et al., 2004) |
| *Tuberosa II* | *S. cajamarquense* Ochoa | cjm | 122$ | GLKS 35328$ | 1 (0) | |
| *Tuberosa II* | *S. canasense* Hawkes | can | 260, 526, 527, 528, 529, 951$, 952, 953 | CPC 2725, cgn960639, CGN 17722, CGN 17672, CGN 17589, CGN 20592$, CGN 18072, CGN23007 | 8 (7) | |
| *Tuberosa III* | *S. candolleanum* P. Berthault | cnd | 530, 531, 532 | PI 498226, CGN 18132, CGN 20603 | 3 (3) | |
| *Canensa, subsect. Basarthrum* | *S. canense* Rydb. | cns | 332&, 333& | CGN 18062&, CGN 18123& | 2 (0) | not from section Petota |
| *Circaeifolia* | *S. capsicibaccatum* Cárdenas | cap | 261, 334, 335, 534, 535, 536, 537, 538, | CPC 3554, CGN 18297, CGN 18268, CGN 18291, CGN 18254, CGN 22388, cgn962457, CGN 18265 | 8 (8) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Cardiophylla* | *S. cardiophyllum* Lindley | cph | 124$, 336, 337, 539, 541, 542 | GLKS 30099$, CGN 18325, CGN 18326, BGRC 55227, CGN 22387, CGN 17697 | 6 (5) | |
| *Yungasensa* | *S. chacoense* Bitter | chc | 125, 126, 127, 246*, 263, 338, 470$, 543, 544, 545, 546, 547, 548, 549, 550, 551 | GLKS 30162, GLKS 30161, GLKS 30180, GLKS 32343*, CPC5901, CGN 18248, CGN 17679$, cgn962709, CGN 18365, CGN 17702, CGN 22384, CGN 18202, CGN 18294, CGN 18338, cgn961764, CGN 22368 | 15 (15) | |
| *Tuberosa II* | *S. chancayense* Ochoa | chn | 1, 552, 553 | VIR 20892, CGN 18036, CGN 18356 | 3 (3) | |
| *Tuberosa II* | *S. chaparense* | chp | 339& | CGN 18060& | 1 (0) | not from section Petota |
| *Tuberosa III* | *S. chaucha* Juz. and Bukasov | cha | 128, 129, 130, 131, 132 | GLKS 30125, GLKS 30115, GLKS 30120, GLKS 30118, GLKS 30119 | 5 (5) | classified as a cultivar group: Chaucha Group within *S. tuberosum* L. (Huaman and Spooner, 2002) |
| *Megistacroloba* | *S. chavinense* Correll | chv | 11 | PI 498235 | 1 (1) | |
| *no information* | *S. chillonanum* Ochoa | chi | 12 | PI 607890 | 1 (1) | |
| *Tuberosa II* | *S. chiquidenum* Ochoa | chq | 49$, 50, 51$ | PI 310989$, PI 473459, PI 310942$ | 3 (1) | |
| *Conicibaccata* | *S. chomatophilum* Bitter | chm | 555, 556, 557, 558, 559, 562, 997 | BGRC 7171, BGRC 18519, CGN 18037, CGN 17712, CGN 20574, CGN 17713, cgn960685 | 7 (7) | |
| *Circaeifolia* | *S. circaeifolium* Bitter | crc | 563, 564 | BGRC 27058, CGN 18133 | 2 (2) | |
| *Circaeifolia* | *S. circaeifolium subsp. quimense* Hawkes and Hjert. | qum | 340, 341, 565, 566, 567, | CGN 18127, CGN 18128, CGN 20643, CGN 22767, CGN 18158 | 5 (5) | |
| *Bulbocastana* | *S. clarum* Correll | clr | 52, 133, 568, 569, 570, | PI 275202, GLKS 32378, | 6 (6) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| | | | 999 | cgn962759, cgn962765, cgn962773, cgn962782 | | |
| *Tuberosa II* | *S. coelestipetalum* Vargas | cop | 134, 135, 306, 307, 572 | GLKS 35433, GLKS 35434, CIP 761755, CIP 761999, CGN 20557 | 5 (5) | |
| *Conicibaccata* | *S. colombianum* Bitter | col | 136, 137$, 573, 574 | GLKS 31536, GLKS 31530$, CGN 18287, CGN 18289 | 4 (3) | |
| *Commersoniana* | *S. commersonii* Dunal | cmm | 265, 575, 576, 577, 578, 1017, 1018, 1019, 1027, 1028, 1039, 1050 | CPC 5861, cgn961592, cgn961597, CGN 18027, CGN 22351, CGN 17988, CGN 18024, CGN 18026, CGN 18327, CGN 18328, GLKS 35340, CGN 23492 | 12 (12) | |
| *Commersoniana* | *S. commersonii* subsp. *malmeanum* (Bitter) | mlm | 139, 266, 579$, 580, 581, 1038, 1045, 1058 | GLKS 35340, CPC 7520, CGN 18329$, CGN 18025, CGN 18215, CGN21353, CGN 22352, cgn962274 | 8 (7) | |
| *Tuberosa III* | *S. curtilobum* Juz. and Bukasov | cur | 140, 141, 142, 143, 267 | GLKS 31620, GLKS 31628, GLKS 35346, GLKS 31627, CPC 7323 | 5 (5) | classified as a cultivar-group: Curtilobum Group of *S. tuberosum* L. (Huaman and Spooner, 2002) |
| *Demissa* | *S. demissum* Lindley | dms | 264*, 299*, 342, 343, 344, 345, 364*, 510*, 582, 584, 585, 586, 587, 748$, 1060 | CPC 5858*, CPC 7069*, CGN 18313, CGN 17820, CGN 20571, CGN 17810, CGN 17829*, cgn960640*, CGN 17797, CGN 17794, CGN 18302, CGN 22378, CGN 20570, CGN 23062$, CGN 17800 | 15 (14) | |
| *Tuberosa III* | *S. x doddsii* Correll | dds | 144, 145, 146, 588, 589, | GLKS 32882, GLKS 32883, GLKS 32880, CGN 20661, CGN 18359 | 5 (5) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa II* | *S. dolichocremastrum* Bitter | dcm | 147, 148, 149, 308, | GLKS 32342, GLKS 35348, GLKS 35349, CIP 762533 | 4 (4) | |
| *Demissa* | *S. x edinense* P. Berthault | edn | 150, 151, 152 | GLKS 25493, GLKS 25492, GLKS 25494 | 3 (3) | |
| *Pinnatisecta* | *S. cardiophyllum* subsp. *ehrenbergii* Bitter | ehr | 153, 154, 155, | GLKS 35331, GLKS 35332, GLKS 32158 | 3 (3) | seen as a separate species *S. ehrenbergii* (Spooner et al., 2004) |
| *Etuberosa* | *S. etuberosum* Lindl. | etb | 591, 593, 594, 595 | CGN 17714, CGN 23066, CGN 18242, CGN 20648 | 4 (4) | not from section *Petota* |
| *Longipedicellata* | *S. fendleri* A. Gray | fen | 156, 157, 158, 159, 160, 596, 597, 598 | GLKS 30425, GLKS 30428, GLKS 30429, GLKS 30433, GLKS 30444, CGN 18116, CGN 18063, CGN 17715 | 8 (8) | synonym of *S. stoloniferum* (Spooner et al., 2004) |
| *Longipedicellata* | *S. fendleri* subsp. *arizonicum* Hawkes | azn | 13, 54, 55, 56, 161, 162 | PI 497996, PI 497999, PI 498000, PI 498001, GLKS 35350, GLKS 35351 | 6 (6) | synonym of *S. stoloniferum* (Spooner et al., 2004) |
| *Etuberosa* | *S. fernandezianum* Phil. | fm | 59, 599, 600, | PI 566756, CGN 18360, CGN 18243 | 3 (3) | not from section *Petota* |
| *Conicibaccata* | *S. flahaultii* Bitter | flh | 14, 15, 61 | PI 570620, PI 583306, PI 583317 | 3 (3) | |
| *Suaveolentia subgenus Basarthrum* | *S. fraxinifolium* | frx | 602& | CGN 21368& | 1 (0) | not from section *Petota* |
| *Tuberosa III* | *S. ganderillasii* Cárdenas | gnd | 16, 62, 163, 270, 346, 603, | PI 597750, PI 597751, GLKS 32423, CPC 7044, CGN 20560, CGN 17590 | 6 (6) | |
| *Conicibaccata* | *S. garcia-barrigae* Ochoa | gab | 17 | PI 498158 | 1 (1) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa III* | *S. gourlayi* Hawkes | grl | 347, 604, 605, 606, 607$, 608, 609, 610, 611, 1000$, 1005, 1006, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1021, 1022, 1029, 1030, 1032, 1033, 1034, 1035, 1036$, 1037, 1042, 1043$, 1044, 1048, 1049, 1051, 1052$, 1053, 1054, 1055$ | CGN 17851, CGN 22705, CGN 17591, CGN 18039, CGN 22380$, cgn961345, CGN 17592, CGN 22336, CGN 21335, cgn961607$, CGN 17872, CGN 17873, CGN 17962, CGN 17963, CGN 17965, CGN 17966, CGN 17967, CGN 17969, CGN 17970, CGN 17971, CGN 18065, CGN 18066, CGN 20585, CGN 20594, CGN 20657, CGN 21332, CGN 21333, CGN 21334, CGN 21336$, CGN 21341, CGN 22340, CGN 22342$, CGN 22343, CGN 23022, CGN 23486, CGN 23497, CGN 23515$, cgn960071, cgn961347, CGN 23794$ | 40 (34) | synonym of *S. leptophyes* (Ochoa, 1990) |
| *Tuberosa III* | *S. gourlayi subsp. pachytrichum x S. leptophyes* | ptr | 612, 613, 614, 615, 616, 617, 618, | cgn18102, cgn18176, bgrc27294, bgrc27295, cgn18188, bgrc7231, bgrc28084 | 7 (7) | synonym of *S. leptophyes* (Ochoa, 1990) |
| *Tuberosa III* | *S. gourlayi* subsp. *pachytrichum x S. leptophyes* | ptr x lph | 409# | BGRC 27293# | 1 (0) | synonym of *S. leptophyes* (Ochoa, 1990) |
| *Tuberosa III* | *S. gourlayi* subsp. *vidaurrei (Cárdenas) Hawkes and hjert.* | vid | 619, 620, 621, 622, 623, 624, 625, 626, | CGN 17848, CGN 17849, CGN 18040, CGN 17850, CGN 18038, CGN 17864, CGN 23024, CGN 23045 | 8 (8) | |
| *Demissa* | *S. guerreroense* Correll | grr | 348, 627 | CGN 18290, GLKS 31513 | 2 (2) | synonym of *S. iopetalum* (Spooner et al., 2004) |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| | *S. hannemanii* | han | 252*, 628, 629, 630, 631, 632, 633 | GLKS 32196*, CGN 17996, CGN 17854, CGN 17997, CGN 20578, CGN 17856, CGN 17858 | 6 (7) | provisional name |
| | *S. hawkesianum* | haw | 166, 167, 634, 635, 636, 637, 638, | GLKS 32762, GLKS 32765, CGN 17888, CGN 17889, CGN 17890, CGN 17891, CGN 17892 | 7 (7) | provisional name |
| *Longipedicellata* | *S. hjertingii* Hawkes | hjt | 349, 350, 639, 640, 641, 643 | CGN 17717, CGN 17718, CGN 22369, CGN 22370, CGN 18345, cgn962453 | 6 (6) | |
| *Tuberosa III* | *S. hondelmannii* Hawkes and Hjert. | hdm | 168, 351, 644, 645, 646, 647, 648 | GLKS 32852, CGN 18106, cgn961918, cgn962199, CGN 18192, CGN 18193, cgn962204 | 7 (7) | |
| *Tuberosa III* | *S. hoopesii* Hawkes and Okada | hps | 169, 650, 651, 652, 653 | GLKS 32885, CGN 18363, CGN 18367, CGN 18368, CGN 18372 | 5 (5) | |
| *Demissa* | *S. hougasii* Correll | hou | 271, 272, 654, 655, | CPC 7050, CPC 2718, CGN 18339, CGN 21361 | 4 (4) | |
| *Yungasensa* | *S. huancabambense* Ochoa | hcb | 18, 170, 353, 354 | PI 365359, GLKS 32441, CGN 18306, CGN 17719 | 4 (4) | |
| *Tuberosa II* | *S. huarochiriense* Ochoa | hro | 309 | CIP 761224 | 1 (1) | |
| *Tuberosa II* | *S. humectophilum* Ochoa | hmp | 171 | GLKS 32829 | 1 (1) | |
| *Piurana* | *S. hypacrarthrum* Bitter | hcr | 311 | CIP 761259 | 1 (1) | |
| *Tuberosa II* | *S. immite* Dunal | imt | 63, 64, 172, | PI 498245, PI 365331, GLKS32819 | 3 (3) | |
| *Tuberosa III* | *S. incamayoense* K.A. Okada and A.M. Clausen | inm | 657, 658, 659, 660, 661, 662, 663, 1016 | CGN 18077, CGN 21320, CGN 17874, CGN 17875, CGN 17968, cgn961363, CGN 22335, CGN 17972 | 8 (8) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Megistacroloba* | *S. infundibuliforme* Phil. | ifd | 664, 665, 666, 667, 668, 1007, 1023 | CGN 17720, CGN 23063, CGN 22334, CGN 23048, cgn960696, CGN 17959, CGN 18079 | 7 (7) | |
| *Demissa* | *S. iopetalum* (Bitter) Hawkes | iop | 273, 670, 671 | CPC 2922, CGN 20561, CGN 20572 | 3 (3) | |
| *Conicibaccata/ Tuberosa II* | *S. irosinum* Ochoa | irs | 66, 312, 313 | PI 583305, CIP 761252, CIP 762259 | 3 (3) | |
| *Pinnatisecta* | *S. jamesii* Torr. | jam | 268*, 274, 355, 672, 673, 674 | CPC 7510*, CPC 7167, CGN 18349, cgn962542, cgn960923, CGN 18346 | 5 (6) | |
| *Tuberosa III* | *S. juzepczukii* Bukasov | juz | 173, 174, 175, 176, 177 | GLKS 25465, GLKS 25467, GLKS 25468, GLKS 25469, GLKS 25470 | 5 (5) | classified as a cultivar-group: Juzepczukii Group within *S. tuberosum* L. (Huaman and Spooner, 2002) |
| *Tuberosa III* | *S. kurtzianum* Bitter and Wittm. | ktz | 275, 276, 675, 676, 677, 678, 995, | CPC 5864, CPC 5889, CGN 22338, cgn961563, CGN 23042, cgn961013, CGN 22353 | 7 (7) | |
| *Conicibaccate* | *S. laxissimum* Bitter | lxs | 178, 679 | GLKS 32439, CGN 22721 | 2 (2) | |
| *Tuberosa II* | *S. leptophyes* Bitter | lph | 356, 357, 680, 682, 683, 684, | CGN 18174, CGN 18140, CGN 18173, CGN 18167, CGN 20611, CGN 18126 | 6 (6) | |
| | *S. leptophyes x S. sparsipilum* | lph x spl | 405# | CGN 20619# | 1 (0) | |
| *Tuberosa I* | *S. leptosepalum* Correll | lps | 19 | PI 607843 | 1 (1) | *S. leptosepalum* is a synonym of *S. stoloniferum* (Spooner et al., 2004) |
| *Polyadenia* | *S. lesteri* Hawkes and Hjert. | les | 20, 21, 358, | PI 558434, PI 558435, CGN 18337 | 3 (3) | |
| *Lignicaulia* | *S. lignicaule* Vargas | lgl | 179, 685 | GLKS 32215, CGN 17723 | 2 (2) | |
| *Conicibaccata* | *S. limbaniense* Ochoa | lmb | 686 | CGN 22720 | 1 (1) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Conicibaccata* | *S. longiconicum* Bitter | lgc | 68, 69, 70, 180, 687, 992 | PI 208780, PI 604093, PI 604095, GLKS 35363, cgn962815, CGN 23561 | 6 (6) | |
| *Juglandifolia* | *S. lycopersicoides* Dunal | lyc | 71&, 72&, 314& | PI 558090&, PI 558091&, CIP 761078& | 3 (0) | not from section *Petota* |
| *Tuberosa I* | *S. macropilosum* Correll | mcp | 23, 74 | PI 607844, PI 607845 | 2 (2) | synonym of *S. verrucosum* (Spooner et al. 2004) |
| *Maglia* | *S. maglia* Schtdl. | mag | 75, 76, 359, 688, | PI 245087, PI 558316, CGN 18064, CGN 22719 | 4 (4) | |
| *Maglia* | *S. maglia x microdontum* | mag x mcd | 406# | CGN 18250# | 1 (0) | |
| *Tuberosa II* | *S. marinasense* Vargas | mrn | 77, 181, 182, 277, 690 | PI 607884, GLKS 35430, GLKS 32281, CPC 7172, CGN 17594 | 5 (5) | |
| *Longipedicellata* | *S. matehulae* Hjert. and T.R. Tarn | mat | 192 | GLKS 35364 | 1 (1) | *synonym of S. hjertingii (Spooner et al., 2004)* |
| *Tuberosa II* | *S. medians* Bitter | med | 183, 691, 692, 693, 694, 695, | GLKS 32226, CGN 21349, CGN 18043, CGN18308, CGN 21343, CGN 18307 | 6 (6) | |
| *Megistacroloba* | *S. megistacrolobum* Bitter | mga | 696, 697, 699, 700 | CGN 23064, CGN 17828, CGN 22347, CGN 20601 | 4 (4) | |
| *Megistacroloba* | *S. megistacrolobum* subsp. *toralapanum* (Cárdenas and Hawkes) | tor | 278, 701, 702, 703, 704, 705, 706 | CPC 1773, CGN 17728, CGN 23006, CGN 18145, CGN 18146, CGN 18147, CGN 18125 | 7 (7) | |
| *Pinnatisecta* | *S. x michoacanum* (Bitter) Rydb. | mch | 185, 279 | GLKS 32346, CPC 3847 | 2 (2) | |
| *Tuberosa III* | *S. microdontum* Bitter | mcd | 360, 707, 708, 958, 959, 994 | CGN 17596, CGN 22382, CGN 18259, CGN 20646, CGN 18047, CGN 20597 | 6 (6) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa III* | *S. microdontum subsp. gigantophyllum* (Bitter) Hawkes and hjert. | gig | 361, 362, 710, 711, 712, 713, 714, 715, 956, 957, 960, 961, 962, 963, 964, 965, 966, 967, | CGN 18046, CGN 18083, CGN 18199, CGN 20639, CGN 18200, CGN 17595, CGN 23050, CGN 21342, CGN 18295, CGN 23511, CGN 20586, CGN 18048, CGN 17597, CGN 18049, CGN 18084, CGN 18003, CGN 18067, CGN 22372 | 18 | synonym of *S. microdontum* Bitter (van den Berg and Spooner, 1992) |
| *Tuberosa I* | *S. minutifoliolum* Correll | min | 24$ | PI 583298$ | 1 (0) | |
| *Tuberosa II* | *S. mochiquense* Ochoa | mcq | 186$, 716, 717, 718, 719 | GLKS 32319$, CGN 20587, CGN 18263, CGN 17731, CGN 21360 | 5 (4) | |
| *Morelliformia* | *S. morelliforme* Bitter and G. Muench | mrl | 78, 79, 187, | PI 619119,PI 545720, GLKS 32245 | 3 (3) | |
| *Conicibaccata* | *S. moscopanum* Hawkes | msp | 25, 81, 188$, 720, 721 | PI 230462, PI 498159, GLKS 35366$, CGN 22355, CGN 18343 | 5 (4) | |
| *Tuberosa II* | *S. multidissectum* Hawkes | mlt | 363, 722, 723, 724, 725, 727, 728, 729, 730, 731, 732 | CGN 17824, CGN 21344, CGN 18330, cgn960739, CGN 17686, CGN 17733, cgn960736, CGN 17825, cgn961613, cgn17840, cgn960967 | 11 (11) | synonym of *S. bukasovii* Juz. f. multidissectum (Hawkes) Ochoa |
| *Tuberosa II* | *S. multiinterruptum* Bitter | mtp | 190 | GLKS 32431 | 1 (1) | |
| *Pinnatisecta* | *S. nayaritense* (Bitter) Rydb. | nyr | 26, 27, | PI 545825, PI 545820 | 2 (2) | synonym of *S. stenophyllidium* (Spooner et al. 2004) |
| *Tuberosa III* | *S. neocardenasii* Hawkes and Hjert. | ncd | 193, 734 | GLKS 32855, CGN 18217 | 2 (2) | |
| *Tuberosa III* | *S. neorossii* Hawkes and Hjert. | nrs | 281, 735, 736, 737, 987* | CPC 6047, CGN 18280, CGN 17599, CGN 18051, CGN | 5 (5) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| | | | | 17763* | | |
| | *S. nigrum* L. | nig | 968& | CGN 21367& | 1 (0) | not from section *Petota* |
| | *S. ochranthum* Dunal | ocr | 194& | GLKS 30919& | 1 (0) | not from section *Petota* |
| *Tuberosa III* | *S. okadae* Hawkes and Hjert. | oka | 283, 365*, 366, 367, 368, 739, 740, 741$, 742, 743, 744, 745, 746, 969, 970 | CPC 7129, CGN 18000*, CGN 18109, CGN 18108, CGN 17998, CGN 18269, CGN 17999, CGN 18279$, cgn962076, cgn962078, CGN 18157, CGN 22709, CGN 18129, CGN 22703, CGN 20599 | 15 (14) | |
| *Tuberosa III* | *S. oplocense* Hawkes | opl | 747, 749, 750, 751, 752, 753, 754, 1001$, 1002, 1003, 1004, 1024, 1031, 1041, 1046, 1056, 1057, | CGN 23049, cgn962217, CGN 21352, CGN 18088, CGN 18085, CGN 21319, CGN 17736$, CGN 17868, CGN 17869, CGN 17870, CGN 18087, CGN 20638, CGN 22324, CGN 22713, CGN 23798, cgn961876, cgn962541 | 18 (17) | |
| *Conicibaccata* | *S. orocense* Ochoa | oro | 28 | PI 583307 | 1 (1) | |
| *Tuberosa II* | *S. orophilum* Correll | orp | 29, 83, 84, 196, 756 | PI 498213, PI 498209, PI 498212, GLKS 35301, cgn962570 | 5 (5) | |
| *Conicibaccata* | *S. otites* Dunal | oti | 30 | PI 570618 | 1 (1) | |
| *Conicibaccata* | *S. oxycarpum* Schiede | oxc | 32, 85, 86, 757 | PI 498026, PI 545776, PI 545779, CGN 18292 | 4 (4) | |
| *Etuberosa* | *S. palustre* Poepp. | pls | 197, 198, 284, 285, 286, 759, 760, 761, | GLKS 35317, GLKS 35319, CPC 7034, CPC 1576, CPC 2451, CGN 18286, CGN 17983, CGN 18241 | 8 (8) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa II* | *S. pampasense* Hawkes | pam | 288, 762, 763, 764 | CPC 6024, CGN 962604, CGN 20575, cgn960051 | 4 (4) | |
| *Longipedicellata* | *S. papita* Rydb. | pta | 369, 370, 642*, 765, 766$, 767, | CGN 17830, CGN 17832, CGN 22385*, CGN18309, CGN 18319$, CGN17831, | 5 (5) | synonym of *S. stoloniferum* (Spooner et al. 2004) |
| *Tuberosa I* | *S. paramoense* Bitter (ex Pittier) | prm | 87, 88 | PI 587114, PI 604202 | 2 (2) | synonym of *S. tuberosum* subsp. *andigena* (Ocoa, 1992) |
| *Conicibaccata* | *S. paucijugum* Bitter | pcj | 33, 89, 90, 199, 768 | PI 561650, PI 583299, PI 561651, GLKS 35372, CGN 18050 | 5 (5) | |
| *Piurana* | *S. paucissectum* Ochoa | pcs | 91, 769$ | pi590922, cgn962622$ | 2 (1) | |
| *Tuberosa III* | *S. phureja* Juz. and Bukasov | phu | 200, 201, 203, 289, 290, 371, 372, 373, 771, 772, 773, 774 | GLKS 31467, GLKS 31468, GLKS 31455, CPC 4188, CPC 4414, CGN 17667, CGN 18301, CGN 18315, CGN 18342, cgn960564, CGN17668, CGN18316 | 12 (12) | classified as cultivar-group: Phureja Group within *S. tuberosum* L. (Huaman and Spooner, 2002) |
| *Pinnatisecta* | *S. pinnatisectum* Dunal | pnt | 204, 231*, 374, 375, 775, 776, 777, 778, 779, 780, 781, 880*, | GLKS 31586, GLKS 32298*, CGN 17745, CGN 17743, CGN 17742, CGN17744, CGN 17740, CGN 23011, CGN 23012, CGN 17741, CGN 18331, CGN 18335* | 10 (12) | |
| *Piurana* | *S. piurae* Bitter | pur | 34, 206 | PI 365365, GLKS 32341 | 2 (2) | |
| *Polyadenia* | *S. polyadenium* Greenm. | pld | 207, 376, 377, 782, 783, 784, 785, | GLKS 35375, CGN 17749, CGN 17746, CGN 17748, CGN 17747, CGN 23013, CGN 23014 | 7 (7) | |
| *Longipedicellata* | *S. polytrichon* Rydb. | plt | 378, 379, 786, 787, 788, 789, | CGN 17750, CGN 17751, cgn962607, CGN 22362, CGN 22361, CGN 18318 | 6 (6) | synonym of *S. stoloniferum* (Spooner et al. 2004) |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Megistacroloba* | *S. raphanifolium* Cárdenas and Hawkes | rap | 208, 209, 291, 380, 790, 791, 792, 793, 794, 797, 798, 799, 800, 801, 976 | GLKS 30637, GLKS 30644, CPC 7090, CGN 17598, cgn960772, CGN 20589, CGN 18300, CGN 18089, cgn961878, CGN 18320, CGN 17752, CGN 18033, CGN 17833, CGN 17835, CGN 17822 | 15 (15) | |
| *Megistacroloba / Tuberosa III* | *S. raphanifolium x S. sparsipilum* | rap x spl | 410# | cgn960058# | 1 (0) | |
| *Tuberosa III* | *S. xrechei* Hawkes and Hjert. | rch | 35 | PI 558227 | 1 (1) | |
| *Tuberosa III* | *S. xrechei x S. microdontum* | rch x mcd | 411# | CGN 20658# | 1 (0) | |
| *Tuberosa III* | *S. xruiz-lealii* Brücher | rzl | 802 | CGN 18117 | 1 (1) | |
| *Pinnatisecta* | *S. xsambucinum* Rydb. | smb | 92 | PI 595478 | 1 (1) | |
| *Megistacroloba* | *S. sanctae-rosae* Hawkes | sct | 803, 804, 805, 806, 807, 1061 | CGN 20576, CGN 22344, CGN 17910, CGN 20564, CGN 17837, cgn961619 | 6 (6) | |
| *Tuberosa II* | *S. sandemanii* Hawkes | snd | 93, 94, 808 | PI 607894, PI 607895, CGN 17600 | 3 (3) | |
| *Conicibaccata* | *S. santolallae* Vargas | san | 36, 809 | PI 607887, CGN 18293 | 2 (2) | |
| *Tuberosa II* | *S. scabrifolium* Ochoa | scb | 37 | PI 365363 | 1 (1) | |
| *Demissa* | *S. schenckii* Bitter | snk | 212, 213, 293, 294, 810, | GLKS 30658, GLKS 30659, CPC 7165, CPC 7164, CGN 18361 | 5 (5) | |
| *Demissa* | *S. xsemidemissum* Juz. | sem | 295 | CPC 7331 | 1 (1) | synonym of *S. demissum* Lindley (Spooner et al., 2004) |
| *Tuberosa III* | *S. xsetulosistylum* Bitter | stl | 214, 811 | GLKS 31014, CGN 20655 | 2 (2) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Juglandifolia sect .Juglandifolium* | *S. sitiens* I. M. Johnst | sit | 38&, 95&, 812& | pi558114&, pi558115&, cgn962632& | 3 (0) | not from section Petota |
| *Circaeifolia* | *S. soestii* Hawkes and Hjert. | sst | 813 | cgn962729 | 1 (1) | |
| *Megistacroloba* | *S. sogarandinum* Ochoa | sgr | 215, 315, 316, 814 | GLKS 35382, CIP 761465, CIP 761586, CGN 17601 | 4 (4) | |
| *Piurana* | *S. solisii* Hawkes | sol | 216 | GLKS 35383 | 1 (1) | |
| *Tuberosa II* | *S. soukupii* Hawkes | sou | 815 | CGN 18061 | 1 (1) | synonym of *S. canasense* (Hawkes 1990) |
| *Tuberosa II* | *S. sparsipilum* (Bitter) Juz. and Bukasov | spl | 382, 383, 384, 816, 817, 818, 819, 820, 821, 972, 973, 975, 978, | CGN 18225, CGN 18230, CGN 18154, CGN 18096, CGN 17838, CGN 18221, CGN 20653, CGN 17758, CGN 20602, CGN 18099, CGN 22702, CGN 18094, CGN 18131 | 13 (13) | |
| | *S. sparsipilum* x *S. leptophyes* | spl x lph | 413# | CGN 18142# | 1 (0) | |
| | *S. sparsipilum* x *S. sucrense* | spl x scr | 414# | cgn960960# | 1 (0) | |
| *Tuberosa III* | *S. spegazzinii* Bitter | spg | 217, 385, 386, 822, 823, 824, 826, 827, 828$ | GLKS 32755, CGN 17759, CGN 17839, cgn960795, CGN 21318, CGN 22707, CGN 21321, CGN 23015, CGN 18034$ | 9 (8) | |
| *Tuberosa III* | *S. stenotomum* Juz. and Bukasov | stn | 218, 219, 296, 387, 388, 829, | GLKS 30762, GLKS 30732, CPC 4741, CGN 18351, CGN 20616, CGN 18161 | 6 (6) | classified as a cultivar-group: Stenotomum Group within *S. tuberosum* L. (Huaman and Spooner, 2002) |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa III* | *S. stenotomum subsp. goniocalyx* (Juz. and Bukasov) Hawkes | gon | 220, 830$, 831, 832, 833$, 834 | GLKS 32703, CGN 17621$, CGN 18314, CGN 18237, CGN 17625$, CGN 17623 | 6 (4) | classified as cultivar-group: Stenotomum Group within *S. tuberosum* L. (Huaman and Spooner, 2002) |
| *Longipedicellata* | *S. stoloniferum* Schtdl. and Bouchet | sto | 221, 297, 298, 389, 390, 554*, 835, 836, 837, 838, 839, 840, 841, 842 | GLKS 30512, CPC 0012, CPC 0028, CGN 17605, CGN 18332, CGN 22718*, PI 205522, CGN 17607, CGN 17606, CGN 18348, cgn962615, CGN 18334, CGN 23072, CGN 18333 | 13 (14) | |
| *Tuberosa III* | *S. xsubandigena* Hawkes | sub | 222 | GLKS 30722 | 1 (1) | synonym of *S .tuberosum* subsp. *andigena* (Ochoa, 1992) |
| *Conicibaccata* | *S. subpanduratum* Ochoa | sup | 223 | GLKS 22873 | 1 (1) | |
| *Tuberosa III* | *S. xsucrense* Hawkes | scr | 391, 843, 844, 845, 846, 847, 848, 849, 850 | CGN 18205, CGN 20628, CGN 20630, CGN 20631, CGN 18187, CGN 20634, CGN 22350, CGN 18206, CGN 18105 | 9 (9) | |
| | *S. sucrense x S. oplocense* | scr x opl | 412# | cgn962507# | 1 (0) | |
| *Conicibaccata* | *S. sucubunense* Ochoa | suc | 39 | PI 583320 | 1 (1) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Tuberosa III* | *S. tarijense* Hawkes | tar | 224, 225, 280*, 392, 852, 853, 854, 855, 856, 857, 858, 859, 860, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, | GLKS 31570, GLKS 31572, CPC 7208*, CGN 17861, CGN 22729, cgn962224, CGN 22714, CGN 18198, cgn960807, cgn960805, cgn960806, CGN 17975, cgn961432, CGN 21337, CGN 23795, cgn961736, CGN 17976, CGN 17974, CGN 17977, CGN 18107, cgn961128, CGN 17978, CGN 17979, cgn961441, CGN 17980, CGN 21338, CGN 17981, cgn961449 , CGN 17982, cgn961451, cgn962690 | 31 (31) | |
| | *S. tarijense x S. amezii* | tar x arz | 418# | cgn962468# | 1 (0) | |
| | *S. tarijense x S. microdontum* | tar x mcd | 419# | cgn960084# | 1 (0) | |
| *Pinnatisecta* | *S. tarnii* Hawkes and Hjert. | trn | 40, 41, 226, 227, 228, 229 | PI 570642, PI 498048, GLKS 32870, GLKS 35384, GLKS 35385, GLKS 32871 | 6 (6) | |
| *Pinnatisecta* | *S. trifidum* Correll | trf | 881, 882 | CGN 22722, CGN 22371 | 2 (2) | |
| *Tuberosa III* | *S. tuberosum* L. | tbr | 232, 233, 234, 883, 884 | GLKS 22827, GLKS 31211, GLKS 31114, CGN 22877, cgn962368 | 5 (5) | |
| *Tuberosa III* | *S. tuberosum subsp. andigena* (Juz. and Bukasov) Hawkes | adg | 235, 237, 238, 239, 240, 241, 242, 243, 885, 886, 887, 888, 889 | GLKS 35027, GLKS 34963, GLKS 34867, GLKS 34149, GLKS 34509, GLKS 34523, GLKS 34630, GLKS34737, cgn960305, CGN 20610, CGN 20614, CGN 23055, cgn960441 | 13 (13) | classified as a cultivar-group: Andigenum Group within *S. tuberosum* L. (Huaman and Spooner, 2002) |
| *Tuberosa II* | *S. tuberosum* subsp. *andigena x S. curtilobum* | adg x cur | 395#, 396#, 397#, 398#, 400# | CGN 18144#, CGN 18149#, CGN 18134#, cgn962564#, | 5 (0) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| | | | | CGN 21355# | | |
| *Conicibaccata* | *S. tundalomense* Ochoa | tnd | 244, 245, 247 | GLKS 35388, GLKS 35389, GLKS 35390 | 3 (3) | |
| *Piurana* | *S. tuquerrense* Hawkes | tuq | 96, 890, | PI 590926, CGN 18353 | 2 (2) | |
| *Tuberosa III* | *S. ugentii* Hawkes and K. A. Okada | ugt | 44, 248, 249, 892, | PI 546029, GLKS 32887, GLKS 32889, CGN 18369 | 4 (4) | |
| *Tuberosa II* | *S. velardei* Ochoa | vlr | 97, 893 | PI 619114, CGN 18324 | 2 (2) | |
| *Tuberosa III* | *S. venturii* Hawkes and Hjert. | vnt | 250, 894, 896, 993, | GLKS 32794, CGN 17761, cgn961508, CGN 17755 | 4 (4) | |
| *Tuberosa III* | *S. vernei* Bitter and Wittm. | vrn | 895*, 897, 898, 899, 00, 901, 902, 903, 904, 05$, 979, 980, 981, 982, 83, 984, 985, 986 | CGN 17762*, CGN 22728, CGN 18111, CGN 21350, CGN 22345, CGN 18112, CGN 18114, CGN 23039, CGN 18278, CGN 17836$, CGN 18110, CGN 21315, CGN 17995, CGN 18113, CGN 18115, CGN 23516, CGN 18277, cgn963094 | 17 (17) | |
| *Tuberosa III* | *S. vernei* subsp. *ballsii* (Hawkes) Hawkes and Hjert. | bal | 906, 907, 908 | CGN 17992, CGN 17993, CGN 17994 | 3 (3) | |
| *Tuberosa I* | *S. verrucosum* Schtdl. | ver | 393, 825*, 909, 910, 911, 912, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 988, 989, 990 | CGN 17768, CGN 18100*, CGN 22326, CGN 22374, CGN 17764, CGN 20567, CGN 17769, CGN 17765, CGN 17773, CGN 17771, CGN 17766, CGN 17770, CGN 17772, cgn960832, cgn960833, CGN 20566, CGN 23017, CGN 17767, CGN 17774 | 19 (19) | |

*Additional file 1. (Continued) List of material used in the AFLP analysis.*

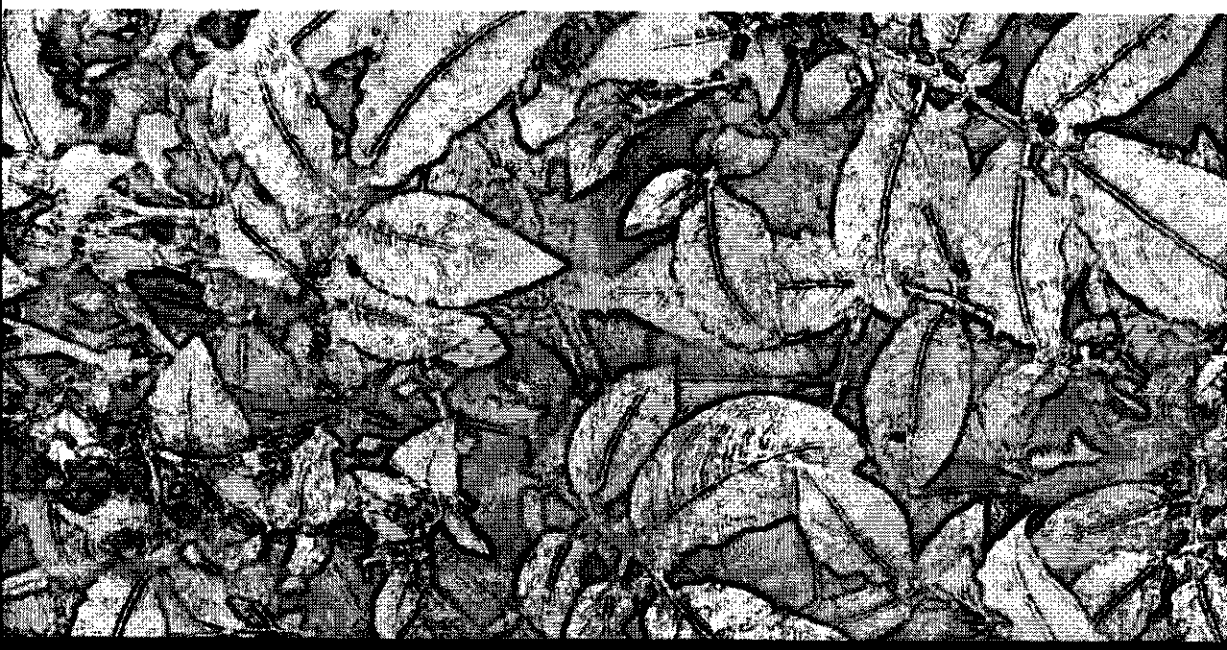
| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 4929 dataset (total nr. of accessions in 916 dataset) | Taxonomic remarks |
|---|---|---|---|---|---|---|
| *Conicibaccata* | *S. violaceimarmoratum* Bitter | vio | 924, 925, 926, | CGN 18296, CGN 20647, CGN 22878 | 3 (3) | |
| *Conicibaccata / Yungasensa* | *S. violaceimarmoratum x S. yungasense* | vio x yun | 420#, 421# | cgn961955#, CGN 18124# | 2 (0) | |
| *Tuberosa III* | *S. virgultorum* (Bitter) Cárdenas and Hawkes | vrg | 927, 928, 929, 930, 931, 932$ | cgn962448, CGN 17775, cgn962072, CGN 20615, cgn962077, CGN 20652$ | 6 (5) | |
| *Tuberosa II* | *S. weberbaueri* | wbr | 254, 300 | GLKS 32725, CPC 6032 | 2 (2) | |
| | *S. weberbaueri x S. yungasense* | wbr x yun | 422# | CGN 20656# | 1 (0) | |
| *Yungasensa* | *S. yungasense* Hawkes | yun | 98, 934$, 935, 936 | PI 614703, CGN 18336$, CGN 20677, CGN 20676 | 4 (3) | |
| | unknown species | spec | 6*, 43*, 123*, 165*, 184*, 205* , 210*, 211*, 253*, 255*, 262*, 287*, 292*, 310*, 329*, 352*, 381*, 394*, 416, 533*, 601*, 649*, 726*, 796*, 891*, 933*, 974*, 998* | PI 498214*, PI 546033*, GLKS 32175*, GLKS 31512*, GLKS 35422*, GLKS 22340*, GLKS 32808*, GLKS 32809*, GLKS 32722*, GLKS 32172*, CPC 7211*, CPC 7328*, CPC 7105*, CIP 761265*, CGN 18249*, CGN 18182*, CGN 17753*, CGN 18121, CGN 20580, CGN 18137, cgn962595, CGN 18053, CGN 17685, CGN 17754, CGN 18364, CGN 18262, CGN 18052, cgn962734 | 28 (28) | |

**Symbols used in Additional file 1**

| | |
|---|---|
| # | recorded hybrid, removed in 916 dataset |
| $ | complete accession removed in the 916 dataset because of conflicting positions in NJ tree |
| & | removed outgroups in 916 dataset: S. lycopersicoides, S. nigrum, S. chaparense, S. sitiens, S. canense, S. fraxinifolium. |
| * | the label of this accession was changed in the 916 dataset after checking the position in the large NJ tree and checking morphology in the greenhouse/field |
| ( ) | the number in parentheses indicates the number of accessions used for the 916 analysis in case of removal or change of accessions |

**Abbreviations for Genebank source codes:**

CPC: Commonwealth Potato Collection, UK
CGN: Centre for Genetic Resources, the Netherlands
cgn: cgn receipt number, Centre for Genetic Resources, the Netherlands
PI: Plant Introduction number, USA
GLKS: Gross Lusewitz, Germany



# CHAPTER 5

## What's in a name? Validation of species labels in *Solanum* section *Petota*

Mirjam M.J. Jacobs, Marinus J.M. Smulders, Ronald G. van den Berg and Ben Vosman

## Abstract

The taxonomy of wild potato species, belonging to section *Petota* of the genus *Solanum*, is known to be problematic. The systematic relationships among species, as expressed in the arrangement of the 19 series as designated by Hawkes and others, are complicated and the group of wild species belonging to *Solanum* section *Petota* seems overclassified. Because many of the 90 presumed species are very similar and are known to exchange genetic material we chose to initially treat them as populations and look for sufficient support for any grouping within the section. A dataset of 566 South-American accessions was analyzed with the program STRUCTURE 2.2 in an 'unsupervised' procedure based only on genetic similarities, assigning individual accessions to inferred clusters based on genetic similarity, rather than taxonomic label. STRUCTURE results showed that at best the section could be arranged in 16 clusters of various size and composition. Within the clusters further subdivision was determined based on maximizing genetic diversity among groups (Fst values) for all available accessions of the species present, testing various arrangements within the separate clusters. The latter analysis included as many as 2767 genotypes. Overall, for 8 species labels support was found for preserving the species status, and for 10 species labels plus five 2-species combinations weak support was found. No support was found for the remaining 43 species labels (and 19 species labels were only represented by only one accession). Some of these species labels occurred in two clusters or in two groups within clusters, which may be indicative of cases of species hybridization. Many of the species labels were distributed across more than one cluster and/or group within clusters, which may indicate misclassifications. Furthermore some species labels appeared only in fixed combinations with another species label without displaying any differentiation between them, which are clear examples of overclassification. Thus, the methodology used here enabled us to estimate the number of supported groups with the section, which turns out to be well below the number of species postulated, and provides a method to distinguish between species labels with and without molecular genetic support. The substructure found within the clusters should not automatically be regarded as taxa. To define or reject species a dataset such as obtained here should be combined with data from morphological surveys, with geographical distribution data, and with information from crossing experiments.

## Introduction

The taxonomy of wild potato species, belonging to section *Petota* of the genus *Solanum*, is known to be problematic (Hawkes 1990; Spooner & Salas 2006; Van den Berg & Jacobs 2007). Identification of species in wild potato material remains difficult and the systematic relationships among the potato species is still unclear. One of the causes for these difficulties is the ability of many species to hybridize easily with other species (Spooner & Salas, 2006). Hawkes (1990) hypothesized that approximately 12% of the 224 tuber-bearing *Solanum* species he recognized, had arisen by hybrid speciation. Second, there is a large amount of phenotypic plasticity, i.e., plants look different in different environments (Spooner & Hijmans, 2001; Spooner *et al.*, 2004). Additionally, taxonomists have had the tendency to classify all the variability in characters they observe in the group of wild potatoes. As a consequence, species boundaries may be based on distinctive morphological characters that are not expressed under all conditions. Hence, numerous species have been described, many of which are extremely similar to each other, and Spooner and Salas (2006) and van den Berg and Jacobs (2007) concluded that the group of wild species belonging to *Solanum* section *Petota* appears overclassified. An example of overclassification within *Solanum* section *Petota* is the so-called Brevicaule complex. Morphological data studied by Van den Berg et al. (1998) failed to distinguish the 30 species in the Brevicaule complex. Molecular results of Miller and Spooner (1999) showed that the Brevicaule complex is paraphyletic and that many taxa should be relegated to synonymy.

Furthermore, also the systematic relationships among species, as expressed in the arrangement of the 19 series, as designated by Hawkes (1990) and others, is hard to determine. Some of the series are difficult to keep apart and other series contain subgroups that could be considered a separate series (van den Berg & Jacobs 2007). To date, the series classification of Hawkes and other previous authors has received only partial cladistic support (Spooner *et al.* 2004). In a previous paper, Jacobs et al. (2008) described the taxonomic structure present in *Solanum* section *Petota* and focused on testing the validity of the series classification and studying the taxonomic structure of the section. The largest dataset ever constructed for *Solanum* section *Petota* was analyzed in a phylogenetic and phenetic manner. Although some of the branches in the resulting trees were supported by jackknife values above 69, both (phenetic and phylogenetic) trees also display a large polytomy containing taxa that seem to be all equally related to each other and to the supported groups.

In the present study, we focus on the overclassification of species. As stressed before, many described species in section *Petota* are extremely similar to each other and in many cases, potato species can only be distinguished by means of combining often minor characters with overlapping character states (Spooner & Van den Berg, 1992a). The number of species in the whole of *Solanum* section *Petota* has already been reduced somewhat due to the application of molecular techniques in potato taxonomy. While Hawkes (1990) still recognized 227 tuber-bearing species (of which 7 were cultivated species) and 9 non tuber-bearing species within section *Petota*, Spooner and Hijmans (2001) recognized only 203 tuber-bearing species, including 7 cultivated species, while Spooner and Salas (2006) reduced the number further to 189 species (including only 1 cultivated species).

In their recent review on section *Petota* taxonomy, Spooner and Salas (2006) speculate on taxonomic changes for several species.

Phylogenetic and phenetic results of our previous study and from others (Hawkes, 1990; Jacobs *et al.*, 2008; Spooner & Salas, 2006; van den Berg & Jacobs, 2007) revealed that many wild *Solanum* species, especially the species that were designated as belonging to the series *Tuberosa*, *Megistacroloba*, and *Yungasensa*, are very closely related. Spooner and van den Berg (1992a) stated already that section *Petota* has many phenetically distinct groups of taxa, generally labeled as species until now, that have the ability to freely exchange genes under artificial conditions and produce advanced generation hybrids. When the line separating populations from species is blurred, it could be more fruitful to consider the individual plants as belonging to one gene pool, rather than to isolated taxa, unless sufficient support can be found for these. We therefore employed a population genetics approach to detect inner structure in the South American polytomy of *Solanum* section *Petota*.

Focusing on species delimitation using a population genetics approach with AFLP markers, we take the accessions as a starting point. To test which accessions may comprise one or more genetic units (or species) we used a Bayesian population clustering approach implemented in the program STRUCTURE 2.2 (Falush *et al.*, 2003; Pritchard *et al.*, 2000a; Pritchard *et al.*, 2000b). STRUCTURE clusters individuals without using a-priori information from their population of origin. The primary assumptions of the model used in STRUCTURE are Hardy-Weinberg equilibrium within populations (or metapopulations) and linkage equilibrium among loci. It attempts to find population groupings that are not in disequilibrium (Pritchard *et al.*, 2000a). Using a Markov chain Monte Carlo (MCMC) algorithm the program assigns individuals to populations and estimates population allele frequencies. The program has been successfully used in many population genetic studies, for example in the research of genetic structure in human population (Rosenberg *et al.*, 2002), in the phylogeography of the sand-dune shrub *America pungens* (Pineiro *et al.*, 2007) and for distinguishing chicken breeds (Rosenberg *et al.*, 2001). Recently, STRUCTURE was also used in studies on phylogenetic relationships among species in the genus *Betula* (Schenk *et al.*, 2008) and on species delimitation in a recent species radiation in turtles (Shaffer & Thomson, 2007).

Accessions within one species are expected to share more genetic material with each other than with accessions from outside the species. As a result, genetic differentiation among species is expected to be higher than within species. Consequently, if we use correct species labels to subdivide an unstructured set of accessions this will lead to an increase of the genetic variation among groups, but if the species labels are incorrect this will not happen, or to a lesser extent. The genetic differentiation among groups (Fst) allowed us to determine which species labels within the observed STRUCTURE groups actually contributed to increased differentiation among labels, and therefore can be considered to have (some) support. This approach of the species delimitation resembles somewhat the view of Shaffer and Thompson (2007) that follows Mayden (1997) and de Queiroz (1998), in that they consider species as segments of evolutionary lineages.

Under this view, species delimitation comes down to identification of metapopulation lineages. The metapopulation lineage species definition leads to operational species delimitation approaches that recognize sets of populations that freely exchange genes in nature but have no or very restricted gene exchange with other sets of populations (Shaffer & Thomson, 2007).

## Material and Methods

### Plant Material

We used the same plant material from genus *Solanum* section *Petota* as described in Jacobs et al. (2008). The plants were grown, young leaf material was harvested and DNA was extracted as described in Jacobs et al. (2008). In total 196 different taxa were sampled. At least 5 accessions from each available species and 5 individual plants per accession were used (in total 4929 genotypes). A condensed *Petota* dataset (Jacobs *et al.*, 2008) was created by choosing a representative genotype from all the accessions available in the original dataset of 4929 genotypes. The nomenclature of the plant material also followed the decisions as made and explained in Jacobs et al. (2008). That means that in some cases we have retained the original labels, even if taxonomic references suggested a change of the species name, but obvious mistakes (mislabeling) have been corrected after preliminary AFLP analyses.

### AFLP

The protocol of Vos et al. (1995) was used to generate AFLP fragments. The plant material was fingerprinted with two EcoRI/MseI AFLP primer combinations: E32/M49 and E35/M48. These primer combinations gave 91 and 131 polymorphic bands, respectively. The AFLP analysis was done on a MegaBACE 2.1 by Keygene. Bands were scored as dominant markers, using the Keygene proprietary software.

### Datasets

For the STRUCTURE analysis of the South-American *Solanum* accessions a dataset was constructed containing 566 samples, representing 90 species/subspecies (information on the accession numbers and geographic origin in Additional file 1) This 566 South-American accessions dataset was a subset of the 916 accessions dataset (which is a condensed dataset of the original 4929 dataset, Jacobs et al (2008), and contained all the accessions of species collected in South America that appear in the large polytomy in Jacobs et al. (2008) and that do not belong to those species groups with high jackknife support: the Acaulia group, Mexican diploid group, diploid Piurana group, tetraploid Piurana group, polyploid Conicibaccata group, diploid Conicibaccata group, Circaeifolia group, Longipedicellata group, and Iopetala group.

The results from one of the 10 STRUCTURE runs on the 566 South-American accessions dataset at K=16 (see Results) with the highest probability (ln P(D)=-41181.7) was used to define subsets for population genetic statistics using AFLP-SURV.

For the AFLP-SURV analysis we used all the available genotypes from the 566 accessions retrieved from the original dataset of 4929 samples (in total 2767 genotypes). Due to technical restrictions it was not possible to use the similar number of genotypes for the STRUCTURE analyses, so therefore these analyses were done with only one representative genotype per accession.

**Data analysis**

*Bayesian clustering*

The 916 accessions condensed dataset and the 566 South-American accessions dataset were analyzed with STRUCTURE 2.2 (Falush *et al.*, 2003; Pritchard *et al.*, 2000a; Pritchard *et al.*, 2000b) in order to test if the species in the datasets form separate clusters or species groups (populations according to STRUCTURE) in an 'unsupervised' procedure (Rosenberg, 2004) based only on genetic similarities, and to assign individual accessions to these groups based on genetic similarity, rather than taxonomic label.

To test whether STRUCTURE was suitable for analyzing the *Solanum* AFLP data a pilot analysis was done on a condensed dataset of 916 samples. Almost all species groups as defined by Jacobs et al. (2008) and smaller supported branches in the NJ tree have their own cluster at K=18 (results not shown). These results confirmed that STRUCTURE can be used for the AFLP dataset. While STRUCTURE was designed for studies on populations, in which individual samples are assumed to be able to exchange genetic material, apparently it can also be used to distinguish accessions from different species that do not exchange genetic material any more. Within *Solanum* section *Petota* many species are still able to hybridize and exchange genes.

We used the approach of coding the dominant markers as described by Falush et al. (2007). The dominant AFLP data were entered by coding both alleles as '1' when the AFLP band was present and both as '0' when the band was absent. We specified '0' as the recessive allele for all the AFLP data. This enables the simultaneous analysis of accessions with different levels of ploidy (Schenk *et al.*, 2008). Evanno et al (2005) showed that results from AFLP with STRUCTURE can be as accurate as with microsatellites. Estimates for the log likelihood were obtained using the admixture model and the assumption that the allele frequencies are correlated. The log likelihood estimates were obtained for 10 replicate runs at each K ranging from K=1 to K=30. For each run, we used a burn-in of 25,000 cycles and a data run of 100,000 cycles.

*Partitioning of genetic variation within and among groups*

The partitioning of genetic variation within and among (i) STRUCTURE clusters of accessions, and (ii) preexisting species labels within these clusters were computed using AFLP-SURV 1.0 (Vekemans *et al.*, 2002). The allelic frequencies at AFLP loci were calculated from the observed frequencies of fragments, using the Bayesian approach by Zhivotovsky (1999) for all the species (assuming diploid species and Hardy-Weinberg equilibrium). We assumed a uniform prior distribution of allelic frequencies.

In order to test whether the clusters found with the software STRUCTURE are genetically differentiated from each other, we computed the proportion of genetic variation among the clusters and among all pairs of clusters (overall and pairwise Fst). We compared the Fst of the 16 clusters with the Fst computed for STRUCTURE clusters at K=10 (a suboptimal population subdivision), with a division based on the original species labels, and with a division based on all 566 separate accessions, to compare the effect of these alternative subdivisions on the partitioning of genetic variation. Significance of the Fst values was tested by 1000 permutations. The confidence limits obtained were used to determine significance of differences between these separate estimates.

Within the 16 STRUCTURE clusters Fst was calculated based on using the preexisting species labels, but with the inclusion of all accessions available for these species labels. The Fst value was compared to that based on all individual accessions. As the contribution to the partitioning of genetic variation could differ among the various species within a cluster (some may form a homogeneous group of genotypes while others in the same cluster may be highly variable and rather resemble a random selection of accessions), the species showing a pairwise Fst of less than the observed overall Fst were merged into one group, and the new species group was included in a new AFLP-SURV analysis. The value of the Fst of this new partitioning was compared to the Fst of the previous species partitioning. The process was repeated for each STRUCTURE cluster, merging species and species groups, until the highest value of the overall Fst was reached for the cluster. Table 2 shows the results of the Fst analysis with AFLP-SURV for each step in each cluster.

## Results

### Clustering of the 566 South-American accessions dataset

The 566 South-American samples dataset was analyzed for K=1 to K=30. Figure 1 shows the average posterior probability Ln(P(D)) for 10 runs as a function of the number of populations K. The largest increase in the posterior probability of the data is found at K=2, but this has no biological meaning. In the runs with K higher than 2 the posterior probability still increases and around K=18, the values seems to decrease slightly (Figure 1). Furthermore, the posterior probability in runs with K=17 or higher became highly variable among runs, and the resulting clustering of accessions became unstable. Contrastingly, at K=16 the clustering results were stable and most clusters had the same composition in all 10 replicate runs. We therefore took K=16 as the optimal K.

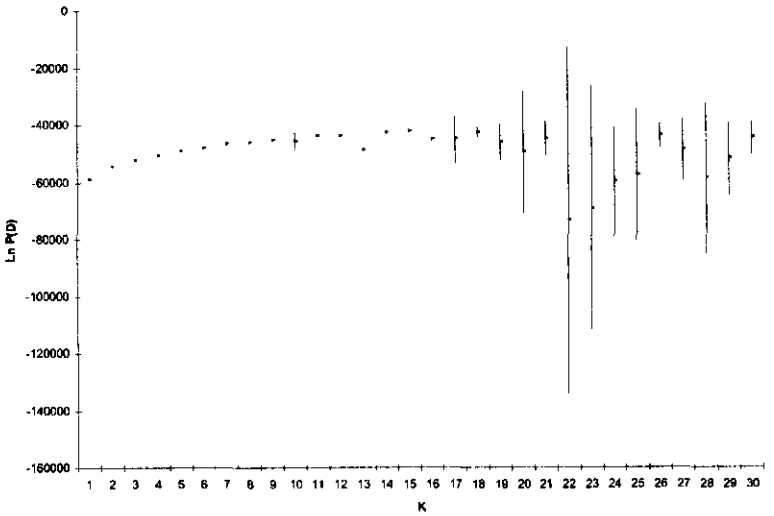



*Figure 1. Mean L(K) ±SD for 10 replicate runs at each level of K proposed clusters.*

The estimated population structure at K=16 is shown in Figure 2. Each individual accession is presented by a thin vertical line, and this line shows colored segments that represent the relative percentage of membership to the K clusters. The individuals are arranged according to their species labels. Some species labels, *S. okadae*, *S. raphanifolium*, *S. verrucosum*, and *S. macropilosum* occupy exclusively one cluster, while many other species labels share a cluster with one or more other species labels, for instance *S. huancabambense* with *S. sogarandinum*. Strikingly, accessions of many species labels appear as members of multiple clusters, like the species labels *S. maglia*, *S. gourlayi*, *S. tarijense* and many others. Finally, there are many individuals that show partial membership to multiple clusters.

In the Additional file 1 details are found on the composition of the clusters and the percentage of memberships per individual accession for these clusters, in the run with the highest probability for K=16 (Ln P(D) =-41181.7). Most clusters defined by STRUCTURE for K=16 are the same in all 10 runs. One of the exceptions is cluster 3, it was found in 3 out of 10 runs as a separate unit. In 7 out of 10 runs it is combined with the accessions of cluster 4.

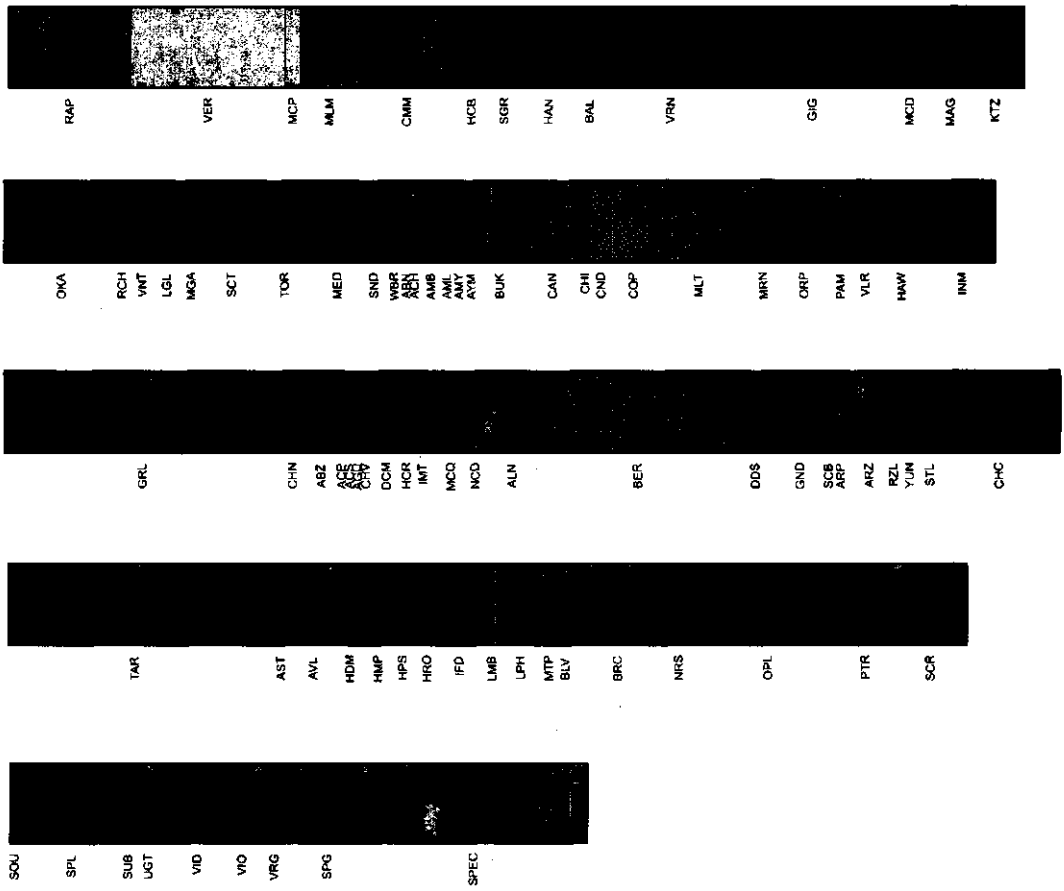



*Figure 2. Estimated population structure for K=16. Each accession is represented by a thin line, which is partitioned in K colored segments that represent the membership to K clusters. The labels below indicate the species labels.*

**Genetic differentiation in the 566 accessions dataset (among accessions, species, and groups)**
In order to test whether the clusters found with STRUCTURE are significantly different gene pools, we computed the Fst among the 16 clusters. We compared the Fst of the 16 cluster arrangement to the Fst of the individual accessions, that of the original 90 species labels, and that of a 10 cluster arrangement. The results are shown in Table 1. The Fst of the 16 cluster arrangement is the highest, representing 31% of the genetic variation among the clusters. The 90 pre-existing species labels explain 29% of the existing genetic variation, but a subdivision in 10 groups already explains 27%. All the Fst values are significantly different from each other. The 556 individual accession arrangement shows the lowest value of Fst, as only 15% of the genetic variation is present among accessions. Thus, there clearly is substructure among these 556 accessions, but there appears to be no support for more than in the order of 16 clusters. The decrease in Fst when moving upwards towards 90 species labels may indicate that at least some of the designated labels are superfluous or incorrect, and thereby create sets that are genetically heterogeneous. This is also confirmed by the results of the STRUCTURE analysis, in which many accessions belonging to the same species labels were placed in different cluster groups.

*Table 1. Genetic differentiation in complete dataset (*28 accessions labeled "unknown species" were excluded in this analysis).*

| | n | Ht | Hw | Hb | Fst | p-value |
|---|---|---|---|---|---|---|
| among the accessions | 538 | 0.3256 | 0.2783 | 0.0473 | 0.1453 | <0.001 |
| among the old species labels | 90 | 0.2632 | 0.1855 | 0.0777 | 0.2953 | <0.001 |
| among the clusters at k=16 | 16 | 0.2077 | 0.1430 | 0.0647 | 0.3124 | <0.001 |
| among the clusters at k=10 | 10 | 0.2023 | 0.1475 | 0.0548 | 0.2733 | <0.001 |

**Genetic differentiation within the STRUCTURE clusters**

As expected, the level of genetic differentiation among the accessions is lower within the clusters (Table 2). The lowest values are for cluster 1, 6 and 15, which mainly or exclusively consist of accessions of only one species label, e.g. cluster 15, which contains only *S. okadae* accessions, has an Fst of 0.0029. The negative values of Fst among the accessions for cluster 1 and 2 are probably caused by missing data for certain loci. Values near zero are consistent with groups that consist entirely of accessions of one homogeneous biological species. Genetic differentiation among species within those clusters that contain accessions from two species labels ranged from 9.8% in cluster 4 to 27.8% in cluster 7. In cluster 4, cluster 10, and cluster 12 the species arrangement only added a small contribution to the genetic differentiation, relative to the value for all accessions separately. Therefore, this analysis does not provide support for the taxa included in these clusters.

In those clusters that contain accessions with more than two species labels, the pairwise Fst of the species groups were compared with the overall Fst of that specific cluster to distinguish among species labels that did represent a meaningful grouping (meaning an increased Fst) and those species labels that did not. The species labels that showed a low pairwise Fst were subsequently merged. In most of the clusters one or two merging steps were sufficient, but in cluster 7, 12, and 14, three cycles were needed, while in cluster 10 and 16 this process took four cycles to reach the maximum Fst. In some clusters this meant that the highest overall Fst was reached when most of the species labels were merged together; this was the case in cluster 10, 14 and 16. In other clusters the optimal Fst was reached at an arrangement that only merged a few of the species in the cluster, while other species remained separate. This was the case in cluster 3, 4 and 13. In cluster 8 no new arrangement yielded a higher Fst.

*Table 2. Genetic diffrentiation within the 16 observed clusters*

| Cluster | Species included in cluster | Sampling units | n | Ht | Hw | Hb | Fst | Species included in the clusters / new arrangements |
|---|---|---|---|---|---|---|---|---|
| 1 | rap (15) | accessions | 15 | 0.2679 | 0.2776 | -0.0098 | -0.0365 | |
| 2 | ver (19), mcp (2) spec 262, spec 287 | accessions | 23 | 0.2515 | 0.2448 | 0.0068 | 0.0272 | |
| | | species | 4 | 0.2765 | 0.2063 | 0.0702 | 0.2377 | |
| | | new groups | 3 | 0.2445 | 0.1551 | 0.0894 | 0.3598 | |
| | | new groups | ***2*** | ***0.2859*** | ***0.1471*** | ***0.1388*** | ***0.4675*** | ***(ver, mcp) (spec 262, spec287)*** |
| 3 | ktz(7), mag (2), oka (8), rch (1), spg (1), vnt (4) | accessions | 22 | 0.2875 | 0.2704 | 0.0171 | 0.0592 | |
| | | species | 6 | 0.2438 | 0.2113 | 0.0325 | 0.1285 | |
| | | new groups | ***3*** | ***0.2*** | ***0.1527*** | ***0.0473*** | ***0.2319*** | ***(ktz, mag, rch) (oka, vnt) spg*** |
| | | new groups | 2 | 0.1686 | 0.1305 | 0.0381 | 0.2231 | |
| 4 | gig(18), mag (2), mcd (6) spg (1) | accessions | 27 | 0.2556 | 0.2457 | 0.0098 | 0.0384 | |
| | | species | 4 | 0.1973 | 0.1777 | 0.0195 | 0.0984 | |
| | | new groups | ***3*** | ***0.213*** | ***0.1898*** | ***0.0231*** | ***0.1059*** | ***(gig, mcd) mag spg*** |
| | | new groups | 2 | 0.3341 | 0.3064 | 0.0277 | 0.0639 | |
| 5 | med (6), snd (3), wbr (2) vio (1) | accessions | 13 | 0.303 | 0.2864 | 0.0166 | 0.0548 | |
| | | species | 4 | 0.2594 | 0.2069 | 0.0525 | 0.2023 | |
| | | new groups | 3 | 0.2626 | 0.1949 | 0.0678 | 0.2559 | |
| | | new groups | ***2*** | ***0.2926*** | ***0.2055*** | ***0.0871*** | ***0.2893*** | ***(snd, wbr, med) vio*** |
| 6 | cmm (12), mlm (7) | accessions | 19 | 0.3357 | 0.3535 | -0.0177 | -0.0528 | |
| | | species | ***2*** | ***0.1763*** | ***0.1395*** | ***0.0368*** | ***0.2084*** | ***cmm mlm*** |
| 7 | abz (4), acp (1), acs (1), agu (1), chn (3), chv (1), dcm (4), hcr (1), imt (3), mcq(4), ncd (2), scb (1) vio (1) spec 205, spec 310, spec 6 | accessions | 30 | 0.3016 | 0.2449 | 0.0567 | 0.1878 | |
| | | species | 16 | 0.2754 | 0.1988 | 0.0766 | 0.2777 | |
| | | new groups | 7 | 0.2448 | 0.143 | 0.1019 | 0.4175 | |
| | | new groups | 6 | 0.245 | 0.1405 | 0.1045 | 0.4276 | |
| | | new groups | ***5*** | ***0.2513*** | ***0.1419*** | ***0.1094*** | ***0.4357*** | ***(acs, agu, hcr, acp, scb, chn, mcq, imt, spec 205, spec 310, spec 6) (dcm, chv) ncd, abz, vio*** |
| | | new groups | 4 | 0.25 | 0.1418 | 0.1082 | 0.4333 | |
| 8 | hcb(4), sgr (3), vio (1) | accessions | 9 | 0.2813 | 0.2319 | 0.0494 | 0.1758 | |

*Table 2. (Continued) Genetic diffrentiation within the 16 observed clusters*

| Cluster | Species included in cluster | Sampling units | n | Ht | Hw | Hb | Fst | Species included in the clusters / new arrangements |
|---|---|---|---|---|---|---|---|---|
| | | species | ***3*** | ***0.2774*** | ***0.1919*** | ***0.0856*** | ***0.3043*** | ***hcb, sgr, vio*** |
| | | new groups | 2 | 0.3053 | 0.2128 | 0.0925 | 0.2948 | |
| **9** | bal(3), han (7), vrn (16), spg (1) | accessions | 27 | 0.2659 | 0.2434 | 0.0226 | 0.0847 | |
| | | species | 4 | 0.2092 | 0.1687 | 0.0405 | 0.1904 | |
| | | new groups | 3 | 0.2133 | 0.1665 | 0.0468 | 0.2114 | |
| | | new groups | ***2*** | ***0.1662*** | ***0.1251*** | ***0.0411*** | ***0.2459*** | ***(bal, vrn, spg) han*** |
| **10** | abn (1), ach(1), amb (3), aml(1), amy (2), aym (1), buk (6), can (7), chi (1), cnd (3), cop (5), hro (1), lmb (1), mlt (11), mrn (5), orp (5), pam (4), sou (1), scr (1), sub (1) vlr (2) vrg (1) spec 998, spec 184, spec 292, spec 533, spec 726, spec 796, spec 394, spec 933 | accessions | 72 | 0.3095 | 0.2881 | 0.0213 | 0.069 | |
| | | species | 31 | 0.2828 | 0.2529 | 0.03 | 0.1058 | |
| | | new groups | 16 | 0.2598 | 0.2171 | 0.0427 | 0.1645 | |
| | | new groups | 7 | 0.2574 | 0.2045 | 0.0529 | 0.2066 | |
| | | new groups | 5 | 0.2796 | 0.2195 | 0.0601 | 0.2143 | |
| | | new groups | ***4*** | ***0.2996*** | ***0.2336*** | ***0.066*** | ***0.2194*** | ***(abn, ach, amb, aml, amy, aym, buk, can, chi, cnd, cop, hro, lmb, mlt, mrn, orp, pam, sou, scr, sub, vlr, vrg, spec 998, spec 184, mtp, spec 292, spec 533, spec 726, spec 796) ach, spec933, spec394*** |
| | | new groups | 3 | 0.2748 | 0.2142 | 0.0606 | 0.2189 | |
| **11** | lgl (2), mga (4), sct (6), tor (7) | accessions | 19 | 0.3231 | 0.31 | 0.0131 | 0.0401 | |
| | | species | 4 | 0.2285 | 0.1885 | 0.04 | 0.1743 | |
| | | new groups | ***3*** | ***0.2421*** | ***0.1968*** | ***0.0453*** | ***0.1858*** | ***(mga, tor) lgl, sct*** |
| | | new groups | 2 | 0.2269 | 0.1877 | 0.0391 | 0.1704 | |
| **12** | arz (5), chc (12), rzl (1), stl(1), yun (3), tar (4), vrn (1) spec 210, spec211, spec329, grl(1) | accessions | 31 | 0.2927 | 0.2716 | 0.0211 | 0.072 | |
| | | species | 11 | 0.2784 | 0.2506 | 0.0278 | 0.0991 | |
| | | new groups | 6 | 0.2236 | 0.1867 | 0.0369 | 0.1646 | |
| | | new groups | 4 | 0.22 | 0.178 | 0.042 | 0.1886 | |
| | | new groups | ***3*** | ***0.2317*** | ***0.1853*** | ***0.0465*** | ***0.1949*** | ***(chc, rzl, stl, grl, spec210, spec211, spec329, yun, arz) vrn, arz*** |

*Table 2. (Continued) Genetic diffrentiation within the 16 observed clusters*

| Cluster | Species included in cluster | Sampling units | n | Ht | Hw | Hb | Fst | Species included in the clusters / new arrangements |
|---|---|---|---|---|---|---|---|---|
| | | new groups | 2 | 0.1772 | 0.1445 | 0.0326 | 0.1831 | |
| 13 | dds (1), grl (15), hps (3), ifd (2), inm (8), ptr (3), vid (3), nrs (2), spg (3), haw (7) | accessions | 47 | 0.3045 | 0.2885 | 0.016 | 0.0526 | |
| | | species | 10 | 0.2252 | 0.18 | 0.0452 | 0.2006 | |
| | | new groups | ***5*** | ***0.2135*** | ***0.1582*** | ***0.0553*** | ***0.259*** | ***(grl, dds, hps, ifd, inm, ptr) vid, haw, spg, nrs*** |
| | | new groups | 4 | 0.2066 | 0.1545 | 0.0521 | 0.2522 | |
| 14 | chc(3), dds((3), tar(27), aln(4), ber(23), gnd(5), stl(1), spec 255, spec 601 | accessions | 69 | 0.2671 | 0.2518 | 0.0153 | 0.0573 | |
| | | species | 9 | 0.2092 | 0.1823 | 0.027 | 0.1276 | |
| | | new groups | 6 | 0.1726 | 0.148 | 0.0246 | 0.1417 | |
| | | new groups | 5 | 0.1766 | 0.1502 | 0.0264 | 0.1479 | |
| | | new groups | ***4*** | ***0.157*** | ***0.1327*** | ***0.0244*** | ***0.1549*** | ***(stl, spec 255, spec 601, yun, chc, dds, tar) aln ber gnd*** |
| | | new groups | 3 | 0.1616 | 0.1383 | 0.0233 | 0.1439 | |
| 15 | oka (7) | accessions | ***7*** | ***0.261*** | ***0.2602*** | ***0.0007*** | ***0.0029*** | |
| 16 | grl (18), hps(2), aln(3), arp(2), dds(1), hmp(1) ptr(5), scr(8), lph(6), ugt(4), vrg(4), opl(17), ifd(5), vid(5), gnd(1), vrn(1), nrs(2), spg(1), spl(13), brc(10), ast(5), blv(3), hdm(6), avl(3) spec352, spec43, spec123, spec165, spec891, spec381, spec649 | accessions | 134 | 0.3169 | 0.302 | 0.015 | 0.0473 | |
| | | species | 31 | 0.2456 | 0.215 | 0.0306 | 0.1245 | |
| | | new groups | 14 | 0.217 | 0.1823 | 0.0347 | 0.1597 | |
| | | new groups | 7 | 0.214 | 0.1742 | 0.0399 | 0.1855 | |
| | | new groups | 6 | 0.2166 | 0.1749 | 0.0418 | 0.1913 | |
| | | new groups | ***5*** | ***0.224*** | ***0.1799*** | ***0.0442*** | ***0.1952*** | ***(grl, hps, aln, arp, dds, hmp ptr, scr, lph, ugt, vrg, opl, ifd, vid, gnd, vrn, nrs, spg, spl, brc, spec416, spec352, spec43, spec123, spec165, spec891, spec381), (ast, blv), hdm, avl, spec649*** |
| | | new groups | 4 | 0.2372 | 0.1932 | 0.0439 | 0.1831 | |

## Discussion

The results given by the analysis with STRUCTURE of over 500 South-American *Solanum* section *Petota* accessions show that the optimal overall subdivision of the accessions is 16 clusters. Fst analysis shows that support for more groups within these clusters can be found when studied more closely, up to a grand total of 47 units (Table 2). Because the majority of the existing genetic variation can be explained already with a subdivision of 10 groups and the 16 group subdivision explains more genetic variation than the 90 species labels, a subdivision in 90 species labels does not seem a correct model to explain the genetic differentiation among the accessions in the present study. This does not automatically mean that 47 is the correct number of species. First, some species may have been represented by one or only few accessions in this study, and therefore may not have appeared as a separate group. Second, and more importantly, genetic differentiation would be expected among separate species but it can also be found among populations within a species (see below).

**Misclassification, overclassification and hybridization**

The Fst value of the 90 species arrangement is lower than that of 16 groups, but still high (0.2953) indicating that the species arrangement does explain considerable genetic variation within the dataset, in excess of that being explained by the accessions. This high value of Fst might be caused by a few correct species labels that differ greatly from the rest, while most other species labels are incorrect. The decrease of the Fst from a 16 cluster arrangement to one based on 90 species indicates that some species labels are incorrect. Detailed inspection of the results of the 16 cluster arrangement enabled to differentiate four types of observations on the preexisting species labels, with concomitant implications for their biological status.

**Support for a few species labels**

First, the STRUCTURE and AFLP-SURV results show some species labels to behave as distinct genetic units. The species labels *S. raphanifolium*, *S. verrucosum* (plus its synonym *S. macropilosum*), *S. commersonii* (plus *S. commersonii* subsp. *malmeanum*) and *S. okadae* were put in exclusive clusters by STRUCTURE. The seven *S. okadae* accessions that appear in cluster 3 together with *S. venturii* accessions turned out to be mislabeled and have been corrected as being *S. venturii* accessions by R. Hoekstra from CGN (personal communication). The species labels *S. microdontum* (including *S. microdontum* subsp. *gigantophyllum* as a synonym) and *S. huancabambense*, and *S. sogarandinum* share their cluster with other accessions from other species, but the optimal partitioning of genetic variation within the cluster shows that they could represent distinct genetic units. The support for these species labels as distinct genetic units, or species, is consistent with the results from Jacobs et al. (2008). According to our results, the following species labels should be preserved as correct labels covering distinctive units: *S. raphanifolium*, *S. verrucosum* (with *S. macropilosum* as synonym), *S. microdontum* (including *S. microdontum* subsp. *gigantophyllum* as a synonym) *S. commersonii* with *S. commersonii* subsp. *malmeanum* as subspecies, *S. okadae* (only the 7 accessions in cluster 15), *S. huancabambense*, and *S. sogarandinum*.

Most of these species are designated as good and possibly stable species by one or sometimes more previous studies (Castillo & Spooner, 1997; Spooner & Salas, 2006; Spooner *et al.*, 1991; Spooner *et al.*, 2004; van den Berg & Spooner, 1992). For the species label *S. huancabambense* no data on species status were found, and *S. okadae* was suggested to be part of the brevicaule complex, but no further information on this remark is found (Spooner & Salas, 2006).

**Weak support for some species labels and combinations of species labels**

Second, some species labels appear in one STRUCTURE cluster, but their accessions do not seem to form distinct genetic units like the species labels described before, or only receive support as distinct units either by STRUCTURE or based on high Fst values, but not both. We designated the support for the species status of these species labels and sometimes combinations of species labels as weak. Weak support was found for *S. kurtzianum*, *S. venturii*, a combination of the species *S. sandemanii*, *S. weberbauerii*, *S. medians*, *S. albornozii*, a combination of *S. chavinense* and *S. dolichocremastrum*, *S. hannemanii*, a combination of *S. vernei* subsp. *balsii* and *S. vernei*, a combination of *S. megistacrolobum* and *S. megistacrolobum* subsp. *toralapanum*, *S. hawkesianum*, *S. alandiae* and *S. gandarillassi*, a combination of *S. boliviense* and *S. boliviense* subsp. *astleyi*, *S. hondelmanii*, and *S. avilesii*. Notably, accessions of certain species were always clustered together by STRUCTURE, while within this group the Fst analysis led to a compete merger of these species labels. For these species labels we conclude that they are superfluous, and prime examples of overclassification. This is the case for *S. microdontum* subsp. *gigantophyllum* (already acknowledged to be a synonym of *S. microdontum* by van den Berg and Spooner (1992)) and *S. microdontum* in cluster 4, *S. kurtzianum* and *S. rechei* in cluster 3, *S. sandemanii*, *S. weberbaueri*, and *S. medians* in cluster 5, *S. vernei* subsp. *balsii* and *S. vernei* in cluster 9, and *S. megistacrolobum* and *S. megistacrolobum* subsp. *toralapanum* in cluster 11.

In many cases where support was found for combining two taxa, in stead of separating them as distinct units, it concerned subspecies of the same species label. On some of the species labels mentioned here, more extensive research has been done previously. Giannatasio and Spooner (1994a; 1994b) studied the species boundaries between *S. megistacrolobum* and *S. megistacrolobum* subsp. *toralapanum* with molecular (RFLP) and morphological data and suggested to preserve *S. megistacrolobum* subsp. *toralapanum* as a distinct subspecies while our analysis does not find support for that. Spooner et al (1997) studied the relationships of *S. boliviense* and *S. astleyi* with morphological data and found *S. astleyi* to be a supspecies of *S. boliviense*. Our data (weakly) support combining the two taxa as one species, but support for subspecies differentiation is not found. The provisional species label *S. hannemanni* was recently investigated with AFLP data by van den Berg and Groendijk-Wilders (2007). They found evidence for the species status of this label. Our recent data also (weakly) support the species status of *S. hannemannii*.

Comparing our data to the list that was published in a review by Spooner and Salas (2006), we see some similarities, but also some incongruencies. For *S. venturii, S. weberbauerii, S. chavinense, S. hannemanii, S. hawkesianum, S. alandiae, and S. hondelmanii*, no information on species status is given by Spooner and Salas (2006), maybe because these species labels were not recognized by the authors. The species labels *S. kurtzianum, S. medians, S. vernei, S. megistacrolobum* (with *S. megistacrolobum* subsp. *toralapanum* as a stable subspecies), and *S. gandarillasii* were seen as phenetically distinct species from South America (Spooner & Salas, 2006).

**No support for species status**

Third, the analysis shows that for several species labels the accessions belonging to that species were scattered across two or even three clusters. Furthermore, some of them appeared in different subclusters when STRUCTURE was run at the appropriate K on the separate cluster. This kind of observations was made on the following species labels: *S. maglia, S. doddsii, S. chacoense, S. gourlayi, S. virgultorum, S. hoopesii, S. augustii, S. tarijense S. vernei, S. infundibuliforme, S. alandiae, S. neorosii, S. sucrense, S. gourlayi* subsp. *pachytrichum*, and *S. violaceimarmoratum*. Many of these species may be the product of hybridization between two and sometimes even more potato species. This is not surprising, as many authors have suggested that many species from *Solanum* sect. *Petota* are the results of hybridization (Hawkes, 1990; Spooner & Salas, 2006). It could also be caused by misclassification due to problematic identification or by an incomplete or vague description of the species. The problems with the identification of species were already addressed by several earlier studies. Spooner et al. (2006) and Spooner and van den Berg (1992) noted that many of the taxa are extremely similar in morphology and many species are distinguished only by minor and often overlapping character states. In total, for 8 species labels that are named here, no previous data on their species status could be found. Spooner and Salas (2006) claimed that *S. maglia, S. chacoense*, and *S. infundibuliforme* would be phenetically distinct species but their source of information is unknown. Fourth, some species appear in (mainly) only one cluster of the STRUCTURE analysis, but the accessions do not form a separate group, neither in the analysis of genetic variation nor in a STRUCTURE analysis of all accessions related to the cluster, not even as part of a fixed combination with one other species label. This involves the species labels *S. mochiquense, S. immite, S. chancayense, in cluster 7, S. canasense, S. bukasovii, S. candolleanum, S. coelestipetalum, S. pampasense, S. ambosinum, S. marinasense, S. multidissectum , S. velardei* in cluster 10, *S. arnezii, S. yungasense*, in cluster 12, *S. incamayoense*, in cluster 13, *S. tarijense, S. berthaultii* in cluster 14, *S. arac-pappa, S. leptophyes, S. ugentii, S. oplocense, S. sparsipilum, and S. brevicaule* in cluster 16. Previous results from a morphological study by Spooner and van den Berg (1992b) suggest that the species labels *S. berthaultii* and *S. tarijense* should be combined. Species label *S. oplocense* was shown to be a well defined species using morphological data by van den Berg et al. (1998) and using molecular data by Miller and Spooner (1999), but it was not distinct in an AFLP study by Spooner et al. (2005). Many species labels mentioned in this category are considered to be part of the brevicaule complex (Miller & Spooner, 1999; van den Berg *et al.*, 1998). This would be valid for the species labels: *S. canasense S. bukasovii, S. candolleanum, S. coelestipetalum, S. pampasense, S. ambosinum, S. marinasense, S. velardei, S. incamayoense, S. leptophyes, S. ugentii* and *S. sparsipilum*.

All the herefore mentioned conclusions are summarized in Table 3, in which for each species label is indicated if evidence for species status was found in this analysis. Overall, for 8 species labels support was found for preserving the species status, and for 10 species labels plus five 2-species combinations weak support was found. No support was found for the remaining 43 species labels (and 19 species labels were only represented by only one accession).

**Correlation between clusters and geography**

The accessions within a cluster usually come from the same geographical region (Additional file 1), which is consistent with a meaningful arrangement of the accessions into groups that may exchange genetic material. For the largest and most complicated clusters (7, 10, 12, 14, 16) the information on the geographic origin of the accessions allows to draw some tentative conclusions. Cluster 16 contains mostly accessions from Argentina and Bolivia from the Southern brevicaule complex and Cluster 10 consist mostly of accessions from Peru (and northern Bolivia) that can be considered as belonging to the Northern brevicaule complex. Cluster 7 contains almost exclusively Peruvian accessions, and some species labels in cluster 7 (*S. albornozii, S. augustii, S. chancayense, S. dolichocremastrum, S. immite*) are associated with the Hawkes series *Piurana*, (Hawkes, 1990) and Spooner and Salas (2006), but Jacobs et al. (2008) could not find support for these species to be included in one of the Piurana species groups. Cluster 14 contains all *S. berthaultii* accessions and almost all *S. tarijense* accessions, plus few accessions with other species labels, which mostly come from Bolivia and Argentina. Cluster 12 contains accessions from various geographical origins, most of them come from Bolivia and Argentina but some are from Peru and Paraguay.

**Evidence against species status of several species labels**

We have searched for overall genetic structure in our South American dataset of 566 accessions and found that a subdivision in 16 clusters produces the largest genetic differences among groups of accessions. This was confirmed by comparing the Fst of several divisions of the dataset. Subsequently, we have gone back to the original species labels within the clusters to test whether they provide the framework which could explain most of the genetic differentiation within a cluster. Some of the species labels already receive support because they distinguished themselves in the Bayesian clustering. For others we could find support, although sometimes weakly, within the defined clusters. For a large number of species labels (43 in total) we could not find any support for species status. The subclusters seem to split certain species labels and put accessions of different species labels together. All these results could be interpreted as evidence against the (sub)species status of these species labels. This is especially true for sub-species labels, as only one of all sub-species labels was supported (*S. commersonii* subsp. *malmeanum* could be differentiated from *S. commersonii*).

**Alternative new taxonomic units?**

The observation of so many incorrect species labels that seem not fit to cover the available genetic variation between the accessions, poses the question of how to classify these accessions anew. Although STRUCTURE recognizes some species as distinguishable units, it would be undesirable to conclude that the other clusters or groups within clusters containing accessions from more than one species label would also represent natural units at the species level. The absolute values of genetic differentiation within clusters vary quite strongly, indicating varying levels of variation within and differentiation between groups. To test the genetic partitioning of new arrangements within the clusters we choose the species labels as test units, as we aimed to test the validity of the existing labels. If, in contrast, the scope would be to exactly define the new groups (or new species), the approach we have presented here should take the accessions as units, and infer groups based on genetic similarities among individual accession, for each of the 16 clusters. Then, this evidence should be combined with information on morphology and geographical distribution, and with information from crossing experiments, to determine which groups merit species status.

Although we have not undertaken any activities yet in that area, we dare to speculate on this issue. For some accessions, it might turn out to be best to consider them as members of a large species complex, in which some metapopulation structure is visible, represented by the clusters and subclusters. Perhaps these accessions are still in the early phase of speciation and there fore lack any genetic or morphological features that can distinguish them from other accessions (Shaffer & Thomson, 2007).This scenario could be valid in the case of Clusters 10, 12 and 16 that contain many different species labels that are sometimes not restricted to one cluster exclusively.

## Acknowledgements

This project was (co)financed by the research program of the Centre for Biosystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research. The acronym AFLP is a registered trademark (AFLP®) of Keygene N.V. and the AFLP® technology is covered by patents and patent applications of KeyGene N.V.

*Table 3. Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| *Tuberosa II* | *S. abancayense* Ochoa | abn | 423 | CGN 18357 | 1 | synonym of *S. bukasovii* (Ochoa, 1999) | only one accession |
| *Tuberosa III* | *S. achacachense* Cárdenas | ach | 99 | GLKS 32830 | 1 | | only one accession |
| *Tuberosa II* | *S. acroscopicum* Ochoa | acs | 100 | GLKS 32436 | 1 | | only one accession |
| *Tuberosa III* | *S. alandiae* Cárdenas | aln | 257, 320, 455, 457, 458, 459, 460 | CPC 7212, CGN 18245, CGN 22349, cgn962384, CGN 20651, CGN 18260, CGN 18264 | 7 | | weak evidence for combination with *S. gandarillasii* |
| *Piurana* | *S. albornozii* Correll | abz | 2, 102, 103, 466 | PI 561637, GLKS 35297, GLKS 35298, CGN 22731 | 4 | | weak evidence |
| *Tuberosa II* | *S. amabile* Vargas | aml | 3 | PI 365356 | 1 | synonym of *S. canasense* (Hawkes, 1990) | only one accession |
| *Tuberosa II* | *S. amayanum* Ochoa | amy | 302, 303 | CIP 763004, CIP 763005 | 2 | | no evidence |
| *Tuberosa II* | *S. ambosinum* Ochoa | amb | 104, 105, 467, | GLKS 32282, GLKS 35299, CGN 18358 | 3 | | no evidence |
| *Tuberosa II* | *S. ancophilum* (Correll) Ochoa | acp | 304 | CIP 761448 | 1 | synonym of *S. rhomboideilanceolatum* Ochoa (Hawkes, 1990) | only one accession |
| *Megistacroloba* | *S. aracc-papa* Juz. | arp | 109, 110 | GLKS 30082, GLKS 30081 | 2 | nomen dubium (Hawkes, 1990) | no evidence |
| *Yungasensa* | *S. arnezii* Cárdenas | arz | 4, 111, 112, 113, 471 | PI 545880, GLKS 32832, GLKS 32833, GLKS 32834, GLKS 32831 | 5 | | no evidence |
| *Megistacroloba* | *S. astleyi* Hawkes and Hjert. | ast | 114, 472, 474, 475, 476 | GLKS 32836, CGN 18207, CGN 18210, CGN 18211, CGN 18212 | 5 | | weak evidence for combination with *S. boliviense* |
| *Tuberosa II* | *S. augustii* Ochoa | agu | 305 | CIP 762631 | 1 | | only one accession |
| *Tuberosa III* | *S. avilesii* Hawkes and Hjert. | avl | 477, 478, 479, | CGN 18255, CGN 18256, CGN 18257 | 3 | | no evidence |
| *Tuberosa II* | *S. aymaraesense* Ochoa | aym | 5 | PI 607896 | 1 | | only one accession |
| *Tuberosa III* | *S. berthaultii* Hawkes | ber | 322, 323, 324, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 561*, 939, 940, 941, 943, 944 | CGN 20644, CGN 20650, CGN 18042, CGN 18074, CGN 18190, CGN 20635, CGN 20636, CGN 22715, CGN18216, CGN 22716, CGN 20645, CGN 18246, CGN 23804, CGN 18228, CGN 22727, BGRC 15479, CGN 17823, CGN | 24 | | no evidence |

*Table 3. (Continued) Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| | | | | 18118, GLKS 31670*, CGN 18189, CGN 23508, CGN 18267, CGN 17716, CGN 23477 | | | |
| *Megistacroloba* | *S. boliviense* Dunal | blv | 496, 498, 499 | CGN 18196, CGN 18070, INTA 73228B | 3 | | weak evidence for combination with *S. astleyi* |
| *Tuberosa III* | *S. brevicaule* Bitter | brc | 327, 505, 506, 507, 509, 1020, 1025, 1026, 1040, 1047, | CGN 18231, CGN 17841, CGN 18226, CGN 18232, CGN 22321, CGN 18030, CGN 18223, CGN18247, CGN22322, CGN22717 | 10 | | no evidence |
| *Tuberosa II* | *S. bukasovii* Juz. | buk | 328, 511, 512, 514, 955, 971, | CGN 17683, CGN 17684, CGN 17737, CGN 17821, CGN 21305, CGN 17738 | 6 | | no evidence |
| *Tuberosa II* | *S. canasense* Hawkes | can | 260, 526, 527, 528, 529, 951$, 952, 953 | CPC 2725, cgn960639, CGN 17722, CGN 17672, CGN 17589, CGN 20592$, CGN 18072, CGN23007 | 7 | | no evidence |
| *Tuberosa III* | *S. candolleanum* P. Berthault | cnd | 530, 531, 532 | PI 498226, CGN 18132, CGN 20603 | 3 | | no evidence |
| *Yungasensa* | *S. chacoense* Bitter | chc | 125, 126, 127, 246*, 263, 338, 470$, 543, 544, 545, 546, 547, 548, 549, 550, 551 | GLKS 30162, GLKS 30161, GLKS 30180, GLKS 32343*, CPC5901, CGN 18248, CGN 17679$, cgn962709, CGN 18365, CGN 17702, CGN 22384, CGN 18202, CGN 18294, CGN 18338, cgn961764, CGN 22368 | 15 | | no evidence |
| *Tuberosa II* | *S. chancayense* Ochoa | chn | 1, 552, 553 | VIR 20892, CGN 18036, CGN 18356 | 3 | | no evidence |
| *Megistacroloba* | *S. chavinense* Correll | chv | 11 | PI 498235 | 1 | | only one accession |
| *no information* | *S. chillonanum* Ochoa | chi | 12 | PI 607890 | 1 | | only one accession |
| *Tuberosa II* | *S. coelestipetalum* Vargas | cop | 134, 135, 306, 307, 572 | GLKS 35433, GLKS 35434, CIP 761755, CIP 761999, CGN 20557 | 5 | | no evidence |
| *Commersoniana* | *S. commersonii* Dunal | cmm | 265, 575, 576, 577, 578, 1017, 1018, 1019, 1027, 1028, 1039, 1050 | CPC 5861, cgn961592, cgn961597, CGN 18027, CGN 22351, CGN 17988, CGN 18024, CGN 18026, CGN 18327, CGN 18328, GLKS 35340, CGN 23492 | 12 | | evidence |
| *Commersoniana* | *S. commersonii* | mlm | 139, 266, 579$, 580, 581, | GLKS 35340, CPC 7520, CGN | 7 | | evidence |

*Table 3. (Continued) Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| | subsp. *malmeanum* Bitter | | 1038, 1045, 1058 | 18329$, CGN 18025, CGN 18215, CGN21353, CGN 22352, cgn962274 | | | (subspecies or species) |
| *Tuberosa II* | *S. dolichocremastrum* Bitter | dcm | 147, 148, 149, 308, | GLKS 32342, GLKS 35348, GLKS 35349, CIP 762533 | 4 | | weak evidence |
| *Tuberosa III* | *S. gandarillasii* Cárdenas | gnd | 16, 62, 163, 270, 346, 603, | PI 597750, PI 597751, GLKS 32423, CPC 7044, CGN 20560, CGN 17590 | 6 | | weak evidence for combination with *S.alandiae* |
| *Tuberosa III* | *S. gourlayi* Hawkes | grl | 347, 604, 605, 606, 607$, 608, 609, 610, 611, 1000$, 1005, 1006, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1021, 1022, 1029, 1030, 1032, 1033, 1034, 1035, 1036$, 1037, 1042, 1043$, 1044, 1048, 1049, 1051, 1052$, 1053, 1054, 1055$ | CGN 17851, CGN 22705, CGN 17591, CGN 18039, CGN 22380$, cgn961345, CGN 17592, CGN 22336, CGN 21335, cgn961607$, CGN 17872, CGN 17873, CGN 17962, CGN 17963, CGN 17965, CGN 17966, CGN 17967, CGN 17969, CGN 17970, CGN 17971, CGN 18065, CGN 18066, CGN 20585, CGN 20594, CGN 20657, CGN 21332, CGN 21333, CGN 21334, CGN 21336$, CGN 21341, CGN 22340, CGN 22342$, CGN 22343, CGN 23022, CGN 23486, CGN 23497, CGN 23515$, cgn960071, cgn961347, CGN 23794$ | 34 | synonym of *S. leptophyes* (Ochoa, 1990) | no evidence |
| *Tuberosa III* | *S. gourlayi subsp. pachytrichum* | ptr | 612, 613, 614, 615, 616, 617, 618, | cgn18102, cgn18176, bgrc27294, bgrc27295, cgn18188, bgrc7231, bgrc28084 | 7 | synonym of *S. leptophyes* (Ochoa, 1990) | no evidence |
| *Tuberosa III* | *S. gourlayi* subsp. *vidaurrei (Cárdenas) Hawkes and Hjert.* | vid | 619, 620, 621, 622, 623, 624, 625, 626, | CGN 17848, CGN 17849, CGN 18040, CGN 17850, CGN 18038, CGN 17864, CGN 23024, CGN 23045 | 8 | | no evidence |
| | *S. hannemanii* | han | 252*, 628, 629, 630, 631, 632, 633 | GLKS 32196*, CGN 17996, CGN 17854, CGN 17997, CGN 20578, CGN 17856, CGN 17858 | 7 | provisional name | weak evidence |
| | *S. hawkesianum* | haw | 166, 167, 634, 635, 636, 637, 638, | GLKS 32762, GLKS 32765, CGN 17888, CGN 17889, CGN 17890, CGN 17891, CGN 17892 | 7 | provisional name | weak evidence |
| *Tuberosa III* | *S. hondelmannii* Hawkes and Hjert. | hdm | 168, 351, 644, 645, 646, 647, 648 | GLKS 32852, CGN 18106, cgn961918, cgn962199, CGN | 7 | | weak evidence |

*Table 3. (Continued) Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| | | | | 18192, CGN 18193, cgn962204 | | | |
| *Tuberosa III* | *S. hoopesii* Hawkes and Okada | hps | 169, 650, 651, 652, 653 | GLKS 32885, CGN 18363, CGN 18367, CGN 18368, CGN 18372 | 5 | | no evidence |
| *Yungasensa* | *S. huancabambense* Ochoa | hcb | 18, 170, 353, 354 | PI 365359, GLKS 32441, CGN 18306, CGN 17719 | 4 | | evidence |
| *Tuberosa II* | *S. huarochiriense* Ochoa | hro | 309 | CIP 761224 | 1 | | only one accession |
| *Tuberosa II* | *S. humectophilum* Ochoa | hmp | 171 | GLKS 32829 | 1 | | only one accession |
| *Piurana* | *S. hypacrarthrum* Bitter | hcr | 311 | CIP 761259 | 1 | | only one accession |
| *Tuberosa II* | *S. immite* Dunal | imt | 63, 64, 172, | PI 498245, PI 365331, GLKS32819 | 3 | | no evidence |
| *Tuberosa III* | *S. incamayoense* K.A. Okada and A.M. Clausen | inm | 657, 658, 659, 660, 661, 662, 663, 1016 | CGN 18077, CGN 21320, CGN 17874, CGN 17875, CGN 17968, cgn961363, CGN 22335, CGN 17972 | 8 | | no evidence |
| *Megistacroloba* | *S. infundibuliforme* Phil. | ifd | 664, 665, 666, 667, 668, 1007, 1023 | CGN 17720, CGN 23063, CGN 22334, CGN 23048, cgn960696, CGN 17959, CGN 18079 | 7 | | no evidence |
| *Tuberosa III* | *S. kurtzianum* Bitter and Wittm. | ktz | 275, 276, 675, 676, 677, 678, 995, | CPC 5864, CPC 5889, CGN 22338, cgn961563, CGN 23042, cgn961013, CGN 22353 | 7 | | weak evidence |
| *Tuberosa II* | *S. leptophyes* Bitter | lph | 356, 357, 680, 682, 683, 684, | CGN 18174, CGN 18140, CGN 18173, CGN 18167, CGN 20611, CGN 18126 | 6 | | no evidence |
| *Lignicaulia* | *S. lignicaule* Vargas | lgl | 179, 685 | GLKS 32215, CGN 17723 | 2 | | weak evidence |
| *Conicibaccata* | *S. limbaniense* Ochoa | lmb | 686 | CGN 22720 | 1 | | only one accession |
| *Tuberosa I* | *S. macropilosum* Correll | mcp | 23, 74 | PI 607844, PI 607845 | 2 | synonym of *S. verrucosum* (Spooner et al. 2004) | no evidence, part of *S. verrucosum* |
| *Maglia* | *S. maglia* Schtdl. | mag | 75, 76, 359, 688, | PI 245087, PI 558316, CGN 18064, CGN 22719 | 4 | | no evidence |
| *Tuberosa II* | *S. marinasense* Vargas | mrn | 77, 181, 182, 277, 690 | PI 607884, GLKS 35430, GLKS 32281, CPC 7172, CGN 17594 | 5 | | no evidence |
| *Tuberosa II* | *S. medians* Bitter | med | 183, 691, 692, 693, 694, 695, | GLKS 32226, CGN 21349, CGN 18043, CGN18308, CGN 21343, CGN 18307 | 6 | | weak evidence |
| *Megistacroloba* | *S. megistacrolobum* Bitter | mga | 696, 697, 699, 700 | CGN 23084, CGN 17828, CGN 22347, CGN 20601 | 4 | | weak evidence for combination with *S.* |

*Table 3. (Continued) Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| | | | | | | | *megistacrolobum* subsp. *toralapanum* |
| *Megistacroloba* | *S. megistacrolobum* subsp. *toralapanum* Cárdenas and Hawkes | tor | 278, 701, 702, 703, 704, 705, 706 | CPC 1773, CGN 17728, CGN 23006, CGN 18145, CGN 18146, CGN 18147, CGN 18125 | 7 | | weak evidence for combination with *S. megistacrolobum* |
| *Tuberosa III* | *S. microdontum* Bitter | mcd | 360, 707, 708, 958, 959, 994 | CGN 17596, CGN 22382, CGN 18259, CGN 20646, CGN 18047, CGN 20597 | 6 | | evidence |
| *Tuberosa III* | *S. microdontum subsp. gigantophyllum* (Bitter) Hawkes and Hjert. | gig | 361, 362, 710, 711, 712, 713, 714, 715, 956, 957, 960, 961, 962, 963, 964, 965, 966, 967, | CGN 18046, CGN 18083, CGN 18199, CGN 20639, CGN 18200, CGN 17595, CGN 23050, CGN 21342, CGN 18295, CGN 23511, CGN 20586, CGN 18048, CGN 17597, CGN 18049, CGN 18084, CGN 18003, CGN 18067, CGN 22372 | 18 | synonym of *S. microdontum* Bitter (van den Berg and Spooner, 1992) | no evidence, part of *S. microdontum* |
| *Tuberosa II* | *S. mochiquense* Ochoa | mcq | 186$, 716, 717, 718, 719 | GLKS 32319$, CGN 20587, CGN 18263, CGN 17731, CGN 21360 | 4 | | no evidence |
| *Tuberosa II* | *S. multidissectum* Hawkes | mll | 363, 722, 723, 724, 725, 727, 728, 729, 730, 731, 732 | CGN 17824, CGN 21344, CGN 18330, cgn960739, CGN 17686, CGN 17733, cgn960736, CGN 17825, cgn961613, cgn17840, cgn960967 | 11 | synonym of *S. bukasovii* Juz. f. multidissectum (Hawkes) Ochoa | no evidence |
| *Tuberosa II* | *S. multiinterruptum* Bitter | mlp | 190 | GLKS 32431 | 1 | | only one accession |
| *Tuberosa III* | *S. neocardenasii* Hawkes and Hjert. | ncd | 193, 734 | GLKS 32855, CGN 18217 | 2 | | no evidence |
| *Tuberosa III* | *S. neorossii* Hawkes and Hjert. | nrs | 281, 735, 736, 737, 987* | CPC 6047, CGN 18280, CGN 17599, CGN 18051, CGN 17763* | 5 | | no evidence |
| *Tuberosa III* | *S. okadae* Hawkes and Hjert. | oka | 283, 365*, 366, 367, 368, 739, 740, 741$, 742, 743, 744, 745, 746, 969, 970 | CPC 7129, CGN 18000*, CGN 18109, CGN 18108, CGN 17998, CGN 18269, CGN 17999, CGN 18279$, cgn962076, cgn962078, CGN 18157, CGN 22709, CGN 18129, CGN 22703, CGN 20599 | 14 | | evidence |
| *Tuberosa III* | *S. oplocense* Hawkes | opl | 747, 749, 750, 751, 752, 753, 754, 1001$, 1002, 1003, 1004, 1024, 1031, | CGN 23049, cgn962217, CGN 21352, CGN 18088, CGN 18085, CGN 21319, CGN 17736$, CGN | 17 | | no evidence |

*Table 3. (Continued) Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| | | | 1041, 1046, 1056, 1057, 1059 | 17868, CGN 17869, CGN 17870, CGN 18087, CGN 20638, CGN 22324, CGN 22713, CGN 23798, cgn961876, cgn962541 | | | |
| *Tuberosa II* | *S. orophilum* Correll | orp | 29, 83, 84, 196, 756 | PI 498213, PI 498209, PI 498212, GLKS 35301, cgn962570 | 5 | | no evidence |
| *Tuberosa II* | *S. pampasense* Hawkes | pam | 288, 762, 763, 764 | CPC 6024, CGN 962604, CGN 20575, cgn960051 | 4 | | no evidence |
| *Megistacroloba* | *S. raphanifolium* Cárdenas and Hawkes | rap | 208, 209, 291, 380, 790, 791, 792, 793, 794, 797, 798, 799, 800, 801, 976 | GLKS 30637, GLKS 30644, CPC 7090, CGN 17598, cgn960772, CGN 20589, CGN 18300, CGN 18089, cgn961878, CGN 18320, CGN 17752, CGN 18033, CGN 17833, CGN 17835, CGN 17822 | 15 | | evidence |
| *Megistacroloba* | *S. sanctae-rosae* Hawkes | sct | 803, 804, 805, 806, 807, 1061 | CGN 20576, CGN 22344, CGN 17910, CGN 20564, CGN 17837, cgn961619 | 6 | | weak evidence |
| *Tuberosa II* | *S. sandemanii* Hawkes | snd | 93, 94, 808 | PI 607894, PI 607895, CGN 17600 | 3 | | weak evidence for combination with *S. weberbaueri* |
| *Tuberosa II* | *S. scabrifolium* Ochoa | scb | 37 | PI 365363 | 1 | | only one accession |
| *Megistacroloba* | *S. sogarandinum* Ochoa | sgr | 215, 315, 316, 814 | GLKS 35382, CIP 761465, CIP 761586, CGN 17601 | 4 | | evidence |
| *Tuberosa II* | *S. soukupii* Hawkes | sou | 815 | CGN 18061 | 1 | synonym of *S. canasense* (Hawkes 1990) | only one accession |
| *Tuberosa II* | *S. sparsipilum* (Bitter) Juz. and Bukasov | spl | 382, 383, 384, 816, 817, 818, 819, 820, 821, 972, 973, 975, 978, | CGN 18225, CGN 18230, CGN 18154, CGN 18096, CGN 17838, CGN 18221, CGN 20653, CGN 17758, CGN 20602, CGN 18099, CGN 22702, CGN 18094, CGN 18131 | 13 | | no evidence |
| *Tuberosa III* | *S. spegazzinii* Bitter | spg | 217, 385, 386, 822, 823, 824, 826, 827, 828$ | GLKS 32755, CGN 17759, CGN 17839, cgn960795, CGN 21318, CGN 22707, CGN 21321, CGN 23015, CGN 18034$ | 8 | | no evidence |
| *Tuberosa III* | *S. tarijense* Hawkes | tar | 224, 225,280*,392, 852, 853, 854, 855, 856, 857, 858, 859, 860, 862, 863, | GLKS 31570, GLKS 31572, CPC 7208*, CGN 17861, CGN 22729, cgn962224, CGN 22714, CGN | 31 | | no evidence |

*Table 3. (Continued) Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| | | | 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, | 18198, cgn960807, cgn960805, cgn960806, CGN 17975, cgn961432, CGN 21337, CGN 23795, cgn961736, CGN 17976, CGN 17974, CGN 17977, CGN 18107, cgn961128, CGN 17978, CGN 17979, cgn961441, CGN 17980, CGN 21338, CGN 17981, cgn961449 , CGN 17982, cgn961451, cgn962690 | | | |
| *Tuberosa III* | *S. ugentii* Hawkes and K. A. Okada | ugt | 44, 248, 249, 892, | PI 546029, GLKS 32887, GLKS 32889, CGN 18369 | 4 | | no evidence |
| *Tuberosa II* | *S. velardei* Ochoa | vlr | 97, 893 | PI 619114, CGN 18324 | 2 | | no evidence |
| *Tuberosa III* | *S. venturii* Hawkes and Hjert. | vnt | 250, 894, 896, 993, | GLKS 32794, CGN 17761, cgn961508, CGN 17755 | 4 | | weak evidence |
| *Tuberosa III* | *S. vernei* Bitter and Wittm. | vrn | 895*, 897, 898, 899, 900, 901, 902, 903, 904, 905$, 979, 980, 981, 982, 983, 984, 985, 986 | CGN 17762*, CGN 22728, CGN 18111, CGN 21350, CGN 22345, CGN 18112, CGN 18114, CGN 23039, CGN 18278, CGN 17836$, CGN 18110, CGN 21315, CGN 17995, CGN 18113, CGN 18115, CGN 23516, CGN 18277, cgn963094 | 17 | | weak evidence for combination with *S. vernei* subsp. *ballsii* |
| *Tuberosa III* | *S. vernei* subsp. *ballsii* (Hawkes) Hawkes and Hjert. | bal | 906, 907, 908 | CGN 17992, CGN 17993, CGN 17994 | 3 | | weak evidence for combination with *S. vernei* |
| *Tuberosa I* | *S. verrucosum* Schtdl. | ver | 393, 825*, 909, 910, 911, 912, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 988, 989, 990 | CGN 17768, CGN 18100*, CGN 22326, CGN 22374, CGN 17764, CGN 20567, CGN 17769, CGN 17765, CGN 17773, CGN 17771, CGN 17766, CGN 17770, CGN 17772, cgn960832, cgn960833, CGN 20566, CGN 23017, CGN 17767, CGN 17774 | 19 | | evidence |
| *Conicibaccata* | *S. violaceimarmoratum* Bitter | vio | 924, 925, 926, | CGN 18296, CGN 20647, CGN 22878 | 3 | | no evidence |
| *Tuberosa III* | *S. virgultorum* (Bitter) Cárdenas and Hawkes | vrg | 927, 928, 929, 930, 931, 932$ | cgn962448, CGN 17775, cgn962072, CGN 20615, cgn962077, CGN 20652$ | 5 | | no evidence |

*Table 3. (Continued) Information on species labels and accessions used in the analyses and suggestions for species status.*

| Series according to Hawkes (1990) | Species | Species abbreviation | Accessions code | Source codes (genebank) | Total nr. of accessions in 566 dataset | taxonomic remarks | Evidence for species status according to the authors |
|---|---|---|---|---|---|---|---|
| *Tuberosa II* | *S. weberbaueri* | wbr | 254, 300 | GLKS 32725, CPC 6032 | 2 | | weak evidence for combination with *S. sandemanii* |
| *Tuberosa III* | *S. x doddsii* Correll | dds | 144, 145, 146, 588, 589, | GLKS 32882, GLKS 32883, GLKS 32880, CGN 20661, CGN 18359 | 5 | | no evidence |
| *Tuberosa III* | *S. x subandigena* Hawkes | sub | 222 | GLKS 30722 | 1 | synonym of | only one accession |
| *Tuberosa III* | *S. x rechei* Hawkes and Hjert. | rch | 35 | PI 558227 | 1 | | only one accession |
| *Tuberosa III* | *S. x ruiz-lealii* Brücher | rzl | 802 | CGN 18117 | 1 | | only one accession |
| *Tuberosa III* | *S. x setulosistylum* Bitter | stl | 214, 811 | GLKS 31014, CGN 20655 | 2 | | no evidence |
| *Tuberosa III* | *S. x sucrense* Hawkes | scr | 391, 843, 844, 845, 846, 847, 848, 849, 850 | CGN 18205, CGN 20628, CGN 20630, CGN 20631, CGN 18187, CGN 20634, CGN 22350, CGN 18206, CGN 18105 | 9 | | no evidence |
| *Yungasensa* | *S. yungasense* Hawkes | yun | 98, 934$, 935, 936 | PI 614703, CGN 18336$, CGN 20677, CGN 20676 | 3 | | no evidence |

***Additional file 1. Information on the 566 South American accessions used in the analyses: their origin and partial membership to each of 16 clusters in STRUCTURE run86.***

| species | code | collection number | country | cluster | Partial membership of each cluster | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *raphanifolium* | RAP206 | OCH 2061 | PER | 1 | 0.003 | 0.002 | 0.956 | 0.004 | 0.002 | 0.002 | 0.005 | 0.005 | 0.001 | 0.007 | 0.001 | 0.002 | 0.003 | 0.001 | 0.003 | 0.003 |
| *raphanifolium* | RAP290 | UGN 3835 | PER | 1 | 0.003 | 0.001 | 0.929 | 0.039 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.004 | 0.001 | 0.002 | 0.003 | 0.005 | 0.002 | 0.002 |
| *raphanifolium* | RAP291 | COR P 218 | PER | 1 | 0.009 | 0.002 | 0.93 | 0.006 | 0.002 | 0.001 | 0.023 | 0.003 | 0.003 | 0.005 | 0.001 | 0.006 | 0.004 | 0.003 | 0.004 | 0.002 |
| *raphanifolium* | RAP380 | PEH 1529 | PER | 1 | 0.006 | 0.004 | 0.946 | 0.003 | 0.008 | 0.007 | 0.003 | 0.002 | 0.003 | 0.002 | 0.001 | 0.005 | 0.002 | 0.004 | 0.003 | 0.001 |
| *raphanifolium* | RAP795 | HAW 694 | PER | 1 | 0.001 | 0.003 | 0.971 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.004 | 0.003 | 0.002 | 0.001 | 0.003 |
| *raphanifolium* | RAP791 | HAW 2470 | PER | 1 | 0.002 | 0.002 | 0.968 | 0.002 | 0.002 | 0.001 | 0.003 | 0.003 | 0.001 | 0.002 | 0.001 | 0.003 | 0.003 | 0.003 | 0.002 | 0.002 |
| *raphanifolium* | RAP792 | HOE BGRC 53019 | PER | 1 | 0.001 | 0.001 | 0.881 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 |
| *raphanifolium* | RAP793 | HVHL 5421 | PER | 1 | 0.004 | 0.001 | 0.773 | 0.002 | 0.003 | 0.002 | 0.009 | 0.181 | 0.001 | 0.005 | 0.002 | 0.002 | 0.002 | 0.006 | 0.003 | 0.003 |
| *raphanifolium* | RAP794 | HVHL 5452 | PER | 1 | 0.003 | 0.001 | 0.949 | 0.016 | 0.004 | 0.001 | 0.004 | 0.004 | 0.003 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 | 0.002 | 0.004 |
| *raphanifolium* | RAP797 | OCH S- 58 | PER | 1 | 0.001 | 0.008 | 0.974 | 0.001 | 0.001 | 0.007 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 |
| *raphanifolium* | RAP798 | PEH 1521 | PER | 1 | 0.004 | 0.002 | 0.961 | 0.002 | 0.003 | 0.002 | 0.004 | 0.002 | 0.002 | 0.002 | 0.001 | 0.003 | 0.004 | 0.002 | 0.003 | 0.002 |
| *raphanifolium* | RAP799 | ROR 762 | PER | 1 | 0.003 | 0.008 | 0.867 | 0.004 | 0.004 | 0.002 | 0.004 | 0.003 | 0.002 | 0.008 | 0.001 | 0.02 | 0.002 | 0.004 | 0.003 | 0.003 |
| *raphanifolium* | RAP800 | ROR 775 | PER | 1 | 0.002 | 0.001 | 0.977 | 0.001 | 0.002 | 0.001 | 0.001 | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 |
| *raphanifolium* | RAP801 | ROR 975 | PER | 1 | 0.002 | 0.002 | 0.961 | 0.006 | 0.002 | 0.003 | 0.001 | 0.005 | 0.002 | 0.002 | 0.001 | 0.01 | 0.004 | 0.005 | 0.002 | 0.003 |
| *raphanifolium* | RAP976 | ROR 183 | BOL | 1 | 0.001 | 0.001 | 0.96 | 0.002 | 0.001 | 0.005 | 0.002 | 0.001 | 0.001 | 0.017 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 |
| *unknown species* | SPEC574 | HHCH 5127 | PER | 1 | 0.3 | 0.028 | 0.351 | 0.002 | 0.013 | 0.006 | 0.006 | 0.019 | 0.009 | 0.004 | 0.032 | 0.097 | 0.122 | 0.003 | 0.009 | 0.004 |
| *chacoense* | SPEC262 | EBS 2084 | CHL | 2 | 0.004 | 0.001 | 0.018 | 0.909 | 0.005 | 0.003 | 0.013 | 0.01 | 0.004 | 0.006 | 0.001 | 0.004 | 0.006 | 0.009 | 0.004 | 0.004 |
| *macroplosum* | MCP23 | RSSV 931 | MEX | 2 | 0.017 | 0.007 | 0.003 | 0.022 | 0.004 | 0.004 | 0.009 | 0.003 | 0.005 | 0.003 | 0.906 | 0.008 | 0.008 | 0.003 | 0.016 | 0.022 |
| *macroplosum* | MCP74 | RSSV 932 | MEX | 2 | 0.009 | 0.008 | 0.002 | 0.002 | 0.002 | 0.005 | 0.005 | 0.002 | 0.003 | 0.003 | 0.929 | 0.009 | 0.01 | 0.002 | 0.008 | 0.002 |
| *pampaserse* | SPEC287 | CPC 7326 | UNKNOWN | 2 | 0.004 | 0.001 | 0.001 | 0.005 | 0.002 | 0.003 | 0.003 | 0.005 | 0.003 | 0.044 | 0.015 | 0.003 | 0.004 | 0.004 | 0.005 | 0.897 |
| *spegazzinii* | VER825 | HOHH 6079 | ARG | 2 | 0.019 | 0.003 | 0.001 | 0.003 | 0.006 | 0.013 | 0.003 | 0.002 | 0.014 | 0.002 | 0.9 | 0.001 | 0.003 | 0.025 | 0.003 | 0.002 |
| *verrucosum* | VER363 | BLS 5828 | MEX | 2 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 | 0.005 | 0.97 | 0.001 | 0.003 | 0.002 | 0.001 | 0.002 |
| *verrucosum* | VER505 | HAW 341 | MEX | 2 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.984 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER910 | HAW 343 | MEX | 2 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.985 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER911 | HAW 756 | MEX | 2 | 0.001 | 0.001 | 0.001 | 0 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.987 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER912 | HAW 1350 | MEX | 2 | 0.013 | 0.006 | 0.003 | 0.001 | 0.003 | 0.002 | 0.014 | 0.003 | 0.007 | 0.004 | 0.864 | 0.009 | 0.007 | 0.006 | 0.055 | 0.003 |
| *verrucosum* | VER914 | HAW 1526 | MEX | 2 | 0.003 | 0.004 | 0.002 | 0.001 | 0.002 | 0.003 | 0.004 | 0.004 | 0.003 | 0.003 | 0.957 | 0.002 | 0.006 | 0.002 | 0.002 | 0.002 |
| *verrucosum* | VER915 | HAW 1532 | MEX | 2 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.985 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER916 | HAW 1542 | MEX | 2 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.985 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER917 | HAW 1548 | MEX | 2 | 0.003 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.003 | 0.966 | 0.001 | 0.005 | 0.011 | 0.001 | 0.002 |
| *verrucosum* | VER918 | HAW 2246 | MEX | 2 | 0.005 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.009 | 0.002 | 0.001 | 0.967 | 0.002 | 0.003 | 0.001 | 0.005 | 0.001 |
| *verrucosum* | VER919 | HHLs 1546 | MEX | 2 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.985 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER920 | UGN 1269 | MEX | 2 | 0.003 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.979 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 |
| *verrucosum* | VER921 | WAC 3320 | MEX | 2 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.982 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER922 | WAC 3321 | MEX | 2 | 0.002 | 0.001 | 0.001 | 0 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.987 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER923 | WAC 3323 | MEX | 2 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.985 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER988 | COR 14217b | MEX | 2 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.987 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER989 | COR 14252 | MEX | 2 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.985 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *verrucosum* | VER990 | PET 925 | MEX | 2 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.983 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 |
| *kurtzianum* | KTZ275 | HHR 3152 x 3384 | ARG | 3 | 0.002 | 0.955 | 0.001 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.013 | 0.004 | 0.001 | 0.003 | 0.008 | 0.001 | 0.001 | 0.002 |
| *kurtzianum* | KTZ276 | HHR 3383 x 3384 | UNKNOWN | 3 | 0.01 | 0.917 | 0.001 | 0.002 | 0.003 | 0.003 | 0.003 | 0.002 | 0.012 | 0.003 | 0.002 | 0.007 | 0.003 | 0.002 | 0.029 | 0.002 |
| *kurtzianum* | KTZ675 | OKA 4265 | ARG | 3 | 0.004 | 0.781 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.001 | 0.171 | 0.006 | 0.001 | 0.003 | 0.002 | 0.002 | 0.008 | 0.011 |
| *kurtzianum* | KTZ676 | OKA 5026 | ARG | 3 | 0.003 | 0.969 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.003 | 0.006 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 |
| *kurtzianum* | KTZ677 | OKA 6004 | ARG | 3 | 0.003 | 0.954 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.004 | 0.004 | 0.001 | 0.004 | 0.002 | 0.018 | 0.001 | 0.001 |
| *kurtzianum* | KTZ678 | OKA 6139 | ARG | 3 | 0.002 | 0.959 | 0.002 | 0.003 | 0.004 | 0.002 | 0.003 | 0.002 | 0.002 | 0.003 | 0.001 | 0.006 | 0.003 | 0.002 | 0.002 | 0.004 |
| *kurtzianum* | KTZ995 | HJT 6316 | ARG | 3 | 0.002 | 0.818 | 0.001 | 0.001 | 0.003 | 0.146 | 0.002 | 0.002 | 0.003 | 0.003 | 0.001 | 0.004 | 0.007 | 0.002 | 0.001 | 0.003 |
| *maglia* | MAG75 | COR C 1 | CHL | 3 | 0.004 | 0.564 | 0.002 | 0.001 | 0.003 | 0.407 | 0.002 | 0.003 | 0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 |
| *maglia* | MAG76 | SCo 4310 | CHL | 3 | 0.004 | 0.759 | 0.001 | 0.001 | 0.005 | 0.211 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.004 | 0.001 |
| *neorossii* | OKA365 | OKA 4392 | ARG | 3 | 0.003 | 0.864 | 0.019 | 0.002 | 0.078 | 0.009 | 0.004 | 0.002 | 0.004 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.005 |
| *okadae* | OKA283 | HOHH 6034 | ARG | 3 | 0.003 | 0.931 | 0.002 | 0.002 | 0.002 | 0.004 | 0.015 | 0.003 | 0.001 | 0.003 | 0.002 | 0.008 | 0.012 | 0.002 | 0.009 | 0.002 |
| *okadae* | OKA366 | HOHH 6033 | ARG | 3 | 0.003 | 0.95 | 0.002 | 0.001 | 0.001 | 0.004 | 0.004 | 0.004 | 0.001 | 0.002 | 0.001 | 0.002 | 0.017 | 0.002 | 0.002 | 0.002 |
| *okadae* | OKA367 | OKA 4908 | ARG | 3 | 0.004 | 0.948 | 0.002 | 0.001 | 0.002 | 0.007 | 0.01 | 0.003 | 0.002 | 0.002 | 0.001 | 0.005 | 0.005 | 0.002 | 0.004 | 0.002 |
| *okadae* | OKA368 | OKA 4388 x 4404 | ARG | 3 | 0.002 | 0.961 | 0.006 | 0.002 | 0.003 | 0.003 | 0.001 | 0.002 | 0.001 | 0.004 | 0.001 | 0.002 | 0.003 | 0.001 | 0.002 | 0.006 |
| *okadae* | OKA740 | OKA 4388 x 4392 | ARG | 3 | 0.004 | 0.971 | 0.001 | 0.001 | 0.005 | 0.002 | 0.001 | 0.001 | 0.004 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 |
| *okadae* | OKA959 | HHR 3741 | ARG | 3 | 0.002 | 0.969 | 0.001 | 0.001 | 0.002 | 0.003 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.005 | 0.001 | 0.004 | 0.001 |
| *rechei* | RCH35 | SCI 4572 | ARG | 3 | 0.006 | 0.86 | 0.002 | 0.009 | 0.009 | 0.055 | 0.01 | 0.002 | 0.011 | 0.002 | 0.003 | 0.023 | 0.001 | 0.004 | 0.003 | 0.001 |
| *sanctae-rosae* | VNT993 | EBS 438 | ARG | 3 | 0.02 | 0.715 | 0.067 | 0.001 | 0.004 | 0.005 | 0.004 | 0.006 | 0.114 | 0.003 | 0.001 | 0.035 | 0.009 | 0.004 | 0.009 | 0.001 |
| *spegazzinii* | SPG386 | PEH 332 x HAW 24 | ARG | 3 | 0.262 | 0.459 | 0.002 | 0.002 | 0.071 | 0.028 | 0.036 | 0.004 | 0.087 | 0.002 | 0.003 | 0.021 | 0.012 | 0.004 | 0.005 | 0.002 |
| *venturii* | VNT250 | HAW 841 | ARG | 3 | 0.017 | 0.941 | 0.002 | 0.001 | 0.006 | 0.003 | 0.005 | 0.004 | 0.004 | 0.003 | 0.003 | 0.001 | 0.004 | 0.002 | 0.004 | 0.001 |
| *venturii* | VNT894 | EBS 457 | ARG | 3 | 0.027 | 0.614 | 0.005 | 0.008 | 0.011 | 0.007 | 0.008 | 0.008 | 0.029 | 0.004 | 0.002 | 0.109 | 0.058 | 0.005 | 0.106 | 0.002 |
| *venturii* | VNT896 | OKA 4392 x 4404 | ARG | 3 | 0.002 | 0.97 | 0.002 | 0.002 | 0.003 | 0.003 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.002 |
| *maglia* | MAG359 | HJR 532 | CHL | 4 | 0.008 | 0.085 | 0.003 | 0.073 | 0.006 | 0.741 | 0.005 | 0.018 | 0.004 | 0.006 | 0.003 | 0.012 | 0.044 | 0.003 | 0.005 | 0.004 |
| *maglia* | MAG688 | CPC 2057 | ARG | 4 | 0.005 | 0.01 | 0.016 | 0.032 | 0.008 | 0.826 | 0.008 | 0.013 | 0.004 | 0.003 | 0.001 | 0.012 | 0.003 | 0.052 | 0.003 | 0.002 |
| *microdontum* | MCD360 | OKA 4478 | ARG | 4 | 0.003 | 0.003 | 0.001 | 0.002 | 0.005 | 0.956 | 0.002 | 0.004 | 0.002 | 0.002 | 0.001 | 0.002 | 0.007 | 0.008 | 0.003 | 0.001 |
| *microdontum* | MCD707 | HHA 6502 | BOL | 4 | 0.007 | 0.004 | 0.002 | 0.003 | 0.093 | 0.784 | 0.003 | 0.002 | 0.009 | 0.002 | 0.001 | 0.01 | 0.011 | 0.014 | 0.006 | 0.008 |
| *microdontum* | MCD708 | HHA 6531 | BOL | 4 | 0.002 | 0.002 | 0.003 | 0.001 | 0.002 | 0.957 | 0.002 | 0.003 | 0.002 | 0.005 | 0.001 | 0.004 | 0.01 | 0.001 | 0.003 | 0.002 |
| *microdontum* | MCD958 | HHA 6650 | BOL | 4 | 0.002 | 0.003 | 0.002 | 0.002 | 0.003 | 0.948 | 0.006 | 0.002 | 0.01 | 0.008 | 0.001 | 0.001 | 0.004 | 0.001 | 0.002 | 0.008 |
| *microdontum* | MCD959 | HHR 3777 | ARG | 4 | 0.004 | 0.006 | 0.016 | 0.001 | 0.002 | 0.953 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.003 |
| *microdontum gigantophyllum* | GIG361 | EBS 2679 CAR | BOL | 4 | 0.002 | 0.005 | 0.001 | 0.001 | 0.002 | 0.972 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.005 | 0.001 | 0.002 | 0.002 |
| *microdontum gigantophyllum* | GIG362 | HOHH 6000 | ARG | 4 | 0.002 | 0.003 | 0.001 | 0.002 | 0.004 | 0.965 | 0.003 | 0.001 | 0.001 | 0.004 | 0.001 | 0.002 | 0.006 | 0.001 | 0.002 | 0.002 |
| *microdontum gigantophyllum* | GIG710 | HAM 174 | BOL | 4 | 0.004 | 0.004 | 0.001 | 0.003 | 0.005 | 0.947 | 0.003 | 0.009 | 0.002 | 0.001 | 0.002 | 0.006 | 0.006 | 0.001 | 0.004 | 0.001 |
| *microdontum gigantophyllum* | GIG711 | HAM 175 | BOL | 4 | 0.005 | 0.002 | 0.001 | 0.001 | 0.003 | 0.964 | 0.002 | 0.001 | 0.008 | 0.001 | 0.001 | 0.002 | 0.006 | 0.001 | 0.001 | 0.001 |
| *microdontum gigantophyllum* | GIG712 | HAM 177 | BOL | 4 | 0.001 | 0.003 | 0.001 | 0.001 | 0.001 | 0.979 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 | 0.002 |
| *microdontum gigantophyllum* | GIG713 | HHR 3881 | ARG | 4 | 0.002 | 0.004 | 0.003 | 0.002 | 0.003 | 0.818 | 0.002 | 0.078 | 0.002 | 0.057 | 0.001 | 0.001 | 0.002 | 0.017 | 0.002 | 0.002 |
| *microdontum gigantophyllum* | GIG714 | HOHH 6012 | ARG | 4 | 0.009 | 0.004 | 0.001 | 0.001 | 0.001 | 0.931 | 0.002 | 0.02 | 0.01 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.004 | 0.009 |
| *microdontum gigantophyllum* | GIG715 | HPR 293 | ARG | 4 | 0.003 | 0.002 | 0.001 | 0.002 | 0.002 | 0.962 | 0.003 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.004 | 0.002 | 0.008 | 0.004 |
| *microdontum gigantophyllum* | GIG956 | COR A 705 | ARG | 4 | 0.003 | 0.001 | 0.001 | 0.001 | 0.001 | 0.957 | 0.008 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.005 | 0.002 | 0.002 |
| *microdontum gigantophyllum* | GIG957 | HAM 179 | BOL | 4 | 0.002 | 0.002 | 0.001 | 0.003 | 0.002 | 0.976 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.008 | 0.002 | 0.001 | 0.001 | 0.001 |
| *microdontum gigantophyllum* | GIG960 | HOF 1976 | ARG | 4 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.982 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 |
| *microdontum gigantophyllum* | GIG961 | OKA 2910 | ARG | 4 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.979 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.004 | 0.002 | 0.001 | 0.001 |
| *microdontum gigantophyllum* | GIG962 | OKA 4820 | ARG | 4 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.98 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 |
| *microdontum gigantophyllum* | GIG963 | OKA 4897 | ARG | 4 | 0.003 | 0.003 | 0.001 | 0.002 | 0.002 | 0.968 | 0.003 | 0.002 | 0.003 | 0.001 | 0.001 | 0.001 | 0.003 | 0.002 | 0.002 | 0.003 |
| *microdontum gigantophyllum* | GIG964 | OKA 5913 | ARG | 4 | 0.004 | 0.008 | 0.003 | 0.006 | 0.002 | 0.828 | 0.002 | 0.045 | 0.007 | 0.029 | 0.001 | 0.002 | 0.033 | 0.005 | 0.004 | 0.02 |
| *microdontum gigantophyllum* | GIG965 | OKA 6327 | ARG | 4 | 0.001 | 0.001 | 0.005 | 0.001 | 0.003 | 0.973 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.003 | 0.002 | 0.003 |
| *microdontum gigantophyllum* | GIG966 | OKA 6840 | ARG | 4 | 0.002 | 0.004 | 0.001 | 0.001 | 0.015 | 0.935 | 0.002 | 0.002 | 0.001 | 0.005 | 0.002 | 0.018 | 0.002 | 0.005 | 0.002 | 0.003 |
| *microdontum gigantophyllum* | GIG967 | PEH 364 | ARG | 4 | 0.001 | 0.004 | 0.001 | 0.001 | 0.002 | 0.973 | 0.002 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.004 | 0.001 | 0.001 | 0.001 |
| *microdontum gigantophyllum* | MCD994 | EBS 1091 | ARG | 4 | 0.006 | 0.005 | 0.001 | 0.002 | 0.016 | 0.942 | 0.013 | 0.001 | 0.002 | 0.003 | 0.001 | 0.001 | 0.004 | 0.001 | 0.001 | 0.001 |
| *spegazzinii* | SPG824 | HHR 3995 | ARG | 4 | 0.044 | 0.139 | 0.001 | 0.004 | 0.033 | 0.355 | 0.031 | 0.02 | 0.273 | 0.008 | 0.001 | 0.008 | 0.063 | 0.002 | 0.004 | 0.016 |
| *medians* | MED163 | HAW 2463 | PER | 5 | 0.001 | 0.001 | 0.001 | 0.001 | 0.008 | 0.001 | 0.002 | 0.007 | 0.001 | 0.003 | 0.001 | 0.008 | 0.004 | 0.954 | 0.004 | 0.003 |
| *medians* | MED691 | EBS 3185 | PER | 5 | 0.006 | 0.002 | 0.002 | 0.014 | 0.003 | 0.014 | 0.01 | 0.003 | 0.003 | 0.003 | 0.001 | 0.008 | 0.002 | 0.878 | 0.004 | 0.046 |
| *medians* | MED692 | HAW 2489 | PER | 5 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.978 | 0.001 | 0.003 |
| *medians* | MED693 | OCH 5032 | PER | 5 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.98 | 0.001 | 0.001 |
| *medians* | MED694 | ROR 1029 | PER | 5 | 0.149 | 0.006 | 0.001 | 0.009 | 0.002 | 0.004 | 0.021 | 0.002 | 0.011 | 0.005 | 0.008 | 0.006 | 0.008 | 0.745 | 0.021 | 0.002 |
| *medians* | MED695 | VIL 211 | PER | 5 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.978 | 0.001 | 0.002 |
| *sandemanii* | SND808 | ROR 726 | PER | 5 | 0.006 | 0.008 | 0.012 | 0.001 | 0.002 | 0.031 | 0.015 | 0.005 | 0.002 | 0.002 | 0.002 | 0.003 | 0.002 | 0.906 | 0.002 | 0.001 |
| *sandemanii* | SND93 | SS 7250 | PER | 5 | 0.008 | 0.009 | 0.002 | 0.003 | 0.004 | 0.026 | 0.012 | 0.006 | 0.003 | 0.002 | 0.002 | 0.004 | 0.002 | 0.915 | 0.002 | 0.001 |
| *sandemanii* | SND94 | SS 7252 | PER | 5 | 0.003 | 0.005 | 0.004 | 0.023 | 0.004 | 0.047 | 0.002 | 0.003 | 0.002 | 0.005 | 0.001 | 0.017 | 0.006 | 0.874 | 0.002 | 0.001 |
| *violaceomarmoratum* | VIO924 | GND BGRC 53015 | BOL | 5 | 0.005 | 0.007 | 0.009 | 0.031 | 0.008 | 0.003 | 0.078 | 0.019 | 0.005 | 0.233 | 0.001 | 0.025 | 0.004 | 0.348 | 0.007 | 0.217 |
| *weberbaueri* | SPEC253 | GLK 159 1 | PER | 5 | 0.021 | 0.006 | 0.017 | 0.007 | 0.01 | 0.008 | 0.008 | 0.04 | 0.007 | 0.01 | 0.388 | 0.003 | 0.02 | 0.15 | 0.007 | 0.317 |
| *weberbaueri* | WBR254 | OCH 5081 | PER | 5 | 0.012 | 0.004 | 0.01 | 0.002 | 0.009 | 0.005 | 0.004 | 0.012 | 0.004 | 0.002 | 0.002 | 0.003 | 0.007 | 0.923 | 0.003 | 0.001 |
| *weberbaueri* | WBR350 | HAW CPC 6032 | PER | 5 | 0.004 | 0.003 | 0.004 | 0.002 | 0.002 | 0.01 | 0.006 | 0.002 | 0.003 | 0.009 | 0.001 | 0.135 | 0.002 | 0.809 | 0.001 | 0.004 |

***Additional file 1. (Continued) Information on the 566 South American accessions used in the analyses: their origin and partial membership to each of 16 clusters in STRUCTURE run86.***

| species | code | collection number | country | cluster | Partial membership of each cluster | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *commersonii* | CMM1017 | HOF BALC 71385 | ARG | 8 | 0.003 | 0.005 | 0.002 | 0.001 | 0.001 | 0.007 | 0.002 | 0.001 | 0.004 | 0.001 | 0.001 | 0.004 | 0.965 | 0.001 | 0.001 | 0.002 |
| *commersonii* | CMM1019 | OKA 5107 x 5075 | ARG | 8 | 0.003 | 0.006 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.001 | 0.001 | 0.972 | 0.001 | 0.003 | 0.001 |
| *commersonii* | CMM1027 | HAW 195 | ARG? | 8 | 0.005 | 0.002 | 0.001 | 0.002 | 0.002 | 0.053 | 0.003 | 0.003 | 0.003 | 0.001 | 0.001 | 0.003 | 0.913 | 0.002 | 0.004 | 0.001 |
| *commersonii* | CMM1028 | HAW 184 x 185 | UNKNOWN | 8 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 | 0.96 | 0.001 | 0.001 | 0.002 |
| *commersonii* | CMM1039 | HAW 184 x 159 | ARG? | 8 | 0.009 | 0.019 | 0.001 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.008 | 0.001 | 0.001 | 0.003 | 0.938 | 0.003 | 0.006 | 0.001 |
| *commersonii* | CMM265 | MDC 2 x MDC 3 | BRA | 8 | 0.003 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.023 | 0.003 | 0.004 | 0.015 | 0.001 | 0.003 | 0.922 | 0.007 | 0.004 | 0.005 |
| *commersonii* | CMM575 | OKA 4584 | ARG | 8 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.981 | 0.001 | 0.001 | 0.001 |
| *commersonii* | CMM576 | OKA 5138 | ARG | 8 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.006 | 0.001 | 0.001 | 0.001 | 0.002 | 0.973 | 0.001 | 0.001 | 0.005 |
| *commersonii* | CMM577 | OKA 5180 | ARG | 8 | 0.002 | 0.003 | 0.001 | 0.001 | 0.003 | 0.001 | 0.003 | 0.002 | 0.003 | 0.003 | 0.001 | 0.002 | 0.968 | 0.001 | 0.002 | 0.002 |
| *commersonii* | CMM578 | OKC 7254 | ARG | 8 | 0.009 | 0.005 | 0.001 | 0.001 | 0.003 | 0.001 | 0.011 | 0.005 | 0.015 | 0.007 | 0.001 | 0.003 | 0.921 | 0.005 | 0.004 | 0.009 |
| *commersonii commersonii* | CMM1018 | OKA 4583 | ARG | 8 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.984 | 0.001 | 0.001 | 0.001 |
| *commersonii commersonii* | CMM1050 | OKA 5258b | ARG | 8 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.004 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.981 | 0.001 | 0.001 | 0.001 |
| *commersonii malmeanum* | MLM1038 | OKA 7281 | ARG | 8 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.981 | 0.002 | 0.001 | 0.001 |
| *commersonii malmeanum* | MLM1045 | OKA 7292 | ARG | 8 | 0.001 | 0.003 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.976 | 0.001 | 0.002 | 0.003 |
| *commersonii malmeanum* | MLM1058 | OKA 7313 | ARG | 8 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.983 | 0.001 | 0.001 | 0.001 |
| *commersonii malmeanum* | MLM139 | BRU 18 | PRY | 8 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.007 | 0.001 | 0.976 | 0.002 | 0.001 | 0.001 |
| *commersonii malmeanum* | MLM266 | CPC BPC 825 | UNKNOWN | 8 | 0.002 | 0.005 | 0.003 | 0.003 | 0.001 | 0.003 | 0.003 | 0.003 | 0.003 | 0.001 | 0.022 | 0.001 | 0.004 | 0.935 | 0.004 | 0.002 | 0.008 |
| *commersonii malmeanum* | MLM580 | OKA 5071 | ARG | 8 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.981 | 0.001 | 0.001 | 0.002 |
| *commersonii malmeanum* | MLM581 | OKA 7270 | ARG | 8 | 0.001 | 0.002 | 0.001 | 0.041 | 0.002 | 0.001 | 0.007 | 0.015 | 0.001 | 0.002 | 0.023 | 0.001 | 0.854 | 0.012 | 0.003 | 0.033 |
| *scrospicipeum* | ACS100 | OCH S- 96 | PER | 7 | 0.005 | 0.002 | 0.008 | 0.002 | 0.002 | 0.002 | 0.003 | 0.004 | 0.002 | 0.012 | 0.001 | 0.002 | 0.003 | 0.002 | 0.004 | 0.946 |
| *albomarginatum* | ABZ102 | SCLp 5030 | ECU | 7 | 0.004 | 0.003 | 0.001 | 0.001 | 0.002 | 0.005 | 0.003 | 0.001 | 0.005 | 0.002 | 0.001 | 0.001 | 0.005 | 0.002 | 0.002 | 0.96 |
| *albomarginatum* | ABZ103 | SCLp 5032 | ECU | 7 | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.003 | 0.002 | 0.006 | 0.002 | 0.001 | 0.001 | 0.003 | 0.002 | 0.001 | 0.97 |
| *albomarginatum* | ABZ2 | SCLp 5033 | ECU | 7 | 0.005 | 0.004 | 0.001 | 0.004 | 0.004 | 0.006 | 0.004 | 0.001 | 0.008 | 0.003 | 0.001 | 0.004 | 0.005 | 0.003 | 0.95 | |
| *albomarginatum* | ABZ486 | OCHS 11007 | ECU | 7 | 0.002 | 0.003 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.001 | 0.003 | 0.002 | 0.001 | 0.001 | 0.006 | 0.002 | 0.002 | 0.968 |
| *ancophilum* | ACP304 | OCH 12086 | PER | 7 | 0.018 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.005 | 0.003 | 0.01 | 0.013 | 0.001 | 0.004 | 0.001 | 0.014 | 0.001 | 0.922 |
| *augustii* | AGU305 | OCHS 12596 | PER | 7 | 0.003 | 0.001 | 0.015 | 0.006 | 0.002 | 0.002 | 0.002 | 0.013 | 0.002 | 0.011 | 0.001 | 0.01 | 0.001 | 0.002 | 0.003 | 0.927 |
| *chancayense* | CHN1 | OCHS 11251 | PER | 7 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.003 | 0.001 | 0.003 | 0.976 |
| *chancayense* | CHN552 | EBS 2807 | PER | 7 | 0.002 | 0.005 | 0.001 | 0.001 | 0.002 | 0.012 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.962 |
| *chancayense* | CHN553 | OCHS 11250 | PER | 7 | 0.007 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.961 |
| *chavinense* | CHV11 | OCH 12072 | PER | 7 | 0.004 | 0.003 | 0.001 | 0.036 | 0.004 | 0.005 | 0.003 | 0.004 | 0.003 | 0.013 | 0.001 | 0.007 | 0.008 | 0.004 | 0.002 | 0.904 |
| *dolichocremastrum* | DCM147 | OCH 12071 | PER | 7 | 0.004 | 0.002 | 0.001 | 0.06 | 0.003 | 0.006 | 0.002 | 0.002 | 0.004 | 0.002 | 0.001 | 0.008 | 0.001 | 0.006 | 0.002 | 0.898 |
| *dolichocremastrum* | DCM148 | OCH 12074 | PER | 7 | 0.012 | 0.001 | 0.002 | 0.083 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.009 | 0.006 | 0.004 | 0.002 | 0.007 | 0.001 | 0.866 |
| *dolichocremastrum* | DCM149 | OCH 13013 | PER | 7 | 0.003 | 0.004 | 0.004 | 0.031 | 0.003 | 0.008 | 0.002 | 0.001 | 0.002 | 0.004 | 0.001 | 0.004 | 0.002 | 0.002 | 0.002 | 0.925 |
| *dolichocremastrum* | DCM308 | OCH 12083a | PER | 7 | 0.002 | 0.004 | 0.001 | 0.023 | 0.004 | 0.004 | 0.002 | 0.001 | 0.001 | 0.014 | 0.002 | 0.008 | 0.002 | 0.002 | 0.002 | 0.932 |
| *hypocrarthrum* | HCR311 | OCH 11692 | PER | 7 | 0.001 | 0.001 | 0.017 | 0.001 | 0.003 | 0.002 | 0.002 | 0.008 | 0.002 | 0.004 | 0.003 | 0.003 | 0.003 | 0.002 | 0.002 | 0.946 |
| *iramite* | IMT172 | OCH S- 53 | PER | 7 | 0.004 | 0.004 | 0.003 | 0.001 | 0.002 | 0.002 | 0.04 | 0.005 | 0.002 | 0.003 | 0.001 | 0.003 | 0.005 | 0.002 | 0.005 | 0.917 |
| *iramite* | IMT63 | OCH 13346 | PER | 7 | 0.016 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.007 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.005 | 0.001 | 0.003 | 0.954 |
| *iramite* | IMT84 | OCH S- 56 | PER | 7 | 0.002 | 0.001 | 0.003 | 0.001 | 0.002 | 0.002 | 0.003 | 0.004 | 0.002 | 0.003 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.968 |
| *mochiquense* | MCQ716 | EBS 2764 | PER | 7 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.977 |
| *mochiquense* | MCQ717 | K 11273 x 18129 | PER | 7 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.983 |
| *mochiquense* | MCQ718 | OCH 1822 | PER | 7 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.981 |
| *mochiquense* | MCQ719 | OCH SCHICK 138 | PER | 7 | 0.002 | 0.002 | 0.005 | 0.001 | 0.001 | 0.002 | 0.001 | 0.004 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.966 |
| *neocardenasii* | NCD193 | SL 32084.1 | BOL | 7 | 0.004 | 0.003 | 0.002 | 0.152 | 0.016 | 0.002 | 0.004 | 0.007 | 0.002 | 0.215 | 0.002 | 0.022 | 0.002 | 0.003 | 0.011 | 0.552 |
| *neocardenasii* | NCD734 | HHA 6496 | BOL | 7 | 0.003 | 0.003 | 0.002 | 0.078 | 0.029 | 0.002 | 0.002 | 0.006 | 0.002 | 0.184 | 0.001 | 0.057 | 0.002 | 0.003 | 0.009 | 0.618 |
| *scabrifolium* | SCB37 | OCH S- 80 | PER | 7 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.003 | 0.002 | 0.001 | 0.274 | 0.003 | 0.007 | 0.001 | 0.017 | 0.003 | 0.677 |
| *unknown species* | SPEC205 | HAW 2443 | PER | 7 | 0.007 | 0.005 | 0.002 | 0.155 | 0.002 | 0.001 | 0.006 | 0.002 | 0.003 | 0.007 | 0.006 | 0.044 | 0.003 | 0.029 | 0.003 | 0.724 |
| *unknown species* | SPEC310 | OCH 11699 | PER | 7 | 0.004 | 0.003 | 0.007 | 0.12 | 0.009 | 0.227 | 0.411 | 0.009 | 0.005 | 0.057 | 0.006 | 0.011 | 0.004 | 0.01 | 0.006 | 0.02 |
| *unknown species* | SPEC6 | OCH 13009 | PER | 7 | 0.042 | 0.012 | 0.002 | 0.003 | 0.026 | 0.024 | 0.158 | 0.005 | 0.072 | 0.002 | 0.002 | 0.01 | 0.065 | 0.009 | 0.006 | 0.001 |
| *violaceimarmoratum* | VIO925 | HHA 6678 | BOL | 7 | 0.008 | 0.002 | 0.013 | 0.003 | 0.004 | 0.008 | 0.022 | 0.017 | 0.004 | 0.194 | 0.001 | 0.116 | 0.004 | 0.286 | 0.013 | 0.325 |
| *huancabambense* | HCB170 | GLK 155 1 | PER | 8 | 0.002 | 0.001 | 0.001 | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.975 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.003 |
| *huancabambense* | HCB18 | OCH S- 49 | PER | 8 | 0.002 | 0.001 | 0.001 | 0.014 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.934 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.036 |
| *huancabambense* | HCB353 | HUA 966 | PER | 8 | 0.002 | 0.001 | 0.001 | 0.001 | 0.01 | 0.001 | 0.001 | 0.001 | 0.002 | 0.972 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.003 |
| *huancabambense* | HCB354 | OCH 21 | PER | 8 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.001 | 0.001 | 0.001 | 0.979 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.003 |
| *sogarandinum* | SGR215 | OCH S- 54 | PER | 8 | 0.001 | 0.002 | 0.009 | 0.019 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.948 | 0.001 | 0.002 | 0.003 | 0.003 | 0.002 | 0.003 |
| *sogarandinum* | SGR315 | OCH 13006 | PER | 8 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.003 | 0.001 | 0.001 | 0.001 | 0.969 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.011 |
| *sogarandinum* | SGR316 | OCH 13336 | PER | 8 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.981 | 0 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 |
| *sogarandinum* | SGR814 | OCH 1440 | PER | 8 | 0.007 | 0.003 | 0.004 | 0.068 | 0.003 | 0.002 | 0.032 | 0.078 | 0.006 | 0.666 | 0.016 | 0.028 | 0.004 | 0.012 | 0.014 | 0.03 |
| *violaceimarmoratum* | VIO926 | OCHS 11901 | BOL | 8 | 0.007 | 0.013 | 0.104 | 0.024 | 0.004 | 0.002 | 0.037 | 0.01 | 0.006 | 0.308 | 0.001 | 0.038 | 0.286 | 0.092 | 0.002 | 0.086 |
| *hannemanii* | HAN628 | OKA 4372 | ARG | 9 | 0.003 | 0.002 | 0.003 | 0.002 | 0.95 | 0.002 | 0.004 | 0.002 | 0.002 | 0.01 | 0.001 | 0.001 | 0.002 | 0.009 | 0.002 | 0.003 |
| *hannemanii* | HAN629 | OKA 4374 | ARG | 9 | 0.002 | 0.002 | 0.003 | 0.001 | 0.875 | 0.001 | 0.002 | 0.002 | 0.002 | 0.003 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 |
| *hannemanii* | HAN630 | OKA 4383a | ARG | 9 | 0.001 | 0.001 | 0.002 | 0.001 | 0.88 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 |
| *hannemanii* | HAN631 | OKA 4385 x 4372 | ARG | 9 | 0.003 | 0.001 | 0.003 | 0.001 | 0.966 | 0.001 | 0.002 | 0.001 | 0.001 | 0.005 | 0.001 | 0.002 | 0.003 | 0.002 | 0.001 | 0.004 |
| *hannemanii* | HAN632 | OKA 4407 x 4481 | ARG | 9 | 0.004 | 0.001 | 0.002 | 0.001 | 0.936 | 0.002 | 0.002 | 0.003 | 0.005 | 0.024 | 0.001 | 0.002 | 0.003 | 0.003 | 0.003 | 0.005 |
| *hannemanii* | HAN633 | OKA 4481 x 4374 | ARG | 9 | 0.004 | 0.002 | 0.018 | 0.001 | 0.958 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 |
| *spegazzinii* | SPG622 | BRU 46 | ARG | 9 | 0.022 | 0.157 | 0.002 | 0.001 | 0.276 | 0.033 | 0.161 | 0.018 | 0.127 | 0.003 | 0.002 | 0.021 | 0.159 | 0.004 | 0.01 | 0.004 |
| *vernei* | VRN895 | HAW 676 | ARG | 9 | 0.035 | 0.064 | 0.002 | 0.001 | 0.581 | 0.008 | 0.016 | 0.005 | 0.242 | 0.003 | 0.003 | 0.02 | 0.003 | 0.002 | 0.014 | 0.001 |
| *vernei* | VRN897 | HOHH 6030 | ARG | 9 | 0.004 | 0.003 | 0.001 | 0.001 | 0.982 | 0.002 | 0.005 | 0.004 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.007 | 0.002 |
| *vernei* | VRN898 | HOHH 6089 | ARG | 9 | 0.003 | 0.003 | 0.002 | 0.002 | 0.955 | 0.004 | 0.003 | 0.002 | 0.005 | 0.001 | 0.001 | 0.002 | 0.004 | 0.001 | 0.009 | 0.001 |
| *vernei* | VRN899 | HOHH 6088 | ARG | 9 | 0.005 | 0.003 | 0.018 | 0.001 | 0.877 | 0.002 | 0.007 | 0.002 | 0.024 | 0.003 | 0.001 | 0.012 | 0.002 | 0.006 | 0.024 | 0.013 |
| *vernei* | VRN900 | HOHH 6031 x 6086 | ARG | 9 | 0.003 | 0.007 | 0.001 | 0.001 | 0.964 | 0.003 | 0.005 | 0.002 | 0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.004 | 0.001 |
| *vernei* | VRN901 | HOHH 6088 x 6032 | ARG | 9 | 0.002 | 0.006 | 0.003 | 0.001 | 0.945 | 0.006 | 0.003 | 0.004 | 0.003 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.013 | 0.001 |
| *vernei* | VRN902 | OKA 5917 | ARG | 9 | 0.005 | 0.003 | 0.002 | 0.001 | 0.926 | 0.009 | 0.003 | 0.024 | 0.002 | 0.002 | 0.001 | 0.002 | 0.007 | 0.003 | 0.016 | 0.001 |
| *vernei* | VRN903 | OKA 7617 | ARG | 9 | 0.002 | 0.002 | 0.001 | 0.001 | 0.962 | 0.004 | 0.001 | 0.002 | 0.001 | 0.002 | 0.014 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 |
| *vernei* | VRN979 | WAC 4070 | ARG | 9 | 0.002 | 0.002 | 0.001 | 0.001 | 0.974 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.003 | 0.002 | 0.003 | 0.001 |
| *vernei* | VRN980 | HHA 3505 | ARG | 9 | 0.025 | 0.005 | 0.001 | 0.001 | 0.724 | 0.046 | 0.007 | 0.007 | 0.149 | 0.003 | 0.002 | 0.003 | 0.003 | 0.001 | 0.01 | 0.012 |
| *vernei* | VRN981 | OKA 5702 | ARG | 9 | 0.028 | 0.014 | 0.003 | 0.001 | 0.842 | 0.011 | 0.008 | 0.007 | 0.056 | 0.003 | 0.001 | 0.008 | 0.005 | 0.003 | 0.005 | 0.004 |
| *vernei* | VRN982 | OKA 5855 | ARG | 9 | 0.15 | 0.008 | 0.001 | 0.006 | 0.796 | 0.003 | 0.001 | 0.002 | 0.011 | 0.004 | 0.001 | 0.002 | 0.001 | 0.003 | 0.005 | 0.002 |
| *vernei* | VRN983 | OKA 5923 | ARG | 9 | 0.011 | 0.002 | 0.002 | 0.001 | 0.947 | 0.007 | 0.003 | 0.002 | 0.005 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.007 | 0.002 |
| *vernei* | VRN984 | OKA 7502 | ARG | 9 | 0.005 | 0.002 | 0.001 | 0.001 | 0.974 | 0.002 | 0.158 | 0.002 | 0.042 | 0.003 | 0.002 | 0.002 | 0.006 | 0.003 | 0.002 | 0.003 |
| *vernei* | VRN985 | OKA 7614 | ARG | 9 | 0.001 | 0.003 | 0.002 | 0.002 | 0.976 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *vernei* | VRN986 | EBS 189 | ARG | 9 | 0.014 | 0.016 | 0.001 | 0.001 | 0.941 | 0.01 | 0.005 | 0.003 | 0.076 | 0.003 | 0.001 | 0.002 | 0.002 | 0.019 | 0.001 | 0.008 |
| *vernei ballsii* | BAL906 | OKA 4366 | ARG | 9 | 0.002 | 0.008 | 0.001 | 0.001 | 0.965 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.007 | 0.001 | 0.002 | 0.005 |
| *vernei ballsii* | BAL907 | OKA 4367 | ARG | 9 | 0.002 | 0.003 | 0.002 | 0.002 | 0.963 | 0.002 | 0.012 | 0.008 | 0.003 | 0.002 | 0.003 | 0.002 | 0.002 | 0.002 | 0.008 | 0.004 |
| *vernei ballsii* | BAL908 | OKA 4427 | ARG | 9 | 0.009 | 0.013 | 0.002 | 0.001 | 0.886 | 0.021 | 0.014 | 0.015 | 0.004 | 0.002 | 0.002 | 0.002 | 0.006 | 0.001 | 0.004 | 0.017 |
| *virgultorum* | HAN252 | GLK 112 1 | BOL | 9 | 0.002 | 0.003 | 0.001 | 0.001 | 0.985 | 0.002 | 0.006 | 0.001 | 0.004 | 0.004 | 0.001 | 0.003 | 0.003 | 0.001 | 0.002 | 0.002 |
| *abancayense* | ABN423 | OCH 7728 ? | PER | 10 | 0.002 | 0.001 | 0.002 | 0.002 | 0.977 | 0.001 | 0.001 | 0.001 | 0.001 | 0 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 |
| *achacachense* | ACH98 | RWH 3 | BOL | 10 | 0.026 | 0.013 | 0.013 | 0.531 | 0.074 | 0.005 | 0.012 | 0.002 | 0.037 | 0.1 | 0.033 | 0.005 | 0.007 | 0.025 | 0.036 | 0.092 |
| *amabile* | AMC3 | OCH S- 34 | PER | 10 | 0.002 | 0.005 | 0.002 | 0.942 | 0.003 | 0.005 | 0.002 | 0.005 | 0.002 | 0.003 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.022 |
| *amayanum* | AMY302 | OCHS 16263 | PER | 10 | 0.009 | 0.006 | 0.001 | 0.766 | 0.064 | 0.007 | 0.002 | 0.002 | 0.009 | 0.109 | 0.002 | 0.002 | 0.007 | 0.008 | 0.005 | 0.003 |
| *amayanum* | AMY303 | OCHS 16265 | PER | 10 | 0.003 | 0.002 | 0.001 | 0.855 | 0.003 | 0.004 | 0.003 | 0.002 | 0.002 | 0.111 | 0.001 | 0.003 | 0.003 | 0.166 | 0.008 | 0.037 |
| *amboseum* | AMB104 | OCH S- 59 | PER | 10 | 0.004 | 0.003 | 0.004 | 0.768 | 0.004 | 0.002 | 0.007 | 0.018 | 0.005 | 0.002 | 0.002 | 0.004 | 0.004 | 0.001 | 0.172 | 0.001 |
| *amboseum* | AMB105 | UGN 4297 | PER | 10 | 0.007 | 0.013 | 0.002 | 0.905 | 0.002 | 0.001 | 0.01 | 0.004 | 0.005 | 0.001 | 0.002 | 0.014 | 0.002 | 0.003 | 0.027 | 0.002 |
| *amposinum* | AMB467 | OCH PI 458405 | PER | 10 | 0.003 | 0.002 | 0.004 | 0.944 | 0.003 | 0.003 | 0.003 | 0.003 | 0.001 | 0.006 | 0.001 | 0.003 | 0.004 | 0.002 | 0.003 | 0.014 |
| *aymarasense* | AYM5 | SS 7257 | PER | 10 | 0.006 | 0.003 | 0.001 | 0.914 | 0.003 | 0.005 | 0.004 | 0.006 | 0.002 | 0.004 | 0.004 | 0.002 | 0.008 | 0.003 | 0.005 | 0.022 |
| *bukasovii* | BUK328 | OCH WAC 3342 | PER | 10 | 0.004 | 0.003 | 0.004 | 0.925 | 0.005 | 0.008 | 0.006 | 0.004 | 0.002 | 0.006 | 0.004 | 0.002 | 0.006 | 0.01 | 0.008 | 0.003 |
| *bukasovii* | BUK519 | HHCH 5220 | PER | 10 | 0.015 | 0.003 | 0.002 | 0.899 | 0.004 | 0.001 | 0.01 | 0.007 | 0.016 | 0.003 | 0.003 | 0.008 | 0.004 | 0.003 | 0.019 | 0.003 |
| *bukasovii* | BUK512 | HHCH 5230 | PER | 10 | 0.005 | 0.004 | 0.006 | 0.719 | 0.003 | 0.002 | 0.003 | 0.014 | 0.014 | 0.041 | 0.002 | 0.011 | 0.117 | 0.042 | 0.004 | 0.011 |
| *bukasovii* | BUK514 | OCH S- 64 | PER | 10 | 0.007 | 0.009 | 0.002 | 0.765 | 0.006 | 0.014 | 0.009 | 0.003 | 0.004 | 0.012 | 0.002 | 0.006 | 0.026 | 0.126 | 0.008 | 0.006 |
| *bukasovii* | BUK955 | HHCH 5225 | PER | 10 | 0.011 | 0.008 | 0.002 | 0.928 | 0.016 | 0.002 | 0.002 | 0.005 | 0.005 | 0.001 | 0.011 | 0.005 | 0.003 | 0.002 | 0.002 | 0.001 |
| *bukasovii* | BUK971 | HHCH 5229 | PER | 10 | 0.022 | 0.011 | 0.006 | 0.718 | 0.03 | 0.003 | 0.005 | 0.022 | 0.014 | 0.013 | 0.002 | 0.008 | 0.009 | 0.028 | 0.015 | 0.006 |
| *canasense* | CAN260 | PEH 1527 | PER | 10 | 0.005 | 0.002 | 0.005 | 0.926 | 0.005 | 0.003 | 0.012 | 0.008 | 0.011 | 0.004 | 0.001 | 0.003 | 0.004 | 0.002 | 0.008 | 0.007 |

***Additional file 1. (Continued) Information on the 566 South American accessions used in the analyses: their origin and partial membership to each of 16 clusters in STRUCTURE run86.***

| species | code | collection number | country | cluster | Partial membership of each cluster | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *canasense* | CAN526 | HHCH 5064a | PER | 10 | 0.001 | 0.001 | 0.001 | 0.969 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.01 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 |
| *canasense* | CAN527 | HHCH 5107 | PER | 10 | 0.005 | 0.004 | 0.003 | 0.705 | 0.014 | 0.004 | 0.003 | 0.02 | 0.002 | 0.019 | 0.001 | 0.006 | 0.023 | 0.157 | 0.003 | 0.032 |
| *canasense* | CAN528 | HHCH 5233 | PER | 10 | 0.001 | 0.001 | 0.002 | 0.979 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.003 |
| *canasense* | CAN529 | ROR 795 | PER | 10 | 0.006 | 0.003 | 0.001 | 0.913 | 0.001 | 0.005 | 0.008 | 0.003 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.045 | 0.001 | 0.004 |
| *canasense* | CAN952 | ASL 6199 | PER | 10 | 0.001 | 0.001 | 0.002 | 0.981 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 |
| *canasense* | CAN953 | COR P 223 | PER | 10 | 0.02 | 0.036 | 0.004 | 0.837 | 0.002 | 0.018 | 0.017 | 0.005 | 0.022 | 0.002 | 0.001 | 0.014 | 0.002 | 0.005 | 0.006 | 0.006 |
| *candolleanum* | CND530 | OCHS 11897 | BOL | 10 | 0.008 | 0.004 | 0.002 | 0.91 | 0.013 | 0.002 | 0.003 | 0.002 | 0.013 | 0.004 | 0.001 | 0.006 | 0.002 | 0.006 | 0.004 | 0.021 |
| *candolleanum* | CND531 | VSOA 68 | BOL | 10 | 0.002 | 0.006 | 0.002 | 0.9 | 0.004 | 0.004 | 0.014 | 0.002 | 0.002 | 0.009 | 0.001 | 0.024 | 0.002 | 0.005 | 0.004 | 0.016 |
| *candolleanum* | CND532 | VSOA 75 | BOL | 10 | 0.004 | 0.003 | 0.001 | 0.956 | 0.004 | 0.002 | 0.001 | 0.005 | 0.002 | 0.005 | 0.003 | 0.003 | 0.002 | 0.003 | 0.002 | 0.005 |
| *chilonanum* | CHI12 | SS 7239 | PER | 10 | 0.001 | 0.001 | 0.001 | 0.986 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *coelestipetalum* | COP134 | OCH 7728 | PER | 10 | 0.002 | 0.001 | 0.002 | 0.977 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.005 | 0.001 | 0.002 |
| *coelestipetalum* | COP135 | OCH 13674 | PER | 10 | 0.002 | 0.001 | 0.002 | 0.974 | 0.002 | 0.001 | 0.002 | 0.004 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 |
| *coelestipetalum* | COP306 | OCH 13716 | PER | 10 | 0.003 | 0.007 | 0.001 | 0.949 | 0.001 | 0.006 | 0.002 | 0.005 | 0.002 | 0.001 | 0.01 | 0.001 | 0.001 | 0.003 | 0.006 | 0.002 |
| *coelestipetalum* | COP307 | OCH 14342 | PER | 10 | 0.002 | 0.001 | 0.001 | 0.928 | 0.023 | 0.001 | 0.003 | 0.001 | 0.001 | 0.007 | 0.001 | 0.004 | 0.002 | 0.004 | 0.002 | 0.016 |
| *coelestipetalum* | COP572 | ROR 893 | PER | 10 | 0.001 | 0.005 | 0.005 | 0.974 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 |
| *hoerchinense* | HRO309 | OCH 11335 | PER | 10 | 0.009 | 0.008 | 0.016 | 0.799 | 0.015 | 0.002 | 0.015 | 0.058 | 0.003 | 0.04 | 0.001 | 0.003 | 0.011 | 0.002 | 0.004 | 0.019 |
| *limbanense* | LMB686 | OCH 5123 | PER | 10 | 0.001 | 0.001 | 0.001 | 0.968 | 0.001 | 0.002 | 0.001 | 0.006 | 0.001 | 0.001 | 0.001 | 0.007 | 0.001 | 0.002 | 0.001 | 0.002 |
| *mannaseense* | MRN181 | OCH 13619 | PER | 10 | 0.002 | 0.002 | 0.057 | 0.913 | 0.003 | 0.004 | 0.002 | 0.005 | 0.002 | 0.002 | 0.001 | 0.001 | 0.003 | 0.003 | 0.001 | 0.001 |
| *mannaseense* | MRN182 | OCH S- 14 | PER | 10 | 0.005 | 0.003 | 0.003 | 0.958 | 0.004 | 0.001 | 0.004 | 0.004 | 0.004 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.004 | 0.002 |
| *mannaseense* | MRN277 | HAW 2474 | PER | 10 | 0.002 | 0.002 | 0.006 | 0.973 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0 | 0.002 | 0.003 | 0.002 | 0.001 | 0.002 |
| *mannaseense* | MRN690 | ROR 779 | PER | 10 | 0.002 | 0.003 | 0.001 | 0.971 | 0.003 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.007 | 0.002 | 0.002 | 0.001 | 0.001 |
| *mannaseense* | MRN77 | SS 7209 | PER | 10 | 0.002 | 0.001 | 0.001 | 0.958 | 0.003 | 0.004 | 0.002 | 0.003 | 0.001 | 0.004 | 0.001 | 0.002 | 0.008 | 0.002 | 0.002 | 0.006 |
| *multidissectum* | MLT363 | PEH 1366 | PER | 10 | 0.019 | 0.023 | 0.005 | 0.712 | 0.01 | 0.019 | 0.008 | 0.004 | 0.007 | 0.031 | 0.005 | 0.012 | 0.045 | 0.056 | 0.021 | 0.022 |
| *multidissectum* | MLT722 | EBS 2420 | PER | 10 | 0.004 | 0.004 | 0.004 | 0.842 | 0.009 | 0.004 | 0.003 | 0.003 | 0.004 | 0.013 | 0.001 | 0.003 | 0.092 | 0.008 | 0.002 | 0.003 |
| *multidissectum* | MLT723 | HAW 96 | PER | 10 | 0.01 | 0.007 | 0.002 | 0.777 | 0.003 | 0.002 | 0.007 | 0.02 | 0.029 | 0.021 | 0.002 | 0.011 | 0.002 | 0.003 | 0.072 | 0.032 |
| *multidissectum* | MLT724 | HAW 2469 | PER | 10 | 0.007 | 0.044 | 0.008 | 0.492 | 0.013 | 0.018 | 0.006 | 0.024 | 0.003 | 0.212 | 0.002 | 0.028 | 0.003 | 0.011 | 0.005 | 0.064 |
| *multidissectum* | MLT725 | HHCH 5068 | PER | 10 | 0.008 | 0.017 | 0.017 | 0.579 | 0.02 | 0.01 | 0.013 | 0.007 | 0.005 | 0.12 | 0.003 | 0.014 | 0.016 | 0.027 | 0.006 | 0.138 |
| *multidissectum* | MLT727 | PEH 1340 | PER | 10 | 0.039 | 0.032 | 0.002 | 0.746 | 0.004 | 0.011 | 0.003 | 0.002 | 0.003 | 0.008 | 0.004 | 0.002 | 0.003 | 0.128 | 0.002 | 0.011 |
| *multidissectum* | MLT728 | PEH 1407 | PER | 10 | 0.007 | 0.004 | 0.003 | 0.525 | 0.002 | 0.002 | 0.003 | 0.005 | 0.003 | 0.047 | 0.003 | 0.005 | 0.003 | 0.382 | 0.001 | 0.005 |
| *multidissectum* | MLT729 | PEH 1583 | PER | 10 | 0.008 | 0.093 | 0.009 | 0.522 | 0.013 | 0.011 | 0.008 | 0.044 | 0.008 | 0.088 | 0.003 | 0.005 | 0.008 | 0.03 | 0.02 | 0.135 |
| *multidissectum* | MLT730 | ROR 591 | UNKNOWN | 10 | 0.337 | 0.002 | 0.002 | 0.476 | 0.007 | 0.002 | 0.003 | 0.005 | 0.014 | 0.008 | 0.002 | 0.004 | 0.001 | 0.132 | 0.004 | 0.002 |
| *multidissectum* | MLT731 | ROR 989 | PER | 10 | 0.01 | 0.004 | 0.009 | 0.951 | 0.004 | 0.005 | 0.002 | 0.002 | 0.003 | 0.01 | 0.001 | 0.003 | 0.005 | 0.009 | 0.001 | 0.001 |
| *multidissectum* | MLT732 | ROR 1020 | PER | 10 | 0.009 | 0.003 | 0.004 | 0.907 | 0.003 | 0.002 | 0.015 | 0.003 | 0.004 | 0.005 | 0.001 | 0.009 | 0.002 | 0.024 | 0.004 | 0.004 |
| *multiinterruptum* | MTP190 | OCH S- 27 | PER | 10 | 0.123 | 0.004 | 0.002 | 0.446 | 0.009 | 0.01 | 0.006 | 0.074 | 0.007 | 0.004 | 0.018 | 0.005 | 0.004 | 0.002 | 0.286 | 0.002 |
| *orophilum* | ORP196 | OCH 12082 | PER | 10 | 0.002 | 0.002 | 0.004 | 0.961 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.005 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.008 |
| *orophilum* | ORP29 | OCH 13020 | PER | 10 | 0.001 | 0.001 | 0.002 | 0.985 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *orophilum* | ORP756 | OCHS 11869 | PER | 10 | 0.007 | 0.011 | 0.002 | 0.694 | 0.002 | 0.009 | 0.009 | 0.025 | 0.01 | 0.18 | 0.004 | 0.008 | 0.009 | 0.02 | 0.002 | 0.008 |
| *orophilum* | ORP83 | OCH 12080 | PER | 10 | 0.001 | 0.001 | 0.002 | 0.977 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.005 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.004 |
| *orophilum* | ORP84 | OCH 13015 | PER | 10 | 0.001 | 0.001 | 0.002 | 0.974 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.005 |
| *pampaeense* | PAM268 | PEH 1420 | PER | 10 | 0.003 | 0.007 | 0.017 | 0.706 | 0.004 | 0.002 | 0.004 | 0.048 | 0.002 | 0.029 | 0.001 | 0.044 | 0.048 | 0.073 | 0.007 | 0.006 |
| *pampaeense* | PAM762 | GLK 108 9 | PER | 10 | 0.009 | 0.003 | 0.013 | 0.887 | 0.003 | 0.003 | 0.005 | 0.003 | 0.002 | 0.002 | 0.001 | 0.025 | 0.005 | 0.012 | 0.023 | 0.002 |
| *pampaeense* | PAM763 | ROR 986 | PER | 10 | 0.003 | 0.003 | 0.005 | 0.57 | 0.002 | 0.002 | 0.004 | 0.103 | 0.002 | 0.017 | 0.002 | 0.007 | 0.004 | 0.264 | 0.007 | 0.004 |
| *pampaeense* | PAM764 | ROR 989 | PER | 10 | 0.009 | 0.006 | 0.008 | 0.792 | 0.005 | 0.004 | 0.007 | 0.007 | 0.008 | 0.007 | 0.001 | 0.015 | 0.028 | 0.095 | 0.004 | 0.004 |
| *soukupii* | SOU815 | OCH 2032 x BLS 166 | PER | 10 | 0.272 | 0.001 | 0.002 | 0.694 | 0.003 | 0.002 | 0.003 | 0.005 | 0.004 | 0.001 | 0.001 | 0.003 | 0.002 | 0.003 | 0.001 | 0.003 |
| *subpositigena* | SLB222 | BLS 6146 | BOL | 10 | 0.002 | 0.001 | 0.002 | 0.981 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 |
| *sucrense* | SCR847 | HAM 131 | BOL | 10 | 0.139 | 0.002 | 0.038 | 0.788 | 0.003 | 0.003 | 0.007 | 0.006 | 0.017 | 0.002 | 0.001 | 0.002 | 0.004 | 0.002 | 0.004 | 0.002 |
| *unknown species* | SPEC164 | OCH 11987 | BOL | 10 | 0.017 | 0.015 | 0.005 | 0.782 | 0.004 | 0.002 | 0.022 | 0.116 | 0.013 | 0.001 | 0.001 | 0.022 | 0.002 | 0.004 | 0.01 | 0.004 |
| *unknown species* | SPEC210 | HHCH 5211 | PER | 10 | 0.029 | 0.009 | 0.003 | 0.12 | 0.004 | 0.334 | 0.408 | 0.005 | 0.014 | 0.006 | 0.002 | 0.007 | 0.003 | 0.025 | 0.029 | 0.002 |
| *unknown species* | SPEC211 | OCH 52 | PER | 10 | 0.005 | 0.331 | 0.008 | 0.001 | 0.003 | 0.03 | 0.003 | 0.006 | 0.002 | 0.026 | 0.001 | 0.009 | 0.003 | 0.566 | 0.002 | 0.003 |
| *unknown species* | SPEC533 | VSOA 86 | BOL | 10 | 0.006 | 0.002 | 0.01 | 0.001 | 0.015 | 0.002 | 0.003 | 0.004 | 0.002 | 0.003 | 0.001 | 0.029 | 0.005 | 0.027 | 0.006 | 0.886 |
| *unknown species* | SPEC726 | HHCH 5226 | PER | 10 | 0.263 | 0.005 | 0.006 | 0.596 | 0.003 | 0.004 | 0.003 | 0.002 | 0.005 | 0.001 | 0.001 | 0.037 | 0.007 | 0.004 | 0.002 | 0.002 |
| *velardei* | VLR893 | OCH 13807 | PER | 10 | 0.002 | 0.001 | 0.001 | 0.852 | 0.001 | 0.001 | 0.002 | 0.004 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.125 | 0.001 | 0.002 |
| *velardei* | VLR67 | CON 725 | CHL | 10 | 0.001 | 0.001 | 0.001 | 0.983 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 |
| *violaceimarmoratum* | SPEC394 | VSOA 7 | BOL | 10 | 0.936 | 0.019 | 0.003 | 0.001 | 0.004 | 0.002 | 0.003 | 0.004 | 0.005 | 0.003 | 0.002 | 0.006 | 0.004 | 0.002 | 0.005 | 0.002 |
| *violaceimarmoratum* | SPEC998 | SFVU 6731B | BOL | 10 | 0.002 | 0.002 | 0.002 | 0.968 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 | 0.004 | 0.001 | 0.003 | 0.004 | 0.002 | 0.001 | 0.003 |
| *virgultorum* | VRG930 | VSA 175 | BOL | 10 | 0.152 | 0.003 | 0.004 | 0.779 | 0.003 | 0.002 | 0.006 | 0.003 | 0.017 | 0.017 | 0.001 | 0.002 | 0.002 | 0.003 | 0.005 | 0.001 |
| *weberbauerii* | SPEC933 | K 3971.14 | PER | 10 | 0.272 | 0.006 | 0.001 | 0.632 | 0.026 | 0.002 | 0.028 | 0.008 | 0.01 | 0.001 | 0.001 | 0.003 | 0.003 | 0.001 | 0.004 | 0.003 |
| *lignicaule* | LGL179 | EBS 1884 | PER | 11 | 0.024 | 0.008 | 0.008 | 0.017 | 0.019 | 0.013 | 0.078 | 0.009 | 0.003 | 0.034 | 0.002 | 0.722 | 0.015 | 0.017 | 0.021 | 0.008 |
| *lignicaule* | LGL685 | HHCH 5113 | PER | 11 | 0.005 | 0.007 | 0.013 | 0.017 | 0.004 | 0.004 | 0.019 | 0.017 | 0.003 | 0.057 | 0.001 | 0.744 | 0.007 | 0.02 | 0.004 | 0.018 |
| *megistacrolobum* | MGA696 | HAM 201 | BOL | 11 | 0.003 | 0.001 | 0.011 | 0.006 | 0.001 | 0.001 | 0.003 | 0.002 | 0.003 | 0.005 | 0.001 | 0.95 | 0.002 | 0.004 | 0.002 | 0.005 |
| *megistacrolobum* | MGA697 | HJR 256 | ARG | 11 | 0.003 | 0.019 | 0.029 | 0.004 | 0.002 | 0.008 | 0.003 | 0.003 | 0.002 | 0.002 | 0.002 | 0.916 | 0.004 | 0.004 | 0.002 | 0.001 |
| *megistacrolobum* | MGA699 | VSAL 133 | BOL | 11 | 0.004 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.003 | 0.003 | 0.004 | 0.001 | 0.961 | 0.003 | 0.003 | 0.002 | 0.006 |
| *megistacrolobum* | MGA700 | VSOA 46 | BOL | 11 | 0.004 | 0.007 | 0.011 | 0.003 | 0.003 | 0.003 | 0.008 | 0.043 | 0.004 | 0.003 | 0.001 | 0.901 | 0.003 | 0.004 | 0.002 | 0.002 |
| *megistacrolobum toralapanum* | TOR726 | CAR CPC 1773 | BOL | 11 | 0.003 | 0.001 | 0.003 | 0.004 | 0.001 | 0.001 | 0.002 | 0.012 | 0.002 | 0.003 | 0.005 | 0.948 | 0.004 | 0.005 | 0.001 | 0.004 |
| *megistacrolobum toralapanum* | TOR701 | HHCH 4556 | BOL | 11 | 0.005 | 0.005 | 0.008 | 0.006 | 0.008 | 0.002 | 0.003 | 0.004 | 0.002 | 0.031 | 0.003 | 0.811 | 0.003 | 0.003 | 0.005 | 0.002 |
| *megistacrolobum toralapanum* | TOR702 | HOF 1851 | ARG | 11 | 0.013 | 0.002 | 0.006 | 0.003 | 0.004 | 0.001 | 0.005 | 0.004 | 0.004 | 0.013 | 0.022 | 0.8 | 0.001 | 0.005 | 0.009 | 0.006 |
| *megistacrolobum toralapanum* | TOR703 | VSAL 134 | BOL | 11 | 0.009 | 0.005 | 0.005 | 0.001 | 0.009 | 0.002 | 0.002 | 0.006 | 0.003 | 0.009 | 0.001 | 0.941 | 0.003 | 0.002 | 0.008 | 0.002 |
| *megistacrolobum toralapanum* | TOR704 | VSAL 136 | BOL | 11 | 0.002 | 0.002 | 0.002 | 0.004 | 0.001 | 0.002 | 0.002 | 0.003 | 0.001 | 0.002 | 0.001 | 0.969 | 0.004 | 0.002 | 0.001 | 0.002 |
| *megistacrolobum toralapanum* | TOR705 | VSLC 158 | BOL | 11 | 0.003 | 0.002 | 0.016 | 0.004 | 0.003 | 0.003 | 0.004 | 0.004 | 0.002 | 0.005 | 0.001 | 0.939 | 0.002 | 0.007 | 0.002 | 0.002 |
| *megistacrolobum toralapanum* | TOR706 | VSOA 19 | BOL | 11 | 0.006 | 0.004 | 0.006 | 0.005 | 0.021 | 0.01 | 0.006 | 0.01 | 0.006 | 0.003 | 0.001 | 0.906 | 0.004 | 0.005 | 0.003 | 0.001 |
| *sanctae-rosae* | SCT1061 | UNKNOWN | UNKNOWN | 11 | 0.002 | 0.016 | 0.004 | 0.007 | 0.025 | 0.008 | 0.007 | 0.003 | 0.002 | 0.004 | 0.001 | 0.899 | 0.002 | 0.002 | 0.003 | 0.015 |
| *sanctae-rosae* | SCT803 | HHR 3939 | ARG | 11 | 0.007 | 0.034 | 0.003 | 0.001 | 0.178 | 0.051 | 0.008 | 0.003 | 0.005 | 0.008 | 0.003 | 0.69 | 0.003 | 0.028 | 0.004 | |
| *sanctae-rosae* | SCT804 | HOHH 6066 | ARG | 11 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.003 | 0.001 | 0.974 | 0.002 | 0.001 | 0.001 | 0.002 |
| *sanctae-rosae* | SCT805 | OKA 8152 x 6153 | ARG | 11 | 0.005 | 0.005 | 0.003 | 0.002 | 0.009 | 0.002 | 0.003 | 0.005 | 0.002 | 0.002 | 0.001 | 0.895 | 0.041 | 0.007 | 0.013 | 0.005 |
| *sanctae-rosae* | SCT806 | PEH 328 | ARG | 11 | 0.002 | 0.001 | 0.001 | 0.001 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.976 | 0.002 | 0.002 | 0.001 | 0.001 |
| *sanctae-rosae* | SCT807 | ROR 41 | ARG | 11 | 0.002 | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.975 | 0.001 | 0.002 | 0.002 | 0.001 |
| *amazii* | ARZ111 | HAO 154 | BOL | 12 | 0.009 | 0.002 | 0.008 | 0.001 | 0.004 | 0.011 | 0.897 | 0.003 | 0.006 | 0.008 | 0.001 | 0.018 | 0.007 | 0.002 | 0.02 | 0.004 |
| *amazii* | ARZ112 | HAO 159 | BOL | 12 | 0.028 | 0.012 | 0.004 | 0.01 | 0.029 | 0.004 | 0.519 | 0.007 | 0.038 | 0.005 | 0.301 | 0.009 | 0.011 | 0.016 | 0.009 | 0.004 |
| *amazii* | ARZ113 | HOHL 297 | BOL | 12 | 0.002 | 0.001 | 0.006 | 0.004 | 0.003 | 0.003 | 0.941 | 0.002 | 0.001 | 0.005 | 0.001 | 0.003 | 0.003 | 0.009 | 0.002 | 0.013 |
| *amazii* | ARZ4 | HAO 157 | BOL | 12 | 0.062 | 0.002 | 0.004 | 0.002 | 0.01 | 0.003 | 0.868 | 0.007 | 0.004 | 0.007 | 0.01 | 0.003 | 0.003 | 0.004 | 0.007 | 0.005 |
| *amazii* | ARZ471 | HAM 126 | BOL | 12 | 0.004 | 0.004 | 0.004 | 0.017 | 0.003 | 0.003 | 0.921 | 0.004 | 0.002 | 0.003 | 0.001 | 0.008 | 0.005 | 0.005 | 0.004 | 0.01 |
| *chacoense* | CHC125 | BRU 6b | ARG | 12 | 0.002 | 0.003 | 0.001 | 0.001 | 0.005 | 0.001 | 0.972 | 0.002 | 0.002 | 0.003 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 |
| *chacoense* | CHC126 | BRU 444 | ARG | 12 | 0.004 | 0.004 | 0.002 | 0.001 | 0.001 | 0.002 | 0.97 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.004 | 0.001 |
| *chacoense* | CHC127 | GLK 8 54 | ARG | 12 | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.97 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.008 | 0.001 |
| *chacoense* | CHC246 | OCHS 11008 | ECU | 12 | 0.004 | 0.007 | 0.001 | 0.001 | 0.014 | 0.002 | 0.933 | 0.005 | 0.004 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 | 0.015 | 0.004 |
| *chacoense* | CHC263 | PEH 1900a x 1669 | ARG | 12 | 0.009 | 0.003 | 0.003 | 0.002 | 0.182 | 0.003 | 0.655 | 0.005 | 0.107 | 0.003 | 0.002 | 0.004 | 0.004 | 0.008 | 0.003 | 0.007 |
| *chacoense* | CHC545 | BRU 57 | ARG | 12 | 0.004 | 0.003 | 0.003 | 0.001 | 0.005 | 0.002 | 0.934 | 0.003 | 0.017 | 0.002 | 0.001 | 0.003 | 0.001 | 0.004 | 0.009 | 0.008 |
| *chacoense* | CHC546 | FCE 104 | ARG | 12 | 0.003 | 0.005 | 0.001 | 0.001 | 0.002 | 0.002 | 0.96 | 0.002 | 0.005 | 0.001 | 0.001 | 0.001 | 0.01 | 0.001 | 0.003 | 0.001 |
| *chacoense* | CHC547 | HAM 181 | BOL | 12 | 0.002 | 0.007 | 0.001 | 0.001 | 0.006 | 0.002 | 0.933 | 0.002 | 0.002 | 0.002 | 0.003 | 0.002 | 0.005 | 0.001 | 0.029 | 0.003 |
| *chacoense* | CHC548 | OCH 15261 | PRY | 12 | 0.009 | 0.003 | 0.002 | 0.001 | 0.008 | 0.004 | 0.951 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.011 | 0.001 |
| *chacoense* | CHC549 | OCH VDO 7-9 | PRY | 12 | 0.01 | 0.007 | 0.004 | 0.002 | 0.003 | 0.002 | 0.929 | 0.002 | 0.012 | 0.002 | 0.002 | 0.006 | 0.004 | 0.003 | 0.008 | 0.002 |
| *chacoense* | CHC550 | OKA 8888 | ARG | 12 | 0.003 | 0.007 | 0.001 | 0.001 | 0.002 | 0.002 | 0.962 | 0.003 | 0.002 | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 | 0.005 | 0.002 |
| *chacoense* | CHC551 | PEH 349 | ARG | 12 | 0.005 | 0.001 | 0.001 | 0.002 | 0.001 | 0.003 | 0.972 | 0.001 | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 |
| *gourlayi* | GRL1034 | OKA 4336 | ARG | 12 | 0.303 | 0.004 | 0.001 | 0.002 | 0.004 | 0.002 | 0.335 | 0.002 | 0.316 | 0.007 | 0.003 | 0.002 | 0.001 | 0.015 | 0.002 | 0.002 |
| *neocardenasii* | TAR280 | CPC BPC 1747 | BOL | 12 | 0.011 | 0.002 | 0.002 | 0.001 | 0.005 | 0.007 | 0.764 | 0.003 | 0.087 | 0.003 | 0.001 | 0.01 | 0.002 | 0.002 | 0.096 | 0.002 |
| *ruiz-lealii* | RZL802 | HJT BGRC 24960 | PER | 12 | 0.011 | 0.003 | 0.001 | 0.002 | 0.014 | 0.005 | 0.916 | 0.004 | 0.003 | 0.002 | 0.002 | 0.005 | 0.004 | 0.002 | 0.024 | 0.003 |
| *setulosiphylum* | STL811 | K 2906 x 5248 | ARG | 12 | 0.004 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.968 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.004 | 0.002 | 0.004 | 0.002 |
| *tarijense* | TAR392 | OKA 5632 | ARG | 12 | 0.004 | 0.002 | 0.002 | 0.001 | 0.005 | 0.006 | 0.789 | 0.009 | 0.006 | 0.002 | 0.003 | 0.005 | 0.005 | 0.002 | 0.157 | 0.003 |
| *tarijense* | TAR868 | HOHH 6022 | ARG | 12 | 0.004 | 0.004 | 0.003 | 0.001 | 0.009 | 0.003 | 0.913 | 0.003 | 0.007 | 0.003 | 0.001 | 0.007 | 0.004 | 0.003 | 0.032 | 0.002 |
| *tarijense* | TAR869 | OKA 4818 | ARG | 12 | 0.014 | 0.004 | 0.003 | 0.002 | 0.009 | 0.005 | 0.874 | 0.005 | 0.009 | 0.004 | 0.002 | 0.021 | 0.009 | 0.002 | 0.036 | 0.002 |
| *unknown species* | SPEC210 | HAW 867 | ARG | 12 | 0.03 | 0.003 | 0.001 | 0.001 | 0.102 | 0.009 | 0.617 | 0.002 | 0.168 | 0.002 | 0.017 | 0.007 | 0.002 | 0.003 | 0.014 | 0.002 |

*Additional file 1. (Continued) Information on the 566 South American accessions used in the analyses: their origin and partial membership to each of 16 clusters in STRUCTURE run86.*

| species | code | collection number | country | cluster | Partial membership of each cluster | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *unknown species* | SPEC211 | HAW 866 | ARG | 12 | 0.004 | 0.028 | 0.001 | 0.001 | 0.004 | 0.001 | 0.037 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.01 | 0.001 |
| *unknown species* | SPEC329 | PEH 1370 | PER | 12 | 0.926 | 0.002 | 0.01 | 0.022 | 0.003 | 0.001 | 0.002 | 0.003 | 0.005 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.012 | 0.005 |
| *vernei* | VRN903 | OKA 5927 | ARG | 12 | 0.01 | 0.001 | 0.001 | 0.001 | 0.396 | 0.005 | 0.561 | 0.003 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.01 | 0.004 |
| *yungasense* | YUN935 | SFVU 6738 x 6732 | BOL | 12 | 0.003 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.838 | 0.011 | 0.002 | 0.006 | 0.002 | 0.003 | 0.005 | 0.009 | 0.01 | 0.002 |
| *yungasense* | YUN938 | SFVU 6738 x 6739 | BOL | 12 | 0.004 | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 | 0.963 | 0.002 | 0.002 | 0.002 | 0.001 | 0.003 | 0.003 | 0.001 | 0.004 | 0.004 |
| *yungasense* | YUN96 | SFVU 6739 | BOL | 12 | 0.003 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.97 | 0.003 | 0.001 | 0.003 | 0.001 | 0.003 | 0.003 | 0.001 | 0.002 | 0.003 |
| *doddsii* | DDS588 | BESP 831 | BOL | 13 | 0.033 | 0.006 | 0.003 | 0.004 | 0.006 | 0.004 | 0.003 | 0.004 | 0.765 | 0.02 | 0.003 | 0.079 | 0.006 | 0.035 | 0.011 | 0.018 |
| *gourlayi* | GRL1005 | OKA 4832 | ARG | 13 | 0.057 | 0.027 | 0.001 | 0.001 | 0.017 | 0.003 | 0.004 | 0.003 | 0.889 | 0.001 | 0.002 | 0.003 | 0.006 | 0.001 | 0.002 | 0.001 |
| *gourlayi* | GRL1006 | OKA 4841 | ARG | 13 | 0.127 | 0.008 | 0.003 | 0.001 | 0.011 | 0.021 | 0.005 | 0.003 | 0.801 | 0.001 | 0.004 | 0.002 | 0.002 | 0.002 | 0.006 | 0.001 |
| *gourlayi* | GRL1013 | OKA 4866 | ARG | 13 | 0.101 | 0.003 | 0.002 | 0.001 | 0.004 | 0.003 | 0.003 | 0.003 | 0.86 | 0.001 | 0.004 | 0.002 | 0.005 | 0.001 | 0.006 | 0.001 |
| *gourlayi* | GRL1015 | OKA 5570 | ARG | 13 | 0.169 | 0.003 | 0.003 | 0.003 | 0.019 | 0.002 | 0.074 | 0.006 | 0.736 | 0.005 | 0.002 | 0.007 | 0.003 | 0.001 | 0.014 | 0.004 |
| *gourlayi* | GRL1030 | HOHH 5991 | ARG | 13 | 0.006 | 0.006 | 0.001 | 0.001 | 0.003 | 0.003 | 0.004 | 0.003 | 0.957 | 0.001 | 0.001 | 0.002 | 0.003 | 0.001 | 0.01 | 0.001 |
| *gourlayi* | GRL1037 | SLU 1 | ARG | 13 | 0.005 | 0.136 | 0.001 | 0.003 | 0.145 | 0.054 | 0.015 | 0.008 | 0.497 | 0.002 | 0.001 | 0.006 | 0.113 | 0.002 | 0.005 | 0.006 |
| *gourlayi* | GRL1044 | HOHH 5980 | ARG | 13 | 0.035 | 0.011 | 0.001 | 0.001 | 0.004 | 0.003 | 0.003 | 0.004 | 0.918 | 0.001 | 0.003 | 0.007 | 0.003 | 0.002 | 0.003 | 0.001 |
| *gourlayi* | GRL1049 | OKA 4925 | ARG | 13 | 0.013 | 0.014 | 0.002 | 0.007 | 0.011 | 0.006 | 0.004 | 0.037 | 0.882 | 0.002 | 0.005 | 0.004 | 0.004 | 0.001 | 0.007 | 0.003 |
| *gourlayi* | GRL347 | HOF 1727 | ARG | 13 | 0.221 | 0.003 | 0.007 | 0.002 | 0.031 | 0.004 | 0.008 | 0.056 | 0.563 | 0.002 | 0.005 | 0.003 | 0.004 | 0.079 | 0.003 | 0.01 |
| *gourlayi* | GRL805 | HOF 1800 | ARG | 13 | 0.009 | 0.002 | 0.001 | 0.001 | 0.002 | 0.003 | 0.01 | 0.007 | 0.94 | 0.002 | 0.001 | 0.002 | 0.01 | 0.005 | 0.005 | 0.002 |
| *gourlayi* | GRL806 | OKA 3801 | ARG | 13 | 0.017 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 | 0.004 | 0.002 | 0.951 | 0.003 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.003 |
| *gourlayi* | GRL808 | OKA 4445 | ARG | 13 | 0.016 | 0.003 | 0.003 | 0.007 | 0.003 | 0.005 | 0.005 | 0.001 | 0.909 | 0.013 | 0.003 | 0.007 | 0.009 | 0.005 | 0.008 | 0.004 |
| *gourlayi* | GRL809 | OKA 4829 | ARG | 13 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.983 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *gourlayi* | GRL810 | OKA 4858 | ARG | 13 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.986 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *gourlayi* | GRL811 | OKA 4873 | ARG | 13 | 0.004 | 0.009 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.973 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.003 | 0.002 |
| *gourlayi pachytrichum* | PTR613 | HAM 26 | BOL | 13 | 0.255 | 0.003 | 0.002 | 0.002 | 0.001 | 0.007 | 0.004 | 0.014 | 0.65 | 0.004 | 0.011 | 0.001 | 0.002 | 0.007 | 0.003 | 0.006 |
| *gourlayi pachytrichum* | PTR616 | HAM 112 | BOL | 13 | 0.366 | 0.003 | 0.003 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.6 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.005 | 0.007 |
| *gourlayi vidaurrei* | VID619 | HOF 1704 | ARG | 13 | 0.005 | 0.001 | 0.002 | 0.001 | 0.003 | 0.002 | 0.002 | 0.017 | 0.952 | 0.002 | 0.003 | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 |
| *gourlayi vidaurrei* | VID620 | HOF 1706 | ARG | 13 | 0.01 | 0.006 | 0.002 | 0.002 | 0.008 | 0.006 | 0.003 | 0.056 | 0.875 | 0.004 | 0.001 | 0.002 | 0.003 | 0.002 | 0.011 | 0.009 |
| *gourlayi vidaurrei* | VID622 | HOF 1724 | ARG | 13 | 0.012 | 0.004 | 0.004 | 0.004 | 0.011 | 0.006 | 0.02 | 0.027 | 0.806 | 0.031 | 0.012 | 0.012 | 0.001 | 0.022 | 0.006 | 0.02 |
| *hawkesianum* | HAW188 | OKA 4849 | ARG | 13 | 0.008 | 0.003 | 0.001 | 0.001 | 0.004 | 0.002 | 0.002 | 0.003 | 0.963 | 0.003 | 0.003 | 0.009 | 0.003 | 0.001 | 0.004 | 0.002 |
| *hawkesianum* | HAW167 | OKA 4911 | ARG | 13 | 0.015 | 0.011 | 0.001 | 0.002 | 0.038 | 0.004 | 0.007 | 0.007 | 0.88 | 0.002 | 0.001 | 0.003 | 0.024 | 0.001 | 0.003 | 0.001 |
| *hawkesianum* | HAW934 | OKA 4913 | ARG | 13 | 0.005 | 0.002 | 0.003 | 0.002 | 0.004 | 0.001 | 0.002 | 0.003 | 0.967 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 |
| *hawkesianum* | HAW935 | OKA 4914 | ARG | 13 | 0.003 | 0.004 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.973 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 |
| *hawkesianum* | HAW838 | OKA 4915 | ARG | 13 | 0.004 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.974 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 |
| *hawkesianum* | HAW837 | OKA 4916 | ARG | 13 | 0.005 | 0.017 | 0.002 | 0.002 | 0.007 | 0.005 | 0.002 | 0.001 | 0.922 | 0.002 | 0.002 | 0.002 | 0.025 | 0.001 | 0.003 | 0.002 |
| *hawkesianum* | HAW836 | OKA 4917 | ARG | 13 | 0.004 | 0.002 | 0.001 | 0.006 | 0.003 | 0.003 | 0.004 | 0.004 | 0.909 | 0.006 | 0.032 | 0.003 | 0.009 | 0.004 | 0.006 | 0.002 |
| *hoopesi* | HPS650 | BESP 823 | BOL | 13 | 0.389 | 0.005 | 0.001 | 0.004 | 0.001 | 0.001 | 0.058 | 0.003 | 0.425 | 0.017 | 0.003 | 0.007 | 0.005 | 0.002 | 0.004 | 0.014 |
| *hoopesi* | HPS652 | SPBE 6684 | BOL | 13 | 0.012 | 0.005 | 0.002 | 0.219 | 0.003 | 0.003 | 0.011 | 0.001 | 0.703 | 0.004 | 0.002 | 0.012 | 0.008 | 0.004 | 0.008 | 0.003 |
| *hoopesi* | HPS653 | SPBE 6685 x 6683 | BOL | 13 | 0.406 | 0.006 | 0.003 | 0.042 | 0.002 | 0.004 | 0.021 | 0.002 | 0.489 | 0.001 | 0.004 | 0.005 | 0.006 | 0.001 | 0.007 | 0.002 |
| *incamayoense* | INM1016 | OKA 4893 | ARG | 13 | 0.012 | 0.003 | 0.002 | 0.003 | 0.082 | 0.01 | 0.018 | 0.007 | 0.82 | 0.006 | 0.001 | 0.003 | 0.011 | 0.002 | 0.016 | 0.004 |
| *incamayoense* | INM657 | HOHH 5995 | ARG | 13 | 0.008 | 0.001 | 0.003 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.97 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.003 | 0.001 |
| *incamayoense* | INM658 | OKA 4824 | ARG | 13 | 0.005 | 0.034 | 0.005 | 0.003 | 0.004 | 0.003 | 0.004 | 0.009 | 0.895 | 0.009 | 0.001 | 0.008 | 0.008 | 0.005 | 0.003 | 0.004 |
| *incamayoense* | INM659 | OKA 4850 | ARG | 13 | 0.003 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.976 | 0.001 | 0.001 | 0.002 | 0.001 | 0.005 | 0.001 | |
| *incamayoense* | INM660 | OKA 4852 | ARG | 13 | 0.014 | 0.002 | 0.001 | 0.002 | 0.003 | 0.013 | 0.003 | 0.003 | 0.931 | 0.002 | 0.001 | 0.006 | 0.004 | 0.003 | 0.011 | 0.002 |
| *incamayoense* | INM661 | OKA 4853 | ARG | 13 | 0.003 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.976 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.003 | 0.002 |
| *incamayoense* | INM662 | OKA 4854 | ARG | 13 | 0.003 | 0.002 | 0.006 | 0.001 | 0.015 | 0.001 | 0.003 | 0.003 | 0.948 | 0.004 | 0.001 | 0.002 | 0.003 | 0.002 | 0.004 | 0.002 |
| *incamayoense* | INM663 | OKA 4855 | ARG | 13 | 0.003 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.98 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 |
| *infundibuliforme* | IFD684 | ALN 84- 8 | BOL | 13 | 0.226 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.005 | 0.004 | 0.735 | 0.005 | 0.001 | 0.002 | 0.002 | 0.002 | 0.005 | 0.003 |
| *infundibuliforme* | IFD667 | OKA 6000 | ARG | 13 | 0.028 | 0.003 | 0.004 | 0.001 | 0.031 | 0.001 | 0.047 | 0.004 | 0.814 | 0.033 | 0.009 | 0.005 | 0.004 | 0.004 | 0.008 | 0.004 |
| *neorossii* | NRS736 | HOF 1874 | ARG | 13 | 0.15 | 0.027 | 0.005 | 0.002 | 0.021 | 0.003 | 0.004 | 0.247 | 0.399 | 0.002 | 0.003 | 0.002 | 0.006 | 0.002 | 0.124 | 0.003 |
| *neorossii* | NRS737 | HOF 1875 | ARG | 13 | 0.034 | 0.02 | 0.001 | 0.005 | 0.008 | 0.013 | 0.004 | 0.432 | 0.435 | 0.003 | 0.005 | 0.003 | 0.004 | 0.002 | 0.026 | 0.004 |
| *spegazzinii* | SPG217 | OKA 4058 | ARG | 13 | 0.025 | 0.004 | 0.002 | 0.002 | 0.077 | 0.265 | 0.169 | 0.014 | 0.311 | 0.01 | 0.001 | 0.008 | 0.047 | 0.002 | 0.007 | 0.035 |
| *spegazzinii* | SPG385 | HHR 3516 | ARG | 13 | 0.005 | 0.005 | 0.004 | 0.006 | 0.026 | 0.005 | 0.143 | 0.098 | 0.589 | 0.005 | 0.002 | 0.027 | 0.051 | 0.018 | 0.011 | 0.008 |
| *spegazzinii* | SPG823 | EBS 2079 | ARG | 13 | 0.046 | 0.02 | 0.001 | 0.007 | 0.273 | 0.05 | 0.01 | 0.046 | 0.356 | 0.006 | 0.001 | 0.011 | 0.008 | 0.021 | 0.008 | 0.134 |
| *spegazzinii* | SPG826 | OKA 3793 x 3794 | ARG | 13 | 0.007 | 0.019 | 0.001 | 0.001 | 0.156 | 0.249 | 0.108 | 0.035 | 0.388 | 0.003 | 0.005 | 0.007 | 0.006 | 0.002 | 0.004 | 0.006 |
| *stenolee* | ALN257 | HHCH 4735 | BOL | 14 | 0.129 | 0.004 | 0.001 | 0.001 | 0.003 | 0.002 | 0.086 | 0.003 | 0.014 | 0.007 | 0.174 | 0.004 | 0.007 | 0.004 | 0.558 | 0.002 |
| *stenolee* | ALN320 | HHA 5657 | BOL | 14 | 0.349 | 0.007 | 0.001 | 0.002 | 0.014 | 0.003 | 0.007 | 0.007 | 0.077 | 0.005 | 0.033 | 0.004 | 0.035 | 0.011 | 0.441 | 0.004 |
| *stenolee* | ALN457 | HHA 6526a | BOL | 14 | 0.124 | 0.002 | 0.004 | 0.003 | 0.004 | 0.007 | 0.09 | 0.013 | 0.012 | 0.005 | 0.003 | 0.003 | 0.018 | 0.003 | 0.163 | 0.004 |
| *stenolee* | ALN458 | HHA 6530 | BOL | 14 | 0.021 | 0.003 | 0.006 | 0.002 | 0.009 | 0.007 | 0.011 | 0.004 | 0.003 | 0.006 | 0.002 | 0.005 | 0.013 | 0.004 | 0.164 | 0.002 |
| *berthaultii* | BER322 | HHA 6542 | BOL | 14 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.003 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.977 | 0.001 |
| *berthaultii* | BER323 | ROR 325 | BOL | 14 | 0.003 | 0.003 | 0.001 | 0.001 | 0.004 | 0.003 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.971 | 0.001 |
| *berthaultii* | BER324 | ROR BGRC 18548 | BOL | 14 | 0.003 | 0.002 | 0.001 | 0.003 | 0.002 | 0.002 | 0.005 | 0.005 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.003 | 0.886 | 0.003 |
| *berthaultii* | BER480 | AST 44 | BOL | 14 | 0.005 | 0.007 | 0.001 | 0.008 | 0.002 | 0.006 | 0.007 | 0.005 | 0.002 | 0.001 | 0.001 | 0.001 | 0.007 | 0.01 | 0.907 | 0.03 |
| *berthaultii* | BER481 | HAM 115 | BOL | 14 | 0.004 | 0.003 | 0.001 | 0.001 | 0.003 | 0.003 | 0.002 | 0.003 | 0.003 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.888 | 0.001 |
| *berthaultii* | BER482 | HAM 141 | BOL | 14 | 0.004 | 0.002 | 0.008 | 0.002 | 0.002 | 0.001 | 0.011 | 0.007 | 0.004 | 0.004 | 0.01 | 0.005 | 0.002 | 0.002 | 0.884 | 0.002 |
| *berthaultii* | BER483 | HAM 146 | BOL | 14 | 0.006 | 0.004 | 0.001 | 0.001 | 0.002 | 0.003 | 0.011 | 0.003 | 0.004 | 0.001 | 0.006 | 0.002 | 0.003 | 0.001 | 0.88 | 0.001 |
| *berthaultii* | BER484 | HHA 6451 | BOL | 14 | 0.016 | 0.006 | 0.002 | 0.001 | 0.003 | 0.008 | 0.004 | 0.004 | 0.005 | 0.001 | 0.007 | 0.002 | 0.002 | 0.002 | 0.893 | 0.004 |
| *berthaultii* | BER485 | HHA 6453 | BOL | 14 | 0.004 | 0.002 | 0.001 | 0.001 | 0.005 | 0.003 | 0.004 | 0.003 | 0.005 | 0.001 | 0.004 | 0.001 | 0.002 | 0.002 | 0.889 | 0.003 |
| *berthaultii* | BER486 | HHA 6540 | BOL | 14 | 0.005 | 0.001 | 0.003 | 0.006 | 0.002 | 0.002 | 0.008 | 0.007 | 0.002 | 0.011 | 0.01 | 0.005 | 0.002 | 0.009 | 0.827 | 0.002 |
| *berthaultii* | BER487 | HHA 6546 | BOL | 14 | 0.003 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.001 | 0.003 | 0.002 | 0.001 | 0.971 | 0.002 |
| *berthaultii* | BER488 | HHA 6552 | BOL | 14 | 0.002 | 0.001 | 0.001 | 0.004 | 0.001 | 0.002 | 0.009 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.886 | 0.002 |
| *berthaultii* | BER489 | HHA 6557 | BOL | 14 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.979 | 0.001 |
| *berthaultii* | BER490 | HHA 6668 | BOL | 14 | 0.005 | 0.002 | 0.001 | 0.003 | 0.002 | 0.003 | 0.007 | 0.002 | 0.003 | 0.003 | 0.001 | 0.002 | 0.005 | 0.007 | 0.891 | 0.002 |
| *berthaultii* | BER491 | HOF 1840 | ARG | 14 | 0.007 | 0.003 | 0.002 | 0.002 | 0.035 | 0.004 | 0.005 | 0.004 | 0.018 | 0.002 | 0.001 | 0.001 | 0.003 | 0.002 | 0.91 | 0.002 |
| *berthaultii* | BER492 | ROR 355 | BOL | 14 | 0.003 | 0.002 | 0.004 | 0.001 | 0.004 | 0.001 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.004 | 0.001 | 0.001 | 0.889 | 0.002 |
| *berthaultii* | BER493 | ROR 382 | BOL | 14 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.977 | 0.001 |
| *berthaultii* | BER494 | VSOA 1 | BOL | 14 | 0.003 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.979 | 0.001 |
| *berthaultii* | BER561 | OCH SCHICK 155 | PER | 14 | 0.003 | 0.002 | 0.002 | 0.001 | 0.003 | 0.001 | 0.007 | 0.01 | 0.002 | 0.001 | 0.001 | 0.002 | 0.003 | 0.001 | 0.886 | 0.005 |
| *berthaultii* | BER939 | HAM 113 | BOL | 14 | 0.006 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.008 | 0.012 | 0.005 | 0.001 | 0.001 | 0.006 | 0.001 | 0.002 | 0.948 | 0.002 |
| *berthaultii* | BER940 | HAM 144 | BOL | 14 | 0.006 | 0.006 | 0.002 | 0.001 | 0.002 | 0.013 | 0.007 | 0.004 | 0.016 | 0.002 | 0.002 | 0.002 | 0.003 | 0.006 | 0.902 | 0.005 |
| *berthaultii* | BER941 | HHA 6446 | BOL | 14 | 0.003 | 0.001 | 0.002 | 0.001 | 0.001 | 0.004 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.975 | 0.001 |
| *berthaultii* | BER943 | HHCH 4746 | BOL | 14 | 0.048 | 0.004 | 0.015 | 0.006 | 0.004 | 0.002 | 0.009 | 0.016 | 0.008 | 0.009 | 0.016 | 0.003 | 0.002 | 0.002 | 0.833 | 0.031 |
| *berthaultii* | BER944 | ROR 351 | BOL | 14 | 0.005 | 0.004 | 0.003 | 0.003 | 0.005 | 0.005 | 0.003 | 0.005 | 0.002 | 0.002 | 0.002 | 0.002 | 0.004 | 0.001 | 0.962 | 0.001 |
| *chacoense* | CHC338 | EBS BGRC 28503 | ARG | 14 | 0.047 | 0.002 | 0.001 | 0.001 | 0.003 | 0.002 | 0.019 | 0.002 | 0.896 | 0.002 | 0.001 | 0.001 | 0.003 | 0.004 | 0.016 | 0.001 |
| *chacoense* | CHC543 | BESP 844 | BOL | 14 | 0.006 | 0.02 | 0.005 | 0.006 | 0.002 | 0.025 | 0.023 | 0.004 | 0.005 | 0.023 | 0.001 | 0.012 | 0.012 | 0.031 | 0.767 | 0.03 |
| *chacoense* | CHC544 | BESP 845 | BOL | 14 | 0.012 | 0.003 | 0.024 | 0.002 | 0.002 | 0.004 | 0.008 | 0.005 | 0.004 | 0.007 | 0.003 | 0.016 | 0.003 | 0.005 | 0.894 | 0.011 |
| *doddsii* | DDS145 | HAO 138 | BOL | 14 | 0.006 | 0.01 | 0.012 | 0.003 | 0.004 | 0.002 | 0.004 | 0.005 | 0.005 | 0.002 | 0.002 | 0.028 | 0.003 | 0.004 | 0.886 | 0.002 |
| *doddsii* | DDS146 | HHCH 4534 | BOL | 14 | 0.02 | 0.004 | 0.023 | 0.001 | 0.012 | 0.004 | 0.006 | 0.003 | 0.011 | 0.004 | 0.004 | 0.007 | 0.003 | 0.001 | 0.896 | 0.002 |
| *doddsii* | DDS588 | HHCH 4736 | BOL | 14 | 0.18 | 0.011 | 0.001 | 0.001 | 0.003 | 0.001 | 0.101 | 0.003 | 0.055 | 0.093 | 0.01 | 0.004 | 0.076 | 0.008 | 0.443 | 0.016 |
| *gandarillasii* | GND163 | HAO 18 | BOL | 14 | 0.352 | 0.006 | 0.003 | 0.001 | 0.006 | 0.006 | 0.039 | 0.007 | 0.007 | 0.003 | 0.002 | 0.011 | 0.004 | 0.001 | 0.547 | 0.004 |
| *gandarillasii* | GND270 | B 715 | BOL | 14 | 0.224 | 0.012 | 0.003 | 0.021 | 0.018 | 0.007 | 0.008 | 0.005 | 0.026 | 0.002 | 0.002 | 0.003 | 0.022 | 0.004 | 0.641 | 0.004 |
| *gandarillasii* | GND346 | ROR 288 | BOL | 14 | 0.129 | 0.006 | 0.008 | 0.002 | 0.005 | 0.003 | 0.127 | 0.004 | 0.029 | 0.038 | 0.002 | 0.009 | 0.003 | 0.004 | 0.611 | 0.023 |
| *gandarillasii* | GND603 | ROR 259 | BOL | 14 | 0.165 | 0.005 | 0.002 | 0.002 | 0.006 | 0.016 | 0.161 | 0.006 | 0.008 | 0.014 | 0.001 | 0.006 | 0.058 | 0.003 | 0.592 | 0.016 |
| *gandarillasii* | GND62 | SFVU 6656 | BOL | 14 | 0.177 | 0.002 | 0.001 | 0.002 | 0.007 | 0.002 | 0.007 | 0.005 | 0.006 | 0.002 | 0.013 | 0.017 | 0.003 | 0.002 | 0.761 | 0.002 |
| *ssikuosslyium* | STL214 | HAW 41 | ARG | 14 | 0.006 | 0.005 | 0.004 | 0.001 | 0.004 | 0.003 | 0.409 | 0.018 | 0.004 | 0.001 | 0.003 | 0.001 | 0.002 | 0.003 | 0.534 | 0.001 |
| *tarijense* | TAR224 | CPC 2667 | BOL | 14 | 0.005 | 0.006 | 0.001 | 0.003 | 0.003 | 0.004 | 0.128 | 0.013 | 0.003 | 0.001 | 0.001 | 0.003 | 0.017 | 0.007 | 0.802 | 0.002 |
| *tarijense* | TAR225 | HAW 17 | BOL | 14 | 0.015 | 0.004 | 0.003 | 0.006 | 0.004 | 0.003 | 0.007 | 0.006 | 0.006 | 0.002 | 0.003 | 0.004 | 0.002 | 0.002 | 0.923 | 0.009 |
| *tarijense* | TAR852 | HAM 117 | BOL | 14 | 0.006 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.003 | 0.002 | 0.001 | 0.002 | 0.002 | 0.005 | 0.001 | 0.001 | 0.97 | 0.001 |
| *tarijense* | TAR853 | HAM 170 | BOL | 14 | 0.003 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.008 | 0.005 | 0.002 | 0.005 | 0.001 | 0.004 | 0.003 | 0.002 | 0.948 | 0.003 |
| *tarijense* | TAR854 | HAM 172 | BOL | 14 | 0.004 | 0.002 | 0.003 | 0.004 | 0.002 | 0.004 | 0.002 | 0.01 | 0.002 | 0.005 | 0.001 | 0.003 | 0.007 | 0.002 | 0.943 | 0.005 |
| *tarijense* | TAR855 | HAM 173 | BOL | 14 | 0.003 | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 | 0.005 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.949 | 0.001 |
| *tarijense* | TAR856 | HAW 810 | ARG | 14 | 0.003 | 0.002 | 0.001 | 0.001 | 0.042 | 0.001 | 0.058 | 0.009 | 0.023 | 0.014 | 0.001 | 0.002 | 0.007 | 0.003 | 0.834 | 0.008 |
| *tarijense* | TAR857 | HHCH 4538 | BOL | 14 | 0.009 | 0.001 | 0.001 | 0.002 | 0.005 | 0.001 | 0.004 | 0.007 | 0.01 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.949 | 0.003 |

***Additional file 1. (Continued) Information on the 566 South American accessions used in the analyses: their origin and partial membership to each of 16 clusters in STRUCTURE run86.***

| species | code | collection number | country | cluster | Partial membership of each cluster | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *tarijense* | TAR858 | HHCH 4574 | BOL | 14 | 0.005 | 0.013 | 0.002 | 0.002 | 0.002 | 0.003 | 0.004 | 0.002 | 0.002 | 0.002 | 0.001 | 0.01 | 0.004 | 0.002 | 0.945 | 0.001 |
| *tarijense* | TAR859 | HOF 1717 | ARG | 14 | 0.003 | 0.036 | 0.002 | 0.001 | 0.002 | 0.003 | 0.004 | 0.006 | 0.002 | 0.002 | 0.001 | 0.011 | 0.005 | 0.002 | 0.914 | 0.009 |
| *tarijense* | TAR860 | HOF 1867 | ARG | 14 | 0.006 | 0.002 | 0.001 | 0.005 | 0.003 | 0.001 | 0.005 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.004 | 0.002 | 0.952 | 0.011 |
| *tarijense* | TAR862 | HOF 1885 | ARG | 14 | 0.002 | 0.007 | 0.001 | 0.001 | 0.007 | 0.007 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.96 | 0.001 |
| *tarijense* | TAR865 | HOF 1896 | ARG | 14 | 0.003 | 0.007 | 0.001 | 0.002 | 0.004 | 0.004 | 0.004 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.004 | 0.001 | 0.961 | 0.002 |
| *tarijense* | TAR864 | HOF 1902 | ARG | 14 | 0.003 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.977 | 0.001 |
| *tarijense* | TAR865 | HOF 2017 | ARG | 14 | 0.002 | 0.006 | 0.013 | 0.001 | 0.016 | 0.01 | 0.013 | 0.003 | 0.004 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.922 | 0.002 |
| *tarijense* | TAR866 | HOF 1713 x 1714 | ARG | 14 | 0.037 | 0.041 | 0.001 | 0.001 | 0.004 | 0.006 | 0.01 | 0.01 | 0.074 | 0.002 | 0.003 | 0.002 | 0.002 | 0.009 | 0.789 | 0.008 |
| *tarijense* | TAR867 | HOF 1886 x 1867 | ARG | 14 | 0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.013 | 0.005 | 0.002 | 0.001 | 0.002 | 0.001 | 0.007 | 0.001 | 0.957 | 0.002 |
| *tarijense* | TAR870 | OKA 5881 | ARG | 14 | 0.02 | 0.007 | 0.004 | 0.001 | 0.03 | 0.005 | 0.019 | 0.006 | 0.066 | 0.001 | 0.002 | 0.002 | 0.016 | 0.001 | 0.82 | 0.001 |
| *tarijense* | TAR871 | OKA 5883 | ARG | 14 | 0.002 | 0.004 | 0.001 | 0.002 | 0.004 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.001 | 0.002 | 0.005 | 0.002 | 0.964 | 0.001 |
| *tarijense* | TAR872 | OKA 5884 | ARG | 14 | 0.003 | 0.003 | 0.001 | 0.002 | 0.002 | 0.005 | 0.003 | 0.003 | 0.005 | 0.003 | 0.001 | 0.01 | 0.007 | 0.004 | 0.946 | 0.002 |
| *tarijense* | TAR873 | OKA 5885 | ARG | 14 | 0.004 | 0.017 | 0.012 | 0.002 | 0.007 | 0.003 | 0.007 | 0.005 | 0.004 | 0.052 | 0.001 | 0.003 | 0.007 | 0.002 | 0.868 | 0.005 |
| *tarijense* | TAR874 | OKA 5886 | ARG | 14 | 0.002 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 | 0.003 | 0.003 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 | 0.974 | 0.001 |
| *tarijense* | TAR875 | OKA 5888 | ARG | 14 | 0.028 | 0.003 | 0.001 | 0.005 | 0.004 | 0.003 | 0.002 | 0.003 | 0.003 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 | 0.939 | 0.002 |
| *tarijense* | TAR876 | OKA 6234 | ARG | 14 | 0.007 | 0.002 | 0.002 | 0.003 | 0.003 | 0.001 | 0.002 | 0.003 | 0.004 | 0.003 | 0.001 | 0.002 | 0.001 | 0.009 | 0.954 | 0.001 |
| *tarijense* | TAR877 | OKA 6239 | ARG | 14 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.001 | 0.003 | 0.005 | 0.001 | 0.002 | 0.001 | 0.001 | 0.002 | 0.002 | 0.972 | 0.001 |
| *tarijense* | TAR878 | OKA 6320 | ARG | 14 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.976 | 0.001 |
| *tarijense* | TAR879 | VSUG 619 | BOL | 14 | 0.039 | 0.004 | 0.008 | 0.001 | 0.003 | 0.002 | 0.009 | 0.003 | 0.005 | 0.009 | 0.003 | 0.032 | 0.002 | 0.007 | 0.888 | 0.002 |
| *unknown species* | SPEC601 | SCH 88 | COL | 14 | 0.021 | 0.004 | 0.007 | 0.066 | 0.005 | 0.003 | 0.001 | 0.001 | 0.009 | 0.002 | 0.023 | 0.002 | 0.031 | 0.002 | 0.003 | 0.002 |
| *yungasense* | SPEC255 | GLK 107 3 | BOL | 14 | 0.011 | 0.005 | 0.008 | 0.005 | 0.007 | 0.008 | 0.003 | 0.04 | 0.004 | 0.005 | 0.364 | 0.003 | 0.012 | 0.179 | 0.003 | 0.326 |
| *okadae* | OKA739 | HHA 6727 | ARG | 15 | 0.003 | 0.002 | 0.003 | 0.002 | 0.001 | 0.002 | 0.002 | 0.971 | 0.002 | 0.001 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.001 |
| *okadae* | OKA742 | VSA 176 | BOL | 15 | 0.009 | 0.002 | 0.002 | 0.002 | 0.003 | 0.008 | 0.002 | 0.954 | 0.006 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.004 | 0.002 |
| *okadae* | OKA743 | VSA 178 | BOL | 15 | 0.002 | 0.001 | 0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.979 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| *okadae* | OKA744 | VSA 180 | BOL | 15 | 0.002 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.002 | 0.975 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 |
| *okadae* | OKA745 | VSOA 38 | BOL | 15 | 0.002 | 0.004 | 0.013 | 0.001 | 0.002 | 0.001 | 0.002 | 0.96 | 0.002 | 0.001 | 0.001 | 0.006 | 0.002 | 0.001 | 0.002 | 0.001 |
| *okadae* | OKA746 | VSOA 42 | BOL | 15 | 0.033 | 0.002 | 0.003 | 0.001 | 0.003 | 0.002 | 0.007 | 0.914 | 0.018 | 0.001 | 0.001 | 0.004 | 0.003 | 0.002 | 0.006 | 0.001 |
| *okadae* | OKA970 | VSOA 39 | BOL | 15 | 0.007 | 0.001 | 0.003 | 0.001 | 0.017 | 0.001 | 0.029 | 0.913 | 0.005 | 0.002 | 0.001 | 0.003 | 0.005 | 0.001 | 0.007 | 0.004 |
| *stenotomum* | ALN455 | HAM 147 | BOL | 16 | 0.652 | 0.004 | 0.002 | 0.002 | 0.001 | 0.002 | 0.036 | 0.005 | 0.005 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.071 | 0.008 |
| *stenotomum* | ALN459 | HHA 6664 | BOL | 16 | 0.653 | 0.002 | 0.016 | 0.003 | 0.002 | 0.003 | 0.022 | 0.006 | 0.002 | 0.002 | 0.002 | 0.003 | 0.006 | 0.022 | 0.251 | 0.002 |
| *stenotomum* | ALN460 | HHA 6456 x 6457 | BOL | 16 | 0.663 | 0.021 | 0.001 | 0.002 | 0.018 | 0.017 | 0.105 | 0.006 | 0.011 | 0.009 | 0.004 | 0.054 | 0.004 | 0.02 | 0.033 | 0.002 |
| *sraoc-paps* | ARP109 | GLK 5 4 | PER | 16 | 0.317 | 0.123 | 0.002 | 0.038 | 0.065 | 0.03 | 0.27 | 0.012 | 0.005 | 0.002 | 0.06 | 0.003 | 0.013 | 0.005 | 0.054 | 0.001 |
| *sraoc-paps* | ARP110 | SCH 150 | PER | 16 | 0.938 | 0.007 | 0.001 | 0.002 | 0.004 | 0.008 | 0.01 | 0.003 | 0.002 | 0.002 | 0.001 | 0.003 | 0.005 | 0.002 | 0.008 | 0.005 |
| *astleyi* | AST114 | HAQ 126 | BOL | 16 | 0.657 | 0.002 | 0.004 | 0.002 | 0.003 | 0.002 | 0.046 | 0.002 | 0.02 | 0.003 | 0.004 | 0.02 | 0.003 | 0.002 | 0.021 | 0.009 |
| *astleyi* | AST472 | HAM 203 | BOL | 16 | 0.957 | 0.002 | 0.002 | 0.001 | 0.002 | 0.003 | 0.005 | 0.004 | 0.007 | 0.003 | 0.003 | 0.004 | 0.002 | 0.001 | 0.004 | 0.002 |
| *astleyi* | AST474 | HAM 207 | BOL | 16 | 0.891 | 0.005 | 0.001 | 0.006 | 0.002 | 0.004 | 0.04 | 0.007 | 0.004 | 0.009 | 0.001 | 0.006 | 0.005 | 0.005 | 0.005 | 0.006 |
| *astleyi* | AST475 | HAM 208 | BOL | 16 | 0.89 | 0.022 | 0.003 | 0.002 | 0.002 | 0.003 | 0.016 | 0.003 | 0.006 | 0.013 | 0.002 | 0.005 | 0.006 | 0.004 | 0.005 | 0.019 |
| *astleyi* | AST478 | HAM 209 | BOL | 16 | 0.93 | 0.015 | 0.005 | 0.001 | 0.002 | 0.003 | 0.003 | 0.003 | 0.007 | 0.003 | 0.007 | 0.002 | 0.007 | 0.001 | 0.01 | 0.001 |
| *avilesii* | AVL477 | HHA 6519 | BOL | 16 | 0.656 | 0.004 | 0.002 | 0.026 | 0.003 | 0.111 | 0.004 | 0.084 | 0.009 | 0.007 | 0.007 | 0.034 | 0.003 | 0.018 | 0.026 | 0.007 |
| *avilesii* | AVL478 | HHA 6521 | BOL | 16 | 0.365 | 0.004 | 0.004 | 0.052 | 0.004 | 0.038 | 0.026 | 0.116 | 0.097 | 0.13 | 0.015 | 0.034 | 0.01 | 0.02 | 0.073 | 0.012 |
| *avilesii* | AVL479 | HHA 6522 | BOL | 16 | 0.909 | 0.002 | 0.001 | 0.001 | 0.003 | 0.003 | 0.024 | 0.004 | 0.007 | 0.008 | 0.002 | 0.003 | 0.004 | 0.004 | 0.021 | 0.002 |
| *bolivianse* | BLV498 | HAM 162 | BOL | 16 | 0.934 | 0.003 | 0.003 | 0.006 | 0.004 | 0.002 | 0.007 | 0.003 | 0.01 | 0.003 | 0.001 | 0.003 | 0.008 | 0.002 | 0.01 | 0.002 |
| *bolivianse* | BLV498 | HVHL 5669 | BOL | 16 | 0.908 | 0.002 | 0.002 | 0.014 | 0.002 | 0.004 | 0.023 | 0.007 | 0.004 | 0.01 | 0.006 | 0.003 | 0.008 | 0.004 | 0.004 | 0.003 |
| *bolivianse* | BLV499 | OKA 5051b | ARG | 16 | 0.826 | 0.004 | 0.002 | 0.003 | 0.004 | 0.003 | 0.025 | 0.004 | 0.009 | 0.013 | 0.008 | 0.009 | 0.054 | 0.021 | 0.014 | 0.003 |
| *brevicaule* | BRC1020 | EBS 2363 OCH | BOL | 16 | 0.959 | 0.002 | 0.01 | 0.002 | 0.002 | 0.003 | 0.005 | 0.001 | 0.004 | 0.001 | 0.001 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 |
| *brevicaule* | BRC1025 | HHA 6620 | BOL | 16 | 0.915 | 0.005 | 0.001 | 0.001 | 0.013 | 0.005 | 0.008 | 0.002 | 0.004 | 0.002 | 0.001 | 0.005 | 0.031 | 0.002 | 0.005 | 0.002 |
| *brevicaule* | BRC1026 | HHA 6619 | BOL | 16 | 0.97 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.004 | 0.003 | 0.002 | 0.002 | 0.005 | 0.001 | 0.002 | 0.001 | 0.003 | 0.002 |
| *brevicaule* | BRC1040 | UGN 4673 | BOL | 16 | 0.92 | 0.002 | 0.002 | 0.003 | 0.014 | 0.002 | 0.003 | 0.002 | 0.009 | 0.011 | 0.001 | 0.002 | 0.002 | 0.002 | 0.003 | 0.022 |
| *brevicaule* | BRC1047 | HHA 6621 | BOL | 16 | 0.966 | 0.002 | 0.001 | 0.001 | 0.004 | 0.002 | 0.003 | 0.004 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 |
| *brevicaule* | BRC327 | HHA 6690 | BOL | 16 | 0.952 | 0.001 | 0.001 | 0.001 | 0.004 | 0.003 | 0.007 | 0.003 | 0.011 | 0.001 | 0.001 | 0.005 | 0.002 | 0.004 | 0.002 | 0.001 |
| *brevicaule* | BRC505 | EBS 2361 OCHK | BOL | 16 | 0.902 | 0.002 | 0.003 | 0.003 | 0.005 | 0.002 | 0.02 | 0.011 | 0.011 | 0.016 | 0.001 | 0.003 | 0.003 | 0.004 | 0.003 | 0.01 |
| *brevicaule* | BRC506 | HHA 6661b | BOL | 16 | 0.969 | 0.001 | 0.001 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.006 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| *brevicaule* | BRC507 | HHA 6701 | BOL | 16 | 0.948 | 0.001 | 0.002 | 0.012 | 0.003 | 0.001 | 0.01 | 0.002 | 0.002 | 0.002 | 0.001 | 0.003 | 0.002 | 0.006 | 0.002 | 0.005 |
| *brevicaule* | BRC509 | UGN 4909 | BOL | 16 | 0.94 | 0.002 | 0.001 | 0.001 | 0.004 | 0.002 | 0.01 | 0.002 | 0.005 | 0.005 | 0.002 | 0.002 | 0.001 | 0.002 | 0.007 | 0.014 |
| *doddsii* | DDS144 | HAQ 19 | BOL | 16 | 0.574 | 0.002 | 0.002 | 0.001 | 0.002 | 0.001 | 0.25 | 0.004 | 0.004 | 0.003 | 0.002 | 0.003 | 0.008 | 0.002 | 0.139 | 0.003 |
| *gandarillasii* | GND18 | SFVU 6624 | BOL | 16 | 0.471 | 0.004 | 0.003 | 0.001 | 0.008 | 0.002 | 0.034 | 0.003 | 0.008 | 0.002 | 0.004 | 0.005 | 0.017 | 0.002 | 0.437 | 0.004 |
| *gourlayi* | GRL1008 | OKA 4346 | ARG | 16 | 0.897 | 0.005 | 0.002 | 0.001 | 0.003 | 0.003 | 0.007 | 0.003 | 0.054 | 0.003 | 0.002 | 0.003 | 0.006 | 0.003 | 0.006 | 0.003 |
| *gourlayi* | GRL1009 | OKA 4349a | ARG | 16 | 0.814 | 0.018 | 0.003 | 0.004 | 0.003 | 0.002 | 0.009 | 0.009 | 0.306 | 0.005 | 0.002 | 0.007 | 0.01 | 0.007 | 0.003 | 0.003 |
| *gourlayi* | GRL1010 | OKA 4495 | ARG | 16 | 0.964 | 0.004 | 0.001 | 0.001 | 0.002 | 0.003 | 0.004 | 0.003 | 0.004 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.004 | 0.001 |
| *gourlayi* | GRL1011 | OKA 4513 | ARG | 16 | 0.954 | 0.002 | 0.001 | 0.001 | 0.003 | 0.002 | 0.002 | 0.003 | 0.019 | 0.003 | 0.002 | 0.001 | 0.002 | 0.001 | 0.002 | 0.002 |
| *gourlayi* | GRL1012 | OKA 4516 | ARG | 16 | 0.839 | 0.006 | 0.002 | 0.016 | 0.006 | 0.003 | 0.004 | 0.006 | 0.034 | 0.016 | 0.007 | 0.024 | 0.004 | 0.025 | 0.004 | 0.003 |
| *gourlayi* | GRL1014 | OKA 4885 | ARG | 16 | 0.545 | 0.018 | 0.002 | 0.001 | 0.005 | 0.018 | 0.011 | 0.003 | 0.372 | 0.001 | 0.002 | 0.005 | 0.003 | 0.001 | 0.003 | 0.006 |
| *gourlayi* | GRL1021 | OKA 6724 | ARG | 16 | 0.844 | 0.002 | 0.002 | 0.001 | 0.004 | 0.003 | 0.006 | 0.008 | 0.313 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.004 | 0.002 |
| *gourlayi* | GRL1022 | OKA 6793 | ARG | 16 | 0.917 | 0.012 | 0.001 | 0.002 | 0.01 | 0.006 | 0.002 | 0.003 | 0.015 | 0.002 | 0.003 | 0.012 | 0.005 | 0.003 | 0.003 | 0.003 |
| *gourlayi* | GRL1029 | HOF 1697 | ARG | 16 | 0.481 | 0.002 | 0.003 | 0.002 | 0.007 | 0.001 | 0.005 | 0.003 | 0.476 | 0.003 | 0.001 | 0.002 | 0.003 | 0.002 | 0.003 | 0.007 |
| *gourlayi* | GRL1032 | OKA 7558 | ARG | 16 | 0.769 | 0.006 | 0.001 | 0.001 | 0.004 | 0.003 | 0.006 | 0.002 | 0.191 | 0.001 | 0.002 | 0.001 | 0.004 | 0.002 | 0.004 | 0.001 |
| *gourlayi* | GRL1033 | OKA 4319 | ARG | 16 | 0.652 | 0.003 | 0.001 | 0.002 | 0.003 | 0.003 | 0.007 | 0.002 | 0.313 | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.003 |
| *gourlayi* | GRL1035 | OKA 4556 | ARG | 16 | 0.925 | 0.003 | 0.001 | 0.001 | 0.003 | 0.011 | 0.003 | 0.006 | 0.024 | 0.009 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 |
| *gourlayi* | GRL1042 | PEH 365 | ARG | 16 | 0.817 | 0.036 | 0.003 | 0.001 | 0.022 | 0.012 | 0.008 | 0.027 | 0.046 | 0.002 | 0.002 | 0.002 | 0.004 | 0.002 | 0.01 | 0.003 |
| *gourlayi* | GRL1048 | OCH K 9 | BOL | 16 | 0.929 | 0.028 | 0.004 | 0.001 | 0.004 | 0.017 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 | 0.001 | 0.003 | 0.001 | 0.003 | 0.001 |
| *gourlayi* | GRL1051 | OKA 5376 | ARG | 16 | 0.691 | 0.013 | 0.001 | 0.001 | 0.004 | 0.002 | 0.012 | 0.007 | 0.243 | 0.002 | 0.001 | 0.01 | 0.002 | 0.002 | 0.006 | 0.001 |
| *gourlayi* | GRL1053 | HOF 1832 | ARG | 16 | 0.576 | 0.019 | 0.002 | 0.001 | 0.005 | 0.002 | 0.002 | 0.003 | 0.348 | 0.004 | 0.001 | 0.004 | 0.01 | 0.002 | 0.02 | 0.002 |
| *gourlayi* | GRL1054 | OKA 4490a | ARG | 16 | 0.886 | 0.002 | 0.001 | 0.001 | 0.006 | 0.002 | 0.002 | 0.003 | 0.087 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 |
| *gourlayi* | GRL604 | HHR 3907 | ARG | 16 | 0.959 | 0.004 | 0.001 | 0.003 | 0.003 | 0.002 | 0.003 | 0.004 | 0.007 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.004 | 0.002 |
| *gourlayi pachylrichum* | PTR612 | AST 62 | BOL | 16 | 0.905 | 0.015 | 0.003 | 0.001 | 0.004 | 0.01 | 0.007 | 0.003 | 0.014 | 0.013 | 0.002 | 0.007 | 0.002 | 0.002 | 0.01 | 0.005 |
| *gourlayi pachylrichum* | PTR614 | HAM 106 | BOL | 16 | 0.963 | 0.003 | 0.001 | 0.001 | 0.003 | 0.004 | 0.004 | 0.002 | 0.003 | 0.001 | 0.002 | 0.001 | 0.004 | 0.001 | 0.006 | 0.002 |
| *gourlayi pachylrichum* | PTR615 | HAM 107 | BOL | 16 | 0.961 | 0.004 | 0.002 | 0.001 | 0.002 | 0.001 | 0.003 | 0.004 | 0.003 | 0.002 | 0.005 | 0.002 | 0.002 | 0.003 | 0.004 | 0.001 |
| *gourlayi pachylrichum* | PTR617 | ROR 194 | BOL | 16 | 0.73 | 0.001 | 0.002 | 0.235 | 0.005 | 0.001 | 0.002 | 0.003 | 0.008 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 |
| *gourlayi pachylrichum* | PTR618 | VSH 2426 | BOL | 16 | 0.972 | 0.002 | 0.002 | 0.001 | 0.003 | 0.001 | 0.002 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.001 | 0.002 | 0.003 | 0.002 |
| *gourlayi vidaurrei* | VID621 | HOF 1716 | ARG | 16 | 0.483 | 0.002 | 0.002 | 0.004 | 0.008 | 0.021 | 0.008 | 0.032 | 0.378 | 0.004 | 0.003 | 0.003 | 0.019 | 0.024 | 0.004 | 0.005 |
| *gourlayi vidaurrei* | VID623 | OKA 3047 | ARG | 16 | 0.812 | 0.004 | 0.004 | 0.001 | 0.03 | 0.008 | 0.013 | 0.087 | 0.018 | 0.002 | 0.008 | 0.005 | 0.003 | 0.003 | 0.005 | 0.002 |
| *gourlayi vidaurrei* | VID624 | OKA 5836 | ARG | 16 | 0.799 | 0.008 | 0.004 | 0.002 | 0.029 | 0.003 | 0.008 | 0.005 | 0.017 | 0.006 | 0.005 | 0.004 | 0.005 | 0.005 | 0.034 | 0.01 |
| *gourlayi vidaurrei* | VID625 | OKA 5869 | ARG | 16 | 0.737 | 0.016 | 0.002 | 0.001 | 0.013 | 0.05 | 0.008 | 0.009 | 0.045 | 0.011 | 0.005 | 0.008 | 0.008 | 0.006 | 0.018 | 0.008 |
| *gourlayi vidaurrei* | VID626 | OKA 6629 | ARG | 16 | 0.647 | 0.051 | 0.003 | 0.005 | 0.03 | 0.005 | 0.007 | 0.019 | 0.198 | 0.002 | 0.004 | 0.002 | 0.004 | 0.004 | 0.01 | 0.011 |
| *hondelmannii* | HDM168 | HAQ 57 | BOL | 16 | 0.732 | 0.004 | 0.134 | 0.001 | 0.005 | 0.002 | 0.006 | 0.01 | 0.014 | 0.011 | 0.052 | 0.004 | 0.003 | 0.002 | 0.011 | 0.01 |
| *hondelmannii* | HDM351 | AST 21 | BOL | 16 | 0.662 | 0.005 | 0.004 | 0.004 | 0.014 | 0.006 | 0.004 | 0.003 | 0.006 | 0.004 | 0.001 | 0.043 | 0.004 | 0.001 | 0.013 | 0.005 |
| *hondelmannii* | HDM645 | HAM 136 | BOL | 16 | 0.854 | 0.009 | 0.013 | 0.007 | 0.027 | 0.008 | 0.005 | 0.009 | 0.008 | 0.005 | 0.006 | 0.022 | 0.015 | 0.001 | 0.005 | 0.005 |
| *hondelmannii* | HDM646 | HAM 138 | BOL | 16 | 0.932 | 0.003 | 0.006 | 0.001 | 0.004 | 0.002 | 0.005 | 0.004 | 0.011 | 0.004 | 0.003 | 0.006 | 0.002 | 0.002 | 0.01 | 0.005 |
| *hondelmannii* | HDM647 | HAM 139 | BOL | 16 | 0.46 | 0.004 | 0.04 | 0.006 | 0.002 | 0.001 | 0.02 | 0.002 | 0.341 | 0.03 | 0.003 | 0.065 | 0.008 | 0.003 | 0.005 | 0.009 |
| *hondelmannii* | HDM648 | HAM 142 | BOL | 16 | 0.873 | 0.005 | 0.023 | 0.001 | 0.011 | 0.003 | 0.006 | 0.003 | 0.008 | 0.005 | 0.006 | 0.011 | 0.013 | 0.002 | 0.025 | 0.003 |
| *hoopesii* | HPS165 | HAQ 161b | BOL | 16 | 0.919 | 0.003 | 0.002 | 0.001 | 0.003 | 0.011 | 0.007 | 0.004 | 0.004 | 0.001 | 0.001 | 0.002 | 0.001 | 0.001 | 0.038 | 0.002 |
| *hoopesii* | HPS651 | SPBE 6683 | BOL | 16 | 0.897 | 0.003 | 0.003 | 0.003 | 0.006 | 0.002 | 0.003 | 0.004 | 0.029 | 0.005 | 0.001 | 0.023 | 0.026 | 0.006 | 0.004 | 0.005 |
| *tumeroforphum* | HMP171 | OCH 13247 | PER | 16 | 0.272 | 0.255 | 0.005 | 0.171 | 0.004 | 0.003 | 0.117 | 0.006 | 0.005 | 0.002 | 0.004 | 0.003 | 0.021 | 0.021 | 0.008 | 0.003 |
| *infundibuliforme* | IFD1007 | OKA 5870 | ARG | 16 | 0.653 | 0.002 | 0.002 | 0.002 | 0.034 | 0.003 | 0.004 | 0.054 | 0.015 | 0.011 | 0.002 | 0.005 | 0.002 | 0.002 | 0.007 | 0.002 |
| *infundibuliforme* | IFD1023 | HOHH 5892 | ARG | 16 | 0.946 | 0.002 | 0.003 | 0.002 | 0.004 | 0.002 | 0.004 | 0.008 | 0.017 | 0.002 | 0.001 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 |
| *infundibuliforme* | IFD665 | HAM 151 | BOL | 16 | 0.571 | 0.001 | 0.001 | 0.004 | 0.005 | 0.004 | 0.014 | 0.002 | 0.372 | 0.004 | 0.002 | 0.003 | 0.002 | 0.001 | 0.007 | 0.006 |
| *infundibuliforme* | IFD666 | OKA 4981 | ARG | 16 | 0.794 | 0.003 | 0.001 | 0.002 | 0.114 | 0.002 | 0.012 | 0.009 | 0.025 | 0.004 | 0.001 | 0.005 | 0.002 | 0.001 | 0.006 | 0.019 |
| *infundibuliforme* | IFD668 | PEH 360 | ARG | 16 | 0.879 | 0.002 | 0.003 | 0.002 | 0.167 | 0.002 | 0.012 | 0.005 | 0.002 | 0.007 | 0.002 | 0.004 | 0.004 | 0.008 | 0.004 | 0.006 |
| *leptophyes* | LPH356 | HAM 14 | BOL | 16 | 0.97 | 0.003 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.004 | 0.001 | 0.002 | 0.003 | 0.002 | 0.003 | 0.002 | 0.001 |
| *leptophyes* | LPH357 | VSOA 99 | BOL | 16 | 0.86 | 0.006 | 0.001 | 0.001 | 0.006 | 0.034 | 0.001 | 0.002 | 0.003 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.005 | 0.001 |
| *leptophyes* | LPH680 | HAM 9 | BOL | 16 | 0.958 | 0.004 | 0.001 | 0.001 | 0.002 | 0.005 | 0.003 | 0.002 | 0.002 | 0.001 | 0.003 | 0.001 | 0.005 | 0.002 | 0.007 | 0.002 |

*Additional file 1. (Continued) Information on the 566 South American accessions used in the analyses: their origin and partial membership to each of 16 clusters in STRUCTURE run86.*

| species | code | collection number | country | cluster | Partial membership of each cluster | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *leptophyes* | LPH682 | VSH 248 | BOL | 16 | 0.833 | 0.006 | 0.001 | 0.003 | 0.002 | 0.002 | 0.005 | 0.019 | 0.013 | 0.001 | 0.002 | 0.002 | 0.003 | 0.002 | 0.005 | 0.002 |
| *leptophyes* | LPH683 | VSLC 146 | BOL | 16 | 0.929 | 0.008 | 0.001 | 0.001 | 0.004 | 0.003 | 0.004 | 0.002 | 0.003 | 0.014 | 0.002 | 0.003 | 0.001 | 0.02 | 0.003 | 0.002 |
| *leptophyes* | LPH684 | VSOA 22 | BOL | 16 | 0.462 | 0.002 | 0.002 | 0.417 | 0.011 | 0.004 | 0.05 | 0.007 | 0.012 | 0.003 | 0.002 | 0.006 | 0.003 | 0.003 | 0.008 | 0.006 |
| *neorossii* | NRS281 | CPC 6047 | UNKNOWN | 16 | 0.515 | 0.024 | 0.002 | 0.003 | 0.11 | 0.011 | 0.048 | 0.018 | 0.202 | 0.019 | 0.002 | 0.006 | 0.005 | 0.006 | 0.016 | 0.012 |
| *neorossii* | NRS735 | HHR 3873 x 3875 | ARG | 16 | 0.746 | 0.003 | 0.002 | 0.006 | 0.021 | 0.004 | 0.007 | 0.009 | 0.117 | 0.006 | 0.002 | 0.004 | 0.005 | 0.004 | 0.06 | 0.004 |
| *oplocense* | OPL1002 | HOF 1612 | ARG | 16 | 0.943 | 0.002 | 0.002 | 0.004 | 0.003 | 0.003 | 0.003 | 0.009 | 0.007 | 0.002 | 0.001 | 0.012 | 0.002 | 0.003 | 0.002 | 0.001 |
| *oplocense* | OPL1003 | OKA 3829 | ARG | 16 | 0.935 | 0.004 | 0.004 | 0.001 | 0.005 | 0.002 | 0.003 | 0.008 | 0.023 | 0.003 | 0.001 | 0.001 | 0.002 | 0.001 | 0.005 | 0.001 |
| *oplocense* | OPL1004 | OKA 3946 | ARG | 16 | 0.952 | 0.002 | 0.002 | 0.001 | 0.005 | 0.003 | 0.005 | 0.006 | 0.015 | 0.002 | 0.001 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 |
| *oplocense* | OPL1024 | HOHN 5892a | ARG | 16 | 0.944 | 0.001 | 0.002 | 0.004 | 0.002 | 0.001 | 0.002 | 0.017 | 0.006 | 0.004 | 0.001 | 0.002 | 0.004 | 0.005 | 0.002 | 0.001 |
| *oplocense* | OPL1031 | HAM 166 | BOL | 16 | 0.902 | 0.003 | 0.007 | 0.001 | 0.015 | 0.003 | 0.005 | 0.006 | 0.029 | 0.002 | 0.003 | 0.004 | 0.003 | 0.003 | 0.012 | 0.002 |
| *oplocense* | OPL1041 | HAW 1921 | UNKNOWN | 16 | 0.934 | 0.004 | 0.001 | 0.001 | 0.003 | 0.013 | 0.005 | 0.003 | 0.003 | 0.001 | 0.001 | 0.002 | 0.008 | 0.004 | 0.011 | 0.003 |
| *oplocense* | OPL1046 | HAM 169 | BOL | 16 | 0.962 | 0.005 | 0.001 | 0.001 | 0.003 | 0.002 | 0.002 | 0.004 | 0.008 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| *oplocense* | OPL1056 | AST 34 | BOL | 16 | 0.856 | 0.01 | 0.003 | 0.001 | 0.008 | 0.008 | 0.046 | 0.092 | 0.004 | 0.005 | 0.007 | 0.006 | 0.005 | 0.001 | 0.034 | 0.002 |
| *oplocense* | OPL1057 | AST 72 | BOL | 16 | 0.857 | 0.012 | 0.003 | 0.001 | 0.047 | 0.008 | 0.018 | 0.003 | 0.016 | 0.002 | 0.002 | 0.013 | 0.003 | 0.002 | 0.01 | 0.004 |
| *oplocense* | OPL1059 | OKA 4502 x 4498 | ARG | 16 | 0.93 | 0.004 | 0.002 | 0.002 | 0.005 | 0.002 | 0.008 | 0.018 | 0.011 | 0.001 | 0.002 | 0.004 | 0.005 | 0.002 | 0.004 | 0.001 |
| *oplocense* | OPL747 | AST 60 | BOL | 16 | 0.911 | 0.006 | 0.004 | 0.001 | 0.039 | 0.002 | 0.008 | 0.005 | 0.003 | 0.008 | 0.001 | 0.002 | 0.001 | 0.002 | 0.004 | 0.001 |
| *oplocense* | OPL749 | HAM 161 | BOL | 16 | 0.428 | 0.003 | 0.003 | 0.002 | 0.017 | 0.001 | 0.008 | 0.415 | 0.078 | 0.002 | 0.001 | 0.03 | 0.002 | 0.002 | 0.005 | 0.003 |
| *oplocense* | OPL750 | HAM 167 | BOL | 16 | 0.86 | 0.006 | 0.027 | 0.003 | 0.002 | 0.002 | 0.002 | 0.05 | 0.004 | 0.003 | 0.003 | 0.006 | 0.003 | 0.006 | 0.003 | 0.002 |
| *oplocense* | OPL751 | HOHN 5895 | ARG | 16 | 0.493 | 0.005 | 0.002 | 0.003 | 0.005 | 0.002 | 0.002 | 0.337 | 0.133 | 0.003 | 0.001 | 0.002 | 0.004 | 0.005 | 0.003 | 0.002 |
| *oplocense* | OPL752 | HOHN 5896 | ARG | 16 | 0.389 | 0.011 | 0.012 | 0.002 | 0.02 | 0.002 | 0.004 | 0.35 | 0.144 | 0.04 | 0.01 | 0.005 | 0.002 | 0.002 | 0.007 | 0.002 |
| *oplocense* | OPL753 | OKA 3951b | ARG | 16 | 0.829 | 0.002 | 0.002 | 0.001 | 0.017 | 0.001 | 0.01 | 0.006 | 0.113 | 0.002 | 0.001 | 0.003 | 0.003 | 0.001 | 0.005 | 0.001 |
| *oplocense* | OPL754 | OKA 4498 | ARG | 16 | 0.947 | 0.002 | 0.006 | 0.004 | 0.004 | 0.001 | 0.006 | 0.003 | 0.003 | 0.001 | 0.002 | 0.003 | 0.003 | 0.006 | 0.003 | 0.009 |
| *sparsipilum* | SPL382 | HHA 6627 | BOL | 16 | 0.951 | 0.002 | 0.002 | 0.002 | 0.004 | 0.015 | 0.003 | 0.003 | 0.002 | 0.002 | 0.001 | 0.002 | 0.003 | 0.003 | 0.002 | 0.003 |
| *sparsipilum* | SPL383 | HHA 6670 | BOL | 16 | 0.968 | 0.005 | 0.001 | 0.001 | 0.002 | 0.002 | 0.001 | 0.001 | 0.003 | 0.001 | 0.002 | 0.001 | 0.004 | 0.002 | 0.003 | 0.001 |
| *sparsipilum* | SPL384 | VSLC 159 | BOL | 16 | 0.535 | 0.003 | 0.002 | 0.425 | 0.003 | 0.003 | 0.004 | 0.006 | 0.003 | 0.002 | 0.002 | 0.003 | 0.002 | 0.001 | 0.007 | 0.001 |
| *sparsipilum* | SPL816 | AST 16 | BOL | 16 | 0.881 | 0.019 | 0.008 | 0.028 | 0.003 | 0.006 | 0.004 | 0.01 | 0.165 | 0.006 | 0.002 | 0.027 | 0.036 | 0.005 | 0.003 | 0.013 |
| *sparsipilum* | SPL817 | EBS 2189 CAR | BOL | 16 | 0.854 | 0.003 | 0.002 | 0.002 | 0.004 | 0.021 | 0.006 | 0.003 | 0.036 | 0.003 | 0.02 | 0.009 | 0.016 | 0.011 | 0.006 | 0.005 |
| *sparsipilum* | SPL818 | HHA 6556 | BOL | 16 | 0.956 | 0.002 | 0.002 | 0.001 | 0.002 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.001 | 0.002 | 0.005 | 0.002 | 0.012 | 0.003 |
| *sparsipilum* | SPL819 | HHA 6669a | BOL | 16 | 0.906 | 0.015 | 0.003 | 0.003 | 0.005 | 0.009 | 0.013 | 0.005 | 0.002 | 0.002 | 0.001 | 0.004 | 0.007 | 0.002 | 0.024 | 0.006 |
| *sparsipilum* | SPL820 | HHCH 5025 | BOL | 16 | 0.966 | 0.003 | 0.001 | 0.001 | 0.002 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.002 | 0.003 | 0.003 | 0.005 | 0.002 |
| *sparsipilum* | SPL821 | VSAL 1206 | BOL | 16 | 0.886 | 0.007 | 0.002 | 0.002 | 0.006 | 0.006 | 0.003 | 0.002 | 0.002 | 0.002 | 0.001 | 0.006 | 0.003 | 0.002 | 0.004 | 0.006 |
| *sparsipilum* | SPL972 | AST 20 | BOL | 16 | 0.854 | 0.004 | 0.002 | 0.075 | 0.002 | 0.006 | 0.006 | 0.004 | 0.006 | 0.002 | 0.002 | 0.012 | 0.012 | 0.002 | 0.003 | 0.006 |
| *sparsipilum* | SPL973 | HHCH 4428 | BOL | 16 | 0.803 | 0.038 | 0.002 | 0.002 | 0.007 | 0.063 | 0.003 | 0.002 | 0.002 | 0.002 | 0.003 | 0.009 | 0.056 | 0.002 | 0.002 | 0.004 |
| *sparsipilum* | SPL975 | HVHL 5636 | BOL | 16 | 0.937 | 0.008 | 0.003 | 0.003 | 0.006 | 0.003 | 0.003 | 0.001 | 0.003 | 0.005 | 0.002 | 0.003 | 0.011 | 0.003 | 0.002 | 0.008 |
| *sparsipilum* | SPL978 | VSOA 57 | BOL | 16 | 0.848 | 0.072 | 0.004 | 0.001 | 0.011 | 0.003 | 0.002 | 0.003 | 0.003 | 0.003 | 0.001 | 0.003 | 0.007 | 0.004 | 0.014 | 0.022 |
| *spegazzinii* | SPG827 | PEH 1900 | ARG | 16 | 0.404 | 0.07 | 0.001 | 0.002 | 0.394 | 0.056 | 0.011 | 0.01 | 0.021 | 0.002 | 0.001 | 0.009 | 0.01 | 0.003 | 0.002 | 0.002 |
| *sucrense* | SCR391 | HAM 192 | BOL | 16 | 0.705 | 0.004 | 0.008 | 0.195 | 0.003 | 0.004 | 0.01 | 0.002 | 0.006 | 0.004 | 0.036 | 0.005 | 0.004 | 0.005 | 0.003 | 0.001 |
| *sucrense* | SCR843 | HAM 66 | BOL | 16 | 0.953 | 0.002 | 0.006 | 0.002 | 0.002 | 0.004 | 0.002 | 0.004 | 0.004 | 0.002 | 0.004 | 0.003 | 0.002 | 0.007 | 0.002 | 0.002 |
| *sucrense* | SCR844 | HAM 86 | BOL | 16 | 0.821 | 0.008 | 0.011 | 0.083 | 0.004 | 0.004 | 0.003 | 0.004 | 0.029 | 0.004 | 0.008 | 0.011 | 0.003 | 0.002 | 0.004 | 0.003 |
| *sucrense* | SCR845 | HAM 88 | BOL | 16 | 0.818 | 0.003 | 0.001 | 0.142 | 0.001 | 0.002 | 0.002 | 0.002 | 0.013 | 0.001 | 0.001 | 0.002 | 0.001 | 0.005 | 0.004 | 0.002 |
| *sucrense* | SCR846 | HAM 102 | BOL | 16 | 0.939 | 0.012 | 0.002 | 0.003 | 0.003 | 0.011 | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 | 0.013 | 0.002 | 0.003 | 0.001 |
| *sucrense* | SCR848 | HAM 195 | BOL | 16 | 0.969 | 0.002 | 0.005 | 0.001 | 0.002 | 0.001 | 0.003 | 0.003 | 0.002 | 0.001 | 0.003 | 0.001 | 0.001 | 0.002 | 0.002 | 0.002 |
| *sucrense* | SCR849 | HAM 197 | BOL | 16 | 0.647 | 0.011 | 0.002 | 0.003 | 0.011 | 0.009 | 0.015 | 0.002 | 0.009 | 0.005 | 0.001 | 0.006 | 0.004 | 0.001 | 0.072 | 0.002 |
| *sucrense* | SCR850 | HVHL 6127 | BOL | 16 | 0.838 | 0.003 | 0.002 | 0.124 | 0.002 | 0.001 | 0.005 | 0.002 | 0.007 | 0.002 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.003 |
| *ugentii* | SPEC881 | BESP 628 | BOL | 16 | 0.752 | 0.004 | 0.001 | 0.001 | 0.006 | 0.011 | 0.088 | 0.004 | 0.007 | 0.002 | 0.001 | 0.002 | 0.003 | 0.003 | 0.131 | 0.003 |
| *ugentii* | UGT248 | HOHL 288 | BOL | 16 | 0.929 | 0.003 | 0.001 | 0.001 | 0.004 | 0.006 | 0.019 | 0.003 | 0.003 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.024 | 0.002 |
| *ugentii* | UGT249 | HOHL 290 | BOL | 16 | 0.821 | 0.003 | 0.002 | 0.001 | 0.028 | 0.005 | 0.044 | 0.005 | 0.003 | 0.002 | 0.002 | 0.002 | 0.002 | 0.001 | 0.077 | 0.002 |
| *ugentii* | UGT44 | HOHL 284 | BOL | 16 | 0.854 | 0.001 | 0.001 | 0.001 | 0.001 | 0.007 | 0.002 | 0.002 | 0.006 | 0.002 | 0.111 | 0.005 | 0.001 | 0.003 | 0.002 | 0.001 |
| *ugentii* | UGT892 | SPBE 6887 | BOL | 16 | 0.592 | 0.007 | 0.01 | 0.004 | 0.037 | 0.006 | 0.047 | 0.003 | 0.008 | 0.01 | 0.006 | 0.006 | 0.003 | 0.003 | 0.258 | 0.003 |
| *unknown species* | SPEC123 | OCHS 11913 | BOL | 16 | 0.835 | 0.002 | 0.001 | 0.002 | 0.004 | 0.001 | 0.02 | 0.008 | 0.003 | 0.006 | 0.081 | 0.002 | 0.008 | 0.006 | 0.012 | 0.007 |
| *unknown species* | SPEC165 | GLK 65 5 | MEX | 16 | 0.762 | 0.004 | 0.002 | 0.002 | 0.004 | 0.002 | 0.003 | 0.003 | 0.002 | 0.001 | 0.003 | 0.14 | 0.007 | 0.002 | 0.036 | 0.003 |
| *unknown species* | SPEC352 | HAM 65 | BOL | 16 | 0.026 | 0.163 | 0.01 | 0.002 | 0.018 | 0.002 | 0.003 | 0.002 | 0.004 | 0.008 | 0.001 | 0.007 | 0.011 | 0.026 | 0.014 | 0.013 |
| *unknown species* | SPEC381 | ROR 791 | PER | 16 | 0.007 | 0.003 | 0.024 | 0.546 | 0.004 | 0.003 | 0.068 | 0.018 | 0.004 | 0.178 | 0.001 | 0.007 | 0.003 | 0.072 | 0.034 | 0.01 |
| *unknown species* | SPEC416 | ROR 10290 | PER | 16 | 0.913 | 0.006 | 0.018 | 0.007 | 0.002 | 0.002 | 0.004 | 0.004 | 0.002 | 0.001 | 0.01 | 0.024 | 0.501 | 0.001 | 0.004 | 0.001 |
| *unknown species* | SPEC43 | HOHL 298 | BOL | 16 | 0.404 | 0.003 | 0.015 | 0.835 | 0.003 | 0.003 | 0.002 | 0.003 | 0.017 | 0.003 | 0.002 | 0.002 | 0.003 | 0.002 | 0.002 | 0.002 |
| *unknown species* | SPEC649 | HHCH 4595 | BOL | 16 | 0.011 | 0.875 | 0.002 | 0.012 | 0.008 | 0.015 | 0.006 | 0.003 | 0.008 | 0.027 | 0.002 | 0.011 | 0.002 | 0.015 | 0.003 | 0.002 |
| *vernei* | NRS887 | SLU 3748 | ARG | 16 | 0.73 | 0.008 | 0.001 | 0.001 | 0.034 | 0.015 | 0.003 | 0.023 | 0.115 | 0.008 | 0.009 | 0.002 | 0.004 | 0.002 | 0.032 | 0.012 |
| *virgultorum* | VRG927 | HHA 6581 | BOL | 16 | 0.799 | 0.005 | 0.002 | 0.008 | 0.015 | 0.008 | 0.045 | 0.002 | 0.003 | 0.005 | 0.076 | 0.01 | 0.008 | 0.003 | 0.01 | 0.006 |
| *virgultorum* | VRG928 | HHCH 4739 | BOL | 16 | 0.695 | 0.005 | 0.001 | 0.002 | 0.002 | 0.003 | 0.006 | 0.001 | 0.002 | 0.002 | 0.001 | 0.003 | 0.06 | 0.002 | 0.01 | 0.004 |
| *virgultorum* | VRG929 | VSA 172 | BOL | 16 | 0.836 | 0.009 | 0.003 | 0.008 | 0.002 | 0.016 | 0.004 | 0.011 | 0.004 | 0.011 | 0.004 | 0.041 | 0.014 | 0.032 | 0.002 | 0.002 |
| *virgultorum* | VRG931 | VSA 177 | BOL | 16 | 0.434 | 0.015 | 0.007 | 0.406 | 0.006 | 0.062 | 0.019 | 0.012 | 0.002 | 0.003 | 0.001 | 0.003 | 0.014 | 0.002 | 0.01 | 0.003 |

# CHAPTER 6

## A novel approach to locate *Phytophthora infestans* resistance genes on the potato genetic map

Mirjam M.J. Jacobs, Ben Vosman, Vivianne Vleeshouwers, Richard R.G. Visser, Betty Henken and Ronald G. van den Berg

## Abstract

Mapping resistance genes is usually accomplished by phenotyping a segregating population for the resistance trait and genotyping it using a large number of markers. Most resistance genes are of the NBS-LRR type, of which an increasing number is sequenced. These genes and their analogs (RGAs) are often organized in clusters. Clusters tend to be rather homogenous, viz. containing genes that share high sequence homology with each other. From many of these clusters the map position is known. In this study we present and test a novel method to quickly identify to which cluster a new resistance gene belongs and to produce markers that can be used for introgression breeding. Recently, a new marker system was developed (termed NBS profiling) that produces markers in resistance genes and their analogs. We used NBS profiling to identify markers in bulked DNA samples prepared from resistant and susceptible genotypes of small segregating populations. Markers co-segregating with resistance can be tested on individual plants and directly used for breeding. To identify the resistance gene cluster a gene belongs to, the fragments were sequenced and the sequences analyzed using bioinformatics tools. Putative map positions arising from this analysis were validated using markers mapped in the segregating population. The versatility of the approach is demonstrated with a number of populations derived from wild *Solanum* species segregating for *P. infestans* resistance. Newly identified *P. infestans* resistance genes originating from *S. verrucosum*, *S. schenckii*, and *S. capsicibaccatum* could be mapped to potato chromosomes 6, 4 and 11 respectively.

## Introduction

Plants are attacked by a wide range of pathogens including viruses, bacteria, oomycetes, fungi, nematodes and insects. They have evolved passive and active ways to defend themselves against these attackers. One of the active defense systems is a type of immunity that is described by the gene for gene resistance theory, which was developed by Flor in the 1940's. It considers the gene causing resistance, the R gene in the host, to be complementary to an Avr (avirulence) gene in the pathogen (Flor, 1971). To date, more than 90 resistance (R) genes have been identified in various plants, by a wide variety of methods including map-based cloning, transposon tagging and homology based DNA library screening (Ingvardsen *et al.*, 2008). Most R genes can be assigned to one of the five major classes of R genes (Dangl & Jones, 2001). The largest of these classes contains genes that encode proteins with a Nucleotide Binding Site and a leucine-rich repeat region (the so called NBS-LRR genes). NBS-LRR resistance genes and their analogs (RGAs) are numerous in plant genomes and are often organized in clusters (AGI, 2000; Michelmore & Meyers, 1998). Many of the R genes in *Solanum* seem to be positioned in relatively few clusters (Gebhardt *et al.*, 1991; Tanksley *et al.*, 1992; Bakker *et al.*, 2003). The common approach to map resistance genes is to construct a mapping population derived from a susceptible and a resistant parent, phenotype the offspring, and then analyze the offspring with molecular markers. As many resistance traits turned out to be controlled by a single gene, more efficient methods have been developed to facilitate the search for markers linked to these genes. Bulked segregant analysis is a method for efficiently identifying markers linked to a specific trait. Two pooled DNA samples of individuals from a segregating population with contrasting phenotypes resulting from a single cross are compared. Michelmore et al. (1991) showed that this approach works well to rapidly identify RAPD and RFLP markers for any trait of interest.

The NBS region of (NBS-LRR) R genes and RGAs contain highly conserved common motifs like the P-loop, the kinase-2 motif and the GLPL motif (Meyers *et al.*, 1999; Meyers *et al.*, 2003; Monosi *et al.*, 2004) These conserved motifs within the NBS-LRR genes have been used successfully to sequence (parts of) NBS regions from various plant species (Collins *et al.*, 1998; Pflieger *et al.*, 1999; Zhang *et al.*, 2007). Van der Linden et al. (2004) developed a method for efficiently tagging NBS-LRR type of resistance genes and their analogs called NBS profiling. NBS profiling is a PCR based method that makes use of primers that target different conserved motifs in the NBS domain. It produces a DNA profile that is highly enriched for R genes and RGAs. Studies in apple (Calenge *et al.*, 2005) and potato, tomato, barley and lettuce (Van der Linden *et al.*, 2004) show that NBS profiling produces markers that are tightly linked to R genes and R gene clusters.

Late blight caused by the oomycete *Phytophthora infestans* is one of the most important and devastating diseases in potato. Currently, late blight is mainly controlled by a combination of disease management strategies, relying heavily on the use of fungicides (Fry, 2007). High disease management costs, environmental concern and the threat of promoting the evolution of resistant populations stimulated the search for R genes that can be used in breeding programs to create resistant cultivars.

In the past, 11 late blight resistance genes from the wild potato species *S. demissum* (Gebhardt & Valkonen, 2001) were introduced into cultivated potato.

As the resistances conferred by these R genes were quickly overcome by the pathogen (Wastie, 1991), the focus of breeders and scientist moved towards germplasm with partial or quantitative resistance (Fry, 2008; Van der Vossen *et al.*, 2005). More recently, the interest in finding new R genes has increased again. The presence of R genes conferring resistance against *P. infestans* in other wild potato species than *S. demissum* was investigated as well. Resistance against *P. infestans* conferred by R genes has been found in *S. pinnatisectum* (Kuhl *et al.*, 2001), *S. bulbocastanum* (Naess *et al.*, 2000) (Park *et al.*, 2005a; Song *et al.*, 2003; Van der Vossen *et al.*, 2003; Van der Vossen *et al.*, 2005), in *S. berthaultii* (Ewing *et al.*, 2000; Rauscher *et al.*, 2006), *S. microdontum* (Sandbrink *et al.*, 2000; Tan *et al.*, 2008), *S. mochiquense* (Smilde *et al.*, 2005), *S. paucissectum* (Villamon *et al.*, 2005) and *S. stoloniferum* (Wang *et al.*, 2008). There are still many other wild species that have not been tested yet for the presence of R genes against *P. infestans*. In the present study, we searched for new *P. infestans* R genes and markers in the wild potato species *Solanum verrucosum*, *Solanum schenckii* and *Solanum capsicibaccatum*. We present and test a novel approach to quickly identify at which chromosome / chromosomal region the targeted resistance gene is located and to obtain markers that can be used for introgression breeding. We illustrate this approach by describing three cases sing wild *Solanum* populations that are segregating for *P. infestans* resistance.

## Material and Methods

### Plant material

The plant material used as parents for the segregating populations were selected from a large screen of around 1000 accessions of mainly wild *Solanum* section *Petota* germplasm. The evaluated material was described by Jacobs et al. (2008). The generated segregating populations used in this study are listed in Table 1, where also details on the crosses and the number of the offspring plants are presented.

*Table 1. Segregating populations used in this study.*

| Population | Parents | Population size | Resistant bulk | Susceptible bulk |
|---|---|---|---|---|
| ver 03-392 | ver 00-3228 x AR 95-2172, Ver 03-392 and Ver 03-394 were both BC2 populations based on a resistant individual of *S. verrucosum* accession CGN 17772 (syn. PI 310966). Ver 03-392 and 03-394 are reciprokal crosses | 12 | 5 | 3 |
| ver 03-394 | ver 00-3229 x AR 95-2172, Ver 03-392 and Ver 03-394 were both BC2 populations based on a resistant individual of *S. verrucosum* accession CGN 17772 (syn. PI 310966) | 16 | 7 | 5 |
| snk 7458 | *S. schenckii* GLKS 30659 x *S. brachycarpum* CGN 18347 | 49 | 10 | 8 |
| cap 7358 | *S. capsicibaccatum* CGN 22388 x *S. circaeifolium* CGN 18133 | 32 | 6 | 4 |

***Phytophthora infestans* isolate and disease testing**

The aggressive and complex *P. infestans* isolate 90128 (race 1.3.4.7.8.11), kindly provided by Prof. Francine Govers (Laboratory of Phytopathology, Wageningen University) was cultured on Bintje leaves or on rye sucrose medium as described previously (Vleeshouwers *et al.* 1999). For disease testing, leaves from 8 to 10 weeks old plants were used. The third and fourth fully stretched leaves (counted from the top) were detached, and placed in water-saturated florists foam. The leaves were inoculated with a zoospore suspension of 50,000 spores/ml and incubated in humid trays. After 6 days, the leaves were examined for occurrence of sporulation, and the lesions sizes (LS) were measured. For each plant genotype, 10 replicates were applied on leaflets, and duplicate experiments were performed.

**DNA extractions and NBS profiling**

After 7 to 8 of weeks of growing, young plant leaves were harvested for DNA extraction. DNA was extracted according to Fulton et al. (1995). NBS profiling was performed as described by Van der Linden et al. (2004), with some minor modifications. The protocol of NBS profiling involves three steps: (1) restriction enzyme digestion of genomic DNA and the ligation of adapters (which in our experiments was done in one and the same incubation step, except when the enzyme *Taq*I was used ), (2) selective amplification of fragments containing an NBS motif using a (degenerated) primer for the conserved domains, and (3) polyacrylamide gel electrophoresis of the amplified fragments. The following 5 restriction enzymes: *Mse*I, *Taq*I, *Rsa*I, *Alu*I, *Hae*III, were used in combination with 5 NBS primers: NBS1, NBS2, NBS3, NBS5a6, and NBS9 (Van der Linden *et al.* 2004; Wang *et al.* 2008b; Mantovani *et al.* 2006; Brugmans *et al.* 2008) resulting in 25 primer-enzyme combinations. NBS profiling was carried out first on the parents and bulks of pooled resistant and susceptible plants from the population of interest. When this first round produced polymorphic bands between the parents and between the bulks, another round of NBS profiling was carried out on DNA of the parents, the bulks, and all individuals separately that constituted the resistant and susceptible bulk using only the primer-enzymes combinations that produced polymorphic bands. When possible, bulks were created using 10 resistant or 10 susceptible individuals (Table 1). In the two small *S. verrucosum* populations ver 03-392 and ver 03-394, only samples from 5 resistant and 3 susceptible and 7 resistant and 5 susceptible plants respectively, could be scored reliably for *P. infestans* resistance.

**Sequence analysis of polymorphic bands**

To determine the sequence of a polymorphic NBS marker, the band was excised from the gel and re-amplified with the same primers that initially produced the band. The PCR conditions were identical to the first PCR of the NBS profiling protocol. Only bands that were clearly separated from surrounding bands were considered. In the case of population 7358 it was necessary to clone the band first because direct sequencing showed that the band consisted of a mixture of two fragments. For this, the PCR products were then ligated into the pGEM-T easy Vector System (Promega). Ligation mixtures were transformed into *E. coli* DH5α, as recommended by the supplier (Invitrogen). Colonies containing a plasmid with insert were used for colony PCR. Clones were sequenced using vector M13 primers. For populations ver03-392, ver03-394, and snk7458 the bands were directly sequenced following reamplification.

Each fragment was sequenced from both sides with the NBS profiling primers using the Big Dye Terminator Kit on an ABI 3700 automated sequencer (Applied Biosystems, USA). DNA sequences were analyzed using DNAstar (Lasergene, Madison, WI, USA). The obtained sequences were compared to the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/), by using BLASTN suite (Altschul *et al.* 1997) from NCBI. The similarity scores with sequences found in the NCBI database were evaluated taking into account the E-value. The E-value is also dependent from the length of the query sequence that can be blasted to a certain sequence in the database. The shorter the sequence, the higher the possibility that the result is due by chance. A similarity was defined as a good hit if it showed a combination of a similarity (identity) score of 75% or higher plus a small expected value of 1.00E-25 or smaller.

**Confirmation of position with PCR- based primers**

To verify the putative map positions for NBS profiling markers deduced from the BLAST analysis, we used flanking markers (mainly CAPS). For each population the flanking markers were tested on the parents, the bulks and the individuals of the bulks. The position of sequenced NBS markers was confirmed with testing CAPS markers. The details on the primers that were used successfully are given in Table 2. To amplify the samples with PCR marker Th21, approximately 10 ng of genomic DNA was mixed in a total volume of 20 µl containing (end-concentration per reaction) 1x PCR buffer, 0.2 mM mixture of all dNTPs, 0.1 pm each primer and 0.1 unit Taq DNA polymerase (Promega). The following PCR protocol was used: a first step of 3 minutes at 96 °C, followed by 30 cycli of 0.5 minutes at 96 °C, 0.5 minutes at 56 °C, and 1 minute at 72 °C, concluding with 10 minutes at 72 °C. With the CAPS markers CP58 and CD67 a slightly different mixture and protocol was used. Approximately 10 ng of DNA was mixed in a total volume of 25µl containing 1 x PCR buffer, 0.12 mM dNTPs, 0.05 pM from each primer, and 0.1unit Super Taq DNA polymerase. The following PCR protocol was used: starting with 4 minutes at 94 °C, followed by 35 cycles of 0.5 minutes at 94 °C, 0.5 minutes of 58 °C, 1.5 minutes at 72 °C. At the end of the protocol, 6 minutes at 72°C were programmed. The presence of PCR products of the correct length was evaluated on a 1.0% agarose gel.

*Table 2 PCR primers used for confirmation*

| Population | Locus | Chrom. | forward primer | reverse primer | Enzyme |
|---|---|---|---|---|---|
| ver 03-392 / ver 03-394 | CD67 | 6 | CCCCTGCAAATCCGTACATA | CCATACGAGTTGAGGGATCG | *Hpyc*HIV, SSiL |
| snk 7458 | Th21 | 4 | ATTCAAAATTCTAGTTCCGCC | AACGGCAAAAAAGCACCAC | *Mbo*1 |
| cap 7358 | CP58 | 11 | ATGTATGGTTCGGGATCTGG | TTAGCACCAACAGCTCCTCT | *Msp*I |

## Results

### NBS Profiling

The populations differed strongly in the number of polymorphic bands in the bulks that showed co-segregation with resistance, ranging from 1 in population cap7358 to 33 in the ver03-392 population (Table 3). For population snk7458, cap7358 and ver03-394 all the primer/enzyme combinations that produced polymorphic bands were tested on the individuals of the bulks. In the population ver03-392, only a selection of primer/enzyme combinations producing polymorphic bands in the bulks was tested on the individuals. An example of an NBS gel for parents, bulks and individuals is given in Figure 1. Not all putative polymorphisms observed in the bulks were validated in the individuals (see Table 3). Most of the bands that were found and confirmed as co-segregating in the second round of NBS profiling on individuals of the bulks were bands in coupling phase, e.g. co-segregating with the resistant phenotype. However, also several bands in repulsion phase, e.g. co-segregating with the susceptible phenotype, were observed. For all NBS markers studied in this paper, the co-segregation of markers and resistance was 100% in the tested individuals, except for the NBS markers ver03-394_9H1, ver03-394_9R1, ver 03-394_9R2 which show 1 (identical) recombinant resistant plant (out of 7 resistant plants) that does not have the specific NBS fragment.

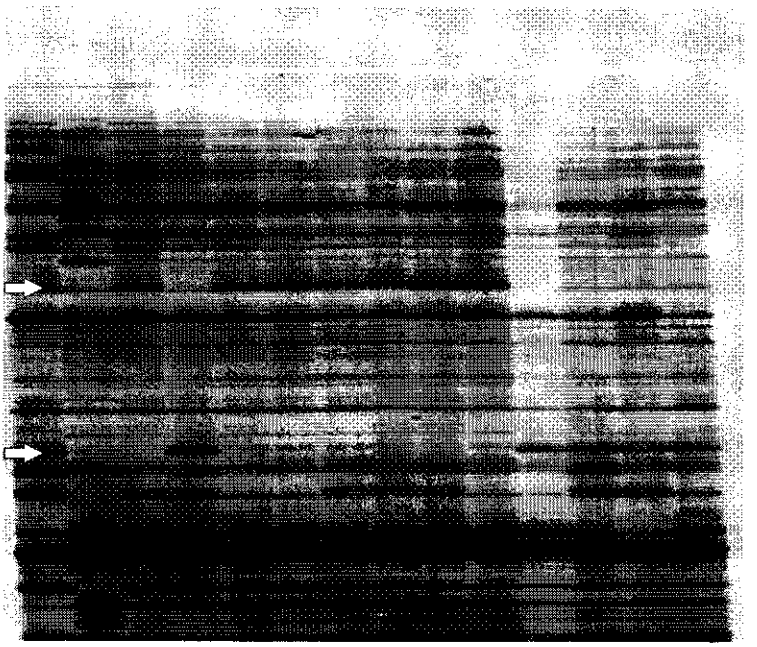


*Figure 1. An example of a part of a NBS profiling gel. This figure shows part of the NBS profiling gel of population snk7458 using NBS2 and Mse. The arrows indicate the segregating NBS profiling bands. The upper arrow points at at band in coupling phase, the lower arrow points at a band in repulsion phase.*

*Table 3 Summarized results NBS profiling on bulks and individuals. Bands in coupling or repulsion phase refers to the number of polymorphic bands that could be reproduced in the individuals that constituted the bulk. For population ver03-392 not all the primer-enzym-combinations that gave polymorphisms in the bulks were tested in the individual NBS profiling step. 8/18 in this case means that 8 out of the*

| Population | Polymorphic bands (bulks) | Bands coupling phase (individuals) | Bands repulsion phase (individuals) |
|---|---|---|---|
| ver 03-392* | 33 | 8/18 | 2/5 |
| ver 03-394 | 19 | 12/13 | 4/6 |
| snk 7458 | 10 | 4/5 | 5/5 |
| cap 7358 | 1 | 1/1 | 0/0 |

**Identification of NBS bands and deduction of mappositions**

The bands co-segregating with the resistance phenotype in the individuals were excised from the gels and sequenced to determine their identity. Sequences obtained were compared to the NCBI database. All the sequences gave hits with NBS related sequences (additional table), so we regard the sequences as RGAs. The best hits (high identity score and low E-value) to *Solanum* sequences are shown in Table 4 and the complete list of the 10 best hits is available as Additional file 1.

For population ver03-392, 2 bands were successfully sequenced, but only band 392-9H1, see Table 4, gave similarity scores higher than 75% with sequences from Genbank. Good hits were found with *Solanum lycopersicum* chromosome 11 clone C11HBa0119D16, complete sequence, and later with *Solanum lycopersicum* DNA, chromosome 8, clone: C08SLm0114A09, complete sequence. For population ver03-394, 5 different NBS bands could be sequenced successfully, 4 bands gave high similarity scores with Genbank sequences. The sequences found in the NCBI database that showed similarity with the NBS markers of ver03-394 were identical to those found for population ver03-392.

For population snk7458, sequences of 5 NBS profiling bands could be successfully retrieved. For all 5 bands, the highest similarity scores were found with *Solanum lycopersicum* BAC clone Clemson_Id 127E11. Park et al. (2005) showed that this BAC clone contains several RGA sequences that are similar to the *Rpi-blb3* gene of *S. bulbocastanum* which is located on chromosome 4 of potato (Park *et al.*, 2005a). Another high homology score with bands from snk7458 was found with a *S. lycopersicum* DNA sequence from clone SL_MboI-40B16, also located on chromosome 4.

For population cap7358, the only band found in the NBS profiling analysis with the bulk and the individual samples was successfully sequenced. When comparing this relatively short band with the sequences in the Genbank database, an identity score of 98% with a E value of 2.00E-25 (Table 4) was found with *S. tuberosum* mRNA for the NL27 protein. Hehl et al. (1999) located the gene encoding the NL27 protein on chromosome 11.

*Table 4. Blast results. Only the best hits (small E value and high maximum identity) to Solanum sequences are shown.*

| Population | NBS band | Length bp. | Accession | Description | E value | Max. ident. |
|---|---|---|---|---|---|---|
| ver 03-392 | 392_9h1 | 382 | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 1.00E-69 | 78% |
| | 392_9h1 | 382 | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 2.00E-68 | 78% |
| ver 03-394 | 394_9h1 | 382 | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 1.00E-69 | 78% |
| | 394_9h1 | 382 | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 2.00E-68 | 78% |
| | 394_9r1 | 284 | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 6.00E-53 | 81% |
| | 394_9r1 | 284 | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 4.00E-49 | 80% |
| | 394_9R2 | 284 | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 5.00E-54 | 81% |
| | 394_9R2 | 284 | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 3.00E-50 | 80% |
| | 394_9T1 | 340 | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 1.00E-49 | 75% |
| | 394_9T1 | 340 | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 6.00E-48 | 75% |
| | 7458_2A2 | 431 | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 1.00E-126 | 91% |
| snk 7458 | 7458_2A2 | 431 | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 1.00E-120 | 85% |
| | 7458_2M2 | 513 | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 2.00E-171 | 94% |
| | 7458_2M2 | 513 | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 2.00E-176 | 86% |

*Table 4. Blast results. Only the best hits (small E value and high maximum identity) to Solanum sequences are shown.*

| Population | NBS band | Length bp. | Accession | Description | E value | Max. ident. |
|---|---|---|---|---|---|---|
| | 7458_2M3 | 431 | CU326358.5 | S.lycopersicum DNA sequence from clone SL_MboI-40B16 on chromosome 4, complete sequence | 2.00E-131 | 94% |
| | 7458_2M3 | 431 | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 7.00E-124 | 87% |
| | 7458_2M4 | 238 | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 2.00E-79 | 94% |
| | 7458_2M4 | 238 | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 2.00E-78 | 94% |
| | 7458_2M5 | 203 | CU326358.5 | S.lycopersicum DNA sequence from clone SL_MboI-40B16 on chromosome 4, complete sequence | 2.00E-37 | 86% |
| | 7458_2M5 | 203 | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 5.00E-33 | 84% |
| cap 7358 | 7358_3M1 (1-10) | 130 | AJ009720.1 | Solanum tuberosum mRNA for NL27 protein | 2.00E-25 | 98% |

**Confirmation of map position of genes**

To verify the deduced map positions, we used markers that were expected to be (closely) linked based on their position on the potato maps (http://gabi.rzpd.de/projects/Pomamo/ and http://www.sgn.cornell.edu/). For each population flanking markers were tested on the parents, the bulks and the individuals of the bulks. For populations ver03-392 / ver03-394, sequence homology suggested that the resistance gene was positioned on chromosome 11. Several CAPS markers for chromosome 11 were tested, but none of those displayed any polymorphisms nor co-segregation. The marker sequence was compared to a sequence database containing NBS profiling marker sequences that were mapped in the SHxRH potato mapping population (van der Linden et al, unpublished results), and was found to be nearly identical to a marker mapped on chromosome 6. This mapping position was confirmed by marker CD67 digested with enzyme *Hpyc*HIV and with enzyme *Ssi*I that both produced a polymorphic band that co-segregated with the resistance (see Figure 2). An extra band is visible in the resistant parent and the resistant offspring in both *S. verrucosum* populations.

The *P. infestans* resistance in population snk7458 was suggested to be located in the same cluster as the *Rpi-blb3* gene on chromosome 4. Several markers for chromosome 4, such as TG370, Th21, TG506R, CT229, T1430, were tested. The parents often showed polymorphisms, but the offspring was almost always homogeneous for the same marker. Therefore, another approach was taken to find segregating markers. The PCR products of markers TG506, AF411807R, T1430, TG370 and Th21 from the 6 resistant en 6 susceptible individuals were sequenced and checked for SNPs. The only polymorphism that was found between the resistant and susceptible individuals was in marker Th21. This SNP was shown to co-segregate with resistance. The SNP is located in an *Mbo* 1 site in the middle of the PCR fragment: TGATC for the susceptible, G[A/G]ATC for the (heterozygous) resistant individuals). A PCR with Th21 followed by digestion with *Mbo*1 on all the available individuals, resulted in an extra band for the resistant parent *S. schenckii* GLKS 30659 and in 32 out of 45 resistance phenotypes. The 8 susceptible phenotypes all lacked the extra band. This means that at least 13 out of 45 resistant plants contain a second gene conferring resistance to *P. infestans*.

To confirm the position of the marker in population cap7358, several SSR and CAPS primers for chromosome 11 were tested on the parents, the bulks and the individuals of this population. The CAPS marker CP58 in combination with restriction enzyme *Msp*I produced an extra band in the resistant parent and in 10 out of 10 susceptible offspring. Nine of ten resistant individuals lacked this extra band.

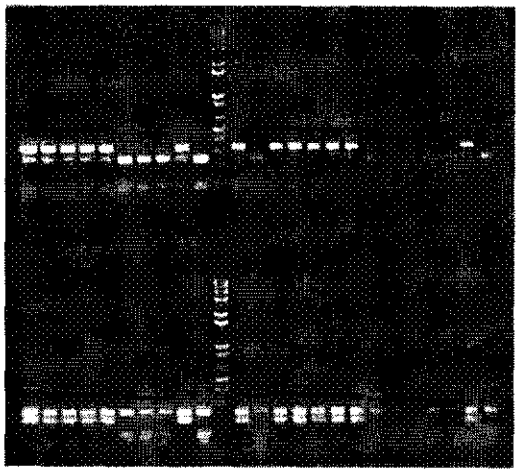


*Figure 2 Marker CD67 shows co-segregation with P. infestans resistance in populations ver03-392 and ver03-394 after digestion with HpycHIV*

## Discussion and Conclusions

**A new strategy for mapping resistance genes**

In the present study we describe a novel approach to identify markers linked to resistance genes that can be used directly for introgression breeding. The first step in the approach consists of producing small populations segregating for *P. infestans* resistance, phenotyping the populations for resistance and composing bulks of resistant and susceptible individuals. Then, the bulks are genotyped using NBS profiling to obtain markers that co-segregate with resistance, followed by sequencing of co-segregating NBS fragments and BLAST analysis to identify the fragment. Combining this information with literature data on mapping of resistance genes results in a suggestion for a putative map position. Finally, the map positions were confirmed using known flanking markers.

Large differences were observed in the number of markers co-segregating with resistance in the 3 different populations, ranging from 1 in population snk7358 to 33 in the ver03-392 population (Table 3). These differences are possibly caused by the position of the targeted R gene. Many polymorphic bands co-segregating with resistance probably mean that they are part of a large cluster of R genes. Little or no polymorphic NBS bands could mean that the parents were closely related and that the targeted R gene has an isolated position. In case no polymorphisms are detected with the 25 primer-enzyme combinations additional enzymes or primers may be tested. Furthermore, there is always a chance that the resistance gene under study is not of the NBS-LRR type but belongs to another class of resistance genes (Ingvardsen *et al.*, 2008). The polymorphisms identified in the bulks that could not be confirmed in the analysis of the individual genotypes were due to differences in band intensity.

The polymorphisms that showed a clear presence/absence of bands could be confirmed. All NBS bands that were sequenced successfully could be annotated as being from a putative NBS-LRR type of resistance gene (see Table 4 and additional Table 1). Correct annotation depends on the availability of sufficient sequence information in the databases. In addition, highly homologous sequences may sometimes be found in different clusters, as was shown for *Mi* and *I2* homologues (Seah *et al.*, 2007; Van der Vossen *et al.*, 2005). This may complicate mapping afterwards, as was shown for fragment 392_9H1, obtained from population ver 392. For this fragment, high homology was found with a sequence of *S. lycopersicum* which had previously been mapped to chromosome 11 (Mueller & Tanksley, 2008; Mueller *et al.*, 2005), while another sequence homolog was retrieved from chromosome 4 (McGuire 2008). The positions on chromosome 11, or chromosome 8 or 4, (Table 4), as suggested by BLAST with NCBI database for several bands from populations ver03-392 and ver03-394, could not be verified. The putative map position on chromosome 6 for this marker, inferred from the high sequence similarity to a mapped NBS profiling marker (Van der Linden et al., unpublished results) could be verified. This indicates that the NCBI database is still far from complete. With the increasing amount of data deposited in public sequence databases and with the progress of the potato genome sequencing (PGSC, http://www.potatogenome.net) it is likely that in the near future new fragments can be mapped more efficient and with higher accuracy.

**Identification and mapping of *P. infestans* resistance genes in *S. verrucosum*, *S. schenckii*, and *S. capsicibaccatum***

In the ver03-392 and ver03-394 populations, the resistance against *P. infestans* is located on chromosome 6, near marker CD67. Population ver03-394 showed one recombinant for the polymorphism found with marker CD67. We have named the gene underlying the resistance *Rpi-ver1*. As we could not find any other co-segregating markers for the resistance, it is not clear whether the gene is positioned downstream or upstream of marker CD67. The marker 67 it self is positioned at 10.50 cM according to the potato map Potato-TXB 1992 v27 (http://www.sgn.cornell.edu/). Resistance against *P. infestans* in *S. verrucosum* has been reported by Van Soest et al. (1984). Rivera-Peña (1989, 1990) studied the occurrence of late blight on naturally occurring populations of wild *Solanum* species on the slope of Nevado de Toluca for many years. He found highly resistant natural populations of *S. verrucosum*. Furthermore, a *P. infestans* resistance screening of the Commonwealth potato collection also yielded a very resistant *S. verrucosum* accession (Bradshaw *et al.*, 2006). Whether any of the genes involved is similar to the *Rpi-ver1* remains to be established. Finally, Liu and Halterman (2006) reported on *P. infestans* resistance in *S. verrucosum*. They have identified a gene sharing 83,5% nucleotide identity with *Rpi-blb1*. It will be interesting to see whether this gene maps to the same position on chromosome 8 as the original *Rpi-blb1* (Van der Vossen et al. 2003) or to the same position as *Rpi-ver1* on chromosome 6, or to a complete new position. There were no indications that the sequences of the NBS markers of populations 03-392 and ver03-394 show any similarity to the original *Rpi-blb1*.

In the snk 7458 population, there are probably 2 resistance genes against *P. infestans* segregating. The offspring consists of far more resistant than susceptible phenotypes (45:8). Furthermore, only 2/3 of the resistant genotypes contained the linked CAPS marker. None of the susceptible phenotypes contained the CAPS marker. The result suggests the possible presence of another gene or QTL that confers the phenotypic resistance. This situation is very similar to that described by Wang et al. (2008) for *S. stoloniferum*. Based on the results, one would also expect the presence of polymorphic NBS bands with less than 100% co-segregation. However, the five NBS markers tested on 6 resistance and 4 susceptible individuals from the bulk showed 100% co-segregation for the NBS markers. It is possible that the polymorphic NBS bands that were linked to the other gene were not discovered in the first round of NBS profiling on the bulked individuals. One gene conferring resistance in the *S. schenckii* population 7458, and mapped in this study is located on chromosome 4, near or on marker Th21 (Table 4). We call the gene underlying the resistance *Rpi-snk1*. The second (non-mapped) gene we named *Rpi-snk2*. The *S. schenckii Rpi*-snk1 gene is an *Rpi-blb3* homolog (Park *et al.*, 2005b) that fully co-segregates with the Th21 marker, and therefore also similar to *Rpi-abpt*, *R2* and *R2 like* that all reside in the same R gene cluster on chromosome 4 and likely belongs to the same family (Park *et al.*, 2005b). According to the phenetic and phylogenetic results of Jacobs et al. (2008) *S. schenckii* is closely related to *S. hougasii*. Accessions from *S. hougasii* are reported to show high resistance against to *P. infestans* (Bradshaw *et al.*, 2006). It will be interesting to see whether these accessions also carry the same resistance genes as found in *S. schenckii*.

In population cap 7358, a gene conferring resistance against *P. infestans* was found on chromosome 11, near marker CP58. The recombination percentage between *P. infestans* resistance and CP58 in population cap 7358 is 5%. The recombination percentage between the *P. infestans* resistance and the NBS marker is 0%. The newly found gene is named *Rpi-cap1*. The position of CP58 is at the top of the chromosome 11, on 0.00 cM according to the data from the map Potato-TXB 1992 V27 (http://www.sgn.cornell.edu/). Another resistance gene that was mapped to the this region is *R-Mc1*, (mapped at 66 cM in the functional map of chromosome 11 of potato for pathogen resistance, as published at http://gabi.rzpd.de/index.shtml. Note that the orientation of this potato map of chromosome 11 is reversed compared to the previously mentioned SGN potato map of chromosome 11) which is a resistance gene against root-knot nematode *Meloidogyne chitwoodi* retrieved from *S. bulbocastanum* (Brown *et al.* 1996). Resistance against *P. infestans* in *S. capsicibaccatum* was reported by Van Soest et al. (1984) and Ruiz de Galarreta et al. (1998) but no further details on sequence or position of R genes were given.

In the mapping populations used in this study no new resistance gene clusters in *Solanum* material were found. The *P. infestans* resistance was derived from different wild *Solanum* species in which previously no resistance genes had been identified (though some species had been reported to express some *P. infestans* resistance). Although we used these relatively unknown sources, it seems that the genes conferring the resistance are linked to known clusters of resistance genes. This may suggest that the present view on the *Solanum* genome is rather exhaustive and that most resistance clusters are already known. In a previous study, new resistance genes in *Solanum* derived from wild *Solanum* species, could also be positioned at already known R gene clusters of the *Solanum* genome. Wang et al. (2008) found that the dominant R genes *Rpi-sto1* (derived from *S. stoloniferum*) and *Rpi-plt1* (from *S. polytrichon*) resided at the same position on chromosome 8 as *Rpi-blb1* in *S. bulbocastanum*. Possibly, the found R genes on known loci contain new alleles but new alleles can be positively identified with the aid of effector proteins (Vleeshouwers *et al.* 2008).

**Acknowledgements**

This project was co-financed by the research program of the Centre for Biosystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research. We thank Agrico Research B.V. for kindly providing populations Ver 03-292 and Ver 03-394. Furthermore, we thank the following persons for their contribution to the experimental part of this project: Roel Hoekstra (Centre for Genetic Resources, The Netherlands), Gerard van der Linden, Linda Kodde, Martijn van Kaauwen, Christel Denneboom, Dirk Budding, Ronald Hutten, Edwin van der Vossen (all WUR Plant Breeding, Wageningen University and Research Centre), Guus Heselmans (Meijer B.V. The Netherlands) and Sjefke Allefs (Agrico Research B.V.) Gerard van der Linden and Sjefke Allefs are thanked for their useful comments on earlier versions of the manuscript.

*Additional file 1. All the BLAST results for the NBS bands.*
**Blast algorithms: Megablast is intended for comparing a query to closely related sequences and works best if the target percent identity is 95% or more, but is very fast. Discontiguous megablast uses an initial seed that ignores some bases (allowing mismatches) and is intended for cross-species comparisons. BlastN is slow, but allows a word-size down to seven bases.*

| NBS band | BLAST algorithm* | Accession | Description | Max score | Total score | Query coverage | E value | Max ident |
|---|---|---|---|---|---|---|---|---|
| 392_2r2 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 67.6 | 67.6 | 17% | 2.00E-08 | 100% |
| 392_2r2 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 67.6 | 67.6 | 17% | 2.00E-08 | 100% |
| 392_2r2 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 67.6 | 67.6 | 17% | 2.00E-08 | 100% |
| 392_2r2 | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 17% | 5.00E-08 | 100% |
| 392_2r2 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 17% | 5.00E-08 | 100% |
| 392_2r2 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 17% | 5.00E-08 | 100% |
| 392_2r2 | discontiguous megablast | AC198104.5 | MACACA MULATTA BAC clone CH250-357M16 from chromosome 11, complete sequence | 42.8 | 42.8 | 23% | 0.54 | 79% |
| 392_2r2 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 17% | 5.00E-08 | 100% |
| 392_2r2 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 17% | 5.00E-08 | 100% |
| 392_2r2 | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 17% | 5.00E-08 | 100% |
| 392_2r2 | nblast | CU326400.10 | S.lycopersicum DNA sequence from clone SL_Mbol-39E17 on chromosome 4, complete sequence | 53.6 | 53.6 | 25% | 3.00E-04 | 82% |
| 392_2r2 | nblast | AF528092.1 | Fagus sylvatica microsatllite FS3-04 sequence | 50 | 50 | 13% | 0.004 | 100% |
| 392_2r2 | nblast | BX950218.13 | Zebrafish DNA sequence from clone DKEY-276J7 in linkage group 6, complete sequence | 44.6 | 44.6 | 15% | 0.16 | 90% |
| 392_2r2 | nblast | AL844190.5 | complete sequence | 44.6 | 44.6 | 15% | 0.16 | 93% |
| 392_2r2 | nblast | AP009592.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08HBa0018C13, complete sequence | 42.8 | 42.8 | 16% | 0.54 | 88% |
| 392_2r2 | nblast | AJ494989.1 | Lipomhopalum tentacularis microsatellite DNA, clone WLT4-27 | 42.8 | 42.8 | 11% | 0.54 | 100% |
| 392_2r2 | nblast | AJ494987.1 | Lipomhopalum tentacularis microsatellite DNA, clone WLT4-19 | 42.8 | 42.8 | 11% | 0.54 | 100% |
| 392_9h1 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 67.6 | 67.6 | 9% | 4.00E-08 | 100% |
| 392_9h1 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 67.6 | 67.6 | 9% | 4.00E-08 | 100% |
| 392_9h1 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 67.6 | 67.6 | 9% | 4.00E-08 | 100% |
| 392_9h1 | discontiguous megablast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 271 | 271 | 86% | 1.00E-69 | 78% |
| 392_9h1 | discontiguous megablast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 268 | 268 | 86% | 2.00E-68 | 78% |
| 392_9h1 | discontiguous megablast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 264 | 264 | 86% | 2.00E-67 | 78% |
| 392_9h1 | discontiguous megablast | CU012067.6 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 259 | 259 | 86% | 8.00E-66 | 78% |
| 392_9h1 | discontiguous megablast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 250 | 250 | 86% | 4.00E-63 | 77% |
| 392_9h1 | discontiguous megablast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 237 | 237 | 79% | 3.00E-59 | 78% |
| 392_9h1 | discontiguous megablast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 87.8 | 87.8 | 40% | 3.00E-14 | 72% |
| 392_9h1 | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 392_9h1 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 392_9h1 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 392_9h1 | nblast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 271 | 271 | 86% | 1.00E-69 | 78% |
| 392_9h1 | nblast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 268 | 268 | 86% | 2.00E-68 | 78% |
| 392_9h1 | nblast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 264 | 264 | 86% | 2.00E-67 | 78% |
| 392_9h1 | nblast | CU012067.6 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 259 | 259 | 86% | 8.00E-66 | 78% |
| 392_9h1 | nblast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 250 | 250 | 86% | 4.00E-63 | 77% |
| 392_9h1 | nblast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 237 | 237 | 79% | 3.00E-59 | 78% |
| 392_9h1 | nblast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 120 | 208 | 41% | 5.00E-24 | 77% |
| 392_9h1 | nblast | AC215460.2 | Solanum lycopersicum chromosome 2 clone C02SLe0091I10, complete sequence | 98.7 | 145 | 58% | 2.00E-17 | 76% |
| 392_9h1 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 392_9h1 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 394_9h1 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 67.6 | 67.6 | 9% | 4.00E-08 | 100% |

*Additional file 1. (Continued) All the BLAST results for the NBS bands.*

| NBS band | BLAST algorithm* | Accession | Description | Max score | Total score | Query coverage | E value | Max ident |
|---|---|---|---|---|---|---|---|---|
| 394_9h1 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 67.6 | 67.6 | 9% | 4.00E-08 | 100% |
| 394_9h1 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 67.6 | 67.6 | 9% | 4.00E-08 | 100% |
| 394_9h1 | discontiguous megablast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 271 | 271 | 86% | 1.00E-69 | 78% |
| 394_9h1 | discontiguous megablast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 268 | 268 | 86% | 2.00E-68 | 78% |
| 394_9h1 | discontiguous megablast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 264 | 264 | 86% | 2.00E-67 | 78% |
| 394_9h1 | discontiguous megablast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 259 | 259 | 86% | 8.00E-66 | 78% |
| 394_9h1 | discontiguous megablast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 250 | 250 | 86% | 4.00E-63 | 77% |
| 394_9h1 | discontiguous megablast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 237 | 237 | 79% | 3.00E-59 | 78% |
| 394_9h1 | discontiguous megablast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 87.8 | 87.8 | 40% | 3.00E-14 | 72% |
| 394_9h1 | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 394_9h1 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 394_9h1 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 394_9h1 | nblast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 271 | 271 | 86% | 1.00E-69 | 78% |
| 394_9h1 | nblast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 268 | 268 | 86% | 2.00E-68 | 78% |
| 394_9h1 | nblast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 264 | 264 | 86% | 2.00E-67 | 78% |
| 394_9h1 | nblast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 259 | 259 | 86% | 8.00E-66 | 78% |
| 394_9h1 | nblast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 250 | 250 | 86% | 4.00E-63 | 77% |
| 394_9h1 | nblast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 237 | 237 | 79% | 3.00E-59 | 78% |
| 394_9h1 | nblast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 120 | 208 | 40% | 5.00E-24 | 77% |
| 394_9h1 | nblast | AC215460.2 | Solanum lycopersicum chromosome 2 clone C02SLe0091I10, complete sequence | 98.7 | 145 | 58% | 2.00E-17 | 76% |
| 394_9h1 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 394_9h1 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 9% | 1.00E-07 | 100% |
| 394_9h2 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 67.6 | 67.6 | 9% | 3.00E-08 | 100% |
| 394_9h2 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 67.6 | 67.6 | 9% | 3.00E-08 | 100% |
| 394_9h2 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 67.6 | 67.6 | 9% | 3.00E-08 | 100% |
| 394_9h2 | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 9% | 9.00E-08 | 100% |
| 394_9h2 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 9% | 9.00E-08 | 100% |
| 394_9h2 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 9% | 9.00E-08 | 100% |
| 394_9h2 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 9% | 9.00E-08 | 100% |
| 394_9h2 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 9% | 9.00E-08 | 100% |
| 394_9h2 | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 9% | 9.00E-08 | 100% |
| 394_9h2 | nblast | AC215367.2 | Solanum lycopersicum chromosome 2 clone C02HBa0044L14, complete sequence | 48.2 | 48.2 | 19% | 0.024 | 75% |
| 394_9h2 | nblast | AC217002.1 | Solanum lycopersicum chromosome 11 clone C11HBa0072I13, complete sequence | 48.2 | 48.2 | 19% | 0.024 | 75% |
| 394_9h2 | nblast | AJ494984.1 | Lipomhopalum tentaculatis microsatellite DNA, clone WLT3-11 | 46.4 | 46.4 | 9% | 0.085 | 91% |
| 394_9h2 | nblast | AP010785.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08HBa0040B24, complete sequence | 44.6 | 44.6 | 19% | 0.3 | 73% |
| 394_9h2 | nblast | CU074337.8 | S.lycopersicum DNA sequence from clone SL_Mbol-143K21 on chromosome 4, complete sequence | 44.6 | 44.6 | 17% | 0.3 | 75% |
| 394_9h2 | nblast | AJ494989.1 | Lipomhopalum tentaculatis microsatellite DNA, clone WLT4-27 | 44.6 | 44.6 | 6% | 0.3 | 100% |
| 394_9h2 | nblast | AJ494987.1 | Lipomhopalum tentaculatis microsatellite DNA, clone WLT4-19 | 44.6 | 44.6 | 6% | 0.3 | 100% |
| 394_9r1 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 65.8 | 65.8 | 12% | 1.00E-07 | 100% |
| 394_9r1 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 65.8 | 65.8 | 12% | 1.00E-07 | 100% |
| 394_9r1 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 65.8 | 65.8 | 12% | 1.00E-07 | 100% |

*Additional file 1. (Continued) All the BLAST results for the NBS bands.*

| NBS band | BLAST algorithm* | Accession | Description | Max score | Total score | Query coverage | E value | Max ident |
|---|---|---|---|---|---|---|---|---|
| 394_9r1 | discontiguous megablast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 215 | 215 | 83% | 6.00E-53 | 81% |
| 394_9r1 | discontiguous megablast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 203 | 203 | 83% | 4.00E-49 | 80% |
| 394_9r1 | discontiguous megablast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 203 | 203 | 83% | 4.00E-49 | 80% |
| 394_9r1 | discontiguous megablast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 203 | 203 | 83% | 4.00E-49 | 80% |
| 394_9r1 | discontiguous megablast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 188 | 188 | 80% | 8.00E-45 | 79% |
| 394_9r1 | discontiguous megablast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 64.4 | 64.4 | 22% | 2.00E-07 | 82% |
| 394_9r1 | discontiguous megablast | AC099789.2 | Homo sapiens chromosome 1 clone RP11-377K22, complete sequence | 42.8 | 42.8 | 13% | 0.79 | 84% |
| 394_9r1 | nblast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 215 | 215 | 83% | 6.00E-53 | 81% |
| 394_9r1 | nblast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 203 | 203 | 83% | 4.00E-49 | 80% |
| 394_9r1 | nblast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 203 | 203 | 83% | 4.00E-49 | 80% |
| 394_9r1 | nblast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 203 | 203 | 83% | 4.00E-49 | 80% |
| 394_9r1 | nblast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 188 | 188 | 80% | 8.00E-45 | 79% |
| 394_9r1 | nblast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 170 | 170 | 83% | 2.00E-39 | 76% |
| 394_9r1 | nblast | AC215460.2 | Solanum lycopersicum chromosome 2 clone C02SLe0091I10, complete sequence | 98.7 | 98.7 | 50% | 1.00E-17 | 76% |
| 394_9r1 | nblast | AC215399.2 | Solanum lycopersicum chromosome 2 clone C02HBa0155E05, complete sequence | 64.4 | 64.4 | 48% | 2.00E-07 | 72% |
| 394_9r1 | nblast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 64.4 | 127 | 22% | 2.00E-07 | 82% |
| 394_9r1 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 64.4 | 64.4 | 12% | 2.00E-07 | 100% |
| 394_9R2 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 67.6 | 67.6 | 12% | 3.00E-08 | 100% |
| 394_9R2 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 67.6 | 67.6 | 12% | 3.00E-08 | 100% |
| 394_9R2 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 67.6 | 67.6 | 12% | 3.00E-08 | 100% |
| 394_9R2 | discontiguous megablast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 219 | 219 | 82% | 5.00E-54 | 81% |
| 394_9R2 | discontiguous megablast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 206 | 206 | 82% | 3.00E-50 | 80% |
| 394_9R2 | discontiguous megablast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 206 | 206 | 82% | 3.00E-50 | 80% |
| 394_9R2 | discontiguous megablast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 206 | 206 | 82% | 3.00E-50 | 80% |
| 394_9R2 | discontiguous megablast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 192 | 192 | 79% | 7.00E-46 | 79% |
| 394_9R2 | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 12% | 7.00E-08 | 100% |
| 394_9R2 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 12% | 7.00E-08 | 100% |
| 394_9R2 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 12% | 7.00E-08 | 100% |
| 394_9R2 | discontiguous megablast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 64.4 | 64.4 | 21% | 2.00E-07 | 82% |
| 394_9R2 | discontiguous megablast | AC099789.2 | Homo sapiens chromosome 1 clone RP11-377K22, complete sequence | 42.8 | 42.8 | 13% | 0.8 | 84% |
| 394_9R2 | nblast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 219 | 219 | 82% | 5.00E-54 | 81% |
| 394_9R2 | nblast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 206 | 206 | 82% | 3.00E-50 | 80% |
| 394_9R2 | nblast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 206 | 206 | 82% | 3.00E-50 | 80% |
| 394_9R2 | nblast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 206 | 206 | 82% | 3.00E-50 | 80% |
| 394_9R2 | nblast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mbol-58F7 on chromosome 4, complete sequence | 192 | 192 | 79% | 7.00E-46 | 79% |
| 394_9R2 | nblast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 174 | 174 | 82% | 2.00E-40 | 76% |
| 394_9R2 | nblast | AC215460.2 | Solanum lycopersicum chromosome 2 clone C02SLe0091I10, complete sequence | 98.7 | 98.7 | 49% | 1.00E-17 | 76% |
| 394_9R2 | nblast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 68 | 132 | 22% | 2.00E-08 | 83% |
| 394_9R2 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 12% | 7.00E-08 | 100% |
| 394_9R2 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 12% | 7.00E-08 | 100% |
| 394_9T1 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 58.4 | 58.4 | 9% | 2.00E-05 | 100% |

*Additional file 1. (Continued) All the BLAST results for the NBS bands.*

| NBS band | BLAST algorithm* | Accession | Description | Max score | Total score | Query coverage | E value | Max ident |
|---|---|---|---|---|---|---|---|---|
| 394_9T1 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 58.4 | 58.4 | 9% | 2.00E-05 | 100% |
| 394_9T1 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 58.4 | 58.4 | 9% | 2.00E-05 | 100% |
| 394_9T1 | discontiguous megablast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 205 | 205 | 88% | 1.00E-49 | 75% |
| 394_9T1 | discontiguous megablast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 199 | 199 | 85% | 6.00E-48 | 75% |
| 394_9T1 | discontiguous megablast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 197 | 197 | 88% | 2.00E-47 | 75% |
| 394_9T1 | discontiguous megablast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 188 | 188 | 87% | 1.00E-44 | 74% |
| 394_9T1 | discontiguous megablast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mboi-58F7 on chromosome 4, complete sequence | 183 | 183 | 88% | 4.00E-43 | 75% |
| 394_9T1 | discontiguous megablast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 178 | 178 | 88% | 2.00E-41 | 74% |
| 394_9T1 | discontiguous megablast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 156 | 156 | 82% | 6.00E-35 | 72% |
| 394_9T1 | nblast | AC171735.3 | Solanum lycopersicum chromosome 11 clone C11HBa0119D16, complete sequence | 205 | 205 | 88% | 1.00E-49 | 75% |
| 394_9T1 | nblast | AP010265.1 | Solanum lycopersicum DNA, chromosome 8, clone: C08SLm0114A09, complete sequence | 199 | 199 | 85% | 6.00E-48 | 75% |
| 394_9T1 | nblast | AC215440.2 | Solanum lycopersicum chromosome 2 clone C02HBa0323C04, complete sequence | 197 | 197 | 88% | 2.00E-47 | 75% |
| 394_9T1 | nblast | AC215465.1 | Solanum lycopersicum chromosome 2 clone C02SLe0123B22, complete sequence | 188 | 188 | 87% | 1.00E-44 | 74% |
| 394_9T1 | nblast | CU457812.8 | S.lycopersicum DNA sequence from clone SL_Mboi-58F7 on chromosome 4, complete sequence | 183 | 183 | 88% | 4.00E-43 | 75% |
| 394_9T1 | nblast | CU012067.8 | S.lycopersicum DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence | 178 | 178 | 88% | 2.00E-41 | 74% |
| 394_9T1 | nblast | AC213008.2 | Solanum lycopersicum chromosome 12 clone LE_HBa-144B17, complete sequence | 156 | 280 | 82% | 6.00E-35 | 72% |
| 394_9T1 | nblast | AC215460.2 | Solanum lycopersicum chromosome 2 clone C02SLe0091I10, complete sequence | 109 | 109 | 57% | 8.00E-21 | 72% |
| 394_9T1 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 57.2 | 57.2 | 9% | 4.00E-05 | 100% |
| 394_9T1 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 57.2 | 57.2 | 9% | 4.00E-05 | 100% |
| 7358_3M1 (1-10) | megablast | AJ009720.1 | Solanum tuberosum mRNA for NL27 protein | 122 | 122 | 53% | 2.00E-25 | 98% |
| 7358_3M1 (1-10) | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 73.1 | 73.1 | 30% | 2.00E-10 | 100% |
| 7358_3M1 (1-10) | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 73.1 | 73.1 | 30% | 2.00E-10 | 100% |
| 7358_3M1 (1-10) | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 73.1 | 73.1 | 30% | 2.00E-10 | |
| 7358_3M1 (1-10) | discontiguous megablast | AJ009720.1 | Solanum tuberosum mRNA for NL27 protein | 127 | 127 | 69% | 9.00E-27 | 91% |
| 7358_3M1 (1-10) | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 71.6 | 71.6 | 30% | 6.00E-10 | 100% |
| 7358_3M1 (1-10) | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 71.6 | 71.6 | 30% | 6.00E-10 | 100% |
| 7358_3M1 (1-10) | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 71.6 | 71.6 | 30% | 6.00E-10 | 100% |
| 7358_3M1 (1-10) | discontiguous megablast | AC171266.26 | Medicago truncatula clone mth2-14o4, complete sequence | 48.2 | 48.2 | 31% | 0.007 | 85% |
| 7358_3M1 (1-10) | discontiguous megablast | AC145513.44 | Medicago truncatula chromosome 8 clone mth2-26m10, complete sequence | 48.2 | 48.2 | 31% | 0.007 | 85% |
| 7358_3M1 (1-10) | discontiguous megablast | EU024838.1 | Fragaria vesca subsp. americana clone fosmid 19M24, complete sequence | 46.4 | 46.4 | 37% | 0.025 | 81% |
| 7358_3M1 (1-10) | nblast | AJ009720.1 | Solanum tuberosum mRNA for NL27 protein | 127 | 127 | 69% | 9.00E-27 | 91% |
| 7358_3M1 (1-10) | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 71.6 | 110 | 50% | 6.00E-10 | 100% |
| 7358_3M1 (1-10) | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 71.6 | 110 | 50% | 6.00E-10 | 100% |
| 7358_3M1 (1-10) | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 71.6 | 110 | 50% | 6.00E-10 | 100% |
| 7358_3M1 (1-10) | nblast | EF091690.1 | Nicotiana tabacum TMV resistance protein N (CN) mRNA, complete cds | 51.8 | 51.8 | 71% | 6.00E-04 | 72% |
| 7358_3M1 (1-10) | nblast | AM470012.2 | Vitis vinifera contig VV78X115420.3, whole genome shotgun sequence | 51.8 | 51.8 | 35% | 6.00E-04 | 84% |
| 7358_3M1 (1-10) | nblast | AM457800.2 | Vitis vinifera contig VV78X040047.2, whole genome shotgun sequence | 51.8 | 51.8 | 35% | 6.00E-04 | 84% |
| 7358_3M1 (1-10) | nblast | AC210656.1 | Populus trichocarpa clone POP061-K03, complete sequence | 50 | 50 | 36% | 0.002 | 82% |
| 7358_3M1 (1-10) | nblast | AM484783.2 | Vitis vinifera contig VV78X072046.2, whole genome shotgun sequence | 50 | 50 | 36% | 0.002 | 82% |
| 7358_3M1 (1-10) | nblast | AC182652.2 | Populus trichocarpa clone Pop1-120B14, complete sequence | 50 | 50 | 36% | 0.002 | 82% |
| 7458_2A2 | megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mboi-40B16 on chromosome 4, complete sequence | 429 | 652 | 86% | 5.00E-117 | 93% |

*Additional file 1. (Continued) All the BLAST results for the NBS bands.*

| NBS band | BLAST algorithm* | Accession | Description | Max score | Total score | Query coverage | E value | Max Ident |
|---|---|---|---|---|---|---|---|---|
| 7458_2A2 | megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 387 | 1492 | 89% | 3.00E-104 | 86% |
| 7458_2A2 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 67.6 | 67.6 | 8% | 4.00E-08 | 100% |
| 7458_2A2 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 67.6 | 67.6 | 8% | 4.00E-08 | 100% |
| 7458_2A2 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 67.6 | 67.6 | 8% | 4.00E-08 | 100% |
| 7458_2A2 | discontiguous megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 461 | 974 | 89% | 1.00E-126 | 91% |
| 7458_2A2 | discontiguous megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 441 | 2674 | 91% | 1.00E-120 | 85% |
| 7458_2A2 | discontiguous megablast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 107 | 107 | 26% | 4.00E-20 | 76% |
| 7458_2A2 | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 8% | 1.00E-07 | 100% |
| 7458_2A2 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 8% | 1.00E-07 | 100% |
| 7458_2A2 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 8% | 1.00E-07 | 100% |
| 7458_2A2 | nblast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 461 | 974 | 89% | 1.00E-126 | 91% |
| 7458_2A2 | nblast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 441 | 2674 | 91% | 1.00E-120 | 85% |
| 7458_2A2 | nblast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 107 | 107 | 26% | 4.00E-20 | 76% |
| 7458_2A2 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 66.2 | 66.2 | 8% | 1.00E-07 | 100% |
| 7458_2A2 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 66.2 | 66.2 | 8% | 1.00E-07 | 100% |
| 7458_2A2 | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 66.2 | 66.2 | 8% | 1.00E-07 | 100% |
| 7458_2A2 | nblast | DQ465448.1 | Arabidopsis thaliana ecotype Tamm_46 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 51.8 | 51.8 | 12% | 0.002 | 81% |
| 7458_2A2 | nblast | DQ465446.1 | Arabidopsis thaliana ecotype Pu2_8 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 51.8 | 51.8 | 12% | 0.002 | 81% |
| 7458_2A2 | nblast | DQ465444.1 | Arabidopsis thaliana ecotype Nfe_17 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 51.8 | 51.8 | 12% | 0.002 | 81% |
| 7458_2A2 | nblast | DQ465442.1 | Arabidopsis thaliana ecotype Lov_2 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 51.8 | 51.8 | 12% | 0.002 | 81% |
| 7458_2A2 | discontiguous megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 461 | 974 | 89% | 1.00E-126 | 91% |
| 7458_2A2 | nblast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 461 | 974 | 89% | 1.00E-126 | 91% |
| 7458_2A2 | discontiguous megablast | CU856252.3 | S.lycopersicum DNA sequence from clone SL_Mbol-127L14 on chromosome 4, complete sequence | 441 | 2681 | 91% | 1.00E-120 | 85% |
| 7458_2A2 | discontiguous megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 441 | 2674 | 91% | 1.00E-120 | 85% |
| 7458_2A2 | nblast | CU856252.3 | S.lycopersicum DNA sequence from clone SL_Mbol-127L14 on chromosome 4, complete sequence | 441 | 2681 | 91% | 1.00E-120 | 85% |
| 7458_2A2 | nblast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 441 | 2674 | 91% | 1.00E-120 | 85% |
| 7458_2A2 | megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 429 | 652 | 86% | 5.00E-117 | 93% |
| 7458_2A2 | megablast | CU856252.3 | S.lycopersicum DNA sequence from clone SL_Mbol-127L14 on chromosome 4, complete sequence | 390 | 1496 | 89% | 2.00E-105 | 86% |
| 7458_2A2 | megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 387 | 1492 | 89% | 3.00E-104 | 86% |
| 7458_2A2 | discontiguous megablast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 107 | 107 | 26% | 4.00E-20 | 76% |
| 7458_2A2 | nblast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 107 | 107 | 26% | 4.00E-20 | 76% |
| 7458_2M2 | megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 610 | 1299 | 96% | 2.00E-171 | 94% |
| 7458_2M2 | megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 555 | 1930 | 97% | 9.00E-155 | 87% |
| 7458_2M2 | discontiguous megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 645 | 1347 | 96% | 0 | 94% |
| 7458_2M2 | discontiguous megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 627 | 3458 | 97% | 2.00E-176 | 86% |
| 7458_2M2 | discontiguous megablast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 246 | 246 | 46% | 7.00E-62 | 80% |
| 7458_2M2 | nblast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 645 | 1347 | 96% | 0 | 94% |
| 7458_2M2 | nblast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 627 | 3458 | 97% | 2.00E-176 | 86% |
| 7458_2M2 | nblast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 246 | 246 | 46% | 7.00E-62 | 80% |
| 7458_2M2 | nblast | DQ465448.1 | Arabidopsis thaliana ecotype Tamm_46 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 10% | 2.00E-04 | 83% |
| 7458_2M2 | nblast | DQ465446.1 | Arabidopsis thaliana ecotype Pu2_8 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 10% | 2.00E-04 | 83% |

*Additional file 1. (Continued) All the BLAST results for the NBS bands.*

| NBS band | BLAST algorithm* | Accession | Description | Max score | Total score | Query coverage | E value | Max ident |
|---|---|---|---|---|---|---|---|---|
| 7458_2M2 | nblast | DQ465444.1 | Arabidopsis thaliana ecotype Nfe_17 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 10% | 2.00E-04 | 83% |
| 7458_2M2 | nblast | DQ465442.1 | Arabidopsis thaliana ecotype Lov_2 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 10% | 2.00E-04 | 83% |
| 7458_2M2 | nblast | AY487228.1 | Arabidopsis thaliana isolate Poo_1 disease resistance protein RPP13 variant gene, complete cds | 55.4 | 55.4 | 10% | 2.00E-04 | 83% |
| 7458_2M2 | nblast | AY487227.1 | Arabidopsis thaliana isolate Pet_1 disease resistance protein RPP13 variant gene, complete cds | 55.4 | 55.4 | 10% | 2.00E-04 | 83% |
| 7458_2M2 | nblast | AY487226.1 | Arabidopsis thaliana isolate Asp_1 disease resistance protein RPP13 variant gene, complete cds | 55.4 | 55.4 | 10% | 2.00E-04 | 83% |
| 7458_2M3 | megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mboi-40B16 on chromosome 4, complete sequence | 449 | 972 | 86% | 4.00E-123 | 94% |
| 7458_2M3 | megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 418 | 1778 | 86% | 1.00E-113 | 87% |
| 7458_2M3 | megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 73.1 | 73.1 | 9% | 9.00E-10 | 100% |
| 7458_2M3 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 73.1 | 73.1 | 9% | 9.00E-10 | 100% |
| 7458_2M3 | megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 73.1 | 73.1 | 9% | 9.00E-10 | 100% |
| 7458_2M3 | discontiguous megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mboi-40B16 on chromosome 4, complete sequence | 477 | 1011 | 86% | 2.00E-131 | 94% |
| 7458_2M3 | discontiguous megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 452 | 2679 | 89% | 7.00E-124 | 87% |
| 7458_2M3 | discontiguous megablast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 109 | 109 | 26% | 1.00E-20 | 76% |
| 7458_2M3 | discontiguous megablast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 71.6 | 71.6 | 9% | 3.00E-09 | 100% |
| 7458_2M3 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 71.6 | 71.6 | 9% | 3.00E-09 | 100% |
| 7458_2M3 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 71.6 | 71.6 | 9% | 3.00E-09 | 100% |
| 7458_2M3 | nblast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mboi-40B16 on chromosome 4, complete sequence | 477 | 1011 | 86% | 2.00E-131 | 94% |
| 7458_2M3 | nblast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 452 | 2679 | 89% | 7.00E-124 | 87% |
| 7458_2M3 | nblast | AK247883.1 | Solanum lycopersicum cDNA, clone: LEFL2039C13, HTC in fruit | 109 | 109 | 26% | 1.00E-20 | 76% |
| 7458_2M3 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 71.6 | 71.6 | 9% | 3.00E-09 | 100% |
| 7458_2M3 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 71.6 | 71.6 | 9% | 3.00E-09 | 100% |
| 7458_2M3 | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 71.6 | 71.6 | 9% | 3.00E-09 | 100% |
| 7458_2M3 | nblast | DQ465446.1 | Arabidopsis thaliana ecotype Tamm_46 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 12% | 2.00E-04 | 83% |
| 7458_2M3 | nblast | DQ465446.1 | Arabidopsis thaliana ecotype Pu2_8 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 12% | 2.00E-04 | 83% |
| 7458_2M3 | nblast | DQ465444.1 | Arabidopsis thaliana ecotype Nfe_17 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 12% | 2.00E-04 | 83% |
| 7458_2M3 | nblast | DQ465442.1 | Arabidopsis thaliana ecotype Lov_2 disease resistance protein RPP13 variant (RPP13) gene, complete cds | 55.4 | 55.4 | 12% | 2.00E-04 | 83% |
| 7458_2M4 | megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 303 | 666 | 81% | 2.00E-79 | 94% |
| 7458_2M4 | megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 69.4 | 69.4 | 16% | 6.00E-09 | 97% |
| 7458_2M4 | discontiguous megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 300 | 1181 | 83% | 2.00E-78 | 94% |
| 7458_2M4 | discontiguous megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mboi-40B16 on chromosome 4, complete sequence | 156 | 312 | 57% | 4.00E-35 | 85% |
| 7458_2M4 | discontiguous megablast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 68 | 68 | 16% | 2.00E-08 | 97% |
| 7458_2M4 | discontiguous megablast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 68 | 68 | 16% | 2.00E-08 | 97% |
| 7458_2M4 | nblast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 300 | 1238 | 83% | 2.00E-78 | 94% |
| 7458_2M4 | nblast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mboi-40B16 on chromosome 4, complete sequence | 156 | 312 | 57% | 4.00E-35 | 85% |
| 7458_2M4 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 68 | 68 | 16% | 2.00E-08 | 97% |
| 7458_2M4 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 68 | 68 | 16% | 2.00E-08 | 97% |
| 7458_2M4 | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 68 | 68 | 16% | 2.00E-08 | 97% |
| 7458_2M4 | nblast | AC200591.4 | Rhesus Macaque BAC CH250-310E4 () complete sequence | 41 | 41 | 11% | 2.2 | 92% |
| 7458_2M4 | nblast | AC092043.3 | Homo sapiens chromosome 3 clone RP11-141M3, complete sequence | 41 | 41 | 21% | 2.2 | 76% |
| 7458_2M4 | nblast | BX957255.9 | Zebrafish DNA sequence from clone CH211-87E17 in linkage group 21, complete sequence | 41 | 41 | 16% | 2.2 | 82% |
| 7458_2M4 | nblast | AL954390.19 | sapiens transmembrane protein SHREW1 and a CpG island, complete sequence | 41 | 41 | 13% | 2.2 | 88% |

*Additional file 1. (Continued) All the BLAST results for the NBS bands.*

| NBS band | BLAST algorithm* | Accession | Description | Max score | Total score | Query coverage | E value | Max ident |
|---|---|---|---|---|---|---|---|---|
| 7458_2M4 | nblast | AJ494987.1 | Liporrhopalum tentacularis microsatellite DNA, clone WLT4-19 | 41 | 41 | 10% | 2.2 | 92% |
| 7458_2M4 | megablast | No significant similarity found | | | | | | |
| 7458_2M4 | discontiguous megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 163 | 309 | 66% | 2.00E-37 | 86% |
| 7458_2M4 | discontiguous megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 149 | 837 | 66% | 5.00E-33 | 84 |
| 7458_2M4 | nblast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 163 | 309 | 66% | 2.00E-37 | 86% |
| 7458_2M4 | nblast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 149 | 837 | 66% | 5.00E-33 | 84% |
| 7458_2M4 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 64.4 | 64.4 | 19% | 2.00E-07 | 94% |
| 7458_2M4 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 64.4 | 108 | 34% | 2.00E-07 | 94% |
| 7458_2M4 | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 64.4 | 105 | 34% | 2.00E-07 | 94% |
| 7458_2M4 | nblast | AC144343.8 | Medicago truncatula clone mth2-12c21, complete sequence | 42.8 | 42.8 | 15% | 0.54 | 90% |
| 7458_2M4 | nblast | AM431500.2 | Vitis vinifera contig VV78X244274.3, whole genome shotgun sequence | 41 | 41 | 14% | 1.9 | 90% |
| 7458_2M4 | nblast | CP000609.1 | Methanococcus maripaludis C5, complete genome | 41 | 41 | 19% | 1.9 | 82% |
| 7458_2M4 | nblast | AM481091.1 | Vitis vinifera, whole genome shotgun sequence, contig VV78X136705.3, clone ENTAV 115 | 41 | 41 | 13% | 1.9 | 92% |
| 7458_2M4 | nblast | AC090961.2 | Bos taurus clone RP42-552E5, complete sequence | 41 | 41 | 19% | 1.9 | |
| 7458_2M5 | megablast | No significant similarity found | | | | | | |
| 7458_2M5 | discontiguous megablast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 163 | 309 | 66% | 2.00E-37 | 86% |
| 7458_2M5 | discontiguous megablast | CU856252.3 | S.lycopersicum DNA sequence from clone SL_Mbol-127L14 on chromosome 4, complete sequence | 149 | 844 | 66% | 5.00E-33 | 84% |
| 7458_2M5 | discontiguous megablast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 149 | 837 | 66% | 5.00E-33 | 84% |
| 7458_2M5 | nblast | CU326358.5 | S.lycopersicum DNA sequence from clone SL_Mbol-40B16 on chromosome 4, complete sequence | 163 | 309 | 66% | 2.00E-37 | 86% |
| 7458_2M5 | nblast | CU856252.3 | S.lycopersicum DNA sequence from clone SL_Mbol-127L14 on chromosome 4, complete sequence | 149 | 844 | 66% | 5.00E-33 | 84% |
| 7458_2M5 | nblast | AF411807.1 | Lycopersicon esculentum BAC clone Clemson_Id 127E11, complete sequence | 149 | 837 | 66% | 5.00E-33 | 84% |
| 7458_2M5 | nblast | DQ978779.1 | Triticum turgidum NBS-related sequence NBS2 350, genomic sequence | 64.4 | 64.4 | 19% | 2.00E-07 | 94% |
| 7458_2M5 | nblast | DQ978778.1 | Triticum turgidum NBS-related sequence NBS3 290, genomic sequence | 64.4 | 108 | 34% | 2.00E-07 | 94% |
| 7458_2M5 | nblast | DQ978777.1 | Triticum turgidum NBS-related sequence NBS2 290, genomic sequence | 64.4 | 105 | 34% | 2.00E-07 | 94% |

# CHAPTER 7

## General Discussion

Mirjam M.J. Jacobs

**Introduction**

Since the discovery of the wild potato species in the late 19th century, many taxonomists have tried to understand the relationships between them and create order in this complicated part of the genus *Solanum*. At first, most of these studies relied on morphological characteristics but, later in the twentieth century experimental methods like cytogenetics and hybridization experiments were used on a limited scale, followed by molecular methods like cpRFLP, AFLP and RAPD in the past 20 years. So, what more can one add to the pile of taxonomic studies on the wild species of section *Petota*?

**Extensive Sampling**

The present study has an unprecedented sampling and number of taxa included. All previous (molecular) studies on the taxonomy of wild potato species treated only a small part of the variation present. For example, Miller and Spooner (1999) investigated the species boundaries in the wild potato *Solanum brevicaule* complex, comprising about 30 species names. Another study using cpDNA restriction site data focused on the relationships of *S. bulbocastanum*, *S. cardiophyllum* and closely related species (Rodriguez & Spooner, 1997). In some studies, the entire width of the section was covered, but the sampling was very restricted. In 1998, the first AFLP study on potatoes was carried out on 19 taxa of section *Petota* (Kardolus, 1998b). Bonierbale used nRFLPs to study 12 wild and 4 cultivated members of section *Petota* (Bonierbale *et al.*, 1990). The most complete effort to unravel the taxonomy of wild potatoes has been undertaken by Spooner and his collaborators (Castillo & Spooner, 1997; Spooner & Sytsma, 1992). These three studies combined, covered 86 species from most of the series of *Solanum* section *Petota*, but this still is only less than half of the total number of species. Undersampling can be a problem when analyzing taxonomic data (Chase *et al.*, 2005; Hillis *et al.*, 2003; Poe, 1998; Pollock *et al.*, 2002). In the present project, we attempt to cover all the available variation in the section *Petota* by including whenever possible at least 5 accessions per species. This resulted in a dataset with 4929 individual plants from 951 accessions representing 196 different taxa. As far as we know this thesis provides results on the largest collection of *Solanum* section *Petota* accessions ever analyzed simultaneously.

**Remaining problems in potato taxonomy and evolution**

Despite the previously mentioned studies, there are still important issues in potato taxonomy that remain to be solved. Chapter 2 of this thesis discusses these issues in detail. One of the main problems is that many described species are extremely similar to each other. In certain groups, there is a lack of distinctive characters and species boundaries are difficult to trace. The underlying causes for these difficulties are gene flow caused by hybridization between species, hybrid speciation, and phenotypic plasticity in different environments (Spooner & Hijmans, 2001). Although many of the molecular and morphological studies in the last decades have helped to reduce the number of accepted names, section *Petota* still seems to be somewhat over-classified (Spooner & Salas, 2006). Many species are supported largely by a range of overlapping character states (polythetic support), as was observed for example in studies of series *Demissa* (Spooner *et al.*, 1995) and series *Longipedicellata* (van den Berg *et al.*, 2002) In many cases, potato species can only be distinguished by means of multivariate analysis of quantitative characters and/or on the basis of geographic origin (Giannattasio & Spooner, 1994; Kardolus, 1998a; van den Berg & Groendijk-Wilders, 1999; van den Berg *et al.*, 1998).

Apart from the problematic species distinction, the other main problem is the higher level taxonomic structure within section *Petota*, notably the series classification. Correll (1962) distinguished 25 series while Hawkes (1990) recognized 19 tuber bearing series plus two non-tuber bearing series. These series vary considerably in the number of species included. These series classifications were based on morphological and crossing data and an "intuitive" interpretation of those data. (Jansky *et al.*, 2008). The boundaries between some series seem unclear. The series classification of Hawkes and previous authors has received only partial cladistic support in any molecular study to date (Spooner *et al.*, 2004). Spooner and Castillo (1997) hypothesized that the section *Petota* consists of 4 clades only.

**Molecular markers for taxonomic study**

One of the aims of the present study was to elucidate the taxonomic relationships between the wild *Solanum* section *Petota* species. In three different chapters both the series classification and the species boundaries of the most complicated part of the section *Petota* were investigated. For this purpose two molecular marker methods were chosen: AFLP and cpDNA. Both markers have different characteristics, can be used for different levels of genetic variation and have their own strengths and limitations. They seem to be complementary based on the results of previous taxonomic studies (Després *et al.*, 2003; Pelser *et al.*, 2003; Small *et al.*, 2004).

Chloroplast DNA (cpDNA) sequences have been used to solve taxonomic problems at different taxonomic levels (Olmstead & Palmer, 1994). Coding regions (like *rbcL*) were used for revealing family level taxonomy and non-coding regions (like *matK*) for lower taxonomic levels. Mutation rates in cpDNA are low, which makes cpDNA valuable for inferring relationships at the interspecies level and above (Palmer, 1987). Mutation rates in non-coding chloroplast sequences are higher than in coding cpDNA regions (Gielly & Taberlet, 1994). A serious limitation of cpDNA is that the chloroplast genome for most plant species is maternally derived. Data resulting from an analysis on the chloroplast genome will therefore only show the evolutionary history of the maternal line. For section *Petota*, no previous systematic study has used cpDNA sequences, but several of them used cpDNA restriction fragments length polymorphisms (RFLPs) (Spooner & Castillo, 1997; Spooner *et al.*, 1991a; Spooner & van den Berg, 1992b).

The application of AFLP has many advantages. It produces highly reproducible data, does not need a priori sequence information and has the ability of high resolution (Jones *et al.*, 1997; Meudt & Clarke, 2007; Wolfe & Liston, 1998 ). AFLP generates fragments at random over the whole genome thus avoiding the risk of generating a gene tree instead of a species tree (Després *et al.*, 2003). It has proven to be a useful method to solve phylogenetic relationships especially at a low taxonomic level (Koopman, 2005; Meudt & Clarke, 2007; Pelser *et al.*, 2003). In potato taxonomy AFLP already has already proven its value. Kardolus et al. (1998a) were the first to apply AFLP in potato taxonomy. In their study they used 53 potato species and showed the efficiency of AFLP by producing no less than 997 markers with only three primer combinations. He concluded from his AFLP results that in *Solanum* section *Petota* the AFLP technique can be applied at or below species level.

The AFLP method has since then successfully been used in more studies on potato taxonomy (Lara-Cabrera & Spooner, 2004; McGregor *et al.*, 2002; Spooner *et al.*, 2005; van den Berg *et al.*, 2002). Despite all the mentioned advantages, AFLP is not undisputed. One of the arguments against the use of AFLP is the possible bias caused by homoplasy (Koopman & Gort, 2004; Meudt & Clarke, 2007). Non-identical but co-migrating bands in the AFLP fingerprints can contribute noise instead of signal to the dataset. Homoplasy becomes a problem mainly when distantly related species are involved. Koopman (2005) showed that in a set of closely related *Lactuca* species sufficient phylogenetic signal was present and concluded that in practice the influence of possible limitations of AFLP, such as co-migration of non-homologous fragments, appears to be limited.

**Backbone approach**

Our original plan was to construct a backbone phylogeny using the cpPDNA sequences from one individual per accession. The detailed phylogeny of the branches would then be resolved by using AFLP data. By doing so, the risk of introducing homoplasy when scoring the AFLP data would be reduced and the scoring would also become easier. Based on the outcome of a pilot study on a subset of 210 genotypes, the definitive scoring strategy for the AFLP reactions would be chosen. This original plan had to be departed for two main reasons. The cpDNA data (sequence data from non coding regions *trn*T- *trn*L-*trn*F and *trn*H-*psb*A) showed surprisingly low variation. Hence the resulting phylogeny also had a low resolution and only a few well-supported large groups could be distinguished: Mexican diploid species, Mexican polyploid species, and a group representing the South American species (Chapter 3). The AFLP data showed sufficient variation, but the results showed several incongruencies with the cpDNA results. Such incongruencies are interesting because they may reveal information on the specific evolutionary history like the occurrence of hybridization events between species or the formation of hybrid species (Vriesendorp & Bakker, 2005).

**Evaluation of the series classification**

To evaluate the series classifications of Hawkes (1990) and the 4 clade hypothesis as proposed by Spooner (1997), a large AFLP dataset of 4929 individual samples was analyzed (Chapter 4). The combined cpDNA/AFLP analysis of the small dataset in Chapter 3 and the AFLP analysis of the large dataset in Chapter 4 show that the taxonomic structure of *Solanum* section *Petota* is highly unbalanced. Some subgroups of the section *Petota* have high support and their inner structure also displays well supported subdivisions. However, a large number of the species cannot be further classified in groups and seem to be equally related to each other and to the supported groups. Roughly, there seem to be three levels of support according to our AFLP results. A number of groups is always well supported, whether the analysis is done in a phenetic or phylogenetic way. This is valid for the group of Mexican diploid species, the group of Mexican tetraploid species, the group of *S. demissum* and *S. acaule* and closely related species, the group of *S. circaeifolium* and *S. capsicibaccatum*, the group of *S. commersonii* and the group of *S. schenckii* and *S. hougasii*.

Another category is formed by groups of species that can be distinguished in the original MP and NJ trees but that display no statistical jackknife support. This applies to the Mexican hexaploid species, the group containing polyploid species belonging to series *Conicibaccata*, the group containing diploid species from series *Piurana*, and the small groups of *S. huancabambense*, *S. kurtzianum*, *S. medians*, *S. mochiquense*, *S. hannemanii*, *S. buesii*, and *S. paucijugum*. The largest part of the phylogenetic and phenetic trees consists of a polytomy and thus seem to contain no structure at all. Although it is possible to identify additional groups (for instance, the group of cultivated potatoes together with species of series *Tuberosa* from Peru) in many of the original NJ and MP trees, the taxonomic signal is not strong enough to show statistical support in the form of high jackknife values.

The above results provide only partial support for the series classification of Hawkes (Hawkes, 1990). Especially, the distinct status of many small series like *Maglia*, *Cuneoalata* and *Lignicaulia* could not be supported, but also the support for the larger South American series like *Megistacroloba* and *Yungasensa was* lacking. Our results also show some discrepancies with the 4 clade hypothesis suggested by Spooner and co-authors based on cpDNA RFLPs. Our AFLP results showed more groups than the four main clades found with cpDNA restriction data, and the groups were not completely analogous. In both our and Spooner's results a clade with the Mexican and Central American diploid species was recognized, but according to the cpDNA RFLP results the diploid species *S. bulbocastanum*, and *S. cardiophyllum* appear together in a separate clade. Our results show a *Piurana* clade but the species composition is different from that of Spooner's third clade which consist of the South American diploid species of series *Piurana*, but also members of series *Conicibaccata*, *Megistacroloba*, *Tuberosa*, and *Yungasensa*. Furthermore, we found different clades and groups for members of series *Acaulia* (including *S. demissum*), series *Conicibaccata*, *Circaeifolia*, and series *Longipedicellata*, while the cpDNA RFLP results combined all these series in one clade together with *S. verrucosum* and members of *Commersoniana, Cuneoalata, Lignicaulia, Maglia, Megistacroloba, Tuberosa,* and *Yungasensa*.

**Informal species groups**

Because the scientific support for the series classification of Hawkes is missing, an alternative is needed to subdivide the section *Petota*. The structure found in the AFLP study in Chapter 4 was used to design a new classification. We propose a classification in informal species groups. This approach is similar to the approach of Spooner et al. (2004) who followed the approach of designating informal species classifications of Whalen (1984) and Knapp (2000). They constructed 11 informal species groups for the North and Central American *Solanum* species. However, many species cannot be accommodated in groups and are intentionally left unclassified. For this reason, an exhaustive and closed classification (Knox, 1998) as requested by the rules of the International Code for Botanical Nomenclature (McNeill *et al.*, 2006) is difficult to apply. Our informal group classification is based on the groups supported in the NJ jackknife tree, because of the higher level of resolution shown in this tree. Because most of our informal classification matches the informal species group classification of Spooner et al.(2004) the names of their informal species groups will be maintained if applicable, and new species groups that were not treated in their study (which was limited to the species of Mexico and Central America) will be added.

We designated 10 informal species groups: Diploid Mexican group, Acaulia group, Iopetala group, Longipedicellata group, Polyploid Conicibaccata group, Diploid Conicibaccata group, Diploid Piurana group, Tetraploid Piurana group, Circaeifolia group and Verrucosa group.

**Hybridization within section *Petota***

A majority of all higher plant species may be derived from past hybridization events and hybridization is considered to be an important phenomenon in angiosperm evolution. Additionally, there is a growing interest in the reconstruction of reticulate patterns, aiming to investigate the origin of putative hybrid species (Vriesendorp & Bakker, 2005). Hybridization between species and hybrid speciation are often mentioned as two of the underlying causes for the complications in the systematics of section *Petota*. (Spooner & Hijmans, 2001; Spooner & Salas, 2006; Spooner *et al.*, 2004; van den Berg & Jacobs, 2007).

According to Hawkes (1990) hybridization and polyploidization would have played an important role in the origin of the group of species known under the series name *Longipedicellata*. The tetraploid members of series *Longipedicellata* would have been the product of a common tetraploid genome species from South America possibly *S. chacoense* with a native diploid primitive Mexican and Central American species (Hawkes, 1990). Also for the hexaploid species belonging to series *Demissa* a similar hybridization origin is suggested. Furthermore, various individual species are suspected to be hybrids of natural crosses between wild potatoes. Spooner and van den Berg (1992a) list 27 taxa that are considered to be hybrids by one or more authors. The hypotheses of the hybridization are based on intermediate morphology, plus data on ploidy levels, distributional data, artificial reconstruction of the hybrids, comparison with putative natural hybrids and reduction in fertility (Spooner *et al.*, 2004). More recent (molecular) studies have discounted the existence of several acclaimed hybrids such as *S. chacoense* (Miller & Spooner, 1996) and *S. raphanifolium* (Spooner *et al.*, 1991b). The claims of hybrid species should be considered as hypotheses and should be treated with caution.

However, very recently, evidence was found for the allopolyploid and hybrid origin of members of series *Longipedicellata* and possible hybrid origin for members of series *Conicibaccata* and the Iopetala group (Spooner *et al.*, 2008). Another clue for the hybrid origin of the series *Longipedicellata* comes from research on R genes conferring resistance against *P. infestans*. Some accessions of the polyploid Central American species *S. stoloniferum*, *S. polytrichon* (synonym of *S. stoloniferum*) and *S. papita* (synonym of *S. stoloniferum*) contain sequentially and positionally conserved *Rpi-blb1* homologues (named after the species *S. bulbocastanum* were this gene initially was discovered) (Wang *et al.*, 2008).

In the studies presented in this thesis, several results indicate the possible presence of hybrid species in section *Petota*. Both the lack of support for the inter-group relationships in the cpDNA and AFLP tree (Chapter 3 and Chapter 4) and the lack of structure found within the South American part of the AFLP tree (Chapter 4) could point at the influence of hybridization and introgression, which would have a homogenizing effect on the relationships between the species (through exchange of genetic material with different closely related taxa) and possibly also the higher taxa.

More direct evidence for the existence of hybridization within section *Petota* was found by comparing AFLP data with plastid cpDNA sequence data, as described in Chapter 3, which revealed many incongruencies between the two datasets. For example, the composition of species of the clade representing series *Piurana* in the cpDNA tree is different from that of the clade representing the *Piurana* series in the AFLP tree. This is also valid for the species composition in the clade representing series *Conicibaccata*. The boundaries of both series *Piurana* and *Conicibaccata* seem to be blurred and unclear. This discrepancy might be explained by assuming that gene flow between species of these series still occurs. The series *Conicibaccata* might also be closely related to series *Longipedicellata* according to the results in Chapter 3.

Furthermore, the accessions of *S. demissum* and *S. acaule* and their closest relatives form one clade in the AFLP tree while they are far apart in the cpDNA tree. In the cpDNA tree, *S. demissum* is placed amidst the *Demissa / Longipedicellata* clade. Previous studies have already suggested that *S. demissum* could be derived from *S. acaule* and an unknown female parent (Kardolus, 1998a; Nakagawa & Hosaka, 2002; Spooner *et al.*, 1995). Based on the present results it would be logical to assume that an unknown species from series *Longipedicellata* or series *Demissa* has acted as a maternal parent. Although the incongruencies between the two dataset were clearly visible, the outcome of the statistical tests was not very consistent. The outcome of the Mantel test confirmed the differences between the dataset, while the outcome of the ILD test pointed in the direction of non-significant incongruencies. Perhaps these differences between the outcomes of the statistical tests are caused by a difference in sensitivity between the tests for difference in resolution between the datasets. Our results show that within section *Petota* hybridization has played an important role in the origin of certain taxa. Further research with methods like FISH or GISH is needed to shed more light on the origin of certain (polyploid) species.

**Species radiation in the Andes**

As put forward in the previous paragraph on hybridization, a lack of support was observed for the phylogenetic relationships between the different species groups found in the NJ and MP trees, and for any systematic structure in the South American part of section *Petota*. Contrastingly, other parts of the (NJ or MP) trees show several well-supported groups with some subdivisions (Chapter 3 and 4). Since the presence of well-supported groups clearly shows that phylogenetic signal is present in our data, the lack of structure in other parts of the trees must have a different source. Previous taxonomic studies on wild potatoes using AFLP (Kardolus, 1998a; Kardolus, 1998b) and cpDNA RFLP (Castillo & Spooner, 1997; Spooner & Sytsma, 1992) and ETS rDNA (Volkov *et al.*, 2003) also revealed difficulties in finding resolution in the group of South American species. From these results and our own data it would be logical to search for a biological explanation for this lack of structure. Similar patterns of poor resolution have been reported in studies on other plant taxa (Crisp *et al.*, 2004; Fishbein *et al.*, 2001; Hughes & Eastwood, 2006; McKinnon *et al.*, 2008; Schmidt-Lebuhn, 2007; Walker *et al.*, 2004). Various explanations for the observed patterns are suggested in these studies like short internal branch lengths, hybridization, and the influence of ecological factors.

A low sequence divergence and lack of resolution was observed in the large Andean clade of the genus *Lupinus* (Hughes & Eastwood, 2006). This might point at a rapid and recent diversification in the Andes. The authors also suggest that *Lupinus* is probably only one example of many (yet unknown) plant radiations that followed the final uplift of the Andes. It is possible that the factors underlying the *Lupinus* diversification are also responsible for the *Solanum* section *Petota* diversification. These factors would be the large scale of the area over which the radiation extends, a repeated fragmentation of high altitude habitats due to quaternary climate fluctuations, the extremely dissected topography, and the habitat heterogeneity. A relatively fast spread of tuber-bearing *Solanum* species over South America, due to the geographic conditions in the Andes (Hughes & Eastwood, 2006) combined with high levels of hybridization, may explain why the phylogenetic links between species are so difficult to establish.

A more general explanation for the lack of resolution is found in the hypothesis put forward by Rokas and Caroll (2006). This hypothesis assumes the ratio in length between internal and external branches in a tree influences its resolvability. Homoplastic characters can mislead the reconstruction of the short stems (characteristic for radiation) by obscuring the true phylogenetic signal. The phylogeny becomes bush-like when the time since the radiation proceeds and the external branches lengthen (Fishbein *et al.*, 2001) If this theory is valid for the situation in section *Petota*, this would imply that the true relationships might never be solved. All these hypotheses suggest the occurrence of a rapid radiation or within a short stretch of time. We doubt that it would be possible to find more resolution with other methods or more markers, and we consider it likely that the polytomy is indicative of the real situation in section *Petota*.

**Species status and over-classification**

According to many contemporary authors that focused on the taxonomy of the wild potatoes, *Solanum* section *Petota* is over-classified (Spooner & Salas, 2006). Many of the described species in section *Petota* are extremely similar to each other and in many cases potato species can only be distinguished by a combination of often minor and overlapping character states (Spooner & van den Berg, 1992a). Following the increased understanding of potato taxonomy due to the application of molecular techniques, the overall number of species in *Solanum* section *Petota* has already been reduced somewhat. While Hawkes (1990) still recognized 227 tuber-bearing species (7 cultivated species included) and 9 non tuber-bearing species within section *Petota*, Spooner and Hijmans (2001) recognized only 203 tuber-bearing species, including 7 cultivated species, and Spooner and Salas (2006) further reduced the number to 189 species (including 1 cultivated species). In this last review on section *Petota* taxonomy, speculations on necessary taxonomic changes for several species are already made (Spooner & Salas, 2006).

A textbook example of presumed over-classification within *Solanum* section *Petota* is the so-called brevicaule complex. Morphological results from van den Berg et al. (1998) failed to distinguish the 30 species in the brevicaule complex. Molecular results showed that the brevicaule complex is paraphyletic and that many taxa should be relegated to synonymy (Miller & Spooner, 1999. The cpDNA data and AFLP trees in Chapter 3 and Chapter 4 display large polytomies for the part of the South American species. Many species cannot be classified in groups in any meaningful way and most of them seem to be equally related to each other and to the supported groups.

The polytomy found in the results of Chapters 3 and 4 is further scrutinized in Chapter 5. By using methods currently prevalent in population genetics we found evidence that in the South American part of the section *Solanum* many species labels do not correspond to species. This seems to be caused by two different phenomena; mis-classification and over-classification. Mis-classification occurs when accessions bearing the same species labels show up in different genetically defined clusters and are combined with accessions with different species labels. Over-classification is defined as the situation when accessions with different species labels are always combined in a genetic cluster, and show no subdivision amongst them. Of both phenomena, various examples were found in the results of Chapter 5. The consequences for potato taxonomy are that a revision of the species status for many species in section *Petota* seems inevitable.

Although it would be preferable to test the cases for suggested synonymy in the future by performing field or greenhouse experiments to obtain reliable morphological characters, we dare to suggest some revisions for certain taxa. Strong evidence for support was found for the species status of 8 out of 90 species labels. For another 9 species labels plus 6 combinations of 2 different species labels weak evidence was found, because results from Chapter 5 do not completely agree while in some cases only a few accessions were analyzed or some accessions behaved like outliers. For 43 species labels no evidence for species status was found. Finally, for 18 species labels it was impossible to draw conclusions on the species status, because the species label was only represented by one accession. These results indicate that the number of species labels in section *Petota* will probably decrease further after future revisions. Based on our results, we expect that the number of taxonomical units will be closer to 40, an appreciable reduction from the 90 species labels included in chapter 5. Furthermore, the potential cases of misclassified accessions need to be examined in more detail.

**General conclusions on the taxonomy of *Solanum* sect. *Petota***

After the analysis of the AFLP data for 1000 accessions of *Solanum* section *Petota* and the cpDNA sequence data of a representative subset the following conclusions can be drawn. Both phenetic and phylogenetic trees show that taxonomic structure within *Solanum* section *Petota* is highly unbalanced. Several branches in the trees show strong support but a large part of the trees consists of a polytomy, mainly formed by species from South America. Support for the series classification of Hawkes was only found for a restricted number of series, and our results show many incongruencies with the 4 clade hypothesis of Spooner and co-workers.

Therefore, an alternative classification in informal species groups is proposed. The branches with support in the NJ jackknife tree are used to classify these informal species groups: Mexican diploids, Acaulia, Iopetala, Verrucosa, Longipedicellata, polyploid Conicibaccata, diploid Conicibaccata, Circaeifolia, diploid Piurana and tetraploid Piurana. Furthermore, incongruencies between the results based on chloroplast DNA, which is inherited maternally, and nuclear AFLP point at hybridizations between several groups of species within *Solanum* section *Petota*. A detailed analysis of the South American polytomy by using population genetics methodology revealed that a classification in 16 groups based on genetic similarity could explain more variation than the old species classification. Moreover, the distribution of a number of the species labels over the groups pointed at misclassification and/or overclassification. Only for 8 species labels strong and for 15 species weak support for their species status could be found. The other unsupported species labels need to be scrutinized further including with other data like morphology, and geographical distribution. Generally, we believe that the lack of structure is not due to any methodological problem but represents the real biological situation caused by hybridization and rapid radiation within section *Petota*.

**Genebanks, taxonomy and breeding**

The recognition of wild potato species from Central and South America as primary sources for resistance against pests, diseases and abiotic stress, has resulted in numerous collecting expeditions, starting with the Russian pioneer expeditions in the 1920s (Hawkes, 1990) to recent ones in the 1990s (Bradshaw *et al.*, 2006; Spooner & Hijmans, 2001). It is widely accepted that geographic gene centers of cultivated plants and their wild relatives could serve as a main source of natural resistance to diseases, insect pests, and nematodes. Additionally, plants grown in these gene centers have long been exposed to local selective pressure and may have developed resistance to local pathogens and insect pests (Leppik, 1970).

Collecting activities in these areas led to the establishment of a number of germplasm collections worldwide. The wild potato species in these genebanks are important for breeding programs being both sources of genetic diversity (base broadening) as well as sources for genetic resistances to diseases, pests and abiotic stresses (Bradshaw, 2007; Hawkes, 1990; Pavek & Corsini, 2001). To provide an optimal use of the biodiversity available in these genebanks it is important that the identifications of the accessions are correct and that the applied classification of section *Petota* reflects the biological situation in the field. Conflicting taxonomies can confuse breeders (Harlan, 1976; Spooner & van den Berg, 1992a). For them, data on crossability is the most important information (Spooner & van den Berg, 1992a), but a stable taxonomy can provide additional information on the interpretation of the morphological and genetic diversity within crossing groups. Insight in the systematic relationships within the tuber-bearing *Solanum* species might help to identify and select the most interesting materials for breeding purposes (Wang *et al.*, 2008). The passport data of the accessions that were used in this thesis were combined with the taxonomic data from our project and sequence data and late blight data from accompanying projects within the CBSG consortium and were made available to users (scientists and breeders. The description of the complete dataset is going to be published in the near future.

**Taxonomy and predictive value**

Besides serving as a general tool for identification of wild germplasm and interpretation of morphological and genetic diversity, taxonomy is considered to be a valuable predictor for certain traits (Jansky *et al.*, 2006; Jansky *et al.*, 2008). The identification of wild populations or genotypes that possess useful traits involves screening of accessions from genebanks. Usually there are restrictions in time and funding that prevent screening of all the samples. It would therefore be valuable to be able to predict which populations would most likely possess specific traits of interest (Hijmans *et al.*, 2003). Only a few studies have investigated the role that taxonomy can fulfill as a predictor for (useful) traits (Burns, 2006; Hijmans *et al.*, 2003; Jansky *et al.*, 2006; Thaler & Karban, 1997). Hijmans et al. (2003) studied the predictive value of taxonomic, geographic and ecological factors for the presence of frost resistance in 1646 wild potato accessions, representing 87 species. A strong (significant) association of frost tolerance with species and a strong association with Hawkes' series (Hawkes, 1990) were found. Other studies, however, did not find such encouraging results.

Recently, Jansky and co-workers studied the predictivity of taxonomy and other related factors for the presence of resistance to early blight (Jansky *et al.*, 2008) and to white mold in wild potatoes (Jansky *et al.*, 2006). In both studies, no consistent association between the presence of resistance and taxonomic or geographic factors could be found. They concluded therefore that neither taxonomic nor geographic data can be used to predict sources of disease resistance. Although in the present study we have not yet tested the existence of possible associations between potato taxonomy, geographic data and late blight resistance, some observations are worth mentioning. In one of the accompanying CBSG projects, V. Vleeshouwers and co-workers screened more than 900 accessions of wild potatoes for late blight resistance (personal communication). They found that late blight resistance, although not always expressed at the same level, was found in many different accessions, bearing various species labels and belonging to various groups of species from section *Petota*. Even more remarkable was the fact that within accessions, variation in resistance to late blight could be found. These observations lead to the assumption that future tests for taxonomic predictivity for late blight resistance in wild potato species will yield negative results. Predicting presence of resistance to diseases seems far more complicated than predicting resistance against abiotic stress like frost. This is probably caused by the complicated interactions between host species and pathogen, combined with other abiotic and biotic factors that can influence this relationship. Possibly, disease gene evolution may occur faster than plant speciation, disrupting a concordance between resistance and taxonomy (Jansky *et al.*, 2006).

The relationships between abiotic stress factors like frost and plant taxonomy is probably more straightforward. The study of Hijmans et al. (2003) on the prediction of the presence of frost tolerance indicates that it is too early to dismiss taxonomy as a possible predictive variable in general. And last but not least important to mention, conclusions on associations between taxonomy and traits also depend on the accuracy, consistency, and relevance of the taxonomic system used (Hijmans *et al.*, 2003).

**Occurrence of R genes against P. infestans in wild potatoes**

Late blight, caused by *Phytophthora infestans* is one of the most important diseases in potato cultivation (Fry, 2007). Disease management relies heavily on disease control chemicals. The management costs to suppress late blight epidemics are high, surpassing 10% of the total value of the crop. Furthermore, only farmers and breeders in developed, rich countries have access to the full range of these expensive treatments. The environmental effects of the large amount of fungicide are yet unknown (Fry, 2008). *P. infestans* is known to be a rapidly adapting organism and given the wide use of fungicides, the risk of selecting fungicide resistant is realistic. This has already happened with the fungicide metalaxyl which proved successful the first years after its introduction, but lost efficiency after spontaneous selection of resistance. All these issues make the benefits of finding one or more durable resistance genes even more evident.

Over the last century, 11 late blight resistance genes were introduced into cultivated potato from the wild species *S. demissum* (Gebhardt & Valkonen, 2001). As the resistances conferred by these R genes were quickly broken by the pathogen (Wastie, 1991), the presence of R genes in other relatives of the cultivated potato was investigated as well. In the following species late blight R-genes or QTLs have been identified and mapped: *S. microdontum* (chromosome 10), *S. mochiquense* (chromosome 9), *S. paucissectum* (chromosome 10/11/12), *S. spegazinni* (chromosome 4/5), *S. pinnatisectum* (chromosome 7), *S. berthaultii* (chromosome 10) and *S. bulbocastanum* (chromosome 4, 6, 8, 10) and *S. stoloniferum* (chromosome 4) (Bisognin *et al.*, 2005; Ewing *et al.*, 2000; Ghislain *et al.*, 2001; Kuhl *et al.*, 2001; Naess *et al.*, 2000; Oberhagemann *et al.*, 1999; Park *et al.*, 2005; Rauscher *et al.*, 2006; Sandbrink *et al.*, 2000; Sliwka *et al.*, 2006; Smilde *et al.*, 2005; Tan *et al.*, 2008; van der Vossen *et al.*, 2003; Villamon *et al.*, 2005; Wang *et al.*, 2008).

However, besides *S. demissum* and the wild species listed above, there are many other wild species available in genebanks that have not been tested yet for the presence of R genes against *P. infestans*. In Chapter 6 from this thesis, we investigated some of these possible sources for *P. infestans* resistance. We developed and tested a novel approach to identify to which cluster a new resistance gene belongs and to obtain markers that can be used for introgression breeding. Using NBS profiling, we searched for markers that are linked to *P. infestans* resistance in resistant and susceptible genotypes of small segregating populations. To identify the relevant resistance gene cluster. The polymorphic NBS fragments are sequenced to identify the relevant resistance gene cluster and the sequences were analyzed using bio-informatics tools. We found *P. infestans* resistance genes in the accessions of the species *S. verrucosum* (chromosome 6), *S. schenckii* (chromosome 4) and *S. capsicibaccatum* (chromosome 11). We also found indications for locations of R genes in accessions of the species *S. weberbauerii*, *S. ehrenbergii*, *S. circaeifolium* and *S. cardiophyllum*, but these could not be confirmed yet and are therefore not discussed in Chapter 6.

These yet unmapped R genes will be investigated further in future projects within CBSG. The species that expressed the *P. infestans* resistance belonged to various species groups; *S. verrucosum* belongs to the species group of Verrucosa, *S. capsicibaccatum* to diploid species group Circaeifolia and *S. schenckii* to the polyploid lopetala group.

Although the *P. infestans* resistance was found in different wild *Solanum* species previously not used or investigated for *P. infestans* resistance, it seems that the genes causing the resistance are linked to known clusters of resistance genes. This would suggest that the present view on the *Solanum* genome is rather exhaustive and that most resistance clusters are already known. In another study, genes derived from wild *Solanum* species could also be positioned at already known R gene clusters of the *Solanum* genome. Wang et al. (2008) found that the dominant R genes *Rpi-sto1* (derived from *S. stoloniferum*) and *Rpi-plt1* (from *S. polytrichon*) resided at the same position on chromosome VIII as *Rpi-blb1* in *S. bulbocastanum*. The fact that identical R genes are shared by *Solanum* species from completely different species group could point at unsuspected genetic relationships between these taxonomic groups (Wang *et al.*, 2008) One of the challenges of future research in R genes in *Solanum* will be the investigation of the relationships between the R genes in different wild potato species. The results can be used to facilitate the use of R genes for breeding programs. Resistance genes in species that cannot be crossed easily with cultivated potato, may have homologues in more advanced species that are easily crossable with cultivated germplasm (Wang *et al.*, 2008). Furthermore, the distribution of R genes and their relationships could also serve as an extra source of information on the evolutionary history of the wild potatoes.

**Perspectives for resistance breeding**

Durable resistance is often associated with horizontal resistance as opposed to vertical resistance. The R genes that are part of the gene for gene relationship are defined as vertical resistance (Ritter *et al.*, 1990). This implies that the general view is that R genes are inferring non-durable forms of resistance (Ellis *et al.*, 2000; Hulbert *et al.*, 2001). This view might not be completely correct, since resistant genotypes exist in natural populations. However, previous experiences with the resistance genes R1-R11 derived from *S. demissum* have fuelled the discussion on the value of the use of R genes in potato breeding (Fry, 2008).

So why do we still invest in the search for new R genes with the intention to use them eventually in cultivated material? The answer to this question has several aspects. First, there are always exceptions to the rule, thus there is still the hope that in the future new R genes can be discovered that infer more durable resistance (Fry, 2007). Secondly, other strategies are being developed in which the application of non durable R genes could be continued. One type of these strategies is known as diversification strategies (Finckh *et al.*, 2007). These strategies encompass mixed cropping in the form of randomly mixed varieties, alternating rows or strips of varieties or strip intercropping of potato with other crop species. The disease reduction would be effected by 3 main factors: dilution of the inoculum, barrier effects, and, most importantly with respect to combining cultivars with different genetic sources of resistance, induction of host defense mechanisms by avirulent spores. The efficiency of cultivar mixtures depends on the availability of several cultivars with high levels of resistance (Pilet *et al.*, 2006). A third possibility is the combination of various sources of resistance within one and the same cultivar. This strategy is also referred to as pyramiding (Pedersen & Leath, 1988).

Another related method is the creation of multilines (Pink, 2002) which are described as mixtures of components which are agronomically similar but differ in a few key traits like resistance. Genetic engineering could be used to synthesize multilines quickly and efficiently by inserting different resistant alleles into superior agronomic genotypes (McDonald & Linde, 2002). However, it does seem probable that even with the introduction of such multilines or cultivars containing several R genes pathogen populations will continue to evolve and respond to these forms of genetic resistance, especially as many of the known R genes have already been broken by the pathogen. For now, it seems that investing time and money in investigating all these possible solutions would be the best strategy. More generally, I agree with Harlan (1976) who argued that stabilizing strategies that tend in the directions of balanced host-pathogen relationships are much to be desired. The pressure on the world food supply is such that modest yearly losses would be far better than occasional disastrous epidemics.

# References

AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796-815.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25, 3389–3402

Bakker E, Butterbach P, Rouppe van der Voort J, *et al.* (2003) Genetic and physical mapping of homologues of the virus resistance gene *Rx1* and the cyst nematode resistance gene *Gpa2* in potato. Theoretical and Applied Genetics 106, 1524-1531.

Bisognin DA, Douches DS, Buszka L, *et al.* (2005) Mapping late blight resistance in *Solanum microdontum* Bitter. Crop Science, 45, 340-345.

Bonierbale M, Ganal W, Tanksley SD (1990) Applications of restriction fragmenth length polymorphisms and genetic mapping in potato breeding and molecular genetics. In: The molecular and Cellular Biology of the Potato (eds. Vayda ME, Park WD), pp. 13-24. CAB International, Wallingford, UK.

Bornet, B., Goraguer F, Joly G, Branchard M, (2002), Genetic diversity in European and Argentinian cultivated potatoes (*Solanum tuberosum* subsp. tuberosum) detected by inter-simple sequence repeats (ISSRs). Genome, 45,481-484.

Bradshaw J, Bryan G, Ramsay G (2006) Genetic Resources (Including Wild and Cultivated *Solanum* Species) and Progress in their Utilisation in Potato Breeding. Potato Research, 49, 49-65.

Bradshaw JE (2007) Potato-Breeding Strategy. In: Potato Biology and Biotechnology, Advances and perspectives (ed. Vreugdehil D), pp. 157-177. Elsevier, Oxford.

Brickell CD, Baum BR, Hetterscheid WLA, *et al.* (2004) International Code of Nomenclature for Cultivated Plants, Seventh Edition. *Acta Horticulturae* 647, 1-123, i-xxi.

Brown CR, Yang C-P, Mojtahedi H, Santo GS, Masuelli R, (1996) Theoretical and Applied Genetics, 92, 572-576.

Brugmans, B, Wouters D, Van Os H, Hutten R, Van der Linden G, Visser RGF, Van Eck HJ, Van der Vossen EAG (2008) Genetic mapping and transcription analyses of resistance gene loci in potato using NBS profiling. Theoretical and Applied Genetics, DOI 10.1007/s00122-008-0871-7.

Bryan GJ, McNicoll J, Ramsay G, et al. (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theoretical and Applied Genetics, 99, 859-867.

Burns JH (2006) Relatedness and environment affect traits associated with invasive and noninvasive introduced *Commelinaceae*. Ecological applications, 16, 1367-1376.

Calenge F, Van der Linden CG, Van de Weg E, *et al.* (2005) Resistance gene analogues identified through the NBS-profiling method map close to major genes and QTL for disease resistance in apple. Theoretical and Applied Genetics, 110, 660-668.

Castillo RO, Spooner DM (1997) Phylogenetic Relationships of Wild Potatoes, *Solanum* Series *Conicibaccata* (Sect. *Petota*). Systematic Botany, 22, 45-83.

Chase MW, Hanson L, Albert VA, et al. (2005) Life History Evolution and Genome Size in Subtribe *Oncidiinae* (*Orchidaceae*). Annals of Botany, 95, 191-199.

Child A, (1990) A synopsis of *Solanum* subgenus *Potatoe* (G. Don) D'Arcy (Tuberarium (Dun.) Bitter (s.l.)), Feddes Repertorium Specierum Novarum Vegni Vegetablilis, 101, 209-235.

Clausen AM, Spooner DM (1998) Molecular support for the hybrid origin of the wild potato species *Solanum x rechei*, Crop Science, 3, 858-865.

Collins NC, Webb CA, Seah S, *et al.* (1998) The isolation and mapping of disease resistance gene analogs in maize. Molecular Plant-Microbe Interactions, 11, 968-978.

Correll DS (1962) The potato and its wild relatives. In: Contributions from the Texas Research Foundation 4, Texas Research Foundation, Renner, Texas.

Crisp M, Cook L, Steane D (2004) Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? Philosophical transactions of the Royal Society of London, 359, 1551-1571.

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature, 411, 826-833.

De Bary A (1876) Researches into the nature of the potato fungus, *Phytophthora infestans*. Journal of the Royal Agricultural Society of England, 12, 357-391.

De Queiroz K (1998) The general lineage concept of species. In: Endless forms: Species and Speciation (eds. Howard J, Berlocher SH), pp. 57-75. Oxford University Press, New York.

De Queiroz K (2005) Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences of the United States of America, 102, 6600-6607.

Debener, T, Salamini F, Gebhardt C (1991) The use of RFLP's (restriction Fragment Length Polymorphisms) detects germplasm introgressions from wild species into potato (*Solanum tuberosum* ssp. *tuberosum*) breeding lines. Plant Breeding, 106, 173-181.

Debener, T, Salamini F, Gebhard C (1999) Phylogeny of wild and cultivated *Solanum* species based on nuclear restriction fragment polymorphisms (RFLPs), Theoretical and Applied Genetics, 79, 360-368.

Del Rio AH, Bamberg JB (2003) The effect of genebank seed increase on the genetics of recently collected potato (*Solanum*) germplasm. American Journal of Potato Research, 80, 215-218.

Després L, Gielly L, Redoutet B, *et al.* (2003) Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. Molecular Phylogenetics and Evolution, 27, 185-196.

Dodds KS (1962) Classification of cultivated potatoes. In: The potato and its wild relatives (ed.: Correll DS) pp.517-539. Contributions from the Texas Research Foundation, Botanical Studies 4:.

Douches DS, Quiros CF (1988) Additional isozyme loci in tuber-bearing *Solanums*: inheritance and linkage relationships, Journal of Heredity, 79, 377-384.

Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance genes. Current Opinion in Plant Biology, 3, 278-284.

Erwin DC, Ribeiro OK (1996) *Phytophthora Disease Worldwide* APS Press, St. Paul, Minneapolis.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology, 14, 2611-2620.

Ewing EE, Simko I, Smart CD, *et al.* (2000) Genetic mapping from field tests of qualitative and quantitative resistance to *Phytophthora infestans* in a population derived from *Solanum tuberosum* and *Solanum berthaultii*. Molecular Breeding, 6, 25-36.

Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics, 164, 1567-1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes, 7, 574-578.

Farris, JS, Kallersjo M., Kluge AG, Bult C (1995). Testing significance of incongruence. Cladistics, 10, 315-319.

Finckh M, Wolfe M, Lammerts van Bueren E (2007) The Canon of Potato Science: 32. Variety Mixtures and Diversification Strategies. Potato Research, 50, 335-339.

Fishbein M, Hibsch-Jetter C, Soltis DE, *et al.* (2001) Phylogeny of *Saxifragales* (Angiosperms, Eudicots): Analysis of a Rapid, Ancient Radiation. Systematic Biology, 50, 817-847.

Flier WG, Grunwald NJ, Kroon LPNM, *et al.* (2003) The Population Structure of *Phytophthora infestans* from the Toluca Valley of Central Mexico Suggests Genetic Differentiation Between Populations from Cultivated Potato and Wild *Solanum* spp. Phytopathology, 93, 382-390.

Flor HH (1942) Inheritance of pathogenicity in *Melampsora lini*. Phytopathology, 32, 653-669.

Flor HH (1971) The current status of the gene for gene concept. Annual review of Phytopathology, 9, 275-296.

Fry WE (2007) The Canon of Potato Science: 10. Late Blight and Early Blight. Potato research, 50, 243-245.

Fry W (2008) *Phytophthora infestans*: the plant (and R gene) destroyer. Molecular Plant Pathology, 9, 385-402.

Fulton T, Chunwongse J, Tanksley S (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Molecular Biology Reporter, 13, 207-209.

Galindo J, Gallegly ME (1960) The nature and sexuality in *Phytophtora infestans*. Phytopathology, 50, 123-128.

Gallegly ME, Galindo J (1958) Mating types and oospores of *Phytophtora infestans* in potato fields in the United States and Mexico. Phytopathology, 22, 274-277.

Gebhardt C, Ritter E, Barone A, *et al.* (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome. Theoretical and Applied Genetics, 83, 49-57.

Gebhardt C, Valkonen JPT (2001) Organization of genes controlling disease resistance in the potato genome. Annual Review of Phytopathology, 39, 79-102.

Ghislain M, Trognitz B, Herrera MD, *et al.* (2001) Genetic loci associated with field resistance to late blight in offspring of *Solanum phureja* and *S. tuberosum* grown under short-day conditions. Theoretical and Applied Genetics, 103, 433-442.

Giannattasio R, Spooner DM (1994a) A reexamination of species boundaries between *Solanum megistacrolobum* and *S. toralapanum* (*Solanum* sect. *Petota*, series *Megistacroloba*): morphological data. Systematic Botany, 19, 89-105.

Giannattasio RB, Spooner DM (1994b) A reexamination of species boundaries and hypotheses of hybridization concerning *Solanum megistacrolobum* and *S. toralapanum* (*Solanum* sect. *Petota*, series *Megistacroloba*): Molecular data. Systematic Botany, 19, 106-115.

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. Molecular Biology and Evolution 11, 769-777.

Görg, R, Schachtschabel U, Ritter E, Salamini F, Gebhardt C (1992) Discrimination among 136 tetraploid potato varieties by fingerprints using highly polymorphic DNA markers. Crop Science, 32, 815-819.

Grun, P, (1990) The evolution of cultivated Potatoes. Economic Botany, 44 (3 suppl.), 39-55.

Grunwald NJ, Flier WG (2005) The Biology of *Phytophthora infestans* at Its Center of Origin. Annual Review of Phytopathology, 43, 171-190.

Hamilton MB (1999) Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Molecular Ecology, 8, 521-523.

Harlan JR (1976) Diseases as a Factor in Plant Evolution. Annual Review of Phytopathology, 14, 31-51.

Hawkes, JG, Hjerting JP (1989) The potatoes of Bolivia: their breeding value and evolutionary relationships. Oxford University Press, Oxford, UK.

Hawkes, JG (1989) Nomenclatural and taxonomic notes on the infrageneric taxa of the tuber-bearing solanums (*Solanaceae*), Taxon, 38, 489-492.

Hawkes JG (1990) The potato: evolution, biodiversity and genetic resources. Belhaven Press, London (United Kingdom).

Hawkes JG, Jackson MT (1992) Taxonomic and evolutionary implications of the Endosperm Balance Number hypothesis in potatoes. Theoretical and Applied Genetics, 84, 180-185.

Hehl R, Faurie E, Hesselbach J, *et al.* (1999) TMV resistance gene N homologues are linked to *Synchytrium endobioticum* resistance in potato. Theoretical and Applied Genetics, 98, 379-386.

Hetterscheid WLA, Brandenburg WA (1995) Culton versus taxon:conceptual issues in cultivated plant systematics, Taxon, 44, 161-175.

Hijmans RJ, Forbes GA, Walker TS (2000) Estimating the global severity of potato late blight with GIS-linked disease forecast models. Plant Pathology, 49, 697-705.

Hijmans RJ, Spooner DM (2001) Geographic distribution of wild potato species. American Journal of Botany, 88, 2101-2112.

Hijmans RJ, Spooner DM, Salas AR, Guarino L, de la Cruz J (2002) Atlas of wild potatoes. Systematic and ecogeographic studies on crop genepools, 10, IPGRI.

Hijmans R, Jacobs M, Bamberg J, Spooner D (2003) Frost tolerance in wild potato species: Assessing the predictivity of taxonomic, geographic, and ecological factors. Euphytica, 130, 47-59.

Hillis, DM, Bull, JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology, 42, 182-192.

Hillis DM, Pollock DD, McGuire JA, *et al.* (2003) Is Sparse Taxon Sampling a Problem for Phylogenetic Inference? Systematic Biology, 52, 124-126.

Hosaka K, Ogihara Y, Matsubayashi M, Tsunewaki K, (1984) Phylogenetic relationship between the tuberous *Solanum* species as revealed by restriction endonuclease analysis of chloroplast DNA. Japanse Journal of Genetics, 59: 349-369.

Hosaka K,(1986) Who is the mother of the potato?- restriction endonuclease analysis of chloroplast DNA of cultivated potatoes. Theoretical and Applied Genetics, 72, 606-618.

Hosaka K, de Zoeten GA, Hanneman Jr RE, (1988) Cultivated potato chloroplast DNA differs from the wild type by one deletion-evidence and implications. Theoretical and Applied Genetics, 75, 741-745.

Hosaka K, Hanneman Jr RE, (1988a) The origin of the cultivated potato based on chloroplast DNA, Theoretical and Applied Genetics, 76, 172-176.

Hosaka K, Hanneman Jr RE, (1988b) Origin of chloroplast DNA diversity in the Andean potatoes, Theoretical and Applied Genetics, 76, 333-340.

Hosaka K, Spooner DM (1992) RFLP analysis of the wild potato species, *Solanum acaule* Bitter (*Solanum* sect. *Petota*). Theoretical and Applied Genetics, 84, 851-858.

Hosaka K, Mori M, Ogawa K, (1994) Genetic relationships of Japanese potato cultivars assessed by RAPD analysis. American Potato Journal, 71, 535-546.

Hosaka K, (1995) Successive domestication and evolution of the Andean potatoes as revealed by chloroplast DNA restriction endonuclease analysis, Theoretical and Applied Genetics, 90, 356-363.

Hosaka K, (2002) Distribution of the 241 bp deletion of chloroplast DNA in wild potato species, American Journal of Potato Research, 79, 119-123.

Hosaka K, (2003) T-type chloroplast DNA in *Solanum tuberosum* L. ssp. tuberosum was conferred from some populations of *S. tarijense* Hawkes. American Journal of Potato Research, 80, 1-32.

Hosaka K, (2004) Evolutionary pathway of T-type Chloroplast DNA in Potato. American Journal of Potato research, 81, 153-158.

Huaman Z, Spooner DM (2002) Reclassification of landrace populations of cultivated potatoes (*Solanum* sect. *Petota*). American Journal of Botany, 89, 947-965.

Hughes C, Eastwood R (2006) Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* 103, 10334-10339.

Hulbert SH, Webb CA, Smith SM, *et al.* (2001) Resistance gene complexes: Evolution and utilization. Annual Review of Phytopathology, 39, 285-312.

Ingvardsen CR, Schejbel B, Lübberstedt T (2008) Functional Markers in Resistance Breeding. In: Progress in Botany, pp. 61-87.

Jacobs M, van den Berg RG, Hoekstra R, Vosman B (2005) Assessing the phylogeny of *Solanum* sect. *Petota* using cpDNA and AFLP, XVII International Botanical Congress, Vienna.

Jacobs MMJ, van den Berg RG, Vleeshouwers VGAA, et al. (2008) AFLP analysis reveals a lack of phylogenetic structure within *Solanum* section *Petota*. BMC Evolutionary Biology, 8, 1-12.

Jansen R, Geerlings H, Van Oeveren AJ, Van Schaik RC, (2001). A Comment on Codominant Scoring of AFLP Markers. Genetics, 158, 925-926.

Jansky SH, Simon R, Spooner DM (2006) A Test of Taxonomic Predictivity: Resistance to White Mold in Wild Relatives of Cultivated Potato. *Crop Science* 46, 2561-2570.

Jansky SH, Simon R, Spooner DM (2008) A Test of Taxonomic Predictivity: Resistance to Early Blight in Wild Relatives of Cultivated Potato. *Phytopathology* 98, 680-687.

Johnston SA, Nijs TPM, Peloquin SJ, Hanneman RE (1980) The significance of genic balance to endosperm development in interspecific crosses. Theoretical and Applied Genetics, 57, 5-9.

Jones CJ, Edwards KJ, Castaglione S, *et al.* (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular breeding, 3, 381-390.

Judelson HS, Blanco FA (2005) The spores of *Phytophthora*: weapons of the plant destroyer. Nature Reviews Microbiology, 3, 47-58.

Juzepczuk SW, Bukasov SM, (1929) A contribution to the question of the origin of the potato. Proceedings USSR Congres on Genetics and Plant and Animal Breeding, 3, 592-611.

Kardolus JP (1998a) A biosystematic analysis of *Solanum acaule*, PhD thesis, Wageningen University.

Kardolus JP, van Eck HJ, van den Berg RG (1998b) The potential of AFLPs in biosystematics: A first application in *Solanum* taxonomy (*Solanaceae*). Plant systematics and evolution, 210, 87-103.

Keen NT (1990) Gene-For-Gene Complementarity in Plant-Pathogen Interactions. Annual Review of Genetics, 24, 447-463.

Knapp S, (1991) A revision of the *Solanum* sessile species group (section *Geminata* pro parte: *Solanaceae*). *Botatical Journal of the Linnean Society,* 105, 179-210.

Knapp S, (2000) A revision of *Solanum thelopedium* species group (section *Anthoresis* sensu Seithe, pro parte): *Solanaceae*. Bulletin of the Natural History Museum London,30, 13-30.

Knox EB (1998) The use of hierarchies as organizational models in systematics. *Biological Journal of the Linnean Society* 63, 1-49.

Koopman WJM, Gort G, (2004) Significance Tests and Weighted Values for AFLP Similarities, Based on *Arabidopsis* in Silico AFLP Fragment Length Distributions. Genetics, 167, 1915-1928.

Koopman WJM (2005) Phylogenetic signal in AFLP data sets. Systematic biology, 54, 197-217.

Kuhl JC, Hanneman RE, Havey MJ (2001) Characterization and mapping of *Rpi1*, a late-blight resistance locus from diploid (1EBN) Mexican *Solanum pinnatisectum*. Molecular Genetics and Genomics, 265, 977-985.

Lapointe F, Legendre P, (1992) A statistical framework to test the consensus among additive trees (cladograms). Systematic Biology, 41, 158-171.

Lara-Cabrera SI, Spooner DM (2004) Taxonomy of North and Central American diploid wild potato (*Solanum* sect. *Petota*) species: AFLP data. Plant Systematics and Evolution, 248, 129-142.

Leppik EE (1970) Gene Centers of Plants as Sources of Disease Resistance. Annual Review of Phytopathology, 8, 323-344.

Lester RN, (1965) Immunolgical studies on the tuber-bearing Solanums. I. Techniques and South American species. Annals of Botany, 29, 609-624.

Liua Z, Halteman D (2006) Identification and characterization of *RB*-orthologous genes from the late blight resistant wild potato species *Solanum verrucosum*. Physiological and Molecular Plant Pathology, 69, 230-239.

Maddison DR, (1991) The Discovery and Importance of Multiple Islands of Most-Parsimonious Trees. Systematic Zoology, 40, 315-328.

Mantovani P, Van der Linden G, Maccaferri M, Sanguineti MC, Tuberosa R (2006) Nucleotide-binding site (NBS) profiling of genetic diversity in durum wheat. Genome, 49, 1473-1480.

Mayden RL (1997) A hierarchy of species concepts: The denouement in the saga of the species problem. In: Species (eds. Clarge MF, Dawah HA, Wilson MR), pp. 381-424. Chapman and Hall, London.

McDonald B, Linde C (2002) The population genetics of plant pathogens and breeding strategies for durable resistance. Euphytica, 124, 163-180.

McGregor CE, van Treuren E, Hoekstra R, van Hintum ThJL (2002) Analysis of the wild potato germplasm of the series *Acaulia* with AFLPs: implications for *ex situ* conservation. Theoretical and Applied Genetics, 104, 146-156.

McGuire S (2008) *S. lycopersicum* DNA sequence from clone LE_HBa-132O11 on chromosome 4, complete sequence. NCBI.

McKinnon GE, Vaillancourt RE, Steane DA, *et al.* (2008) An AFLP marker approach to lower-level systematics in *Eucalyptus* (*Myrtaceae*). American Journal of Botany, 95, 368-380.

McNeill J, Barrie FR, Burdet HM, *et al.* (2006) International Code of Botanical Nomenclature. ARG Gantner Verlag, Ruggell, Liechtenstein.

Meudt HM, Clarke AC (2007) Almost forgotten or latest practice? AFLP applications, analyses and advances. Trends in plant science, 12, 106-117.

Meyer S, Nagel A, Gebhardt C (2005) PoMaMo--a comprehensive database for potato genome data. *Nucleic Acids Reearch.* 33, D666-670.

Meyers BC, Dickerman AW, Michelmore RW, *et al.* (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant Journal 20, 317-332.

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-Wide Analysis of NBS-LRR-Encoding Genes in Arabidopsis. Plant Cell, 15, 809-834.

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease reistance genes by bulked segregant analysis. A rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences of the United States of America, 88, 9828-9832.

Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Research, 8, 1113-1130.

Miller JT, Spooner DM (1996) Introgression of *Solanum chacoense* (*Solanum* sect. *Petota*): Upland Populations Reexamined. Systematic Botany, 21, 461-475.

Miller JT, Spooner DM (1999) Collapse of species boundaries in the wild potato *Solanum brevicaule* complex (*Solanaceae*, *Solanum* sect. *Petota*): molecular data. Plant Systematics and Evolution, 214, 103-130.

Monosi B, Wisser R, Pennill L, Hulbert S (2004) Full-genome analysis of resistance gene homologues in rice. Theoretical and Applied Genetics, 109, 1434-1447.

Mueller LA, Tanksley SD, Giovannini JJ, *et al.* (2005) The tomato sequencing project, the first cornerstone of the International *Solanaceae* Project (SOL) Genomics 153-158.

Mueller LA, Giovannoni JJ, van Eck HJ, Stack S, Tanksley SD (2008) Tomato Genome Sequencing. In: (Submitted 12-MAR-2008) Plant Breeding and Genetics, Cornell.

Murashigi T, Skoog F (1962) A revised medium for rapid growth and bio-assay with tobacco tissue cultures. Physiology of the Plant, 15, 473–497.

Naess SK, Bradeen JM, Wielgus SM, *et al.* (2000) Resistance to late blight in *Solanum bulbocastanum* is mapped to chromosome 8. Theoretical and Applied Genetics, 101, 697-704.

Nakagawa K, Hosaka K (2002) Species relationships between a wild tetraploid potato species, *Solanum acaule* Bitter, and its related species as revealed by RFLPs of chloroplast and nuclear DNA. American Journal of Potato Research, 79, 85-98.

Nei M, Li WH, (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences (USA), 18, 298-303.

Niederhauser JS, Millis WR (1953) Resistance of *Solanum* species to *Phytophthora infestans* in Mexico. Phytopathology, xx, 456-457.

Nixon KC (1999) The Parsimony Ratchet: A New Method for Rapid Parsimony Analysis. Cladistics, 15, 407-414.

Oberhagemann P, Chatot-Balandras C, Schafer-Pregl R, *et al.* (1999) A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. Molecular Breeding, 5, 399-415.

Ochoa CM (1990) The Potatoes of South America: Bolivia. Cambridge University Press, UK.

Ochoa CM (1999) Las papas de sudamerica: Peru. International Potato Center, Lima, Peru.

Olmstead RG, Palmer JD (1994) Chloroplast DNA Systematics: A Review of Methods and Data Analysis. American Journal of Botany, 81, 1205-1224.

Palmer JD (1987) Chloroplast DNA Evolution and Biosystematic Uses of Chloroplast DNA Variation. *The American Naturalist* 130, Supplement 6, 6-29.

Park TH, Gros J, Sikkema A, *et al.* (2005a) The late blight resistance locus *Rpi-blb3* from *Solanum bulbocastanum* belongs to a major late blight R gene cluster on chromosome 4 of potato. Molecular Plant-Microbe Interactions, 18, 722-729.

Park TH, Vleeshouwers V, Hutten RCB, *et al.* (2005b) High-resolution mapping and analysis of the resistance locus *Rpi-abpt* against *Phytophthora infestans* in potato. Molecular Breeding, 16, 33-43.

Pavek J, Corsini D (2001) Utilization of potato genetic resources in variety development. American Journal of Potato Research, 78, 433-441.

Pedersen WL, Leath S (1988) Pyramiding Major Genes for Resistance to Maintain Residual Effects. Annual Review of Phytopathology, 26, 369-378.

Pelser PB, Gravendeel B, van der Meijden R (2003) Phylogeny reconstruction in the gap between too little and too much divergence: the closest relatives of *Senecio jacobaea* (*Asteraceae*) according to DNA sequences and AFLPs. Molecular phylogenetics and evolution, 29, 613-628.

Pflieger S, Lefebvre V, Caranta C, *et al.* (1999) Disease resistance gene analogs as candidates for QTLs involved in pepper-pathogen interactions. Genome, 42, 1100-1110.

Pilet F, Chacon G, Forbes GA, *et al.* (2006) Protection of Susceptible Potato Cultivars Against Late Blight in Mixtures Increases with Decreasing Disease Pressure. Phytopathology, 96, 777-783.

Pineiro R, Aguilar JF, Munt DD, Feliner GN (2007) Ecology matters: Atlantic-Mediterranean disjunction in the sand-dune shrub Armeria pungens (Plumbaginaceae). Molecular Ecology, 16, 2155-2171.

Pink DAC (2002) Strategies using genes for non-durable disease resistance. Euphytica 124, 227-236.

Poe S (1998) The effect of taxonomic sampling on accuracy of phylogeny estimation: test case of a known phylogeny. Molecular Biology and Evolution, 1086-1090.

Pollock DD, Zwickl DJ, McGuire JA, *et al.* (2002) Increased Taxon Sampling Is Advantageous for Phylogenetic Inference. Systematic Biology, 51, 664-671.

PRAP - parsimony ratchet analyses using PAUP*. [http://www.nees.uni-bonn.de/downloads/PRAP]

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of Population Structure Using Multilocus Genotype Data. Genetics, 155, 945-959.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association Mapping in Structured Populations. The American Journal of Human Genetic, 67, 170-181.

Provan J, Powell W, Dewar H, Bryan G, Machray CC, Waugh R, (1999) An extreme cytoplasmic bottleneck in the modern European cultivated potato (*Solanum tuberosum*) is not reflected in decreased levels of nuclear diversity. Proceedings of the Royal Society of London, 266, 633-639.

Quiros CF, McHale N, (1985) Genetic analysis of isozyme variants in diploid and tetraploid potatoes. Genetics, 111, 131-145.

Raimondi JP, Peralta IE, Masuelli RW, Feingold S, Camadro EL (2005) Examination of the hybrid origin of the wild potato *Solanum ruiz-lealii*. Plant Systematics and Evolution, 253, 33-51.

Raker C, Spooner DM (2002), Chilean tetraploid cultivated potato, *Solanum tuberosum*, is distinct from the Andean populations: microsatellite data. Crop Science, 42,1451-1458.

Rauscher GM, Smart CD, Simko I, *et al.* (2006) Characterization and mapping of *RPi-ber*, a novel potato late blight resistance gene from *Solanum berthaultii*. Theoretical and Applied Genetics, 112, 674-687.

Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmatic gene flow in plants. Evolutionary Trends in Plants, 5, 65-83.

Ritter E, Gebhardt C, Salamini F (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics, 125, 645-654.

Rivera-Peña A, Molina-Galan J (1989) Wild tuber-bearing species of *Solanum* and incidence of *Phytophthora infestans* (Mont.) de Bary on the western slopes of the volcano Nevado de Toluca. 1. *Solanum* species. Potato Research, 32, 181-195.

Rodriguez A, Spooner DM (1997) Chloroplast DNA Analysis of *Solanum bulbocastanum* and *S. cardiophyllum*, and evidence for the distinctiveness of *S. cardiophyllum* subsp. *ehrenbergii* (Sect. *Petota*). Systematic Botany, 22, 31-43.

Rokas A, Carroll SB (2006) Bushes in the Tree of Life. *PLoS Biology* 4, e352.

Rosenberg NA, Burke T, Elo K, et al. (2001) Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes From 20 Chicken Breeds. Genetics, 159, 699-713.

Rosenberg NA, Pritchard JK, Weber JL, et al. (2002) Genetic structure of human populations. Science 298, 2381-2385.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Molecular Ecology Notes, 4, 137-138.

Ruiz de Galarreta J, Carrasco A, Salazar A, *et al.* (1998) Wild *Solanum* species as resistance sources against different pathogens of potato. Potato Research, 41, 57-68.

Sandbrink JM, Colon LT, Wolters PJCC, Stiekema WJ. (2000) Two related genotypes of *Solanum microdontum* carry different segregating alleles for field resistance to *Phytophthora infestans*. Molecular Breeding, 6, 215-225.

Schenk MF, Thienpont C, Koopman WJM, Gilissen LJWJ, Smulders MJM (2008) Phylogenetic relationships in *Betula* (*Betulaceae*) based on AFLP markers. Tree Genetics and Genomics, online may 2008.

Schmidt-Lebuhn AN (2007) Using amplified fragment length polymorphism (AFLP) to unravel species relationships and delimitations in *Minthostachys* (*Labiatae*). Botanical Journal of the Linnean Society, 153, 9-19.

Seah S, Telleen AC, Williamson VM (2007) Introgressed and endogenous *Mi-1* gene clusters in tomato differ by complex rearrangements in flanking sequences and show sequence exchange and diversifying selection among homologues. Theoretical and Applied Genetics, 114, 1289-1302.

Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. Systematic Biology, 56, 896-906.

Shaw J, Lickey EB, Beck JT *et al.* (2005) The tortoise and the hare II: relative utility of 21 noncoding Chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany, 92, 142-166.

Sliwka J, Jakuczun H, Lebecka R, *et al.* (2006) The novel, major locus *Rpi-phu1* for late blight resistance maps to potato chromosome IX and is not correlated with long vegetation period. *Theoretical and Applied Genetics* 113, 685-695.

Small RL, Cronn RC, Wendel JF (2004) L.A.S.JOHNSON REVIEW No.2. Use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany, 17, 145-170.

Smilde WD, Brigneti G, Jagger L, *et al.* (2005) *Solanum mochiquense* chromosome IX carries a novel late blight resistance gene *Rpi-moc1*. Theoretical and Applied Genetics, 110, 252-258.

Song J, Bradeen JM, Naess SK, *et al.* (2003) Gene RB cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proceedings of the National Academy of Sciences, 100, 9128-9133.

Spooner DM, Sytsma KJ, Conti E (1991a) Chloroplast DNA evidence for genome differentiation in wild potatoes (*Solanum* sect. *Petota*: *Solanaceae*). American Journal of Botany, 78, 1354-1366

Spooner DM, Sytsma KJ, Smith JF (1991b) A molecular reexamination of diploid hybrid speciation of *Solanum raphanifolium*. Evolution 45, 757-764.

Spooner DM, Sytsma KJ (1992) Reexamination of Series Relationships of Mexican and Central American Wild Potatoes (*Solanum* sect. *Petota*): Evidence from Chloroplast DNA Restriction Site Variation. Systematic Botany 17, 432-448.

Spooner DM, van den Berg RG (1992a) An analysis of recent taxonomic concepts in wild potatoes (*Solanum* sect. *Petota*). Genetic Resources and Crop Evolution 39, 23-37.

Spooner DM, van den Berg RG (1992b) Species limits and hypotheses of hybridization of *Solanum berthaultii* Hawkes and *S. tarijense* Hawkes: morphological data. Taxon 41, 685-700

Spooner DM, Anderson GJ, Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (*Solanaceae*). American Journal of Botany, 80, 676-688.

Spooner DM, van den Berg RG, Bamberg JB (1995) Examination of Species Boundaries of *Solanum* series *Demissa* and Potentially Related Species in series *Acaulia* and series *Tuberosa* (sect. *Petota*). Systematic Botany 20, 295-314.

Spooner DM, Tivang J, Nienhuis J, *et al.* (1996) Comparison of four molecular markers in measuring relationships among the wild potato relatives *Solanum* section *Etuberosum* (subgenus *Potatoe*). Theoretical and Applied Genetics, 92, 532-540.

Spooner DM, Castillo RT (1997) Reexamination of series relationships of South American wild potatoes (*Solanaceae*: *Solanum* sect. *Petota*): evidence from chloroplast DNA restriction site variation. American journal of botany, 84, 671-685.

Spooner DM, Hijmans RJ (2001) Potato systematics and germplasm collecting, 1989-2000. American Journal of Potato Research, 78, 237-278.

Spooner DM, van den Berg RG, Miller JT (2001) Species and series boundaries of *Solanum* series *Longipedicellata* (*Solanaceae*) and phenetically similar species in ser. *Demissa* and ser. *Tuberosa*: implications for a practical taxonomy of sect. *Petota*. American Journal of Botany, 88, 113-130.

Spooner DM, Van den Berg RG, Rodríguez A, Bamberg J, Hijmans, RJ, Lara-Cabrera SI (2004) Wild Potatoes (*Solanum* section *Petota*; *Solanaceae*) of North and Central America. In Systematic Botany Monographs. (Ed. McPherson GD, Prather LA, Ranker TA, Reznicek AA) USA, The American Society of Plant Taxonomists.

Spooner DM, Stephenson SA, Ballard HE, Polgar Z (2004) DNA Sequences of single-copy waxy gene support allopolyploid origins of wild potatoes (*Solanum* section *Petota*), Abstract Plant and Animal Genome Conference.

Spooner DM, Nunez J, Rodriguez F, Naik PS, Ghislain M, (2005b) Nuclear and chloroplast DNA reassessment of the origin of Indian potato varieties and its implications for the origin of the early European potato. Theoretical and Applied Genetics, 110, 1020-1026.

Spooner DM, Mclean K, Ramsay G, Waugh R, Bryan GJ (2005a) A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proceedings of the National Academy of Sciences of the U.S.A., 41, 14649-14699.

Spooner DM, Salas A (2006) Structure, biosystematics, and genetic resources. In: Handbook of potato production, improvement, and post-harvest management. (ed. Gopal J, Khurana SMP), pp. 1-39. Haworth's Press, Inc., Binghampton, New York.

Spooner DM, Rodríguez F, Polgár Z, *et al.* (2008) Genomic Origins of Potato Polyploids: GBSSI Gene Sequencing Data. Crop Science, 48, S27-S36.

Stephenson SA, Spooner DM, Ballard HE, (2004) Nuclear ribosomal internal transcribed spacer region (ITS) DNA sequence phlyogeny of wild potatoes (*Solanum* section *Petota*). Abstract Plant And Animal Genome Conference.

Stewart CNJ, Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques, 14, 748-750.

Sukhotu T, Kamijima O, Hosaka K (2004) Nuclear and chloroplast DNA differentiation in Andean potatoes. Genome, 47, 46-56.

Sukhotu T, Hosaka K (2006) Origin and evolution of Andigena potatoes revealed by chloroplast and nuclear DNA markers. Genome, 49, 636-647.

Swaminathan MS, Howard HW (1953) The cytology and genetics of the potato (*Solanum tuberosum*) and related species. Bibliographica Genetica, 16, 192pp.

Swofford DL, (2002) PAUP*4.0 Beta, Phylogenetic analysis Using Parsimony (* and other methods), Software program, Sinauer associates, Sunderland, Massachusetts.

Tan MYA, Hutten RCB, Celis C, *et al.* (2008) The *RPi-mcd1* Locus from *Solanum microdontum* Involved in Resistance to *Phytophthora infestans*, Causing a Delay in Infection, Maps on Potato Chromosome 4 in a Cluster of NBS-LRR Genes. Molecular Plant-Microbe Interactions, 21, 909-918.

Tanksley SD, Ganal MW, Prince JP, *et al.* (1992) High Density Molecular Linkage Maps of the Tomato and Potato Genomes. Genetics, 132, 1141-1160.

Thaler JS, Karban R (1997) A Phylogenetic Reconstruction of Constitutive and Induced Resistance in *Gossypium*. The American Naturalist, 149, 1139-1146.

Thatcher LF (2005) Plant defence responses: what have we learnt from *Arabidopsis*? Functional plant biology, 32, 1-.

Ugent D, (1966) Hybrid weed complexes in *Solanum*, section *Tuberarium*, PhD dissertation. University of Wisconsin, Madison, USA.

Van den Berg RG, Spooner DM (1992) A reexamination of infraspecific taxa of a wild potato, *Solanum microdontum* (*Solanum* sect. *Petota*: *Solanaceae*). Plant Systematics and Evolution, 182, 239-252.

Van den Berg RG, Miller JT, Ugarte ML, *et al.* (1998) Collapse of morphological species in the wild potato *Solanum brevicaule* complex (*Solanaceae:* sect. Petota). American Journal of Botany 85, 92-109.

Van den Berg RG, Groendijk-Wilders N (1999) Numerical analysis of the taxa of series *Circaeifolia* (*Solanum* sect. *Petota*). In: Proceedings of the International Conference *Solanaceae* IV: July 1994, Adelaide (eds. Nee M, Symon DE, Lester RN, Jessop JP), pp. 213-226. The Royal Botanical gardens, Kew.

Van den Berg RG, Groendijk-Wilders N, Zevenbergen MJ, Spooner DM, (2001a) Molecular systematics of *Solanum* series *Circaeifolia* (*Solanum* section *Petota*) based on AFLP and RAPD markers, In Solanaceae V: Advances in taxonomy and utilization (ed. van den Berg RG, van der Weerden GM, Barendse GWM, Mariani C). pp.73-84 Nijmegen University Press, The Netherlands.

Van den Berg RG, Bryan GJ, del Rio A, Spooner DM (2001b) Reduction of species in the wild potato *Solanum* section *Petota* series *Longipedicellata*: AFLP, RAPD and chloroplast SSR data, Theoretical and Applied Genetics, 105, 1109-114.

Van den Berg R, Bryan G, del Rio A, Spooner DM (2002) Reduction of species in the wild potato *Solanum* section *Petota* series *Longipedicellata*: AFLP, RAPD and chloroplast SSR data. *Theoretical and Applied Genetics* 105, 1109-1114.

Van den Berg RG, Groendijk-Wilders N (2007) AFLP data support the recognition of a new tuber-bearing Solanum species but are uninformative about its taxonomic relationships. Plant Systematics and Evolution, 269, 133-143.

Van den Berg RG, Jacobs MMJ (2007) Molecular Taxonomy. In: Potato Biology and Biotechnology (ed. Vreugdehil D), pp. 55-76. Elsevier Oxford, UK.

Van der Linden CG, Wouters D, Mihalka V, *et al.* (2004) Efficient targeting of plant disease resistance loci using NBS profiling. Theoretical and Applied Genetics, 109, 384-393.

Van der Vossen E, Sikkema A, Hekkert BTL, *et al.* (2003) An ancient R gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *Plant Journal* 36, 867-882.

Van der Vossen EAG, Gros J, Sikkema A, *et al.* (2005) The *Rpi-blb2* gene from *Solanum bulbocastanum* is an Mi-1 gene homolog conferring broad-spectrum late blight resistance in potato. Plant Journal 44, 208-222.

Van Eijk M, Broekhof J, Van der Poel H, et al. (2004) SNPWaveTM : a flexible multiplexed SNP genotyping technology. Nucleic Acids Research, 32, e47.

Van Ooijen G, van den Burg HA, Cornelissen BJC, Takken FLW (2007) Structure and function of resistance proteins in solanaceous plants. Annual Review of Phytopathology 45, 43-72.

Van Os H, Andrzejewski S, Bakker E, *et al.* (2006) Construction of a 10,000-Marker Ultradense Genetic Recombination Map of Potato: Providing a Framework for Accelerated Gene Isolation and a Genomewide Physical Map. Genetics 173, 1075-1087.

Van Soest L, Schöber B, Tazelaar M (1984) Resistance to *Phytophthora infestans* in tuber-bearing species of *Solanum* and its geographical distribution. Potato Research 27, 393-411.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Molecular Ecology 11, 139-151.

Villamon FG, Spooner DM, Orrillo M, *et al.* (2005) Late blight resistance linkages in a novel cross of the wild potato species *Solanum paucissectum* (series *Piurana*). Theoretical and Applied Genetics, 111, 1201-1214.

Vleeshouwers VGAA, van Dooijeweert W, Keizer LCP, Sijpkes L, Govers F, Colon LT (1999) A Laboratory Assay for *Phytophthora infestans* Resistance in Various *Solanum* Species Reflects the Field Situation European Journal of Plant Pathology, 105, 241-250.

Vleeshouwers, VGAA, Rietman H, Krenek P, *et al.* (2008). Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora Infestans* avirulence Genes. PLoS ONE 3(8): e2875. doi:10.1371/journal.pone.0002875.

Volkov RA, Zanke C, Panchuk II, Hemleben V (2001) Molecular evolution of 5S rDNA of Solanum species (sect. Petota): application for molecular phylogeny and breeding, *Theoretical and* Applied Genetics, 103, 1273-1282.

Volkov RA, Komarova NY, Panchuk II, Hemleben V (2003) Molecular evolution of rDNA external transcribed spacer and phylogeny of sect. *Petota* (genus *Solanum*). Molecular Phylogenetics and Evolution, 29, 187-202.

Vos P, Hogers R, Bleeker M, et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, 23, 4407-4414.

*Vriesendorp B, Bakker FT (2005) Reconstructing patterns of reticulate evolution in angiosperms: what* can we do? Taxon, 54, 593-604.

Walker JB, Sytsma KJ, Treutlein J, *et al.* (2004) *Salvia* (*Lamiaceae*) is not monophyletic: implications for the systematics, radiation, and ecological specializations of *Salvia* and tribe *Mentheae*. American Journal of Botany, 91, 1115-1125.

Wang M, Allefs S, van den Berg R, *et al.* (2008a) Allele mining in *Solanum*: conserved homologues of *Rpi-blb1* are identified in *Solanum stoloniferum*. Theoretical and Applied Genetics, 116, 933-943.

Wang M, Van den Berg RG, Van der Linden GC, Vosman B (2008b) The utility of NBS profiling for plant systematics: a first study in tuber-bearing *Solanum* species. Plant Systematics and Evolution, DOI 10.1007/s00606-008-0087-y

Wastie RL (1991) Breeding for resistance. In: *Phytophthora infestans*, the cause of late blight of potato, pp. 193-224. (eds. Ingram DS, Williams PH).

Wendel JF, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and Molecular Evolution. In: Molecular Systematics of Plants, II, pp. 265-296 (ed.: Soltis DE, Soltis PS, Doyle JJ).

Whalen MD (1984) Conspectus of species groups in *Solanum* subgenus *Leptostemonum*. *Gentes Herbarum*, 12, 179-282.

Wolfe AD, Liston A (1998 ) Contributions of PCR-based methods to plant systematics and evolutionary biology. In: Molecular Systematics of Plants II (eds. Soltis DE, Soltis PS, J.J. D), pp. 43-86. Kluwer Academic Publishers, Boston

Zhang JF, Yuan YL, Niu C, *et al.* (2007) AFLP-RGA markers in comparison with RGA and AFLP in cultivated tetraploid cotton. Crop Science 47, 180-187.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology, 8, 907-913.

# Summary

Potato is an economically important crop. It was recognized early that the wild relatives of the cultivated potato could provide crossing material to improve the cultivated material and hence their botany and taxonomy have been the subject of intensive study since the 19th century. However, the taxonomy of the tuber-bearing species is complicated by phenomena like polyploidization, hybridization and morphological plasticity. Furthermore, crossing barriers between certain species are presumably influenced by an unknown mechanism called EBN (Embryo Balance Number) which also adds to the confusion.. Most of the early taxonomic studies relied on morphological observations and, later, on a limited scale, on experimental methods like cytogenetics and hybridization experiments. More than 200 species and many infraspecific taxa in *Solanum* section *Petota* have been described. These taxa have been classified in groups called series, with different authors recognizing a varying number of series, often with different circumscriptions.

One of the most popular authorative treatments on potato taxonomy was given by Hawkes (1990). Although he and many other excellent taxonomists have achieved a great task in describing numerous species, classifying them into series and providing morphological keys, many issues in potato taxonomy remain to be solved. Difficulties such as identification using morphological keys, over-classification of parts of section *Petota* and problems with series classification still exist. In Chapter 1, we first provide a short summary on the history of potato taxonomy and in chapter 2 we give a more detailed review of the molecular studies on potato taxonomy and their goals and achievements. The molecular methods used in potato taxonomy are diverse: cytology data, serology data, isozyme data, restriction site data like RFLP, and AFLP, and primer-based data like RAPD and SSR. The application of molecular methods in potato taxonomy has offered more possibilities to solve complicated issues and improve our understanding of the taxonomy of the potato. However, the taxonomic studies on potato applying molecular methods have one flaw: most of them do not cover the complete width of the variation of the group of wild potato species and in many cases only one or very few accessions were sampled for each taxon which could well have influenced the outcome.

The first goal of this thesis is to elucidate the systematic relationship of wild, tuber bearing *Solanum* species. To cover the width of the variation in *Solanum* section *Petota* and to prevent the risk of undersampling, a very large set of potato accessions was sampled (retrieved mainly from the CGN , supplemented from many other genebanks), resulting in 916 accessions representing more than 190 species. Whenever possible, each species was represented by at least 5 different accessions and each accession by at least 5 genotypes. This resulted in a dataset of approximately 5000 genotypes, the largest ever constructed for *Solanum* section *Petota*. The sampled plants were genotyped using 2 AFLP primer combinations which yielded 222 AFLP markers. In Chapters 3, 4, and 5 we have used this dataset or parts of it to address some of the taxonomic problems mentioned: the higher level taxonomy and the presumed over-classification of section *Petota*.

In chapter 3 two different markers were used that, according to previous studies, should provide different levels of taxonomic resolution. Chloroplast DNA (cpDNA) sequence data and nuclear AFLP data were used for phylogeny reconstruction in *Solanum* section *Petota*. A comprehensive set of accessions (199 accessions from 174 taxa), covering the section as widely as possible, was chosen. The chloroplast regions *trn*TLF and *psbA/trn*H were sequenced. The AFLP data were taken as a subset from the large AFLP dataset. The plant material used for the cpDNA sequences was exactly the same as the plant material analysed with AFLP. Both dataset were analysed separately following phenetic and phylogenetic methods and separate trees reflecting the relationships among the accessions were produced. This approach allowed a direct comparison of the outcome of the cpDNA analysis and AFLP analysis. In Chapter 4 the complete dataset of 4929 genotypes was used for an extensive AFLP analysis. For this large dataset a NJ tree could be produced. For the phylogenetic analyses and estimations of statistical support a condensed dataset of 916 genotypes representing all the available accessions was created. Due to the results in Chapter 3 the original plan, to construct a cpDNA backbone phylogeny and resolving the detailed phylogenetic structure with AFLP results, had to be departed. This backbone strategy would have facilitated the scoring of AFLP bands and prevented possible risk of introducing homoplasy. But the cpDNA data showed far less resolution than the AFLP results, so only a few groups could be classified. More importantly, the cpDNA results showed several main incongruencies with the AFLP results.

The results from the combined cpDNA/AFLP analysis of the subset (Chapter 3) and the AFLP analysis of the large dataset (Chapter 4) show that the taxonomic structure of *Solanum* section *Petota* is highly unbalanced. Some subgroups of the section *Petota* have high support and their inner structure also displays supported subdivisions, while a large number of the species cannot be further classified into taxonomic groups. These species seem to be equally related to each other and to the supported groups as displayed by a large unresolved polytomy in the trees. Only partial support for the series classification of Hawkes was found and the data also showed some discrepancies with the 4 clade hypothesis of Spooner and co-authors. Our AFLP results showed more groups than the four main clades found with cpDNA restriction data, and the groups were not completely analogous. Because both the series classification and the 4 clade hypothesis are found to be deficient, a new alternative classification, one of informal species groups, was proposed. This approach was intentionally informal (in contrast to a closed classification as required by the rules of the International Code for Botanical Nomenclature) because many species could not be accommodated in any group. Our informal classification can be viewed as a significant extension of a previous informal species group classification by Spooner and co-authors for the North and Central American members of section *Petota*. Based on group support as provided by the NJ jackknife tree, 10 informal species groups (Diploid Mexican group, Acaulia group, Iopetala group, Longipedicellata group, Polyploid Conicibaccata group, Diploid Conicibaccata group, Diploid Piurana group, Tetraploid Piurana group, Circaeifolia group and Verrucosa group) could be distinguished.

Hybridization between species and hybrid speciation are often mentioned as the underlying causes for the complications in the systematics of section *Petota*. Although previous studies have showed that the claims of hybrid species should be considered with caution, recently some evidence for the possible hybrid origin of members of series *Longipedicellata*, series *Conicibaccata*, and series *Iopetala* was published. Several results presented in this thesis support the possible presence of hybridization in section *Petota*. The comparison of nuclear AFLP data with maternally inherited cpDNA sequence data revealed important incongruencies. This indicates unexpected gene flow between species and species groups. Furthermore, both the lack of support for the inter-group relationships in the cpDNA and AFLP tree and the lack of structure found within the South American part of the AFLP tree could point at the influence of hybrization and introgression, which could have a homogenizing effect on the relationships between the species and possibly also the higher taxa. The indications of hybridization in our and previous results should be confirmed by further research. The observed lack of support for the relationships between the different species groups found in the NJ and MP trees, and the lack of support for any systematic structure in the South American part of section *Petota* could also be caused by other processes than hybridization.

A general lack of phylogenetic signal as a possible cause can be rejected because of its obvious presence in the well-supported groups in the same trees and similar results (lack of resolution for part of section Petota) from previous taxonomic studies. A more likely explanation could be found in biological causes. Similar patterns of poor resolution have been reported in studies on other plant taxa. The possible rapid and recent diversification in the Andes of *Lupinus* for example, would have been caused by a combination of ecological and geographical factors and fluctuations of these factors in time. These underlying factors could also have influenced the evolution of wild potatoes possibly in combination with the above mentioned hybridization. Another but partly overlapping general explanation is given by the hypothesis that assumes that the ratio in length between internal and external branches in a tree influences its resolvability. Homoplastic characters can mislead the reconstruction of the short stems (characteristic for radiation) by obscuring the true phylogenetic signal. Both hypotheses suggest the occurrence of a rapid radiation or many speciation events within a short stretch of time.

The other main issue in potato taxonomy, over-classification, is treated in Chapter 5. According to many contemporary authors who focused on the taxonomy of the wild potatoes, *Solanum* section *Petota* is over-classified and speculations on taxonomic changes for several species labels have already been made in previous publications. The polytomy of the mainly South American species found in the results of Chapters 3 and 4 is further scrutinized in Chapter 5. By using methods currently prevalent in population genetics studies, evidence was found that in the South American part of the section *Solanum* many species labels do not correspond to species. This seems to be caused by two different phenomena; misclassification and over-classification. We defined a case as misclassification when accessions with identical species labels appear in different groups and are combined with accessions with different species labels. When accessions with different species labels are always combined together in a group this would be defined as overclassification. For many species labels analyzed in chapter 5 one or both phenomena could be observed.

A revision of the species status of many species in section *Petota* is much needed. Clear support for species status was found for only 8 species labels and for another 9 species labels (plus 6 combinations of 2 different species labels) weak evidence was found. For the other species labels investigated no support for species status could be found. Based on these results, we expect that no more than 46 taxonomical units will be found instead of the initial 90 species labels examined in this chapter. Further research is needed to elaborate the results and identify potential alternative taxonomical units but our information on the species status of several taxa can serve as a firm handhold for future investigations.

The recognition of wild potato species from Central and South America as primary sources for resistance against pests, diseases and abiotic stress, has resulted in numerous collecting expeditions, and to the establishment of a number of potato germplasm collections worldwide. The wild potato species in these genebanks are important for breeding programs, being sources of genetic diversity (base broadening) as well as sources for genetic resistances to diseases, pests and abiotic stresses. For breeders, data on crossability is the most important information but a stable taxonomy can provide additional valuable information for the interpretation of the morphological and genetic diversity within crossing groups. Insight in the systematic relationships within the tuber-bearing *Solanum* species might help to identify the most interesting materials for breeding purposes. The taxonomic results in our study were combined with the original passport data of the used accessions and with information from other accompanying projects within the CBSG consortium. All this information was made available to users (scientists and breeders).

The second goal of this thesis, next to elucidating the taxonomy of section *Petota*, was to search for new *Phytophthora infestans* resistance (R) genes in wild potato species. *P. infestans* causes the most notorious disease, late blight, in potato cultivation. Late blight has the ability to destroy entire fields of potato in a few weeks or even days. Despite the intensive control management with fungicides complemented with other measures, it is still increasingly difficult and costly to control. This and the possible risk of developing fungicide resistant strains of the pathogen and the unknown environmental burden of the fungicides, make the development of resistant cultivars much desired. Natural populations of wild potato species were observed to show resistance against *P. infestans*, especially *S. demissum* plants in Mexico. Over the last century, 11 late blight R genes from *S. demissum* were introduced into cultivated potato but the resistances were quickly broken by the pathogen.

The presence of R genes in some other wild species has been investigated but many wild species available in genebanks have not been tested yet for the presence of R genes against *P. infestans*. In Chapter 6 we investigated some of these possible sources for *P. infestans* resistance. We developed and tested a novel approach to identify to which cluster a new R gene belongs and to obtain markers that can be used for introgression breeding. Mapping R genes is usually accomplished by producing a mapping population that is phenotyped for the resistance trait and genotyped using a large number of markers. Our approach is novel in using a method called NBS (Nucleotide Binding Site) profiling, that specifically targets R genes and their resistance gene analogues (RGAs).

We searched for markers (NBS bands) that are linked to *P. infestans* resistance in small segregating populations. To identify the resistance gene cluster a targeted gene belongs to, the NBS bands were sequenced and the sequences were analyzed using bio-informatics tools. Putative map positions arising from this analysis are validated using already mapped markers in the segregating population. The versatility of the approach is demonstrated on a number of populations derived from wild *Solanum* species segregating for *P. infestans* resistance. We found *P. infestans* resistance genes in accessions of *S. verrucosum* (chromosome 6), *S. schenckii* (chromosome 4) and *S. capsicibaccatum* (chromosome 11). The species that expressed the *P. infestans* resistance belonged to various species groups; respectively to the Verrucosum, Circaeifolia and Iopetala group. Although the *P. infestans* resistance was found in several wild *Solanum* species previously not used or investigated for *P. infestans* resistance, it seems that the genes causing the resistance are linked to known clusters of R genes and in one case show high homology with a known R gene. In another recent study, genes derived from wild *Solanum* species could also be identified as homologues of already known and mapped R genes. Identical R genes, shared by *Solanum* species from completely different species groups, could indicate unsuspected genetic relationships and they could facilitate the use of R genes for breeding programs. Resistance genes in species that cannot be crossed easily with cultivated potato may have homologues in more advanced species that are easily crossable with cultivated germplasm. In Chapter 7 (General discussion) the value of R genes and the application of different strategies using R genes in potato cultivation are briefly discussed.

# Samenvatting

De aardappel is vanuit agrarisch en economisch oogpunt een belangrijk gewas, en de botanische en taxonomische achtergrond wordt al sinds de 19e eeuw intensief bestudeerd. Reeds vroeg realiseerde men zich dat de wilde verwanten van de gecultiveerde aardappel konden dienen als kruisingsmateriaal waarmee bestaande cultivars konden worden verbeterd. Taxonomie is de wetenschap van het indelen en in de biologie betekent dit het beschrijven, indelen en benoemen van organismen. Het indelen gebeurt bijna altijd op basis van veronderstellingen over natuurlijke verwantschap. Het onderzoek naar de verwantschap tussen de verschillende wilde verwanten van de aardappel wordt echter bemoeilijkt door verschijnselen als polyploidisatie, hybridisatie en morfologische plasticiteit. Daarnaast bestaan er tussen bepaalde soorten kruisingsbarrières die worden veroorzaakt door een onbekend mechanisme, EBN genaamd, wat verder bijdraagt aan de verwarring.

In Hoofdstuk 1 wordt eerst een korte samenvatting van de geschiedenis van de aardappeltaxonomie gegeven en in Hoofdstuk 2 wordt een overzicht gegeven van de moleculaire studies in de aardappeltaxonomie, de doelstellingen en de behaalde resultaten. De meeste eerdere taxonomische studies bestonden uit morfologische waarnemingen en op beperkte schaal uit experimentele methoden zoals cytogenetica, and hybridisatie-experimenten. Binnen het genus *Solanum* sectie *Petota*, waartoe de cultuuraardappel en haar wilde verwanten behoren, zijn meer dan 200 soorten en vele infraspecifieke taxa beschreven. De soorten zijn ingedeeld in groepen die series worden genoemd. Verschillende auteurs/deskundigen erkennen een verschillend aantal series met vaak afwijkende beschrijvingen. Eén van de meest gebruikte en invloedrijke aardappeltaxonomische studies is gepubliceerd door prof. J.G. Hawkes in 1990. Ondanks de inspanningen van Hawkes and vele andere uitstekende taxonomen bestaan er nog steeds problemen met de identificatie van soorten met behulp van morfologische determinatiesleutels, met de overclassificatie van grote delen van sectie *Petota*, en er is onduidelijkheid over de series classificatie.

Het toepassen van moleculaire methoden in aardappeltaxonomie heeft de mogelijkheden voor het oplossen van ingewikkelde kwesties en het begrijpen van de taxonomie vergroot. Echter, de meeste taxonomische studies in aardappel hebben een gedeelde tekortkoming: ze bestrijken niet de gehele breedte van de aanwezige variatie in de sectie *Petota* en in de meeste gevallen worden slechts één of zeer weinig accessies gebruikt wat van invloed kan zijn op de uitkomst van de analyses.. Het eerste doel van dit proefschrift tracht de taxonomische relaties tussen wilde *Solanum* soorten binnen sectie *Petota* te verhelderen. Om de breedte van de variatie te dekken en ter preventie van het nemen van een te kleine steekproef, werd een grote set aardappelaccessies bemonsterd (grotendeels verkregen van de Nederlandse genenbank CGN, en aangevuld met accessies van vele andere internationale genenbanken). In totaal werden er 916 accessies gebruikt, die meer dan 190 soorten vertegenwoordigen. Elke soort is, indien mogelijk, vertegenwoordigd door ten minste 5 verschillende accessies en elke accessie door tenminste 5 genotypen. Dit resulteerde in de constructie van de grootste dataset ooit (van ongeveer 5000 genotypen) voor *Solanum* sectie *Petota*.

De bemonsterde planten werden geanalyseerd met behulp van 2 AFLP primer combinaties die uiteindelijk 222 bruikbare merkers opleverden. In de hoofdstukken 3, 4 en 5 gebruiken we deze dataset of delen ervan om enkele eerder genoemde taxonomische vraagstukken op te lossen: de taxonomie van de aardappelsoorten op hoger niveau en de veronderstelde overclassificatie van sectie *Petota*.

In Hoofdstuk 3 maken we gebruik van 2 verschillende soorten merkers voor het reconstrueren van de fylogeny van genus *Solanum* sectie *Petota*: chloroplast (cp) DNA sequentie data en nucleaire AFLP data. Deze bezitten, volgens eerdere studies, een verschillend niveau van oplossend vermogen. Amplification Fragment Length Polymorphism (AFLP) is een techniek om het genoom van organismen van elkaar te kunnen onderscheiden en genotyperen. Het uiteindelijke resultaat zijn streeppatronen die worden genalyseerd op hun overeenkomsten en verschillen. Voor het verkrijgen van DNA sequentie data wordt de volgorde van de nucleotiden van een bepaald stuk genoom bepaald (in dit geval uit het genoom in de bladgroenkorrels) de volgorde van de nucleotiden (bouwstenen van het DNA molecuul). Een set van 199 accessies, afkomstig van 174 verschillende taxa, werd gebruikt in de analyses. De sequenties van de chloroplast DNA regio's *trn*TLF and *psb*A/*trn*H werden bepaald. De AFLP gegevens werden als een subset uit de grote AFLP dataset gehaald. Voor de chloroplast DNA sequenties en de AFLP data analyse werd exact hetzelfde plant materiaal gebruikt.

Beide datasets werden apart geanalyseerd op een fenetische en fylogenetische manier en er werden aparte fylogenetische bomen geproduceerd om de onderlinge verwantschap visueel weer te geven. Deze aanpak maakt het mogelijk om de uitkomsten van de cpDNA analyse en de AFLP analyse te vergelijken. Vanwege de resultaten in Hoofdstuk 3, moest het originele plan, dat bestond uit de constructie van een grove ruggengraat fylogenie met behulp van cpDNA data, en het invullen van de gedetailleerde fylogenetische structuur met behulp van AFLP, worden verlaten. Deze ruggengraat fylogenie aanpak zou de scoring van AFLP banden hebben vergemakkelijkt en zou het mogelijke risico op het introduceren van homoplasie hebben verminderd. Helaas vertoonde de chloroplast DNA data veel minder oplossend vermogen dan de AFLP resultaten, en slechts enkele groepen konden worden onderscheiden. Bovendien gaven de chloroplast DNA resultaten belangrijke verschillen met de AFLP resultaten.

In Hoofdstuk 4 werd de complete dataset van 4929 genotypen gebruikt voor een uitgebreide AFLP analyse. Het was mogelijk om voor de grote oorspronkelijke dataset (4929 genotypen) een fenetische Neighbour Joining boom te construeren. Vanwege technische beperkingen, was het echter nodig om voor de fylogenetische analyses en het berekenen van statistische sterkte/onderbouwing van de gevonden taxonomische structuur, een gecodenseerde dataset te creëren. De gecondenseerde dataset bestond uit 916 genotypen die alle aanwezige accessies vertegenwoordigen.

De resultaten van de gecombineerde chloroplast/AFLP analyse van de subset (Hoofdstuk 3) en de AFLP analyse van de grote dataset (hoofdstuk 4) laten zien dat de taxonomische structuur van *Solanum* sectie *Petota* zeer ongebalanceerd is. Sommige bestaande subgroepen van de sectie *Petota* worden statistisch goed ondersteund, alsmede hun interne structuur, terwijl daarnaast een groot aantal soorten niet verder kan worden ingedeeld in taxonomische groepen. Deze soorten zijn onderling allemaal even verwant aan elkaar en aan de wel goed ondersteunde groepen. Voor de series classificatie van Hawkes werd slechts gedeeltelijke ondersteuning gevonden en onze resultaten vertoonden ook enkele belangrijke verschillen met de 4 clade hypothese van Spooner en coauteurs. Onze AFLP resultaten onderscheidden meer dan 4 groepen en de gevonden groepen zijn niet volledig analoog. Omdat zowel de series classificatie als de 4 clade hypothese met onze resultaten slechts gedeeltelijk kunnen worden bevestigd, stellen we een alternatieve open classificatie voor, daarbij gebruikmakend van informele soortsgroepen. Deze aanpak is opzettelijk informeel en open (in tegenstelling tot een gesloten classificatie zoals vereist volgens de regels van de International Code for Botanical Nomenclature), omdat veel soorten (tot nu toe) in geen enkele groep kunnen worden ondergebracht. Onze informele classificatie kan worden beschouwd als een aanzienlijke uitbreiding van de eerdere informele soortsgroep classificatie voor de Noord- en Centraal Amerikaanse soorten van de sectie *Petota* door Spooner en coauteurs. Gebaseerd op de gevonden statistische ondersteuning voor de verschillende groepen in de NJ jackknife boom, konden 10 informele soortsgroepen worden onderscheiden: de Diploïde Mexicaanse groep, Acaulia groep, Iopetala groep, Longipedicellata groep, Polyploïde Conicibaccata groep, Diploïde Conicibaccata groep, Diploïde Piurana groep, Tetraploïde Piurana groep, Circaeifolia groep en Verrucosa groep.

Hybridisatie tussen soorten en soortsvorming door hybridisatie worden vaak genoemd als potentiële oorzaken voor de problemen in de taxonomie van de sectie *Petota*. Hoewel eerdere studies hebben aangetoond dat voorzichtigheid geboden is bij beweringen over de hybride oorsprong van soorten, is er recentelijk bewijs gepubliceerd voor de mogelijke hybride oorsprong van leden van de series *Longipedicellata,* series *Conicibaccata, en* series *Iopetala*. Ook verschillende resultaten uit dit proefschrift suggereren de aanwezigheid van hybridisatie in de sectie *Petota*. De vergelijking van nucleaire AFLP data met maternaal overgeërfd chloroplast DNA sequenties onthulde belangrijke incongruenties. Dit wijst op genetische uitwisseling tussen soorten en soms zelfs tussen soortsgroepen. Daarnaast kan zowel het gebrek aan ondersteuning voor de relaties tussen de groepen in de chloroplast DNA en de AFLP boom en het gebrek aan structuur in het Zuid-Amerikaanse deel van de AFLP boom worden beschouwd worden als een aanwijzing voor de invloed van hybridisatie en introgressie op de evolutionaire geschiedenis binnen sectie *Petota*. Beide processen kunnen de relaties tussen de soorten en mogelijk ook de relaties tussen taxa op hogere niveaus vertroebelen. De aanwijzingen voor het bestaan van hybridisatie in onze en in eerdere resultaten zouden echter moeten worden getoetst met meer onderzoek, want zij kunnen ook worden veroorzaakt door andere processen dan hybridisatie alleen. Een gebrek aan fylogenetisch signaal kan worden verworpen als een mogelijke oorzaak vanwege de overduidelijke aanwezigheid van fylogenetisch signaal in de goed ondersteunde groepen in dezelfde bomen en de vergelijkbare resultaten in eerdere taxonomische studies. Een meer waarschijnlijke verklaring kan gevonden worden in biologische oorzaken. Vergelijkbare patronen van slecht op te lossen fylogeniëen werden ook gerapporteerd in studies in andere planten.

De mogelijk snelle en relatief recente diversificatie van het genus *Lupinus* in de Andes regio bijvoorbeeld, zou mogelijk veroorzaakt kunnen worden door een combinatie van ecologische en geografische factoren en de fluctuaties van deze factoren in het verloop van de tijd. Deze factoren kunnen ook van invloed geweest zijn op de evolutie van wilde aardappelsoorten, mogelijk in combinatie met de al eerder genoemde hybridisatie. Een andere, maar gedeeltelijk overlappende, verklaring wordt gegeven door de hypothese die stelt dat de ratio van de lengte tussen de interne en externe takken in een boom bepaalt of de boom kan worden opgelost. Homoplastische kenmerken kunnen de reconstructie van korte takken (kenmerkend voor zogenaamde soortsradiaties) misleiden door het verstoren van het werkelijke fylogenetische signaal. Beide hypotheses suggereren het optreden van een snelle radiatie of vele soortvormingsgebeurtenissen binnen een korte tijdspanne.

Een andere belangrijke kwestie in aardappel taxonomie; de overclassificatie, wordt besproken in Hoofdstuk 5. Volgens verschillende hedendaagse auteurs die zich hebben beziggehouden met de taxonomie van de wilde aardappelsoorten, is *Solanum* sectie *Petota* overgeclassificeerd en een herziening van een aantal soortsnamen is hard nodig. De polytomie bestaande uit hoofdzakelijk Zuid Amerikaanse soorten, zoals gepresenteerd in de resultaten van Hoofdstukken 3 and 4, wordt verder onderzocht in Hoofdstuk 5. Met behulp van een analysemethode die gangbaar is in het onderzoek van de populatiegenetica, werd bewijs gevonden dat vele soortslabels, behorend tot het Zuid Amerikaanse deel van de sectie *Petota,* niet overeenkomen met "echte" soorten. Dit lijkt veroorzaakt te worden door 2 verschillende fenomenen, misclassificatie en overclassificatie. We definiëren een geval als misclassificatie indien accessies met identieke soortslabels in de analyse in verschillende groepen terechtkomen en gecombineerd worden met accessies met andere soortslabels. Indien accessies met verschillende soortslabels altijd worden gecombineerd als een hechte groep, kan dit worden beschouwd als overclassificatie. Voor veel soortslabels die in Hoofdstuk 5 werden geanalyseerd, kon één van beide fenomenen worden geobserveerd. Goede ondersteuning voor het hebben van soortsstatus werd alleen gevonden voor 8 soortslabels en voor 9 andere soortslabels werd zwakke ondersteuning gevonden. Voor de overige soortslabels kon geen bewijs voor soortsstatus worden gevonden. We verwachten, op basis van deze resultaten, dat slechts 46 taxonomische eenheden kunnen worden onderscheiden, in plaats van de oorspronkelijke 90 soortslabels die werden onderzocht. Meer onderzoek is nodig om de resultaten uit te werken en potentiële alternatieve taxonomische units te identificeren. Onze resultaten kunnen gebruikt worden als een goede basis voor een dergelijk toekomstig onderzoek.

De erkenning van wilde aardappelsoorten uit Centraal- en Zuid Amerika als primaire bronnen voor resistenties tegen plagen, ziektes en abiotische stress heeft wereldwijd geresulteerd in het organiseren van vele verzamelexpedities en tot het oprichten van een aantal collecties (zogenaamde genenbanken) met genetisch materiaal van wilde aardappelsoorten. De aardappelsoorten die zich in deze genenbanken bevinden zijn belangrijk voor veredelingsprogramma, zowel als basis voor algemene genetische diversiteit (verbreden van de genetische basis) als bron voor specifieke resistenties.

Voor veredelaars zijn gegevens over kruisbaarheid verreweg de meest belangrijke informatie maar desondanks kan een stabiele taxonomie toegevoegde waarde hebben voor de interpretatie van morfologische en genetische diversiteit binnen kruisingsgroepen. Inzicht in de systematische relaties tussen de wilde knoldragende *Solanum* soorten kan helpen om het meest interessante materiaal voor veredeling te identificeren. De taxonomische resultaten uit onze studie werden gecombineerd met de originele paspoort data van de gebruikte accessies en met informatie van zuster projecten binnen het CBSG consortium. Al deze informatie was en is nog steeds toegankelijk voor gebruikers (wetenschappers en veredelaars) binnen het CBSG.

Het tweede doel van dit proefschrift, naast het verhelderen van de taxonomie van de sectie *Petota*, was de zoektocht naar nieuwe *Phytophthora infestans* resistentie (R) genen in wilde aardappel soorten. *P. infestans* is de oorzaak van de meest beruchte ziekte in aardappelproductie. De aardappelziekte (er bestaat geen officiële Nederlandse naam) is in staat om complete aardappelvelden te vernietigen binnen slechts enkele weken of dagen. Ondanks intensieve bestrijdingsprogramma's met fungiciden aangevuld met andere maatregelen, is het nog steeds zeer moeilijk en prijzig om de ziekte te bestrijden. Daarnaast is er een mogelijk risico op het ontwikkelen van pathogene lijnen die resistent zijn tegen fungiciden. Vanwege deze zaken plus de onbekende belasting van het milieu door het gebruik van fungiciden is de ontwikkeling van resistente cultivars zeer gewenst. In natuurlijke populaties van wilde aardappel soorten, voornamelijk in populaties van *S. demissum* planten in Mexico, werd natuurlijke resistentie tegen *P. infestans* geobserveerd. In totaal werden er afgelopen decennia 11 *P. infestans* resistentie genen vanuit *S. demissum* ingebracht in aardappelrassen door middel van kruisingen, maar alle resistenties werden snel doorbroken door de ziekteverwekker.

De aanwezigheid van R genen in andere wilde aardappelsoorten is wel onderzocht maar nog lang niet alle wilde soorten die beschikbaar zijn in de genenbanken zijn getest voor de aanwezigheid van *P. infestans* resistentiegenen. In Hoofdstuk 6 hebben we enkele van deze mogelijke nieuwe bronnen voor *P. infestans* resistentie onderzocht. We ontwikkelden en testten een nieuwe aanpak voor de identificatie en positionering van nieuwe resistentiegenen. Daarnaast produceerden we merkers die kunnen worden gebruikt in introgressie veredeling. De positie van resistentie genen wordt normaliter bepaald door middel van het maken van een grote karteringspopulatie, die wordt gefenotypeerd voor een specifieke eigenschap, zoals een resistentie, en geanalyseerd met behulp van een groot aantal merkers. Onze aanpak is nieuw omdat we gebruikmaken van een methode, NBS profiling genaamd (Nucleotide Binding Site), die specifiek is gericht op resistentiegenen en resistentie gen analogen (RGAs). We zochten naar merkers (NBS profiling banden) die gekoppeld zijn met *P. infestans* resistentie in kleine splitsende populaties. Om te ontdekken tot welk resistentiegen-cluster het beoogde gen behoort, werd de DNA sequentie van de betreffende NBS banden bepaald. Vervolgens werden de sequenties geanalyseerd door middel van bioinformatica toepassingen. Dit resulteerde in potentiële karteringsposities die werden getoetst in de splitsende populaties door middel van reeds eerder gekarteerde merkers. De veelzijdigheid van deze aanpak wordt aangetoond in een aantal populaties die afstammen van wilde *Solanum* soorten en die uitsplitsen voor *P. infestans* resistentie. We vonden *P. infestans* resistentie genen in accessies van *S. verrucosum* (chromosoom 6), *S. schenckii* (chromosoom 4) and *S. capsicibaccatum* (chromosoom 11).

Deze wilde soorten die de *P. infestans* resistentie vertonen behoren tot verschillende soortsgroepen; respectievelijk tot de Verrucosum, Circaeifolia en Iopetala groep. Hoewel *P. infestans* resistentie werd gevonden in verscheidene wilde *Solanum* soorten die niet eerder werden gebruikt of onderzocht op de aanwezigheid van *P. infestans* resistentie, lijkt het erop dat de genen die deze resistentie veroorzaken, behoren tot bekende resistentiegen clusters. In één geval vertoont het gevonden gen zelfs zeer grote homologie met een reeds bekend resistentiegen. In een andere recente studie, werden genen afkomstig van wilde *Solanum* soorten reeds geïdentificeerd als homologen van reeds bekende en gekarteerde resistentiegenen. Identieke resistentiegenen uit verschillende *Solanum* soorten van zeer verschillende soortsgroepen kunnen een aanwijzing zijn voor (onverwachte) genetische verwantschap en daarnaast zouden ze het gebruik van resistentiegenen voor veredelingsprogramma's kunnen faciliteren. Resistentiegenen in soorten die niet makkelijk te kruisen zijn met de gecultiveerde aardappels kunnen mogelijk homologen hebben in meer geavanceerde soorten die wel makkelijk te kruisen zijn met cultivars. In Hoofdstuk 7 (Algemene discussie) wordt het nut van resistentiegenen, en de toepassing van verschillende veredelingsstrategieën, waarbij gebruik wordt gemaakt van resistentiegenen, kort besproken.

# Dankwoord

De wandeling is bijna ten einde. Ik ben eigenlijk al aan een nieuwe wandelroute begonnen, maar routes lopen vaak in elkaar over. Op deze plek wil ik iedereen bedanken die stukken van de route samen met mij heeft gewandeld, of die mijn pad heeft gekruist of mijn zere voeten heeft verzorgd.

Allereerst wil ik mijn begeleiders bedanken. Ben en Ronald. Jullie waren een uitstekend team, en vulden elkaar waar nodig aan. Als ik bij de een niet terecht kon, stond de ander klaar om te discussiëren of zo nodig steun te verlenen. Soms was het nodig om mijn optimisme bij te stellen naar de realiteit, maar veel vaker nog vonden jullie het nodig om mijn realisme te ontdoen van onnodige visioenen over "beren op de weg". Discussies met ieder van jullie apart, maar ook samen, waren altijd leerzaam en aangenaam, en soms was er zelfs tijd voor filosofische bespiegelingen.

Dan wil ik graag mijn beide promotoren, Marc Sosef en Richard Visser bedanken. Jullie stonden, met name in de eindfase, altijd klaar voor mij. Ondanks jullie ongetwijfeld volle agenda's kreeg ik altijd snel reactie op alles wat ik jullie stuurde en jullie kritische opmerkingen hebben mijn proefschrift zeker verbeterd.

Naast mijn officiële begeleiders zijn er ook een aantal mensen waarmee ik veel heb gepraat en overlegd en die met hun discussie veel aan de ontwikkeling van het proefschrift hebben bijgedragen. Mijn dank gaat uit naar Rene, Gerard, Eric en Herman. Daarnaast ben ik ook veel dank verschuldigd aan Roel Hoekstra van CGN voor al het (voor)werk van mijn project en de nuttige adviezen en discussies over planten en databases die we van tijd tot voerden. Dank ook aan Bastienne Vriesendorp van lsg Biosystematiek, Theo van Hintum van CGN, Jaap Buntjer van Keygene en Jack Leunissen van de lsg Bioinformatica die tijd hebben gestoken in lastige en interessante zaken die we in dit project tegenkwamen. Veel dank aan Jeroen Engelberts van SARA reken-en netwerkdiensten in Amsterdam, zonder zijn hulp was ik niet in staat geweest mijn (te) grote dataset te analyseren.

Ik ben veel dank verschuldigd aan de mensen die het meeste praktische werk op zich hebben genomen. Rolf Mank en Mariëlle Sengers van Keygene die de DNA isolatie van de planten en de AFLP fingerprints hebben gedaan. Vivianne, Marcel, Dirk, Bernadette en collega's, heel erg bedankt voor de enorme klus van het opkweken van alle plantjes, toetsen van de planten en het maken van populaties. In het bijzonder wil ik bedanken: Christel en Betty en later ook Henry, Linda en Martijn voor alle labwerkzaamheden met NBS profiling en overige experimenten die nodig waren. Zonder jullie geen data, en dus geen proefschrift. Tijdens de periode dat ik zelf nog in het lab stond heb ik veel begeleiding gekregen van Hanneke, Danny, Yolanda, Wendy, Martijn, Christel, Gerda S., Gerda U. en nog vele andere mensen.

Mijn werk maakte deel uit van het CBSG, een netwerk waarin wetenschappers en mensen uit het bedrijfsleven samenwerken. Ik wil het CBSG, secretariaat (dames bedankt voor jullie snelle en efficiente reacties en geregel!) en management, bedanken maar vooral ook alle mensen waarmee ik heb samengewerkt binnen het CBSG. Mijn speciale dank gaat uit naar Vivianne Vleeshouwers, Edwin van der Vossen, Ralp van Berloo, Erin Bakker en Aska Goverse, Francine Goverts, Gerard van der Weerden, Titti Mariani, Tomeck, en Patrick Butterbach. Veel van mijn dank gaat ook uit naar de vertegenwoordigers van de veredelingsbedrijven die zich met de aardappelprojecten bezighielden, en met name Sjefke Allefs en Marielle Muskens (Agrico) en Guus Heselmans (Meijer BV). Zonder jullie betrokkenheid en inbreng had mijn proefschrift er veel slechter uitgezien, bedankt hiervoor.

Voor de nodige ontspanning en gezelligheid tijdens de werkweek zou ik eigenlijk een ontelbare hoeveelheid mensen willen bedanken, maar ik zal er slechts een paar noemen. Allereerst veel dank aan de dames uit het "kippenhok"; Adriana, Chang, Eveline, Colette, Marleen, Brigitte en Yulia waarmee ik veel lief en leed mee gedeeld heb. Daarnaast wil ik Paul, Clemens en Martijn bedanken die regelmatig gezellig kwamen buurten in ons kippenhok. Verder wil ik het clubje van lunchgangers bedanken van Rene, Henk, Paul en Martijn aangevuld met dames uit "het kippenhok". Verder wil ik al mijn collega's van Plant Research International, leerstoelgroep Plantenveredeling en leerstoelgroep Biosystematiek bedanken voor de fijne werksfeer, de goede samenwerking en de gezellige feestjes en uitstapjes.

Er bestaat ook leven buiten de wetenschap. Ik wil iedereen buiten mijn werkomgeving bedanken voor het begrip, geduld en de belangstelling voor mijn project. Mijn huisgenoten, vrienden en kennissen wil ik bedanken voor de gezelligheid, steun en gezelschap in goede en slechte tijden. Sjaak, bedankt voor je steun, geduld en liefde in de laatste fase van dit project, je hebt het promoveren en het leven lichter gemaakt en ik hoop dat we de wandeling samen kunnen voortzetten.

Als laatste, maar zeker niet als minste, wil ik mijn ouders bedanken hun steun en vertrouwen. Dankzij jullie geloof in mijn kunnen en natuurlijk "het voeden en opvoeden" kan ik deze periode op een mooie manier afsluiten, dit proefschrift is dus eigenlijk ook een beetje van jullie!

# Curriculum Vitae

Mirjam Jacobs werd geboren op 5 december 1975 in Sittard en groeide op in Nieuwstadt. Ze doorliep het VWO op Serviam college in Sittard en begon in 1994 aan de studie Biologie aan de Wageningen Universiteit. Ze koos na het behalen van haar propaedeuse voor de specialisatie populatie-ecologie. Ze deed een afstudeervak populatiegenetica bij de leerstoelgroep Populatiegenetica waarbij ze onderzoek deed naar de fitness kosten voor fungicide resistentie en compenserende mutaties in de schimmel *Aspergillus nidulans*. Hierna volgende een stage Plantentaxonomie bij het International Potato Centre in Lima, Peru. Het onderwerp van de stage was een onderzoek naar mogelijke associaties tussen vorstresistentie in wilde aardappelsoorten en taxonomische en geografische variabelen. Terug in Nederland volgde nog een afstudeervak Plantentaxonomie bij de leerstoelgroep Biosystematiek met als onderwerp de taxonomische analyse van de wilde verwant van de slaplant; *Lactuca altaica*, met behulp van numerieke taxonomie, literatuurstudie en moleculaire technieken. Gedurende haar studie was ze actief in de Jeugdbond voor Milieu- en Natuurstudie (JNM). Ze studeerde af in januari 2001 en was achtereenvolgens werkzaam als technisch onderwijsassistent natuurkunde en scheikunde op scholengemeenschap Pantarijn, secretaresse juridische afdeling Rabobank Nederland, en administratief medewerker bij verzekeringsmaatschappij Menzis. Vanaf augustus 2002 werkte ze als botanisch analist bij de leerstoelgroep Experimentele Plantensystematiek, Universiteit van Amsterdam, bij het Europese ANGEL project waarin onderzoek werd gedaan naar de effecten van introgressie van gecultiveerde sla-genen op wilde populaties van *Lactuca serriola*.

In maart 2004 begon ze als AIO bij Plant Research International en de leerstoelgroepen Biosystematiek en Plantenveredeling. Het onderzoek richtte zich op het verhelderen van de taxonomische relaties tussen de wilde verwanten van de aardappel en daarnaast op het zoeken naar nieuwe bronnen (in wilde aardappelsoorten) van resistentie tegen de aardappelziekte *Phytophthora infestans*. Het onderzoek maakt deel uit van het Centre for Biosystens Genomics (CBSG), een consortium waarin onderzoekers en bedrijven nauw samenwerken. Na het afronden van haar onderzoek zal ze haar loopbaan voortzetten als project adviseur Life Sciences bij Senter Novem in Den Haag.

# Education Statement of the Graduate School
# Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

**Issued to:** Mirjam M.J. Jacobs
**Date:** 31 October 2008
**Group:** Laboratory of Biosystematics and Laboratory of Plant Breeding
Wageningen University and Research Centre

| 1) Start-up phase | date |
|---|---|
| ▶ **First presentation of your project** | |
| Oral presentation "Identifying new sources of resistance in wild accessions of Solanum via relationship determination | September 02, 2004 |
| ▶ **Writing or rewriting a project proposal** | |
| ▶ ***Writing a review or book chapter*** | |
| Book Chapter: "Molecular Taxonomy" in "The Potato"; edited by D. Vreugdehil, published in 2007 | 2007 |
| ▶ **MSc courses** | |
| ▶ **Laboratory use of isotopes** | |
| *Subtotal Start-up Phase* | *7.5 credits** |

| 2) Scientific Exposure | date |
|---|---|
| ▶ **EPS PhD student days** | |
| EPS PhD student day 2004 , University of Amsterdam | June 03, 2004 |
| *EPS PhD student day 2005, Radboud University Nijmegen* | *June 02, 2005* |
| EPS PhD student day 2007, Wageningen University | September 13, 2007 |
| ▶ **EPS theme symposia** | |
| EPS theme 4 Genome Plasticity symposium 2006, Radboud University, Nijmegen | December 08, 2006 |
| EPS theme 4 Genome Plasticity symposium 2007, Leiden University, Leiden | December 07, 2007 |
| ▶ **NWO Lunteren days and other National Platforms** | |
| NWO-ALW Plant sciences meeting Lunteren | April 02-03, 2007 |
| ▶ **Seminars (series), workshops and symposia** | |
| CBSG 2004 plus Potato Cluster Summit meeting | November 12, 2004 |
| *CBSG 2005 Summit plus Potato Cluster meeting* | *Februrary 21-22 , 2005* |
| CBSG 2006 Summit meeting | March 06-07, 2006 |
| CBSG 2006 Potato Cluster Meeting | October 05, 2006 |
| CBSG 2007 Summit meeting | February 06-07, 2007 |
| NHN seminar day | April, 20, 2007 |
| CBSG 2007 Potato Cluster meeting | August 31, 2007 |
| CBSG 2008 Summit plus Potato Cluster Meeting | March 17-18, 2008 |
| ▶ **Seminar plus** | |
| ▶ **International symposia and congresses** | |
| Solanaceae Workshop, Wageningen, The Netherlands | September 19-21, 2004 |
| Plant Gems Workshop Amsterdam, The Netherlands | September 20-23, 2005 |
| International Botanical Congress, Vienna, Austria | July 17-23, 2005 |
| 5th International Symposium on the Taxonomy of Cultivated Plants, Wageningen, The Netherlands | October 15-19, 2007 |
| ▶ **Presentations** | |
| poster presentation on Solanaceae Workshop | September 19-21, 2004 |
| Oral Presentation CBSG 2004 plus Potato Cluster Summit meeting | November 12, 2004 |
| Oral Presentation CBSG 2005 Summit plus Potato Cluster meeting | Februrary 21-22 , 2005 |
| Poster presentation International Botanical Congress, Vienna | July 17-23, 2005 |
| Poster presentation Plant Gems Workshop Amsterdam | September 20-23, 2005 |
| Oral Presentation CBSG 2006 Summit meeting | March 06-07, 2006 |
| Oral Presentation CBSG 2006 Potato Cluster Meeting | October 05, 2006 |
| Oral Presentation CBSG 2007 Summit meeting | February 06-07, 2007 |
| Oral presentation at NWO-ALW Plant sciences meeting 2007 | April 03, 2007 |
| Oral Presentation CBSG 2007 Potato Cluster meeting | August 31, 2007 |
| Oral presentation at 5th International Symposium on the Taxonomy of Cultivated Plants | October 15-19, 2007 |
| Oral Presentation at EPS theme 4 Genome Plasticity | December 07, 2007 |
| Oral Presentation CBSG 2008 Summit plus Potato Cluster Meeting | March 17-18, 2008 |
| ▶ **IAB interview** | |
| ▶ **Excursions** | |
| *Subtotal Scientific Exposure* | *19.5 credits** |

| 3) In-Depth Studies | date |
|---|---|
| ▶ **EPS courses or other PhD courses** | |
| Springschool : Bioinformatics Data Triple I: Information, Integration, Interpretation | March 31, April 01-02 2004 |
| Molecular Phylogenies: reconstruction and interpretation | October 18-22, 2004 |
| Bioinformatics-A User's Approach | March 13-16, 2007 |
| ▶ **Journal club** | |
| PRI Plant Breeding PhD Journal Club 2004-2008 | 2004-2008 |
| ▶ **Individual research training** | |
| *Subtotal In-Depth Studies* | *6.6 credits** |

| 4) Personal development | date |
|---|---|
| ▶ **Skill training courses** | |
| Scientific Writing | May-June 2005 |
| Career Perspectives | October-December 2007 |
| ▶ **Organisation of PhD students day, course or conference** | |
| ▶ **Membership of Board, Committee or PhD council** | |
| *Subtotal Personal Development* | *3.6 credits** |

| TOTAL NUMBER OF CREDIT POINTS* | 37.2 |
|---|---|

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 credits

** A credit represents a normative study load of 28 hours of study*

This PhD project was (co)financed by the research program of the Centre for Biosystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative (NGI) / Netherlands Organization for Sscientific Research (NWO).

Cover design: Mirjam Jacobs and Sjaak Keuvelaar
Thesis layout: Mirjam Jacobs

Printed at Wöhrmann Print Service in Zutphen, the Netherlands