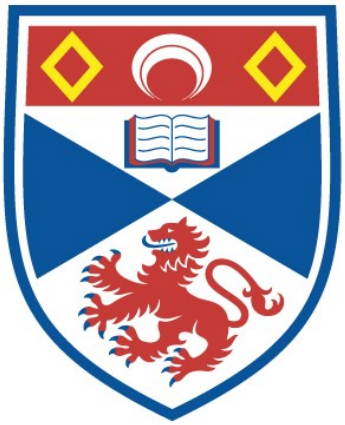
MOLECULAR SYSTEMATICS OF THE NEOTROPICAL TUBEROUS LEGUME 'PACHYRHIZUS' RICH. EX DC, THE YAM BEAN

Jaime Eduardo Estrella Engelmann

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1998

UNIVERSITY OF ST. ANDREWS

SCHOOL OF ENVIRONMENTAL AND EVOLUTIONARY BIOLOGY

Molecular Systematics of the Neotropical Tuberous Legume *Pachyrhizus* Rich. ex DC, the Yam Bean

by

Jaime Eduardo Estrella Engelmann

A thesis submitted to the
University of St. Andrews for
the degree of Doctor of Philosophy

Supervisor: Dr. Richard J. Abbott

St. Andrews, October 1998

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

This thesis is dedicated to my wife María Fernanda,
for unwavering support.

"She gave me eyes, she gave me ears;
And humble cares, and delicate fears;
A heart, the fountain of sweet tears;
And love, and thought, and joy."

William Wordsworth (1770 - 1850)

To those who mean the very most,
my parents Jaime and Hanne,
my brothers Federico and Pablo,
and my aunt Janine.

ABSTRACT

The Neotropical genus *Pachyrhizus* Rich. ex DC. (the yam beans) is one of the few legume genera with edible tuberous roots. Two of its five species (*P. ahipa* and *P. tuberosus*) are cultivated by rural communities in the Andean and Amazonian regions, while a third species (*P. erosus*) is grown on a large scale in Central America for the domestic and international market. Current breeding work in Costa Rica, Ecuador, Denmark, Portugal and Tonga is aimed at increasing the potential use of the genus as a crop. The objective of this research was to construct a phylogeny for the genus based on molecular variation, and to establish relationships and levels of genetic diversity among *Pachyrhizus* species.

A phylogeny based on chloroplast (cp) DNA variation separated the genus into two evolutionary branches, i.e. a Mesoamerican and a South American branch, reflecting a consistent phylogeographic pattern of species distribution and dispersal. The wild species *P. ferrugineus* (resolved as the most primitive within the genus) together with wild and cultivated *P. erosus* represented the Mesoamerican branch. Materials of Andean and Amazonian distribution, namely *P. panamensis*, the *P. tuberosus* complex and the highly derived species *P. ahipa* made up the South American evolutionary branch of the genus. *P. panamensis* was the most primitive taxon in this group.

A phylogeny constructed from sequence variation of the internal transcribed spacer (ITS) region of nuclear ribosomal DNA complemented the cpDNA phylogeny to a broad extent, separating *Pachyrhizus* species into similar clades and reflecting a congruent phylogeographical distribution. The ITS phylogeny also indicated that *P. panamensis* and wild taxa of *P. tuberosus* and *P. erosus* probably originated after rapid radiation from a continuously distributed early ancestor (i.e. ecotypes of *P. ferrugineus*). Subsequently, speciation may have accompanied divergent adaptation to dissimilar ecological niches in two directions: to areas with conspicuous annual dry seasons and deciduous forests in Mesoamerica, which in turn resulted in the origin and specialisation of *P. erosus* and its several primitive landraces and cultivars; and, to the tropical and subtropical rainforests and Andean valleys, giving rise to the different

aided by domestication and man-made selection. Phenograms produced from RAPD variation were congruent to a broad extent with the phylogenies obtained from cpDNA and ITS variation, confirming the affinity between *P. ahipa* and *P. tuberosus*, and revealing the existence of three gene pools in Mesoamerica and Mexico.

P. tuberosus appears to have played a significant role in the evolution of the genus. Multiple lines of molecular evidence suggest that this species might also be ancestral to *P. erosus*, as a separate lineage. Moreover, an early, wild ancestor of *P. tuberosus* was also closely related to other wild species (i.e. *P. panamensis* and *P. ferrugineus*); and, some of the cultigen types of the *P. tuberosus* complex encompass the early ancestry of the highly advanced species *P. ahipa*.

Finally, a broad range of potentially exploitable genetic diversity was identified in this study, enabling both the implementation (or continuation) of *in situ/ex situ* conservation strategies, and a more efficient progress towards genetic breeding of yam beans.

DECLARATION

I, Jaime Eduardo Estrella Engelmann, hereby certify that this thesis, which is approximately 62000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

October, 1998. *Jaime E. Estrella E.*

STATEMENT

I was admitted as a research student to the School of Biological and Medical Sciences, University of St. Andrews, in September, 1995; and, as a candidate for the degree of Ph.D. in the School of Environmental and Evolutionary Biology in September, 1996.

October, 1998. *Jaime E. Estrella E.*

CERTIFICATE

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

October, 1998. *Dr. Richard J. Abbott*

COPYRIGHT

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my supervisor, Dr. Richard J. Abbott, for his valuable advice and enthusiasm, and for patiently going through earlier drafts of this thesis; very special thanks are also due to him for laboratory facilities over the past three years. I also wish to thank Dr. Marten Sørensen (Royal Veterinary and Agricultural University, Copenhagen) for his continued support in providing plant material, valuable literature and suggestions, but - most importantly - encouragement and fruitful discussion during germplasm collections in the quest for *jíquimas* and *ashipas*. I am grateful for a *Sir Harold Mitchell* Scholarship awarded by the University of St. Andrews, and also for an ORS/CVCP award, which allowed me to spend three years at St. Andrews conducting the research reported in this thesis.

I would like to thank INIAP (Instituto Nacional Autónomo de Investigaciones Agropecuarias) and DENAREF (Departamento Nacional de Recursos Fitogenéticos y Biotecnología) in Ecuador, for collaborative support. I am grateful to the germplasm banks and farmers of Central and South America for their generous gifts of seed. In addition, I am indebted to Dr. Amanda Gillies, Dr. Kirsten Wolff and Mr. David Forbes for their valuable expertise and thoughtful assistance during this project. I would like to thank all my friends and colleagues in the laboratory for assisting me with the molecular techniques and for providing an enjoyable working environment. Thanks are also due to Dr. Martin Ryan and Mr. Alex Houston for assistance with DNA sequencing analysis. The support of the Roger and Sarah Bancroft Clark Charitable Trust during the final stage of my Ph.D. studies is also appreciated.

Finally, I offer my sincere gratitude to María Fernanda, my wife, and to my family in Ecuador for valiant support, patience and understanding.

October, 1998.

Contents

CHAPTER 1

GENERAL INTRODUCTION

"... Además de las papas, que son lo principal, son ocas, yanaocas, camotes, jíquimas, yucas, cochucho, cavi, totora, maní y otros cien géneros que no me acuerdo... De aquellas raíces que dije algunas son comida ordinaria; otras sirven para refrescar, como la jíquima, que es muy fría y húmeda; y en verano, en tiempo de estío refresca y apaga la sed; para sustancia y mantenimiento, las papas y ocas hacen ventaja..."

(Padre José Acosta, a Spanish chronicler describing the diversity found in the New World and the use of yam beans by the Amerindians in 1590; cited in Yacovleff, 1933).

1.1. Introduction

Many traditional agroecosystems found in South America are major *in situ* repositories of crop genetic diversity; this native germplasm is crucial to both developing and industrialised countries. Due to encroaching agricultural modernisation and environmental degradation, crop genetic diversity is decreasing worldwide in these ecosystems (IPGRI, 1995). Agricultural research has traditionally focused on a few staple crops (e.g. potato, maize, rice and wheat), while relatively little attention has been placed on minor or neglected crops. Therefore, these minor crops have failed to attract significant financial support and, consequently, limited information is available on many of their basic aspects, hindering their development and sustainable conservation. Research is urgently needed to document levels of genetic variation and causes of genetic erosion in underutilised species.

This study will draw attention to a genus of growing interest and attractiveness, that is attributed with the highest production of leguminous tubers on a world basis (Sørensen, 1990), i.e. the genus *Pachyrhizus* Richard ex DC - the yam beans. It is intended that the investigations described in this thesis will contribute to: **(1)** identifying possible untapped genetic diversity for breeding programmes; **(2)** promoting an understanding of the evolution and biosystematics of the genus; and, **(3)** detecting gaps in existing germplasm collections and conservation approaches.

1.2. Features of the genus *Pachyrhizus*

Yam bean, a Neotropical legume with edible tuberous roots, is an important food crop with a long history of cultivation. The genus is extensively cultivated in the Americas, both as a garden crop and, in the case of *P. erosus* (L.) Urban, for export to North America (Sørensen, 1996). It has been successfully introduced to different pantropical regions, with wide acceptance in south-east Asia and western Africa. The genus currently comprises five species (Sørensen, 1988 & 1996); of these, three are cultivated: *P. ahipa* (Weddell) Parodi, *P. erosus* (L.) Urban and *P. tuberosus* (Lam.)

Sprengel. The other two species, *P. ferrugineus* (Piper) Sørensen and *P. panamensis* Clausen are only found in the wild.

Both cultivated and wild species of *Pachyrhizus*, as well as wild forms of cultivated species are high yielding and exhibit a wide tolerance towards differences in soil fertility, altitude and variations in precipitation rate (Grum *et al.*, 1991 & 1996; Heredia, 1994). In Mexico, for example, the average yields of *P. erosus* obtained in the state of Guanajuato are 60 - 80 t/ha (flood irrigated fields), and similar yields are produced in the lowland area of the state of Nayarit. Fresh tuber yields produced on dry land range from 35 - 60 t/ha. Field trials in Mexico (flood irrigated), Costa Rica (dry land) and Tonga (dry land) exhibited yields of 100 - 145 t/ha (Heredia G., 1994; Heredia Z. and Heredia G., 1994; Morera, 1994; Nielsen, 1995). Furthermore, the performance of 13 landraces observed in Tonga during two consecutive years gave consistent results in the range of 33.1 - 72.0 t/ha (fresh tubers) with a production of 2.72 - 7.46 t/ha dry matter. Similarly, yield sizes of two wild accessions were 24.8 - 51.5 t/ha (fresh tubers) and 3.47 - 5.73 t/ha dry matter.

Phenotypic variation within *Pachyrhizus* is considerable. The genus comprises perennial plants ranging from robust vines with trifoliate leaves to small erect bushes; the growth cycle ranges from less than five months to more than a year (Sørensen, 1996). Landraces can be found with multiple laterally produced tuberous roots, while relatively advanced cultivars possess a single vertically-produced tuberous root (Sørensen, 1996; Sørensen *et al.*, 1997). Tuberous root quality can range from freshly consumed forms with low dry matter content (<10%, as is the case for most of the *ashipa* cultivar group; see section 1.3.3) to forms consumed exclusively when cooked due to a high dry matter content (25 - 30%, i.e. the *chuin* cultivar group; section 1.3.3).

In addition, Sørensen (1996) and Sørensen *et al.* (1997) have recently identified several morphologically distinct landrace/primitive cultivar groups in the genus; their existence is probably the result of geographically isolated domestication processes and man-made selection, as in the case of the various groups of *P. tuberosus* found in the

Amazonian basin, or the red-seeded, white flowered landraces of *P. erosus* with varying degrees of lobed leaflets in the Yucatán peninsula.

1.3. Description of species and cultigen groups

The five species within the genus are morphologically delimited by a series of common characteristics (Sørensen, 1988 & 1996). Among these, the most conspicuous are: vines or semi-erect herbaceous to somewhat lignified perennial plants with one or more tuberous roots; trifoliate leaves with stipules and pinnately arranged leaflets with caducous stipels. The inflorescence is a simple to complex raceme, depending on the species, including flowers with a papilionaceous corolla. The straight legume is septate between seeds, which can be squared, relatively flattened, or rounded kidney-shaped. Seed colour ranges from olive green and deep maroon to black, or from black and white to mottled cream.

1.3.1. The cultivated species: *Pachyrhizus erosus* (L.) Urban

Vernacular names: jícama (Spanish), Mexican yam bean (English), mexikanische Knollenbohne (German), sinkama (Philippines), man kaeo (Thai).

P. erosus is a herbaceous vine with great variation in the outline of its leaflets, from dentate to palmate. The species is defined by its glabrous petals, the number of flowers (4 - 11) per lateral inflorescence axis and an inflorescence length of 8 - 45 cm. Morphological characters specific to the legume of *P. erosus* are size (6 - 13 cm x 8 - 17 mm), reduction of the strigose hairs at maturity and change in colour (from pale brown to olive-green / brown). Furthermore, a number of seed characters are also specific to *P. erosus*, including colour (from olive-green to brown or reddish-brown) and shape (flat and square to rounded, but never reniform). Tuberous roots produce one or more turnip-shaped to elongated tubers (up to 30 cm diameter and 25 cm long; Figure 1.1), with light to dark brown surface, white, whitish yellow to reddish internally (Sørensen, 1988).

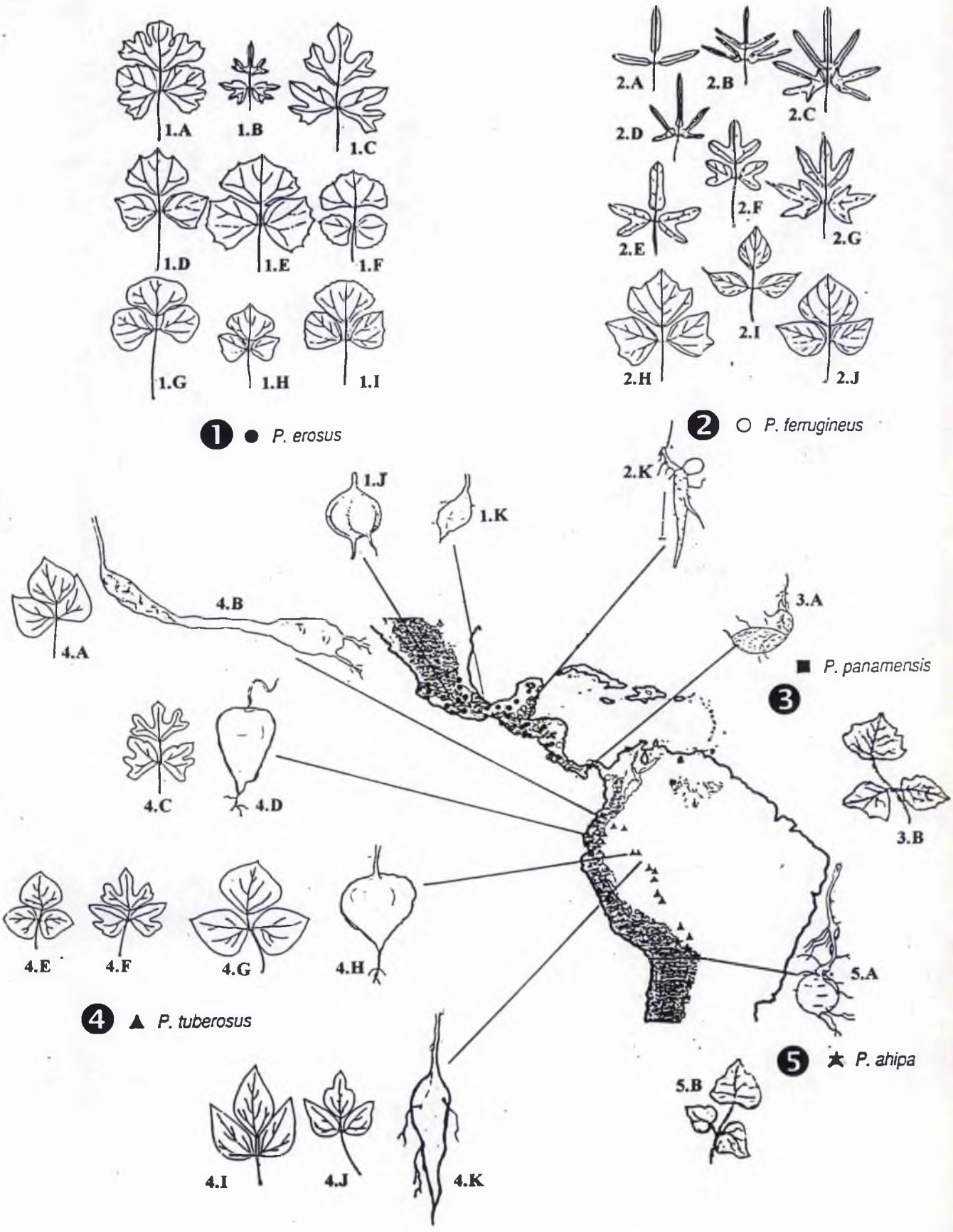


Figure 1.1. Neotropical distribution of phenotypic variation and morphotypes of the five species within the genus *Pachyrhizus* (after Sørensen, 1988 & 1996; see next page for details).

The greatest morphological variation among landraces/cultivars is found in Mexico and Guatemala, where both entire and deeply lobed leaflets are recorded, and tuber shape varies from smooth to lobed or cleft surfaces, with watery-translucent or milky-cloudy juice, a bland or sweet flavour and both thin and thick skin. Flower colour varies from light-dark violet to white. In addition, large differences in earliness exist, so that material with red or maroon-coloured seeds and white flowers, which is often found in the Yucatán peninsula and southern Guatemala, tends to flower later than other genotypes (Sørensen, 1996).

Taxonomic differences between wild and cultivated genotypes of *P. erosus* are difficult to make, due to the vast number of ephemeral populations in Mexico and Central America originating from previous cultivation. Nevertheless, wild material exhibits generally smaller leaf size; increased hairiness of leaves and legumes; smaller, often elongated and irregular shape of the tuberous root; and, a dark brown colour of the tuber surface. Both wild and cultivated genotypes have dehiscent legumes, due to the fact that the crop has never been selected for grain legume characteristics (Sørensen, 1996).

Figure 1.1. Continued. Variation in the outline of leaflets and tuberous roots:

Pachyrhizus erosus: (1.A) Jutiapa, Guatemala; (1.B) Guanacaste, Costa Rica; (1.C) Guanacaste, Costa Rica; (1.D) Cartago, Costa Rica; (1.E to 1.G) Oaxaca, Mexico; (1.H) Yucatán, Mexico; (1.I) Nayarit, Mexico; (1.J) average tuber shape of cultivated accessions (both mono- and multituberous landraces exist); (1.K) tuber from wild population.

P. ferrugineus: (2.A) Belize, Belize; (2.B) Cayo, Belize; (2.C) M. Pine Ridge, Belize; (2.D) Cayo, Belize; (2.E) Stann Creek, Belize; (2.F) Ch. Pine Ridge, Belize; (2.G) Zelaya, Nicaragua; (2.H) Cayo, Belize; (2.I) Alta Verapaz, Guatemala; (2.J) El Paraíso, Honduras; (2.K) tuberised root shape.

P. panamensis: (3.A) tuberised root shape; (3.B) Madden Dam, Panama.

P. tuberosus: Wild populations: (4.A) Los Ríos, Ecuador; (4.B) tuber shape (monotuberous type only). *Jíquima* cultivar group: (4.C) Manabí, Ecuador; (4.D) tuber shape (generally monotuberous). *Ashipa* cultivar group: (4.E) Limoncocha, Ecuador; (4.F) Cusco, Peru; (4.G) San Martín, Peru; (4.H) tuber shape (both mono- and multituberous types exist). *Chuin* cultivar group: (4.I) Loreto, Peru; (4.J) San Antonio, Perú; (4.K) tuberous root (invariably monotuberous).

P. ahipa: (5.A) tuber shape (monotuberous type, multituberous root is rarely found); (5.B) Tarija, Bolivia).

Habitat and present distribution

P. erosus can be found in areas with annual dry seasons in Mexico and Central America, along deciduous forest edges and in scrub vegetation (Figure 1.1). It grows on soil types ranging from deep clay to sandy loam from 0 - 1750 m a.s.l., in areas with mean annual precipitation rates from 250 - 500 mm to over 1500 mm.

The species is widely cultivated in the central and southern states of Mexico (e.g. Nayarit, Guanajuato, Morelos and Veracruz) as well as in Yucatán and Quintana Roo, where presumably pre-Columbian landraces of *P. erosus* have been introduced. Plants are also found as escapes from cultivation in these areas. The same applies to El Salvador and north-western Honduras, where cultivation of the crop is widely practised (M. Grum, pers. comm. in Sørensen, 1996). In Guatemala, limited cultivation is practised today, mainly in the southern states of the country. The plant is often found either as a relic from earlier cultivation or as wild material. This general situation probably also applies in Honduras and Nicaragua, where little or no cultivation is currently practised. Wild materials have been reported in the central and southern states of Mexico, central and western Guatemala, El Salvador, western Honduras, western Nicaragua and north-western Costa Rica (Sørensen, 1988). In general, in the states where a wild distribution is recorded, different cultivars are often found as escapes.

1.3.2. The cultivated species: *Pachyrhizus ahipa* (Wedd.) Parodi

---

Vernacular names: ajipa, ahipa (Spanish); Andean yam bean (English); andine Knollenbohne (German).

The Andean yam bean is distinguished morphologically from the other species by being a herbaceous, erect to semi-erect plant with entire leaflets, very short inflorescences (5- 9 cm) and racemes, and with a general absence of lateral axes (i.e. simple racemes; the number of flowers per lateral raceme, if present, is as low as 2 - 6). The legume is 13 - 17 cm long x 11 - 16 mm wide, and almost circular in cross-

section when immature; its seeds are black, lilac, maroon or black and white/cream mottled, rounded and kidney-shaped. Seeds are never olive-green or red, nor flattened and square (Sørensen *et al.*, 1997). This species is also unique due to the fact that both twining/trailing, semi-erect (indeterminate) to short bushy erect growth habits (determinate genotypes) exist. The tuberous root is turnip-shaped (6 - 8 cm diameter and 10 - 15 cm long) with a greyish-brown colour externally and white internally (Figure 1.1). Both multituberous and monotuberous plants have been identified, though the monotuberous form appears to be dominant (Ørting, 1996; Sørensen, 1997).

Habitat and present distribution

The species is found in sporadic cultivation in Bolivia and the northernmost Argentinean provinces of Jujuy and Salta along the eastern side of the subtropical Andean valleys between 1000 and 3000 m a.s.l. with an annual rainfall ranging from 500 - 1500 mm (Sørensen, 1996). The cultivation of *P. ahipa* in Peru is either not practised or possibly restricted to a few valleys around Tarapoto according to Dr. C. Arbizu (Centro Internacional de la Papa, Peru) (Sørensen,1997). There are no records of plants that are known to be wild; a wild progenitor of *P. ahipa* has yet to be identified and its geographical origin is still unclear. However, Ing. Agr. J. Rea (Centro de Comunicación y Desarrollo Andino, Bolivia) has recently claimed to have found wild germplasm of *P.* ahipa near Sorata, Bolivia. In addition, Dr. D. Debouck (Centro Internacional de Agricultura Tropical, Colombia) mentions other possible locations where a wild progenitor may be found, mainly the Peruvian valleys of Apurimac, Ene and Mantaro, due to favourable climatic and edaphic conditions in those areas. Until true wild material becomes available for morphological and molecular analyses - if such material exists - the phylogeny of the species would be difficult to determine (Sørensen, 1996; Ørting, 1996). Recent studies have also revealed that growth habit follows a geographic distribution from north to south latitudes with the strongest vine-like genotypes occurring in the north (near La Paz, Bolivia) and the smallest, bushy landraces to the south in the Bolivian departments of Chuquisaca and Tarija and the Argentinean provinces of Jujuy and Salta.

1.3.3. The cultivated species: *Pachyrhizus tuberosus* (Lam.) Sprengel

---

Vernacular names: ashipa/namoe/capamu/iwa, chuin, jíquima (Spanish; names applied to the three different cultivars groups found in Ecuador, Peru and Bolivia); potato bean or Amazonian yam bean (English); amazonische Knollenbohne (German); mbacucú (Guaraní). ***NB:*** *Sørensen et al. (1997) provide a comprehensive list of names used by numerous South American ethnic groups.*

This species has probably the broadest intraspecific morphological variation and exhibits the greatest plant size within the genus. Stems of more than 10 m and terminal leaflets of 280 x 260 mm have been recorded. The legumes are also larger than those of the other species (255 x 23 mm) and are conspicuously compressed between seeds. The seeds (12 x 14 mm) are black, black and white mottled or orange-red in colour, kidney-shaped. Both multi- and monotuberous genotypes exist with roots up to 30 cm in diameter and 20 cm or more long (Sørensen, 1988 & 1996).

Recent discoveries of additional cultivar forms (Sørensen *et al.*, 1997) have demonstrated the existence of three different groups of this species, prompting the occurrence of a *P. tuberosus* complex, as follows: **(1)** the typical cultivar form consisting of a herbaceous, strong, climbing vine; **(2)** a newly discovered group of cultivars endemic to areas along Río Ucayali in Peru, represented by herbaceous, basally semi-woody climbing vines reaching less length than the typical cultivar form; and **(3)** the bushy, non-climbing cultivar form currently exclusive to the province of Manabí (western Ecuador), characterised by its determinate growth and deeply lobed lateral leaflets. From this point on, these cultigens will be referred to as *ashipa, chuin* and *jíquima,* respectively (Figure 1.1).

The *ashipa* cultivar group

The prototype plant belonging to this group present in the Amazonian basin and Caribbean Islands is a vine or liana, normally climbing other crops grown in association. It comprises cultivars with long internodes, entire leaflets, long racemes (up to 360 mm), and rounded-reniform seeds.

According to local growers (Sørensen,1996), the *ashipa* cultivar group can be further subdivided into two types: **(1)** the multituberous, comprising two subgroups ('*ashipa* 1' and '*ashipa* 2') and **(2)** a monotuberous type (containing one subgroup only, '*ashipa* 3'). Both '*ashipa* 1' and '*ashipa* 2' produce several large oblong tubers per plant, with the tubers spreading out laterally, the difference between these two types being colour and sweetness of the tuber flesh: '*ashipa* 1' (locally known as *ashipa negro*) has brown peel, white flesh and reduced sweetness, whereas '*ashipa* 2' (*ashipa marrón*) has light brown-yellow peel, yellow-coloured flesh with a much sweeter taste. '*Ashipa* 3' produces a single large turnip-shaped vertical tuber with whitish flesh and little sweetness. All forms have kidney-shaped seeds of an orange-red or brown-red to dull black colour. In addition, both white and violet flowered forms exist (Sørensen *et al.*, 1997).

The *chuin* cultivar group

Chuin (as it is named locally) is a basally semi-woody climbing vine, 2 - 7 m in length, only known from a small area along the Río Ucayali, both upstream and downstream from the town of Requena, Peru. The lateral leaflets are entire acuminated with shallow lobes; the terminal leaflet is rhomboid, occasionally shortly lobed, features never observed in the other *P. tuberosus* groups. This cultigen type differs from the majority of the *ashipas* by: **(1)** it is invariably monotuberous; **(2)** its vertically produced tuber, which resembles a radish or a relatively thickened carrot; **(3)** its particular high dry matter content (which is comparable to that of manioc roots); **(4)** its leaf morphology - the occasionally slightly dentate lateral leaflets have a completely different length:width ratio from those of both the *ashipa* and *jíquima* groups; and, **(5)** the morphology of the legume and seeds (Sørensen *et al.*, 1997).

Similarly to the *ashipa* group, the *chuins* may be further subdivided according to the colour of both the skin and flesh of the tuberous root, as follows: '*chuin* 1' has white skin and flesh (locally known as *chuin blanco*); '*chuin* 2' has yellow skin and flesh (*chuin amarillo*); and '*chuin* 3' has dark purple to violet skin and white flesh (*chuin morado*).

The *jíquima* cultivar group

Unlike the two previous groups, the *jíquima* (pronounced *hee-ki-ma*) is a smaller, non-climbing, determinate, bushy and fast-maturing plant. It comprises cultivars with short internodes, deeply lobed leaflets and racemes interrupted by leaflets with abruptly alternating left- and right-turning growth (Sørensen, 1990). *Jíquima* is grown exclusively in the seasonally dry coastal province of Manabí, western Ecuador, but also very rarely in the province of Los Ríos in the same country.

This cultivar group has been exhaustively collected and evaluated (Grum *et al.*, 1991; INIAP, 1995; Sørensen, 1996) confirming the high level of uniformity in the germplasm from different Ecuadorian localities. The only variability so far recorded is a slight difference in leaf outline and flower colour: the Manabita material has white flowers, while the Los Ríos material has white- and violet-flowered forms. Their growth habit and tuber shape/quality do not vary. The thick turnip-shaped, fleshy tuberous root of the *jíquima* can weigh up to 3 - 4 kg with an average weight of 2 - 3 kg; it is generally monotuberous, but occasionally multituberous with two or three tubers per plant. Their white-cream coloured pulp is very succulent, crisp and tasty. Seeds have always a dull black colour.

The origin of this distinct group within *P. tuberosus* remains uncertain, since there are no other genotypes (wild or cultivated) with a similar growth habit in the surroundings. The nearest materials are the wild populations of *P. tuberosus* in the Ecuadorian protected areas/ genetic reserves Río Palenque (Los Ríos) and La Perla (border between Pichincha and Los Ríos) (pers. obs.), and the wild materials of *P. panamensis* in the province of Guayas. These two are robust vines with inflorescences and legumes of different morphology (Sørensen, 1996).

Wild germplasm of *Pachyrhizus tuberosus*

All wild accessions of this species collected to date originate on the western slopes of the Andes, and no cultivated material has been identified in the immediate vicinities

(INIAP, 1995; Sørensen, 1996). Wild material has a close morphological resemblance to some of the genotypes seen within the *ashipa* cultivar group, i.e. a climbing vine, and to the wild species *P. ferrugineus*. Individuals have evergreen foliage, are all of the monotuberous type, with a very elongated tuber shape and high dry matter content (40 -50%).

In general, wild genotypes are large climbing vines up to 20 m long growing in areas with secondary vegetation (near streams or rivers, often among windfalls or cut down trees). These plants have a long primary, somewhat tuberous root (± 2 m) which continues into thin stringy roots, like a rosary (Figure 1.1). The lower regions of the stems are usually defoliated and tend to become lignified. In contrast to various landraces with non-dehiscent pods, plants from wild populations which reach the upper level of the canopy have dehiscent pods, possibly as an adaptation for enhancing seed dispersal (pers. obs.; Sørensen *et al.*, 1997). The darkish brown coloured pods are prominently pubescent, *P. ferrugineus*-like; seeds are rounded reniform, plane surfaced, with a light-brown colour and 8 - 10 mm long x 9 - 11 mm wide.

Whether these wild populations are in fact true remnants of a west Andean *P. tuberosus* provenance, or whether they represent escaped/introduced material or regressive forms, may be determined only by molecular analyses (see further sections).

Habitat and present distribution (including cultivar groups)

P. tuberosus is widely found in the Amazonian region of South America (Figure 1.1) and appears to be native to the western area of this region (Sørensen, 1988). In general, plants grow in tropical to subtropical evergreen rainforests with an annual rainfall of up to 4100 mm, at altitudes ranging from 0 - 1900 m a.s.l.; they form occasionally dense tangles.

There are some difficulties in determining the extent of the natural distribution of *P. tuberosus* because of its very long history of cultivation in South America. However,

it may be deduced that in isolated areas where little or no variation has been recorded, e.g. pockets of the Amazonian basin, such morphological cultivar uniformity may be the result of a single introduction from neighbouring communities. The present distribution towards the east is not clear, due to encroaching agriculture practised in the Brazilian states of Mato Grosso and Minas Gerais. However, there is little doubt that the species was originally found in the lower reaches of the Amazon, the departments of La Paz and Beni in Bolivia and the north-eastern lowlands of Paraguay along the Río Paraná (L. Ramella, pers. comm., in Sørensen, 1996). To the west, along the Pacific in the semi-arid Ecuadorian province of Manabí, the *jíquimas* are found in cultivation, unfortunately undergoing rapid genetic erosion.

The cultivation of the *ashipa* group takes place in Venezuela, Colombia, Ecuador and Bolivia. It is also grown in Peru by the same ethnic group (i.e. the Cocamas) that grow the *chuins*, and further downstream, near Iquitos. As mentioned before, the *chuins* are distributed along the Río Ucayali, near Iquitos, an area which by far exhibits the greatest diversity among the landraces and cultivars of the *P. tuberosus* complex. *Chuin* is cultivated in a permanently humid climate on flood plains inundated by nutrient-rich rivers.

Finally, wild material has so far been collected only in five localities in the western slopes of the Ecuadorian Andes, growing in disturbed areas, i.e. secondary evergreen rainforest; some of these places are now transformed into pastures or oilpalm (*Elaeis guineensis*), banana (*Musa acuminata* and *M.* x *paradisiaca*) or sugarcane (*Saccharum officinarum)* plantations (Sørensen *et al.*, 1997). It is quite likely, however, that this sample concentration may have been exaggerated by the frequent prospections carried out by INIAP - Ecuador, and RVAU - Denmark during the Yam Bean Project. Hence, further surveys on both slopes of the Ecuadorian Andes and the Amazonian region may yet reveal areas where closely related wild populations are located (Sørensen, 1997).

1.3.4. The wild species *Pachyrhizus ferrugineus* (Piper) Sørensen

This wild species is the only one in the genus which is evergreen (except for the wild accessions of *P. tuberosus*) and perennial above ground (NB: all species have perennial tuberous roots). *P. ferrugineus* comprises semi-woody to woody vines with both tuberous root(s) and stem(s) more lignified than in the other species (Sørensen, 1990 & 1997). The root is less tuberous, although greenhouse-produced plants exhibit tubers of reasonable size (± 0.5 kg); the surface of the root has a dark brown colour, but is whitish brown inside. The morphological variation in leaflet outline is considerable, even within populations, and is greater than in *P. erosus*. The leaflets are subcoriaceous, occasionally relatively pubescent, with reddish-brown strigose hairs. The morphology of the inflorescence differs from the other species in the large number of flowers per lateral raceme (8 -21) and the length of the main raceme axis (≤ 860 mm). The length:width ratio of the legumes is markedly different (80 - 130 mm x 12 - 23 mm). The prominent reddish-brown strigose hairs and the lack of constriction between seeds are also distinguishing pod characters. Seeds are rounded (13 x 13 mm), never reniform, brownish-red and laterally compressed, and are therefore quite distinct from the seed shapes of the remaining species, except for those of wild populations of *P. tuberosus* (Sørensen, 1988 & 1997).

An additional interesting feature of this species is its very slow multiplication rate. Seed production is low, generally less than 100 seeds per plant. Also, the period from germination to maturity of the seeds is the longest within the genus, and in many cases exceeds 10 months (Sørensen, 1996). Most likely, these features are a natural adaptation process to unfavourable habitat conditions.

Habitat and present distribution

The species is ecologically associated with evergreen to deciduous rainforests with soil types ranging from deep clay to coarse sand-clay-loam. It has been recorded from 0 - 1600 m a.s.l. in areas with over 1500 mm rainfall, often in the vicinity of rivers. Its pods appear to possess high resistance to humidity, a useful character when

considering *P. tuberosus* under Amazonian conditions (Sørensen, 1997). *P. ferrugineus* is distributed along the Atlantic coast of Central America from the Mexican state of Veracruz to Panama. From this latter point it spreads to the Colombian department of Chocó on the Pacific coast, as shown in Figure 1.1.

1.3.5. The wild species *Pachyrhizus panamensis* Clausen

The second species only found in the wild, *P. panamensis*is, is also a herbaceous vine distinguished by all parts of the plant being covered by white to light brown pilose hairs, including wing and petals, and a low number of flowers per lateral raceme (4 - 7). The legume is hirsute to sericeous with white hairs, retaining its pubescence at maturity. Seeds are the smallest produced in the genus (6 - 7 mm), rounded to slightly reniform in shape and olive-green in colour. The species has a somewhat elongated root with brown epidermis and greenish white cortex (Sørensen, 1988).

Habitat and present distribution

On the basis of existing records and the very few available accessions, this wild species exhibits a particular disjunct distribution which is typical for most Central American dry forest species, e.g. *Bombacopsis trinitensis*, according to Gentry (1982) and Sørensen (1988). Plants thrive at the edges of deciduous forests, in low shrubbery and open grassy slopes, comprising areas with at least one dry season per year lasting 2 - 3 months. It has been recorded at localities from 0 - 800 m a.s.l. with 250 - 1500 mm rainfall.

P. panamensis exhibits a scattered distribution which extends from central and Pacific Panama into the coastal deciduous forest of Santa Marta, Colombia. It has not been recorded in central Colombia, but is present in the provinces of El Oro and Guayas in the western part of central Ecuador (Figure 1.1). Recent expeditions have confirmed the presence of *P. panamensis* in this latter Ecuadorian province and in the southern area of the state of Bolívar, Venezuela (INIAP, 1996; Sørensen, 1996). Therefore, this species may originally have been distributed relatively uniformly from its

northernmost present localities in Panama to the relatively dry coastal plains in western Ecuador.

1.4. Evolutionary history and biogeography of the genus

The legume family (Fabaceae) is an ancient group in North America dating back at least to the Palaeocene, i.e. 60 Ma, even though fossils to suggest so are scarce, and, when available, are difficult to interpret (Sousa & Delgado, 1993). The group was well diversified by the Eocene; most of the genera are elements that are found today in hot dry seasonal climates, forming part of the lower to medium-sized tropical deciduous forest, although are also found in more humid and warm to cold habitats (Ramamoorthy *et al.*, 1993).

It is possible that the genus *Pachyrhizus* had its origins in the Mexico-Central America area and spread to South America during the Pliocene - Pleistocene (Sousa & Delgado, 1993), implying that *Pachyrhizus* or an ancestor of it may have reached the Mesoamerican area by way of Laurasia. Hence, this hypothesis is contrary to that proposed for genera such a *Dioclea*, *Calopogonium* and their close relatives (Ramamoorthy *et al.*, 1993), which come from the south.

There is good evidence that root and tuber crops were domesticated independently in three regions: **(1)** Southeast Asia and its geographic continuation - the Sunda Islands, Papua New Guinea, Oceania; **(2)** Africa - Madagascar; and, **(3)** Tropical America (León, 1977). Planting of such crops as an agricultural practice was probably first developed in tropical zones. Hawkes (1986) and Harris D. (1973) pointed out that the food-reserve systems of the wild ancestors of such plants must have developed in response to well-marked dry seasons of 5 - 7½ months, otherwise there would have been no reason for such food-reserve systems to have evolved. Seed cultures, according to these authors, would be better suited to areas with more extended dry seasons.

For that reason, the origins of tropical root and tuber crops have to be sought not in the rainforests, where continual humidity allows year-round vegetative growth (and where natural selection would not favour the development of underground storage organs), but in the summer-green rainforests and woodlands with a well-marked dry season (e.g. deciduous regions), where the development of underground starchy food reserves would help a plant survive the dry season and regenerate quickly when rain returns (Sauer cited by Hawkes, 1986).

Such crops, once domesticated, would have been taken into the tropical rainforests at a later stage after the agricultural practices associated with them had become well understood. This implies that agriculture may have come later to the Amazonian basin than to the summer-green rainforests and thorn-scrub areas, where at least some of the lowland root and tuber crops were originally domesticated (Hawkes, 1986).

The history of the yam beans as a plant crop has been recorded from quite an early date, enriched mainly from: **(1)** archaeological data (pottery, plant residues, mummy bundles, embroideries, etc.); **(2)** detailed ethnobotanical references and chronicles during the Spanish Conquest and onwards (including pre- and post-conquest practices and linguistic evidence); and, **(3)** extensive herbarium material. Detailed information of these issues, which goes beyond the scope of this study, can be found in: Montenegro (1740); Urbina (1906); Anonymous (1904); Yacovleff (1933); Yacovleff & Muelle (1933); Yacovleff & Herrera (1934); Herrera (1942); O'Neale & Whitaker (1947); León (1969 & 1977); and, Ugent *et al.* (1986), among others.

1.4.1. *Pachyrhizus erosus*

According to pre- and early post-Columbian references to the cultivation of this species, the 'xicama' (Aztec), 'maen-chicam' (Maya) or 'guyati' (Zapotec) was cultivated by all major civilisations including the Toltec, Olmec, Aztec and Mayan (Martínez, 1979; Yacovleff, 1933). It is known to have been cultivated by the Aztecs in Central Mexico and by the Mayans in the Yucatán Peninsula.

Sørensen (1996) found high uniformity among the cultivars collected in geographically/ climatically/ edaphically isolated areas outside the presumed original distribution of the species (such as the Yucatán peninsula and Central Mexico), suggestive that those cultivars are remnants of ancient ones introduced from southern Mexico, Guatemala and/or from regions further south in Central America, where wild populations exist. The origin of the present landraces and cultivars in Mexico, El Salvador and Guatemala is still unclear, although preliminary molecular analysis (Philips, 1994; Estrella *et al.*, 1998) has indicated different origins of the Mexican and Mesoamerican landraces.

Reports from the 19th century cited by Sørensen (1996) suggest that *P. erosus* only occurred outside the Neotropics after Columbian contacts. The first area of introduction and extensive cultivation was probably the Far East, namely southeast Asia and southern China. The route of introduction is relatively unclear, but there is little doubt that it must have been via old Spanish colonies: from the Philippines to the coastal regions of China, Vietnam, the former French Indochina and Thailand, then India and finally along the west coast of Africa.

Finally, a well documented fact is the return of *P. erosus* to its original distribution area after 300 - 350 years by the French scientist Perrottet, who collected seed material from the Far East for a later 'introduction' to the French Guyana. Thus, it may be possible that some cultivars known from the French Caribbean may have 'travelled' around the world, whereas others may 'only' have crossed over from Mesoamerica.

1.4.2. *Pachyrhizus ahipa*

The recorded history of the Andean yam bean in cultivation indicates that, in contrast to the other two cultivated species, it has never been associated with shifting cultivation. The earliest indications of its use as a crop are remains of tuberous roots found in the 'mummy bundles' of the Paracas Necrópolis (southern Peru), belonging to the Nasca culture (Yacovleff, 1933; Ugent *et al.*, 1986). Embroideries and pottery from the Mochica and Nasca cultures provide further evidence for its cultivation

dating from a pre-Incaic period (200 BC - 600 AD). However, at low altitudes, such representations may have belonged to the *jíquima* cultivar group, due to its similar growth habit (Herrera, 1942; O'Neale & Whitaker, 1947; Sørensen, 1996).

Information contained in the manuscripts and chronicles of Oviedo and Valdéz, at around 1535, confirm pre-Columbian cultivation of the crop. Furthermore, some authors have erroneously identified archaeological records as *P. tuberosus*; nevertheless, the typical growth habit, inflorescence morphology and pods allow a positive identification as *P. ahipa* (Ugent *et al.*, 1986). Sauer (1950) mentions also the crop as one of the common elements of the Andean terraced agriculture in Peru. The crop he was referring to had to be *P. ahipa*, since none of the three cultivar groups within the *P. tuberosus* complex is cultivated at altitudes above 1800 m a.s.l. (Sørensen, 1996 & 1997).

Therefore, substantial evidence exists to confirm that *P. ahipa* was known and cultivated by more than one Amerindian group of the pre-Columbian culture, with a distribution limited to Andean valleys. There are no definite records of the presence of the crop in northern Peru, but due to the proximity of the present Bolivian landraces it may be assumed that such Peruvian material existed/exists.

The Argentinean genotypes collected recently (Ørting *et al.*, 1996) probably originated from seeds introduced from southern Bolivia. Bolivian farm labourers working in Argentina recall importing seed material from Bolivia when visiting relatives (Sørensen, 1990; Ørting *et al.*, 1996); those genotypes belong to the erect bushy type found in Tarija, southern Bolivia.

Germplasm of Bolivian origin exhibits a broad genetic diversity. Almost all landraces/primitive cultivars from the northern departments of La Paz and Cochabamba possess a conspicuous degree of genetic and morphological variation in earliness, growth rate of vegetative and reproductive shoots, and internodal length (Ørting, 1996). In contrast, the single Argentinean landrace/cultivar is very short, has erect bushy growth, reduced seed set and good tuber growth. Germplasm of known

origin most strongly resembling this Argentinean material has been recorded in Tarija and Chuquisaca, southern Bolivia. A comprehensive prospection in southern Bolivia and northern Argentina (Ørting *et al.*, 1996; Ørting, 1996) revealed that all germplasm in this area has determinate growth habit, which distinguishes it from *P. ahipa* landraces with indeterminate growth habit found only in northern areas of Bolivia.

1.4.3. *Pachyrhizus tuberosus*

The plant was already known in Peru for its edible tuberous root in the pre-agricultural period (12200 - 8500 BC), according to a description of León (1987). Probably the first domestication processes took place along the eastern slopes of the Peruvian Andes, at the upper reaches of the Amazonian rivers (Flores Paitán, pers. comm. in Sørensen, 1996). Alternative centres of origin may become evident once additional historic, geographic and molecular information on the remaining distribution areas is available.

The earliest description of *P. tuberosus* and its uses is by Padre José de Anchieta in his *Chartas Inéditas* at the island of São Vicente, Brazil, in 1556 (Arruda & Peckolt cited by Sørensen *et al.*, 1997). Anchieta recorded that the indigenous people cultivated *P. tuberosus* because of the starchy and nourishing tuberous roots, and that the seeds were poisonous. Later descriptions confirm that the plant was quite commonly cultivated in large areas of the humid tropics in South America (Burmann & Pinto cited by Sørensen *et al.*, 1997).

In contrast to *P. ahipa*, *P. tuberosus* has been recorded from the far south-east (the Guaraní communities along the Río Paraná, Paraguay), the northern departments of La Paz and Beni (the Guaraní indians in Bolivia) and as far north as the Andean mountains in Venezuela and the tropics in French Guyana (Sørensen, 1996; Sørensen *et al*, 1997). During colonial times, *P. tuberosus* was well known in Brazil, mainly in the states of Rio de Janeiro, Minas Gerais, São Paulo and Espíritu Santo, where it was used as food for slaves, equivalent to the use of yam (*Dioscorea* spp.).

P. tuberosus is believed to be the only species of the genus introduced to areas outside the continent in pre-Columbian times. Very likely, the Arawak ethnic group in Venezuela, Guyana, Surinam and French Guyana introduced it into the Carribean islands together with other tuberous crops of South American origin (Sørensen, 1996). Introductions have been recorded to the islands of Trinidad, Española (Hispañola), Puerto Rico, Jamaica and Cuba.

There are no definite records of introduction into Central America or tropical areas outside the Neotropics, except the recent ones carried out by the partners of the Yam Bean Project and described elsewhere (Sørensen 1990a, 1991, 1994 & 1995). Finally, seeds of *P. tuberosus* from Trinidad were apparently distributed to the botanic gardens of Calcutta, Ceylon, Brisbane, Melbourne, Sydney and Adelaide (Kew Bulletin 1889 cited by Sørensen, 1996). However, it has not been possible to confirm this information, as no herbarium specimens of this species have been found outside the Neotropics. An explanation could be the misidentification of the material distributed, with the seeds being in fact *P. erosus* and not *P. tuberosus*, as the taxonomic confusion between these two species has a long history (Yacovleff & Herrera, 1935; Sørensen, 1988 & 1996; Sørensen *et al.*, 1997).

No substantiated records exist for the particular cases of the *chuins* and *jíquimas*. *Chuin* is nowadays known only to be grown by people of the Cocama ethnic group, or their descendants, and by the Shipibo Amerindian communities in Amazonian Peru. On the other hand, farmers that grow *jíquima* in Manabí lost their original pre-Columbian identity much earlier. In several cases, tubers are harvested and consumed during special celebrations (e.g. *Corpus Cristi* festivals), showing an interesting combination of Spanish culture and traditional Amazonian/Andean agricultural practices.

Of the three cultivated species, *P. erosus* and *P. ahipa* must be regarded as cultivars principally selected for cultivation at higher or at least drier areas with different complex cropping systems involving at least maize and common bean, due mainly to the uniform appearance of their tubers and the non-twinning habit of *P. ahipa*.

Furthermore, the two species have the longest recorded history of cultivation. The lack of historical records for *P. tuberosus* could also be attributed to the physical conditions in the Amazon and the cultural traditions (Sørensen *et al.*, 1997). This species, in contrast to the previous two, was selected for a different agricultural system where a monotuberous plant does not have any major advantage. The conditions in the Amazon include shade, high humidity and aggressive weed growth (Salick, 1989), and the vast majority of materials of the *P. tuberosus* complex are perfectly suited to these conditions (the exception to this being the uniformly monotuberous *jíquima*, selected for the dry lowlands in Ecuador, and possibly Peru).

When the first plants of *P. tuberosus* were collected for domestication, the selection criteria that resulted in the development of the early original cultivars probably included taste, yield, ease of cultivation, and probably number and shape of tubers per plant in a wide but abrupt geographical area (Salick, 1989; Sørensen *et al.*, 1997). Thus, it is very likely that *P. tuberosus*, like other root and tuber crops, has been domesticated on several occasions at different locations and by various ethnic groups during pre-Columbian times (León, 1977). Such a 'multilocational' domestication is also supported by linguistic evidence, i.e. some 12 different language groups each using independent or non-related vernacular names for the crop (Sørensen *et al.*, 1997).

1.4.4. *Pachyrhizus ferrugineus*

Information regarding the history and biogeography of this species is scarce due to its wild taxonomic status and the low economic and ethnobotanic use. The species is ecologically associated with rainforest lacking a notorious seasonal dry period. In contrast to *P. erosus* and *P. panamensis*, where the stems wilt to ground level during dry seasons, *P. ferrugineus* has no need for this adaptation. This is most likely the reason why the roots are the least tuberous within the genus (Sørensen, 1988 & 1996). Furthermore, its high disease and pest resistance, as well as its strong ecological association with soil types low in available phosphorus, are clear signs of a species of wild taxonomic status.

P. ferrugineus is known from Mexico, Central America and further south in Panama and Colombia. It has been introduced to Cuba, Martinique and Trinidad, where it would appear to have escaped from the botanical gardens to which it was distributed. No records are available of incipient selection nor domestication of this species. The only usage is the one reported by local farmers in Guatemala in which seeds are occasionally used as a vermifuge, probably because of their rotenone content (Sørensen, 1996).

1.4.5. *Pachyrhizus panamensis*

The species may have originally been distributed all the way from its northernmost present distribution area in Panama to the relatively dry coastal plains of Guayas and El Oro, Ecuador. No records are available of incipient selection, domestication or uses of this species.

1.5. Taxonomy and previous systematic studies

The genus *Pachyrhizus* (from the Greek, *pachys* = thick and *rhiza* = root) is placed taxonomically in the legume family (Fabaceae), subfamily Faboideae, tribe Phaseoleae, subtribe Diocleinae and comprises five species, according to Lackey (1977) and Sørensen (1988).

1.5.1. Pre- and post-Linnaean references to the genus

The genus originated in the Neotropics, but there was floristic and taxonomic confusion for a long period due to the early introduction of *P. erosus* to regions in the Palaeotropics, for instance, the Far East (Sørensen, 1988). One of the first botanical references to the species was made by Plukenet in 1696, who described a plant from Mexico as '*Phaseolus nevisensis*', nowadays designated as *P. erosus*. This reference was the basis for the species '*Dolichus erosus*' described by Linnaeus in his Species Plantarum in 1753, stating that this plant originated in the New World, i.e. the

Neotropics. The origin of the species was changed to India with the publication of Linnaeus' Species Plantarum second edition in 1763, this time under the name '*Dolichus bulbosus*'. Du Petit-Thouars published a comprehensive description of six plants listed by Rumphius and included '*Cacara bulbosa*' Rumphius as a synonym of '*Dolichus bulbosus*' in 1806 (Sørensen, 1988).

The generic name that is now the accepted one was originally used by L.C.M. Richard for a herbarium specimen of '*Pachyrhizus angulatus*', an illegitimate species name. De Candolle used the same spelling, with a single '*r*', in his first publication of the name in 1825. Sprengel, in 1826, was the first to introduce the incorrect spelling of *Pachyrrhizus*. Later, when the generic name *Pachyrhizus* was favoured over the 'barbaric' name '*Cacara*', the erroneous spelling '*Pachyrrhizus*' was retained. According to the present botanical code, however, the spelling used by L.C.M. Richard is correct, since it is the original one. Further details regarding pre- and post-Linnaean references to the genus and the origin of the species names can be found in Sørensen (1988, 1990 & 1996).

1.5.2. The genus

Although the genus has been subjected to previous taxonomic revisions, its taxonomy remained relatively confused, especially in regard to the South American species. For instance, this situation applies to Clausen's revision (1945) due to the shortage of herbarium material caused by the Second World War. In addition, the narrow species concept held by this author contributed to the considerable complexity of this work, e.g. made obvious by the vast number of infraspecific taxa included. Sørensen (1988) conducted a new revision of the genus justified by the availability of new material, mainly from European herbaria, and germplasm collected over the past 50 years.

The generic delimitation and taxonomic position accepted in Sørensen's revision (1988) is in agreement with the views by Verdcourt (1970) and Lackey (1981). The genus has a unique structure of the stigma and style, as the short hairs on the adaxial side of the ovary extend almost to the stigma, forming a 'beard' along the incurved

style, and the stigma has a median to subterminal globular process on the adaxial side (Sørensen, 1988). These two characters, together with the tuberous roots, define the genus as of homogeneous identity.

Systematic examinations of the phylogeny and interrelationships of *Pachyrhizus* at the generic and subtribal level have so far been limited to the studies of : **(1)** canavanine and chromosome number by Lackey (1977 & 1980), who placed *Pachyrhizus* in the subtribe Diocleinae; and **(2)** isoflavonoid phytoalexins by Ingham (1979 & 1990), who suggested a close affinity between *Pachyrhizus* and the Palaeotropical genus *Neurautanenia*. According to Ingham (1979 & 1990) this relationship could justify the transfer of *Pachyrhizus* to the subtribe Glycininae, suggesting also that the two genera *Pachyrhizus* and *Calopogonium* (both Diocleinae) may bridge the gap between the genera *Neurautanenia* (Phaseolinae) and *Pueraria* (Glycininae).

A molecular analysis carried out by Bruneau *et al.* (1990) studied the implications of a chloroplast DNA inversion as a subtribal character in the Phaseoleae. In this study, they used three cpDNA probes (ranging from 635 bp to 1.2 kb) and found a 78 kb DNA inversion encompassing most of the large single copy region of the chloroplast genome, which defined a monophyletic group that comprised most genera in this tribe. Six genera (*Calopogonium* Desv., *Canavalia* DC., *Cleobulia* Mart. ex Benth., *Dioclea* Kunth, *Galactia* P. Browne and *Pachyrhizus*) examined within the subtribe Diocleinae lacked the inversion. These results were consistent with morphological data supporting the subtribal classification proposed by Lackey (1981).

Finally, a palynological study of the genus conducted by Sørensen (1989) revealed that the interspecific variation was sufficient to allow doubtless identification to species level. Not surprisingly, the greatest infraspecific variation was detected in the pollen grains from the cultivated species.

1.6. Uses of yam bean

Yam bean is a leguminous plant, but - unlike its distant relatives such as pea, common bean, soybean and peanut - it is mainly grown for its underground parts (National Research Council, 1989). The succulent, flavoursome and crisp tuberous roots of the cultivated species are used as a vegetable, raw or cooked. *P. erosus* and *P. tuberosus* tubers can also be dried and ground to obtain a high quality flour. Immature pods of *P. erosus* can be boiled and used as a vegetable, which is not possible in *P. tuberosus* due to the presence of irritant hairs in the pods (Sørensen, 1996).

1.6.1. An important past use

Archaeological material of yam bean is rare because of ways of consumption and production and also because plant material is prone to rotting in tropical conditions. Interestingly, the Mayan name for yam bean is *chicam*, which according to Patiño (1964) may refer to the act of chewing the root. Other archaeologists have mentioned the consumption of the dried root by pre-Columbian Mexican Indians. In western South America, where conditions for preserving archaeological plant material are better, *P. tuberosus* - or probably *P. ahipa* - have been found in at least four Preceramic sites (Yacovleff, 1933). Clearly, all these records need to be re-checked as to species, but it seems that one or perhaps two of the cultivated species (*P. tuberosus* and *P. erosus*) may have been brought into cultivation during the pre-Inca and Inca kingdoms as long ago as 10 000 BC.

This important past use contrasts with the presence of toxic compounds (rotenones, pachyrrhizin, pachyrrhizone, among others) in almost all parts of the plant except for the tubers (National Research Council, 1989; Sørensen, 1990; Krishnamurti and Seshadri, 1966). Domestication of yam bean, in contrast with that of many other crops (e.g. potatoes and pulses) would thus have induced little changes from the wild ancestors in that nutritional perspective, suggesting either detoxification practices (in order to use pods and seeds) or skilful knowledge about harmless parts of the plant, i.e. the roots, or at certain periods, e.g. unripe pods (Debouck, 1994).

1.6.2. Properties of the genus

Biological nitrogen fixation

Like other members of the legume family, *Pachyrhizus* has an efficient symbiosis with nitrogen-fixing *Rhizobium* and *Bradyrhizobium* bacteria, providing plants with a source of natural fertiliser (Halafihi *et al.*, 1994). In contrast with many grain legumes, a considerable amount of the fixed nitrogen is returned to the soil if the vegetative above-ground parts are left in the field. Thus, the crop forms an integral part of a sustainable land-use system, from both ecological and socioeconomic standpoints (Halafihi *et al.*, 1994; Sørensen, 1996). Tests to quantify the actual amount of nitrogen fixed by symbiosis have been conducted by Castellanos *et al.* (1996) using accessions of *P. ahipa* and *P. erosus*, giving surprising yields in the range of 58 - 215 kg N/ha. Approximately 50% of the N harvested was accumulated in the tuberous roots; in addition, the amount of nitrogen recorded in the residue (hay) ranged from 60 - 150 kg/ha. These values equal or outyield the quantities in practically all grain legumes.

Nutritional aspects

Yam bean has a nutritional composition superior to that found in most non-leguminous root crops. *P. erosus* cultivars can yield a mean crude protein content of 8.3% of the dry weight; similar values have been even recorded for wild accessions (8.0%) and a value of 5.1% has been obtained in *P. ahipa* tubers (Grum *et al.*, 1991). The major carbohydrate present in the mature tuber is starch. Photomicrographs of *P. erosus* starch show it to consist of spherical granules with a mixture of small and comparatively larger granules. The average size of a granule is 3.82 μm in *P. erosus* with a high digestibility score including more than 80% of starch being digested by glycoamylase in 24 hours (Tadera *et al.*, 1984).

P. erosus is a good source of vitamin C (17.7 mg/100 g fresh weight), thiamine (0.06), riboflavin (0.02) and niacin (0.2). The following average composition of *P. erosus*

fresh tubers has been reported: Water (87.0 g/100 g fresh weight), nitrogenous compounds (1.3 g), lipids (0.2 g), non-nitrogenous compounds (7.6 g), fibres (0.7 g), minerals (0.28 g), other components (2.92 g). Similar analyses report the following findings in tubers from field experiments: 9.8% soluble sugars, 28.9% dry matter and 9.5% protein on DM basis in accessions belonging to the *chuin* cultivar group (Sørensen, 1996).

Industrial and other aspects

Provided that an optimal method of preserving the crisp texture of the processed tubers can be developed, yam bean tubers may well be marketed as an attractive product to be used in various dishes, and also as a snack. The latter is in fact one of the common ways of use in Mexico, by which tubers are sliced into sticks and sprinkled with lime juice and chilli, although it can also be used as a vegetable either as fresh tubers (added directly to salad dishes or preserved in vinegar), cooked or stirfried (Sørensen, 1996; Heredia Z. & Heredia G., 1994).

The presence of adenine, choline, rotenone, erosone, pachyrhizid, isoflavonoid phytoalexins and saponin has also been reported (Sørensen, 1995). The figures for the total amount of extractable rotenone in mature seeds vary considerably: a recent analysis of *P. erosus* recorded a range of 0.03 - 0.11% (Lackhan, 1994), although some Chinese landraces may contain as much as 0.5 - 1.0%. Rotenone can be used as a biological pesticide (Halafihi, 1994); however, at present the world market for this use is limited mainly to the production of flea powders, and further studies are needed to identify alternative applications, e.g. as a plant protective agent and in the cleaning of eutrophied lakes (Sørensen, 1996).

Agronomy and potential for crop improvement

Cultivated materials are easily propagated by seed and, except for good manuring of the soil before planting, they require little attention. Plants can also be propagated using small tubers, which greatly reduces the growing time. In Mexico and South

America, floral buds and flowers are pruned to encourage large and sweet roots (National Research Council, 1989; Sørensen, 1996). Furthermore, the tubers are handled, stored and marketed in general terms like potatoes. Apparently, one additional advantage for the farmers is that *Pachyrhizus* functions like cassava (*Manihot esculenta* Crantz) in the tropical lowlands: a food that can be 'stored' in the ground, harvested when needed, and one that is almost unaffected by poor management and environmental conditions.

Almost 10 years ago, a report of an *ad hoc* panel of the National Research Council in USA (1989) stated that yam bean remained a primitive crop with no concerted effort to collect or use germplasm, despite its potential in the tropics. During the last years, the main focus on *Pachyrhizus* has been the development of new attractive cultivars based on interspecific hybrid combinations (Heredia G., 1994; Morera, 1994; Sørensen, 1994). The advantageous agricultural traits present in *P. tuberosus* (e.g. determinate growth in the *jíquima* cultivar group, vigorous growth of the *ashipa* landraces and high dry matter content of the *chuin* cultivars) may all be transferred or combined with traits available in *P. erosus* and *P. ahipa*, in order to create improved varieties with high adaptational qualities to tropical and subtropical areas. More recently, breeding programmes in INIFAP-Mexico have produced high yielding and uniformly shaped tuber varieties such as 'San Miguelito', 'Cristalina', 'Agua Dulce', 'San Juan' and 'Vega de San Juan'. Similar efforts are being made in Costa Rica, Tonga, Denmark and Ecuador. Three options are clearly open for the continued improvement of the crop: **(1)** further exploration and collection of wild and geographically localised landraces, for enrichment of the available genetic basis; **(2)** additional studies on interspecific hybridisation (to date four of the five species, excluding *P. ferrugineus*, have been successfully hybridised); and, **(3)** modern biotechnological methods.

Ethnobotanical and economic data

Several ethnobotanical and anecdotal uses have been described for yam beans as discussed in detail elsewhere (Sørensen, 1990, 1996, 1997; Sørensen *et al.*, 1997). For

example, seeds have been reported to be used for the treatment of human skin diseases; the cure of fevers, pruritis and mange; the control of cattle louse, and several crop pests and diseases (Yang & Tang, 1988; Adjahossou & Sogbenon, 1994; Halafihi, 1994). In interviews with Bolivian farmers (Ørting *et al.*, 1996), it was repeatedly stated that the consumption of *P. ahipa* tubers has a cleansing effect upon the body, is beneficial to the lungs, and cures infections of the air passages. Likewise, shaman Bolívar Santi in Pastaza, Ecuador, mentioned the beneficial effects of *ashipa* tuber consumption, such as improved lactation in breastfeeding mothers and a curative effect on digestive ailments in children (INIAP, 1992 unpublished).

Economic data for yam bean exist for various production areas and markets. Gross incomes ranging from US$ 700 - 2400/ha have been reported in Thailand; in Mexico, the gross income from irrigated fields is US$ 5250 with a marketable yield of 40 t/ha and a net profit of US$ 2500 (A. Heredia, pers. comm. in Sørensen, 1996). In Ecuador, in contrast, there is a decreasing agronomic interest in the crop and tubers are marketed rarely, and used mainly for home consumption or as an occasional present for neighbours. The average price of tubers is US$ 0.20 - 0.30/kg (Sørensen *et al.*, 1997).

1.7. Aims and objectives of this study

The main aim of the research reported in this thesis was to construct a molecular phylogeny of the genus *Pachyrhizus* so as to gain a better understanding of its systematics and past evolution. This was initially achieved by means of a restriction fragment length analysis of chloroplast (cp) DNA variation within and between a wide biogeographic range of *Pachyrhizus* taxa. CpDNA variation was examined both by conventional Southern blotting/probing and by restriction site analysis of cpDNA fragments. The results of these analyses have been compared with those obtained from nuclear variation, namely sequencing of the ITS region. In addition, a survey of RAPD (randomly amplified polymorphic DNA) variation was conducted to aid resolution of the phylogeny within *Pachyrhizus* species, especially within the *P. tuberosus* complex. As a final objective of the current work, it was hoped that the

outcome of these analyses would assist in the interpretation of possible past hybridisation events, the description of unknown cpDNA haplotypes, as well as taxon-specific molecular markers that could be used in the future for reliable identification of problematic taxa and in marker-assisted breeding programmes.

CHAPTER 2

MATERIALS AND METHODS

"Agriculturalists taking part in a fertility ceremony". One of the 26 men figures represented on an *unku* (a sleeveless shirt or tunic) dating from the Early Nasca Period (200 BC - 600 AD).

Reader's left hand: maize (*Zea mays*); *reader's right hand:* yam bean (*Pachyrhizus tuberosus*).

The frequent occurrence of root crops on this garment probably indicates that the inhabitants were heavily dependent upon them for their chief source of carbohydrates (O'Neale & Whitaker, 1947).

2.1. Origin of plant material

Pachyrhizus germplasm used in this study has been generated in the context of the Yam Bean Project (STD Programme, European Union), a multidisciplinary research initiative started in 1982, at the time when the genus *Pachyrhizus* was being taxonomically reviewed for a second occasion (Sørensen, 1988). Since then, several collecting trips to Central and South America have been carried out by the several partners of the Yam Bean Project, the national programmes of plant genetic resources and international centres (CGIAR). To these collections, numerous accessions have been donated by institutions, genebanks, local communities and farmers, so that the total number of accessions of *Pachyrhizus* currently available approximates to 250. This germplasm collection includes both wild and cultivated material of the five species, and is representative of the variability formed within different geographic, ecologic and climatic areas in the Americas.

Leaf material for analysis was made available from 88 accessions of the five *Pachyrhizus* species, and also from single accessions of *Calopogonium caerulum* (Benth.) Sauv., *C. mucunoides* Desv. and *Canavalia ensiformis* (L.) DC. The three latter species were included as outgroups in the phylogenetic analyses, in that they represent taxa that are systematically close, yet distantly enough related to *Pachyrhizus* for rooting phylogenetic trees (Ferguson, 1981; Goldblatt, 1981; Goldblatt, 1981a; Lackey, 1981; Sørensen, 1988). Descriptions and passport data of the accessions examined in this research are presented in Table 2.1.

2.2. Cultivation of plant material

Seeds of most accessions listed in Table 2.1 were sown in a 1:1 compost/vermiculite mixture (*Levington* Medium Structure Compost; Graded Horticultural *Vermiperl*) contained in 8 cm diameter pots. Pots were placed in a glasshouse with temperature maintained at 21 ± 3°C and a 16 h photoperiod supplied by metal halide lamps (*Thorn*™; 400 W). Pots were watered regularly and after germination plants grew

well. Alternatively, leaf samples were supplied by germplasm banks, breeding programmes and universities. Prior to shipment, leaves of each accession were harvested from 5 - 10 healthy plantlets, pooled together, and dried with silicagel at an approximate 1:10 rate, i.e. 1 g of leaf material : 10 g silicagel (Chase & Hills, 1991).

Table 2.1. Accessions of *Pachyrhizus* used in the molecular analyses.

***P. ahipa*:**

Accession number†	Passport data	Molec. Analysis‡
AC102	Bolivia, Tarija, Tarija	4
AC201	Bolivia, La Paz, Luribay, Anquioma (2450 masl)	1, 2, 3, 4
AC202	Bolivia, La Paz, Luribay, Anquioma (2450 masl)	4
AC203	Bolivia, La Paz, Luribay, Asambo (2500 masl)	4
AC204	Bolivia, La Paz, Luribay, Anquioma (2450 masl)	4
AC205	Bolivia, Cochabamba, Machaca, Sanchu Pampa (2250 masl)	4
AC207L	Bolivia, Cochabamba, Machaca, S. Pampa (2200 masl; lilac/violet seeds)	4
AC207S	Bolivia, Cochabamba, Machaca, S. Pampa (2200 masl; black seeds)	4
AC208	Bolivia, Cochabamba, Machaca, Muro Capilla (2800 masl)	1, 3, 4
AC209	Bolivia, La Paz, Tirata, Río Esquina (2200 masl)	4
AC209BR	Bolivia, La Paz, Tirata, Río Esquina (2200 masl; black & white seeds)	4
AC209GS	Bolivia, La Paz, Tirata, Río Esquina (2200 masl; lilac/mixed-colour seeds)	4
AC213	Bolivia, La Paz, Irupana, Cikilini (1750 masl)	4
AC214	Bolivia, La Paz, Arce, Río Chinchico (2900 masl)	4
AC215	Bolivia, La Paz, Arce, Lioja (2450 masl)	1, 3, 4
AC216	Bolivia, La Paz, Arce, Lioja (2450 masl)	4
AC220	Argentina, Jujuy, La Posta, Perico (900 masl)	1, 2, 3, 4
AC222	Bolivia, Tarija, San Lorenzo (2000 masl)	1, 3, 4
AC223	Bolivia, Chuquisaca, Caraparí de Pilaya (1200 masl)	4
AC225	Bolivia, Chuquisaca, C. de Pilaya (1200 masl)	4
AC226	Bolivia, Chuquisaca, C. de Pilaya (1200 masl)	1, 3, 4
AC227	Bolivia, Chuquisaca, C. de Pilaya (1200 masl)	4
AC228	Bolivia, Chuquisaca, C. de Pilaya (1200 masl)	4
AC230	Bolivia, Chuquisaca, Hornillos (1100 masl)	4
AC231	Bolivia, Chuquisaca, Hornillos (1100 masl)	1, 2, 3, 4
AC525	Bolivia, Ayopaya (1900 masl)	4
AC526	Argentina, Salta, Santa Victoria, El Condado (1450 masl)	4

† : Accessions follow the number codes of the Yam Bean Project, i.e. A = *P. ahipa*, E = *P. erosus*, F = *P. ferrugineus*, P = *P. panamensis*, T = *P. tuberosus*; C = cultivated material, and W = wild status.

‡ : Molecular approach used (1 = RFLP variation of total cpDNA; 2 = restriction site variation within a PCR-amplified cpDNA region; 3 = ITS sequencing; 4 = RAPD survey).

Table 2.1. Accessions of *Pachyrhizus*. Continued.

***P. erosus*:**

Accession number†	Passport data	Molec. Analysis‡
EC006	Mexico, Oaxaca (market)	4
EC032	Mexico, Yucatán, Kantunil (100 masl)	1, 2, 3, 4
EC033G	Mexico, Yucatán, Unión Libre (100 masl; light-dark olive green seeds)	4
EC109	Malaysia, Kuala Lumpur (market)	4
EC120	Guatemala, Jutiapa (900 masl)	4
EC201	Mexico, Guanajuato, Celaya (CAEB), *San Miguelito* cultivar (1750 masl)	4
EC205	Mexico, Guanajuato, S. Miguel Ocotopán, *Agua Dulce* cultivar (1740 masl)	4
EC214	Guatemala, Petén, San Andrés (100 masl)	4
EC236	Mexico, Morelos, Jojutla (seed store)	4
EC250	Guatemala, Petén, San Andrés (300 masl)	4
EC502	México, Guanajuato, Celaya (CAEB), *Cristalina* cultivar (1750 masl)	1, 3, 4
EC506	Mexico, Yucatán, Dzan (100 masl)	4
EC509	Costa Rica, Cartago, San Juan (800 masl)	1, 2, 3, 4
EC510	Mexico, Campeche, Los Pueblos (0 - 200 masl)	4
EC511	Mexico, Chiapas, Tapachula (approx. 1000 masl)	4
EC531	Mexico, Oaxaca, Oaxaca	1, 3, 4
EC534	Mexico, Nayarit (from CAEB) (1750 masl)	4
EC558	México, Nayarit, Tuxpán, Coamiles, *Nayarit* type (approx. 15 masl)	1, 2, 3
EC559	Mexico, Nayarit, Santiago Ixcuintla, *Nayarit* type (approx. 14 masl)	4
EC560	México, Nayarit, Ahuacatlán, *Agua Dulce* type (?) (approx. 1000 masl)	4
EC565	Philippines, University of Philippines, Los Baños, College of Agriculture	1, 3, 4
EW051	Costa Rica, Guanacaste, Finca Pacífica (80 masl)	1, 3, 4
EW203	Mexico, Veracruz, between Coatepec - Xalapa, Chavarillo (760 masl)	1, 2, 3, 4
EW223	Costa Rica, Guanacaste, Playa de Cocos (100 masl)	4
EW354	Costa Rica, Guanacaste, Liberia - Bahía Culebra (100 masl)	4
EWHue	Guatemala, Huehuetenango, Nentón - La Democracia (Dr. M. Sørensen)	4
EWPro	Guatemala, El Progreso, Ciudad de Guatemala - Salama (Dr. M. Sørensen)	4

***P. ferrugineus*:**

Accession number†	Passport data	Molec. Analysis‡
FW237	Martinique, Saint Pierre, Botanical Garden, wild (escaped)	1, 2, 3, 4
FWLoc1	Guatemala, Quirigua Ruins	4
FWLoc2	Guatemala, Petén, Cerro Ruso	4
FWLoc7	Guatemala (Pozo del Santo, carretera San Luis - Las Casas)	1, 2, 3, 4
FWGU4	Costa Rica (FWMGCU4, Dr. M. Grum)	4

***P. panamensis*:**

Accession number†	Passport data	Molec. Analysis‡
PW055	Panama, Panama, Maddem Dam	1, 2, 3, 4
PWTM58	Ecuador, Guayas, Guayaquil, Chongón (300 masl)	1, 2, 3, 4
PWTM59	Ecuador, Guayas, Guayaquil, Chongón (300 masl)	4

† : Accessions follow the number codes of the Yam Bean Project, i.e. A = *P. ahipa*, E = *P. erosus*, F = *P. ferrugineus*, P = *P. panamensis*, T = *P. tuberosus*; C = cultivated material, and W = wild status.

‡ : Molecular approach used (1 = RFLP variation of total cpDNA; 2 = restriction site variation within a PCR-amplified cpDNA region; 3 = ITS sequencing; 4 = RAPD survey).

Table 2.1. Accessions of *Pachyrhizus*. Continued.

***P. tuberosus*:**

Accession number†	Passport data	Molec. Analysis‡
TC118	Haiti, Nord Este, Citadelle (ashipa, orange red seeds)	4
TC309	Ecuador, Morona Santiago, San Carlos (900 masl) (ashipa, red-brown seeds)	4
TC350	Peru, Loreto, Jenaro Herrera (chuin morado)	1, 2, 3, 4
TC353	Perú, Loreto, Conta Manillo (chuin amarillo, yellow-orange-brown seeds)	4
TC354	Peru, Loreto, Conta Manillo (chuin blanco, yellow-orange-brown seeds)	1, 2, 3, 4
TC355	Peru, Loreto, Puerto Peru, (chuin morado)	4
TC531	Peru, San Martín (from UNICAMP, Brasil) (ashipa, orange-red seeds)	4
TC532	Bolivia, Beni, Yacuma (218 masl) (ashipa)	4
TC533	Bolivia, La Paz, Iturralde (630 masl) (ashipa, brown seeds)	4
TC536	Brasil, Minas Gerais (from INPA, Manaus) (ashipa, dull black seeds)	1, 2, 3, 4
TC538	Peru, Cusco, Kosnipata (650 masl) (ashipa, dull orange seeds)	4
TC550	Ecuador, Manabí, Sozote (200 masl) (jíquima, dull black seeds)	1, 2, 3, 4
TC552	Ecuador, Manabí, Rocafuerte (48 masl) (jíquima)	4
TC553	Ecuador, Manabí, Santa Ana (50 masl) (jíquima)	1, 2, 3, 4
TC554	Ecuador, Manabí, Rocafuerte (50 masl) (jíquima)	4
TC556	Ecuador, Pastaza, Puyo (960 masl) (ashipa, red-brown seeds)	1, 2, 3, 4
TC557	Ecuador, Pastaza, Diez de Agosto (1100 masl) (ashipa, red-brown seeds)	4
TCNA06	Ecuador, Napo, Archidona, S. P. de Ushpayacu (550 masl)	4
TCNA07	Ecuador, M. Santiago, Macas, playa río Copueno (850 masl)	4
TCNA09	Ecuador, M. Santiago, Gualaquiza, Mercedes Molina (750 masl)	4
TCNA10	Ecuador, Z. Chinchipe, Zumbi (750 masl)	1, 3, 4
TW558	Ecuador, Los Ríos, Buena Fe, Río Palenque (300 masl) (greenish red seeds)	1, 2, 3
TWTM48	Ecuador, Cañar, General Morales, Tomebamba (900 masl)	1, 2, 3, 4
TWNanI	Ecuador, Nanegalito - Nanegal (population I)	4
TWNanII	Ecuador, Nanegalito - Nanegal (population II)	1, 2, 3, 4
TWToaI	Ecuador, Pichincha, Río Toachi (population I)	4

Outgroups:

Accession number†	Passport data	Molec. Analysis‡
OUTcc	Panama (coll. M. Sørensen), no vernacular name (*Calopogonium caeruleum*)	1, 2, 3
OUTce	University of Copenhagen, Botanical Garden, vernacular name: *jack bean* (*Canavalia ensiformis*)	1, 3
OUTcm	Belgium, Meise, Botanical Garden, vernacular name: *calopo* (*Calopogonium mucunoides*)	1, 3

† : Accessions follow the number codes of the Yam Bean Project, i.e. A = *P. ahipa*, E = *P. erosus*, F = *P. ferrugineus*, P = *P. panamensis*, T = *P. tuberosus*; C = cultivated material, and W = wild status.

‡ : Molecular approach used (1 = RFLP variation of total cpDNA; 2 = restriction site variation within a PCR-amplified cpDNA region; 3 = ITS sequencing; 4 = RAPD survey).

In some particular cases, seeds were grown in vermiculite substrate (*Vermiperl*) contained in 35 x 55 cm plastic trays (Optipot, *Karri-Tray*®) which were placed in a *Conviron* growth cabinet. Environmental conditions were programmed as follows:

27°C constant, 58% RH and a 7000 lux/16 h photoperiod. This applied especially for wild and cultivated materials where seed availability was reduced (e.g. wild accessions of *P. tuberosus*) or when plant growth under glasshouse conditions was poor (i.e. for some cultivars within the *P. tuberosus* complex). In all instances, and whenever possible, leaflets were harvested from 6 - 10 week-old plantlets, dried with silicagel in a vacuum dessicator for one week and stored at -20°C prior to DNA extraction.

2.3. DNA extraction and purification

2.3.1. DNA extraction

Numerous problems were encountered during extraction of total genomic DNA from *Pachyrhizus*, including: **(1)** partial or total DNA degradation due to endonucleases, that were present mainly in adult and senescent leaf tissues; **(2)** co-isolation of polysaccharides, sugar and other carbohydrates, which blurred DNA washes with NH_4 acetate and consequently decreased final yields; **(3)** co-isolation of polyphenols, which caused damage to DNA or inhibited the activity of restriction enzymes and polymerase used in this study; and, **(4)** DNA extracts were frequently brown coloured, probably due to the presence of quinonic compounds, i.e. oxidising agents harmful to DNA structure.

To optimise a DNA extraction procedure for *Pachyrhizus* attention was focused on: **(1)** the concentration of mercaptoethanol used as a reducing agent to inhibit oxidation processes; **(2)** speed of centrifugation, as a crucial factor to obtain sufficiently pure DNA preparations; **(3)** use of a method applicable to a broad range of ecotypes and genotypes that differed in biochemical composition of leaf tissues; and, **(4)** cost, speed, ease of use and low hands-on requirements.

Total genomic DNA was extracted using a 2% CTAB method that was scaled to fit within a 2.0 ml Eppendorf tube. The protocol follows the isolation procedure described by Wolff *et al.* (1994) with several modifications and optimised empirically. For each accession, dried leaf material from 5 - 10 individuals was pooled and

approximately 60 - 80 mg of the mixture was flash frozen in liquid nitrogen, ground to a fine powder with a disposable plastic pestle and thoroughly mixed with 1.32 ml of extraction buffer (2% hexa-decyltrimethylammonium bromide, i.e. CTAB; 20 mM EDTA; 1.4 M NaCl; 0.1 M Tris-HCl, pH 8.0; 1% PVP-40T; 0.1% mercaptoethanol) preheated to 60°C. A small amount of sterile alumina was added to aid grinding. Eppendorf tubes containing the homogenate were incubated at 60°C for 30 min and allowed to cool for 5 min.

Half the volume of CTAB, i.e. 0.67 ml of chloroform : iso-amyl alcohol (CI, 24:1) was added to the mixture, vortexed shortly and spun at 8000 rpm for 10 min in a microfuge (Heraeus Instrum.). These steps were repeated once using the uppermost aqueous layer and adding a small amount of TLC-silicagel 60H (Merck). The supernatant was brought into a clean Eppendorf tube and 5 µl RNAse (10 mg/ml, pre-boiled) were added, allowing a 30 min incubation with shaking at room temperature. 1 ml of ice-cold 96% ethanol was added to each sample and then mixed by inverting the tube to a single phase solution to precipitate DNA; samples were left for 1 - 2 h at -20°C before the pellet was spun down at 13000 rpm for 2 min.

DNA was further precipitated and washed by the addition of, first, 76% ethanol/ 0.2 M Na acetate for 50 min, and, second, 76% ethanol / 10 mM NH_4 acetate for 5 - 10 min. Pellets were spun again at 13000 for 3 min, the supernatant discarded, and left to dry at room temperature for 20 min before dissolving in 500 µl TE (10 mM Tris-HCl, pH 7.6; 1 mM EDTA) at 4°C overnight.

2.3.2. DNA purification

Following extraction, DNA was purified using a 7.5 M NH_4 acetate treatment as described by Weising *et al.* (1994). The purified DNA was finally pelleted using ice-cold 96% and 70% ethanol (as detailed above), drained and re-suspended in a final volume of 200 µl TE buffer. DNA was kept at 4°C until further use.

This procedure provided satisfactory yields of high molecular weight DNA (i.e. 150 - 250 µg DNA/g dry leaf tissue) for all *Pachyrhizus* genotypes examined.

2.4. Estimation of DNA concentration

The method used for estimating DNA quantity and quality was based on UV-induced fluorescence emitted by DNA-ethidium bromide complexes (Sambrook *et al.*, 1989; Weising *et al.*, 1994). Known volumes of DNA samples were electrophoresed in 0.5X TBE buffer (45 mM Tris-borate; 1mM EDTA; 27.5 g boric acid to adjust pH to 8.0) using 0.8% agarose gels stained with ethidium bromide; the resulting DNA band was compared to a standard (calf thymus DNA, 125 ng/µl, SIGMA). Intact and high molecular weight genomic DNA was always resolved into one bright band, while DNA seen below this band had suffered degradation. Very low molecular weight material might be RNA, which usually resulted as a smear within the lower part of the gel, and was removed by digestion with RNAse, as described in section 2.3.1.

Gels were photographed under UV light using Polaroid 667 film and/or photodocumented with a *Herolab* gel image analysis system (*Scotlab* EASY STORE software, 1997); values for DNA concentration in the samples were obtained in ng/ul. This procedure allowed DNA quantification and at the same time an estimation of the extent of degradation (quality) and contamination (co-isolated proteins, RNA, etc.) of the extracts.

2.5. RFLP variation of total cpDNA genome

2.5.1. Restriction digestion, running of agarose gels and Southern blotting

Once concentrations had been calculated, DNA of 29 representative *Pachyrhizus* accessions and three outgroups (see Table 2.1) was digested in turn with 22 restriction enzymes (RE). Of these, 15 REs comprising 4- and 6-base recognition sequences cut reliably the DNA of all material studied. Details of the 15 endonucleases successfully employed in the phylogenetic analyses are listed in Table 2.2.

The restriction procedure was as follows: a 500 ng aliquot from each sample was digested with 6 units of RE according to the manufacturer's instructions using acetylated BSA (bovine serum albumin, 10 mg/ml) to enhance RE's activity. Sterile distilled water was also added to a final volume of 30 µl. Digestion of DNA took place in 96-well plates (*Corning*®) overnight at the recommended temperature (Table 2.2) and the reaction was terminated by the addition of 1/5 volume of loading buffer (0.25% bromophenol blue; 60 mM EDTA; 30% glycerol). Plates were stored at 4°C prior to electrophoresis.

Table 2.2. Details of restriction enzymes used in the RFLP analysis of total cpDNA variation.

Restriction enzyme	Recognition sequence†	Incubation temperature (°C)	Heat inactivation‡
Alu I	AG↓CT	37	+
Bam HI	G↓GATCC	37	+
Bgl II	A↓GATCT	37	-
Cfo I	GCG↓C	37	partial
Cla I	AT↓CGAT	37	+
Dde I	C↓TNAG	37	partial
Dra I	TTT↓AAA	37	+
EcoR I	G↓AATTC	37	+
EcoR V	GAT↓ATC	37	+
Hae III	GG↓CC	37	-
Hinf I	G↓ANTC	37	-
Hpa II	C↓CGG	37	-
Sau3A I	↓GATC	37	+
Tru9 I	T↓TAA	65	-
Xba I	T↓CTAGA	37	-

†: the location of cleavage sites is described in a 5'...3' orientation (N = A or C or G or T).
‡: heat inactivation test (+ = 95% inactivation by incubation at 65°C for 15 min; - = no inactivation).

The DNA fragments produced were separated using 1% (for 6-base enzymes) or 1.4% agarose gels (for 4-base cutters) electrophoresed overnight in 0.5X TBE buffer. Ethidium bromide was added to a final concentration of 0.3% and a standard (*Hind* III digests of λ DNA, 10ng/µl; NBL Ltd.) was used to enable calculation of DNA fragment sizes using Templeton's logarithm (software LENGTH, 1988). The separated fragments were transferred from agarose gels on to nylon membranes by Southern blotting (Southern, 1975). The gels were trimmed (i.e. empty lanes and the region

above the wells), photodocumented (as in section 2.4) and immersed in denaturation buffer (1.5 M NaCl; 0.5 NaOH) for 30 min to produce ssDNA fragments. Following denaturation, gels were rinsed twice with distilled water and immersed in neutralisation buffer (1.5 M NaCl; 0.5 M Tris-HCl, pH 7.2; 1 mM EDTA) for another 30 min. Gels were then placed on a blotting apparatus (as illustrated in Nicholl, 1994), and were left overnight to allow DNA transfer onto the nylon membrane using 20X SSC (3 M NaCl; 0.3 M Tri- Na citrate) as the transfer buffer.

The next day, nylon filters were rinsed in 2X SSC, left to air-dry for 5 - 10 min, UV cross-linked for 30 s and baked for 2 h at 80°C in an oven. Membranes were stored between sheets of filter paper at room temperature under dry conditions prior to DNA-DNA hybridisation.

2.5.2. Use of chemiluminescent labelled cpDNA probe

Probe labelling and purification

Chemiluminescent labelled *Vigna radiata* cpDNA probes were produced following a combination of the immunochemical protocols described by the manufacturer (Boehringer Mannheim) and by Hoisington *et al.* (1994). This non-radioactive system uses digoxigenin (DIG), a steroid hapten, coupled to dUTP, UTP or ddUTP to label DNA for hybridisation and subsequent luminescent detection. Thus, a library of heterologous mung bean (MB) probes (described in Table 2.3) were labelled in turn using this procedure in conjunction with anti-DIG/Fab fragments and CSPD, a chemiluminescent substrate which leads to the emission of 477 nm wavelength light after enzymatic dephosphorylation by alkaline phosphatase.

Approximately 120 ng (for the first membrane, adding 60 ng for each additional membrane used) of probe DNA was made up with sterile distilled water to a final volume of 13.5 µl and boiled for 10 min. DNA probe was immediately quenched on ice to prevent it from reannealing before adding the following solutions: 2 µl of *Klenow* 10X buffer (100 mM Tris-HCl, pH 8.5; 100 mM $MgCl_2$), 2 µl hexanucleotide

10X mix (OL; 1 mM Tris-HCl, pH 7.5; 1 mM EDTA; 90 OD units/ml), 2 µl DIG labelling mix (1 mM dATP; 1 mM dCTP; 1 mM dGTP; 0.65 mM dTTP; 0.35 mM DIG-dUTP; alkali-labile, pH 7.5) and 0.5 µl (= 2.5 u) *Klenow* (fragment of *Escherichia coli*, DNA polymerase I).

Table 2.3. Description of the *Vigna radiata* (mung bean) cpDNA probes used.

Probe†	Size (kb)
MB1	16.2
MB2	18.8
MB5	7.5
MB6	1.2
MB7	11.1
MB8	7.0
MB9	5.6
MB10	7.8

†: The vector of all MB probes listed is the plasmid pBR322, which carries genes for tetracycline and ampicilin resistance, and is 4363 bp in length. *Pst* I digests of this plasmid are tetracycline resistant; the bacterial host is probably DH5α, a strain of *Escherichia coli.*

NB: Probes MB3, MB4, MB11 and MB12 were not available at the time of the cpDNA survey.

Labelled cpDNA probe was purified before use by first precipitating it with 2.5 µl of 4 M LiCl and 75 µl of cold 96% ethanol left at -20°C for 2 h before spinning down at 13000 rpm for 15 min and discarding the supernatant. It was then washed with 50 µl of cold 70% ethanol, before spinning down again and discarding the supernatant. Finally, the probe was dissolved in 50 µl TE, incubated for 30 min at 37°C and stored at -20° until required.

Pre-hybridisation and hybridisation treatment

Up to six nylon filters, with a separating mesh in between, were placed in a hybridisation cylinder (*Techne UK*) adding 20 ml for the first membrane and extra aliquots of 2 ml for additional membranes of pre-hybridisation buffer (5X SSC; 0.1% Na-Sarkosyl; 0.02% SDS; 0.5% blocking reagent, Boehringer Mannheim). Pre-hybridisation was carried out for 2 h in a *Techne* hybridiser at 65°C.

Purified probe was then boiled for 5 min and rapidly placed on ice before adding it to pre-warmed (65°C) hybridisation buffer, made up with the same ingredients as the pre-hybridisation buffer. The pre-hybridisation buffer was poured out of the cylinder and replaced with a mixture of hybridisation buffer (10 ml for the first membrane and an extra 1 ml for each additional membrane) and labelled probe. The filters were left to hybridise overnight at 65°C.

Post-hybridisation washes and detection

After hybridisation, membranes were removed from the cylinder and washed first in low stringency washing solution (2X SSC; 0.1% SDS) for 5 min twice. A second wash involved a medium stringency solution (1X SSC; 0.1% SDS) carried out twice for 15 min at 65°C each wash and under constant agitation. Membranes were rinsed in a mixture of buffer 1 (0.1 M maleic acid; 0.15 M NaCl, pH 7.5) and 0.3% Tween 20 before adding 25 ml/membrane buffer 2 (a 1:20 dilution of blocking reagent in buffer 1) and left on an orbital shaker (80 rpm) at room temperature for 30 min. For each membrane, 1.7 µl anti-DIG were added to 25 ml of fresh buffer 2 and left for 30 min at room temperature.

Membranes were washed twice in a mixture of buffer 1 and 0.3% Tween 20 (200 ml/ membrane) for 15 min each wash. Another wash was carried out in buffer 3 (20 ml/ membrane; 0.1 M Tris-HCl, pH 9.5; 0.1 M NaCl; 50 mM $MgCl_2$) for 5 min. Finally, membranes were incubated for 5 min in 25 ml CSPD solution at room temperature and air-dried briefly.

Luminography and filter stripping

Membranes were prepared for luminography by wrapping them separately in plastic hybridisation bags (to prevent drying out) and further incubation at 37°C for 15 min (to enhance the luminescent reaction). Bags were placed in autoradiograph cassettes with X-ray films placed on top and exposed from four up to 24 h at room temperature

before the films were developed. The resulting luminographs were stored in a cool dry place until needed for the evaluation banding patterns.

After luminographs were successfully developed, the probe was stripped off the membranes by washing twice for 15 min at 37°C in a mixture of 0.2 M NaOH and 0.1% SDS. Membranes were finally rinsed in 2X SSC under constant agitation and stored in a sealed plastic bag at -20°C, until required for the next sequential probing.

2.6. Restriction site variation within a PCR amplified cpDNA region

Twenty *Pachyrhizus* accessions and one outgroup were used in this study; plant material subjected to analysis was a subgroup of that examined for total cpDNA variation, with emphasis on accessions of the *P. tuberosus* complex. Details of the taxa assayed are given in Table 2.1.

2.6.1. PCR amplification

Since the introduction of thermostable DNA polymerases in 1988, the use of PCR in research and clinical laboratories has increased tremendously. The method is based on the enzymatic *in vitro* amplification of DNA (Cherfas, 1990) using a very low amount of template DNA (in the ng range). PCR is characterised by its high speed, selectivity and sensitivity (Bachmann, 1994; Newton and Graham, 1994). In order to amplify a particular DNA sequence, two single-stranded oligonucleotide primers are designed, which are complementary to motifs on the template DNA. The primer sequences are chosen so as to allow base-specific binding to the template in reverse orientation. Addition of a thermostable DNA polymerase in a suitable buffer system and cyclic programming of template denaturation, primer annealing and polymerisation steps produces an exponential amplification of the desired sequence between the priming sites (Weising *et al.*, 1994; Hillis *et al.*, 1996).

PCR was used to amplify non-coding regions that separate coding sequences in the cpDNA molecule using a set of universal primers (Taberlet *et al.*, 1991; Demesure *et*

al., 1995). Primers were anchored within the highly conserved tRNA genes and in the large single copy region of cpDNA, which is characterised by a higher substitution rate than the inverted repeat region (Wolfe *et al.*, 1987). The universal primers used are described in Table 2.4.

Table 2.4. Universal primers used for amplifying non-coding sequences of cpDNA (Source: Demesure *et al.*, 1995).

Lab. code	Primer 1	Primer 2	Length (bp)	Anneal. temp. (°C)
HK	*trn*H [TrNA-His (GUC)] 5'-ACGGGAATTGAACCCGCGCA-3'	*trn*K [tRNA-Lys (UUU) exon 1] 5'-CCGACTAGTTCCGGGTTCGA-3'	1690	62.0
KK	*trn*K [tRNA-Lys (UUU) exon1] 5'-GGGTTGCCCGGGACTCGAAC-3'	*trn*K [tRNA-Lys (UUU) exon2] 5'-CAACGGTAGAGTACTCGGCTTTTA-3'	2580	53.5
CD	*trn*C [tRNA-Cys (GCA)] 5'-CCAGTTCAAATCTGGGTGTC-3'	*trn*D [TrnA-Asp (GUC)] 5'-GGGATTGTAGTTCAATTGGT-3'	3000	58.0
CS	*psb*C [psII 44 kd protein] 5'-GGTCGTGACCAAGAAACCAC-3'	*trn*S [tRNA-Ser (UGA)] 5'-GGTTCGAATCCCTCTCTCTC-3'	1680	57.0
ML	*trn*M [tRNA-Met (CAU)] 5'-TGCTTTCATACGGCGGGAGT-3'	*rbc*L [RuBisCO large subunit] 5'-GCTTTAGTCTCTGTTTGTGG-3'	2900	59.0

An aliquot (50 - 150 ng) of purified DNA was used as template for PCR amplification. The PCR reaction mixture also contained 0.4 µl (2 mM) $MgCl_2$, 2.5 µl (100 µM) of each of the four dNTPs (HT Biotechnology Ltd), 0.4 µl (1 pmol/µl) of each primer, 5.0 µl (1X) *Dynazyme* polymerase 10X buffer, 0.7 µl (= 1.4 u) of *Dynazyme* (Flowgen) and sterile distilled water made up to a final volume of 50 µl. The PCR reaction mixture was overlaid with two drops of mineral oil to prevent evaporation. Amplification was achieved in a MJ Research thermocycler programmed as follows: one cycle of 4 min at 94°C; 35 cycles of 45 s at 92°C, 45 s at 53.5 - 62°C (depending of the primers used) and 2 - 4 min at 72°C (depending of the fragment length). A final cycle of 10 min at 72°C was used to complete extension of any remaining products before storing samples at 4°C. A known volume of the PCR product was electrophoresed through a 0.8% TBE agarose gel, stained with ethidium bromide, visualised by UV illumination and compared to a standard (calf thymus DNA, 125 ng/µl, SIGMA).

2.6.2. Restriction digestion and visualisation of fragments

After estimating DNA concentration, PCR products were digested with a battery of restriction enzymes (Table 2.5). Aliquots of 5 - 7 µl (= 500 ng) of PCR product were subjected in turn to restriction digestion with 0.3 µl (= 6 u) of RE, 3 µl 10X restriction enzyme buffer, 0.3 µl BSA (10 mg/ml) and sufficient sterile distilled water to make up a total volume of 30 µl; the mixture was left from 1 h to overnight at 37°C to complete digestion. The reaction was terminated by the addition of 7 µl buffer (0.25% bromophenol blue; 60 mM EDTA; 30% glycerol) and samples were stored at 4°C prior to electrophoresis.

The resulting fragments, together with a standard DNA size marker (1 kb DNA ladder, Gibco BRL), were electrophoresed at 60 V for 1 h and 110 V for 7 h using 2% TBE agarose gels, stained with ethidium bromide. Restriction site variation among taxa was photodocumented (*Scotlab* EASY STORE software), fragment length was evaluated, and presence/absence of a restriction site was interpreted as the character state for phylogenetic analysis.

Table 2.5. Restriction enzymes used to analyse variation within a PCR amplified cpDNA region.

Name	Recognition sequence†	Heat inactivation‡
Cfo I	GCG$^{\downarrow}$C	partial
Hae III	GG$^{\downarrow}$CC	-
Hind III	A$^{\downarrow}$AGCTT	+
Hpa II	C$^{\downarrow}$CGG	-
Hsp92 II	CATG$^{\downarrow}$	+
Rsa I	GT$^{\downarrow}$AC	+

†: The location of cleavage sites is described in a 5'...3' orientation. All enzymes used have functional activity at 37°C.

‡: Heat inactivation test ('+' = 95% inactivation by incubation at 65°C for 15 min; '-' = no inactivation).

2.7. Sequencing of the ITS region of nuclear rDNA

2.7.1. Principles of sequencing methods

Although protein sequencing became routine (albeit costly and labour-intensive) for the study of protein molecular evolution by the late 1950s, nucleic acid sequencing did not become commonplace in studies of molecular systematics until the 1980s. In fact, until the mid-1970s, DNA stretches of only 15 - 20 bp in length had been sequenced. Breakthroughs in nucleic acid sequencing were published almost simultaneously by Maxam and Gilbert (1977) and Sanger *et al.* (1977). DNA sequencing is now widely used in studies of plant systematics and evolution. The main principle behind the sequencing technique is that two ssDNA molecules that differ in length by just a single nucleotide can be separated into distinct bands by electrophoresis in polyacrylamide gels (Brown, 1994).

The sequence of a deoxyribonucleic acid molecule can be elucidated using either chemical or enzymatic techniques, also known as Maxam-Gilbert and Sanger dideoxy sequencing methods, respectively. In this research, the second method was used and is explained briefly as follows.

The enzymatic method of sequencing is based on the ability of a DNA polymerase to extend a primer, hybridised to the template that is to be sequenced, until a chain-terminating nucleotide is incorporated. Each sequence determination is carried out as a set of four separate reactions, each of which contains all four deoxyribonucleotide triphosphates (dNTPs) supplemented with a limiting amount of a different dideoxyribonucleotide triphosphate (ddNTP). Because ddNTPs lack the 3'-OH group necessary for chain elongation, the growing oligonucleotide is terminated selectively at G, A, T or C, depending on the respective dideoxy analogue in the reaction. The relative concentrations of each of the dNTPs and ddNTPs can be adjusted to give a nested set of terminated chains over several hundred to a few thousand bases in length. The resulting fragments, each with a common origin but ending on a different

nucleotide, are separated according to size by high resolution denaturing polyacrylamide gel electrophoresis.

2.7.2. PCR amplification

For comparative purposes, PCR amplification and further sequencing was carried out using DNA templates of the same 29 *Pachyrhizus* accessions studied when analysing total cpDNA variation (section 2.5). Three outgroups were also included in order to determine the direction of change of character-state transformations, and to root the phylogeny. Plant material used in this experiment is listed in Table 2.1.

PCR involved amplification of the internal transcribed region (ITS) of the 18S-26S nuclear ribosomal DNA gene (Figure 2.1). This region has proven to be a useful tool for phylogeny reconstruction in plants by DNA sequencing (Suh *et al.*, 1992; Wojciechowski *et al.*, 1993; Baldwin *et al.*, 1995; Bayer *et al.*, 1996).

The PCR reaction consisted of 10 µl of 10X reaction buffer, 2 µl (2 mM) $MgCl_2$, 10 µl of a 2mM dNTP mixture in equimolar ratio, 2 µl (1 pmol/µl) of each primer, 150 - 200 ng of DNA template, 2 µl (= 4 u) of *Dynazyme* polymerase and sterile distilled water, all in a total volume of 100 µl. The reaction mixture was sealed with a drop of mineral oil (SIGMA) to prevent evaporation during thermal cycling. Amplicons were produced using a MJ Research thermocycler via 46 cycles of denaturation (97°C for 1 min), primer annealing (48°C for 1 min) and extension (72°C for 3 min). A 7-min final extension at 72°C followed cycle 46. Primers used in this PCR reaction were ITS4 and ITS5 (*Cruachem*, UK) as described in White *et al.* (1990) and Baldwin (1992). Primer description is presented in Figure 2.1.

2.7.3. Cleaning the amplified product

PCR products were cleaned using a *Wizard™ PCR Preps DNA Purification System* (Promega) before cycle sequencing, to remove contaminants such as primer-dimers

and amplification primers. Cleaning was carried out according to the manufacturer's instructions.

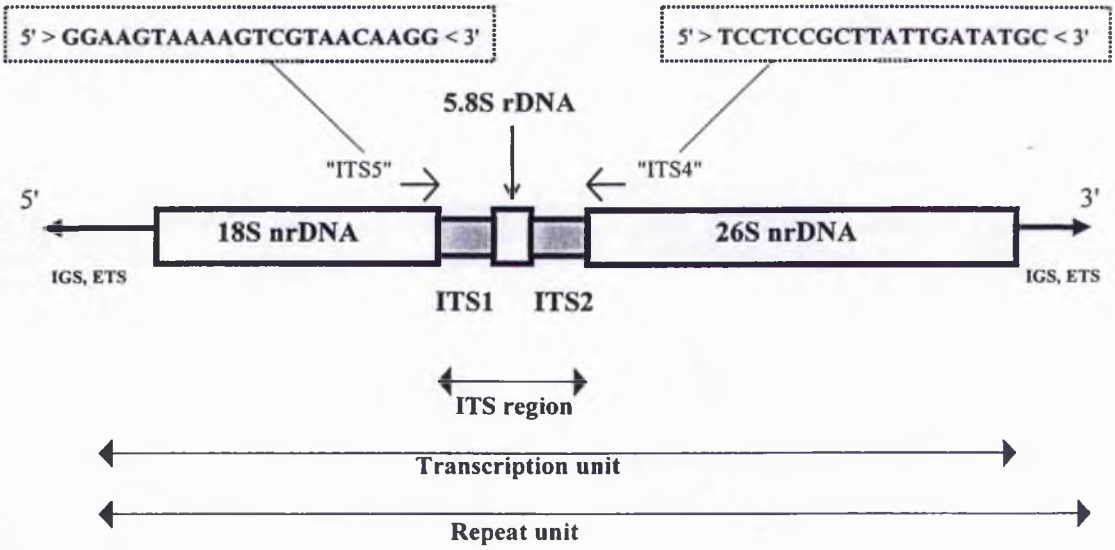


Figure 2.1. Structure of nuclear ribosomal DNA. Positions of the internal transcribed spacer (ITS) regions relative to the 18S, 5.8S and 26S rDNA genes and the intergenic spacer (IGS) are presented. Relative positions of primers ITS4 and ITS5 used in PCR and sequencing are indicated, along with their sequences (after White *et al.*, 1990 & Baldwin, 1992).

The product of each successful PCR reaction, excluding the mineral oil, was transferred to a clean 1.5 ml Eppendorf tube. 100 µl of purification buffer were added to the tube and the components were mixed by vortexing. 1 ml of resin was then added and mixed by vortexing three times over a 1 min period.

A Wizard minicolumn was set up for each PCR product with a 2 ml syringe barrel attached on top and a lidless 1.5 ml Eppendorf tube placed at the bottom. The mixture was pipetted into the syringe barrel and was gently pushed through the minicolumn using a plunger; the syringe was then detached from the minicolumn and the plunger removed. Later, the syringe barrel was reattached to the column and 2 ml 80% isopropanol were pipetted into the syringe and slowly pushed through the column.

Once again, the barrel was removed and a 1.5 ml Eppendorf tube was attached to the bottom of the minicolumn, which were together spun down at 11000 rpm for 20 s to dry the column.

The minicolumn was transferred to a new 1.5 Eppendorf tube and 50 µl of sterile distilled water were added to elute the DNA. The column was left at room temperature for at least 1 min before being centrifuged at 11000 rpm for 20 s, in order to elute the bound DNA fragment. A subsample of purified DNA (5 µl) was electrophoresed on 0.8% TBE agarose gels to estimate its concentration; the remaining DNA was stored at -20°C prior to automated sequencing.

2.7.4. Cycle sequencing and automated sequencing

The sequencing technique used in this research is known as cycle sequencing (Figure 2.2). This technique is based on the dideoxynucleotide chain-termination method of Sanger *et al.* (1977) but utilises a linear polymerase reaction to amplify labelled DNA that is complementary to the target DNA (Murray, 1989; Craxton, 1991). Moreover, sequencing of a double stranded DNA molecule is possible avoiding the need to obtain single stranded DNA as a sequencing template, with the additional advantage that it allows sequencing of a fragment produced directly by PCR (Brown, 1994).

There are a number of types of automated sequencing; the one used in this research was based on Sanger sequencing with flourescent (rather than radioactively) labelled DNA fragments. These fragments were detected during electrophoresis with the use of a tuneable laser. The laser is stationery with respect to the electrophoresis apparatus, and fragments are recorded as they pass a single point. The process is 'automatic' in that no autoradiographs have to be inspected visually nor results recorded manually; instead, the sequence is recorded directly into a computer file in the form of a chromatograph, which may be interpreted (usually by a combination of software programmes) into a DNA sequence for further edition (e.g. inspection of gaps and noise) and alignment.

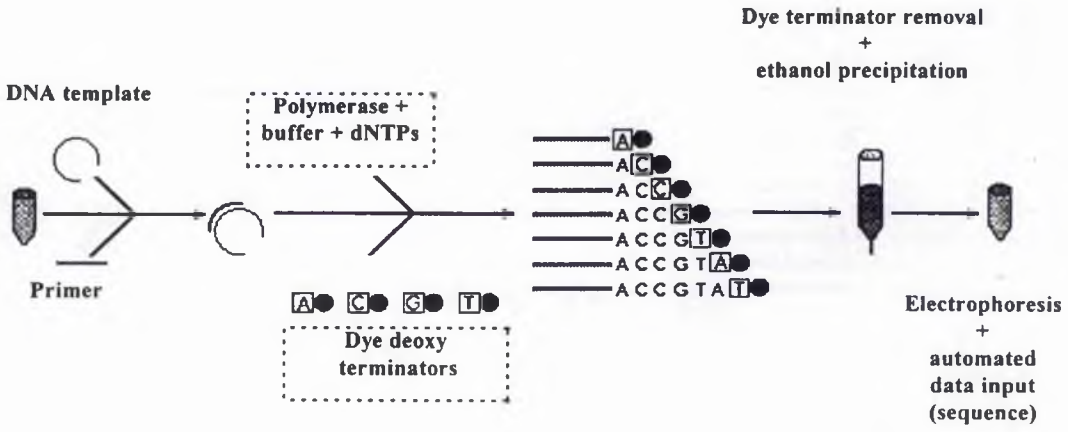


Figure 2.2. Cycle sequencing: A PCR product (DNA template) is mixed with a single primer, dNTPs, fluorescent dye-labelled ddNTPs and a thermostable DNA polymerase. Linear amplification is performed on a thermocycler via 25 cycles of denaturation, annealing and extension, as in standard PCR. Once these cycles are completed, the cycle sequencing product is purified and electrophoresed through polyacrylamide gels; nucleotide sequence is detected with an automated sequencer with input on a computer file (after Hillis *et al.*, 1996; see text for further details).

2.7.5. Automated sequencing chemistry and protocol used

Automated sequencing was conducted using an ABI PRISM™ 377 DNA sequencer which collects data at rates of up to 7.2 kb/h and uses a different fluorescent dye for each extension reaction (i.e. A, C, G and T) so that all four reactions are run in a single lane. Two types of fluorescent sequencing chemistries are currently available: dye primer, where the primer is fluorescently labelled; and, dye terminator, where the dideoxy terminators are labelled. In this research ABI dRhodamine 'Big Dye' terminators were used with *AmpliTaq FS* (a mutant form of *Taq* DNA polymerase) as cycle sequencing enzyme (Figure 2.2).

50 ng of purified PCR product (section 2.7.3) and 5 pmol of primer together in a final volume of 12 µl of sterile distilled water were used for each sequencing reaction. To this, a *Big Dye™ Terminator Ready Reaction Mix* consisting of buffer, enzyme,

dNTPs and labelled dideoxies was added. The reaction was carried out in one tube as each of the four dideoxies is labelled with a different colour fluorescent *Taq*; the mixture went through 25 cycles of denaturation (10 s at 96°C), annealing (5 s at 50°C) and extension (4 min at 60°C) in a thermocycler. The completed PCR reaction was ethanol precipitated using 3M Na acetate (pH 4.6) and 95% ethanol, and left to stand at room temperature for no longer than 10 min.

DNA was spun down at 13000 rpm for 10 min and the supernatant was totally removed. The same procedure was repeated using 70% ethanol and the DNA was finally dried in a vacuum centrifuge for 5 min. Samples were mixed with 3 µl of loading buffer (deionised formamide; blue dextran; 25 mM EDTA, pH 8.0) and loaded on to a 4% polyacrylamide gel (*Long Ranger™*, FMC).

Successful runs rendered two data files per accession (TXT and ABI formats), containing the nucleotide sequence. For each accession, two separate sequencing reactions were produced, one with each primer, i.e. a forward (5'...3') and a reverse (3'...5') sequence, enabling sequencing both complementary DNA strands.

2.8. RAPD analysis

Random Amplified Polymorphic DNAs (RAPDs) are generated by PCR using single, usually decamer, primers that amplify arbitrary fragments of DNA from priming sites throughout the entire genome (Williams *et al.,* 1990). The speed and technical ease with which large numbers of samples can be analysed, and the fact that there is no need for prior knowledge of any sequence information, have made this technique particularly attractive for examining genetic relationships, characterising species and cultivars, and identifying duplicates within germplasm collections. In addition, the RAPD procedure requires only small amounts of DNA and is less costly and labour intensive than other DNA marker methodologies (Rafalsky and Tingey, 1993).

2.8.1. Plant material and RAPD protocol

A total of 85 accessions (Table 2.1) were examined with the RAPD assay for the purposes of identification and assessment of genetic structure and relationships within and between taxa in the genus *Pachyrhizus*. Moreover, a set of 16 primers with G+C contents between 50 - 70% was employed; these included 10-mer arbitrary sequences from kits OPA, OPB and OPH (Operon Technologies, Inc.). Primers are described in Table 2.6. A much fuller description of the RAPD analysis conducted in this study is given in chapter 5.

Table 2.6. Oligonucleotides sequence of the random genomic primers used in the RAPD examinations (supplied by Operon Technologies, Inc.).

Primer name	Sequence (5' - 3')	G+C content (%)
OPA02	TGCCGAGCTG	70
OPA07	GAAACGGGTG	60
OPA10	GTGATCGCAG	60
OPA13	CAGCACCCAC	70
OPB01	GTTTCGCTCC	60
OPB04	GGACTGGAGT	60
OPB07	GGTGACGCAG	70
OPB08	GTCCACACGG	70
OPH01	GGTCGGAGAA	60
OPH02	TCGGACGTGA	50
OPH03	AGACGTCCAC	60
OPH04	GGAAGTCGCC	70
OPH05	AGTCGTCCCC	70
OPH08	GAAACACCCC	60
OPH09	TGTAGCTGGG	60
OPH14	ACCAGGTTGG	60

The RAPD procedure used in this research follows a combination of methodologies described by Weising *et al.* (1994) and Wolff *et al.* (1997), with several modifications and optimised empirically. Amplification conditions of RAPD fragments are as follows:

The 25 µl RAPD reaction mix contained 2.5 µl (12.5 ng) template DNA; 17.8 µl sterile distilled water; 0.2 µl (0.2 pmol/µl) of a single decanucleotide; 1.25 µl (100 µM) dNTPs; 0.5 µl (2mM) $MgCl_2$; 0.25 µl (= 0.5 U) *Dynazyme* polymerase and 2.5 µl reaction 10X buffer provided by the manufacturer. The mixture was overlaid by 40 µl

of mineral oil to prevent evaporation. RAPD reactions were accommodated in disposable 96-well plates placed on a MJ Research thermocycler programmed as follows: one initial cycle of 3 min denaturation at 94°C; 45 cycles of 15 s denaturation at 94°C, 45 s of annealing at 36°C (RAMP 36°C with 0.4°C/sec) and 1.5 min extension at 72°C. A final cycle of 4 min at 72°C was used to complete extension of any remaining products, and samples were stored at 4°C before electrophoresing the final products.

Amplification products were separated in 1.4% agarose gels in 0.5X TBE buffer (60 V for 30 min and 100 V for 3.5 h) and detected by staining with ethidium bromide; a 1 kb ladder (Gibco BRL) was used as a marker. Gels were photographed under UV light with Polaroid film 667 and/or photodocumented with a gel image analysis system (EASY STORE software).

2.8.2. Homology test among comigrating RAPD fragments

The use of RAPDs for comparative purposes relies on the assumption that similarity of fragment size is a dependable indicator of homology (Rieseberg, 1996). To test the validity of this assumption, pairwise comparisons among 212 comigrating fragments produced by the RAPD assays (see chapter 5) were performed as follows:

Selected fragments (see chapter 5 for their complete description) were isolated from the agarose gel in which they were resolved by first removing sections of gel of 0.5 x 0.5 cm containing the putative homologous DNA band. This was done under UV light using a sterilised razor. A small piece of Whatman paper (grade 17) was then folded as a funnel, pre-soaked in TE buffer and placed on top of a bottomless 0.5 ml Eppendorf tube. This funnel-like artifact was placed on top of a 1.5 ml tube and DNA was separated from the solid phase in which it was contained (i.e. agarose, salts) by centrifugation at 8000 rpm for 15 s.

Following separation from the gel matrix, DNA was re-amplified using the same thermocycler programme described above and its respective primer. The resulting

product was divided into three aliquots and digested with two restriction endonucleases that have 4 bp recognition sequences: *Hae* III (GG$^{\downarrow}$CC) and *Hinf* I (G$^{\downarrow}$ANTC). After separation of the digested DNA by electrophoresis, fragments with congruent restriction profiles for both enzymes were considered homologous (Fritsch & Rieseberg, 1992). Where congruent profiles were observed for only one of the two enzymes, the DNAs were tested with a third 4-cutter, *Dde* I (C$^{\downarrow}$TNAG). Identical profiles for two of the three enzymes was considered evidence for homology. Although this approach was successful for most of the fragments tested, several small fragments contained few or no restriction sites and, thus, results were considered ambiguous.

2.9. A preliminary examination of chromosome numbers

Karyotype evolution depends on chromosome mutations, that is, the occasional spontaneous errors of chromosome replication and separation which lead to numerical or structural changes in them. A wide range of mutation types occur, but most are inherently unstable or deleterious at least under the conditions prevailing at the time (Dyer, 1979). Those which survive are of particular evolutionary significance and are widely studied and described.

Much of the phylogenetically useful variation in chromosome number and morphology may be attributable to 'Robertsonian translocations' (fusions and fissions of chromosomes at their centromeres) and two different kinds of inversions (i.e. pericentric inversions, which involve the centromere; and, paracentric inversions, which occur outside the centromeric region; Sessions & Kezer, 1987). Details of these phenomena go beyond the scope of this section and can be found elsewhere (Dyer, 1979; Goldblatt, 1981; White, 1973).

All Papilionoideae so far examined cytologically have most likely had a polyploid ancestry (Goldblatt, 1981). This initial phase of polyploidy is probably very ancient and may have taken place in the late Cretaceous, when major groups of the Fabaceae began differentiating and were probably evolving rapidly into new habitats.

Additionally, an early cycle of polyploidy is evident in many primitive groups of angiosperms (Stebbins, 1970). These events may be correlated with the creation of new habitats following climatic changes that accompanied the opening of the Indian Ocean 150 million years ago, and the separation of Africa and South America which began 130-125 million years ago. The subsequent cytological history seems to involve some descending aneuploidy in every evolutionary line, but it is most pronounced in Papilionoideae, in which predominantly herbaceous genera (e.g. *Pachyrhizus*) have achieved relatively low base numbers.

With some exceptions, the subtribe Diocleinae (to which *Pachyrhizus* belongs) comprises a fairly uniform cytological assemblage, with n = 11 predominant in all genera (Goldblatt, 1981). This basic chromosome number has been confirmed for *Pachyrhizus* in different studies (Lackey, 1980; Goldblatt, 1981; Sørensen, 1988). Similar chromosomal counts have also been reported previously, in which *P. erosus*, *P. ahipa* and *P. tuberosus* also showed n = 11 (Roy, 1933; Clausen, 1945; and, Brücher, 1977, respectively).

In the present study, a methodology for chromosome counting was optimised empirically. Emphasis was placed on the three recently discovered cultigen groups of the *P. tuberosus* complex (Sørensen *et al.*, 1997), since no chromosome counts for these have been reported previously. The technique (see protocol in Appendix 1) yielded reliable squash preparations (i.e. visible chromosomes during microscope observations) in approximately 12 hours; unfortunately, lack of high magnification in the optical systems to hand and the typical small chromosome size in the genus hindered accurate count of stained preparations. It was hoped to achieve a more complete description of the genus by confirming chromosome numbers in these new groups; however, results were considered ambiguous and, thus, are not included in this study.

CHAPTER 3

PHYLOGENETIC RELATIONSHIPS WITHIN *PACHYRHIZUS* SPECIES BASED ON CHLOROPLAST DNA VARIATION

A page from Darwin's Notebook of 1837 showing his tree-diagram of evolution.

"...organised beings represent a tree, irregularly branched." (Darwin, cited by D. Young, 1992).

3.1. Introduction

The past decade has seen a blossoming of molecular biological approaches to the study of angiosperm phylogeny. The two primary sources of molecular variation tapped for phylogenetic purposes have been the chloroplast genome and the nuclear ribosomal DNA repeat region (Clegg & Zurawski, 1992; Baldwin, 1992; Olmstead & Palmer, 1994). The mitochondrial genome in plants has been little used for phylogenetic studies in contrast to animal systematics, where it has played a central role (Avise, 1994).

Zuckerkandl and Pauling (1965) were the first to propose that various proteins and DNA sequences might evolve at constant rates over time, and thereby provide internal biological timepieces for dating past evolutionary events, i.e. the molecular clock concept. Since then, molecular phylogenetics has rapidly developed from the study of protein sequences (e.g. the *rubisco* enzyme; Chan & Wildman, 1972) to the analysis of genomic DNA and individual genes, such as the *rbc*L gene in the chloroplast (Herdenberger *et al.*, 1988). The genetic information generated from these macromolecules has been in turn used to address numerous aspects of the behaviour, life history and evolutionary relationships of organisms.

In this introduction, a description of the structure and function of chloroplast (cp) DNA will be given, and the features that make it a valuable tool for phylogenetic analysis will be summarised. Moreover, a brief review of methods for data analysis and their applications will be presented, emphasising areas where cpDNA has been demonstrated to be useful in phylogeny reconstruction.

3.1.1. Structure of cpDNA

The chloroplast genome varies little in size, structure and gene content among angiosperms (Olmstead and Palmer, 1994). Typically, the chloroplast genome occurs as a closed circular molecule ranging in size from 135 to 160 kilobase pairs (kb) and is characterised by a large, ca. 25 kb, inverted repeat, which divides the remainder of the

genome into one large (LSC) and one small (SSC) single copy region which are approximately 87 kb and 18 kb long, respectively (Figure 3.1; Palmer, 1985; Crawford, 1990; Olmstead & Palmer, 1994). Substantially smaller cpDNA genomes have been documented in which one copy of the inverted repeat is missing (e.g. a group within the Fabaceae and all conifers), and in *Epifagus* (Orobanchaceae) a non-photosynthetic, parasitic plant whose chloroplast genome has suffered massive deletions, including the loss of numerous genes (de Pamphilis & Palmer, 1990). A larger chloroplast genome (217 kb) has been documented in *Pelargonium* (Geraniaceae), but the size increase is due primarily to a greatly expanded inverted repeat without representing an increase in genome complexity (Palmer *et al.*, 1987).

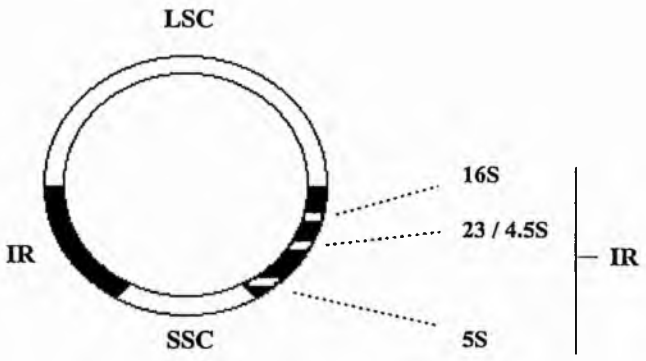


Figure 3.1. Generalised diagram of the chloroplast DNA molecule in higher plants. Shaded areas denote the inverted repeat containing, among others, the genes for rRNA (open areas). LSC: large single-copy region; SSC: small single-copy region; IR: inverted repeat (see text for further details).

The chloroplast, and therefore the DNA contained in this organelle, is suggested to be of symbiotic origin (Palmer *et al.*, 1988). This became apparent due to the clustering of genes of similar function (e.g. the ribosomal RNA genes; Figure 3.1) into polycistronic operons, a distinctive property of prokaryotic genomes. The number of genes present in a typical chloroplast molecule has been estimated at 139 (Shinozaki *et al.*, 1986), 120 (Palmer *et al.*, 1988) and 113 (Olmstead & Palmer, 1994). The genes present include four rRNA genes, 30 - 31 tRNA genes and 79 - 100 putative protein-coding genes mostly of unknown or poorly understood function. These

proteins are mainly involved in photosynthesis and the transcription/translation mechanism within the chloroplast; the large subunit of the enzyme ribulose-1,5-biphosphate carboxylase (*rbc*L for short) is also coded by cpDNA genes.

Changes in gene order and content in the chloroplast genome are rare and typically arise through inversion of a portion of the genome or through the loss of a gene (Olmstead & Palmer, 1994). For example, the differences in gene order and content between *Marchantia*, a nonvascular land plant, and *Nicotiana*, a flowering plant, consist of one 30 kb inversion and the absence in one or the other genome of five protein genes and one tRNA gene (Ohyama *et al.*, 1986; Wolfe & Sharp, 1988). Similarly, the differences between *Oryza*, a monocot, and *Nicotiana*, a dicot, consist of three inversions, the absence of three protein genes, and some gene duplications and rearrangements associated with movement of the ends of the inverted repeat (Sugiura, 1989), representing derived changes in the rice lineage.

3.1.2. Advantages of the use of cpDNA in phylogeny

CpDNA is assumed to be predominantly maternally inherited (in most angiosperms), although biparental (ca. 20% of angiosperm species) and paternal (conifers) inheritance also occurs. However, transmission is essentially clonal as recombination has not been observed in land plants, even in the case of biparental inheritance (Hillis *et al.*, 1996). This means that the molecule is inherited clonally, without crossing over during meiosis and, consequently, cpDNA contains high levels of historical information making it an excellent marker for evolutionary studies (Palmer, 1987).

CpDNA has a very conservative mode of evolution, which would suggest that little intraspecific variation might be expected (Palmer, 1987). This slow rate of change in its sequence and structure is reflected in the low levels of within- and among-population variation apparent from most of the early studies (Wagner *et al.*, 1987; Neale *et al.*, 1988). In addition, the evolution of cpDNA is on average 5-fold slower than that of plant nuclear genes. For instance, the substitution rate for cpDNA protein coding genes varies from 0.2 - 1.0 x 10^{-9} / site / year. Thus, there is a high degree of

conservation in size, structure and gene content of the cpDNA molecule and a change in any of these characters might be expected to have significant phylogenetic implications (Clegg & Zurawski, 1992; Palmer, 1987).

Several studies suggest that the silent substitution rate for plant mitochondrial DNA is only one third of the rate for cpDNA, while the rate for cpDNA is about half that of nuclear DNA (Wolfe *et al.*,1987). The relatively slow rate of nucleotide substitution in cpDNA reduces the problem of parallel and convergent evolution when comparing genomes of congeneric species (Crawford, 1990; Hillis *et al.*, 1996). However, this slow rate may also be a problem causing variation to be inadequate for resolving relationships among closely related species. On the other hand, it is often possible to make valid comparisons among distantly related genera within a large and diverse family (e.g. Asteraceae, Fabaceae), due to the highly conserved nature of cpDNA. Different parts of the cpDNA molecule evolve at different rates, making it feasible to compare taxa at various hierarchical levels by using different regions of the molecule (Crawford, 1990).

The chloroplast molecule is present in many copies per cell (as many as 20 - 200 copies in each mature chloroplast), making it easy to isolate in sufficient quantities from very small amounts of plant material (Palmer, 1987). As few as 10 g fresh weight is adequate for obtaining partially purified cpDNA. Methods employing total DNA require less than 1 g fresh weight (Doyle and Doyle, 1987; Doyle and Dickson, 1987). The cpDNA within individuals appears to represent a homogeneous assemblage, that is, there is normally no evidence of heterogeneity in size or structure of the molecule within a plant (Crawford, 1990).

The small size of the cpDNA molecule makes it possible to visualise on a single gel all the fragments produced by digestion with many of the common four- and six-base restriction enzymes (Palmer, 1987). This is ideal because it often allows a systematist to make at least preliminary comparisons of differences between species and to gain valuable data with minimal time, effort and material. Visualisation of fragments in gels with UV light after staining with ethidium bromide is feasible, however, only if

the cpDNA has been purified. The common approach is the use of Southern blots and filter hybridisations which allow more efficient comparisons (e.g. Palmer *et al.*, 1983; Jansen & Palmer, 1988; Doebley & Wendel, 1989; Van Dijk & Bakx-Schotman, 1997, among others).

The lack of frequent structural changes (inversions, transpositions, deletions and insertions) in the chloroplast genome makes it relatively easy to work in comparative studies. This is true because restriction pattern differences between species usually result from mutations at restriction sites rather than from structural changes. If the latter were frequent it would be difficult and time-consuming to do comparative studies involving a number of taxa (Crawford, 1990).

With regard to studies at higher taxonomic levels (interfamilial and higher), cpDNA data have to be interpreted with care because of accumulated length changes and/or convergent site gains or losses (Crawford, 1990). Such statement, however, has to be determined for each group of plants. Then, an alternative method for studying these higher order relationships can be sequencing of particular parts of the chloroplast genome (Olmstead and Palmer, 1994). Structural changes in the chloroplast genome (especially inversions) have been detected in several large families and tribes of flowering plants. Some of the most completely studied cases are an inversion in the Asteraceae (Jansen and Palmer, 1987) and in the subtribe Phaseolinae of the legume tribe Phaseoleae (Bruneau *et al.*, 1990).

3.1.3. Restriction site mapping of cpDNA: applications and limitations

The analysis of restriction fragment variation has been the main method for the study of cpDNA in phylogeny. The methods have evolved from the direct comparison of gels containing restriction fragments of purified DNA (e.g. Palmer & Zamir, 1982, in their studies of *Lycopersicon*) to the more explicit comparative mapping of restriction sites using Southern blot hybridisation and cloned probes, spanning much or all of a previously mapped genome (e.g. Liu & Furnier, 1993, in *Populus*; Llaca *et al.*, 1994, in *Phaseolus*). A more recent alternative method is based on PCR-amplifying a 2 to

4- kb segment of size conserved, but rapidly evolving (in sequence) cpDNA, digestion with various restriction enzymes and the direct comparison of the resulting fragments on gels. This approach is particularly attractive for groups where the amount of DNA is limiting or where many rearrangements make mapping difficult (e.g. Demesure *et al.*, 1996, in *Fagus*); moreover, the technique is fast, no blotting is required, and it avoids the use of radioisotopes or chemiluminescent labelled probes (Olmstead & Palmer, 1994).

Olmstead and Palmer (1994) present several advantages of restriction site comparison which recommend its use in phylogenetic studies:

- The procedure is technologically simple relative to many methods in molecular biology, enabling the researcher to compare many taxa simultaneously.

- The chloroplast genome is large enough that many sites can be sampled per enzyme; enzymes that cut 100 times per genome can be mapped readily.

- Cleavage sites represent a nearly random sample of the cpDNA, thereby contributing to fulfil an important (yet often unappreciated) assumption of independence of characters. However, Doyle (1992) mentions that the entire chloroplast genome is inherited as a single linkage group, and, therefore, the pattern of its divergence may not reflect species divergence if hybridisation and introgression have resulted in the transfer of cpDNA from one lineage to another.

- Most of the phylogenetically informative variation occurs at restriction sites present in noncoding regions. The value of nonfunctional characters for interpreting phylogeny has long been appreciated.

- Data sets derived from restriction site variation studies of congeneric species or among closely related genera exhibit a remarkably low level of homoplasy (Palmer *et al.*, 1988 and references therein).

- There is sufficient variation in substitution rate in different parts of the genome for applications at a moderate, but limited, range of taxonomic levels. In this sense, recent studies have demonstrated that the inverted repeat region of cpDNA, which accounts for approximately 20% of the total genome complexity of most angiosperms, can be mapped reliably for entire orders and even subclasses. This is mainly due to the significantly lower synonymous substitution rate observed in the inverted repeat, as shown in the previous section.

- Restriction site analysis is relatively free from artefacts due to sample contamination compared with PCR sequencing studies, where so-called "PCR nightmares" (Olmstead & Palmer, 1994) can occur, i.e. inexplicable results during tree construction become obvious, probably from inadvertent amplification of a sequence from the wrong, contaminating DNA.

The analysis of restriction fragment variation of cpDNA has been widely employed in elucidating taxon relationships in phylogenetic studies both at high and low taxonomic levels. At the higher level for example, Schwarzbach and Kadereit (1995) used this approach to investigate relationships between the North American desert genera of the Papaveraceae, subfamilies Platystemonoideae and Papaveroideae. In agreement with earlier results, they found that the first subfamily was nested within the second one and that most taxa under study were characterised by a large number of generic autapomorphies but only a few informative synapomorphies. Furthermore, this was interpreted as strong evidence for a rapid radiation event caused by major climatic changes in the past and accelerated morphological evolution under arid climatic conditions.

At the lower taxonomic level, many studies have examined cpDNA variation to elucidate the phylogeny of species within particular plant genera, e.g. *Leucaena* (Harris *et al.*, 1994), *Erythrina* (Bruneau, 1996) and *Stylosanthes* (Gillies & Abbott, 1996). Liston (1992) examined the *rpo*C1 and *rpo*C2 genes to determine cpDNA divergence and phylogenetic relationships in *Astragalus*; he found low levels of cpDNA divergence among the Californian annual species, being consistent with the

low levels of allozyme divergence detected in his comparative studies. Phylogenetic analysis suggested that the annual growth habit was possibly the ancestral character state in the North American aneuploid legume *Astragalus*.

The origin and evolution of a hybrid species complex in the genus *Brassica* was explored through restriction site analysis of the chloroplast genome (Palmer *et al.*, 1983). In this study, a detailed cpDNA phylogeny enabled identification of the maternal parent for most of the amphidiploids examined and permitted a quantitative resolution of the relative time of hybridisation as well as the relative divergence of the diploid parents. Contradictory chloroplast and nuclear phylogenies obtained for the two amphidiploid accessions of *B. napus* among the 22 accessions used in this study, lead to the hypothesis that introgressive hybridisation had also figured in their recent evolution.

Similarly, cpDNA studies have been of value in determining the origin of polyploid species. For example, Soltis *et al.* (1989) surveyed populations of diploid and autotetraploid plants of *Heuchera micrantha* for variation at cpDNA restriction sites. Fourteen restriction site mutations and three deletions were detected. Four of the mutations were found in single populations while the other 10 occurred in two to six populations. The results showed quite clearly that autotetraploids of this species have arisen several times independently with diploid and tetraploid populations sharing different restriction site mutations. It was argued convincingly that such distribution of cpDNAs was not caused by hybridisation between diploid and tetraploid plants.

CpDNA restriction site variation has also been useful in clarifying colonisation events and the evolution of plant species. For example, Francesco-Ortega *et al.* (1996) demonstrated that the Macaronesian endemic species *Argyranthemum* (Asteraceae) is a monophyletic group that has speciated very recently. One of the two major cpDNA lineages detected is restricted to the northern part of the Macaronesian archipelago (Madeira, Desertas and Selvagens), while the other includes taxa endemic to the southern area of the archipelago (the Canary islands). Two major radiations were recognised within the latter lineage: one of these was restricted to ecosystems

influenced by the north-eastern trade winds, while the other occurred at sites unaffected by these winds.

In summary, cpDNA restriction site analysis can be used in diverse ways with applications that span from answering phylogenetic questions in both wild and cultivated taxa, at extreme taxonomic levels, to the determination of the parentage of hybrids and the study of the evolution of polyploids, introgression and gene flow. However, there are some factors that constrain the usefulness of restriction site analysis in phylogenetic studies, as pointed out by Olmstead and Palmer (1994):

- The conservative nature of cpDNA evolution places a practical lower limit to phylogenetic analysis. Closely related species in many studies are identical for the restriction enzymes used.

- At greater molecular distance an upper limit is also reached, where restriction site homology can no longer be determined with confidence. Molecular divergence may be too great to permit comparative mapping of the whole genome in families that are either old or that have accelerated rates of cpDNA evolution. For example, comparative mapping within and between orders in the subclass Asteridae has generally proven unreliable, at least for single copy regions of the genome (Jansen cited by Olmstead and Palmer, 1994).

- Groups of plants with greatly rearranged cpDNAs (e.g. in the Fabaceae, Geraniaceae and Orobanchaceae) are poor candidates for whole genome mapping studies, although the rearrangements themselves can sometimes be exploited for phylogenetic purposes. Similarly, the absence of one copy of the inverted repeat in several plant groups prohibits their inclusion in comparative mapping studies that are based solely on the inverted repeat, because the rate of evolution for that portion of the genome may be dramatically greater in those taxa. Of course, the problem can be circumvented by restriction site analysis of PCR-amplified, unarranged portions of the genome, as described before.

- A substantial amount of DNA (10 - 100 μg) is needed for digestion of genomic DNA required for whole-genome mapping studies. The tissue required to extract enough DNA may be limiting as in the case of extinct, rare or hard-to-find species, or even poorly preserved. Again, PCR approaches may offer a possible technical solution.

- Scoring and handling of data are still largely manual and cumbersome tasks. Adding new data to an existing data set requires comparing anew all maps, if not the original autorads, from all data sets to include sites that were uninformative in the initial data set. This difficulty has led some researchers simply to assay new accessions for previously identified informative variation, unfortunately eliminating the possibility of identifying formerly undetected relationships based on previously uninformative restriction site variation. This major shortcoming of cpDNA mapping is largely superseded by sequencing studies, in which new sequences can be added easily to an existing data set.

3.1.4. Phylogenetic inference

Several procedures are used in phylogenetic inference of cpDNA as summarised below. Data obtained from the analysis of restriction fragments can be classified as either **(1)** character data and **(2)** distance data. The most frequently used methods for analysing these data are parsimony and distance methods (Swofford & Olsen, 1990). Procedures for testing the reliability of phylogenetic trees include statistical approaches such as bootstrapping and jack-knife methods (Felsenstein, 1988; Swofford *et al.*, 1996).

Types of data

- **Character data:** these are characters that have distinct states, for example a particular band (= the character) on a luminograph (or autorad) produced after hybridisation with a cpDNA probe is either present or absent (= the character state). Character data are assumed to be independently variable and homologous, i.e. all

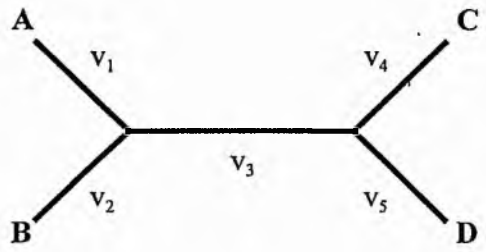
states of a particular character that are observed in a series of taxa are assumed to have been derived from an equivalent state in an ancestral taxon (Swofford & Olsen, 1990). Character data are usually scored as a series of '1's (presence) and '0's (absence) in a binary data matrix (as shown in the results section, this chapter).

- **Distance data:** these specify relationships between pairs of taxa with identical taxa being separated by zero distance. They reflect a measure of the degree of dissimilarity between two taxa or genes which are said to be identical to each other if they are separated by zero dissimilarity (or conversely, 100% similarity).

For tree construction, distance data can be divided into two types, additive distance data and ultrameric distance data. A tree formed from additive distance data is a tree where the evolutionary distance between any pair of taxa would be equal to the sum of the length of the branches connecting them, as shown in Figure 3.2A (Swofford & Olsen, 1990). However, the true topology of an additive tree can only be obtained if no character changes its state more than once (Fitch, 1981), i.e. additivity can only be obtained when there is no homoplasy. If or when the data contain some homoplasy, new calculations will give a different set of values for branch lengths. Conflicting evidence will increase branch lengths and eventually may change branching order. One of the widely used methods to construct phylogenetic trees from additive distance data is the neighbour-joining (NJ) method (Saitou and Nei, 1987). Details of this procedure are described elsewhere.

In contrast, a tree generated from ultrameric data is a tree in which the distance between any two taxa is equal to the sum of the branches joining them (additive distance) and this tree can be rooted so that all taxa are equidistant from the root (Figure 3.2B); thus it is assumed that a molecular clock is operating at the same rate in all lineages. Based on this principle, a tree is constructed by connecting the least distant pair of taxa and then adding successively more distant taxa until all taxa have been joined into the tree. A common method used to construct trees from ultrameric data is the Unweighted Pair Group Mean Analysis (UPGMA; Sneath & Sokal, 1973).

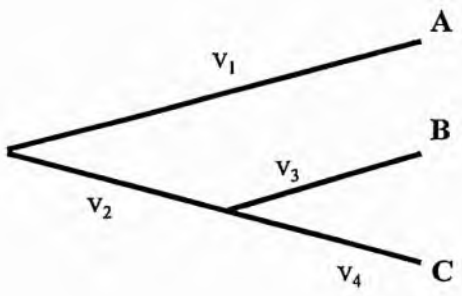
(A)



Additive properties:

$dAB = v_1 + v_2$
$dAC = v_1 + v_3 + v_4$
$dAD = v_1 + v_3 + v_5$
$dBC = v_2 + v_3 + v_4$
$dBD = v_2 + v_3 + v_5$
$dCD = v_4 + v_5$

(B)



Additive properties:

$dAB = v_1 + v_2 + v_3$
$dAC = v_1 + v_2 + v_4$
$dBC = v_3 + v_4$

Ultrameric properties:

$v_3 = v_4$
$v_1 = v_2 + v_3 = v_2 + v_4$

Figure 3.2. Types of tree construction. **(A)** An additive tree showing the relationships among taxa A, B, C and D. Under this scheme of additive distances no assumption about rooting is made, thus the tree is unrooted. dAB equals the distance between taxa A and B. **(B)** An ultrameric tree showing the relationships among taxa A, B and C. In addition to having additive distance properties where all taxon to taxon distances are the total of branch lengths joining them, every common ancestor is the same distance from all of its descendants. For instance, the most recent common ancestor of A and B is v_1 from A and $v_2 + v_3$ from B, therefore $v_1 = v_2 + v_3$; and, the most recent ancestor of B and C is v_3 from B and v_4 from C, therefore $v_3 = v_4$ (Source: Avise, 1994).

On the other hand, restriction data from bands on auto- or luminographs can be analysed with any of the following methods:

For **character analysis** (Bremer, 1991):

- **Fragment direct analysis (FDA):** filters are hybridised with a total cpDNA probe to pick out all the chloroplast bands. Presence and absence of the bands are then scored, with no reference to their homology.

- **Fragment occurrence analysis (FOA):** cpDNA probes from specific regions of the genome are used separately in an attempt to reduce the amount of homoplasy. Presence and absence of fragments from each region are scored and the data kept separate.

- **Site occurrence analysis (SOA):** the locations of site mutations in the cpDNA are mapped from the fragment data and used as characters. Their presence and absence are scored within each taxon.

- **Site mutation analysis (SMA):** loss or gain of restriction sites at particular places in the cpDNA sequence are employed as character states.

For **distance analysis**, both fragment data and site data can be used to calculate genetic distances between pairs of taxa and used to construct phylogenetic trees (Swofford & Olsen, 1990; Avise, 1994).

Phenetic and cladistic approaches to phylogeny reconstruction

Two schools of thought exist as to the approaches for taxonomic classification which differ both in the assumptions they make and the conclusions that can be drawn from the results they produce. These two approaches are phenetics and cladistics and have been described in detail elsewhere, going beyond the scope of this section. Therefore, they are discussed only briefly below.

Phenetic or **numerical taxonomy** can be traced back to Adanson, a contemporary of Linnaeus. However, it was not until the publication of several papers in the late 1950s that phenetics became established as a school of taxonomy, based on the assumption that overall similarity was the key to an accurate classification (Stace, 1989). The aim of phenetics was to overcome intuitive methods which considered that certain characters should be treated as more important in a classification and consequently given more weight than others. Phenetic classification aims to be objective, explicit and repeatable both in the evaluation of taxa and taxon recognition. This is achieved by producing consistent data matrices, examining large character sets and weighting all characters equally (Siebert, 1992).

A phenetic classification requires calculation of a measure of the affinity between taxa and the results are usually presented as a phenogram, i.e. a bifurcating dendrogram. All taxa, known as operational taxonomic units (OTUs) are terminal nodes, i.e. none appear at an interior node, and are clustered together at different hierarchical levels determined by overall percentage of similarity or dissimilarity. Evolutionary relationships between taxa, however, are not examined as no attempt is made to distinguish between homologous and homoplasious characters. If a taxon has accumulated many autapomorphies (uniquely derived evolutionary characters), then it might be clustered further apart from other taxa, simply because of these autapomorphies. Moreover, similarities possibly due to parallel and convergent evolution are ignored (Panchen, 1992). Cain and Harrison (1960) defined phenetic relationships as "an arrangement by overall similarity, based on all available characters without any weighting".

Cladistic or **phylogenetic systematics** was first formulated by Willi Hennig in 1950 based on the idea that in sexually reproducing organisms, diversification in evolution is due to speciation, and speciation either happens or does not. He also considered that most speciation was dichotomous and that sister species shared unique characters (Panchen, 1992). Unlike phenetics, a cladistic approach to classification has no need to look at measurement of genetic distance and a cladogram is representative of the

pattern of speciation events culminating in the species to be classified, i.e. their evolutionary relationships.

The principle works if each past speciation event is detectable in the present complement of characters of each taxon under consideration. A species must have autapomorphies to be taxonomically valid, but it will also have characters that it shares uniquely with the species from which it most recently diverged. These are the synapomorphies (shared, derived characters) that unite sister groups in the cladogram.

Cladistic analysis differs from phenetic analysis in the way that homologous and synapomorphic characters are detected and treated. In phenetics, no consideration is made of synapomorphies, symplesiomorphies (shared, ancestral characters) and whether characters are truly homologous (as a similarity due to common ancestry). Phenetic groupings are defined by all the character states held in common, while cladistic groupings are defined by those characters that are unique to them. Phenetics uses homoplasic characters without even recognising them as such, whereas in a cladistic analysis homoplasy is recognised and, if possible, removed (Panchen, 1992).

In summary, phenetic analysis classifies taxa based on their overall similarity without any attempt to assume evolutionary relationships. In contrast, cladistic analysis provides a description of taxa based on their inferred evolutionary relationships (Stace, 1989). When deciding which approach to use in the analysis of molecular data, it is necessary to understand the nature of the data themselves. For example, the phenetic approach is a reasonable instrument when either the rules of evolutionary change in a character set are not well understood or if the proportion of homoplasic and homologous characters cannot be determined (e.g. a RAPD analysis). On the other extreme, a cladistic approach is reasonable for cpDNA data analysis, as it is possible to determine the mode of cpDNA inheritance in a taxon (Harris & Ingram, 1992; Soltis *et al.*, 1990) and non-homologous characters can be identified and removed (Bremer, 1991).

3.1.5. Objectives and aims of the cpDNA analysis in *Pachyrhizus*

The primary goal of the survey of cpDNA restriction site variation reported here was to produce a phylogenetic tree for *Pachyrhizus* taxa. It was intended that trees would be constructed using both restriction fragment and site data based on an organelle genome for comparison with phylogenetic trees derived from nuclear gene variation (nrDNA and RAPDs), which are described in the next chapters.

A further objective was to compare and contrast the levels of variation resolved by: **(1)** restriction site variation of total cpDNA genome based on hybridisation and probing procedures; and, **(2)** restriction site analysis of PCR-amplified cpDNA regions. The identification of species-specific markers from the restriction fragment patterns was also given priority, since they could be used to identify difficult taxa or to explain evolutionary events.

No extensive research has been previously carried out on *Pachyrhizus* using molecular techniques, such as RFLP analysis of cpDNA. This chapter aims, therefore, to improve our current understanding of the evolution and relationships within *Pachyrhizus* based on an analysis of cpDNA variation.

3.2. Materials and methods

3.2.1. Plant material

For restriction site analysis employing Southern blotting, 29 accessions of the five species within *Pachyrhizus* together with three outgroup taxa were used, whereas for analysis within PCR-amplified regions of the cpDNA genome, a subgroup of these accessions (21 *Pachyrhizus* taxa and outgroup OUTcc) was employed. A complete list of these accessions and their origins are given in Table 2.1. Details of plant growth and care are described in Chapter 2 (sections 2.1 and 2.2).

Plant material used in the analyses comprises a representative range of ecogeographical diversity and variation within the germplasm collection of the Yam Bean Project (Sørensen, 1994). Accessions were chosen based on criteria of previous taxonomic, agronomic and physiological studies (Sørensen, 1990, 1991 & 1996; Grum *et al.*, 1991; Márquez & Morera, 1992). Herbarium specimens for most taxa examined are available at the Royal Veterinary and Agricultural University (RVAU, Botanical Section, KVL - Denmark).

3.2.2. Experimental methods

All experimental methods for the analysis of cpDNA variation were carried out as described in Chapter 2. DNA extraction, purification and quantification from leaf samples were performed as outlined in sections 2.3 and 2.4. Further, experimental conditions for restriction digestion, Southern blotting, hybridisation, probing and production of luminographs were carried out as described in section 2.5. Methods used for PCR-RFLP analysis of cpDNA regions are summarised in section 2.6.

3.2.3. Phylogeny reconstruction using the parsimony approach

Parsimony is probably one of the most popular methods used in cladistic analysis. It can best be described as a criterion for estimating a parameter from observed data

based on the principle of minimising the number of events needed to explain the data (Avise, 1994; Hillis *et al.*, 1996). In phylogenetic analysis, the optimal tree under the parsimony criterion is the one that requires the smallest number of evolutionary (character-state) changes to explain the observed differences among OTUs.

The method is based on the hypothesis that simple character changes are more likely to have occurred than more complicated ones during the evolution of a group of taxa (Swofford, 1993). Furthermore, when there is no reason to think otherwise, two characters that appear to be the same should be treated as homologous. However, if the character is clearly not homologous, i.e. it supports conflicting groups, the explanation that is the simplest should be chosen, which in turn is the one that requires the smallest number of homoplasious characters and character loss (Swofford *et al.*, 1996).

There are five different forms of parsimony which operate by selecting the tree(s) that has the shortest branch length for a particular data set, but differ in their assumptions about character changes:

- **Wagner parsimony:** this method and Fitch parsimony are the simplest methodologies. The Wagner method, formalised by Kluge and Farris (1969), assumes that characters are measured on an interval scale; thus it is appropriate for binary, ordered multistate and continuous characters. Change from one character state (A) to another (D) involves changes through intervening character states (B and C) in the transformation series.

- **Fitch parsimony:** in this approach, transformation from one character state to another can proceed in an unordered fashion, therefore not involving changes through intervening character states (A can become D immediately without going through B and C). Fitch (1966) generalised the method to allow analysis of unordered multistate characters, e.g. nucleotide and protein sequences. Both Fitch and Wagner approaches permit free reversibility; that is, change of character states

in either direction is assumed to be equally probable, and character states may transform freely from one state to another and back again.

- **Dollo parsimony:** in this procedure, each shared derived character state (synapomorphy) is uniquely derived and appears only once in the tree. Character state reversal is allowed, but once a state is reversed it cannot reappear (i.e. parallel or convergent gains of the derived condition are not allowed). DeBry and Slade (1985) and others (Swofford *et al.*, 1996 and references therein) have suggested this parsimony model as more appropriate for restriction site data. The Dollo criterion can be applied to binary or linearly ordered multistate characters for which an ancestral condition (polarity) can reasonably be hypothesised. However, it can also be applied to unrooted trees as well (i.e. an unrooted Dollo approach), which is particularly convenient for restriction-site characters since it does not require the construction of a hypothetical ancestor, only the inclusion of one or more outgroup taxa (for further details see Swofford *et al.*, 1996).

  The assumptions made by both Wagner and Fitch methods are probably unreasonable for restriction site characters, since the loss of an existing restriction site is more probable than a parallel gain of the same site at any particular location, making thus the Dollo criterion more suitable for restriction site analysis. However, if derived states are polyphyletic, even occasionally, Dollo parsimony can render inaccurate phylogenies (Swofford & Olsen, 1990).

- **Camin-Sokal parsimony** (1965)**:** this was the first discrete-character parsimony approach to be described. It makes the strongest assumption of any of the methods discussed so far: evolution is irreversible. It is very unlikely to justify such an assumption for any type of molecular data and, therefore, this procedure is rarely used.

- **Transversion parsimony:** the method is primarily used for nucleotide sequences, assigning greater weight to transversions than transitions, without going so far as to give transitions zero weight (Swofford *et al.*, 1996).

3.2.4. Tree construction from parsimony analysis

Details of tree construction from parsimony analysis have been broadly described and thoroughly reviewed elsewhere (Avise, 1994; Felsenstein, 1988; Hillis *et al.*, 1996; Swofford, 1991; Swofford *et al.*, 1996; Soltis *et al.*, 1990). Thus, these approaches are discussed very briefly below, based on the aforementioned references.

Exhaustive search

In this approach, an initial tree for the first n taxa is constructed and a next $(n + 1)$ taxon is added and evaluated in every topology. Then each additional taxon is added and every single tree topology is evaluated as subsequent taxa are added. The main difficulty with this type of search is that the number of trees increases rapidly with the addition of further taxa. Exhaustive methods are not generally useful for more than 10 or 11 taxa, since they generate over two and 34 million trees for 10 and 11 taxa, respectively (Swofford, 1991 & 1993).

Branch-and-bound search

This is another exact algorithm for the identification of all optimal trees and closely resembles the exhaustive search. It differs in that the length of each tree is not calculated at the time of its construction, thus considerably reducing computing time. In addition, this approach employs a search procedure which has a provision for discarding trees without evaluating them in detail.

Several factors influence the running time of the branch-and-bound algorithm, with quality of the data being perhaps the most important one. Large data sets with little homoplasy will run quickly because most paths of the search tree are terminated early. The speed with which the length of each tree can be evaluated, a function of the character types, is also important. For example, ordered (Wagner) characters are much faster than unordered characters. Finally, for obvious reasons, the speed of the available computer is also critical to the running time (Swofford, 1993).

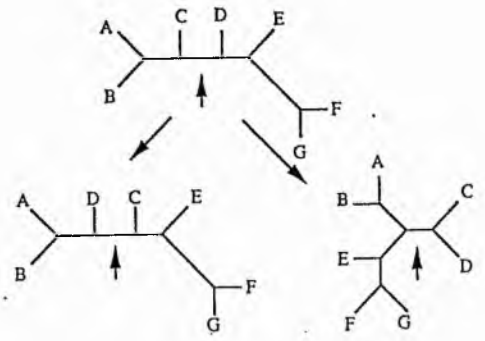
Heuristic methods

When a data set is too large to permit the use of exact methods, a heuristic approach is recommended which operates by sacrificing the guarantee of optimality in favour of reduced computing time. The search begins by surveying only a small sample of all possible tree topologies and the optimal tree is the shortest of these. A tree is then constructed and rearranged so as to bring it closer to the optimum; once no further alterations can improve the tree, the analysis is terminated. Heuristic methods have proven to be very effective by using this principle. Moreover, two basic strategies can be used:

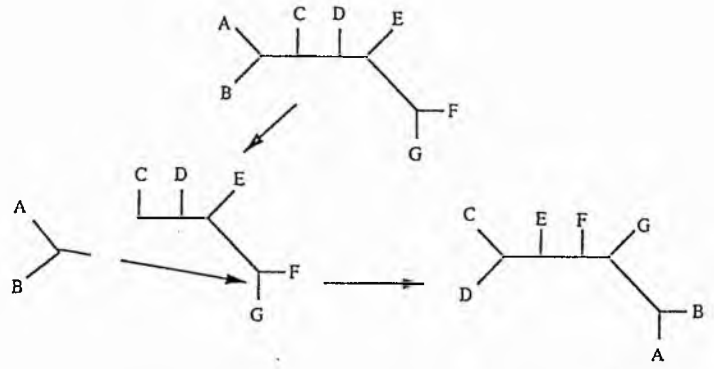
- **Stepwise addition:** this is the common method for obtaining a starting point for further rearrangement of additional taxa to a growing tree. Three taxa are chosen for the initial tree, then one of the unplaced taxa is selected for next addition. The trees resulting from the addition of a fourth taxon are evaluated and the one with the optimal score is saved for the next 'round'. Next, a fifth taxon is placed along one of the five possible branches on the tree saved from the previous round. The evaluation procedure is repeated with the best tree saved for the next round, and this process is concluded when all taxa have been joined to the growing tree.

- **Branch swapping:** in this strategy, the initial estimate provided by stepwise addition is subjected to a series of predefined rearrangements until the shortest tree is found. These rearrangements are performed until the tree cannot be improved any further, which in turn is assumed to be the optimum.

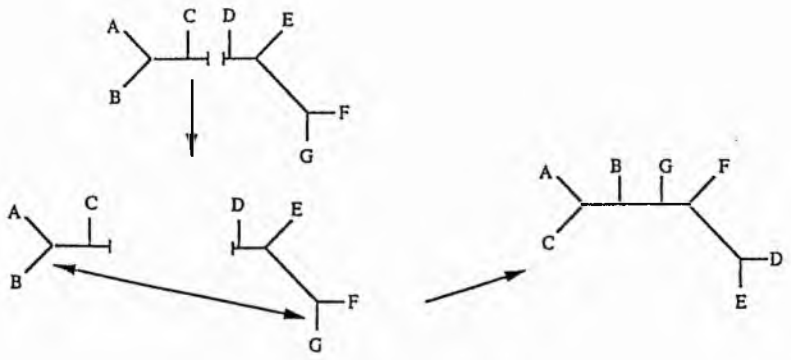
Figure 3.3 illustrates the stages completed in a heuristic approach and the three branch swapping algorithms as implemented in the software programme PAUP (Phylogenetic Analysis Using Parsimony; Swofford, 1993): the nearest neighbour interchanges (NNI); subtree pruning and regrafting (SPR); and, tree bisection and reconnection (TBR). A brief description of these algorithms is also shown in the same figure.



(A) Nearest neighbour interchanges (NNI). Each interior branch of the tree defines a local region of four subtrees connected by the interior branch. Interchanging a subtree on one side of the branch with one from the other constitutes an NNI. Two such rearrangements are possible for each interior branch.



(B) Sub-tree pruning and regrafting (SPR). A subtree is pruned from the tree (e.g. the subtree containing terminal nodes A and B as indicated). The subtree is then regrafted to a different location on the tree. All possible subtree removals and reattachment points are evaluated.



(C) Tree bisection and reconnection (TBR). The tree is bisected along a branch, yielding two disjoined subtrees. The subtrees are then reconnected by joining a pair of branches, one from each subtree. All possible bisections and pairwise reconnections are evaluated.

Figure 3.3. Schematic representation of the heuristic parsimony searches. The three different approaches to branch swapping are described (after Swofford *et al.*, 1996).

Outgroup comparison

An important concept among the optimality criteria in parsimony methods is the use of an outgroup. An outgroup is any taxon used in phylogenetic analysis that is assumed to be outside the group of taxa under study (Swofford *et al.*, 1996). Incorporation of an outgroup is useful for assigning the direction of change to character-state transformations and for determining the root of a phylogenetic tree.

An outgroup is often chosen as a sister group bearing in mind that it is genealogically most closely related to the remaining taxa (i.e. the ingroup), but must not be the ancestor of the ingroup. Swofford *et al.* (1996) emphasise, however, that the assignment of taxa to the outgroup constitutes an automatic assumption that the remaining taxa are monophyletic, an assumption that should be justified by evidence extrinsic to the phylogenetic data at hand. If this assumption is wrong, the tree will be rooted incorrectly.

Consensus trees

A consensus tree is a hierarchical summary of all relationships described by the equally parsimonious trees produced after use of an algorithm, as the ones previously described. However, a consensus tree does not necessarily give the best estimate of phylogenetic relationships among groups; it only summarises them and thus must be interpreted with caution. A large number of polytomies (i.e. unresolved regions in the phylogenetic tree) may become evident in the consensus tree when there is much disagreement among the rival trees it is summarising (Baum, 1992; Swofford *et al.*, 1996). In general, a consensus tree is longer than the minimal trees it describes, since the consensus is less resolved than any of the minimal trees.

There are different types of consensus trees (Figure 3.4), with the strict consensus, semi-strict consensus and 50% majority rule trees as the most widely used (Swofford, 1993). A strict consensus tree describes species groupings (clades) that are present in all rival trees. Therefore, it is the most conservative consensus and the easiest to

interpret, but unfortunately it may be too strict and result in a completely unresolved consensus for trees that only differ by the placement of one taxon.

The semi-strict consensus tree can be best explained by the example given in Figure 3.4. Trees having an ABC trichotomy or an (AB)C dichotomy, with A and B always together, will result in a semi-strict consensus where AB will be retained. In this circumstance, a semi-strict consensus will conserve the AB relationship while a strict consensus will not.

In contrast, the majority rule consensus defines groups that appear in a predefined percentage of the rival trees it describes (50% in this case). In turn, this means that a clade may be retained even if some of the trees do not resolve it (Swofford, 1993; Figure 3.4).

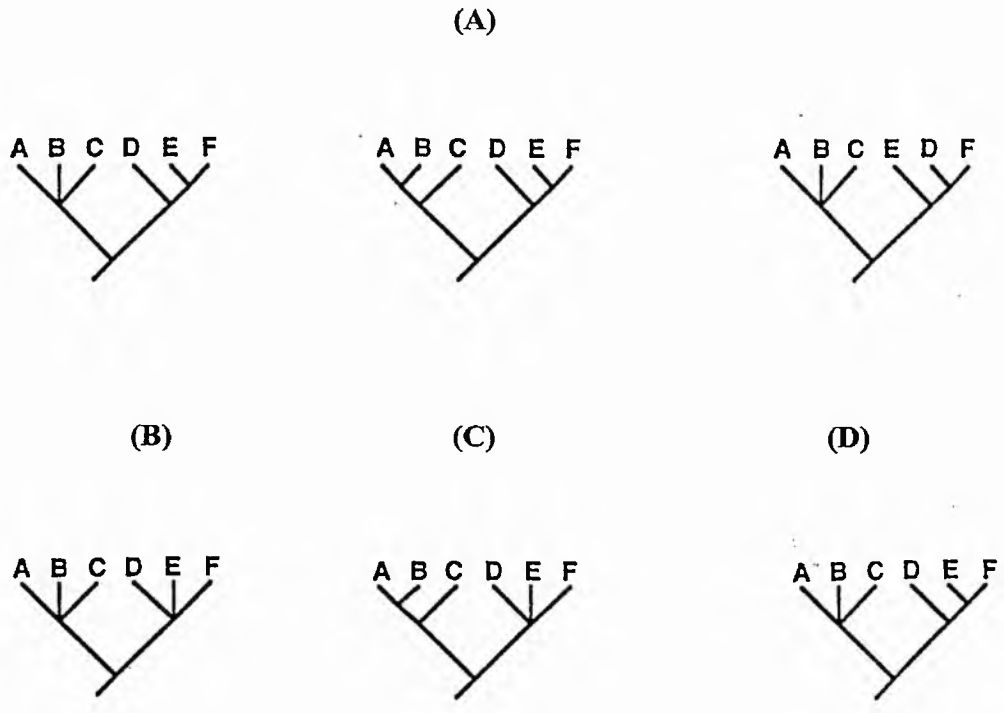


Figure 3.4. Types of consensus trees. **(A)** Rival trees to be summarised by each consensus tree; **(B)** strict consensus tree; **(C)** semi-strict consensus tree; and, **(D)** 50% majority rule consensus tree.

Robustness of an inferred tree: goodness-of-fit statistics

Several statistics can be calculated to determine the *goodness of fit* of phylogenetic trees with the data sets they are describing (Swofford, 1993). The most widely used measure of robustness is the consistency index (CI), which provides an indication of how well a particular tree topology explains the data. In simple terms, a transformation series with little or no homoplasy will yield a high CI value (1 being the maximum), while those with high homoplasy have a lower value (0 as a minimum). CIs are calculated on the basis of synapomorphies only, expressed as the minimum number of changes or steps necessary if all data agreed (m) divided by the actual number of steps (s) in the tree (i.e. $CI = m / s$; Kluge and Farris, 1969).

Another useful goodness-of-fit statistic is the retention index (RI), which indicates how well characters fit the tree that describes them. Farris (1989) described the RI to express the amount of synapomorphy in a data set by examining the actual amount of homoplasy as a function of the maximum possible homoplasy, or in other words, the rate of similarities in a tree due to synapomorphies. In addition, there is also the homoplasy index, which provides an indication as to the amount of homoplasy in a tree (Swofford, 1991 & 1993). All of these statistics can be determined for individual characters in addition to entire data sets.

On the other hand, the robustness of a specific clade within a tree can also be determined by: **(1)** an observation of the number of synapomorphies that support each branch in the tree; and, **(2)** by obtaining a decay index. The latter one is a useful index of support for a monophyletic group obtained by calculating the difference in tree lengths between the shortest trees that contain a group versus those that lack the group (Bremer, 1988). Further details of these and additional indices (e.g. the topology-dependent permutation tail probability, T-PTP test) can be found in Armstrong *et al.* (1994) and Hillis *et al.* (1996).

Reliability of inferred trees

The question of certainty of the historical relationships represented in a particular tree has been addressed by assigning confidence limits to its branches. To this purpose, two procedures have been widely used: the bootstrap and the jack-knife methods.

The **bootstrap method** (Felsenstein, 1985) is a non-parametric resampling method (Hillis *et al.*, 1996) which operates by estimating the variance of the sampling distribution by repeatedly resampling data from the original data set. This method is based on the mathematical principle of constructing a series of fictional matrices; data from the original matrix may be present once, more than once, or not at all in the new matrix. Each bootstrap data set is then analysed using a heuristic or branch-and-bound search to produce a tree or a set of trees (in the present study, heuristic searches were used). This procedure is repeated a predetermined number of times (normally 100) on the random samples, and the percentage of occurrence of a particular group or component that appears in the consensus of the bootstrapped trees is regarded as an index of support for monophyly in that group.

This method, although used widely in phylogenetical studies, cannot be considered as a true confidence limit in a statistical sense. Moreover, it has created controversy (Van Dongen, 1995; Felsenstein & Kishino, 1993; Hillis & Bull, 1993), as a necessary condition for the bootstrap to be valid is that configurations in the characters are independently and identically distributed. This is not true in all evolutionary processes or in the case of multistate characters recoded into binary data.

The **jack-knife** procedure (Mueller & Ayala, 1982) operates in a similar way to the bootstrap, but is based on gene frequency data. This method resamples the original data set by eliminating k data points at a time and recomputing the estimate from the remaining $n - k$ observations. New trees are constructed from the resulting reduced matrix and a cluster from the original tree is confirmed as robust if it appears in the new tree.

3.2.5. Data analysis and phylogeny reconstruction based on cpDNA variation in *Pachyrhizus*

RFLP variation across the total cpDNA genome

Data were recorded from the banding patterns obtained on luminographs. The number of base pairs sampled was calculated with Palmer's equation (1985) using an estimated size of 150 kb for the cpDNA molecule:

$$\frac{\text{(Total no. of fragments/enzyme/probe) x (No. of bases in enzyme recognition site)}}{\text{150 kb}} \times 100$$

Two types of analysis were performed on the data for tree construction. The first methodology used restriction fragments as characters, i.e. an FOA approach as described in section 3.1.4; the second analysis required the interpretation of the series of bands in terms of site and length mutations, i.e. an SOA method. Fragment size was calculated using Templeton's logarithm (software LENGTH, 1988).

Each taxon was characterised in terms of data produced by 90 probe-enzyme combinations (PECs), i.e. a result of sequential hybridisations involving 15 restriction enzymes (Table 2.2) and six probing sessions with MB1, MB2, MB5+6, MB7, MB8 and MB9+10. Fragments and sites were scored as present (1) or absent (0), and organised on a worksheet fashion as a raw data matrix. For FOA examination, all autapomorphies and monomorphic bands were removed prior to phylogenetic analysis, i.e. only synapomorphies were considered so as to achieve the principles of a cladistic approach (see section 3.1.4). For the SOA analysis, synapomorphic fragments were represented as characters based on the interpretation of the observed site mutations.

The resulting data matrices were processed by the software programme PAUP ver. 3.1.1 on an Apple Macintosh *Power PC 7200/75* using Dollo parsimony (see earlier in this chapter). Due to the large number of taxa to be analysed a heuristic search was conducted followed by branch swapping and the TBR algorithm. A strict consensus

tree was calculated to summarise the results from all equally parsimonious trees constructed with either fragment or site data. In addition, PAUP was used to calculate three goodness-of-fit statistics: the consistency, retention and homoplasy indices. These measures were useful not only in comparing characters on a single tree, but on multiple trees as well. They also provided a reliable way to estimate which characters support which hypotheses of topology (Swofford, 1993) when different assumptions were made.

Finally, an indication of the robustness of the clades in each tree was obtained by determining the number of synapomorphies that supported each branch; moreover, the bootstrap procedure (simple taxon addition, TBR branch-swapping & Dollo parsimony) was also implemented to assign statistical confidence to hypotheses of relationship in the cladograms. To evaluate further the relative robustness of clades found in the equally parsimonious trees, strict consensus were constructed of all trees up to one step longer than the shorter trees, then of all trees two steps longer, and so forth, until the consensus eventually collapsed to an unresolved bush. This procedure yielded a decay index (Bremer, 1988), indicating the number of steps that must be added before each clade present in the minimum length trees is no longer unequivocally supported.

Restriction site variation within a PCR-amplified cpDNA region

The methodology described in this section is based on PCR amplification of DNA regions within the chloroplast genome, subsequent restriction with a battery of endonucleases and the visualisation of the resulting fragments. The efficiency of this approach has already been demonstrated in phylogenetic and population genetic studies (e.g. Arnold *et al.*, 1991; Petit *et al.*, 1993; Demesure *et al.*, 1996).

Data were recorded from hard copies of gels obtained after photodocumentation with a *Herolab* image analysis system. Fragment size was estimated by comparison with a 1 kb DNA ladder and the analysis was conducted in the same way as reported above for total cpDNA variation, i.e. both FOA and SOA approaches were used. Phylogeny

reconstruction based on parsimony methods followed the same procedures and parameters (consensus trees, CI, RI, HI, the bootstrap, etc.) used for total cpDNA genome, as described in the previous section.

3.3. Results: RFLP variation across the total cpDNA genome

Pachyrhizus DNA was completely digested by most restriction enzymes used in this study. However, the enzymes *Bam*HI, *Dra*I and *Xba*I did not always cut reliably and only produced fully digested fragments for some PECs (probe x enzyme combinations). Nonetheless, these enzymes were included in the analysis, as they generated additional information for the reconstruction of phylogenetic relationships. Moreover, *Bam*HI and *Xba*I produced taxon-specific cpDNA fragments (see section 3.3.2).

The eight MB cpDNA probes used in this study successfully hybridised with *Pachyrhizus* DNA on the nylon membranes to reveal RFLP variation such as that illustrated in Plates 3.1 to 3.5. A total of 678 restriction fragments were detected by the analysis. Of these, 20 (3%) were uniform across all accessions sampled (i.e. monomorphic), as shown for example in Plate 3.1 ; 529 (78%) fragments were shared between accessions (i.e. synapomorphic) and thus cladistically informative (e.g. Plate 3.2); and, 129 (19%) were unique to a particular accession, mainly those of *P. ferrugineus* and the three outgroups included in the study.

The high number of synapomorphic fragments obtained in this study is due to the utilisation of eight four-base cutters (out of a total of 15 enzymes) and the inclusion of three outgroups. After an SOA analysis of the data (Bremer, 1991), these 529 fragments were interpreted as being representative of 30 site mutations and 13 length mutations, as shown in Plates 3.2 to 3.5. No major structural rearrangements, such as large insertions, deletions or inversions were observed in the 90 PECs examined in this section.

The 15 restriction enzymes used in this RFLP variation survey sampled 3060 bp of *Pachyrhizus* DNA (Table 3.1), which represent approximately 2.038% of the cpDNA genome, assuming an average size of 150 kb (Palmer *et al.*, 1988). The highest number of base pairs was sampled by enzyme *Dde*I (a 4-base cutter), with 444 bp

(0.295%), while the lowest number was obtained by *Xba*I (a 6-base cutter) where only 42 bp (0.028%) were sampled.

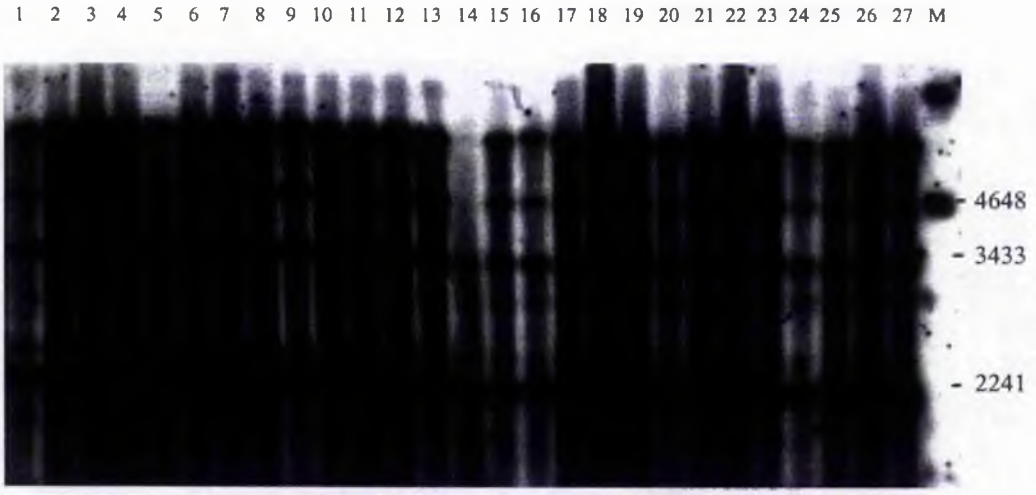


Plate 3.1. *Pachyrhizus* cpDNA digested with *Cfo*I and probed with MB2. No detectable variation was observed among the fragments at positions 4648 bp, 3433 bp and 2241 bp.

From left to right: **1,** AC215; **2,** AC220; **3,** AC226; **4,** AC231; **5,** EC032; **6,** EC502; **7,** EC509; **8,** EC531; **9,** EC558; **10,** EC565; **11,** EW203; **12,** EW051; **13,** FW237; **14,** FWLoc7; **15,** PW055; **16,** PWTM58; **17,** TC556; **18,** TC536; **19,** TC350; **20,** TC354; **21,** TC553; **22,** TC550; **23,** TCNA10; **24,** TWNanII; **25,** TW558; **26,** TWTM48; **M,** molecular size standard (*Hind*III digests of λ DNA; NBL Ltd.).

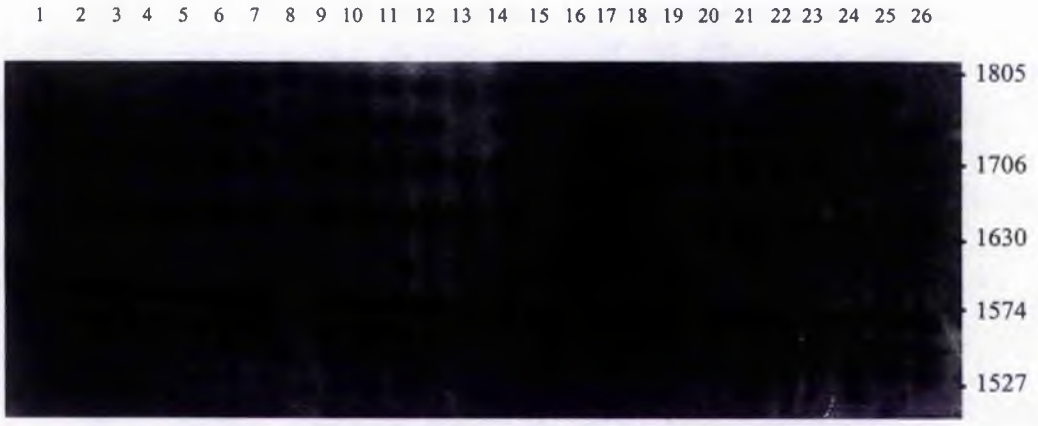


Plate 3.2. *Pachyrhizus* cpDNA digested with *Dde*I and probed with MB5+6. Several site mutations were detected, one of which is depicted here. An extra site in the 1805 bp fragment replaced it with a 1630 bp fragment and a ~175 bp fragment (not detected). The 1630 bp fragment was present in all *P. ahipa* accessions (lanes 1 - 4), TC553 (a jíquima, lane 21), TWNanII (lane 24) and TWTM48 (lane 26). See Table 3.3 for further details.

From left to right: **1,** AC215; **2,** AC220; **3,** AC226; **4,** AC231; **5,** EC032; **6,** EC502; **7,** EC509; **8,** EC531; **9,** EC558; **10,** EC565; **11,** EW203; **12,** EW051; **13,** FW237; **14,** FWLoc7; **15,** PW055; **16,** PWTM58; **17,** TC556; **18,** TC536; **19,** TC350; **20,** TC354; **21,** TC553; **22,** TC550; **23,** TCNA10; **24,** TWNanII; **25,** TW558; **26,** TWTM48.

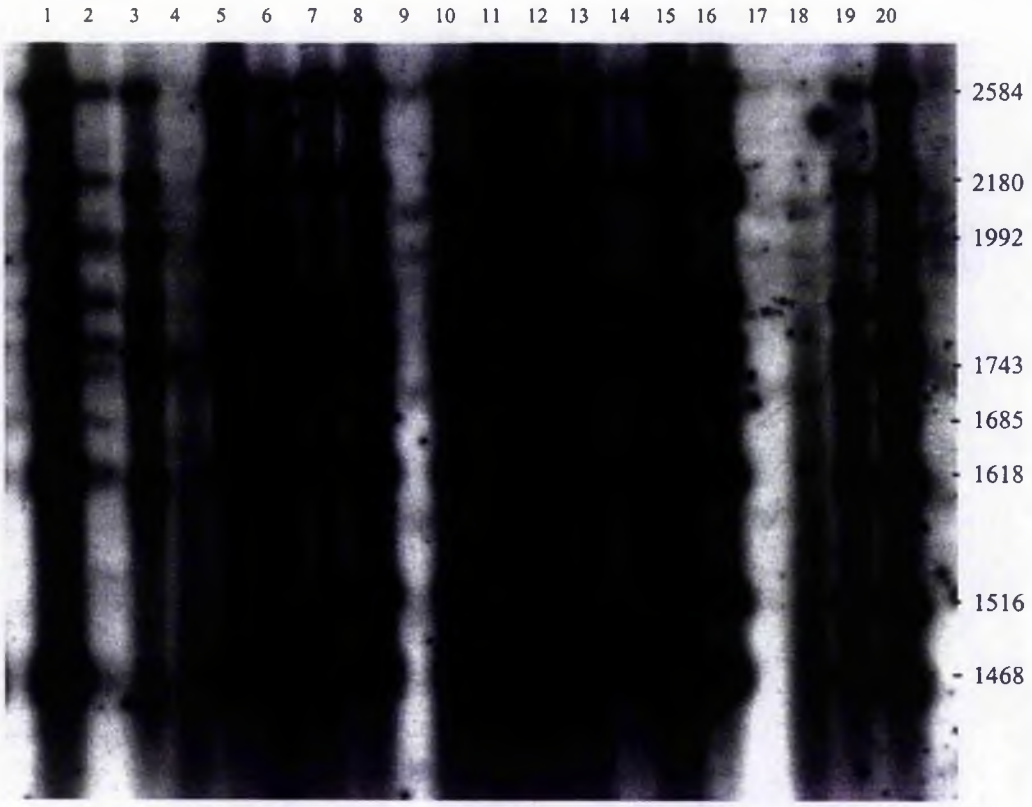


Plate 3.3. *Pachyrhizus* cpDNA digested with *Hae*III and probed with MB9+10. Two site mutations were detected, one of which is depicted here. An extra restriction site in the 1992 bp fragment gave rise to a 1516 bp fragment (lanes 7 - 12, 14 - 16, 19 & 20) and the expected 476 bp fragment was not detected (see Table 3.3 for further details). Blank lanes are the result of a low hybridisation with accessions having poor-quality DNA and, therefore, these were repeated in a subsequent probing session.

From left to right: **1,** EC509; **2,** EC531; **3,** EC558; **4,** EC565; **5,** EW203; **6,** EW051; **7,** FW237; **8,** FWLoc7; **9,** PW055; **10,** PWTM58; **11,** TC556; **12,** TC536; **13,** TC350; **14,** TC354; **15,** TC553; **16,** TC550; **17,** TCNA10; **18,** TWNanII; **19,** TW558; **20,** TWTM48.

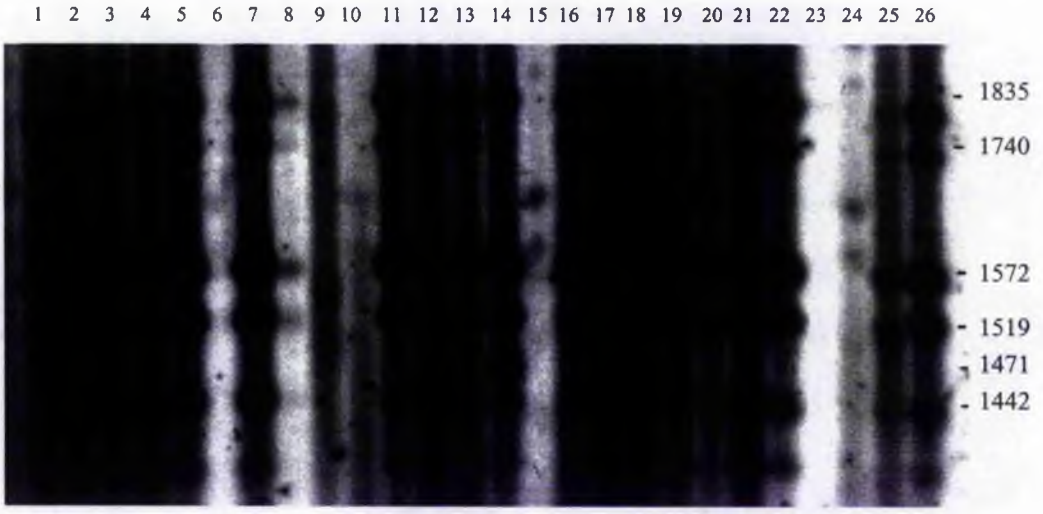


Plate 3.4. *Pachyrhizus* cpDNA digested with *Hpa*II and probed with MB7. One site mutation specific to *P. ferrugineus*, FWLoc7 (lane 14), is illustrated. An extra restriction site in the 1740 bp fragment gave rise to a 1471 bp fragment and a small, ~269 bp band (not detected).

From left to right: **1,** AC215; **2,** AC220; **3,** AC226; **4,** AC231; **5,** EC032; **6,** EC502; **7,** EC509; **8,** EC531; **9,** EC558; **10,** EC565; **11,** EW203; **12,** EW051; **13,** FW237; **14,** FWLoc7; **15,** PW055; **16,** PWTM58; **17,** TC556; **18,** TC536; **19,** TC350; **20,** TC354; **21,** TC553; **22,** TC550; **23,** TCNA10; **24,** TWNanII; **25,** TW558; **26,** TWTM48.

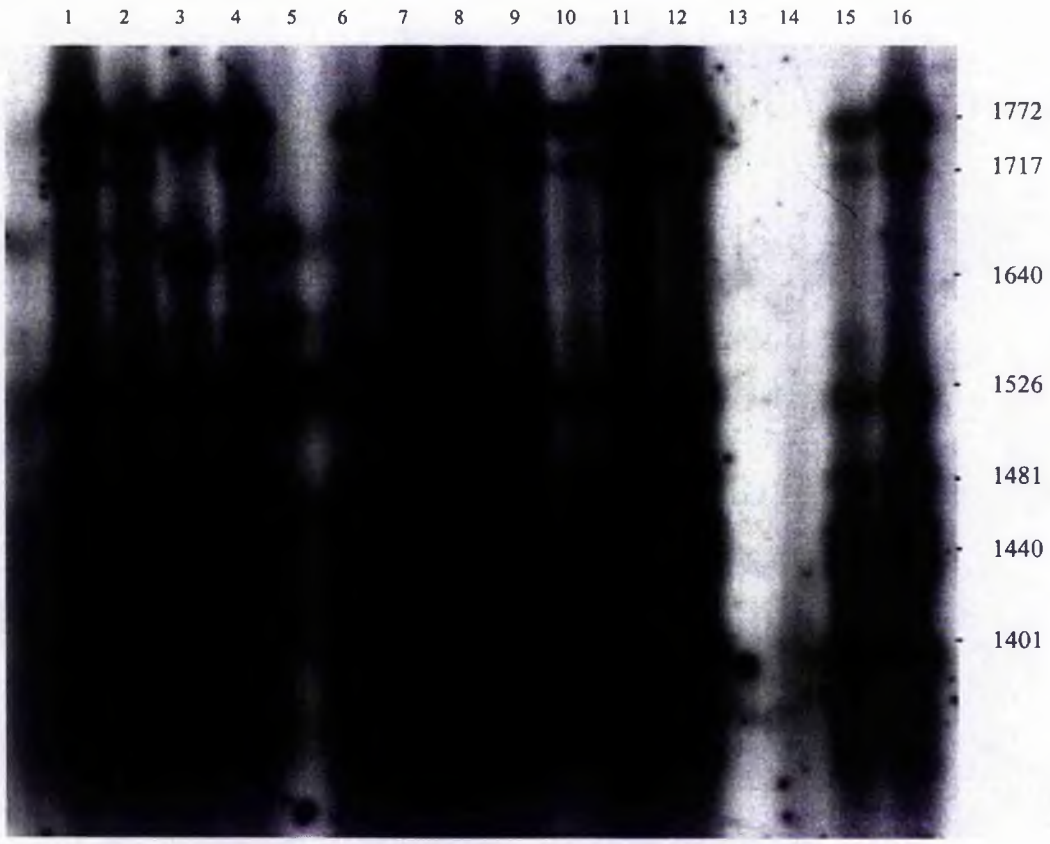


Plate 3.5. *Pachyrhizus* cpDNA digested with *Sau*3AI and probed with MB9+10. One site mutation specific to *P. ferrugineus*, FW237 (lane 3), was detected. An extra restriction site in the 1717 bp band gave rise to a 1640 bp fragment; the small ~77 bp fragment was not detected. Blank lanes are the result of a low hybridisation with accessions having poor-quality DNA and, therefore, these were repeated in a subsequent probing session.

From left to right: **1,** EW203; **2,** EW051; **3,** FW237; **4,** FWLoc7; **5,** PW055; **6,** PWTM58; **7,** TC556; **8,** TC536; **9,** TC350; **10,** TC354; **11,** TC553; **12,** TC550; **13,** TCNA10; **14,** TWNanII; **15,** TW558; **16,** TWTM48.

Table 3.1. Number of base pairs sampled by the 15 restriction enzymes used in the total cpDNA RFLP variation survey in *Pachyrhizus*.

Restriction enzyme	Total number of bands detected	Number of base pairs sampled	Proportion of total cpDNA genome (%)
Alu I	33	132	0.088
Bam HI	43	258	0.172
Bgl II	12	72	0.048
Cfo I	51	204	0.136
Cla I	12	72	0.048
Dde I	111	444	0.295
Dra I	20	120	0.080
EcoR I	62	372	0.248
EcoR V	18	108	0.072
Hae III	78	312	0.208
Hinf I	42	168	0.112
Hpa II	103	412	0.274
Sau3A I	69	276	0.184
Tru9 I	17	68	0.045
Xba I	7	42	0.028
Total	**678**	**3060**	**2.038**

Figure 3.5 summarises the distribution of variation detected in this survey within the *Pachyrhizus* cpDNA genome. As it can be seen, the most variable region of the genome was found to be that detected by probes MB9+10 (in the LSC region), where 1.045 mutations/kb were detected. The least variable region was that scanned by probe MB8 (LSC region) with 0.286 mutations/kb.

3.3.1. Phylogenetic analysis

Phylogenetic analysis was conducted on both FOA and SOA data sets. Only the results of the SOA are presented since phylogenetic trees showed similar topology for both data sets, with the main difference occurring in regard to the clustering of two accessions of *P. panamensis*, which were more basal in the SOA analysis. Presence or absence of restriction sites are listed in Table 3.2 for the 29 accessions of *Pachyrhizus* and the three outgroups (*Calopogonium caeruleum, C. mucunoides* and *Canavalia ensiformis*) analysed. Details of each character included in the matrix (from Table 3.2) are provided in Table 3.3. Presence/absence of the initial 529 synapomorphic restriction fragments (FOA method), as well as their description, are shown in Appendices 2 and 3.

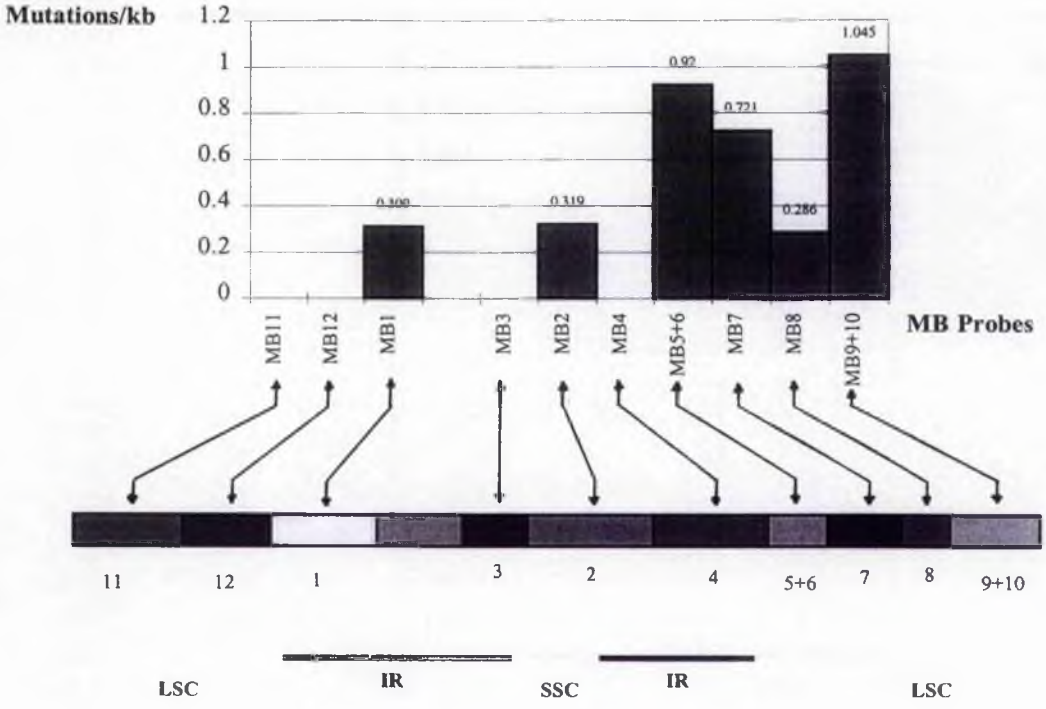


Figure 3.5. Distribution of the number of mutations per kilobase detected by cpDNA probes in *Pachyrhizus*. A linearised *Pst*I restriction map of the *Vigna* genome and probes used is also shown (the two heavy lines beneath the map represent the inverted repeat: IR; LSC: large single-copy region; SSC: small single-copy region). MB probes 3, 4, 11 and 12 were not available and, thus, are not illustrated.

A further attempt at phylogenetic analysis which included the 13 length mutations (described in Table 3.3) as additional characters resulted in a cladogram which was insufficiently resolved and, therefore, uninformative. Thus, it was decided not to include length mutations in the SOA approach.

Phylogenetic analysis of the 30 restriction site characters resulted in the production of four equally parsimonious trees that were 90 steps in length. One of these trees is shown in Figure 3.6. These cladograms had CI, HI and RI values of 0.333, 0.667 and 0.902, respectively; thus, there is a substantial amount of phylogenetic signal in the cpDNA data. A strict consensus tree (Figure 3.7) was calculated and resulted in a well resolved cladogram, although some polytomies were observed at terminal nodes.

Table 3.2. Restriction site data from analysis across total cpDNA genome used in the construction of *Pachyrhizus* phylogenetic trees.

Taxa †	Character number ‡
	11111111112222222223
	12345678901234567890123456789 0
	11111111112222222333333344
	89012345678901247890134567801 2
AC201	0001100111001010101000000000011
AC208	0001100111001010101000000000011
AC215	0001100111001010101000?000011
AC220	0001100111001010101000000000011
AC222	0001100011001010101000000000011
AC226	0001100111001?01010100000000011
AC231	0001100011001?01010100000000011
EC032	0000000000011000101100000001010
EC502	0000000000011000001100000001010
EC509	0000000000011000001100000001010
EC531	0000000000011000001100000001010
EC558	0000000000011000001100000001010
EC565	0000000000011000001100000001010
EW051	0000000000011000001100001001010
EW203	0000000010110000010000000001001
FW237	1111010010110110101100000000101
FWLoc7	0011011010110110101011101100001
PW055	0111000011???001010100000000011
PWTM58	001100?01???011010100000000011
TC350ch	0010000000000010011000000000001?
TC354ch	0001000000000010101000?000011
TC536as	0011000000000010101000000000011
TC550ji	0001000000000010101000000000011
TC553ji	0001100011001010101000000000011
TC556as	0011000000000010101000000000011
TCNA10	0001000000000010011000?000011
TW558	0011000000000010101000?000011
TWNanII	0011100011001010011000?000011
TWTM48	0011000011010010101000000000011
OUTcc	1010111111111110000111101101?1
OUTce	0110011000100101000?0111111?1
OUTcm	1010111011111010001011001101?1

†: Key to accessions is listed in Table 2.1. Coding of accessions of the *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima.

‡: Numbers in this header refer to the description of characters in Table 3.3; note that length mutations have not been included. The binary data represent presence (1) or absence (0) of a particular site mutation; missing data are indicated by '?'.

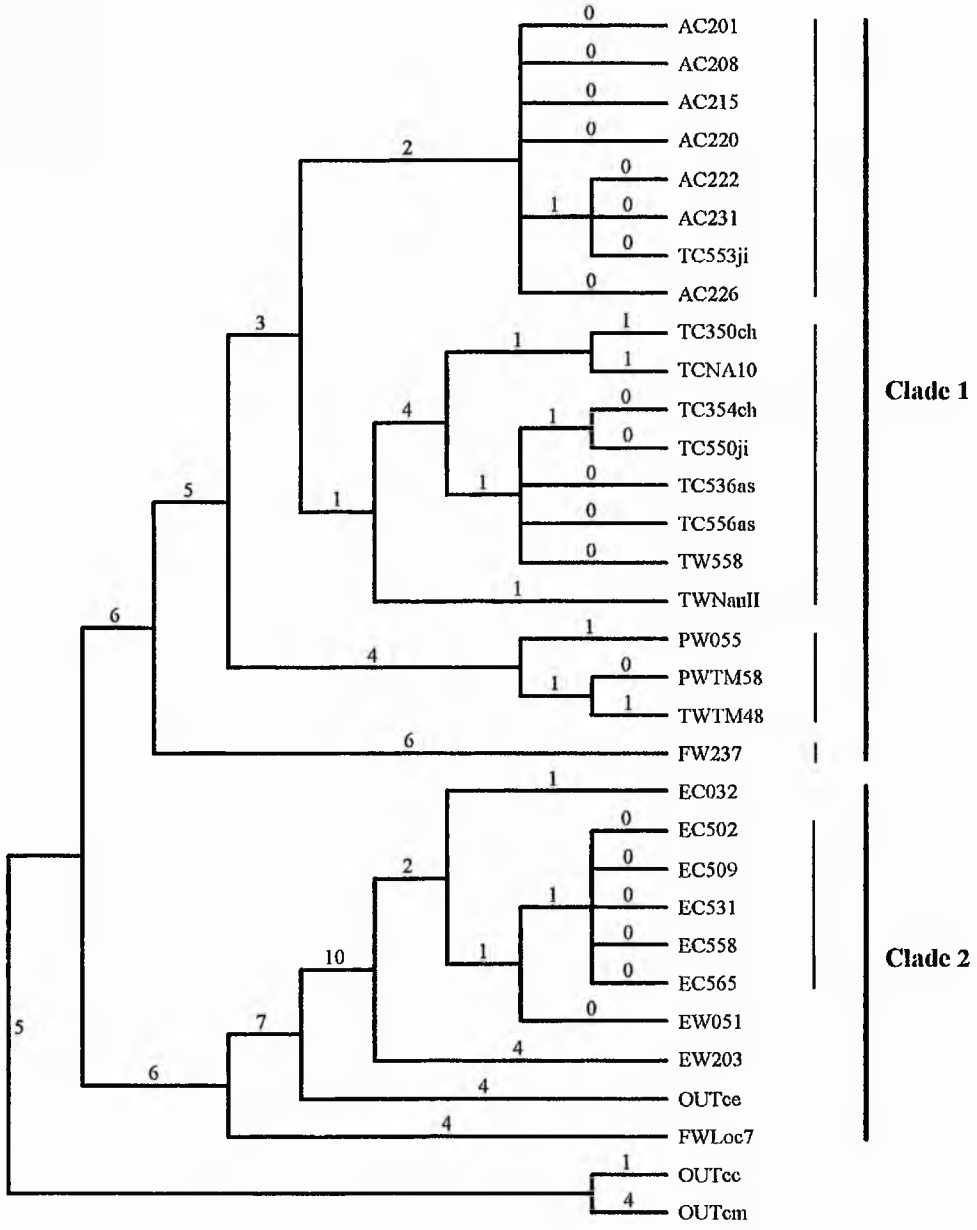


Figure 3.6. One of the four equally parsimonious trees generated from RFLP variation across total cpDNA genome in *Pachyrhizus* (Length = 90 steps; CI = 0.333; HI = 0.677; RI = 0.902). Numbers above branches indicate the number of synapomorphies supporting each of them (see text for further details).

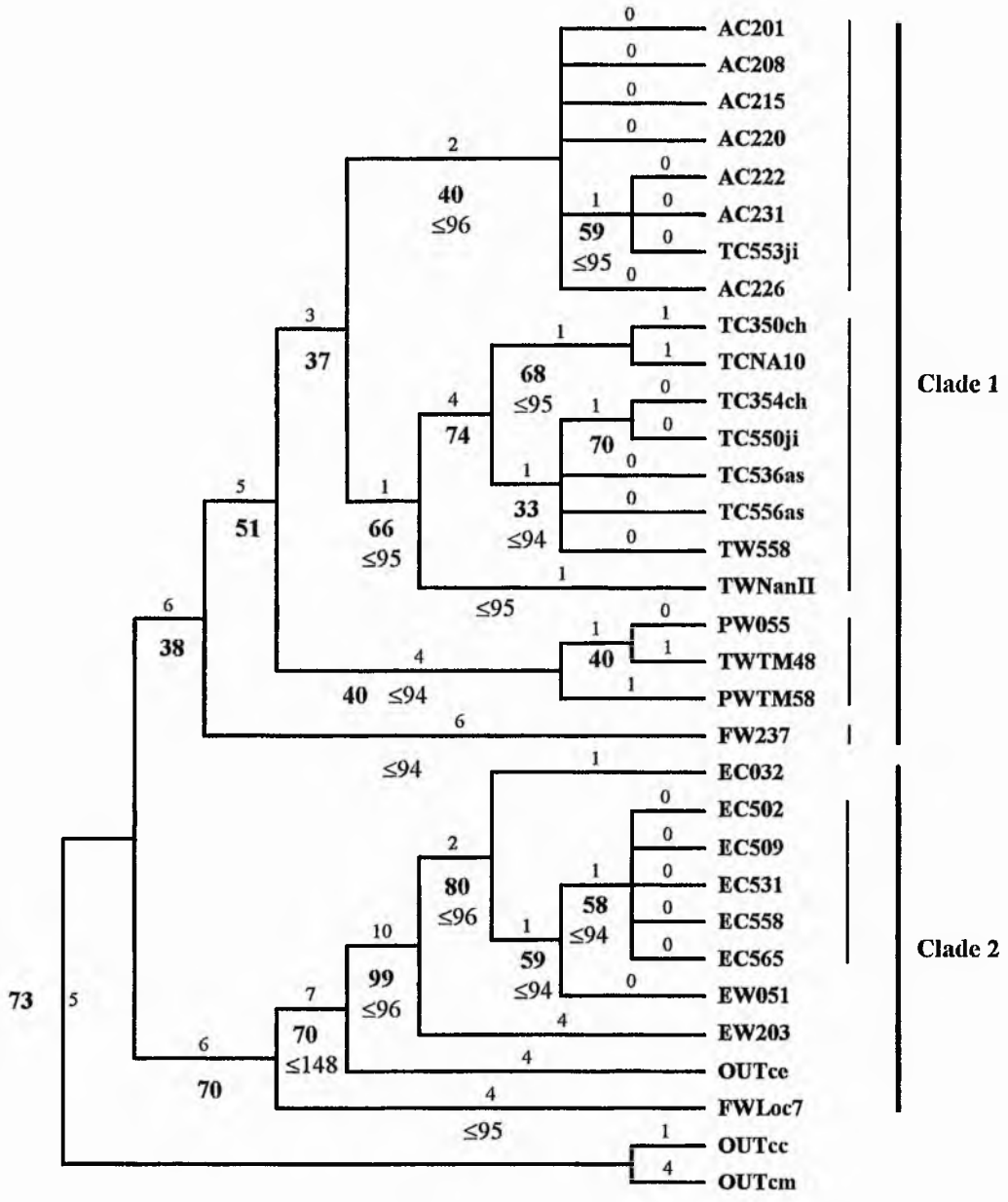


Figure 3.7. Strict consensus of four equally parsimonious trees generated from *Pachyrhizus* total cpDNA restriction site data (Length = 91 steps; CI = 0.323; HI = 0.677; RI = 0.897). Numbers above branches indicate synapomorphies; numbers below branches are bootstrap percentages (in boldface) and decay values. Clades and subclades are also indicated (see text for details).

Table 3.3. Description of the 43 restriction site and length mutations obtained in the study of total cpDNA RFLP variation.

Charac.	Probe	Enz.	Mutation † (bp)	Charac.	Probe	Enz.	Mutation † (bp)
1	MB1	‡	L: 1681 → 1619	23	MB1	‡	L: 1712 → 1697
2	MB2	‡	L: 1659 → 1631	24	MB5+6	*Alu*I	S: 2499 ≈ 1833 + ?
3	MB7	‡	L: 1587 → 1576	25	MB1	‡	L: 1556 → 1513
4	MB1	‡	L: 1770 → 1638	26	MB9+10	‡	L: 2265 → 2079
5	MB2	‡	L: 1846 → 1836	27	MB9+10	*Hae*III	S: 1992 ≈ 1516 + ?
6	MB2	‡	L: 1807 → 1789	28	MB9+10	*Hae*III	S: 1685 ≈ 1387 + ?
7	MB9+10	‡	L: ? → 2499	29	MB9+10	*Hin*fI	S: 1601 ≈ 1537 + ?
8	MB7	*Dra*I	S: 1925 ≈ 1626 + ?	30	MB5+6	*Hpa*II	S: 1915 ≈ 1842 + ?
9	MB2	*Cfo*I	S: 1851 ≈ 1828 + ?	31	MB5+6	*Hpa*II	S: 1723 ≈ 1478 + ?
10	MB7	*Cfo*I	S: 3020 ≈ 2931 + ?	32	MB7	‡	L: 2081 → 2070
11	MB7	*Cfo*I	S: 1943 ≈ 1922 + ?	33	MB7	*Hpa*II	S: 1740 ≈ 1471 + ?
12	MB7	*Cfo*I	S: 1565 ≈ 1544 + ?	34	MB8	*Hpa*II	S: 1730 ≈ ?
13	MB9+10	*Cfo*I	S: 2960 ≈ 1584 + 1524	35	MB9+10	*Hpa*II	S: 1626 ≈ 1333 + ?
14	MB1	*Dde*I	S: 1730 ≈ 1669 + ?	36	MB9+10	*Hpa*II	S: 1473 ≈ ?
15	MB2	*Dde*I	S: 1944 ≈ 1851 + ?	37	MB9+10	*Hpa*II	S: 1442 ≈ ?
16	MB2	*Dde*I	S: 1822 ≈ 1818 + ?	38	MB5+6	*Sau*3AI	S: 1397 ≈ 1332 + ?
17	MB5+6	*Dde*I	S: 1805 ≈ 1630 + ?	39	MB8	‡	L: 1593 → 1579
18	MB5+6	*Dde*I	S: 1711 ≈ 1706 + ?	40	MB9+10	*Sau*3AI	S: 1717 ≈ 1640 + ?
19	MB5+6	*Dde*I	S: 1699 ≈ 1574 + ?	41	MB9+10	*Tru*9I	S: 1319 ≈ ?
20	MB5+6	*Dde*I	S: 1553 ≈ 1527 + ?	42	MB9+10	*Tru*9I	S: 1223 ≈ 1089 + ?
21	MB7	*Dde*I	S: 1666 ≈ 1524 + ?	43	MB9+10	‡	L: 1453 → 1446
22	MB9+10	*Dde*I	S: 1794 ≈ 1482 + ?				

†: The letter 'S' denotes a site mutation, whereas 'L' represents a length mutation. A question mark indicates that a particular fragment(s) was missing or not detected.

‡: Length mutations were detected as a similar size change of the restriction fragments in question produced by the restriction enzymes when using the same MB probe.

A retention index (RI) of 0.897 was obtained for the strict consensus tree, which is a good representation of similarity of taxa in a particular clade due to synapomorphies. However, homoplasy (HI) and consistency (CI) indices of 0.677 and 0.323, respectively, were obtained and results have to be interpreted with care. Therefore, the bootstrap method (100 replicates) and the decay index were implemented to obtain estimates of reliability for monophyletic groups produced after phylogenetic analysis (Figure 3.7).

The SOA method for analysis of restriction fragment data resolved *Pachyrhizus* as a monophyletic genus, i.e. derived from a single ancestral group, with *P. ferrugineus* as the most basal species. The strict consensus tree divided taxa into two clades and was rooted by the addition of three outgroups (Figure 3.7). These two clades were: **(1)** a group containing taxa of South American distribution (namely *P. ahipa*, *P. tuberosus*

and *P. panamensis*) and one accession of *P. ferrugineus*; and, **(2)** a clade containing all *P. erosus* taxa, outgroup OUTce and one accession of *P. ferrugineus*.

Each clade is described in turn below using the number codes of the Yam Bean Project for the different accessions, i.e. A = *Pachyrhizus ahipa*, E = *P. erosus,* F = *P. ferrugineus*, P = *P. panamensis*, T = *P. tuberosus*; C = cultivated material, and W = wild status.

Clade 1

This group of 20 accessions was clearly separated from the rest of the genus by six synapomorphies. Within this clade, four subclades were identified. The first subclade contained all accessions of *P. ahipa* clustering together (40% bootstrap value); of these, two accessions clustered with TC553, a jíquima belonging to the *P. tuberosus* complex. The node that supported this relationship was present in 59% of the bootstrap samples, supporting a moderate relationship among these accessions (Figure 3.7).

A second subclade contained only accessions of the *P. tuberosus* complex that clustered tightly together. This subclade was separated from the remaining ones by one synapomorphy and supported by a bootstrap value of 66% (Figure 3.7), which depicts a reliable representation of relationships. However, within the subclade, two accessions of the ashipa cultivar group (TC536 and TC556) together with TW558 resolved as a polytomy, with only 33 out of 100 bootstraps supporting this group of accessions.

A third subclade contained both accessions of *P. panamensis* examined in this study (PW055 and PWTM58) and the remaining wild accession of *P. tuberosus* (TWTM48). A moderate bootstrap value (40 out of 100 sets) identified this subclade as monophyletic and, therefore, genetic relationships should be treated with caution. A final subclade contained one accession only (FW237), positioned as the most basal taxon within clade 1 (Figure 3.7).

Clade 2

This clade, clearly supported by six synapomorphies and a bootstrap value of 70%, contained 10 accessions, i.e. all accessions of *P. erosus*, one of the outgroups and one accession of *P. ferrugineus*. With the exception of accession EC032, all EC taxa had identical cpDNA and were resolved as one subclade (with 58 bootstrap replicates present out of 100). Both accessions of wild *P. erosus* were placed outside this subclade with EW203 as the most basal taxon within this species. Relationships among *P. erosus* accessions were supported by a high bootstrap value of 99% (Figure 3.7). Outgroup OUTce represents one of the most basal accessions in this second clade, supported by a relatively high bootstrap value of 70% (see 'Discussion' in this chapter). Finally, *P. ferrugineus* (accession FWLoc7) was the most ancestral species within clade 2.

Analysis of the decay index of the two plastome clades within the phylogeny of *Pachyrhizus* indicated that some of the structure of the minimal tree (93 steps) was lost when the consensus of trees $\leq$ 94 (one step longer than the minimal tree) was examined. Internal structure of the major clades was lost as trees further away from the minimal tree were considered; this happened mainly at terminal nodes. For instance, six accessions of *P. tuberosus* and two of *P. panamensis* in clade 1 were no longer resolved into dichotomies after 94 steps (Figure 3.7), and the subclade containing the *P. tuberosus* complex was no longer resolved after 95 steps. However, the two major clades were not lost until the consensus of trees $\leq$ 96 steps, i.e. three steps longer than the minimal tree, was examined (Figures 3.7 and 3.8). That is, the phylogenetic tree which has been recovered in this study had a moderate internal stability. Furthermore, the *P. erosus* plastome clade was strongly supported even in the consensus of trees $\leq$ 148 steps (55 steps longer than the minimal tree), after which the distinction among *Pachyrhizus* accessions (i.e. the ingroup) had been totally lost.

In general terms, the use of outgroups allows rooting such that the ingroup is monophyletic. However, rooting such that the ingroup, i.e. *Pachyrhizus*, was monophyletic with the three outgroups resolving as a sister group was possible in only

two of the taxa used for this purpose. The third outgroup was resolved within clade 2 (see 'Discussion' in this chapter).

3.3.2. Species/complex specific fragments

Thirty four fragments were identified as specific to particular taxa (Table 3.4). Restriction fragments were detected as specific to: **(1)** an accession within a species (e.g. the 1828 bp fragment in FW237 produced by PEC MB2 x *Cfo*I); **(2)** a species (e.g. a 1794 bp fragment produced by PEC MB9+10 x *Dde*I that was specific to *P. erosus*); or, **(3)** a species group or complex (e.g. a 1711 bp fragment generated by PEC MB5+6 x *Dde*I, that was exclusive to accessions of the *P. tuberosus* complex and *P. ahipa*).

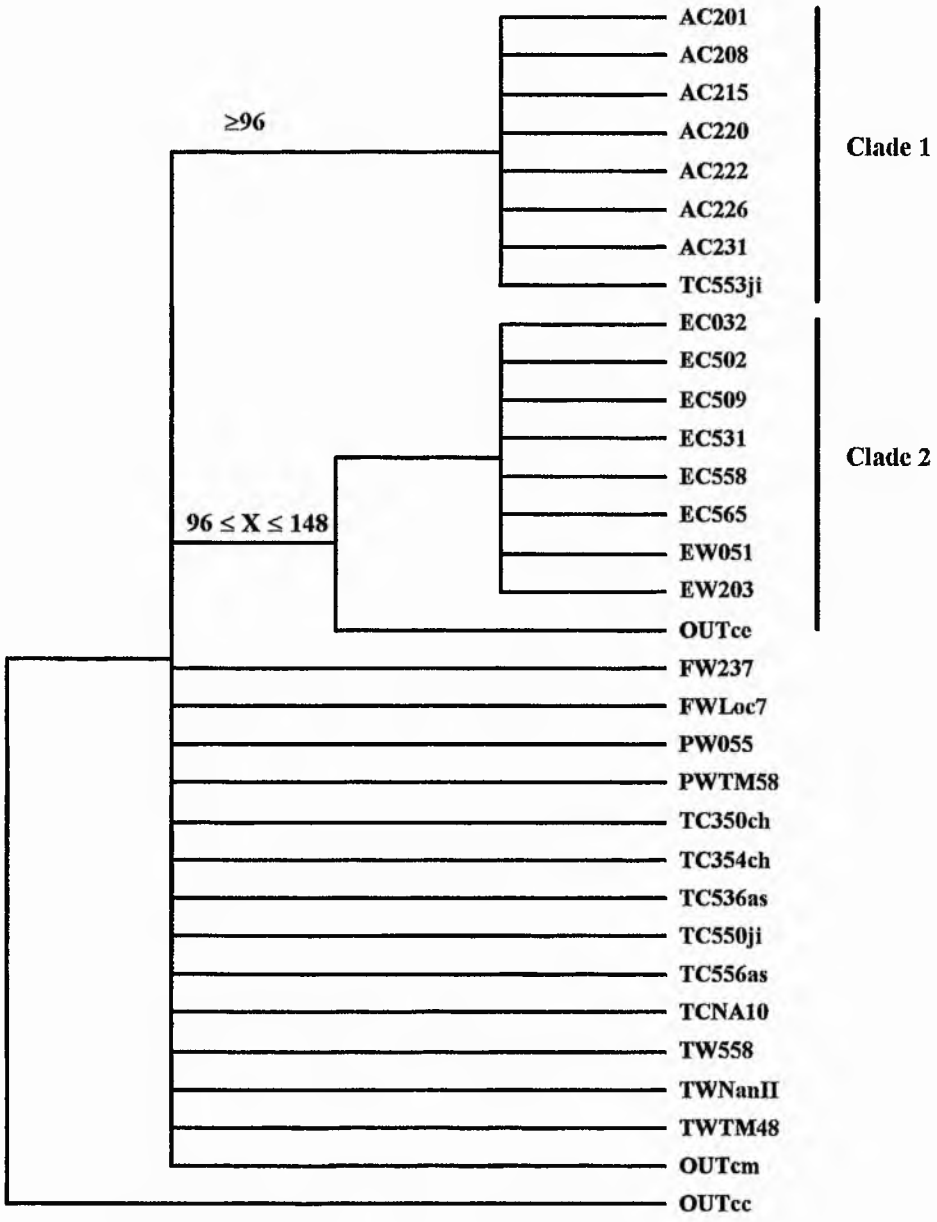


Figure 3.8. Tree showing the order of 'decay' of clades. This tree is the strict consensus of trees whose length is 96. At 97 steps the clade containing all AC accessions and TC553 was no longer resolved, but the *P. erosus* clade was still present. The tree became an unresolved 'bush' at 56 steps away from the minimal tree, i.e. a total length of 149 (see text for further details).

Table 3.4. Species/complex specific restriction fragments identified in a survey of total cpDNA variation in *Pachyrhizus*.

Probe	Enzyme	Fragment size (bp)	Details of specificity
MB1	*Bam*HI	1638	EW203 & TC350
MB9+10	*Bam*HI	2499	*P. ferrugineus* accessions (FW237 & FWLoc7)
MB2	*Cfo*I	1828	FW237
MB7	*Cfo*I	1943	*P. erosus* accessions & TC350
MB7	*Cfo*I	1544	*P. ahipa* accessions, TC553 (*jiquima*) & TWNanII
MB9+10	*Cfo*I	1790	FWLoc7
MB1	*Dde*I	2259	FWLoc7
MB1	*Dde*I	1669	FWLoc7
MB1	*Dde*I	1805	*Chuins:* TC350 & TC354
MB2	*Dde*I	1822	AC226
MB5+6	*Dde*I	1711	*P. ahipa* accessions & *P. tuberosus* complex
MB5+6	*Dde*I	1669	*P. ahipa* accessions & *P. tuberosus* complex
MB5+6	*Dde*I	1630	*P. ahipa* accessions, TC553 (*jiquima*) & TWNanII
MB5+6	*Dde*I	1527	TC553 (*jiquima*) & TWNanII
MB7	*Dde*I	1764	EW203
MB9+10	*Dde*I	1794	*P. erosus* accessions
MB5+6	*Alu*I	1833	EC032
MB1	*Bgl*II	1513	AC201 & TC354 (*chuin*)
MB9+10	*Cla*I	2079	FWLoc7
MB2	*Hae*III	1644†	*P. panamensis* accessions & FW237
MB9+10	*Hae*III	1992	*P. erosus* accessions & TC350 (*chuin*)
MB9+10	*Hinf*I	1601	EW203
MB2	*Hpa*II	1603	TC550 (*jiquima*)
MB2	*Hpa*II	1549	FW237
MB5+6	*Hpa*II	1723†	FWLoc7
MB7	*Hpa*II	1991	FW237
MB7	*Hpa*II	1471	FWLoc7
MB8	*Hpa*II	1730†	FWLoc7
MB9+10	*Hpa*II	1473†	FWLoc7
MB9+10	*Hpa*II	1442†	FWLoc7
MB9+10	*Sau*3AI	1717†	FW237
MB9+10	*Sau*3AI	1559	*P. panamensis* accessions
MB9+10	*Tru*9I	1319	*P. ferrugineus* accessions & EW203
MB9+10	*Xba*I	1446	AC201, AC220, TC556 (*ashipa*) & TC553 (*jiquima*)

†: Absent fragments representing site mutations, with loss of the expected smaller bands.

3.4. Results: Restriction site variation within a PCR-amplified cpDNA region

At the outset of this study it was intended to amplify *Pachyrhizus* DNA using both a larger number of universal primers (Demesure *et al.*, 1995) and accessions. Amplification conditions and concentrations had to be optimised empirically, and PCR amplification was only successful for five of the nine sets of primers tested. The same 32 accessions used for analysis of total cpDNA variation (see previous section) were subjected to PCR amplification; of these, 21 accessions (including one outgroup; see Table 2.1) yielded a PCR product which was subsequently used for the PCR-RFLP assay described in this section. The remaining primer sets and accessions either failed, or yielded a very faint PCR product or a multiple-band PCR amplicon.

The six restriction enzymes used in this study successfully digested *Pachyrhizus* DNA to reveal RFLP variation such as that illustrated in Plates 3.6 to 3.10. A total of 142 restriction fragments were detected by the analysis. Of these, 64 (45%) were monomorphic, as shown for example in Plate 3.6; 50 fragments (35%) were synapomorphic and thus cladistically informative (e.g. Plate 3.7); and, 28 fragments (20%) were unique to a particular accession, mainly those of *P. ferrugineus* and the operational outgroup included in the study.

The six restriction enzymes used in this PCR-RFLP variation survey sampled 568 bp of *Pachyrhizus* DNA (Table 3.5), which represent approximately 0.38% of the cpDNA genome, assuming a 150 kb average size. The highest number of base pairs was sampled by enzyme *Hsp*92II, with 112 bp (0.075%), while the lowest number was obtained by *Hind*III where only 72 bp (0.048%) were sampled.

Figure 3.9 summarises the distribution of variation detected in this survey within the *Pachyrhizus* cpDNA genome. The most variable PCR-amplified region of the genome was found to be that amplified by primer pair KK, where 0.00465 mutations/bp were detected. The least variable region was that amplified by primer pair HK, with 0.00059 mutations/bp.

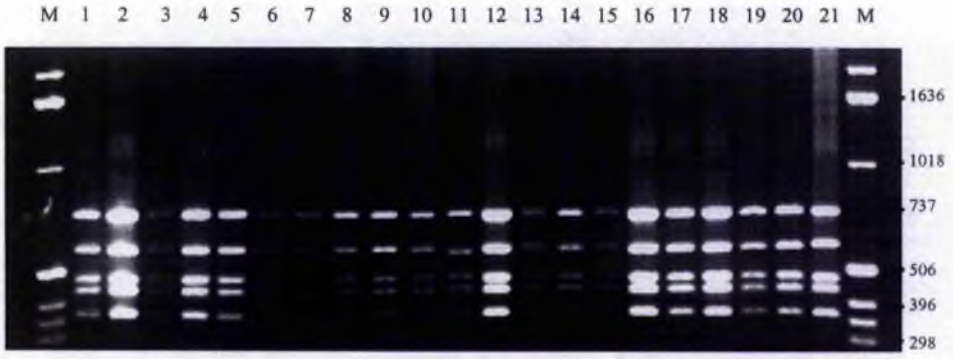


Plate 3.6. Restriction digests of a cpDNA amplified fragment using primer pair ML and enzyme *Cfo*I. Note the five monomorphic bands in positions 737 bp, 577 bp, 474 bp, 448 bp and 375 bp.

From left to right: **1**, AC201; **2**, AC220; **3**, AC231; **4**, EC032; **5**, EC509; **6**, EC558; **7**, EW203; **8**, FW237; **9**, FWLoc7; **10**, PW055; **11**, PWTM58; **12**, TC556; **13**, TC536; **14**, TC350; **15**, TC354; **16**, TC553; **17**, TC550; **18**, TWNanII; **19**, TW558; **20**, TWTM48; and, **21**, OUTcc. **M**, molecular size standard (1 kb ladder, Gibco).

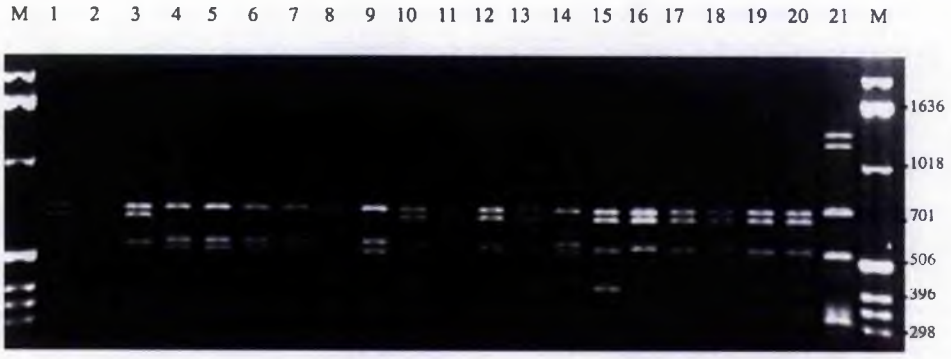


Plate 3.7. Restriction site variation within the cpDNA fragment amplified with primer pair KK and digested with *Rsa*I. A 701 bp fragment is present in all accessions of *P. ahipa* (lanes 1 - 3), *P. panamensis* (lanes 10 & 11) and the *P. tuberosus* complex (except for accession TC350; lanes 12 - 13 & 15 - 20). Note also the 572 bp and 540 bp bands in all *P. erosus* taxa (lanes 4 - 7) and accessions FWLoc7 (lane 9) and TC350 (lane 14). See text and Table 3.7 for further details.

From left to right: **1**, AC201; **2**, AC220; **3**, AC231; **4**, EC032; **5**, EC509; **6**, EC558; **7**, EW203; **8**, FW237; **9**, FWLoc7; **10**, PW055; **11**, PWTM58; **12**, TC556; **13**, TC536; **14**, TC350; **15**, TC354; **16**, TC553; **17**, TC550; **18**, TWNanII; **19**, TW558; **20**, TWTM48; and, **21**, OUTcc. **M**, molecular size standard (1 kb ladder, Gibco).

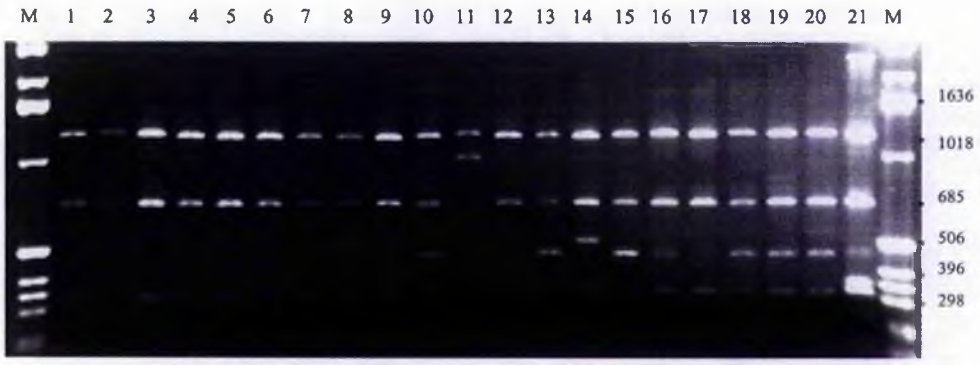


Plate 3.8. Restriction digests of the 2580-bp PCR-amplified cpDNA fragment in *Pachyrhizus*; primer pair KK and restriction enzyme *Hsp92*II were used. An extra site in the 998 bp fragment (PWTM58, lane 11) replaced it with a 685 bp band (lanes 1 - 10 & 12 - 21) and a small fragment (missing). Bands observed near the 517 bp area were partial digests. See text and Table 3.7 for further details.

From left to right: **1**, AC201; **2**, AC220; **3**, AC231; **4**, EC032; **5**, EC509; **6**, EC558; **7**, EW203; **8**, FW237; **9**, FWLoc7; **10**, PW055; **11**, PWTM58; **12**, TC556; **13**, TC536; **14**, TC350; **15**, TC354; **16**, TC553; **17**, TC550; **18**, TWNanII; **19**, TW558; **20**, TWTM48; and, **21**, OUTcc. **M**, molecular size standard (1 kb ladder, Gibco).

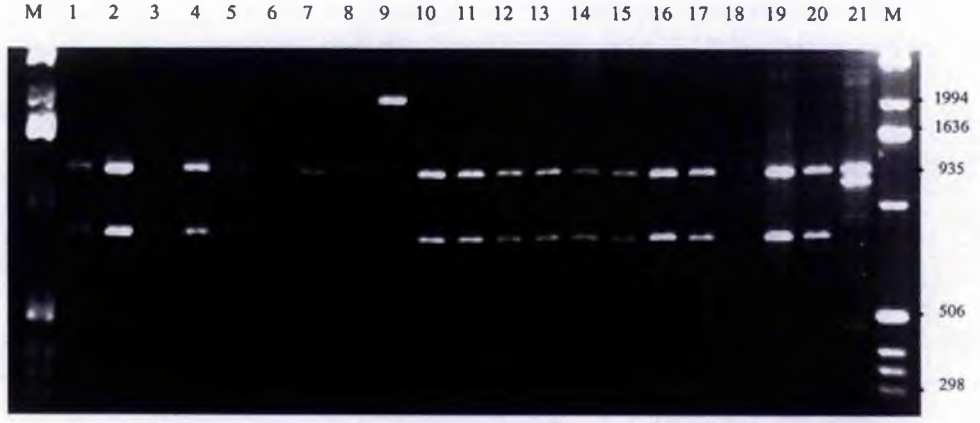


Plate 3.9. Restriction digests of the 3000-bp PCR-amplified cpDNA fragment in *Pachyrhizus*; primer pair CD and restriction enzyme *Rsa*I were used. Two autapomorphies are evident: **(1)** a 1994 bp fragment in lane 9 (FWLoc7); and, **(2)** a 935 bp band in lane 21 (outgroup OUTcc). See text and Table 3.7 for further details.

From left to right: **1**, AC201; **2**, AC220; **3**, AC231; **4**, EC032; **5**, EC509; **6**, EC558; **7**, EW203; **8**, FW237; **9**, FWLoc7; **10**, PW055; **11**, PWTM58; **12**, TC556; **13**, TC536; **14**, TC350; **15**, TC354; **16**, TC553; **17**, TC550; **18**, TWNanII; **19**, TW558; **20**, TWTM48; and, **21**, OUTcc. **M**, molecular size standard (1 kb ladder, Gibco).

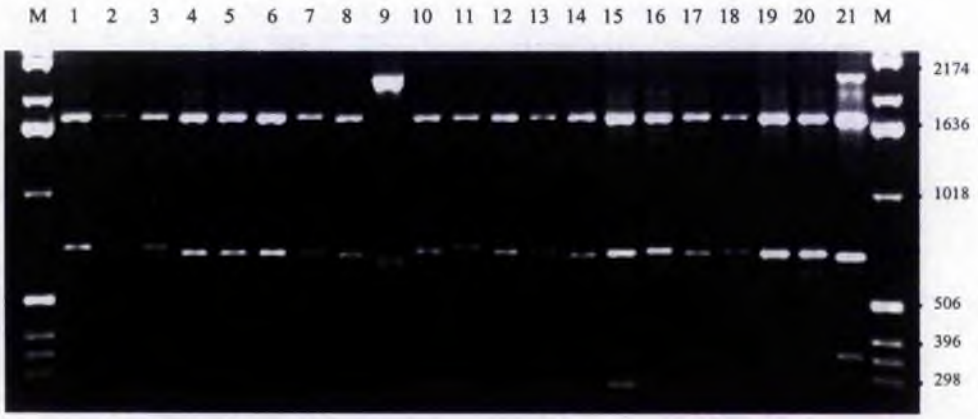


Plate 3.10. Restriction site variation within the cpDNA fragment amplified using primer pair KK and digested with *Hae*III. Autapomorphic site absence for accession FWLoc7 (lane 9; position 2174 bp) and presence for the remaining *Pachyrhizus* taxa (position 1686 bp). Small bands were the result of partial restriction enzyme digestion. See text and Table 3.7 for further details.

From left to right: 1, AC201; **2**, AC220; **3**, AC231; **4**, EC032; **5**, EC509; **6**, EC558; **7**, EW203; **8**, FW237; **9**, FWLoc7; **10**, PW055; **11**, PWTM58; **12**, TC556; **13**, TC536; **14**, TC350; **15**, TC354; **16**, TC553; **17**, TC550; **18**, TWNanII; **19**, TW558; **20**, TWTM48; and, **21**, OUTcc. **M**, molecular size standard (1 kb ladder, Gibco).

Table 3.5. Number of base pairs sampled by six restriction enzymes used in the PCR-RFLP variation survey in *Pachyrhizus*.

Restriction enzyme	Total number of bands detected	Number of base pairs sampled	Proportion of total cpDNA genome (%)
Cfo I	20	80	0.053
Hae III	25	100	0.067
Hind III	18	72	0.048
Hpa II	25	100	0.067
*Hsp*92II	28	112	0.075
Rsa I	26	104	0.070
Total	**142**	**568**	**0.380**

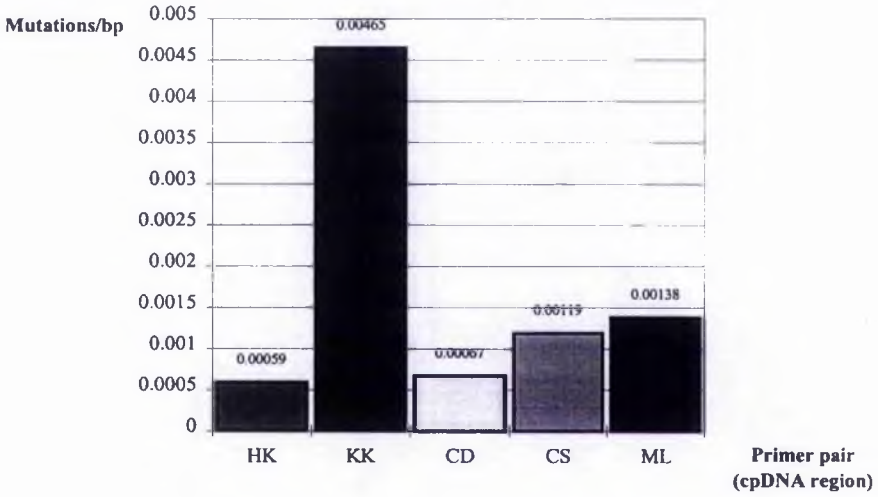


Figure 3.9. Distribution of the number of mutations per base pair detected by six restriction enzymes in five PCR-amplified cpDNA regions (description of primer pairs is shown in Table 2.4).

3.4.1. Phylogenetic analysis

Phylogenetic analysis was conducted on both FOA and SOA data sets. In this section, however, only the results of the SOA approach are presented since phylogenetic trees produced from this method showed better resolution with taxa arranged into dichotomies. In contrast, phylogenetic trees generated from an FOA analysis resolved most taxa of *P. ferrugineus, P. tuberosus* and *P. panamensis* into polytomies, and were largely uninformative.

After an SOA analysis of the data, the 50 synapomorphic fragments were interpreted as being representative of 19 site mutations (e.g. Plates 3.9 and 3.10); length mutations were not detected in the five cpDNA regions examined. No major structural rearrangements, such as large insertions, deletions or inversions were observed in the 30 primer/enzyme combinations studied in this section.

Presence or absence of restriction sites are listed in Table 3.6 for the 20 accessions of *Pachyrhizus* and one outgroup (*Calopogonium caeruleum*). Details of the 19 characters included in the matrix (from Table 3.6) are provided in Table 3.7. Presence/absence of the 50 initial characters and their description, i.e. the synapomorphic restriction fragments for the FOA approach, are shown in Appendices 4 and 5.

Phylogenetic analysis of the 19 restriction site characters resulted in the production of 33 equally parsimonious trees that were 27 steps in length. One of these trees is shown in Figure 3.10. These trees had CI, HI and RI values of 0.704, 0.296 and 0.934, respectively, which reflect a low degree of homoplasy and a substantial amount of phylogenetic signal in the cpDNA data. The strict consensus tree showed a well resolved topology (Figure 3.11); however, a polytomy occurred among several accessions of the *P. tuberosus* complex, due possibly to the low number of synapomorphic characters detected in the accessions of this species.

SOA of PCR-amplified chloroplast regions resolved *Pachyrhizus* as a monophyletic genus with *P. ferrugineus* as the most basal taxon. The strict consensus tree divided accessions into two clades and was rooted by the addition of outgroup OUTcc (Figure 3.11). These two clades were: **(1)** a group of taxa of South American distribution (*P. ahipa*, *P. tuberosus* and *P. panamensis*) and one accession of *P. ferrugineus*; and, **(2)** a clade containing all accessions of *P. erosus*, and one accession of *P. tuberosus* and *P. ferrugineus* each (TC350 and FW237, respectively). In general terms, these clades were highly congruent with those obtained after SOA of total genomic cpDNA with the main difference occurring in regard to the positioning of *P. ahipa* and *P. tuberosus* accessions (see below).

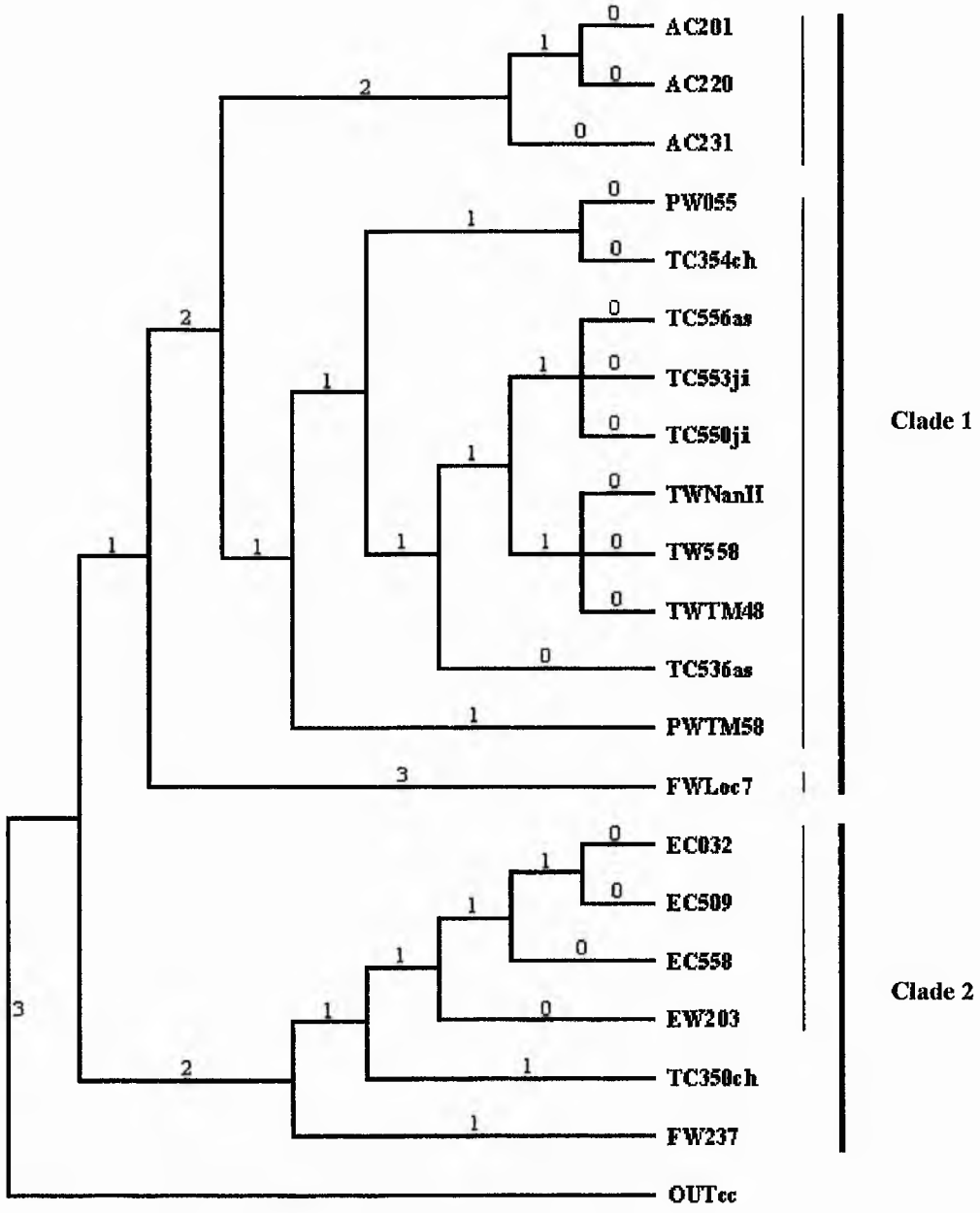


Figure 3.10. One of the 33 equally parsimonious trees derived from restriction site analysis of PCR-amplified cpDNA regions in *Pachyrhizus* (Length = 27 steps; CI = 0.704; HI = 0.296; RI = 0.934). Numbers above branches indicate the number of synapomorphies supporting them (see text for further details).

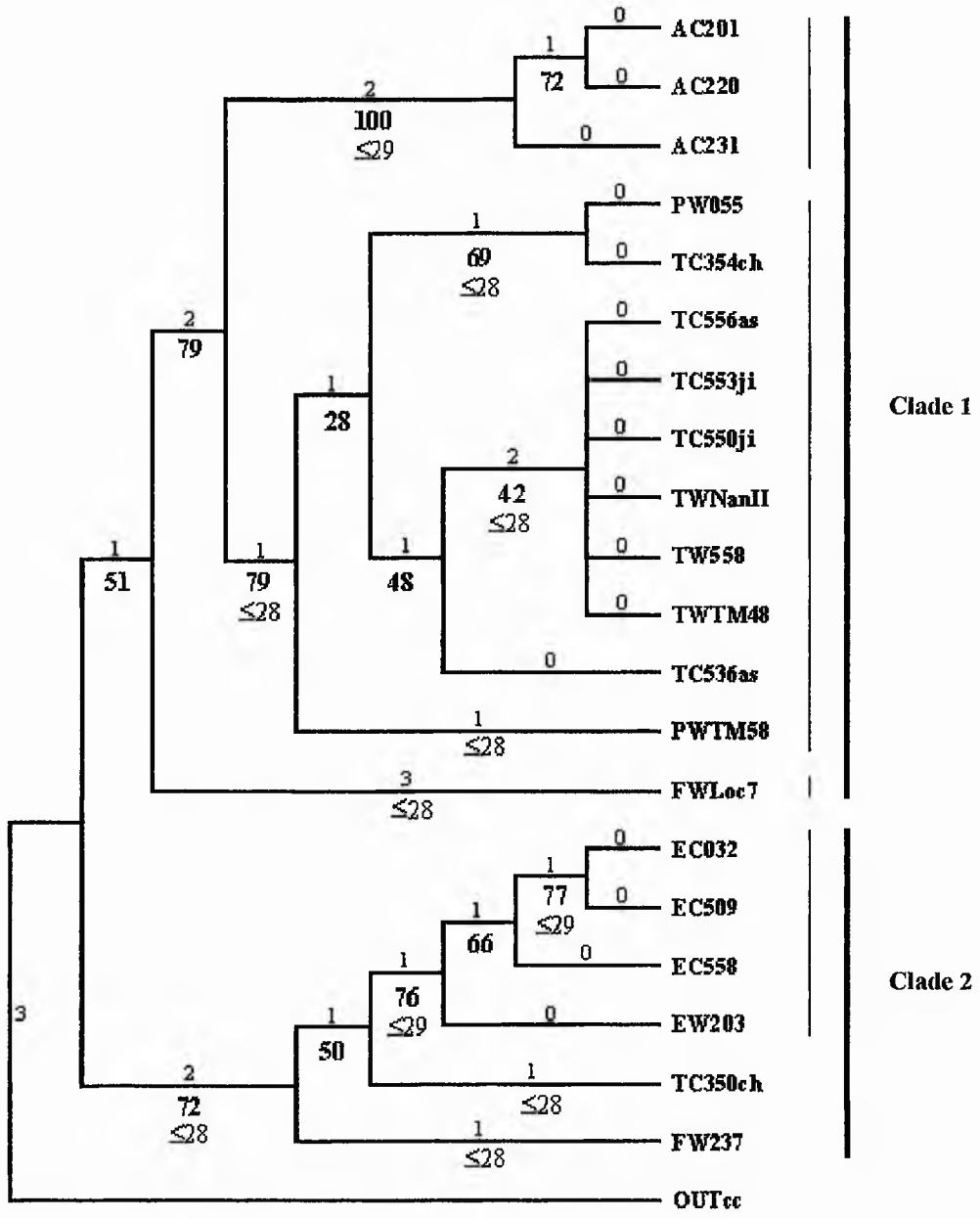


Figure 3.11. Strict consensus of 33 equally parsimonious trees generated from restriction site analysis of PCR-amplified cpDNA regions in *Pachyrhizus* (Length = 27 steps; CI = 0.704; HI = 0.296; RI = 0.934). Numbers above branches indicate synapomorphies; numbers below branches are bootstrap percentages (in boldface) and decay values. Clades and subclades are also indicated (see text for details).

Table 3.6. Restriction site data obtained from PCR-amplified regions of *Pachyrhizus* cpDNA.

Taxa†	Character‡
	1111111111
	1234567890123456789
AC201	0?11?10000011000100
AC220	0?11?10000011000100
AC231	0?11?10000011001100
EC032	0?11?01001111000100
EC509	0?11?01001111000100
EC558	0?11?01001111001100
EW203	0?11?01011111001100
FW237	1?11000011111001100
FWLoc7	0?01000010110001110
PW055	0010000010011001100
PWTM58	0010100110?11001100
TC556as	0010100010011000100
TC536as	0010100010011001?100
TC350ch	0111000011111001?100
TC354ch	0010000010011001?100
TC553ji	0010100010011000100
TC550ji	0010100010011000100
TWNanII	0010100010011000101
TW558	0010100010011000101
TWTM48	0010100010011000101
OUTcc	0010100011110?110010

†: Key to accessions is listed in Table 2.1. Coding of accessions of the *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima.

‡: Numbers in this header refer to description of characters shown in Table 3.7. The binary data represent presence (1) or absence (0) of a particular site mutation; missing data are indicated by '?'.

Clade 1

This group of 14 accessions was separated from the rest of the genus by one synapomorphy; the node that clustered taxa in this clade was present in 51% of the bootstrap samples, supporting a moderate relationship among these accessions (Figure 3.11). Within this first clade, three subclades were distinguished. A first subclade contained all *P. ahipa* accessions (AC) clustering tightly together (100% bootstrap value), being more basal than taxa of *P. tuberosus* and *P. panamensis* that were resolved in other subclades. The node that separated this subclade from the remaining

of clade 1 was strongly supported with 79% of the bootstrap samples being present and two synapomorphies.

Table 3.7. Description of the 19 restriction site mutations obtained in the study of PCR-amplified cpDNA regions of *Pachyrhizus*.

Character	Primer pair †	Restriction enzyme	Mutation (bp) ‡
1	HK	*Hsp*92II	150 ≈ 100 + 50
2	KK	*Cfo*I	460 ≈ 260 + 150
3	KK	*Hae*III	2174 ≈ 1686 + 676
4	KK	*Hae*III	267 ≈ 150 + ?
5	KK	*Hind*III	435 ≈ 398 + ?
6	KK	*Hpa*II	442 ≈ 300 + ?
7	KK	*Hpa*II	442 ≈ 200 + ?
8	KK	*Hsp*92II	998 ≈ 685 + 490
9	KK	*Hsp*92II	298 ≈ ?
10	KK	*Rsa*I	701 ≈ 572 + ?
11	KK	*Rsa*I	556 ≈ 540 + ?
12	CD	*Hae*III	2529 ≈ (462 + 450) + ?
13	CD	*Rsa*I	1994 ≈ 1244 + 791
14	CS	*Hpa*II	442 ≈ 395 + ?
15	CS	*Rsa*I	1287 ≈ 1215 + ?
16	ML	*Cfo*I	251 ≈ ?
17	ML	*Hae*III	1927 ≈ 951 + 882
18	ML	*Hpa*II	1268 ≈ 636 + 587
19	ML	*Hsp*92II	724 ≈ 621 + ?

†: Key and description of the primer pairs used in this survey are shown in Table 2.4.

‡: A question mark indicates that a particular fragment(s) was missing or not detected.

A second subclade contained all accessions of the *P. tuberosus* complex and the two PW accessions examined in this study (Figure 3.11). This subclade was separated from the other subclades by one synapomorphy and was strongly supported by a bootstrap value of 79 out of 100 random replicates. Both PW accessions were resolved as the most basal taxa within this subclade. However, one of these PW accessions (namely PW055) clustered together with TC354 (a chuin) and had identical cpDNA. Finally, seven accessions of *P. tuberosus* were poorly resolved at terminal levels of this subclade; six of them exhibited identical cpDNA and were resolved into a polytomy (42% bootstrap value). TC536 (an ashipa) was basal to this group and the node supporting this relationship had a low bootstrap value of 48%.

The last subclade contained one accession only (FWLoc7), as the most basal taxon within clade 1. However, the inclusion of FWLoc7 within clade 1 was not well supported as it was no longer resolved in this clade when the consensus of trees that were 28 steps in length was examined (i.e. one step away from the minimal length; Figure 3.11).

Clade 2

This second clade contained a total of six accessions and the node that clustered them was strongly supported by a bootstrap value of 72% (Figure 3.11). In this clade, all *P. erosus* accessions were resolved together as a subclade and the node that clustered taxa of this species showed a bootstrap value of 76%. In addition, one accession of *P. tuberosus* (TC350, a chuin) resolved as a sister taxon to the *P. erosus* subclade; this sister-group relationship was supported by a moderate bootstrap value of 50%. Finally, one accession of *P. ferrugineus* (FW237), as in clade 1, was the most basal taxon within this second group (72% bootstrap value).

Analysis of the decay index of the two plastome clades within the phylogeny indicated that most of the structure of the minimal tree (27 steps) was lost when the consensus of trees ≤ 29 (only two steps longer than the minimal tree) was examined. At this stage all taxa were no longer resolved except for the *P. erosus* and *P. ahipa* subclades (Figure 3.12). The subclade containing *P. panamensis* taxa and the *P. tuberosus* complex was no longer resolved after 28 steps, i.e. just one step longer than the minimal tree (Figures 3.11 and 3.12). After calculating a strict consensus of trees ≤ 30 steps (three steps longer than the minimal tree), the cladogram became an unresolved bush; subsequently, the distinction among *Pachyrhizus* accessions (i.e. the ingroup) had been totally lost. Therefore, decay values indicated that the internal structure of the phylogenetic tree recovered in this study was not particularly stable.

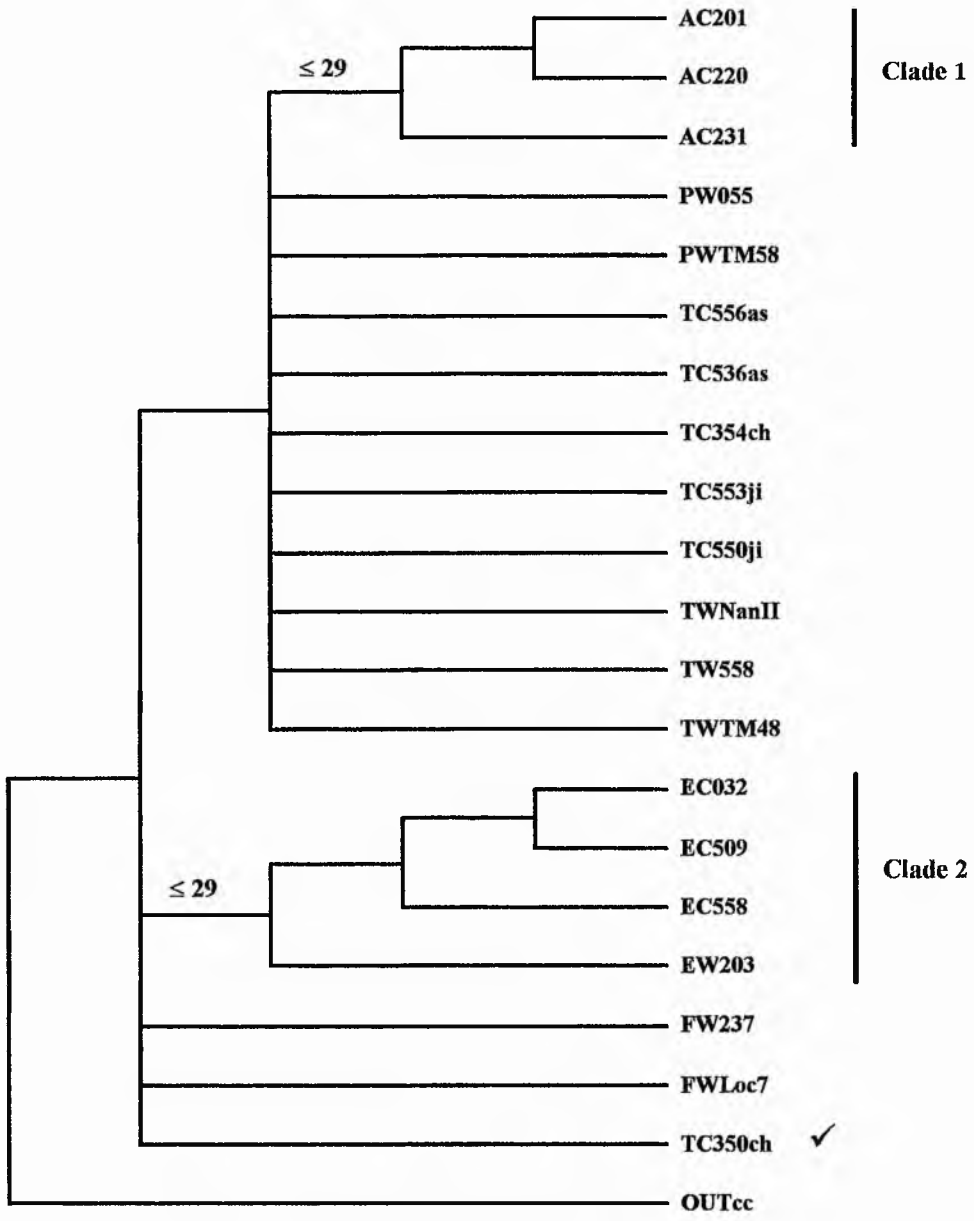


Figure 3.12. Tree showing the order of 'decay' of clades. The tree illustrated here is the strict consensus of trees whose length is 29. Almost all taxa were no longer resolved, except for the *P. erosus* and *P. ahipa* subclades (note that TC350 is no longer resolved in clade 2). The tree became an unresolved 'bush' at three steps away from the minimal tree length (see text for further details).

Finally, and in addition to the two clades described, the cladogram was rooted by the use of outgroup OUTcc which imparted direction upon the tree and allowed polarity determination (Figures 3.10 to 3.12).

3.4.2. Species/complex specific fragments

Eleven fragments were identified as specific to particular taxa (Table 3.8); some examples are shown in Plates 3.11 and 3.12. Restriction fragments were detected to be specific to: **(1)** an accession within a species (e.g. the 1994 bp fragment in PW055 produced by primer/enzyme combination CD x *Rsa*I); **(2)** a species (e.g. a 621 bp fragment produced by primer pair ML after digestion with *Hsp92*II, and specific to TW accessions); or, **(3)** a species group or complex (e.g. a 490 bp fragment generated by the combination KK x *Hsp92*II, that was present in accessions of the *P. tuberosus* complex and *P. panamensis*).

Table 3.8. Species/complex specific restriction fragments identified in a survey of PCR-amplified cpDNA regions in *Pachyrhizus*.

Primer pair	Enzyme	Fragment size (bp)	Details of specificity
HK	*Hsp*92II	150†	FW237
KK	*Cfo*I	750	TW accessions
KK	*Cfo*I	460	*P. panamensis* accessions & *P. tuberosus* complex
KK	*Hae*III	1686†	FWLoc7
KK	*Hind*III	435	FW accessions, PW055 & *chuins* (TC350, TC354)
KK	*Hsp92*II	998	PWTM58
KK	*Hsp92*II	490	*P. panamensis* accessions & *P. tuberosus* complex
KK	*Rsa*I	572	*P. erosus* accessions, FWLoc7 & TC350 (*chuin*)
CD	*Rsa*I	1994	PW055
ML	*Hpa*II	1268†	FWLoc7
ML	*Hsp92*II	621	TW accessions

†: Absent fragments representing site mutations, with loss of the expected smaller bands.

3.5. Results: an approach of combined restriction site data analysis

A further analysis involved the combination of restriction site characters obtained from both studies of the cpDNA genome, i.e. merging the 30 SOA site mutations obtained from the study of RFLP variation across the total cpDNA genome (section 3.3.1) with the 19 SOA site mutations yielded by the PCR-RFLP approach (section 3.4.1). The resulting data set was subjected to phylogenetic analysis using PAUP to investigate if the resolution of species relationships might be improved.

The Dollo algorithm produced eight equally parsimonious trees each of 102 steps (CI = 0.471; HI = 0.529; RI = 0.887); one of these is presented in Figure 3.13. The strict consensus of these trees is illustrated in Figure 3.14 and includes bootstrap percentages and decay values mapped along certain branches (viz., those that define clades critical to interspecific relationships). CI, HI and RI values of 0.466, 0.534 and 0.884, respectively, represent a moderate robustness of this consensus tree conferring reliability to the phylogenetic relationships it described.

In general terms, and as expected, phylogenetic relationships suggested by this analysis were similar to those revealed in the separate analyses of cpDNA variation, with the main difference occurring in the positioning of accessions of *P. panamensis* and accessions belonging to the *P. tuberosus* complex (see below for details). Phylogenetic analysis of a combined 49-character data matrix resolved *Pachyrhizus* as a monophyletic genus, with *P. ferrugineus* as the most ancestral species. The strict consensus tree was rooted by the use of the operational outgroup OUTcc (Figure 3.14) and reveals four clades, as follows:

Clade 1

This first clade contained all accessions of South American distribution, except for accession TC350 that was present in clade 2. This group of 13 accessions was separated from the rest of the genus by two synapomorphies; furthermore, the node that clustered taxa in this clade was present in 41% of the bootstrap samples,

supporting moderate relationships among these accessions (Figure 3.14). Within this first clade, three subclades were distinguished. The first subclade contained all AC accessions clustering tightly together (93% bootstrap value) and one accession of *P. tuberosus* (TC553, a jíquima) as a sister taxon to these *P. ahipa* accessions (53% bootstrap value). The node that separated this subclade from the remaining accessions in clade 1 was supported by a bootstrap value of 43% and three synapomorphies.

A second subclade contained accessions of the *P. tuberosus* complex. This subclade was separated from the other subclades by four synapomorphies and was supported by a bootstrap value of 43 out of 100 random replicates. Within this subclade, all TC accessions were resolved into a polytomy at terminal branches (bootstrap, 25%), whereas TW accessions clustered more internally in the cladogram as sister taxa to these TC materials.

The last subclade contained both PW accessions as the most basal group within clade 1 (40% bootstrap value); the node that separated this group from the remaining subclades was supported by a bootstrap value of 41% and two synapomorphies (Figure 3.14).

Clade 2

Clade 2 showed the highest bootstrap percentages in the cladogram and contained a total of five accessions; the node that united them was strongly supported by eight synapomorphies and a bootstrap value of 100% (Figure 3.14). In this clade, all *P. erosus* accessions were resolved together as a subclade and the node that clustered taxa of this species showed a bootstrap value of 100%. In addition, one accession of *P. tuberosus* (TC350, a chuin) is strongly allied with this *P. erosus* subclade supported by a bootstrap value of 100%.

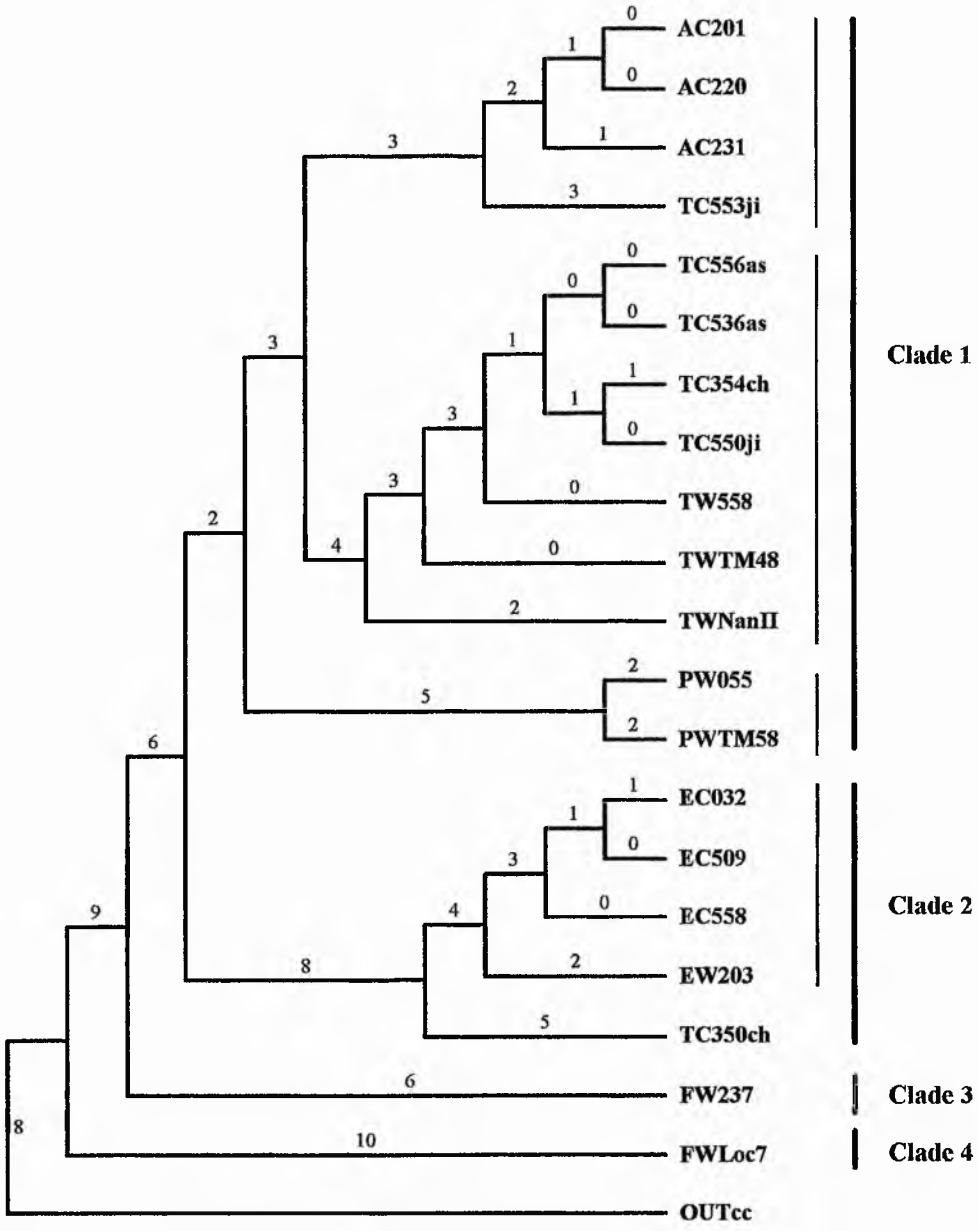


Figure 3.13. One of the eight equally parsimonious trees produced by SOA using a combined data matrix including a total of 49 restriction site characters (Length = 102 steps; CI = 0.471; HI = 0.529; RI = 0.887). Numbers above branches indicate the number of synapomorphies supporting them (see text for further details).

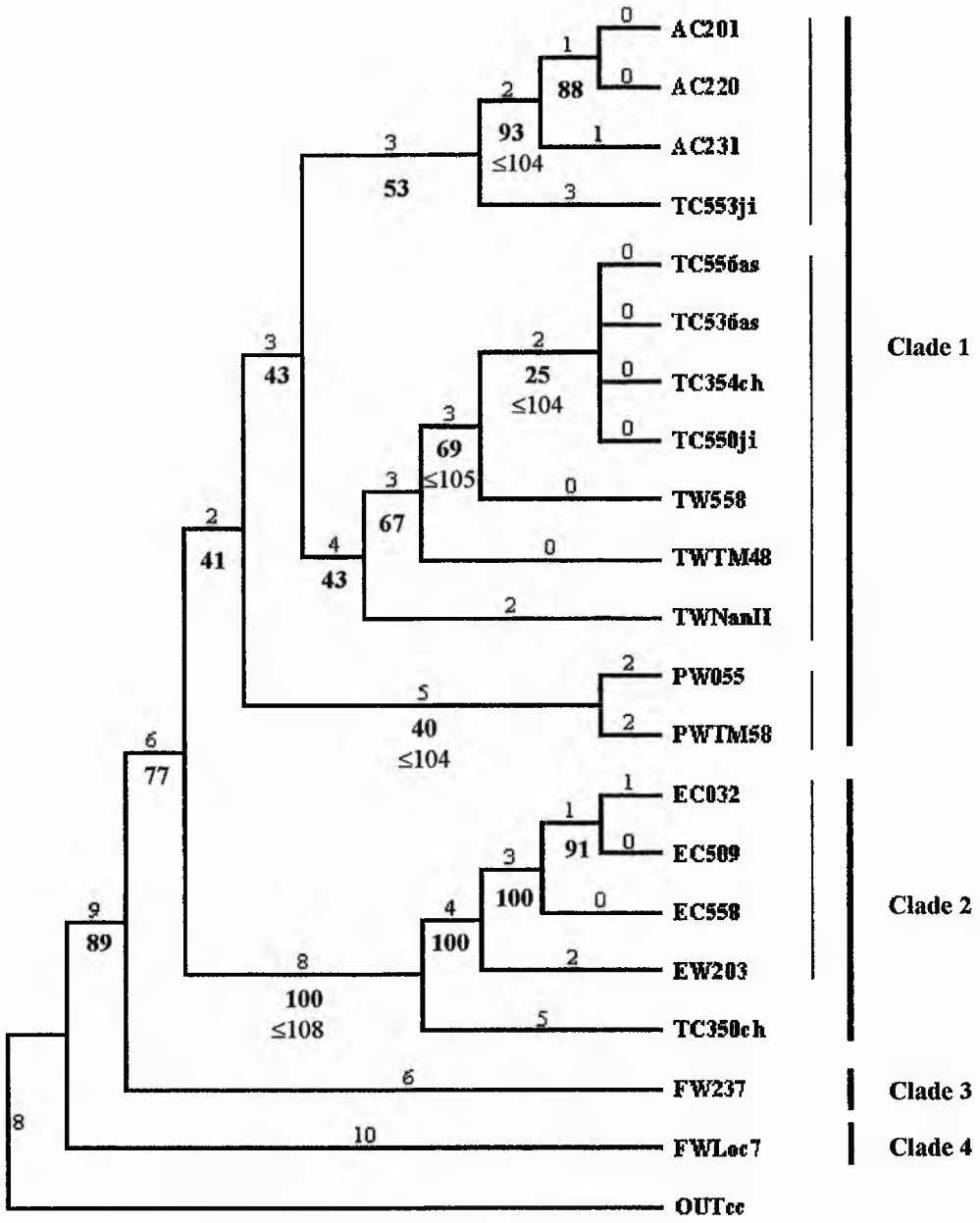


Figure 3.14. Strict consensus of eight equally parsimonious trees produced by SOA using a combined data matrix with a total of 49 cpDNA restriction site characters (Length = 102 steps; CI = 0.466; HI = 0.534; RI = 0.884). Numbers above branches indicate synapomorphies; bootstrap values (%, in boldface) and decay indices are mapped below each branch. Clades and subclades are also indicated (see text for details).

Clades 3 and 4

Both accessions of *P. ferrugineus* are the most basal taxa in the cladogram, with FW237 and FWLoc7 defining clades 3 and 4, respectively (Figure 3.14).

As in the two previous phylogenetic analyses, decay indices indicated that some of the internal structure of the major clades was lost as trees further away from the minimal tree were considered. When the consensus of trees ≤ 104 step was examined, i.e. two steps longer than the minimal tree, most of the four clades were no longer resolved (for example, branches containing all AC and PW taxa collapsed into polytomies). At this same level, the subclade containing the *P. tuberosus* complex was still recovered as a polytomy comprising five accessions (Figure 3.15.A), but was no longer resolved after 105 steps. Clade 3 (FW237) was rapidly lost when the consensus of trees ≤ 103 steps was examined, i.e. just one step longer than the minimal tree (Figures 3.14 and 3.15.A). The *P. erosus* clade and accession FWLoc7 (clade 4) were strongly supported in the consensus of trees ≤ 108 steps (six steps longer than the minimal tree), after which the distinction among accessions of the ingroup (*Pachyrhizus*) was totally lost (Figure 3.15.B). Therefore, the tree recovered after phylogenetic analysis of the combined 49-character data matrix had a moderate internal stability.

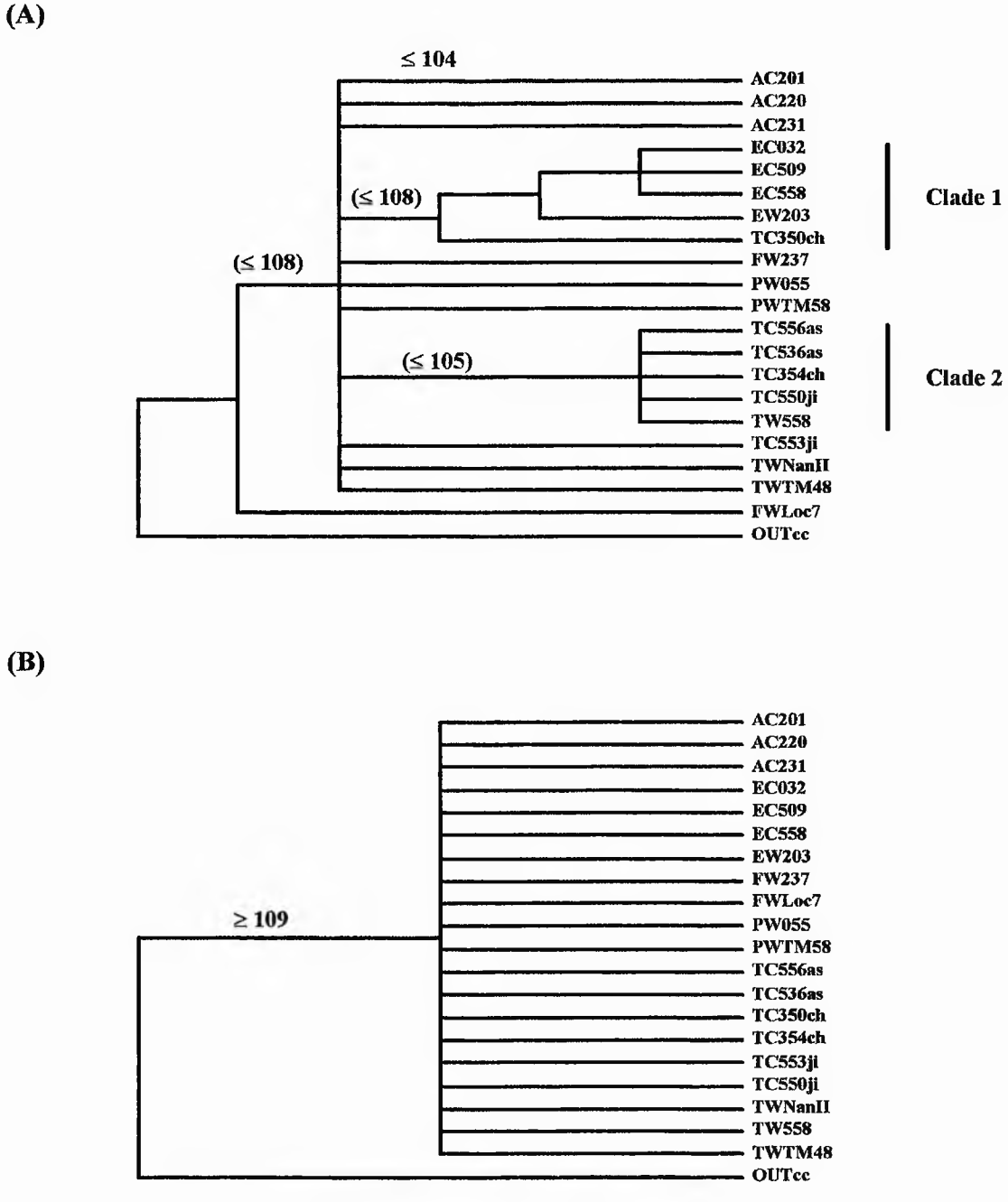


Figure 3.15. Trees showing the order of 'decay' of clades. **(A)** This tree is the strict consensus of trees whose length is 104; note that most of the four clades has been lost. At 104 steps the clade containing all AC accessions and TC553 was no longer recuperated, but the *P. erosus* clade and part of the *P. tuberosus* complex were still present. **(B)** The tree became an unresolved 'bush' at seven steps away from the minimal tree, i.e. a length of 109 (see text for further details).

3.6. Discussion

3.6.1. RFLP variation across the total cpDNA genome

Nature of cpDNA variation in *Pachyrhizus*

A maximum of nine site mutations was obtained with any one of the 15 restriction enzymes used in this study (Table 3.3), which made scoring of restriction fragment variation straightforward. Bremer (1991) expressed concern at the different ways of scoring data in cpDNA-based studies and analysed this problem in detail. Four possible data scoring methods have been described: two RFLP methods, namely FDA and FOA; and, two site occurrence methods which include SOA and SMA (see also section 3.1.4). Restriction fragments (i.e. RFLP methods) tend not to be used for parsimony analysis very often, due to the problem of introducing high levels of homoplasy in the data. In addition, restriction fragments may not evolve independently (i.e. three restriction fragments representing one site mutation), therefore introducing bias into the results (Swofford & Olsen, 1990). Palmer (1987) has reviewed the advantages of site occurrence over RFLP methods of analysis; such advantages have lead most workers in cpDNA systematics to use site occurrence approaches, particularly the SOA method.

In the present study of RFLP variation in *Pachyrhizus* cpDNA, it was possible to use successfully both fragment occurrence (FOA) and site occurrence (SOA) approaches in phylogeny reconstruction. A comparative examination of the use of the FOA and SOA methods yielded similar topologies, with consensus trees showing a much clearer resolution after SOA. In this latter approach, comparison of restriction sites allowed for interpretation of fragment pattern differences as individual mutations that affected the presence/absence and position of restriction sites. The great majority of mutations identified were restriction site changes assumed to be due to single nucleotide substitutions within the 4- or 6-bp site surveyed. Several length mutations were also resolved, but were omitted from phylogenetic analysis. Given the increased confidence in homology involved when using site data, the extra effort over

Pachyrhizus fragment data is well worth the investment. However, the dangers of including non-homologous characters in parsimony analyses seem to be overestimated, and choice of the analysis method should be dependent on a trade-off between accuracy and resources.

Phylogenetic relationships

The present study has demonstrated that cladistic analysis of RFLP variation across the total cpDNA genome can contribute to the reconstruction of evolutionary relationships in *Pachyrhizus*. The strict consensus tree (Figure 3.7) provides valid phylogenetic hypotheses for relationships among the five species. Polytomies observed in the terminal nodes of clades 1 and 2 represent regions where clustering differences occurred; however, most taxon relationships were well resolved. The two clades identified in this study are discussed below.

Clade 1: A South American evolutionary branch

This first clade comprised a total of 20 taxa, most of them of South American distribution, except for FW237 and PW055 that were originally collected in Central America. The clade was clearly separated from the rest of the genus and appears to have split away early in the evolutionary history of *Pachyrhizus*.

Clade 1 contained accessions of *P. ahipa*, *P. tuberosus* and *P. panamensis* plus one additional accession of *P. ferrugineus* (FW237) that represents the basal-most species within this group. Despite the lack of support for its internal structure, as revealed by moderate bootstrap values ranging from 33% to a maximum of 74%, the clade remains distinct from the rest of the genus and contains four subclades (Figure 3.7).

One subclade within clade 1 contained all accessions of *P. ahipa* and one accession of the *P. tuberosus* complex, supported by a bootstrap value of 40%. The clustering of TC553, a jíquima, with two Bolivian accessions of *P. ahipa* (AC222 and AC231) was supported by a bootstrap value of 59% and a decay index of ≤ 95. There are no

records of *P. ahipa* plants that are known to be undoubtedly wild; moreover, a wild progenitor of this species has yet to be identified and its geographical origin remains obscure (Sørensen, 1996). The phylogeny produced from SOA across the cpDNA genome would suggest that *P. ahipa* is possibly derived from *P. tuberosus*, since the latter one resolved closely within it.

Relationships between *P. ahipa* and *P. tuberosus* have been considered previously (Døygaard & Sørensen,1998; Ørting *et al.*, 1996; Sørensen *et al.*, 1997) without formulation of a clear hypothesis as to their phylogeny. Furthermore, palynological similarities have indicated a closer relation of *P. ahipa* with *P. panamensis* than with *P. tuberosus* (Sørensen, 1989). Brücher (1989) suggested that *P. ahipa* was probably selected from its wild material growing in *cejas de montaña* (≈ cloudforests). It is, therefore, of interest that *P. ahipa* and *P. tuberosus* taxa have clustered together within a subclade of the strict consensus tree (Figure 3.7); this affinity suggests that *P. ahipa* is possibly derived from materials of the *P. tuberosus* complex, providing new information concerning the evolution of the genus.

A second subclade within clade 1 comprised eight accessions of the *P. tuberosus* complex - namely ashipas, chuins, one jíquima and two TW accessions, identifying them as a monophyletic entity (Figure 3.7). This subclade was supported by one synapomorphy and was recovered in 66% of bootstrap replicates (decay index ≤ 95). All accessions in this subclade occur sympatrically in the Amazonian basin, except for TC550 (a jíquima) and both TW accessions collected in western Ecuador. The different cultigen types did not cluster together in this subclade; for example, the two chuins examined in this study clustered either with a jíquima (TC550) or with an escape (TCNA10) collected in Zamora Chinchipe, Ecuador, which resembles an ashipa type.

All accessions representing wild material of *P. tuberosus* occur on the western slopes of the Andes, without any record of cultivated material having been identified in their immediate surrounding areas. Morphologically, the wild material of this species is very similar to the ashipa genotypes and its foliage is evergreen (i.e. *P. ferrugineus*-

like). Present cpDNA data did not distinguish TW558, an accession of wild provenance (western Ecuador), from TC536 and TC556, both ashipas from the multituberous type present in the Amazonian basin. However, a bootstrap value of 33% and decay index ≤ 94 indicate that this close relationship between these wild and cultivated forms of *P. tuberosus* should to be interpreted with caution. These findings, therefore, provide some basis to a very close relationship between the ashipas and certain wild *P. tuberosus* taxa.

In addition, the positioning of TW accessions in separate regions of clade 1 (clustering either with accessions of *P. tuberosus* or *P. panamensis*) would suggest that these wild populations are in fact true remnants of Andean origin (and not man-mediated introductions or escaped forms), which in turn gave rise to the diversity within the *P. tuberosus* complex and possibly *P. ahipa*. It should be noted that RFLP variation across the cpDNA genome appears to suggest the presence of two different lineages of TW taxa placed in widely separated regions of clade 1. A first TW lineage was resolved together with *P. panamensis*, while a second lineage was present in the *P. tuberosus* complex (Figure 3.7).

A third subcluster in clade 1 contained accessions of wild status only, i.e. of *P. panamensis* (PW) and *P. tuberosus* (TW) taxa (Figure 3.7). This subclade, although present in only 40% of the bootstrap replicates (decay index ≤ 94), occurred in all equally parsimonious trees at basal nodes. Hence, it is possible that both species originated from the same early ancestor (i.e. *P. ferrugineus*) and have later diverged parapatrically as a response to environmental changes (deciduous vs. evergreen rainforest) and, in the case of *P. tuberosus*, subsequent cultivation and man-made selection.

The similarity of PW and TW taxa identified by cpDNA data in this third subcluster could possible be suggestive of a phenomenon of vicariance, with two vicariants (i.e. similar taxa occupying current separate geographical/ecological areas) represented by *P. panamensis* and wild *P. tuberosus*. Stace (1989) explains that vicariants may arise in various ways; however, in the light of the current data two hypotheses apply: **(1)** a

formerly and continuously distributed taxon (i.e. ecotypes of *P. ferrugineus*) becoming separated into different geographic areas and there undergoing divergent evolution; or, **(2)** parallel evolution of two species (i.e. PW and TW taxa) from a common ancestor (*P. ferrugineus*) in two separate areas (deciduous and evergreen rainforests in southern regions of Mesoamerica and northern areas of South America, respectively).

A final subclade, comprised only of accession FW237, occurred as the most basal (i.e. primitive) taxon within clade 1. Surprisingly, both FW accessions examined in this study did not cluster together, suggestive of intraspecific cpDNA diversity (see later), but were resolved always at basal, rather than terminal, nodes of the cladogram.

Clade 2: A Mesoamerican evolutionary branch

All *Pachyrhizus* taxa located in the second, strongly supported clade of the strict consensus tree have a Central American distribution, i.e. occur from Costa Rica to Mexico, except for EC565, an accession introduced to the Philippines. Outgroup OUTce was also resolved within clade 2. The anomalous positioning of this outgroup might be a consequence of an artefact of the scoring method used (rather than a phylogenetic event *per se*). Genomic DNA of the three outgroups only became available later in the analysis and, consequently, data of their respective luminographs (which included the same size marker and additional *Pachyrhizus* taxa to enable correct band size evaluation) were incorporated in a subsequent stage. Care was taken to ensure that fragments of the same size and detected with adjacent probes were scored only once. Despite all efforts, non-independent fragments might apparently have been scored resulting in an artificial increase in homology. Therefore, OUTce will be ignored for the purposes of this discussion.

Within clade 2, a subclade comprising five EC accessions with identical cpDNAs was evident, although only a moderate bootstrap value of 58% and a decay index ≤ 94 supported this grouping. These five accessions did not differ in their cpDNA, despite their different geographical origin/provenance and the fact that they represent different landraces. Therefore EC502 (= var. Cristalina) and EC558 (= Nayarit type) were both

placed within this subclade. This low degree of differentiation may reflect one of the potential limitations in using cpDNA at the generic level, i.e. the lack of sufficient variation (Crawford, 1990). Thus, while the conservative nature of cpDNA removed convergent site gains and losses as a problem, this same conservatism prevented a complete resolution of relationships among all taxa studied.

Interestingly, accession EC032 was not resolved within the EC subclade described above. In contrast, this accession collected in the Yucatán peninsula, Mexico, was positioned at a more basal level, strongly supported by 80% of the bootstrap replicates and a decay index of ≤ 96. Therefore, two separate 'lineages' of cultivated *P. erosus* were evident within the consensus tree: firstly a group of cultivars from southern Mexico and regions further south in Central America, and second, a cultivar(s) from the Yucatán peninsula. This finding concurs with field observations in Central Mexico and Yucatán by Sørensen (1996), where a high uniformity of cultivars from these latter areas was evident, in contrast to the considerable diversity observed in southern localities. The same author suggested that such uniformity within the Central Mexico and Yucatán regions might be the result of earlier introductions from southern localities, as also observed recently in a RAPD analysis (Estrella *et al.*, 1998). However, the current cpDNA data, though supporting the presence of these two groups of cultivated *P. erosus*, did not explain satisfactorily the hypothesis of an early origin of *P. erosus* in strictly southern localities, since a wild accession from Mexican origin/provenance (i.e. EW203, from a northern locality) was resolved as a more basal branch (clade 2; Figure 3.7).

Finally, clade 2 also contained the two EW accessions analysed in this study although they did not cluster together; moreover, both taxa were separated by nodes containing cultivated *P. erosus* accessions (as described above), suggestive of different centres of origin and domestication for this Mesoamerican species, or - alternatively - an instance of reticulate evolution (see later). This range of cpDNA diversity observed between wild and cultivated *P. erosus* taxa has also been detected frequently in other wild ancestor-cultivated descendant combinations in many agronomically important crop species (Doebley, 1992; Llaca *et al.*, 1994).

One accession of *P. ferrugineus* (FWLoc7, collected in Guatemala) was resolved as the most basal taxon within *Pachyrhizus*. *P. ferrugineus* is known to be a highly divergent species comprising several ecotypes (Sørensen, 1990 & 1996). Similarly, a significant level of intraspecific cpDNA variation was also detected, as evidenced by accessions of this species occurring in widely separated regions of the cladogram (Figure 3.7).

The occurrence of such intraspecific variation may be due to the wide geographic distribution of this wild species during an early evolutionary stage, leading to the production of distinct intraspecific cpDNA races as populations diverged. Current cpDNA data would preliminary suggest the presence of (at least) two separate lineages within *P. ferrugineus*. In parallel to this molecular evidence, morphological studies (Sørensen, 1988; Døygaard & Sørensen,1998) could support similar 'lineages' by traits such as the significant variation in leaflet outline (even within populations) and the wide distribution of the species in ecosystems with ample soil and habitat variability ranging from evergreen to deciduous rainforests. Further surveys of cpDNA variation encompassing a greater number of FW ecotypes might help determine whether the proposed evolutionary pathways are correct.

Intraspecifc cpDNA variation

Some intraspecific cpDNA variation was detected during this study, and resulted in accessions of the same species occurring in separated regions of the strict consensus tree (e.g. FW and TW taxa in Figure 3.7). Such cpDNA variation may reflect the wide geographic distribution of accessions within the species concerned, having originated during periods when populations within taxa were isolated from each other, thus allowing molecular differences to become more evident (Gillies & Abbott, 1996; Soltis *et al.*, 1992). On the other hand, it is also possible that the cpDNA diversity found in these species is the result of interspecific cytoplasmic gene flow, leading to the replacement of a species cytoplasm in certain populations with that of another species. Chloroplast, and therefore cpDNA, capture via hybridisation and

introgression is now recognised as a widespread phenomenon in plants (Avise, 1994 and references therein), and has been shown to be the cause of intraspecific cpDNA variation in a broad range of taxa.

Species/complex specific fragments

FOA analysis revealed a total of 34 restriction fragments (Table 3.4) that uniquely identified a species or a group of species. For instance, 14 fragments were specific to *P. ferrugineus*; six were specific to members of both *P. ahipa* and the *P. tuberosus* complex; and, one fragment was exclusive to accessions of *P. panamensis*. Taxa were identified either by: **(1)** the presence of unique bands; or, **(2)** the absence of bands that were present in all other taxa. In this latter case, absent bands were an indication of a unique restriction site mutation in a particular accession, with the resulting bands present in regions further down the relevant luminograph or, occasionally, not detected.

The use of an increased number of restriction enzymes could evidently aid identification of further molecular markers within the *Pachyrhizus* species examined. Such molecular markers could be employed for accurate cultivar identification, efficient selection of parents for the development of new varieties, and IPR (intellectual property rights) protection of new cultivars. These markers might also be important in the field of conservation biology by revealing phylogenetic relationships within and among rare or endangered *Pachyrhizus* populations. Such phylogenetic assessments can range from parentage evaluations in captive breeding programmes to the identification of major sources of regional phylogeographic diversity around which management guidelines and natural reserves might be established.

3.6.2. Restriction site variation within a PCR-amplified cpDNA region

Nature of cpDNA variation in *Pachyrhizus*

This analysis was conducted on a subset of the accessions included in the survey of RFLP variation across the cpDNA genome. A total of 20 *Pachyrhizus* accessions and one outgroup (OUTcc) yielded a satisfactory PCR product for subsequent restriction digestion; the remaining accessions either failed or yielded a very faint PCR product.

A maximum of four site mutations was obtained with any one of the six restriction enzymes used in this survey (Table 3.7), which made scoring of fragment variation straightforward, therefore limiting homoplasy levels. It was possible to use both FOA and SOA approaches on data obtained from the PCR-RFLP analysis to reconstruct evolutionary trees, despite the low level of synapomorphic characters and the reduced number of base pairs sampled (35% and 568, respectively). However, all accessions of wild provenance (i.e. FW, PW and TW) showed better resolution after analysis of restriction site mutations and, therefore, the phylogenetic tree based on the SOA approach is discussed in this section.

Phylogenetic relationships

The strict consensus tree produced from the SOA data set (Figure 3.11) was not fully resolved, but nonetheless enough variation was found in the five cpDNA regions amplified to resolve two major clades.

Clade 1: A South American evolutionary branch

Clade 1 comprised a total of 14 accessions, most of them of South American distribution, except for FWLoc7 and PW055 originally collected in Mesoamerica. The clade was clearly separated from the rest of the genus and appears to have split away early in the evolutionary history of *Pachyrhizus*. It contained all accessions of

P. ahipa, the *P. tuberosus* complex and *P. panamensis,* as well as one accession of *P. ferrugineus* (FWLoc7) that represented the basal-most species within this clade.

Clade 1 in this study comprised the same species/accessions as its counterpart obtained after analysis across total genomic cpDNA (Figure 3.7). However, there were some striking contrasts in arrangement, as follows:

- All *P. ahipa* accessions clustered together, but were resolved as a more basal subclade to the remaining taxa (a high bootstrap value of 100% and a decay index ≤ 29 supported this hypothesis of monophyly; Figure 3.11). *P. tuberosus* was not resolved as a sister taxon, as suggested by the analysis across the total cpDNA genome.

- The two PW accessions did not cluster together, although their placement was still at basal-most nodes within the main subclade of clade 1. Surprisingly, accession PW055 clustered together with a chuin, TC354 (69% bootstrap value; decay index ≤ 28), and both had identical cpDNAs. Again, this phylogenetic relationship could be a reflection of intraspecific cpDNA variation or, alternatively, interspecific cytoplasmic gene flow, as discussed previously.

- The *P. tuberosus* complex was poorly resolved, as evidenced by low bootstrap values; furthermore, the subclade containing it collapsed after only one step away from the minimal tree (Figure 3.12). In general, placement of *P. tuberosus* accessions examined within the complex was anomalous, impaired by polytomies present at terminal nodes. It is clear that more restriction site characters are needed to aid resolution among the cultigen types of this species. However, it is noteworthy to mention that all TW accessions were grouped together (although poorly supported by 42% bootstrap value) and showed identical cpDNAs. This clustering contrasts markedly when compared with the dispersed placement of TW taxa within the strict consensus tree produced from the RFLP analysis across the total cpDNA genome (Figure 3.7).

- A final subcluster comprised one accession of *P. ferrugineus*, as the most basal (i.e. primitive) taxon within the clade. Surprisingly, the strict consensus tree showed accession FWLoc7 in this node instead of FW237 (Figure 3.11), as suggested previously by the analysis across the cpDNA genome. However, both FW accessions were resolved always at basal, rather than terminal, nodes of the cladogram.

Clade 2: A Mesoamerican evolutionary branch

All *Pachyrhizus* taxa located in a second clade of the phylogenetic tree produced from PCR-RFLP data have a Central American distribution, i.e. from Costa Rica to Mexico, except for accession TC350, a chuin collected in the Peruvian rainforest. All *P. erosus* accessions clustered tightly together as a monophyletic group (76% bootstrap value); in addition, the node supporting monophyly between TC350 and the *P. erosus* accessions was moderately supported by a bootstrap value of 50% and a decay index ≤ 28.

Within clade 2, a small subclade comprising all accessions of *P. erosus* was evident, with EW203, from Mexico, resolving as the most basal taxon in this subclade. On the basis of these cpDNA data, no difference existed between EC032 (from the Yucatán peninsula) and EC509 (Costa Rica), despite their different geographical provenance and the fact that they represent different landraces. This situation contrasts with the phylogenetic relationship proposed by the previous analysis across total genomic cpDNA, where accession EC032 was resolved as a separate phylogenetic entity to most *P. erosus* taxa, suggestive of two groups of cultivated *P. erosus*.

Finally, one accession of *P. ferrugineus* was the most basal taxon within the Mesoamerican evolutionary branch, namely FW237 (Figure 3.11).

Intraspecifc cpDNA variation

As in the analysis of RFLP variation across the cpDNA genome, a degree of intraspecific cpDNA variation was also detected by the PCR-RFLP analysis, and resulted in accessions of the same species occurring in separated regions of the strict consensus tree (e.g. *P. panamensis* accessions; Figure 3.11). Such cpDNA variation may reflect the wide geographic distribution of ancestral species, having originated during periods when populations within taxa were isolated from each other, which allowed molecular differences to become more conspicuous. On the other hand, cpDNA capture via hybridisation and introgression cannot be ruled out.

In the case of *P. ferrugineus*, for example, both accessions examined in this study did not cluster together, suggestive again of intraspecific cpDNA diversity or an instance of reticulate evolution (see later); however, both were resolved always at basal, rather than terminal, nodes of the cladogram. Alternatively, cpDNA data appear to suggest the presence of separate ancestral lineages of *P. ferrugineus* within the evolution of the genus.

Species/complex specific fragments

FOA analysis revealed a total of 11 restriction fragments (Table 3.8) that uniquely identified a species or a group of species. For instance, three fragments were specific to *P. ferrugineus*; two were specific to members of both *P. panamensis* and the *P. tuberosus* complex; and, two fragments were exclusive to wild accessions of *P. tuberosus*. Again, taxa were identified either by: **(1)** the presence of unique bands; or, **(2)** the absence of bands that were present in all other taxa. These absent bands were an indication of a unique restriction site mutation in a particular accession, with loss of the resulting bands due to their small size.

3.6.3. An approach of combined restriction site data analysis

An obvious consideration to expand the level of phylogenetic resolution and to aid identification of further molecular markers is the use of an increased number of restriction enzymes. Although low levels of divergence in some cases precluded formulation of robust hypotheses of interspecific relationships, chloroplast genomes were grouped into two well resolved clades by means of site occurrence analysis of RFLP variation across total genomic cpDNA and of PCR-RFLP analysis, respectively (Figures 3.7 and 3.11). So far, the two resulting cpDNA genealogies have identified two major evolutionary branches that were largely congruent, and both are also compatible with species groupings based on crossing relationships and morphologic variation (Døygaard & Sørensen, 1998; Márquez & Morera, 1992; Sørensen, 1988).

An analysis combining the cpDNA data matrices of the two previous phylogenetic treatments (sections 3.3 and 3.4) was justified because the higher-order structure of the trees produced by each independent analysis was similar, the ultimate aim being to investigate if the resolution of species relationships might be improved.

Phylogenetic relationships

Cladistic analysis of the combined cpDNA data matrix produced a highly resolved strict consensus tree (Figure 3.14), that contained four major plastome clades. Most polytomies observed in either of the two previous phylogenetic analyses were no longer present after combining the two data sets. Bootstrap values and decay indices indicated that the internal structure of the tree was stable, demonstrating - therefore - that the combination of characters improved resolution of evolutionary relationships in the genus.

Clade 1: A South American evolutionary branch

The South American evolutionary branch of the genus was contained within clade 1 and comprised a total of 13 accessions. This clade, which appears to have split away

early in the genealogy of the genus, contained all accessions of *P. ahipa*, of the *P. tuberosus* complex and of *P. panamensis*. This latter species represented the most basal taxon within the clade. Clade 1 comprised three subclades with the same species/accessions groupings obtained previously in the two independent cpDNA analyses. Within this clade it was evident that:

- All *P. ahipa* taxa clustered strongly together (93% bootstrap; decay index ≤ 104) and were resolved with one accession of *P. tuberosus* (TC553, a jíquima) as sister taxon; this grouping was resolved as a more basal subclade to the *P. tuberosus* complex. The affinity between *P. ahipa* and *P. tuberosus* depicted by the consensus tree of combined data is interesting from the evolutionary standpoint, since, of all cultivated yam bean species, *P. ahipa* is so far the only one known for its absence of wild ancestral material; moreover, within this latter species both determinate and indeterminate genotypes exist, resembling some of the cultivars of the *P. tuberosus* complex. This could support the hypothesis that *P. ahipa* is a species derived from *P. tuberosus*, where probably the primary selection criterion during subsequent domestication was its use in the cropping systems of the high Andean terraced fields (none of the cultivar groups of the *P. tuberosus* complex are cultivated at altitudes above 1800 m a.s.l.; the jíquima is a lowland cultivar, whereas the present main cultivation of *P. ahipa* takes place above 2000 m a.s.l.).

- The *P. tuberosus* complex showed better resolution, with all cultivar forms placed at terminal nodes (Figure 3.14), while wild *P. tuberosus* taxa resolved more basally. However, the three cultivar types were resolved still into a polytomy (bootstrap value of 25%) with no clear distinction among them.

- Both PW accessions clustered together (40% bootstrap; decay index ≤ 104) and their placement was still basal-most relative to clade 1, which concurs with the pathways proposed by the analysis across the total cpDNA genome (Figure 3.7). *P. panamensis* in the strict consensus tree was resolved as an independent clade, basal to the remaining species within the South American evolutionary branch (Figure 3.14).

Clade 2: A Mesoamerican evolutionary branch

This Central American evolutionary branch was strongly supported in the cladogram (bootstrap values ranging from 91 to 100%) and was identical to the one produced by the PCR-RFLP based phylogeny reported previously. Again, accession TC350, a chuin collected in the Peruvian rainforest, was resolved as the most basal taxon within this clade. Interestingly, the node supporting the placement of this accession was present in 100% of the bootstrap replicates, with a decay index ≤ 108 (Figure 3.14). *P. tuberosus* might, therefore, be a possible progenitor of two groups: **(1)** *P. ahipa* (see clade 1); and, **(2)** *P. erosus*, which during its evolution underwent dispersal towards the central/northern parts of Mesoamerica and, then, to Mexico.

Clades 3 and 4

The two accessions of *P. ferrugineus* examined, FW237 and FWLoc7, represented clades 3 and 4, respectively. In contrast to the previous independent phylogenetic treatments, the combined analysis of cpDNA data placed these two accessions consecutively at the most basal nodes in the phylogeny, supported by high bootstrap percentages (Figure 3.14).

Intraspecifc cpDNA variation in the combined analysis

As expected, a degree of intraspecific cpDNA variation was also detected in this study, evident by accessions of the same species placed in separated regions of the strict consensus tree; this situation applied to taxa of *P. ferrugineus* and *P. tuberosus* (Figure 3.14). Owing to a strong propensity for hybridisation in many plant taxa, the possibility of widespread reticulate evolution (Grant, 1981; Avise, 1994) cannot be discarded. In this phenomenon, phylogenies might be characterised as anastomotic (i.e. netlike), rather than strictly dichotomous and branching, with a disruption of the expected hierarchical patterns normally established by taxonomy, systematics, etc.

In the three sections of this chapter, the cpDNA molecule could be revealing such cases of reticulation because its clonal transmission allows particular ancestral sources to be identified without the complication of recombination that can lead to a mosaic ancestry for the nuclear genome. In addition, there might have been several biological factors that facilitated cytoplasmic (relative to nuclear) exchange across *Pachyrhizus* species. Several authors (e.g. Doyle, 1992; Carlson & Chelm, 1986; Rieseberg & Soltis, 1991) have reviewed the literature on instances of intra- and interspecific cpDNA capture attributable to introgressive hybridisation and concluded that reticulate evolution is indeed a widespread phenomenon in plants.

A major concern that arises is whether the gene trees presented here reflect the organismal phylogeny (the issue of gene tree versus species tree). Assuming that the genes compared are truly homologous, gene trees and organismal phylogenies can differ because of retention of ancestral polymorphisms, or reticulation among populations (i.e. gene flow) or species (i.e. hybridisation). Therefore, conclusions drawn from non-recombining genetic systems, such as the cpDNA genome, should be treated with caution, because the effects of reticulation are potentially retained through subsequent generations. In addition, variation among loci might also be expected because of stochastic variation (Ball *et al.*, 1990) and differences in effective population size (e.g. organellar genomes versus nuclear genes). Thus, differences in gene trees among populations or closely related species can also arise because of lineage sorting effects (e.g. Slade *et al.*, 1994).

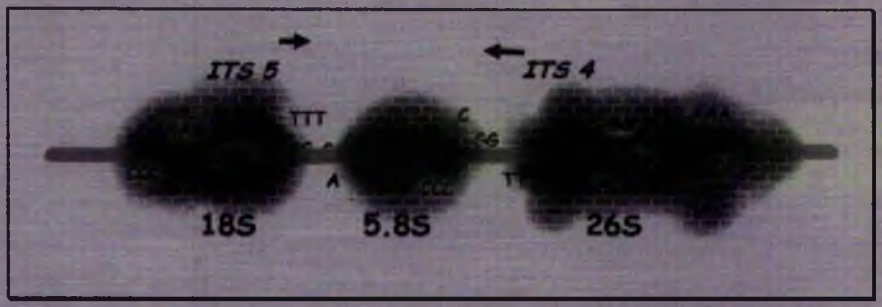
Both theory and practice suggest that these effects can be (partially) overcome by combining data across a larger number of loci (Pamilo and Nei, 1988; Slade *et al.*, 1994). In this study, the PCR-RFLP based analysis brought forward high congruencies as well as some ambiguities with the evolutionary pathways proposed by the survey of RFLP variation across the total chloroplast genome. This added a compelling rationale for combining data from different sources to achieve a more accurate phylogenetic reconstruction in *Pachyrhizus*.

Despite the phenomena described above, it is important to note that, in the light of the current cpDNA evidence, *P. tuberosus*, as a phylogenetic entity, appears to play an important role in the relationships so far described. The consensus trees are clearly suggesting that *P. tuberosus* might not only be present in the South American evolutionary branch: an additional lineage of this species was probably involved in the early parentage of *P. erosus*, in which case the separation of these two lineages must have occurred a long time ago, given the number of cpDNA mutations which now separate them.

In summarising this section, it is now clear that phylogenetic phenomena in *Pachyrhizus* will be best analysed and confirmed through multiple lines of evidence involving different sources of molecular (and other) markers. Barriers to reproduction between closely related taxa seldom are absolute (Avise, 1994) and often appear differentially 'semipermeable' to cytoplasmic and various nuclear alleles. Thus, a varied fabric of gene genealogies (rarely evident from morphological assessment alone) will help clarify and confirm the evolutionary pathways and instances of reticulation discussed before. The following chapters are focused in that direction.

CHAPTER 4

VARIATION WITHIN THE ITS REGION OF NUCLEAR rDNA TO INFER EVOLUTIONARY PATHWAYS IN *PACHYRHIZUS*



The ITS Region - graffiti.
J. Estrella

4.1. Introduction

All organisms have DNA sequences within their genomes that code for ribosomal RNAs (rRNAs), which are essential components for cellular protein synthesis. In plants, ribosomal DNA (rDNA) is found within the nuclear, mitochondrial and chloroplast genomes. The ubiquity of rRNA in organisms and the development of techniques for determining the primary nucleotide sequence of the DNA that encodes rRNA molecules have prompted the use of rRNA as a tool for inferring evolutionary pathways of plants. However, not all regions of DNA evolve at the same rate, and consequently some regions are useful for comparisons at or below the genus level, while other regions are only useful at the family level or above (Hamby & Zimmer, 1992).

Two rRNA gene families occur within the nuclear genome as tandemly repeating arrays. The first encodes the ribosomal rRNA 18S-5.8S-26S subunits, while the second encodes the 5S subunit. It is thought that both arrays were originally clustered into one transcription unit but have separated during evolution, with the 5S gene family breaking from this initial structural arrangement by unequal crossing over (Gerbi, 1986). In the work reported in this chapter, attention was focused on the internal transcribed spacer (ITS) within the 18S-5.8S-26S gene family.

4.1.1. Structure, function and organisation of nrDNA

In higher plants, nrDNA that encodes the 18S-5.8S-26S subunits is a mid to highly repetitive DNA sequence arranged in tandem repeats at loci on one or more chromosomes. Only among closely related species are chromosomal locations similar. Each repeat unit contains a transcribed region which is separated from adjacent repeat units by a long, non-transcribed intergenic spacer (IGS) region (Hamby & Zimmer, 1992). There are three highly conserved ribosomal RNA regions and two non-coding spacer regions within the transcribed region of nrDNA. These three highly conserved regions that comprise the ribosomal RNA gene are arranged in a 5'- 18S-5.8S-26S -3' order with the spacers (designated as ITS1 and ITS2) flanking the 5.8S region (see

Figure 2.1 in chapter 2). In addition, an external transcribed spacer region (ETS) is located at the beginning of the 5' end of the transcribed unit. The whole region is transcribed as a single large precursor, and is subsequently processed into the 18S, 5.8S and 26S functional rRNA forms (Rogers and Bendich, 1987). The intergenic spacer (IGS) is, for convenience, divided into three subregions, i.e. a series of tandem subrepeats, which are flanked by a 3' end on one side and a 5' end at the other side (Appels and Dvorak, 1982). This repeated sequence varies interspecifically in length generally from 100 to 200 bp, while within species its length normally varies slightly (Hamby & Zimmer, 1992; Jorgensen & Cluster, 1988). The variation in length within the IGS is due to alteration in the number and length of the subrepeats it contains.

The main function of rRNAs is in protein synthesis. It was previously thought that rRNAs served primarily as a scaffolding for ribosomal proteins, but more recent evidence suggests that rRNA molecules are the basic functional element of the ribosome and that the proteins serve to mediate interactions between mRNA, tRNA and rRNA (Gerbi, 1985; Hamby & Zimmer, 1992). The ITS region is part of the transcriptional unit of nrDNA, but the spacer segments of the transcript are not incorporated into mature ribosomes. Instead, ITS1 and ITS2 regions of the nrDNA transcript appear to function, at least in part, in the maturation of nrRNAs (Baldwin *et al.*, 1995).

In vivo mutational analyses in yeast (*Saccharomyces cerevisiae*) indicate that deletions of certain regions within ITS1 can inhibit the production of mature small and large subunit rRNAs (Baldwin *et al.*, 1995 and references therein), whereas certain deletions or point mutations in ITS2 prevent or reduce processing of large-subunit rRNAs. In addition to their role as self-splicing group I introns, it seems probable that ITS1 and ITS2 are under some evolutionary constraint in structure and sequence, as suggested by size and G+C content comparisons among angiosperms (Baldwin *et al.*, 1995).

Tandem repeats of nrDNA are usually arranged in a 'head-to-tail' configuration with the total length of a repeat array ranging from 6.0 to 18.5 kb (Avise, 1994; Appels & Honeycutt, 1986). One percent or more of the nuclear genome may be rDNA and

from 1000 to 10000 copies of the sequence can exist in a plant cell. In fact, it can comprise as much as 10% of the total plant DNA (Hemleben *et al.*, 1988; Jorgensen & Cluster, 1988). The rDNA copy number in plants can be as much as 20-fold greater than that within animal genomes, and may also vary up to four-fold within a species. The reason for this high level of variation is unknown, but it has been suggested that it may be a response to environmental stress (Rogers and Bendich, 1987).

4.1.2. Evolution of nrDNA

A remarkable feature of rDNA is the overall sequence homogeneity among members of the gene family. If all parts of the genome were evolving independently, comparisons of nucleotide sequence between members of the same gene family within a species should show a similar level of divergence as would comparisons of the same gene between two closely related species, assuming that the duplication events creating the gene family preceded the divergence of the two species. However, studies consistently show that this is not the case for rDNA (Arnheim cited by Hamby and Zimmer, 1992).

In a study of *Xenopus laevis*, Brown *et al.* (1972) first demonstrated by hybridisation tests that the several hundred rDNA repeats were essentially identical at both the coding and the intergenic regions, but that when the rDNAs were hybridised to those of *X. borealis* a much lower level of overall similarity was found. Whereas the coding regions were still highly conserved, the IGS regions were sharply divergent, although within each species the IGS was conserved. This same motif of conserved coding regions and non-conserved intergenic spacers with species-specific mutations has been identified in the rDNA of all species studied (Dover & Flavell, 1984). The process by which this pattern of intraspecific homogeneity and interspecific heterogeneity is maintained was earlier known as horizontal evolution (Brown *et al.*, 1972) or coincidental evolution (Hood *et al.*, 1975), and nowadays is called concerted evolution (Zimmer *et al.*, 1980).

The mechanisms of concerted evolution, i.e. tandem repeat units in each rDNA array evolving at the same rate, are primarily unequal crossing over or unequal exchange (Figure 4.1), and gene conversion (Figure 4.2). To achieve overall homogeneity, one or both of these processes (and possibly others) must take place within each individual locus, between rDNA loci on homologous chromosomes, and between rDNA loci on non-homologous chromosomes (Dover, 1982; Arnheim, 1983). Several studies have shown that the processes of unequal exchange and gene conversion alone or combined can eventually lead to the fixation of a mutant gene within a population, even with only one or a few original copies of the mutant (Smith, 1974; Hamby & Zimmer, 1992 and references therein).

Theoretically, gene conversion can proceed in either direction when a heteroduplex is recognised, that is, the mutant may be converted to wild type or vice versa. However, if there is even a small bias in one direction or the other, the rate of concerted evolution can be increased significantly. In this sense, the term molecular drive (Dover, 1982) has been coined to describe the process of gene family homogenisation and fixation due to unequal crossing over and biased gene conversion. Transposition may also play an important role in molecular drive, but it has not yet been demonstrated as a mechanism in the concerted evolution of rDNA families.

4.1.3. ITS sequence comparisons and evolutionary rates

In all flowering plants reported to date, ITS1 and ITS2 sizes fall within a similar range (Baldwin *et al.*, 1995), with lengths of less than 300 bp for both spacers (ITS1 ranging from 187 to 298 bp, and ITS2 from 187 to 252 bp), in contrast to much longer spacers reported in other eukaryotes, e.g. some vertebrates (Stewart *et al.*, 1983, in a study of *Xenopus*; Goldman *et al.*, 1983, in mice). Given the fact that the 5.8S subunit has been reported invariant in length (mostly 163 or 164 bp), the entire ITS region appears to be normally under 700 bp in angiosperms.

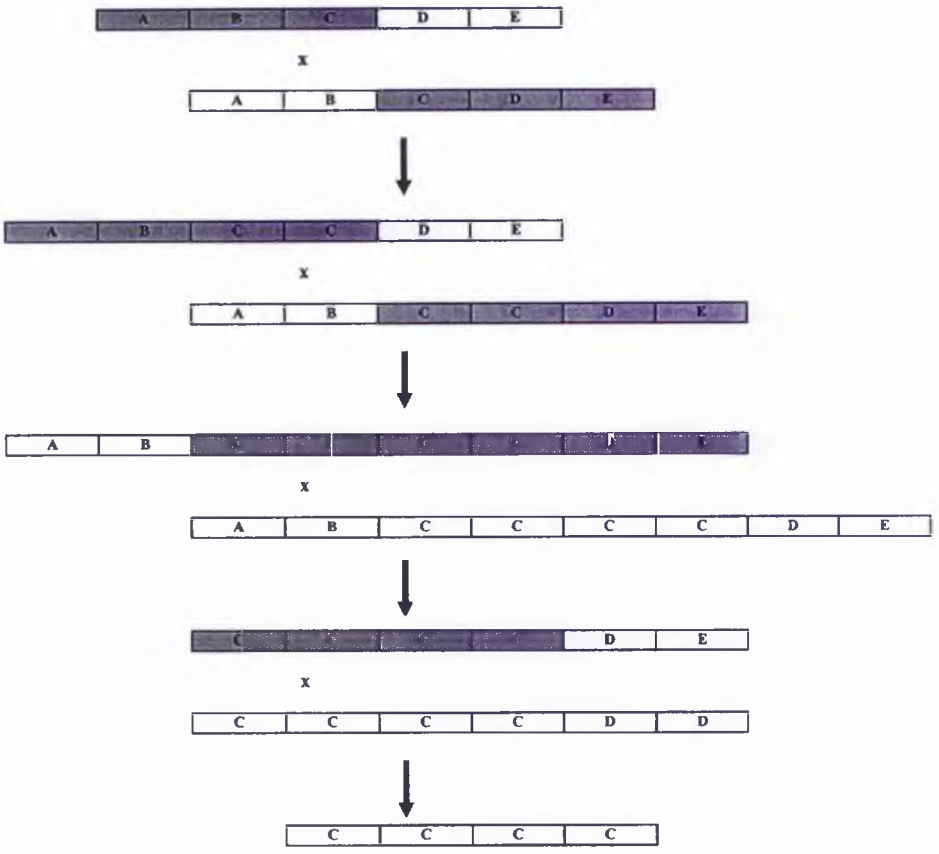


Figure 4.1. Concerted evolution by unequal crossing over (after Li & Graur, 1991). A hypothetical model within a multigene family composed of five variants (**A** to **E**) is shown. Shaded repeat units will form recombinant repeat units in daughter chromosomes; repeated cycles of unequal crossing over events cause the duplicated genes on each chromosome to become progressively more homogenised. For example, after an initial unequal change, a duplication of the **C** type unit (**A-B-C-C-D-E**) occurs in one daughter chromosome while the other chromosome loses the **C** type (not shown). As this process is repeated, the daughter chromosome will become more homogenised and finally only one type of the multigene family (**C** type) will spread to fixation throughout the gene family.

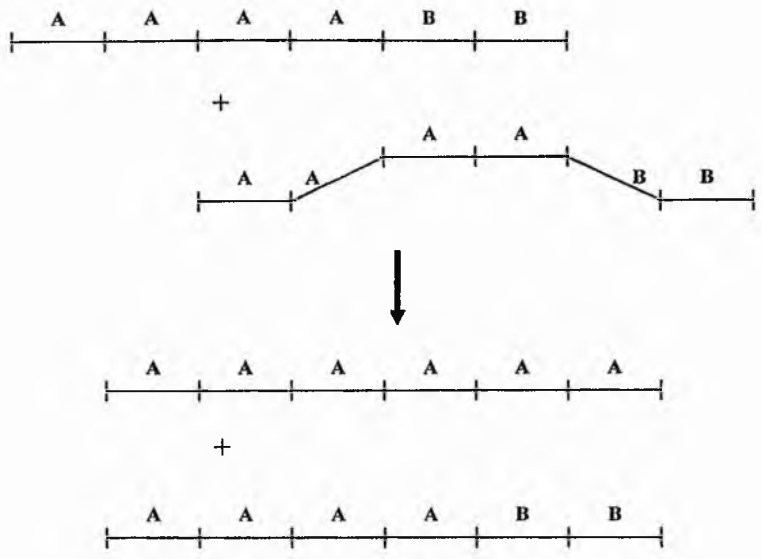


Figure 4.2. Concerted evolution by gene conversion, a non-reciprocal recombination process in which two sequences interact in such a way that one is converted by the other (after Li & Graur, 1991). A hypothetical non-allelic model is shown where two wild type repeats are converted into a mutant type. As a result, the first daughter gene family becomes more homogeneous than the parental gene family while there is no change in the second daughter gene family. In this example, the B repeat type of the first chromatid was converted into the A type; as a result, one daughter chromatid has all the A repeat type while the other chromatid maintained the parental repeat type. Thus, gene conversion changes the frequencies of the two types of repeats in only one of the daughter chromosomes, but does not alter the total number of repeats in either chromosomes.

Relative sizes of the two spacers vary between and, in at least some cases, within families, with little indication of a broad-scale phylogenetic pattern (Baldwin *et al.*, 1995; Hamby & Zimmer, 1992). Indeed, ITS1 is consistently longer than ITS2 (or rarely equal in length) in all available sequences of, for example, Asteraceae, Brassicaceae, Malvaceae, Onagraceae, Ranunculaceae, Salicaceae, Saxifragaceae and Winteraceae. Conversely, ITS2 is larger than ITS1 in all sequences of Betulaceae, Cucurbitaceae, Scrophulariaceae and Viscaceae reported to date. On the other hand, the two spacers are nearly equal in length in the reported sequences from the

Solanaceae, whereas in the large families Fabaceae, Poaceae and Rosaceae ITS1 may be longer or shorter than ITS2.

In most investigated groups of angiosperms, ITS1 pairwise divergence values are similar on average to those of ITS2, with the greatest disparity in divergence comparisons of ITS1 and ITS2 occurring in the Asteraceae (Baldwin *et al.*, 1995). Distances between ITS1 sequences of taxa in the Hawaiian silversword alliance (Madiinae) are generally similar to ITS2 distances from the same species comparisons. In contrast, ITS1 pairwise divergence values in *Krigia* (Lactucaceae) are twice those of ITS2 (Kim & Jansen, 1994). Other reported Asteraceae fall largely between these extremes. Average pairwise distances between ITS1 sequences exceed those between ITS2 sequences by 1.3 to 1 in *Calycadenia* (Baldwin, 1993) and 1.5 to 1 in a broader study of Madiinae (Baldwin, 1992). Outside Asteraceae, *Epilobium* (Onagraceae) and *Gossypium* (Malvaceae) show much higher pairwise divergence between ITS1 sequences than between sequences of the ITS2 (Baum *et al.*, 1994; Wendel cited by Baldwin *et al.*, 1995).

However, Baldwin *et al.* (1995) do not consider these comparisons of average pairwise distance values to be an adequate measure of relative evolutionary rates. Moreover, relative rate tests (e.g. Muse & Weir, 1992) are better suited for this purpose. Average distances are based on a set of non-independent comparisons because they do not take phylogenetic relationships into account; this non-independence imparts some doubt on the statistical significance of differences in distance values. Furthermore, average distances mask variance in relative distances, which is considerable in these studies.

ITS sequences appear to have evolved more slowly in some ancient woody groups than in herbaceous, primarily annual taxa of comparatively recent origin (Baldwin *et al.*, 1995; Hamby & Zimmer, 1992). Similar levels of ITS sequence divergence were found between taxa in lineages that diverged in the early Tertiary or Cretaceous, e.g. genera of the Winteraceae (Suh *et al.*, 1993) and subgroups within *Notofagus* (Manos, 1993), and between taxa in herbaceous lineages that presumably diverged in the

Pliocene or Pleistocene (e.g. genera of the Madiinae; Baldwin, 1992). Such correlations between plant life-form and apparent rates of molecular evolution have been noted from cpDNA data as well (e.g. Brunsfeld *et al.*, 1994; Clegg & Zurawski, 1992), although the basis for this pattern is unclear. A generation-time explanation for differences in molecular evolutionary rates (with longer generation times resulting in slower molecular evolution), although somewhat consistent with patterns observed in animals, is problematical in plants, wherein open development might allow fixation of mutations in vegetative meristems as readily as in reproductive cells that arise from such meristems (Klekowski, 1988; Baldwin *et al.*, 1995). Additional attention by molecular evolutionists is needed to achieve a more detailed description of these complex patterns of evolutionary rates.

4.1.4. Favourable properties of the ITS region as a tool for evolutionary studies

Several general properties of the ITS region encourage its use for phylogenetic studies in angiosperms. First, along with the other members of the nrDNA multigene family, the ITS region is highly repeated in the plant nuclear genome. This abundance in copy number promotes detection, amplification, cloning and sequencing of nrDNA (Baldwin *et al.*, 1995). Second, and very significant for phylogeny reconstruction, this gene family undergoes rapid concerted evolution, promoting intragenomic uniformity of repeat units and, therefore, accurate reconstruction of species relationships from these sequences (Hamby & Zimmer, 1992). As a result, direct sequencing of pooled nrDNA PCR products can be used to extract phylogenetic information in many species (Baldwin *et al.*, 1995). Moreover, concerted evolution and sexual recombination may promote nrDNA uniformity within interbreeding populations and thereby minimise the importance of intrapopulational sampling in phylogenetic studies.

Third, the small size of the ITS region (< 700 bp) and the presence of highly conserved sequences flanking each of the two spacers make this region easy to amplify, even from herbarium material, using universal eukaryotic primers (White *et al.*, 1990; Figure 2.1). In many instances, these primers have been used to generate single stranded DNA for sequencing directly from genomic DNA, bypassing a

separate double stranded DNA amplification procedure (Wojciechowski *et al.*, 1993; Baldwin *et al.*, 1995). This same approach can be used even in more stream-lined double stranded sequencing methods, e.g. cycle sequencing, generating high quality ITS sequences, normally without evidence of divergent repeat-types within individuals (Baldwin *et al.*, 1995). However, such homogeneity within individuals does not always occur and might result occasionally in sequence variants within DNA accessions, as reported, for example, in *Gentiana* (Yuan *et al.*, 1996) and in *Saintpaulia* (Möller & Cronk, 1997).

4.1.5. Use of rDNA ITS sequence variation in plant evolution and systematics

In general, nuclear rDNA, particularly the ITS region, has proved to be valuable in phylogeny reconstruction and in the study of reticulate evolution and the origin of polyploids. Some examples of its use in investigating these different aspects of plant evolution are discussed below.

Phylogenetic reconstruction

Available data indicate that ITS sequences are phylogenetically useful at various intrafamilial levels in angiosperms (depending on the lineage), but are unlikely to retain sufficient evolutionary signal or alignability for the examination of relationships among species in different plant families (Baldwin *et al.*, 1995). However, this generalisation is valid only to the extent that family rank implies an ancient origin, which is not true in all cases. Furthermore, low levels of ITS variation in some ancient plant groups raise the possibility that the ITS region may prove useful for the appraisal of relationships between old families that have experienced exceptionally low rates of spacer evolution. The arbitrariness of taxonomic rank is a major limitation to any general statement about taxonomic limits on the phylogenetic utility of ITS sequences (Baldwin *et al.*, 1995; Hamby & Zimmer, 1992).

Baldwin *et al.* (1995) have cited several plant families (e.g. Asteraceae, Fabaceae, Rosaceae, Saxifragaceae, Viscaceae and Polemoniaceae) where ITS sequences were

utilised effectively for examining relationships within genera and among closely related genera. Also, within species, ITS sequences have been used successfully for investigating relationships among allopatric or disjunct populations. For example, up to 4.3% ITS sequence divergence was found between individuals from conspecific, allopatric populations in *Calycadenia* (Baldwin, 1993). Moreover, it became evident that in this genus the ITS region had evolved primarily by point mutations, based on the moderately high levels of sequence divergence between and within species, and even among subspecies.

The small number of nucleotide positions available for phylogenetic analysis in both ITS spacers is often compensated for by the high levels of variation found in ITS1 and ITS2. In several studies, ITS sequences are reported to be much more variable than the total cpDNA from the same set of DNA accessions (Baldwin *et al.*, 1995), an instance that has been found, for example, in the families *Astragalus* (Wojciechowski *et al.*, 1993), *Madiinae* (Baldwin, 1992), *Rudbeckckiinae* (Urbatsch & Baldwin, 1993) and *Viburnum* (Donoghue & Systma, 1993), among others.

Reticulate evolution and the origin of polyploids

Since nrDNA is inherited biparentally, it is useful for studying hybridisation, introgression and reticulate evolution in plants. Unlike cpDNA, nrDNA data can provide direct evidence of reticulate evolution if concerted evolution fails to act across repeat units contributed by different parental species (Baldwin *et al.*, 1995; Chase *et al.*, 1993). For instance, such lack of sequence homogenisation may occur if: **(1)** the hybridisation event was recent; **(2)** nrDNA repeats are at different loci in the parental taxa, and interlocus gene conversion is inoperative in their hybrid; or, **(3)** the hybrid is asexual. The parentage of suspected early generation hybrids may be resolved simply by screening for presence or absence of restriction sites diagnostic for ITS sequences of each of the putative parental species. In such cases, additivity for the parental restriction patterns provides excellent evidence of hybridity. However, resolution of ancient hybridisation is likely to require more detailed analysis of ITS variation, e.g. by sequencing of ITS clones (Baldwin *et al.*, 1995).

If concerted evolution fails to homogenise ITS paralogues (i.e. those at different chromosomal loci) through a series of speciation events, the possibility of unknowingly sampling sequences with different evolutionary histories is a real danger to phylogenetic analysis (Sanderson & Doyle, 1992). On the other hand, if such paralogues are retained in most or all members of a species lineage, thorough sampling of these sequences can offer independent estimates of organismal phylogeny, and even a means of rooting a portion of the tree in the absence of outgroup data (Iwabe *et al.*, 1989; Baldwin *et al.*, 1995). In other words, non-homogenised paralogues represent positive phylogenetic opportunities along with some potential danger.

One example of the use of ITS in a study of genetic diversity and reticulate evolution comes from the work of Soltis *et al.* (1991) in the genus *Heuchera* (Saxifragaceae). Here, cpDNA restriction site variation had suggested that both northern and southern populations of *Tellima grandiflora* in the USA were distantly related. In contrast, ITS data strongly indicated that both groups of populations were conspecific, as was also indicated by their morphology and allozyme data (Soltis & Kuzoff, 1995). Consequently, it was postulated that introgressive hybridisation between *T. grandiflora* and a species of *Mitella* had led to chloroplast capture of the *Mitella* plastome by some populations of *T. grandiflora*, thus causing the high level of cpDNA divergence found within this species (Soltis *et al.*, 1991).

Ribosomal genes have been used in several investigations dealing with the origins of polyploid taxa, particularly in cereals (Appels *et al.*, 1980; Saghai-Maroof *et al.*, 1984). One study examined spacer length variation within populations and among species of *Triticum* (Appels & Dvorak, 1982). These authors analysed a 130 bp repeat unit found within the spacer region; two out of 11 of the 130 bp variants were sequenced, and the lowered thermal stabilities of heterologous versus homologous hybrids were employed for estimating sequence differences among the other 130 bp variants. Subsequently, different cultivars of hexaploid *T. aestivum* and tetraploid *T. dicoccoides* were assayed for sequence differences in the spacer region, and it was estimated that differences from the 'standard' cultivar *Chinese Spring* ranged from 0.6

to 2.2% at the nucleotide sequence level. These findings, along with cytological evidence, suggested that factors other than simple hybridisation have been involved in the origin of hexaploid wheat. It was suggested that structural changes and deletion of some nrDNA genes in the diploid and tetraploid genomes must have occurred prior to domestication of the polyploid wheats, that is, over 10000 years ago.

To summarise, ITS characters have aided the understanding of plant evolution by providing: **(1)** corroboration of unexpected findings and the resolution of conflict between data sets (e.g. morphological, cpDNA-based vs. ITS sequences); **(2)** improved resolution of species relationships (e.g. aiding clarification of taxonomic, biogeographic and cytological data); **(3)** direct resolution of reticulate evolution; and, **(4)** evidence of the parentage of polyploids. For a more detailed discussion of these various aspects, the reader is referred to Baldwin *et al.* (1995) and references therein. In a more broader context, nuclear ribosomal RNA genes have provided much of the molecular data for phylogenetic reconstructions among several branches in the Tree of Life (Hillis & Dixon, 1991; Mindell & Honeycutt, 1990; Maddison & Maddison, 1996).

The increasing number of nrDNA-based studies (e.g. the reviews by Hamby & Zimmer, 1992; Avise, 1994; and, Baldwin *et al.*, 1995) attests to their value and potential, combined with the relative simplicity of automated sequencing methods. Therefore, there seems to be little question that sequencing of the highly conserved regions encoding nrDNA is a powerful tool in many fields of biological study.

4.1.6. Objectives and aims of sequencing the ITS region of *Pachyrhizus* species

The primary goal of sequencing the nrDNA of *Pachyrhizus* species was three-fold. First, to describe the features of the ITS region and the nature of its molecular evolution; second, to reconstruct a phylogeny of the genus and establish relationships between the five species using taxa from different Neotropical localities; and, third, to compare the inferred phylogeny with those generated from the cpDNA analyses (chapter 3).

4.2. Materials and methods

4.2.1. Plant material

Twenty nine accessions of *Pachyrhizus* were subjected to analysis, comprising ecotypes and cultivars representing the different geographic, ecological and climatic Neotropical areas where the genus is distributed. In addition, a total of three taxa of the genera *Calopogonium* and *Canavalia* were used as outgroups. A complete list of plant material included in the survey is given in Table 2.1. All accessions employed had previously been examined by means of a restriction analysis of cpDNA variation. Details of plant cultivation and preparation of leaf samples are as described in chapter 2 (section 2.2).

4.2.2. DNA extraction, ITS amplification and sequencing

All experimental procedures were carried out as described in Chapter 2. Extraction, purification and estimation of DNA concentration were performed as described in sections 2.3 and 2.4. DNA was isolated from individual plants (as suggested in Campbell *et al.*, 1995; Wendel *et al.*, 1995; Möller & Cronk, 1997; and Wojciechowski *et al.*, 1993) to minimise amplification of multiple nrDNA repeat types or even ITS length or major sequence variants within a DNA accession. Amplification of the ITS region, cleaning of the amplified products, as well as automated cycle sequencing were performed as outlined in section 2.7.

4.2.3. Sequence analysis and alignment

For each accession a consensus sequence was produced using its respective forward and reverse sequencing reaction. The sequences from both reactions were aligned using the options COMPARE TWO SEQUENCES, CREATE SHADOW and COMPUTE CONSENSUS SEQUENCE in the multiple alignment programme Sequence Navigator™ version 1.0.1. (Perkin Elmer, Applied Biosystems Division, CA, USA) with minor manual adjustments.

DNA sequences were aligned manually by sequential pairwise comparisons using the following options of the multiple sequence alignment editor and shading utility GeneDoc© version 2.4 (Nicholas & Nicholas, 1997): ARRANGE SEQUENCES, AUTO SHADING MODES (set to CHEMICAL PROPERTY MODE) and RESIDUE DISPLAY MODE (set to NORMAL and DIFFERENCES MODES). Subunit and spacer boundaries of the DNA sequences were determined by comparison to the corresponding boundaries in *Vicia faba* (Yokota *et al.*, 1989) and *Vigna radiata* (Schiebel & Hemleben, 1989) obtained from sequences available at GENBANK SEQUENCE DATABASE (WWW site; accession numbers X17535 and X14337, respectively).

Finally, the CLUSTAL W computer software package was used to complete alignment of both the ingroup (i.e. *Pachyrhizus*) and outgroup taxa. This alignment required incorporation of minor gaps over the ITS1, 5.8S, and ITS2 regions. The G+C content of the three regions was determined by inspection, and nucleotide sequence divergence among taxa was calculated using the DISTANCE MATRIX option in PAUP, based on unambiguously alignable regions.

4.2.4. Data analysis and phylogeny reconstruction based on ITS sequence variation

Alignment required interpretation of minor gaps which appeared in the sequences of different taxa through the ITS region. Wojciechowski *et al.* (1993) described two ways in which indels (insertions and deletions of nucleotides) can be incorporated in the phylogenetic analysis of a group of taxa. Each gap position can be treated as a missing data item, or alternatively as a new character, i.e. the fifth base. Treating gaps as missing data allows information to be retained on base substitutions occurring in those taxa within the indel region. However, it will exclude information regarding the evolutionary events or transformation involved in the insertion or deletion of bases. On the other hand, scoring indels as separate characters will increase the risk of overweighting them in the analysis, if adjacent gaps are non-independent due to erroneous decisions made during alignment (Baum *et al.*, 1994). In this study indels were scored as missing data.

Regions in which alignment was ambiguous were eliminated from analysis. Phylogenetic trees were generated from unordered character states (Fitch parsimony) using PAUP run on an Apple Macintosh *Power PC 7200/75.* The ACCTRAN option was chosen as a method for optimising unordered characters. Invariant sites and strictly autapomorphous base changes were also ignored in the phylogenetic reconstruction (IGNORE UNINFORMATIVE CHARACTERS option), following the recommendation of Bayer *et al.* (1996). Character state changes were weighted equally. Owing to the relatively large number of taxa, heuristic searches were conducted using the options BRANCH SWAPPING, TBR and MULPARS.

Sets of equally parsimonious trees were summarised using strict consensus. Descriptive statistics reflecting the amount of phylogenetic signal in the parsimony analyses were given by the consistency (CI), homoplasy (HI) and retention (RI) indices. Bootstrapping, taken as an index of support for individual clades (Felsenstein, 1985), was implemented in PAUP using 100 replicates of heuristic searches with TBR swapping and MULPARS. The decay index for individual clades, i.e. the number of additional evolutionary steps required before at least one of the possible trees fails to resolve a particular sister group relationship, was calculated by examining the strict consensus of all equally parsimonious trees one or more steps longer.

4.3. Results

4.3.1. DNA sequence analysis and repeat-unit variation

After alignment of forward and reverse sequences, it was only possible to obtain unambiguous consensus sequences for 14 accessions out of the total of 32 examined (Table 4.1). The remaining accessions exhibited sequences with high percentages of alignment ambiguities (> 50% of unresolved nucleotide sites), rendering them uninformative for phylogenetic analysis. This was an unexpected and surprising result, since PCR products obtained after amplification and purification procedures were resolved in every case as a single, sharp, double-stranded DNA band.

It is likely that these unresolved bases represent genuine polymorphisms within an individual, since they occur in both forward and reverse sequencing reactions, as was also found following ITS sequencing in *Gentiana* (Yuan *et al.*, 1996) and *Saintpaulia* (Möller & Cronk, 1997). In addition, individual DNA sequences probably exhibited some level of potential polymorphism at nucleotide sites, i.e. two bands resolving at a single position of the gel that could indicate multiple nrDNA repeat types including either ITS length variants or major sequence variants, as suggested by Baldwin (1992 & 1993).

Despite every effort to include all 32 ITS consensus sequences in preliminary alignments and phylogenetic analyses, this only led to confusion. For example, in several instances accessions of *P. ferrugineus* and *P. panamensis* (i.e. the most ancestral species as suggested by cpDNA-based phylogenies; chapter 3) were resolved at terminal nodes clustering together with *P. ahipa* and *P. tuberosus* taxa. Thus, it was decided to restrict alignment and phylogenetic reconstruction to only the unambiguous ITS sequences obtained for 13 representatives of *Pachyrhizus* and one outgroup (Table 4.1).

4.3.2. ITS structure, size and composition

Within the species of *Pachyrhizus* examined, the ITS1 sequence was found to be consistently shorter than that of ITS2; the same was also true for the outgroup *Calopogonium mucunoides*. Among *Pachyrhizus* DNAs, ITS1 varied in length from 203 bp in *P. ahipa* to 205 bp in *P. erosus*, *P. tuberosus* and *P. panamensis* (Table 4.1). ITS2 varied from 219 bp in *P. ahipa* to 221 bp in *P. tuberosus* (accession TC536). As regards the 5.8S subunit, most species surveyed contained a sequence of 164 bp, which is consistent in length with most angiosperms; however, some variation was detected with values ranging from 162 bp in accession FW237 to 165 bp in accessions AC208 and FWLoc7 (Table 4.1). Outgroup OUTcm exhibited the shortest ITS1 sequence (202 bp) while both 5.8S and ITS2 sequences slightly exceeded the length ranges of *Pachyrhizus* species (167 bp and 224 bp, respectively).

Table 4.1. Size, structure and composition of the ITS region of nrDNA in 13 *Pachyrhizus* accessions and one outgroup (*Calopogonium mucunoides*) successfully sequenced.

Accession number	Length (bp)				G+C content
	ITS1	5.8S	ITS2	Total	(%)
AC201	204	164	219	587	53.5
AC208	203	165	219	587	53.5
EC558	205	164	220	589	51.3
EC565	205	164	220	589	49.6
EW203	205	164	220	589	53.9
FW237	204	162	220	586	50.5
FWLoc7	204	165	220	589	53.1
PW055	205	164	220	589	52.3
PWTM58	205	164	220	589	54.5
TC350 (chuin)	205	164	220	589	50.8
TC536 (ashipa)	205	164	221	590	52.5
TC553 (jíquima)	205	164	220	589	53.1
TWTM48	205	164	220	589	53.8
Total ingroup	**2660**	**2132**	**2859**	**7651**	**682.4**
Average ingroup	**204.6**	**164.0**	**219.9**	**588.5**	**52.2**
OUTcm	202	167	224	593	50.9

The percentage of G+C content in *Pachyrhizus* ranged from 49.6% (accession EC565) to 54.5% (in PWTM58), whereas G+C content for the outgroup was 50.9% (Table

4.1). The aligned sequences of the entire ITS1-5.8S-ITS2 region are presented in Table 4.2. Alignment of the ITS1 sequences of *Pachyrhizus* created one gap at two positions (sites 2 and 170) equivalent to 0.98% of nucleotide sites. The gap at position 170 was autapomorphic (Table 4.2). Similarly, the aligned 5.8S subunit sequences created additional gaps at positions 213 and 248, representing 1.22% of sites, with the gap at position 213 autapomorphic. Finally, the aligned ITS2 sequences required inclusion of one gap at positions 510 and 597 (0.91% of sites), neither of which were strictly autapomorphic.

Inclusion of outgroup OUTcm among the aligned *Pachyrhizus* sequences created eight additional gaps at the following positions: 4 and 133 in the ITS1 spacer; 282, 318 and 319 in the 5.8S subunit; and, positions 388, 469, 571 and 572 in the ITS2 spacer. All gaps were autapomorphic and represented 1.36% of nucleotide sites.

Table 4.2. Aligned nucleotide sequences of the ITS region in the 18-26S nuclear ribosomal DNA from 13 representatives of *Pachyrhizus* and one outgroup species. Columns are nucleotide sites and rows are individual DNA sequences; sites 1 to 600 are numbered in 5' to 3' order from the 18S subunit / ITS1 border to the ITS2 / 26S subunit border.

Taxa†	Nucleotide sites‡
	ITS1 →
	1 2 3 4 5 6 0 0 0 0 0 0 . . . . . .
AC201	T-G-GGTTNCAGACGAATGCCAGCGGCGNATCGTTCAATCACCCCGAGAGGAGGCACCGT
AC208	T-G-GGTTCCAGACGAATGCCAGCGGCGAATCGTTCAATCACCCCGAGAGGAGGCACCGT
EC558	TCG-GGTTCCAGANGAATCCCAGCAGCGAATCGTTCAATCACCCCAAGAGGAGGGACCGT
EC565	TCG-GGTTCCAGACGAATCCCAGCAGCNAATCGTTCAATCACCCCAAmAGGAGGGACCGT
EW203	TCG-GGTTCCAGACGAATCCCAGCAGCGAATCGTTCAATCACCCCGAGAGGAGGGACCGT
FW237	T-G-GGTTCCAGACGAATCCCAGCsGCGAATCGTTCAATCACCCCGAGAGGAGGCACCGT
FWLoc7	T-G-GGTTCNAGACGAATCCCAGCGGCGAATCGTTCAATCACCCCGAGNGGAGGCACCGT
TC350ch	TCG-GGTTCCAGACGAATGCCGGCAGCGAATCGTTCAATCACCCCAAGAGGAGGGACCGT
TC536as	TCG-GGTTCCAGACGANTGCCGGCAGCGAATCGTTCAATCACCCCAANAGGAGGGACCsT
TC553ji	TCG-GGNTCCAGNCGAATGCCGGCAGCGNyTCGTTCAATCACCCCAAGAGGAGGGACCGT
TWTM48	TCG-GGTTCCAGACGAATCCCAGCAGCGAATCGTTCAATCACCCCGAGAGGAGGGACCGT
PW055	TCG-GGTTCCAGACGAATCCCAGCAGCGAATCGTTCAATCACNCCGAGAGGAGGNACCGT
PWTM58	TCG-GGTTCCAGACGAATCCCAGCAGCGAATCGTTCAATCACCCCGAGAGGAGGGACCGT
OUTcm	--GAGGTTGNACACGAATTTTAGGAGCGAkTGTTTCAATCACCCCTA-AGTAGGCCGGGT

Table 4.2. Aligned DNA sequences. Continued.

Taxa†	Nucleotide sites‡
	7 8 9 1 1 1 0 0 0 0 1 2 0 0 0 . . . . . . ** * * *
AC201	TGGAGTGGCGTTACACCTCGAGGGGCTGGCGGGCTTGGGATCGTGCAACGTCACGTCTCG
AC208	TGGAGTGGCGTTACACCTCGAGGGGCTGGCGGGCTTGGCATCGTGCAACGTCACGTCTCG
EC558	yrGAGTGrysTTACACCTCGAGGGGCCGGCGGGsyTGGGATCsTGsrwCGkCACGTCTCG
EC565	TGGAGTGkwrTTwCACCTCGAGGGGCCGGCGGGCTTGGGATCGTGCAACGTCACGTCTCG
EW203	TGGAGTGGCGTTACACCTCGAGGGGCCGGCGGGCTTGGGATCGTGmAACGTCACGTCTCG
FW237	TGGAGTGrCkTATCsCCTNGGGGGGCCGGCGGGCTTGGGATCGTGCAACGTCANGTCTCG
FWLoc7	TGGAGTGGCGTATCACCTCGGGGGGCCGGCGGGCTTGGGATCGTGCAACGyCACGTCTCG
TC350ch	TGGAGTGGwGTTACACCTCGAGGGGCTGGCGGGCTTGGGATCsTGsrwCGkCACGTCTCG
TC536as	TGGAGTGGCGTTACACCTCGAGGGGCTGGCGGGCTTGGGATCGTGCAAmGTCACGTCTCG
TC553ji	wGGAGTGGCGTTACACCTCGAGGGGCTGGCGGGCTTGGGATCGTrCAAsGTCACGTCTCG
TWTM48	TGGAGTGGCGTTACACCTCGAGGGGCCGGCGGGCTTGGGATCGTsCAAmGTCACGTCTCG
PW055	TmmAGTGGCNTTACACCTCGAGGGGCCGGCGGGCTTCGGANCGTGCAANGTCANTTCTCG
PWTM58	TGGAGTGGCGTTACACCTCGAGGGGCCGGCGGGCTTCGGATCGTGCAACGTCACGTCTCG
OUTcm	GGGAGTGG-NTTACACCTAGGGGGGCCGGCGGGCTTGGGATAGTGrAACGTCACGTCTCG

Taxa†	Nucleotide sites‡
	1 1 1 1 1 1 3 4 5 6 7 8 0 0 0 0 0 0 . . . . . . * * ** **
AC201	GCCAACTGCCTC-CCCTTGCGTTGGGCAGTGGCCTCGGCCCTTGCTCGACAACCmCAAAA
AC208	NCCAACTGCyTC-CCCTTGCGTTGGGCAGTGGCCTGGGCCCTTGCTCGA-AACCACAAAA
EC558	GCCAACTGCmTy-CCCTTGCGTTGGGCAGTCGCCTCGGCCCTTGCkCkkCAACCACAAAA
EC565	GCCAACTkCCTC-CCCTTGCGTTkGGCAGsGsCCTCrGsCCyTksyCGGsmACCmCmAAA
EW203	GCCAACTGCCTC-CCCTTGCGTTGGGCAGTGGGCTTGGCCCTTGCTCGACAACCmCAAAA
FW237	GCCAACTGCCTC-CCCTTGCGTTkGGCAGTGGCyTkGGCCyTTGCTCGACAACCCCAAAA
FWLoc7	GCCAATTGCCCC-CCCTTGCGTTGGGCAGTGGCCmGGsCCCTTGCCCGArAANCCCAAAC
TC350ch	GCCAACTGCmTy-CCCTTGCGTTGGGCAGsGGCCTCGsCCCTTGCTCGACAACNmCAAAA
TC536as	GCCAACTGCCTC-CCCTTGCGTTGGGCAGmGGCCTCGGCCCTTGTTCGACAACCACAAAA
TC553ji	GCCAACTGCCTC-CCCyTGCGTTGGGCAGTGGCCTCGGCCCTTGCTCGACAACCACAAAA
TWTM48	GCCAACTGCCyC-CCCyTGCGTTGGGCAGTGGCCTCGGCCCTTGCTCGACAACCCCAAAA
PW055	GCCAATTGCCCC-CCCTTGCNTTkGGCAGTGNCNTCGGCCCTTGCTCGACAACCCCAAAA
PWTM58	GCCAACTGCCTC-CCCTTGCGTwGGGCAGTGGCCTCGGCCCTTGCTCGACAACCCCAAAA
OUTcm	ACCAACTwCCTCCCCCTTGAGTTGGG-AGTGGCCTCGGCCCTTGCTCGACAACCCCAAAA

Table 4.2. Aligned DNA sequences. Continued.

Taxa†	Nucleotide sites‡

```
                                                5.8S
                                                →
                    1         2         2         2         2         2
                    9         0         1         2         3         4
                    0         0         0         0         0         0
                    .         .         .         .         .         .
                                        *                   *
AC201       ACCCGGATCTTCGTGTGCCAAGrAATCAAAACATGTTTGTGAAGGGCAATTCTCGTGGGC
AC208       CCCCGGATCTTCGTGTGCCAAGGAATCAAAACATGTTTGTNAAGGGCAATTCNCGTGGGC
EC558       CCCCGGsGGTTCGTGTGCCAAGGAATCsAAACATGTTTGTGAAGTGCAATTCyyGTGGGC
EC565       CCCCGGsssTTCGTGTGCCAAGrAATCAAAACATGTTTGTGAAGTGCAATTCTCGTGGGC
EW203       ACCCGGCGCTTCGTGTGCCAAGGAATCAAAACATGTTTGTGAAGTGCAATTCTCGTGGGC
FW237       CCCCGGCGCTTCTGTGCCCAAGGAATCAAAAN-TGTTGNTGAAGTGCAATNCTCNTGGGC
FWLoc7      CCCCGGCGCTTCTGTGCCCAAGGAATGAAAACATGTTGGTGAAGTGCAATTCTCGTGGGC
TC350ch     CCCCGGsGCTTCkTGTGCCAAGGAATCNAAACATGTTTGTGAAGTGCAATTCTCGTGGGC
TC536as     CmCCGGCGCTTCGTGTGCCAArGAATCAAAACATGTTTGTGAAGTGCAATTCTCGTGGGC
TC553ji     CCCCGGCGCTTCGTGTGCCAAGGAATCAAAACATGTTTGTGAAGTGCAATTCTCGTGGGC
TWTM48      CCCCGGCssTTCGTGTGCCAAGGAATCAAAACATGkTTGTGAAGTGCAATTCTCGTGGGC
PW055       CCCCGGCGCTTNGTGTGCCAAGGAATTAAAACATGTTTNTGAAGTrCNATTCTCGNGGGC
PWTM58      CCCCGGCGCTTCGTGTGCCAAGGAATCAAAACATGTTTGTGAAGTGCAATTCTCGTGGGC
OUTcm       CCCCGGCGCTTCGTGTGCCAAGGAATCAAAACATGTTTGTGAAGGwCAATTCCCGTGGGC
```

Taxa†	Nucleotide sites‡

```
                    2         2         2         2         2         3
                    5         6         7         8         9         0
                    0         0         0         0         0         0
                    .         .         .         .         .         .
AC201       TCGGAGA-CGATGTCCCCACGAGCGGTCGTTCTTCACGATA-CAATTGTATACTTCCAAA
AC208       CGGGAGACCGATGTCCCACCNAGCGGTCGTTCTTCACGATA-CAATTGTATACTTCCAAA
EC558       TCGGAGA-CGATGTCCCCACGArCGGTCGTTCTTCACGATA-CAATTGkATwCATCCAAA
EC565       TCGGAGA-CrATGTCCCNACrAGCGGTCGTTCTTCACrATA-CAATTGTATACATCCAAA
EW203       TCGGAGA-CGATGTCCCCACGAGCGGTCGTTCTTCAAGATA-CAATTGTATACATCCAAA
FW237       TGGGAGA-CGTTGTCCCACNGAGNGGTNGTNCTTCNNGATA-CAATTGTATACATCCAAA
FWLoc7      CGGGAGACCrATGTCCCACCGAGCGGTCGTTCTTCACGATA-CAATTGTATACATCCAAA
TC350ch     TCGGAGA-CrwTGTCCCCACGAGCGGTCGTTCTTCAmGATA-CAATTGTATwCTTCCAAA
TC536as     TCGGAGA-CrwTGTCCCCACrAGCGGTCGTTCTTCACrATA-CAATTGTATACTTCCAAA
TC553ji     TCGGAGA-CrATGTCCCCACrAGCGGTCGTTCTTCACrATA-CAATTGTATACTTCCAAA
TWTM48      TCGGAGA-CGATGTCCCCACGAGCGGTCGTTCTTCACGATA-CAATTGTATACATCCAAA
PW055       TCGGAGA-CGATGTCCCACCGAGCGGTCGTTCTTCACGATA-CTATTGTATACATCCAAA
PWTM58      TCGGAGA-CGATGTCCCCACGAGCGGTCGTTCTTCACGATA-CAATTGTAGACATCCAAA
OUTcm       CGGGAGAC-GATGTCCCACCGNGCTGTCGTTCTTCACGATAGCAATTGTCTACTTCCAAA
```

Table 4.2. Aligned DNA sequences. Continued.

Taxa†	Nucleotide sites‡
	3 3 3 3 3 3 1 2 3 4 5 6 0 0 0 0 0 0 . . . . . . * * * * *
AC201	AAACCTTTCGGCAACGA--ATATCTGGGCTCTTGGATCCCTAAAGACCGAAGGCAAATGC
AC208	AAACCTNTCGGCAACGA--ATATCTGGGCTCTTGGATCCCTAAAGACCGAAGGCAAATGC
EC558	AAkCCTTTCGGCAACGA--ATATCTGGGCTCTTGGATCCCTAAAGAACGAAGGCAAATGC
EC565	AArCCTTTCGCCAACGr--ATATCTsGGCTCTTGGATCrCTAAAGAACGAAGGCAAATGC
EW203	AAACCTTTCGGCAACGA--ATATCTkGGCTCTTGGATCCCTAAAGAACGAAGGCAAwTGC
FW237	ANACCTTTCGGCAACGA--A-ATCTGGNCCCTTNGATCCCTAAAGACCGAAGGCAAANGN
FWLoc7	AAACCTTTCGGCAACGA--ATATCTGGGCCCTTGGATCCCTAAAGACCGANGsCAAATGC
TC350ch	AAkCCTTTCGGCAACGA--ATATCTyGGCTCTTGGATCsCTAAAGACCGAAGNCAAATGC
TC536as	AAsCCTTTCGGCAACGA--ATATCTGGGCTCTTGGATCCCTAAAGACCGAAGGCAAATGC
TC553ji	AAACCTTTCGGCAACGr--ATATCTGGGCTCTTGGATCrCTAAAGACCGAAGGCAAATGC
TWTM48	AAACCTTTCGGCAACGA--ATATCTGGGCTCTTGGATCCCTNAAGAACGAAGGCAAATGC
PW055	AAACCTTTCGGCAACGA--ATATNTGGGCTCTTGGATCCCTAAAGAACGAAGGCAAATGC
PWTM58	AAACCTTTCGTCAACGA--ATATCTGGGCTCTTGGATsCCTAAArACCGAAGGCAAATGC
OUTcm	AAACCTGwrCGCAACGAGGATATCTGGGATATTGGATCCCTANAyACCGAAGGCGAATGC

Taxa†	Nucleotide sites‡
	ITS2 → 3 3 3 4 4 4 7 8 9 0 1 2 0 0 0 0 0 0 . . . . . . * * * * *
AC201	GATACTGGGTCAAACTTGGACATTCCC-ATAACCATCAAGTCTTTGAAACCAAGTTGGGC
AC208	GATACTGGGTCAAACTTGGACATTCCC-ATAACCATCAAGTCTTTGAAACCAAGTTGGGC
EC558	GATACTGGGTCAkACTTGGACATTCCC-ATkACCATCGNGTCTTTGAAACCAAGTTGGGC
EC565	GATACTGGGTCAAACTTGNArATTCCC-ATkACCATCGAGTCTTTGAAACCAAGTTGGGC
EW203	GATACTGGGTCAAACTTGGACATCCCC-yTAACCATCGAGTATTTGAACGCmAGTTGCGC
FW237	GANACTGGTNNAAACTNGNACATTCCC-ATAACCANCAAGTATTTNAACGCAAGNTGGGN
FWLoc7	GATACTGGTTCAAACTTGGACATTCCC-ATAACCATCAAGTATTTGAACGCAAGTTGGGC
TC350ch	GATACyGGGTkAAACwTGsAwATTyCC-ATkACCATCGAGTCTTTGAAACCAAGTTGGGC
TC536as	GATACTGGGTCAAACTTGGACATTCCC-ATAACCATCGAGTCTTTGAAACCAAGTTGGGC
TC553ji	GATACTGGGTCArACTTGGACATTCCC-ATAACCATCGAGTCTTTGAAACCAAGTTGCGC
TWTM48	GATACTGGGTCAAACTTGGACATCCCC-ATAACCATCGAGTATTTGAACGCAAGTTGCGC
PW055	GATACTGGGTCAAACNTGGACATCyCC-ATNACCATCGAGTATTTGAACGCAAGTTGCGC
PWTM58	GATACTGGGTCAAACTTGGACATCCCC-AkAACCATCGAGTATTTGAACGCAAGTTGCGC
OUTcm	GATACTGGGTCAAACTTGGATATAANAAmTAACNATCAAGTCTTTTAACGCAAGTTGCGC

Table 4.2. Aligned DNA sequences. Continued.

Taxa†	Nucleotide sites‡
	4 4 4 4 4 4 3 4 5 6 7 8 0 0 0 0 0 0 . . . . . . * * *
AC201	CCAAAGCCATTAGGCTTAGGGCACGCCTTACTGGGTGTCCACACTTyT-kACCCTCCACA
AC208	CCAAAGCCATTAGGCTGAGGGCACGCCTTACTGGGTGTCCACACTTkT-CACCCTCCACA
EC558	CCAAAGGCATTAGGCTTAGGGCACGCCTGCCTGGGTGTCACAAATTGT-CACCCTCCANA
EC565	CCAAAGGCATTAGGCTTAGGGCACGCCTGCCTGGGTGTCACACATTGT-CACCCymCACr
EW203	CCAAAGCCATTAGGCTGAGGGCACGCCyGCCTGGGTGTCACACATTGT-CACCCTCCACA
FW237	CCAAAGCCATTANGCTTAGGGCACGCCTGCCTNGGTNNCACACATTGT-CACCCTNCACA
FWLoc7	CCAAAGCCATTAGGCTTAGGGCACGCCTGCCTkGGTGTCACACATykT-CACCCTCCACA
TC350ch	CCAAAGCCATTAGGsTTAGGGCACGCCTTGCTGGGTGTNACACATTGT-CACCCTCNANA
TC536as	CCAAAGCCATTAGGCTTAGGGCACGCCTTACTGGGTGTCCACACTTGT-kACCCTCCACA
TC553ji	CCAAAGCCATTAGGsTTAGGGCACGCCTTGCTGGGTGTCACACATTGT-CACCCTCCACA
TWTM48	CCAAAGCCATTAGGCTrAGGGCACGCCTGCCTGGGTGTCACmCATTGT-CACCCTCCACA
PW055	CCAAAGCCATTAGGCTGAGGGCACGCCTGCCTGGGTGTCACACATTGN-CACCCTCCACA
PWTM58	CCAAAGGCATTAGGCTGAGGGCACGCCTGCCTGGGTGTCACACATTGT-CACCCTCyACA
OUTcm	CCAAAGGCATTAGGGTGAGGGCCCGCCTTACTNGGTGNCACACATTGTTCACCCTCCACA

Taxa†	Nucleotide sites‡
	4 5 5 5 5 5 9 0 1 2 3 4 0 0 0 0 0 0 . . . . . . *
AC201	CGCNACACCACTGTGCAGGAGGAGGGTTG-TGCTGGCTTCCCGCGAGGCCCGACTCGCGG
AC208	CGCAAAACCACTGTGCAGNAGGAGGGTTG-TGCTGGCTTCCCGCGAGCACCGACTCGCGG
EC558	CGCACACACACTGTACAGGAGGAGGGTGTATGCTGGTTTCCCGCGAGCACCGACTCGCGG
EC565	srCACACACACTGTACAGGAGGAGGGTGTATGCTGGTTTCCCGCGAGCGCCGACTCGCGG
EW203	CGCACACACACTGTGCAGGNGGrGGGTGTATGCTGGTTTCCCGCGAGCACCGACTCGCGG
FW237	CGCACCCCCACTGCGCAGGAGGAGGGTTNATGCTGGTTTCCNGCGAGCCCCwACTNGCGT
FWLoc7	CGCACACACACTGTGCANGAGGAGGGTTTATGCTGGTTTCCCGCNAGCACCGACTCGCGG
TC350ch	sGCACACACACTGTGCAGGAGGAGGGTTTATGCTGGCTTCCCGCGAGCACCGACTCGCGG
TC536as	CGCACACACACTGTGCAGGAGGAGGGTTTATGCTGGCTTCCCGCGAGCACCGACTCGCGG
TC553ji	CsCACACACACTGTGCAGGAGGAGGGTTTATGCTGGCTTCCCGCGAGCACCGACTCGCGG
TWTM48	CGCACACACACTGTGCAGNAGGAGGGTGTATGCTGGTTTCCCGCGArCACCGACTCGCGG
PW055	sGCACCCCCACTGTGCAGGAGGGGGGTTTATGCTGGTTTCGCGCGAGCCCCCACTCGCGG
PWTM58	CGCACACACACTGTACAGGAGGAGGGTGTATGCTGGTTTCCCGCGAGCACCGACTCGCGG
OUTcm	CGCAAACCCATTGTGCAGGAGGGGGNTGCATGCTGGTTTCCCGCGAGCCCCCACTCGCGG

Table 4.2. Aligned DNA sequences. Continued.

Taxa†	Nucleotide sites‡
	5 5 5 5 5 6 5 6 7 8 9 0 0 0 0 0 0 0 . . . . . . * ** * *
AC201	TTGGTTAAAAATCGAGTTCGCAACCNTTTT--CGTNGTGAAAAATTGGTGGATGGT-AAC
AC208	TTGGTTAAAAATCGAGTTCGCAACCTTNTT--CGTCGTGAAAAATTGGTGGATGGT-AAC
EC558	TTGGTTAAAAATCGANwTCGCNACCTTTTT--CGTCGNGATAAATTGGTGGkTGGT-AyC
EC565	TTkGTTAAAAATCGAGTTCGCAACCTkTTT--yGTCGTGATAAATTGGTGGATGGT-rCC
EW203	TTGGTTAAAAATCCAGTTCGCAACCTTTTT--CGTCGTGAAAAATTGGTGGATGGT-AAC
FW237	TTGGTTAAAAATCGAGTTCGCAACCTTTNT--CGTCGTGAAAAATTGGTGGATGGT-AAC
FWLoc7	TTGGTkAAAAATCGAGTTCGCAACCTTTTT--CGTNGTGAAAAATTkNTGGATGrm-AAC
TC350ch	TTGGTTCCAAATCGAGTTCGCAACCTTTTT--CGTCGTGAAAAATTGGTGGATGGT-AAC
TC536as	TTGGkTAAAAATCGAGTTCGCNACCTTTTT--CGTCGTGAAAAATTGGTGGATGGTAAAC
TC553ji	TTkGTTCCAAATCCAGTTCGCANCCTTTTT--CGTCGTGAAAAATTGGTGGATGGT-AAC
TWTM48	TTGGTTAAAAATCGAGTTCGCAACCTTTTT--CGTCGTGATAAATTGGTGGATGGT-AAC
PW055	TTGGTTAAAAANCGAGTTCGCAACCNTTTT--CGTCGTGAAAAATTGGTGGATGGT-AAC
PWTM58	TTGGTTAAAAATCGAGTTCGCAACCTTTTT--CGTCGTGATAAATTGGTGGATGGT-AAC
OUTcm	TTGGTTAAAAATTGAyTTC-CAACCkTTTTATCGTCGTGAAAAATTGGkGwkTGGTTAAC

†: Key to accessions is listed in Table 2.1. Coding of accessions of the *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima.

‡: Nucleotide sequence displayed from 5' to 3'. The beginning of the ITS1 region (sites 1 - 207), the 5.8S subunit (208 - 375) and ITS2 region (positions 376 - 600) are indicated by arrows. Sequence symbols: A = dATP; C = dCTP; G = dGTP and T = dTTP. Coding of ambiguous sites follows IUPAC nomenclature: hyphens = gaps; k = G or T; m = A or C; r = A or G; s = C or G; w = A or T; y = C or T; N = aNy base/nucleotides of unknown identity.

*: Nucleotide positions (35 sites) excluded from phylogenetic analyses due to alignment ambiguities.

4.3.3. ITS nucleotide site variation and sequence divergence

Following alignment of the ITS sequences, a character matrix of 600 sites was required to align *Pachyrhizus* and outgroup DNAs. However, it was necessary to exclude 35 positions (marked '*'; see Table 4.2) prior to phylogenetic analysis because of alignment ambiguities. Of the remaining 565 unambiguous aligned positions, 245 or 43.4% were variable, i.e. possessed at least one nucleotide difference in at least one DNA. Approximately 40% of these sites were contained within ITS1, 20.8% in the 5.8S subunit and 39.2% in ITS2.

Of these variable characters, 108 (44.1%) were phylogenetically informative, i.e. possessed nucleotide states shared by at least two DNAs. Among these variable positions, ITS2 accounted for most of this variation (40.8%) compared to 19.4% in the 5.8S subunit and 39.8% in ITS1.

Rates of ITS sequence divergence were determined using the DISTANCE MATRIX option in PAUP excluding unalignable and undetermined sites. In addition, sites with fixed nucleotide character states in all sequences were compared, i.e. those sites without gaps or polymorphisms in any of the aligned sequences. Divergence rates of ITS1, ITS2 and of combined ITS1-5.8S-ITS2 sequences for *Pachyrhizus* species and the outgroup are shown in Table 4.3.A-C. Within the ingroup, sequence divergence between pairs of species ranged from 1.9% (TC350ch vs. TC536as; TWTM48 vs. PWTM58) to 38.9% (FWLoc7 vs. TC536as; FWLoc7 vs. TC553) for ITS1 (Table 4.3.B). For ITS2, higher values were obtained ranging from 3.7% (TWTM48 vs. PWTM58) to 50.0% (AC201 vs. PWTM58), as shown in Table 4.3.C.

Finally, for combined ITS1-5.8S-ITS2 sequences divergence values spanned from 2.9% (TWTM48 vs. PWTM58) and 4.4% (EW203 vs. TWTM48) to 33.1% (AC201 vs. FWLoc7). Among ingroup and outgroup accessions, sequence divergence varied from 26.9% (PW055 vs. OUTcm) to 37.3% (TC536as vs. OUTcm), as illustrated in Table 4.3.A.

4.3.5. Phylogenetic analysis

Fitch parsimony and TBR branch swapping analysis of phylogenetically informative ITS region sites generated five equally parsimonious trees (Figure 4.3). These trees each required 162 evolutionary steps. The consistency index of each tree was 0.66, excluding uninformative sites and combining sequences identical at potentially informative positions; the retention and homoplasy indices were 0.678 and 0.34, respectively.

Table 4.3. Pairwise divergence between ITS region sequences from 13 *Pachyrhizus* and one outgroup DNAs.

4.3.A. Combined ITS1-5.8S-ITS2 matrix.†

‡	AC201	AC208	EC558	EC565	EW203	FW237	FWLoc7
1 AC201	-	0.074	0.287	0.294	0.272	0.287	0.331
2 AC208	10	-	0.296	0.304	0.274	0.274	0.274
3 EC558	39	40	-	0.066	0.176	0.309	0.309
4 EC565	40	41	9	-	0.176	0.279	0.272
5 EW203	37	37	24	24	-	0.221	0.235
6 FW237	39	37	42	38	30	-	0.074
7 FWLoc7	45	37	42	37	32	10	-
8 TC350ch	32	33	24	19	28	40	42
9 TC536as	18	21	25	25	29	42	43
10 TC553ji	29	30	26	25	23	41	41
11 TWTM48	34	34	18	18	6	29	28
12 PW055	37	37	30	28	16	26	29
13 PWTM58	37	37	19	18	10	31	31
14 OUTcm	46	38	47	45	40	43	49

‡	TC350ch	TC536as	TC553ji	TWTM48	PW055	PWTM58	OUTcm
1 AC201	0.235	0.132	0.213	0.250	0.272	0.272	0.343
2 AC208	0.244	0.156	0.222	0.252	0.274	0.274	0.286
3 EC558	0.176	0.184	0.191	0.132	0.221	0.140	0.351
4 EC565	0.140	0.184	0.184	0.132	0.206	0.132	0.336
5 EW203	0.206	0.213	0.169	0.044	0.118	0.074	0.299
6 FW237	0.294	0.309	0.301	0.213	0.191	0.228	0.321
7 FWLoc7	0.309	0.316	0.301	0.206	0.213	0.228	0.366
8 TC350ch	-	0.103	0.074	0.191	0.250	0.206	0.343
9 TC536as	14	-	0.110	0.184	0.257	0.206	0.373*
10 TC553ji	10	15	-	0.162	0.228	0.176	0.328
11 TWTM48	26	25	22	-	0.088	0.029	0.291
12 PW055	34	35	31	12	-	0.110	0.269*
13 PWTM58	28	28	24	4	15	-	0.291
14 OUTcm	46	50	44	39	36	39	-

†: Values in the upper right half of the matrix indicate proportions of divergent sites to a total of 600 sites in each comparison (mean distances, adjusted by PAUP for missing data). The actual numbers of unambiguous divergent sites from pairwise sequence comparisons (absolute distances) appear in the lower left half of the matrix. Highlighted numbers indicate extreme values of the divergence range within *Pachyrhizus*. * = extreme values of ingroup vs. outgroup pairwise comparisons.

‡: Key to accessions is listed in Table 2.1. Coding of accessions of the *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima.

Table 4.3. Sequence pairwise divergences. Continued.

4.3.B. ITS1 matrix.†

‡	AC201	AC208	EC558	EC565	EW203	FW237	FWLoc7
1 AC201	-	0.038	0.185	0.222	0.130	0.259	0.333
2 AC208	2	-	0.189	0.226	0.151	0.245	0.302
3 EC558	10	10	-	0.037	0.093	0.278	0.352
4 EC565	12	12	2	-	0.130	0.278	0.315
5 EW203	7	8	5	7	-	0.204	0.296
6 FW237	14	13	15	15	11	-	0.074
7 FWLoc7	18	16	19	17	16	4	-
8 TC350ch	9	9	6	5	8	16	20
9 TC536as	8	8	6	5	7	17	21
10 TC553ji	8	8	6	8	7	17	21
11 TWTM48	7	8	3	5	2	11	14
12 PW055	10	11	7	9	6	14	14
13 PWTM58	8	9	5	6	3	12	16
14 OUTcm	13	14	12	11	11	16	21

‡	TC350ch	TC536as	TC553ji	TWTM48	PW055	PWTM58	OUTcm
1 AC201	0.167	0.148	0.148	0.130	0.185	0.148	0.250
2 AC208	0.170	0.151	0.151	0.151	0.208	0.170	0.275
3 EC558	0.111	0.111	0.111	0.056	0.130	0.093	0.231
4 EC565	0.093	0.093	0.148	0.093	0.167	0.111	0.212
5 EW203	0.148	0.130	0.130	0.037	0.111	0.056	0.212
6 FW237	0.296	0.315	0.315	0.204	0.259	0.222	0.308
7 FWLoc7	0.370	0.389	0.389	0.259	0.259	0.296	0.404*
8 TC350ch	-	0.019	0.056	0.111	0.185	0.130	0.231
9 TC536as	1	-	0.037	0.111	0.185	0.130	0.269
10 TC553ji	3	2	-	0.111	0.185	0.130	0.231
11 TWTM48	6	6	6	-	0.056	0.019	0.192*
12 PW055	10	10	10	3	-	0.056	0.250
13 PWTM58	7	7	7	1	3	-	0.212
14 OUTcm	12	14	12	10	13	11	-

†: Values in the upper right half of the matrix indicate proportions of divergent sites to a total of 207 sites in each comparison (mean distances, adjusted by PAUP for missing data). The actual numbers of unambiguous divergent sites from pairwise sequence comparisons (absolute distances) appear in the lower left half of the matrix. Highlighted numbers indicate extreme values of the divergence range within *Pachyrhizus*. * = extreme values of ingroup vs. outgroup pairwise comparisons.

‡: Key to accessions is listed in Table 2.1. Coding of accessions of the *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima.

Table 4.3. Sequence pairwise divergences. Continued.

4.3.C. ITS2 matrix.†

‡	AC201	AC208	EC558	EC565	EW203	FW237	FWLoc7
1 AC201	-	0.074	0.426	0.444	0.481	0.296	0.333
2 AC208	4	-	0.370	0.407	0.389	0.315	0.296
3 EC558	23	20	-	0.074	0.278	0.296	0.222
4 EC565	24	22	4	-	0.278	0.278	0.222
5 EW203	26	21	15	15	-	0.204	0.130
6 FW237	16	17	16	15	11	-	0.074
7 FWLoc7	18	16	12	12	7	4	-
8 TC350ch	19	16	13	11	16	15	11
9 TC536as	8	7	15	17	18	17	13
10 TC553ji	19	16	15	15	12	15	11
11 TWTM48	24	19	12	12	3	10	6
12 PW055	22	21	18	17	7	6	9
13 PWTM58	27	22	10	10	5	12	8
14 OUTcm	26	23	23	24	18	17	20

‡	TC350ch	TC536as	TC553ji	TWTM48	PW055	PWTM58	OUTcm
1 AC201	0.352	0.148	0.352	0.444	0.407	0.500	0.481
2 AC208	0.296	0.130	0.296	0.352	0.389	0.407	0.426
3 EC558	0.241	0.278	0.278	0.222	0.333	0.185	0.426
4 EC565	0.204	0.315	0.278	0.222	0.315	0.185	0.444
5 EW203	0.296	0.333	0.222	0.056	0.130	0.093	0.333
6 FW237	0.278	0.315	0.278	0.185	0.111	0.222	0.315
7 FWLoc7	0.204	0.241	0.204	0.111	0.167	0.148	0.370
8 TC350ch	-	0.204	0.074	0.278	0.315	0.315	0.426
9 TC536as	11	-	0.204	0.296	0.370	0.352	0.500*
10 TC553ji	4	11	-	0.241	0.296	0.278	0.426
11 TWTM48	15	16	13	-	0.130	0.037	0.352
12 PW055	17	20	16	7	-	0.167	0.296*
13 PWTM58	17	19	15	2	9	-	0.352
14 OUTcm	23	27	23	19	16	19	-

†: Values in the upper right half of the matrix indicate proportions of divergent sites to a total of 225 sites in each comparison (mean distances, adjusted by PAUP for missing data). The actual numbers of unambiguous divergent sites from pairwise sequence comparisons (absolute distances) appear in the lower left half of the matrix. Highlighted numbers indicate extreme values of the divergence range within *Pachyrhizus*. * = extreme values of ingroup vs. outgroup pairwise comparisons.

‡: Key to accessions is listed in Table 2.1. Coding of accessions of the *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima.

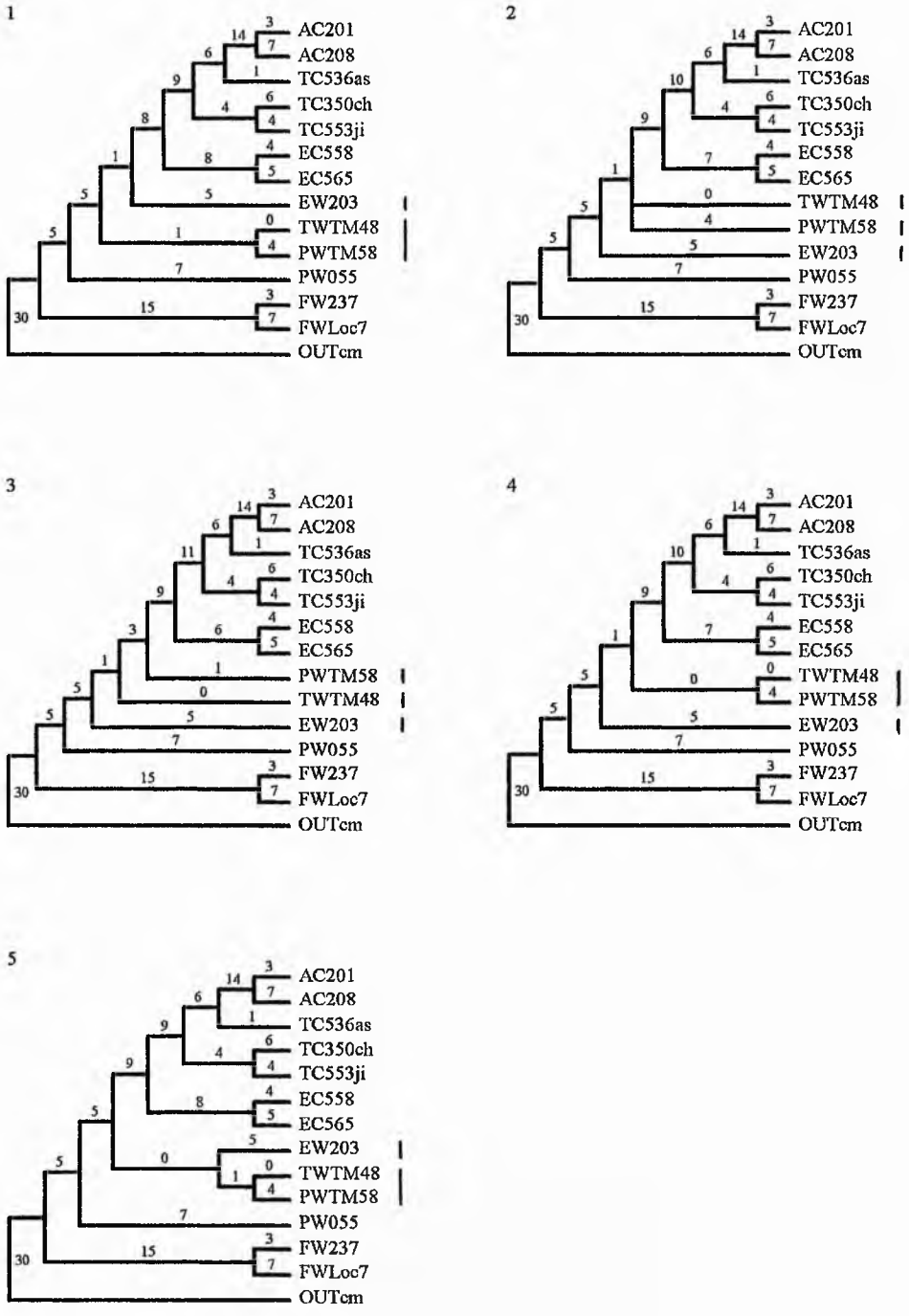


Figure 4.3. The five equally parsimonious trees generated from analysis of ITS-region nucleotide sites among *Pachyrhizus* species and outgroup OUTcm. Numbers above branches indicate branch length, i.e. the number of mutations excluding uninformative changes (Tree length = 162 steps; CI = 0.660; HI = 0.340; RI = 0.678). These trees differ in the positioning of EW, PW and TW taxa, as illustrated (see text for further details).

The five ITS trees differed topologically in their resolution of three accessions: EW203, PWTM58 and TWTM48. In three out of the five equally parsimonious trees, PWTM58 and TWTM48 clustered tightly together, while in the remaining trees these two accessions were resolved either into polytomies or as separate taxa at more basal nodes. Similarly, positioning of accession EW203 in the ITS trees was variable with a resolution either at basal nodes or nested among EC, TW and PW taxa (Figure 4.3).

The strict consensus of the five ITS trees is presented in Figure 4.4. The collapse of phylogeny branches to calculate this consensus tree (CI = 0.656; HI = 0.344 and RI = 0.673) required one additional step, i.e. a total length of 163 evolutionary steps.

Based on the taxa analysed for ITS sequence variation, the strict consensus tree indicates that *Pachyrhizus* is monophyletic, a result supported by 30 base-pair changes, i.e. the number of mutations excluding uninformative changes (Figure 4.4). Parsimony analysis showed also that ITS sequences within the genus were divided into two major clades: **(1)** a group containing all *P. ferrugineus* accessions, being the most ancestral in the phylogeny; and, **(2)** a clade composed of the four remaining *Pachyrhizus* species. Bootstrap values for these consensus clades ranged from 52% to 100%. Each clade is discussed in turn below.

Clade 1

This group of 11 accessions was clearly separated from the rest of the genus by five synapomorphies and supported by a bootstrap value of 52% (Figure 4.4). Within clade 1, six subclades were distinguished. A first subclade contained all cultivated accessions of *P. ahipa* and *P. tuberosus* clustering strongly together (83% bootstrap value) and unambiguously separated from the remaining subclades by 10 synapomorphies. Moreover, in this ITS subclade *P. tuberosus*, namely accession TC536 (an ashipa), is sister to *P. ahipa*, a phylogenetic relationship strongly supported by 82% of the bootstrap replicates. In this same subclade, TC350 and TC553 clustered together (chuin and jíquima, respectively; 80% bootstrap value) and were sister taxa to the aforementioned accessions.

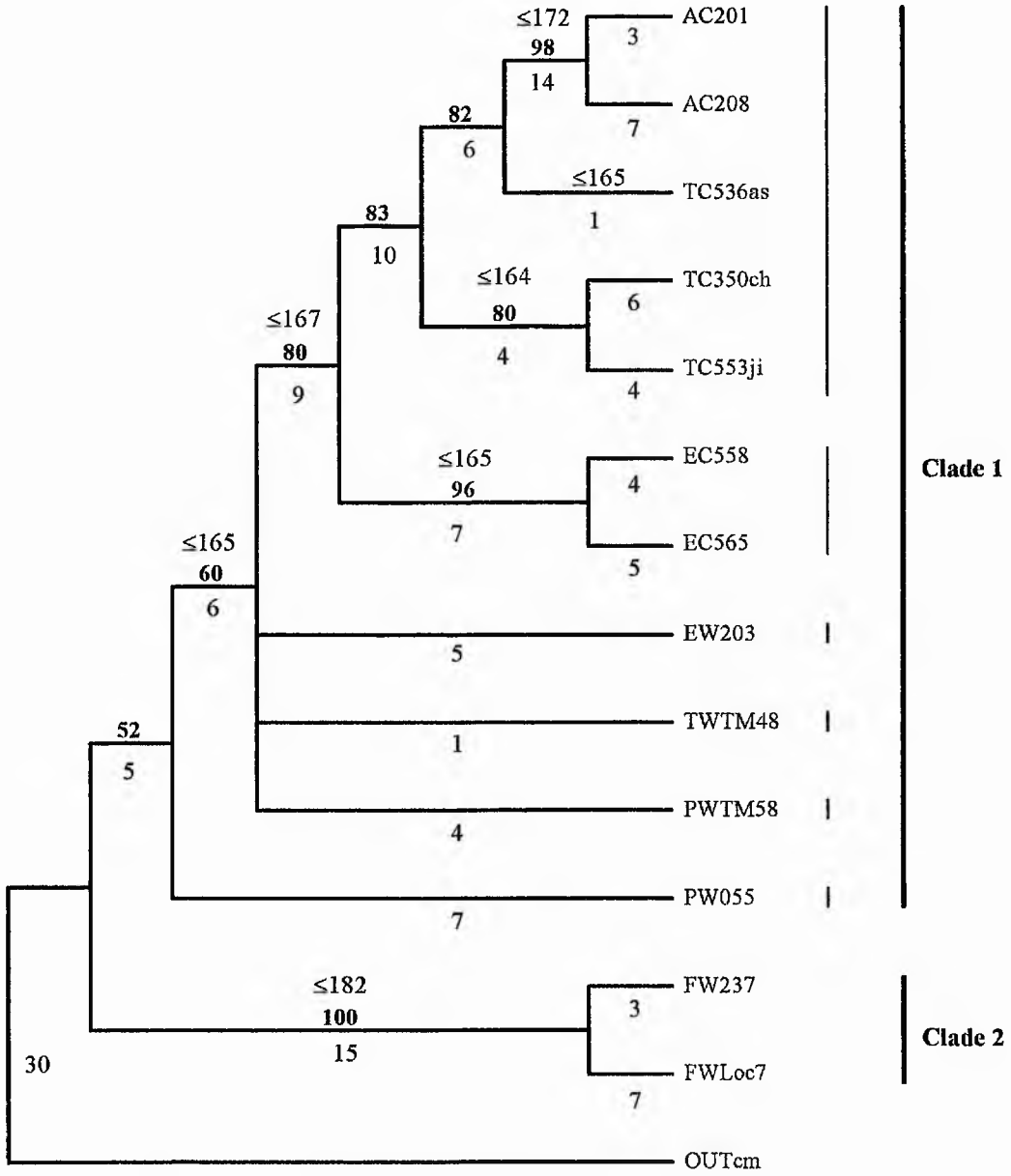


Figure 4.4. Strict consensus of the five equally parsimonious Fitch ITS trees shown in Figure 4.3. Numbers above branches indicate bootstrap percentages (in boldface) and decay values; numbers below branches are synapomorphies (Tree length = 163 steps; CI = 0.656; HI = 0.344; RI = 0.673). The two ITS clades are represented by heavy lines; the six subclades observed in clade 1 are also indicated (see text for further details).

A second subclade contained both cultivated accessions of *P. erosus*. This subclade was separated from the previous subclade discussed by seven unambiguous base-pair changes (i.e. synapomorphies) and supported by a bootstrap value of 96%. This EC subclade was more basal to the first subclade described above, and was sister group to AC and TC accessions (Figure 4.4). Moreover, all cultivated taxa examined in this study (AC, EC and TC) clustered at a strongly supported node (nine base-pair changes) with a bootstrap value of 80%.

The third and fourth subclades contained one accession of wild origin each, i.e. EW203 and TWTM48. Both taxa were ambiguously positioned among the five equally parsimonious ITS trees possibly because of a reduced number of synapomorphic changes and low sequence divergence (= 4.4%, Table 4.3.A). Therefore, these accessions were resolved as polytomies in the strict consensus (Figure 4.4).

The last two subclades each contained an accession of *P. panamensis* which resolved as the most basal taxon within clade 1; of these, the subclade comprising PW055 was distinct from the remaining subclades by seven mutations and rooted clade 1 (Figure 4.4).

Clade 2: *P. ferrugineus*

Clade 2 exhibited the highest bootstrap percentage in the cladogram with a value of 100%. This second clade contained the two accessions of *P. ferrugineus* examined, which resolved as the most ancestral taxon within the genus. A total of 15 mutations, excluding uninformative changes, defined this species as a separate phylogenetic entity. On the other hand, the node that separated *P. ferrugineus* from the rest of the genus had the lowest bootstrap confidence level (52%) in the cladogram (Figure 4.4).

An analysis of decay indices indicated that some of the internal structure of the two ITS clades was lost as trees further away from the minimal consensus tree were considered. When the consensus of trees $\leq$ 164 steps was examined (just one step

longer than the minimal tree), accessions TC350 and TC553 were rapidly lost from clade 1. At this level, the subclade containing *P. ahipa* accessions and TC536 was still recovered (Figure 4.4). At a consensus length ≤ 165 steps most of the internal structure was lost with accessions of *P. erosus*, *P. panamensis* and *P. tuberosus* no longer resolving in clade 1. The *P. ahipa* subclade was strongly supported and collapse of its phylogeny branches required 10 additional evolutionary steps, i.e. a consensus length of trees ≤ 173 (Figure 4.5.A). *P. ferrugineus* accessions in clade 2 exhibited markedly the strongest support (decay value ≤ 182; Figure 4.5.B), and after 20 steps the distinction among accessions of the ingroup had been totally lost. It is evident that the tree recovered after phylogenetic analysis of ITS nrDNA sequences has, in general, a good and reliable internal stability.

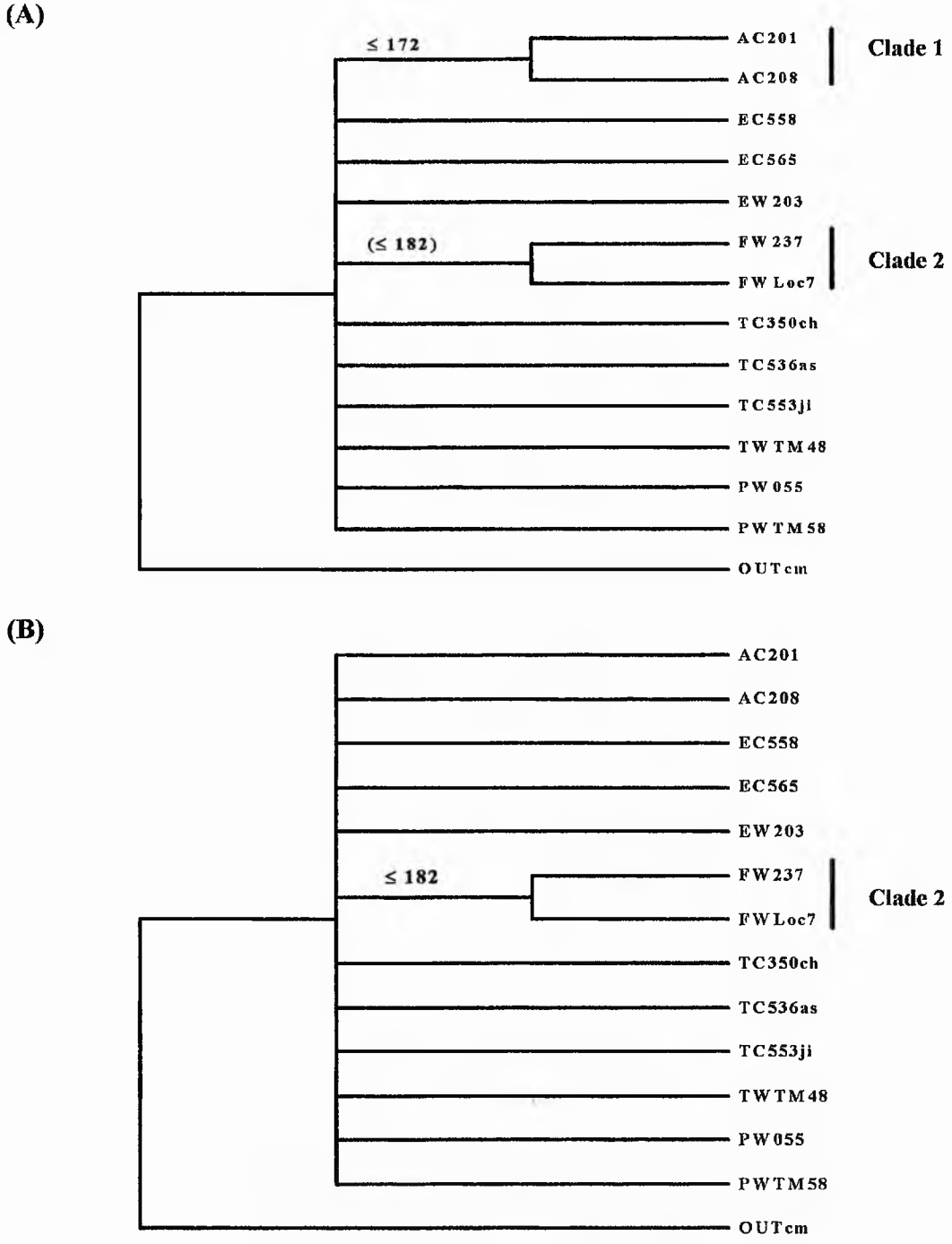


Figure 4.5. ITS trees showing the order of 'decay' of clades. **(A)** This tree is the strict consensus of trees whose length is equal to 172 steps; note that all but *P. ahipa* accessions of clade 1 have been lost. At 173 steps clade 2, containing all *P. ferrugineus* accessions, was still present. **(B)** Strict consensus of trees whose length is 182 steps; *P. ferrugineus* accessions are still clustering together. The tree became an unresolved 'bush' at 21 steps away from the minimal tree, i.e. a length of 183 (see text for further details).

4.4. Discussion

4.4.1. General characteristics of ITS in *Pachyrhizus*

The sizes of the ITS regions in *Pachyrhizus* fell within the range of those reported for other legume genera. For the taxa examined, ITS1 ranges from 203 to 205 bp in length in *Pachyrhizus*, compared to lengths of 221 to 231 bp in *Astragalus* (Wojciechowski *et al.*, 1993), 205 bp in *Vigna* (Schiebel & Hemleben, 1989) and 235 bp in *Vicia* (Yokota *et al.*, 1989). Similarly, ITS2 ranges from 219 to 221 bp in *Pachyrhizus*, compared to 207 - 217 in *Astragalus*, 220 bp in *Vigna* and 208 in *Vicia*. The size of the *Pachyrhizus* 5.8S subunit (164 bp) is the same as the length of this highly conserved region reported for other flowering plants and legumes, e.g. in Madiinae (Baldwin, 1992) and *Vicia* (Yokota *et al.*, 1989).

The level of divergence of ITS sequences among *Pachyrhizus* species (2.9% to 33.1%; Table 4.3.A) is comparable to sequence divergence values reported in *Saintpaulia* and *Streptocarpus* (0% to 27.2%; Möller & Cronk, 1997), in *Gentiana* (1.1% to 48.9%; Yuan *et al.*, 1996) and in *Antennaria* (0% to 14%; Bayer *et al.*, 1996). Moreover, in the case of *Pachyrhizus*, ITS2 was more variable than ITS1 (Tables 4.3.B and 4.3.C), although both regions were sufficiently variable to make the ITS region a useful tool for phylogenetic reconstruction at the species level.

4.4.2. G+C content and secondary structure formation vs. sequence alignment

Overall G+C content values for ITS sequences of *Pachyrhizus* species presented here (49.6% to 54.4%; Table 4.1) are comparable to similar values reported in other angiosperms, particularly legume genera (50% to 59%) (Wojciechowski *et al.*, 1993; Schiebel & Hemleben, 1989; and, Yokota *et al.*, 1989). These *Pachyrhizus* values are towards the middle of the spectrum for plants which has been reported to range approximately from 36% to 75% by Baldwin *et al.* (1995).

According to Baldwin *et al.* (1995), a practical concern about taxa with G+C rich ITS regions is the difficulty of obtaining interpretable DNA sequences because of the propensity for strong intrastrand Watson-Crick base pairing during sequencing reactions and electrophoresis. Indeed, sequencing of the ITS region can be complicated by within-strand Watson-Crick base pairing, because of alteration of the polymerase activity or of the electrophoretic mobility of DNA fragments in sequencing gels. Spacer segments with G+C richness may form secondary structures (i.e. subunit boundaries brought into close proximity within a processing domain; González *et al.*, 1990) under some reaction conditions, e.g. low-temperature sequencing reactions or lack of use of dGTP analogues. However, in some instances such intramolecular pairing is expected based on the probable functional behaviour of transcripts of these nrDNA sequences (Baldwin *et al.*, 1995; Thweatt & Lee, 1990).

To summarise, G+C richness in ITS sequences can lead to ambiguous consensus sequences and further alignment complications because of non-independent and/or overlapping DNA segments. The possibility of this having occurred in some *Pachyrhizus* accessions cannot be ruled out, e.g. the 16 accessions not included in this survey due to alignment ambiguities. Alternatively, within-individual variation observed in some accessions may be explained by multiple rDNA loci, extensive interlocus gene conversion (Sytsma & Schaal, 1990) or, perhaps, extreme PCR amplification bias favouring one repeat-type unit (Baldwin, 1992). Therefore, ITS sequences have to be examined with care, although the relative ease with which such sequence information can now be obtained has demonstrated that it is a powerful nuclear DNA resource for comparison with, for example, cpDNA phylogenetic data. The importance of such comparisons for resolving chloroplast capture via introgression, hybrid speciation, lineage sorting and evolutionary parallelism has become increasingly evident (Rieseberg and Soltis, 1991).

4.4.3. Molecular evolution of ITS in *Pachyrhizus*

The ITS region in *Pachyrhizus* has evolved primarily by point mutations, judging from the moderately high levels of ITS sequence divergence between and even within

species (Table 4.3), the reduced proportion of sites that required gaps (from 0.91% to 1.22%; section 4.3.2), and the absence of evident ITS length variants within the DNA accessions examined. Length variation of ITS sequences among *Pachyrhizus* taxa surveyed here ranged from 1 bp to 4 bp (Table 4.1). This variation was due primarily to short, scattered indels of 1 bp to 2 bp. No case of long length variation was observed, such as that in *Lisianthius*, in which a length variation of 100 bp in ITS1 was reported (Systma & Schaal, 1990).

Nucleotide substitution was the main source of sequence divergence in the ITS. It has been shown that ITS1 is generally more variable than ITS2 both in length and substitution, e.g. in Asteraceae (Baldwin, 1992 & 1993) and in Winteraceae (Suh *et al.*, 1993). This was, however, not the case in *Pachyrhizus* where the numbers of variable characters in both spacers were relatively similar (39.8% and 40.8% for ITS1 and ITS2, respectively) and where ITS1 displayed slightly lower divergence values than ITS2 among pairwise sequence comparisons (Tables 4.3.B and 4.3.C).

In addition, a low fixation of ITS length mutations since divergence of a particular species under examination was evident from the small proportion of nucleotide sites wherein insertion of gaps was necessary to align *Pachyrhizus* sequences. Such sequence conservation among ITS sequences of closely related species is not surprising, given that evidence from experimental and computer-simulation studies (Thweatt & Lee, 1990; González *et al.*, 1990) suggests that ITS sequences are under some evolutionary constraint because of an important role in processing mature rRNAs from primary transcripts. Secondary, 'crucifix' or 'tRNA-like core' structures (Venkateswarlu & Nazar, 1991) assumed by both ITS units in the primary rRNA transcripts may be critical to rRNA maturation by bringing the ends of the 18S, 5.8S and 26S rRNA regions into close proximity for processing.

These levels of structural stability and conservatism in the ITS region were convenient for the proposed phylogenetic study, wherein assessment of positional homologies (= alignment) among *Pachyrhizus* DNA sequences was considered absolutely critical.

Although the internal transcribed spacers are thought to be important in post-transcriptional processing and thus conserved to some extent (Baldwin *et al.*, 1995; Möller and Cronk, 1997), the levels of sequence variation between *Pachyrhizus* taxa were considerable and somewhat similar to intergeneric levels found in other angiosperms. For example, in the subfamily Apioideae (Apiaceae) sequence differences among genera ranged from 0% to 33.2% (Downie and Katz-Downie, 1996); and, in Brassicaceae sequence differences between *Sinapis alba* L. and *Arabidopsis thaliana* (L.) Heynh. were 24.3% for ITS1 and 18.9% for ITS2 (Rathgeber & Capesius, 1989). The high sequence divergence found for *Pachyrhizus* might indicate that either: **(1)** the genus is comparatively old, suggestive also of a relatively continuous distribution of the early ancestors in the Neotropical centre(s) of origin; or, **(2)** that ITS sequences have evolved particularly rapidly in the genus. However, within wild *Pachyrhizus* species sequence divergence was low (e.g. 2.9% for TWTM48 vs. PWTM58; or, 4.4% for EW203 vs. TWTM48; Table 4.3.A), which may indicate simultaneous and rapid radiations from a common ancestor of the wild species (see next section).

4.4.4. ITS phylogenetic relationships and phylogeography within *Pachyrhizus*

The present study reveals that ITS sequences have been useful in reconstructing a molecular phylogeny for *Pachyrhizus* taxa. Cladistic analysis of ITS sequences in the genus provided a well resolved strict consensus tree (Figure 4.4), which illustrates a hierarchical summary of relationships shown by the five equally parsimonious ITS trees. Polytomies observed in the basal nodes of clade 1 represent regions where clustering differences occurred, possibly attributable to low sequence divergence values in certain DNA accessions; however, most taxon relationships within the strict consensus were well resolved. Several conclusions can be drawn about the biogeographic history of the genus from the consensus cladogram. ITS1-5.8S-ITS2 variation resolved two major clades comprised of species from different geographical regions, as discussed below.

Clade 1

This first ITS clade comprised a total of 11 taxa of South and Central American distribution; all *Pachyrhizus* species but *P. ferrugineus* were included in this group clearly separated from the rest of the genus which appears to have split away early in its evolutionary history. Clade 1 has a strongly supported internal structure, as revealed by high bootstrap values ranging from 52% to 98%, and contains six subgroups in the strict consensus (Figure 4.4).

A first subclade contained cultivated accessions of South American distribution only including the cultigen types of the *P. tuberosus* complex and a single, highly derived *P. ahipa* group. TC536, an ashipa, resolved as sister group to *P. ahipa* (AC201 and AC208), supported by a high bootstrap value of 82% (yet, after two steps away the branch containing TC536 in the consensus tree collapsed, i.e. a decay index ≤ 165). The current ITS evidence presented here would suggest that *P. ahipa* is possibly derived from *P. tuberosus*, in view of their close resolution as sister taxa. Placement of these two species together is congruent with their close association in the phylogeny obtained after analysis of total genomic cpDNA (chapter 3); however, the plastome phylogeny proposed a jíquima (western Ecuador) as the sister taxon of *P. ahipa*, whereas the ITS phylogenetic signal points to an ashipa (Amazonian lowlands). It remains to be determined whether this conflict between ITS and cpDNA trees is attributable to homoplasy, interspecific hybridisation, chloroplast capture or lineage sorting.

A second subclade, which comprised cultivated materials of *P. erosus*, was also clearly defined within the strict consensus tree (Figure 4.4) and identified these accessions as a monophyletic entity. This Mesoamerican group, paraphyletic to the first subclade described above, was supported by seven synapomorphies and was recovered in 96% of the bootstrap replicates (decay index ≤ 165). Again, this *P. erosus* subclade has high consonance with a similar grouping observed in the previous cpDNA-based phylogenies. In addition, it is not surprising that the ITS sequences of *P. erosus* differ from those of the subclade *P. ahipa/P. tuberosus* since they reflect the

differences in morphology and geographical distribution that are also evident between the two groups.

The resolution of relationships among the remaining subclades within clade 1 is not completely satisfactory, and a polytomy for wild species is present in the strict consensus tree (Figure 4.4). Nevertheless, several interesting findings are evident. One intriguing finding is the 'anomalous' placement of EW, PW and TW taxa, observed in the five equally parsimonious trees and in their consensus (Figures 4.3 and 4.4). This suggests that after the ancestor of these wild species split from the most primitive phylogenetic entity within the genus, i.e. *P. ferrugineus*, it became isolated to radiate rapidly in different Mesoamerican ecosystems in close proximity to the centre(s) of origin. Moreover, insufficient time has passed for the accumulation of enough mutations within the ITS region of the rDNA gene to satisfactorily differentiate these wild species. Besides, the ancestral groups that diverged early in the Neotropics may have remained diminished, while those splitting off later radiated rapidly and generated the bulk of diversity now evident further away from the putative areas of origin. It would certainly be of interest to seek a more accurate picture of such rapid radiations using faster evolving DNA sequences, and to compare rates of molecular evolution in cultivated relative to wild materials of *Pachyrhizus*.

Another interesting finding concerns the positioning of the wild accession of *P. tuberosus* included in the ITS survey. Indeed, the association of TWTM48 with other wild taxa at basal levels of clade 1 was surprising as it was considered to be more closely related to the cultigen types of the *P. tuberosus* complex, placed at terminal levels of the same clade. This could stem for concerted evolution having occurred within wild populations of *P. tuberosus* such that their ITS homogenised towards their ancestor (viz. *P. panamensis* or *P. ferrugineus*; Figure 4.4). Further analysis, involving a larger number of representative samples of wild *P. tuberosus* is required to investigate this possibility in greater detail.

An interesting feature of the strict consensus tree was the basal-most positioning of PW and TW accessions. PW055 is always basal to clade 1 which could possibly

indicate that the other wild taxa are derived from an early ancestor of *P. panamensis*, after splitting from *P. ferrugineus*. With the exception of this latter species, *P. panamensis* and wild *P. tuberosus* were the most primitive in the genus, which is congruent with the phylogenetic treatments described previously. In addition, this supports the phytogeographic hypothesis postulated from cpDNA data, i.e. that both species (*P. panamensis* and *P. tuberosus*) originated from a continuously distributed early ancestor (i.e. ecotypes of *P. ferrugineus*) after rapid radiation, and diverged later parapatrically as a response to environmental changes (deciduous vs. evergreen rainforest) into two evolutionary branches in the following fashion: FW, EW and EC taxa in Mesoamerica; and, PW, TW, TC and AC taxa in South America. Domestication and subsequent man-made selection aided later the specialisation of primitive landraces and cultivars that are now evident.

Clade 2: *P. ferrugineus*

The two accessions of *P. ferrugineus* (FWLoc7, collected in Guatemala and FW237 from Martinique) were resolved together as the most basal taxa within *Pachyrhizus* (Figure 4.4). *P. ferrugineus* is known to be a highly divergent species comprising several ecotypes (Sørensen, 1990 & 1996), which exhibit considerable morphological variation (e.g. the significant variation in leaflet outline, even within populations, across its wide range of geographic distribution). In contrast, intraspecific ITS sequence divergence values for *P. ferrugineus* are low (7.4%; Table 4.3.A-C). This is not surprising since highly repetitive genes such as rDNA may maintain homogeneity by concerted evolution, in the manner of unequal crossing-over or gene conversion (Arnheim, 1983). Multigene family members evolving in the concerted mode should show a greater degree of interspecific variation than intraspecific variation. Since the ITS regions are functionally less constrained than the coding regions, then it is very likely that homogenisation through the aforementioned mechanisms occurred in the accessions of clade 2.

The considerable divergence between *P. ferrugineus* and the other *Pachyrhizus* species based on ITS sequence comparisons tends also to be reflected in comparisons

of interspecific compatibility and morphology. For example, all species of *Pachyrhizus* with the exception of *P. ferrugineus* are cross-compatible (Sørensen, 1996), resulting in fertile interspecific hybrids (however, no naturally occurring hybrids have been recorded in areas where two species co-occur). In addition, *P. ferrugineus* is the only species in the genus which is evergreen and where the parts of the plant above ground are perennial. Clearly, the ability to interbreed and the herbaceous botanical nature of the remaining four species is a long-retained, apomorphic character state, a distinction also supported by ITS sequence data.

ITS data have provided additional insights into the evolutionary history of yam beans. The strict consensus tree depicts at least three major independent radiations from a common ancestor (*P. ferrugineus*) that took place in Central and South America. Under this hypothesis, the first radiation led to a simultaneously-generated mosaic of species (PW, TW and EW); a second adaptive radiation involved differentiation of a Mesoamerican evolutionary branch (i.e. *P. erosus* and its dispersal to northern habitats) and a South American evolutionary branch. A third radiation led to the specialisation of the South American group along the Andean mountain ranges and Amazonian basin; this recent, most specialised clade successfully spread into very different niches creating the bulk of diversity evident within the cultivated *P. tuberosus* complex and *P. ahipa*. Moreover, these adaptive radiations appear to have been associated with key changes in vegetative and floral morphology, probably driven by ecological factors (e.g. extreme change of habitat conditions), and, in a subsequent stage, domestication and selection.

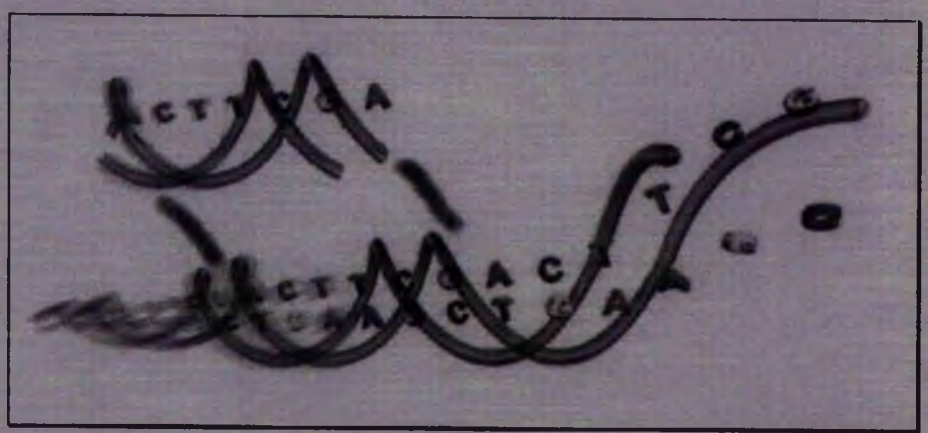
Finally, it is important to consider the potential impact of hybridisation on phylogenetic estimation. Hybridisation, especially if it is ancient, can be difficult to detect from ITS sequence data because concerted evolution rapidly homogenises nrDNA repeats (Hillis & Dixon, 1991). Multiple hybridisations between the same taxa, followed by concerted evolution and lineage sorting of the ITS region (Neigel & Avise, 1986) could lead to discordance between phylogenies based on ITS and morphology. Hybrid derivatives could eventually be of two types, one with the nrDNA repeat of one parent and the other with the nrDNA repeat of the other parent.

Wendel, Schnabel and Seelanan (1995) demonstrated that interlocus concerted evolution can occur bidirectionally subsequent to hybridisation. In their study, fixation of Old World and New World nrDNA repeats in different cotton (*Gossypium*) allopolyploids was observed with consequent discordance between organismal and gene phylogenies. In the light of the current ITS data, it would certainly be of interest to determine the occurrence of such instances in *Pachyrhizus*.

Additional data are required to bring into better focus overall evolutionary relationships within *Pachyrhizus*. Combined analysis of diverse data and inclusion of additional accessions will hopefully amplify phylogenetic signal, dampen random noise and confirm or refute the proposed phylogenetic hypotheses.

CHAPTER 5

USE OF RAPDs IN A PHENETIC ANALYSIS OF *PACHYRHIZUS* TAXA



RAPD priming sites
J. Estrella

5.1. Introduction

Molecular markers have been developed into powerful tools to analyse genetic relationships and genetic diversity within and between species. In addition to several other approaches using polymorphic DNA markers, the technique of Random Amplified Polymorphic DNA (RAPD; Williams *et al.*, 1990) is used in genetics, evolution and systematics to determine taxonomic identity, assess kinship relationships, analyse mixed genome samples, create specific probes and test phylogenetic hypotheses (Bachmann, 1994; Hadrys *et al.*, 1992).

5.1.1. Principles and advantages of the RAPD technique

PCR with single, short (usually 10-mer) arbitrary primers relies on the statistical chance that the complementary primer sites occur somewhere in the genome as inverted repeats enclosing a relatively short stretch of DNA (up to a few thousand base pairs), such that the DNA fragment can be amplified allowing comparison with similar fragments obtained from other individuals, populations, DNA accessions, etc. (Hillis *et al.*, 1996; Bachmann, 1994). Primer sites are thought to be randomly distributed and polymorphism between RAPD profiles can be attributed to a range of processes, including nucleotide substitution (which create or destroy primer sites), and insertion, deletion or inversion of either priming sites or segments between priming sites (Williams *et al.*, 1993; Weising *et al.*, 1994). In addition to polymorphisms due to presence/absence of a DNA product and their size distribution, polymorphisms with respect to product intensity can also be expected within a RAPD profile (Caetano-Anollés *et al.*, 1991; Williams *et al.*, 1990). This variation in product intensity might be the result of low copy numbers of products, competition between RAPD sites, heterozygosity or partial mismatching of primer sites (Adams & Demeke, 1993; Bachmann, 1994; Hadrys *et al.*, 1992; Williams *et al.*, 1993).

RAPD analysis is a simple procedure. Nanogram amounts of total genomic DNA are subjected to PCR using one synthetic oligonucleotide. No prior knowledge of the genome subjected to analysis is required; and, the profile of amplification products

depends on the template-primer combination which is reproducible for any given combination. Amplification products, once resolved on agarose gels, serve as dominant genetic markers which are inherited in a Mendelian fashion (Williams *et al.*, 1990); presence of a particular band is dominant, while absence is recessive (Tingey & del Tufo, 1993). Amplification of non-nuclear RAPD markers is negligible due to the relatively small size of non-nuclear genomes (Hadrys *et al.*, 1992).

The resolution of RAPDs offers several advantages over other molecular techniques detecting DNA variation in plants. The main advantages include the suitability for work on anonymous genomes and its application to problems where only small quantities of DNA are available. This is an attractive option especially when working with limited material such as herbarium specimens, rare or endangered plants (where destructive sampling is not desirable), or individuals with relatively few leaves (Weising *et al.*, 1994; Hillis *et al.*, 1996). The method is quick and efficient in that it can be completed within hours, and is fully automated such that a large number of samples can be handled simultaneously. Further, the analysis of RAPDs involves a nonradioactive assay which requires a simple experimental set-up with ease of use in a modestly equipped laboratory.

The RAPD technique is an attractive complement to conventional fingerprinting aimed at major progress in two directions: **(1)** increase in analytical power per unit effort; and, **(2)** simplification of technology and, ultimately, reduction in expense (Weatherhead & Montgomerie, 1991; Hadrys *et al.*, 1992). Many reports maintain that the RAPD technique is considerably cheaper than other molecular approaches (Hadrys *et al.*, 1992; Rafalski & Tingey, 1993; Williams *et al.*, 1990 & 1993); however, this is not always the case. For example, Ragot and Hoisington (1993) compared radioactive and nonradioactive RFLPs vs. RAPDs in terms of cost and time efficiency, based on maize (*Zea mays* L.) genotyping experiments. They found that the increase in total cost with increasing numbers of samples was higher for RAPDs than for RFLPs. RAPDs were generally more cost and time efficient for studies involving small sample sizes (~ 25 to 50 individuals), while RFLPs had the advantage for larger sample sizes. However, the number of times Southern blots were reused and

the cost of DNA polymerase per reaction, etc., affected the relative merits of using RFLPs and RAPDs, and neither protocol held absolute advantage over the other. It was considered that the efficiency of the technique depended on factors such as labour cost, automation, re-use of reagents, etc. (Ragot and Hoisington, 1993).

5.1.2. Assumptions and limitations of RAPDs

As with other molecular assays, the basic assumptions of RAPD fragment analyses are that the characters in question are heritable, repeatable and independent. Violation of these assumptions impacts significantly on phylogeny reconstruction and population genetic studies (Dowling *et al.*, 1996; Bachmann, 1994).

The assumption of heritability has two components: fidelity of transmission and mode of inheritance. The first of these is most likely to be violated when using rapidly evolving characters such as VNTR loci, because of their high mutation rates (Jeffreys *et al.*, 1988). Knowledge of the mode of inheritance is critical for RAPDs, which are often dominantly expressed (Hadrys *et al.*, 1992). Without this information, it may be impossible to distinguish alleles of a single codominant locus from independent products of non-homologous loci (Riedy *et al.*, 1992), unless progeny testing is conducted.

Repeatability is a problem that has raised concern in the use of RAPDs. The technique is sensitive to reaction conditions and often generates spurious and unrepeatable products if experimental parameters are not carefully standardised (Dowling *et al.*, 1996; Bachmann, 1994; Hadrys *et al.*, 1992). These unclear and non-reproducible fragments, which may derive from non-specific priming or from heteroduplex formation between related amplification products (or other secondary structure artefacts, which can prevent normal amplification patterns) are not useful as genetic markers. Moreover, several authors (e.g. Innis *et al.*, 1990) have reported the presence of 'ghost' bands, i.e. the production of DNA fragments in the total absence of template in a RAPD reaction. Standardising experimental conditions, concentrating on unequivocal and consistent polymorphisms, and treating these as individual

qualitative markers is probably the simplest and safest approach for most investigations (Weising *et al.*, 1994; Bachmann, 1994).

The assumption of independence of characters has both technical and biological (e.g. linkage) dimensions. This assumption is potentially violated at the technical level for characters generated by several molecular approaches, especially those using band-sharing data (Dowling *et al.*, 1996; Danforth & Freeman-Gallant, 1996). In RAPDs, where the genetic basis of specific fragments is unknown, it is impossible to assign fragments of specific mobility to a particular locus without progeny testing. In addition, RAPD products might not be independent especially when amplified fragments are associated with repetitive sequences in the genome, or when allelic relationships are unknown, or if there is heteroduplex formation (Ayliffe *et al.*, 1994).

Non-independence of characters can cause significant errors in phylogeny reconstruction, and careful selection of appropriate statistical methods for data analysis is therefore crucial. The discussion of methods for correcting non-independence, such as estimates of covariance, parametric t-tests, permutation tests, confidence limits, etc., go beyond the scope of this section and are treated elsewhere (e.g. Danforth & Freeman-Gallant, 1996; Felsenstein, 1983; Hillis *et al.*, 1996 and references therein).

Primer size also represents a potential difficulty of the RAPD technique. It may be expected that primers of short length will amplify an unreasonable large number of sequences and that larger primers will amplify too few sequences to be routinely informative. Beyond a certain primer size (~ 15-mer) increasing primer length may also increase non-specific primer annealing, consequently increasing the probability of random non-reproducible amplification patterns (Bachmann, 1994). Most studies using standard RAPD conditions have found 10-mer primers to be an appropriate size (Weising *et al.*, 1994; Hillis *et al.*, 1996). In addition, a G+C content of the primer similar to the G+C content of the analysed genome, maximises the frequency of binding sites and hence amplification products.

Finally, template impurities require consideration. Since primers amplify DNA from virtually all sources, amplification may also include foreign co-isolates, infections and parasites. Fortunately, this may not create problems when clean plants have been raised under controlled conditions or when DNA has been carefully extracted and purified. However, it should always be kept in mind when conclusions are based on anonymous bands (Weising *et al.*, 1994; Bachmann, 1994).

5.1.3. Homology of RAPD fragments

Homology of RAPD fragments is a crucial assumption and, consequently, is treated separately. The RAPD approach assumes that amplified fragments are unique, i.e. that the procedure does not amplify two distinct fragments which comigrate on gels because of similar size (Hadrys *et al.*, 1992). If two individuals exhibit fragments with identical mobilities, it is assumed that these fragments identify homologous stretches of DNA, a consideration that may not hold for anonymous or rapidly mutating segments of DNA. Under some circumstances, the fragments compared may be paralogous (homology via gene duplication) rather than orthologous (homology via speciation); this is a problem since non-homologous products may be amplified by a single primer, leading to misinterpretation of banding patterns and gross errors in subsequent analyses (Rieseberg, 1996; Dowling *et al.*, 1996).

Orthology of putative DNA amplimers should be tested rather than assumed, particularly in comparisons between species (Black, 1993). Southern analysis using the specific RAPD product in question as a hybridisation probe, or alternatively, cleaving gel-isolated products with restriction enzymes and observing congruent band profiles have been suggested as useful tools to test this assumption (Dowling *et al.*, 1996; Fritsh & Rieseberg, 1992; Rieseberg, 1996).

5.1.4. Applications of RAPD studies to plant diversity and evolution

Due to its advantages and technical simplicity, the procedure that resolves RAPDs has been applied widely in various studies of plant diversity; for example, in studies of

plant population genetics, taxonomic identity, construction of gene linkage maps, fingerprinting plant genomes for cultivar identification, and the determination of hybrids. The method has also been employed in the definition of core collections, the study of hybrid origin of species, and in phylogeny reconstruction. Examples of some of these applications are presented below, followed by a brief description of recent RAPD studies in *Pachyrhizus*.

Population genetics

RAPDs have contributed considerably in studies of plant population genetics. For example, Wolff *et al.* (1997) used RAPDs to study subpopulation genetic structure in one population of the short-lived perennial *Alkana orientalis* (L.) Boiss. from the Sinai Desert, Egypt. Results from previous studies suggested that bee pollinator behaviour was likely to cause limited gene dispersal and that subpopulations might have diverged from each other genetically (Willmer *et al.*, 1994). Seven RAPD primers were used and differences between subpopulations were found for several of the 45 polymorphic fragments scored. Population subdivision became evident from cluster analysis, and genetic distances revealed that there was significant genetic differentiation between all subpopulations; however, it was also found that more extensive gene flow appeared to have taken place within the population than was expected. Finally, the study indicated that the pattern of both pollen and seed flow were probably influential in moulding the genetic subdivision of the population of *A. orientalis* (Wolff *et al.*, 1997).

Similarly, RAPDs were used to study the population genetic structure of the forest tree *Eucalyptus globulus* (Nesbitt *et al.*, 1995), a species that has been divided taxonomically into four subspecies. In many localities where it occurs, it appears to be intermediate in morphology between subspecies *globulus* and the remaining three subspecies. However, RAPD analysis did not reveal an intermediate phenotype for specimens from these localities. Moreover, within *E. globulus*, the majority of RAPD variation was greater within rather than between populations. In contrast, in American cranberry (*Vaccinium macrocarpon*), a vegetatively spreading clonal plant that

reproduces sexually by selfing, RAPD profiling revealed many polymorphisms within and among populations (Stewart & Excoffier, 1996). Quantification of population genetic homogeneity in both of these studies had important repercussions in their conservation biology and ecology, since genetically homogeneous populations are considered to be less stable and flexible than heterogeneous ones (Millar & Libby, 1991).

Fingerprinting plant genomes and estimation of genetic relatedness

Unequivocal identification of plant cultivars is important for practical breeding purposes as well as for related areas like plant proprietary rights protection. Further, assessment of genetic diversity among cultivars and their wild relatives has recently attracted attention in efforts to cope with the commonly encountered reduction of diversity resulting from sampling and breeding processes (IPGRI, 1995; Weising *et al.*, 1994). Cultivar identification and estimation of genetic relatedness has been efficiently achieved with RAPD profiling in a large number of crops, e.g. in *Malus* (Koller *et al.*, 1993), *Rubus* (Parent *et al.*, 1993), *Musa* (Kaemmer *et al.*, 1992), *Phaseolus* (Gepts *et al.*, 1993) and *Carica* (Sharon *et al.*, 1992), among many others. Only two representative examples are given here; for additional information the reader is addressed elsewhere (e.g. Weising *et al.*, 1994).

As in many other perennial crops, proper cultivar identification in cocoa (*Theobroma cacao*) traditionally relies on morphological characters that cannot be assessed until tree maturity. Using RAPD analysis with 10 primers, a set of 13 genotypes were successfully distinguished at the seedling stage (Wilde *et al.*, 1992). Wild cocoa species proved to be relatively dissimilar to the cultivated genotypes, whereas the latter exhibited various levels of relatedness. Bandsharing values among the investigated plants ranged from 51% to 86%.

Commercial cultivation of cocoa is based on a very restricted genepool. To assess the amount and distribution of genetic variation in the species, a total of 25 accessions from three populations collected in Ecuador and Peru was analysed with 25 RAPD

primers (Russell *et al.*, 1993). PCO analysis clearly discriminated between the geographical origins of the germplasm under study. Further, Shannon's index of phenotypic diversity was used to quantify the level of polymorphism detected and to partition it into between- and within-population components. On average, diversity was higher within rather than between populations.

The definition of core collections

Major collections of important crop plants are held in genebanks around the world. They are repositories of biodiversity and a valuable source of genes for plant breeders and farmers. Workers at many of these institutions face problems in the efficient conservation and use of germplasm because of the large numbers of accessions in a collection (IPGRI, 1995). Two complementary solutions have been proposed to enhance the efficiency with which such collections can be handled: **(1)** the identification and removal of duplicates; and, **(2)** the definition of core collections. In the latter case, a subset of germplasm, comprising approximately 10% of the total collection, would be selected to represent, with minimal redundancy, as much as possible of the diversity within the whole collection (Virk *et al.*, 1995).

Accessions within a core collection would be the first to be supplied in response to requests by breeders for material from the collection. Due to its smaller size, the core collection could also be the target of back-up conservation, as it could be transferred and stored at other institutions (Vaughan & Jackson, 1994). Moreover, the proposal of a core collection does not involve a reduction in the actual size of the whole collection, but optimises the efficiency of its conservation, evaluation and use by prioritising representative subsets of accessions for special attention (Frankel, 1984; Virk *et al.*, 1995). In this regard, molecular markers (e.g. RAPDs) have proved useful for the identification of duplicates and the definition of core collections.

An example of the production of a core collection aided by molecular markers is presented in Virk *et al.* (1995). They applied the RAPD technology to accessions of rice (*Oryza* spp.) obtained from the major world collection held at IRRI, the

International Rice Research Institute. A total of 24 primers were used which yielded 83 reproducible marker bands. Cluster analysis using the UPGMA and TWINSPAN methods was employed to generate dendrograms, which readily separated accessions into two major groups. Whereas this RAPD classification did not always correlate exactly with classifications based on morphology, they agreed well with those based on isozymes and crossing data. These findings resulted in an immediate practical application by rice breeders who used the RAPD technique to provide a fast and reliable method for classifying relatively uncharacterised accessions of *O. japonica* or *O. indica*. In a wider context, RAPD variation provided data for establishing guidelines and strategies for effective germplasm management.

Construction of plant phylogenies

The use of RAPDs in the study of phylogeny and species relationships has become increasingly widespread. For example, Graham and Nichol (1995) carried out cluster analysis in species of raspberry (*Rubus*) using RAPD markers. The 13 species used represented three subgenera (*Idaeobats, Eubats* and *Anoplobats*) of a total of 12 recognised in *Rubus*. Ten primers were used in their experiment, which generated 372 polymorphic markers; the resulting dendrograms separated the species into the three subgenera with the exception of *R. macrae*, a rare tropical species traditionally placed within the subgenus *Idaeobats*. Further, RAPD data indicated that *R. macrae* had only 26% similarity to other species of the subgenus *Idaeobats*, which was equivalent to its level of similarity to species of *Eubats*.

RAPD markers can be used successfully in studies aimed at elucidating the origin of hybrid species. For example, Wang *et al.* (1994) used RAPDs to investigate the hybrid origin of *Paulownia taiwaniana*, a fast-growing timber species restricted to East Asia (Chen, 1986). Genomic DNA of this species and its two putative parental taxa was amplified using 23 primers. A total of 351 fragments were produced and of these 265 (75.5%) were polymorphic. Almost all PCR-amplified products of *P. taiwaniana* were shared by either *P. fortunei* or *P. kawakamii*, or both; and, the number of polymorphic fragments shared by *P. taiwaniana* and *P. fortunei* was about

equivalent to those shared by *P. taiwaniana* and *P. kawakamii*. These findings, together with the results of a cpDNA variation survey, provided conclusive evidence pointing to a hybrid origin of *P. taiwaniana,* with *P. fortunei* and *P. kawakamii* as parental taxa (Wang *et al.*, 1994).

RAPD analysis has also been used by Halward *et al.* (1992) to examine systematic relationships in *Arachis hypogaea* and related wild species. The study was conducted on two peanut cultivars, 25 divergent lines of *A. hypogaea* and 29 diploid species of wild *Arachis*. No evident RAPD variation was revealed between the cultivars, nor the 25 divergent lines of *A. hypogaea*; nevertheless, the wild materials were uniquely identified with most primers. In addition, cladistic and phenetic analyses of the RAPD data set gave nearly identical results which agreed with previous classifications based on morphology, isozymes and RFLP markers.

Despite the increasing popularity of RAPDs as a tool for phylogeny reconstruction, the assumptions made about the data that are generated (see previous sections) need to be treated with caution. Firstly, the presence of non-homologous comigrating fragments might provide information on convergence rather than close relationships between taxa; therefore, homology has to be tested (Dowling *et al.*, 1996; Rieseberg, 1996). Secondly, RAPDs should not strictly be used in a cladistic analysis, although this has been done in several studies (e.g. in autumn buttercup, *Ranunculus* spp.; Van Buren *et al.*, 1994). This is because RAPDs are generated from anonymous regions of the genome and homoplasic characters will therefore be present in unknown proportions. In contrast, phenetic analysis (section 3.1.4) can be performed with RAPD markers, i.e. clustering taxa on the basis of their overall similarities (Stace, 1989). In some studies, cladistic and phenetic analyses have given similar results (e.g. in Mexican *Pinus*, Furman *et al.*, 1997; in *Stylosanthes*, Gillies & Abbott, 1998), so although phenograms should not necessarily represent a common ancestry, they sometimes do.

RAPD studies in the genus *Pachyrhizus*

So far, three RAPD studies have been carried out on *Pachyrhizus*, aimed either at elucidating species relationships or at morphological and agronomical characterisation of representative groups of accessions. These surveys, performed under the aegis of the European Union- funded Yam Bean Project (Sørensen, 1996), have provided further insight into the genetic relationships and evolutionary history of the genus. Some of these findings are presented below.

Phillips (1994) carried out a preliminary RAPD analysis using 79 accessions and 13 primers (Operon Technologies Inc.). Four of these primers yielded consistent and repeatable results and a total of 79 fragments were produced, of which, 45 (57%) were unique to a species or group of species. Phenograms were generated using both UPGMA and Neighbour-Joining methods. RAPD screening discriminated groups of similar accessions, both between and within species. Moreover, accessions of *P. erosus* and *P. tuberosus* were classified into distinct species-specific groups and a discrete Mexican grouping of cultivated *P. erosus* was revealed. Further studies were recommended to include variation within accessions and to assess the level of confidence with which intra-specific groupings might be accepted.

A subsequent investigation carried out by Estrella *et al.* (1998) screened four of the five species within *Pachyrhizus* using 65 accessions and six 10-mer oligonucleotides. Neighbour-Joining and PCO analyses were used as agglomerative methods to cluster accessions; the bootstrap method was employed to estimate confidence limits of internal branches of the consensus tree. Again, polymorphisms produced in this study discriminated between accessions of *P. erosus* and *P. tuberosus*. Within *P. erosus*, a discrete grouping of Mexican cultivars was revealed, which agreed with the proposal that *P. erosus* in central Mexico and the Yucatan peninsula is derived from a highly limited introduction of germplasm. In addition, RAPD markers were useful for discriminating between accessions belonging to the jíquima and ashipa cultigen types of the *P. tuberosus* complex.

Finally, the research project 'Systematic Characterisation of Genetic Diversity of *Pachyrhizus erosus* and *P. tuberosus*' has been conducted during these two last years at CATIE (Centro Agronómico Tropical para la Investigación y la Enseñanza, Costa Rica). A RAPD analysis conducted by Ing. M.Sc. César Tapia (pers. comm.) used 10 primers from kits OPA, OPH and UBC (Operon Technologies Inc. & University of British Columbia) over a group of 31 accessions representing cultivated materials of the *P. tuberosus* complex. A total of 33 polymorphic fragments were generated and UPGMA divided accessions into four well-resolved clusters. RAPD polymorphisms discriminated the three cultigen types within *P. tuberosus*, with accessions TC118 and TC525 (both ashipas) resolving as the basal-most taxa in the dendrogram. Additional molecular and morphological analyses (the latter one based on 70 descriptors) are currently on its way.

5.1.5. Objectives and aims of the survey of RAPD variation in *Pachyrhizus*

The objectives and aims of the work reported in this chapter were three-fold:

First, to use RAPDs to generate phenograms describing overall similarities between selected *Pachyrhizus* accessions. Taxa were chosen following two criteria: **(1)** selection of accessions that were included in well-resolved clades of the previous cpDNA and ITS-based phylogenies, thus providing a comparative estimate of the reliability of RAPD analysis for the resolution of evolutionary relationships; and, **(2)** the inclusion of accessions from poorly-resolved regions of the previous phylogenetic treatments, aiming to obtain a clearer description of inter- and intraspecific relationships within the genus.

Second, several experiments were carried out to assess homology among RAPD bands produced by the same primer but in different accessions. Detailed testing of the putative homology of comigrating bands was considered necessary before such an assumption could be accepted.

Finally, the potential of RAPDs as a fast and cost-efficient source of molecular markers for the identification of *Pachyrhizus* taxa was examined. It was hoped that the present RAPD study would provide a source of taxon specific fragments for future investigations on genetic relationships and aid molecular plant breeding.

5.2. Materials and methods

5.2.1. Plant material

A total of 85 accessions of *Pachyrhizus*, comprising ecotypes, landraces and cultivars from a broad range of geographic, ecological and climatically different Neotropical areas were subjected to analysis. A complete list of plant material surveyed is presented in Table 2.1. Some accessions examined had previously been analysed for cpDNA restriction site and ITS sequence variation. Details of plant cultivation and preparation of leaf samples were as described in chapter 2 (section 2.2).

5.2.2. Experimental details

All experimental procedures were carried out as described in chapter 2. Extraction, purification and estimation of DNA concentration were performed as described in sections 2.3 and 2.4. RAPD amplification proceeded as described in section 2.8.1. A total of 16 primers were initially screened for polymorphisms in a pilot survey (Table 2.6); of these, eight revealed consistent RAPD profiles and were selected for final study (Table 5.2). Primers were chosen on the basis of the regularity with which they amplified their target sequences; only primers that gave clear and reproducible banding patterns were used. A total of 50% of the primer-accession combinations, chosen at random, was repeated once to ensure reproducibility of results.

5.2.3. Homology assessment among comigrating RAPD fragments

To confirm that RAPD bands scored among *Pachyrhizus* accessions were homologous, restriction digests were performed on purified RAPD fragments following the protocol described in section 2.8.2. Identical profiles produced by different restriction enzymes were considered evidence for homology. The percentage of comigrating bands that were homologous (H) was estimated using the following equation:

$$H = (N_c / N_h) \times 100$$

In this equation, N_h represents the total number of bands tested for homology and N_c is the number of these bands that gave congruent restriction patterns within a particular pairwise comparison of comigrating fragments. A second approach to test homology, i.e. transfer of comigrating fragments onto nylon membranes by Southern blotting and subsequent hybridisation, was not particularly successful and thus was not used further.

5.2.4. Data analysis and phylogeny reconstruction

Presence/absence of each scorable fragment was recorded in a binary data matrix and the frequency of each band was determined by inspection. Pairwise genetic similarities were calculated from the data matrix using Jaccard's coefficient, $F = M_{xy} / (M_t - M_{xy0})$, where M_{xy} represents the number of fragments shared between two accessions, M_t the total number of bands in the data matrix and M_{xy0} the number of bands in the data matrix that were not evident in either of these accessions (Virk *et al.*, 1995). The basis for the presence of RAPD bands is not clearly understood, but it is considered that the absence of a shared band may not necessarily imply shared ancestry. Therefore, Jaccard's coefficient was considered the best algorithm to use for data analysis, as it calculates pairwise genetic distances based only on the shared presence of bands (Armstrong *et al.*, 1994; Gillies *et al.*, 1997).

The matrix of genetic distances was obtained using options M1-14 and M2-21 of the software package RAPDISTANCE, version 1.04 (Armstrong *et al.*, 1994); only polymorphic DNA fragments were considered. A measure of genetic distance was calculated as 1 - *F*. Cluster analysis was performed on the matrix of pairwise distances with both the Neighbour-Joining (Saitou & Nei, 1987) and UPGMA (Unweighted pair group method with arithmetic averages; Sneath & Sokal, 1973) methods. Further, options SEQBOOT, NEIGHBOUR and CONSENSE from PHYLIP for Windows, version 3.5 (Felsenstein, 1993), were used to build a consensus tree, summarising the relationships of 100 bootstrap replicates. Only the Neighbour-

Joining (NJ) generated phenogram was subjected to the bootstrap procedure due to a computer-time constraint. Phenograms were visualised and printed with the software programme TREEVIEW, version 1.2a (Page, 1996).

The permutation tail probability (PTP) test was carried out using option M3-32 of RAPDISTANCE, version 1.04, to test the internal structure in the data set. This test determines if the data set is meaningful for phylogenetic analysis by testing if a tree calculated from a set of genetic distances reflects a tree-like signal in the data, or is merely an artefact of the algorithm used for phylogeny reconstruction. The procedure consists of comparing the minimum length tree derived from the RAPD data set to that derived from the same data after character state assignments have been randomly permuted within each character (Armstrong *et al.*, 1994; Swofford *et al.*, 1996). The PTP test can be applied to trees because a tree calculated from random distances will have a larger total branch length than a tree with an appropriate, hierarchical structure. The random permutations are repeated many times to obtain an estimate of the mean total branch length of trees representing the randomised data, and also the standard deviation of that mean. Thus, the 'tree-likeness' of the original tree (unrandomised; calculated by pairwise distances) is assessed as a Z-value, namely the difference between the total branch length of the original tree and the mean of the random trees, expressed as the number of standard deviations of the randomised trees (Armstrong *et al.*, 1994; Furman *et al.*, 1997).

5.3. Results

5.3.1. PCR amplification of genomic DNA

Sixteen random decamer primers were initially used in a pilot study (data not shown). Eight of these, with G+C contents ranging from 50% to 70%, showed consistent amplification. Consistency of the fragment patterns was checked in all primer-accession combinations that were repeated at random and was found to be 100% in all RAPD profiles, ensuring reproducibility of results. Details of presence/absence of the DNA fragments in the accessions surveyed are presented in Appendix 6. PCR amplification yielded 148 RAPD bands (excluding 23 monomorphic bands), ranging in size from 506 to 6300 bp.

The total numbers of RAPD fragments amplified from *Pachyrhizus* species in this survey are shown in Table 5.1. More DNA bands were amplified with genomic DNA of cultivated *P. tuberosus* and *P. ahipa* (50 and 51 bands, respectively) than with genomic DNA from the other species. The lowest numbers of RAPD fragments were scored for accessions of *P. ferrugineus* and *P. panamensis* (38 bands in both species). The number of shared fragments ranged from 37 (for FW accessions) to 50 (TC accessions). In addition, the average of amplified fragments/primer ranged from 2.9 (for PW accessions) to 14.1 (for TC accessions). The average number of bands per species was 8.3 for the whole RAPD survey.

Table 5.1. Number of RAPD fragments amplified from *Pachyrhizus* with eight 10-mer primers.

Description	Total number of RAPD fragments							
	AC	EC	EW	FW	PW	TC	TW	Average
Scored fragments	51	49	42	38	38	50	41	**44.1**
Shared fragments	46	49	42	37	38	50	41	**43.3**
Average fragments/primer	12.6	10.8	7.1	4.5	2.9	14.1	6.0	**8.3**
Unique fragments/species	5	0	0	1	0	0	0	**0.8**
Accession-specific fragments*	10	9	23	15	11	16	25	**15.5**

*: Fragments that are present in one accession only but not another within a species, but might be present in another species.

With respect to unique DNA fragments that were amplified in only one species, five and one such bands were scored for AC and FW accessions, respectively (Table 5.1). No unique DNA fragments were observed for EC, EW, PW, TC and TW accessions in this study. Accession-specific products were observed for all species surveyed; the number of accession-specific fragments ranged from 9 to 25, with an average of 15.5 (Table 5.1).

The number of polymorphic RAPD fragments produced per primer over the 85 accessions examined is shown in Table 5.2. Primer OPB08 produced most polymorphic fragments (24), whereas OPH14 produced the least (14). With respect to polymorphic DNA fragments that were amplified per species, cultivated material (i.e. AC, EC and TC accessions) produced consistently more of such fragments than wild taxa (i.e. EW, FW, PW and TW). Further, PW accessions showed the lowest number of bands (23), whereas TC accessions had the highest value (113 fragments; Table 5.2). However, it should be noted that these differences might have been affected by the fact that only 17 accessions of wild status (out of a total of 85) were included in the survey.

Table 5.2. Number of polymorphic DNA fragments produced over material surveyed.

Primer	Number of polymorphic DNA fragments								
	AC	EC	EW	FW	PW	TC	TW	Total*	Average
OPA02	18	12	7	7	1	15	9	**21**	**9.9**
OPA13	10	11	8	2	3	15	5	**17**	**7.7**
OPB07	13	13	9	6	5	18	9	**21**	**10.4**
OPB08	15	11	8	7	4	17	8	**24**	**10.0**
OPH02	15	11	7	5	3	13	5	**19**	**8.4**
OPH03	9	10	8	5	3	13	5	**16**	**7.6**
OPH05	13	10	7	2	2	13	4	**16**	**7.3**
OPH14	8	8	3	2	2	9	3	**14**	**5.0**
Total	**101**	**86**	**57**	**36**	**23**	**113**	**48**	**148**	**66.3**

* Number of independent bands scored over all taxa.

5.3.2. Homology assessment among comigrating RAPD fragments

Homology was interpreted as sequence similarity assessed by consistency of restriction digest fragments obtained from bands sampled from different taxa. A total of 212 RAPD fragments were isolated and purified to test their homology among the group of 85 *Pachyrhizus* accessions. Details of the 212 fragments, arranged in 106 pairwise comparisons, are given in Table 5.3. RAPD reactions of the purified DNA confirmed that bands were of the same size as the original samples. Restriction digests that were successful and congruent for two of the three enzymes used (see section 2.8.2) were considered as evidence for homology. Examples of congruent and incongruent restriction patterns obtained in this study are shown in Plate 5.1.

Restriction digests produced identical restriction patterns for 82 of the 106 pairwise comparisons (marked as '✓', Table 5.3). In two instances (i.e. comparisons 9 and 10) the small, 1018 bp- fragments contained no restriction sites; thus, results were considered ambiguous and were scored as incongruent. Overall, these results gave an estimated value of 77.36% homology for comigrating marker bands, as follows:

$$H = (164 / 212) \times 100 = 77.36\%$$

It is assumed that the remaining 22.64% will introduce noise into the RAPD data set and thereby hinder the probability of generating accurate estimates of genetic relationships. Within species, 74.2% of comigrating bands from different accessions were found to be homologous (i.e. 49 of the 66 intraspecific pairwise comparisons), whereas homology was slightly increased to 77.5% for comigrating bands sampled from different species (i.e. 31 of the 40 interspecific comparisons). It should be noted that this slight increase in homology is unfortunately due to the fact that more intraspecific than interspecific comparisons were sampled during experimental set-up. Additionally, of the 82 pairwise comparisons with congruent restriction profiles, 50 (60.9%) were sampled from accessions of the same species, whereas 32 pairwise comparisons (39.1%) corresponded to homologous bands sampled from different species. Overall, comigrating RAPDs are more likely to be homologous when they

originate from accessions within *Pachyrhizus* species than from between species. In view of this, phylogenetic relationships have to be interpreted with caution and confidence estimates for tree structure and reliability (e.g. the bootstrap procedure and the PTP test) should be calculated.

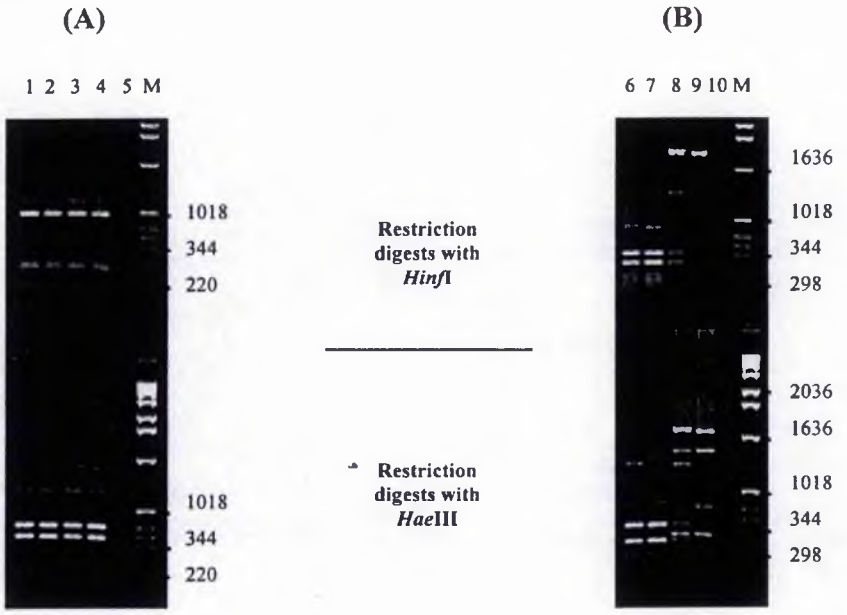


Plate 5.1. Homology test for RAPD fragments.

(A) Congruent restriction fragment profiles. From left: Pairwise comparison 62 [lane 1 (AC207L-3800-H14) vs. lane 2 (AC207S-3800-H14)]; pairwise comparison 63 [lane 3 (AC208-3800-H14) vs. lane 4 (AC220A-3800-H14). Lane 5 (blank); M, molecular size standard (1 kb DNA marker, Gibco).

(B) Incongruent restriction fragment profiles. From left: Pairwise comparison 32 [lane 6 (AC222-3500-B8) vs. lane 7 (EC560-3500-B8)]; pairwise comparison 106 [lane 8 (FWLoc7-4850-B8) vs. lane 9 (PWTM58-4850-B8)]. Lane 10 (blank); M, molecular size standard (1 kb DNA marker, Gibco).

Table 5.3. Homology test among comigrating RAPD fragments.

Pairwise comparison*		Restriction profiles†			H‡
No.	Description	*Hae*III	*Hinf*I	*Dde*I	
1	EC560 4300 H3 vs. EW223 4300 H3	I	I		
2	PWTM58 4072 H3 vs. TC354 4072 H3	C	C		✓
3	EC560 3750 H3 vs. EW203 3750 H3	C	C		✓
4	PWTM58 3650 H3 vs. TC354 3650 H3	C	C		✓
5	PWTM58 4300 H5 vs. TC350 4300 H5	C	C		✓
6	TCNA09 4150 H5 vs. TWToaI 4150 H5	I	I		
7	PWTM58 3400 H5 vs. TC350 3400 H5	C	C		✓
8	TCNA09 3400 H5 vs. TWToaI 3400 H5	I	I		
9	EC560 1018 H3 vs. EW203 1018 H3	A	A		
10	PWTM58 1018 H3 vs. TC354 1018 H3	A	A		
11	PWTM58 2700 H5 vs. PWTM59 2700 H5	C	C		✓
12	EC033G 4600 H2 vs. EC565 4600 H2	C	C		✓
13	AC222 4800 H2 vs. EC559 4800 H2	C	I	I	
14	TCNA06 4800 H2 vs. TC550 4800 H2	C	I	I	
15	TC557 4800 H2 vs. EC236 4800 H2	I	C	I	
16	PWTM58 3200 H3 vs. TCNA06 3200 H3	C	C		✓
17	TC550 3200 H3 vs. TC557 3200 H3	C	C		✓
18	AC222 3400 H5 vs. PW055 3400 H5	C	C		✓
19	TCNA06 3400 H5 vs. TC550 3400 H5	C	C		✓
20	TC557 3400 H5 vs. EW354 3400 H5	C	C		✓
21	PW055 3800 H14 vs. AC222 3800 H14	C	C		✓
22	TC550 3800 H14 vs. TCNA06 3800 H14	C	C		✓
23	AC222 3300 A2 vs. PW055 3300 A2	C	C		✓
24	TCNA06 3300 A2 vs. TC550 3300 A2	C	C		✓
25	TC557 3300 A2 vs. EW203 3300 A2	C	C		✓
26	AC222 3300 A13 vs. PW055 3300 A13	C	C		✓
27	TCNA06 3300 A13 vs. TC550 3300 A13	I	C	I	
28	TC557 3300 A13 vs. EC531 3300 A13	I	C	C	✓
29	AC222 1800 B7 vs. PW055 1800 B7	I	C	I	
30	TCNA06 1800 B7 vs. TC550 1800 B7	I	C	I	
31	TC557 1800 B7 vs. EW051 1800 B7	C	C		✓
32	AC222 3500 B8 vs. EC560 3500 B8	I	I		
33	TC118 3500 B8 vs. TCNA09 3500 B8	I	C	I	
34	EW051 3100 B8 vs. TC309 3100 B8	C	C		✓
35	EC201 4800 H2 vs. EC534 4800 H2	C	C		✓
36	EC565 3500 H2 vs. FWLoc7 3500 H2	C	C		✓
37	TC531 4800 H2 vs. TC553 4800 H2	C	C		✓
38	EC201 3450 H3 vs. EW203 3450 H3	C	C		✓
39	EW354 3450 H3 vs. FW237 3450 H3	I	C	I	
40	PWTM58 3450 H3 vs. PWTM59 3450 H3	C	I	C	✓
41	EC534 4150 H5 vs. EC510 4150 H5	C	C		✓
42	EW223 4150 H5 vs. FW237 4150 H5	C	C		✓
43	FW237 3400 H5 vs. FWLoc7 3400 H5	C	C		✓
44	EC534 4300 H14 vs. EC510 4300 H14	C	C		✓
45	EW223 4300 H14 vs. FW237 4300 H14	C	C		✓
46	FW237 3800 H14 vs. PWTM58 3800 H14	C	C		✓
47	EC201 5400 A2 vs. EW203 5400 A2	C	C		✓
48	EC109 5400 A2 vs. FW237 5400 A2	C	C		✓
49	PWTM59 5400 A2 vs. TC355 5400 A2	C	C		✓
50	EC201 6300 A13 vs. EW223 6300 A13	C	C		✓
51	EC109 6300 A13 vs. FW237 6300 A13	C	C		✓
52	TC353 6300 A13 vs. TC354 6300 A13	C	C		✓
53	EC201 1800 B7 vs. EW203 1800 B7	C	C		✓

*: Band identification in each pairwise comparison is presented in an 'accession - fragment size - primer' fashion.

†: Codes for restriction profiles: A = ambiguous results due to small size of the RAPD fragment tested for homology; C = congruent restriction site profiles for a particular pairwise combination; I = incongruent restriction site profiles for a particular pairwise combination.

‡: H = Homology; ✓ = congruent restriction profiles for at least two of the three restriction enzymes used.

Table 5.3. Homology test among comigrating RAPD fragments. Continued.

Pairwise comparison*		Restriction profiles†			H‡
No.	Description	*Hae*III	*Hin*fI	*Dde*I	
54	EC109 1800 B7 vs. EC205 1800 B7	C	C		✓
55	FW237 3700 B7 vs. PW055 3700 B7	C	C		✓
56	EC201 2300 B8 vs. EW203 2300 B8	C	C		✓
57	EC109 2300 B8 vs. FW237 2300 B8	C	C		✓
58	PW055 2300 B8 vs. PWTM58 2300 B8	I	C	I	
59	AC214 3150 H2 vs. AC215 3150 H2	C	C		✓
60	AC216 3150 H2 vs. AC222 3150 H2	C	C		✓
61	AC228 3150 H2 vs. AC230 3150 H2	C	C		✓
62	AC207L 3800 H14 vs. AC207S 3800 H14	C	C		✓
63	AC208 3800 H14 vs. AC220A 3800 H14	C	C		✓
64	AC225 3800 H14 vs. AC226 3800 H14	C	C		✓
65	AC220A 3450 H3 vs. AC526A 3450 H3	C	C		✓
66	AC201 4300 H3 vs. AC208 4300 H3	C	C		✓
67	AC222 3450 H3 vs. AC223 3450 H3	C	C		✓
68	AC220A 4072 H5 vs. AC223 4072 H5	C	C		✓
69	AC215 4072 H5 vs. AC208 4072 H5	C	C		✓
70	AC222 4072 H5 vs. AC228 4072 H5	C	C		✓
71	AC220A 3300 A2 vs. AC226 3300 A2	C	I	I	
72	AC231 3300 A2 vs. AC213 3300 A2	C	I	I	
73	AC208 3300 A2 vs. AC209 3300 A2	C	C		✓
74	AC216 4750 A13 vs. AC227 4750 A13	I	C	I	
75	AC208 4750 A13 vs. AC207L 4750 A13	C	C		✓
76	AC202 4750 A13 vs. AC203 4750 A13	C	C		✓
77	AC220A 1800 B7 vs. AC226 1800 B7	C	C		✓
78	AC216 1800 B7 vs. AC202 1800 B7	C	C		✓
79	AC102 1800 B7 vs. AC204 1800 B7	C	C		✓
80	AC220A 3600 B8 vs. AC226 3600 B8	C	C		✓
81	AC215 3600 B8 vs. AC213 3600 B8	C	I	I	
82	AC208 4850 B8 vs. AC207S 4850 B8	C	I	I	
83	EC559 3150 H2 vs. EWPro 3150 H2	C	C		✓
84	PW055 3100 H2 vs. TC118 3100 H2	C	C		✓
85	TC532 3100 H2 vs. TWNanII 3100 H2	C	C		✓
86	AC526A 4300 H3 vs. EC006 4300 H3	C	C		✓
87	EC511 4300 H3 vs. TC309 4300 H3	C	C		✓
88	TCNA07 4300 H3 vs. TCNA09 4300 H3	C	C		✓
89	AC102 1636 H5 vs. AC209BR 1636 H5	C	C		✓
90	TC354 1636 H5 vs. TC553 1636 H5	I	C	I	
91	TCNA09 1636 H5 vs. TWToaI 1636 H5	C	C		✓
92	EC006 3600 H14 vs. EC033G 3600 H14	C	C		✓
93	EC560 3600 H14 vs. EW051 3600 H14	C	C		✓
94	FWLoc7 3300 H14 vs. PW055 3300 H14	C	C		✓
95	AC231 4072 A2 vs. EC006 4072 A2	C	C		✓
96	EC565 4072 A2 vs. EW203 4072 A2	C	C		✓
97	FWLoc1 4072 A2 vs. TC350 4072 A2	C	C		✓
98	EC560 4800 A13 vs. EW051 4800 A13	C	C		✓
99	FWLoc1 4800 A13 vs. PW055 4800 A13	C	C		✓
100	PWTM59 4800 A13 vs. TC118 4800 A13	C	C		✓
101	EC565 5090 B7 vs. TC554 5090 B7	C	I	I	
102	EC565 4850 B7 vs. EWHue 4850 B7	C	I	I	
103	TC531 4850 B7 vs. TC550 4850 B7	C	C		✓
104	AC526A 4850 B8 vs. EC006 4850 B8	C	C		✓
105	EC565 4850 B8 vs. EW051 4850 B8	C	C		✓
106	FWLoc7 4850 B8 vs. PWTM58 4850 B8	I	I		

*: Band identification in each pairwise comparison is presented in an 'accession - fragment size - primer' fashion.

†: Codes for restriction profiles: A = ambiguous results due to small size of the RAPD fragment tested for homology; C = congruent restriction site profiles for a particular pairwise combination; I = incongruent restriction site profiles for a particular pairwise combination.

‡: H = Homology; ✓ = congruent restriction profiles for at least two of the three restriction enzymes used.

5.3.3. Taxon-specific RAPD marker production in *Pachyrhizus*

Some of the brightest fragments of the RAPD profiles obtained in this survey were found to be taxon-specific, and could be used as molecular markers for taxon identification or even in future molecular breeding. For example, RAPD fragments (Appendix 6) were detected to be specific to: **(1)** an accession within a species (e.g. the 4000 bp fragment in AC230 produced by primer OPA02); **(2)** a species (e.g. the 2700 bp fragment produced by primer OPB08 that was specific to *P. ferrugineus*); or, **(3)** a species group or complex (e.g. the 3100 bp fragment amplified by primer OPA13, that was exclusive to accessions of the *P. tuberosus* complex and *P. ahipa*).

5.3.4. Species relationships

A first step in the phylogenetic analysis was to test whether or not the data set to be analysed had a hierarchical structure (Armstrong *et al.* 1994; Swofford *et al.*, 1996; Furman *et al.*, 1997). The PTP analysis was carried out on the RAPD data set to test for departures from randomness. The length of the original tree was 18.3, with a mean length of the random trees of 29.4 (PTP value = 118.48). A minimum tree length as small as that observed for the original RAPD data set was not observed among the randomly permuted data sets and trees created by RAPDISTANCE, which allowed rejection of the null hypothesis (i.e. the tree is merely an artefact of the algorithm) and was strong evidence that the data set had a hierarchical structure.

Phenograms were constructed using Neighbour-Joining (NJ) and UPGMA cluster analysis for all 85 accessions included in the survey. The matrix of genetic distances of pairwise comparisons between accessions, based on the proportion of shared fragments, is presented in Appendix 7. The resulting phenograms are described in turn below.

NJ phenogram

Intra- and interspecific relationships resolved by Neighbour-Joining cluster analysis are summarised in Figure 5.1. Phenetic analysis of RAPD data generated a consensus tree that separated *Pachyrhizus* accessions into seven main clusters. Five of these (clusters 3 to 7) were resolved at basal nodes and were composed exclusively of accessions of wild status (with the exception of TC556, TCNA09 and TCNA10 in cluster 3). The remaining two clusters contained all cultivated materials of *P. tuberosus* and *P. ahipa* plus accession TWTM48 (cluster 1); and, materials of wild and cultivated *P. erosus* plus accession TC350 (cluster 2). Within these two latter clusters accessions were split into subspecific groups, as analysed below.

Within cluster 1, accessions of *P. ahipa* were clearly separated into three subspecific groups (denoted as '*A*'; Figure 5.1). Of these, a first subspecific group of 14 accessions (*A1*) was resolved as an independent subcluster; a second group of four *P. ahipa* accessions (*A2*) resolved as a sister entity to six *P. tuberosus* accessions (mainly ashipas and jíquimas); and, a third subspecific group of nine *P. ahipa* accessions (*A3*) was rooted by accession TC309 (an ashipa). In general, bootstrap values in these groups were relatively high, which depicted a reliable and robust representation of the relationships described by the tree; however, the node that separated TC309 from the *P. ahipa* accessions of the third subspecific group was present in only 22% of the bootstrap subreplicates and, therefore, relationships and genetic similarities between these accessions should be treated with caution.

Similarly, accessions of *P. tuberosus* showed also an evident split into four subspecific groups (denoted as '*T*'; Figure 5.1) in cluster 1 with bootstrap values ranging from 10 to 97%. Relationships indicated by the NJ tree must be treated with caution when taking into account the bootstrap confidence intervals in this cluster.

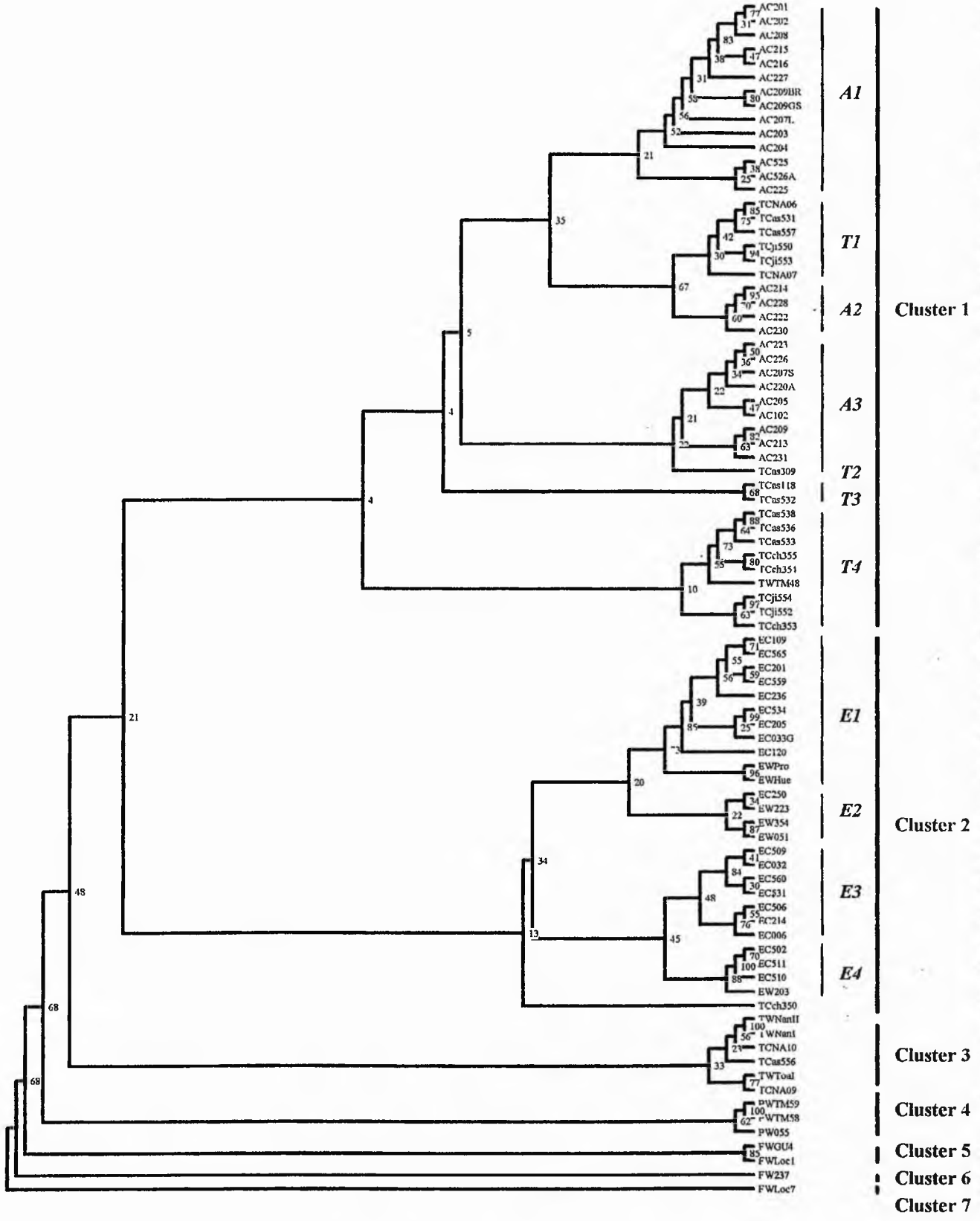


Figure 5.1. Unrooted NJ consensus tree constructed using Jaccard's algorithm showing relationships of the five species of *Pachyrhizus*. Numbers in nodes denote bootstrap values. Seven clusters and their subgroups are indicated: *A1* to *A3* = *P. ahipa*; *E1* to *E4* = *P. erosus*; *T1* to *T4* = *P. tuberosus* (Coding of accessions as for Table 2.1; *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima).

For example, the nodes that cluster accessions of two of the four *P. tuberosus* subspecific groups (i.e. *T4* and *T1*) are present in only 10 and 30 of the 100 bootstraps, respectively. These data suggest that the most basal of the nodes in the *P. tuberosus* groupings are not very reliable (a situation that also applies to cluster 1 as a whole); however, most of the terminal clusters are very robust, e.g. the two subclusters containing accessions TC550 and TC553 (both jíquimas); and, TC536 and TC538 (both ashipas), which occurred in 94% and 88% of the total bootstraps, respectively (see groups *T4* and *T1* in cluster 1).

With the sole exception of TC350, accessions of *P. ahipa* and *P. tuberosus* (in cluster 1) were clearly separated from *P. erosus* taxa which made up cluster 2. The node that separated these two clusters was present in only 21% of the bootstrap subreplicates and, therefore, genetic relationships should be treated with care; moreover, the node that separated TC350 from the remaining *P. erosus* accessions was not robust either (i.e. it was present in only 13 out of 100 subreplicates). *P. erosus* taxa in cluster 2 were clearly divided into a dichotomy, although only supported by a moderate bootstrap value of 34%. These two groups and their subgroups (marked as '*E*', Figure 5.1) are described in turn below.

The first group of *P. erosus* accessions was subdivided into two subgroups. The first of these subgroups (*E1*) comprised nine cultivated and two wild accessions of *P. erosus*; moderate to high bootstrap values ranging from 20% to 99% supported reasonably strong relationships among these accessions (cluster 2, Figure 5.1). A second, more basal, subgroup (*E2*) was composed mainly of EW accessions (from Costa Rica) with the exception of EC250 (Guatemala). The node that separated these two subgroups was present 20 times out of 100 sets.

The second *P. erosus* group in cluster 2 also contained two subgroups. The first subgroup (*E3*) comprised seven cultivated accessions of *P. erosus* and no wild materials. Again, moderate to high bootstrap values ranging from 30% to 84% supported reasonably strong relationships among these accessions (cluster 2, Figure

5.1). A second, basal subgroup (*E4*) was composed mainly of EC accessions, with the exception of EW203. Interestingly, all accessions in this second subgroup were of Mexican provenance. The node that separated these two subgroups of *P. erosus* in cluster 2 was present in 45% of the bootstrap samples, supporting a reasonably strong relationship between these accessions.

Cluster 3 comprised exclusively *P. tuberosus* taxa basally positioned to accessions in clusters 1 and 2. It contained three wild accessions (TWNanI, TWNanII and TWToaI), two of undetermined status (i.e. escaped/wild, namely TCNA09 and TCNA10) and one ashipa (TC556), supported by bootstrap values spanning from 23 to 100%. The node that separated this cluster from clusters 1 and 2 occurred in 48% of the total bootstraps, which reasonably supports the split of accessions in cluster 3 as a separate phylogenetic entity.

Cluster 4 included the three accessions of *P. panamensis* examined in this study, clustering tightly as a monophyletic entity. PWTM58 and PWTM59 (both from Ecuador) were strongly resolved together with a bootstrap value of 100%, whereas PW055 (Panama) was basal to this close affinity between the Ecuadorian accessions (62% bootstrap value). Finally, four *P. ferrugineus* accessions placed in clusters 5, 6 and 7 were resolved as the most primitive taxa within the genus. The node that separated these accessions from the remaining species was present in 68% of the bootstrap subsamples, providing strong support for these genetic relationships.

UPGMA phenogram

The phenogram generated by UPGMA cluster analysis of 85 *Pachyrhizus* accessions showed them to be grouped into six main clusters (Figure 5.2). Cluster 1 comprised a group of 39 accessions representing all *Pachyrhizus* species except *P. panamensis*. Within this cluster, a total of 24 accessions of cultivated *P. tuberosus* and *P. ahipa* were resolved as sister groups in one subcluster, whereas 11 *P. erosus* accessions were positioned in a second, more basal, subcluster. Two accessions of FW and EW each were also included in this first cluster and were resolved as its basal-most taxa. As in

the NJ generated phenogram, there is an evident split of *P. ahipa* and *P. tuberosus* accessions into subspecific groups within cluster 1.

Cluster 2 comprised a total of 35 accessions, most of them of cultivated status (namely AC, EC and TC taxa), although three accessions of wild status (EW203, EW223 and TWTM48) were also included. Again, 19 accessions of *P. ahipa* and *P. tuberosus* were resolved as sister taxa with *P. erosus* as a separate subcluster. Additionally, one accession of *P. tuberosus* (TC350) was also included in this *P. erosus* subcluster and was placed at a basal position. Two ashipas (TC118 and TC532) were resolved as the most basal taxa in this second cluster.

Cluster 3 comprised a total of six accessions of *P. tuberosus* (Figure 5.2); of these, three accessions were of wild status, one (TC556) was an ashipa and the remaining two are regarded as of undetermined status (i.e. wild origin or escapes).

Cluster 4 was made up of all three accessions of *P. panamensis* examined in this study, with PW055 most basal. Finally, clusters 5 and 6 comprised the remaining two accessions of *P. ferrugineus* examined and were resolved as the most primitive taxa in the genus (Figure 5.2).

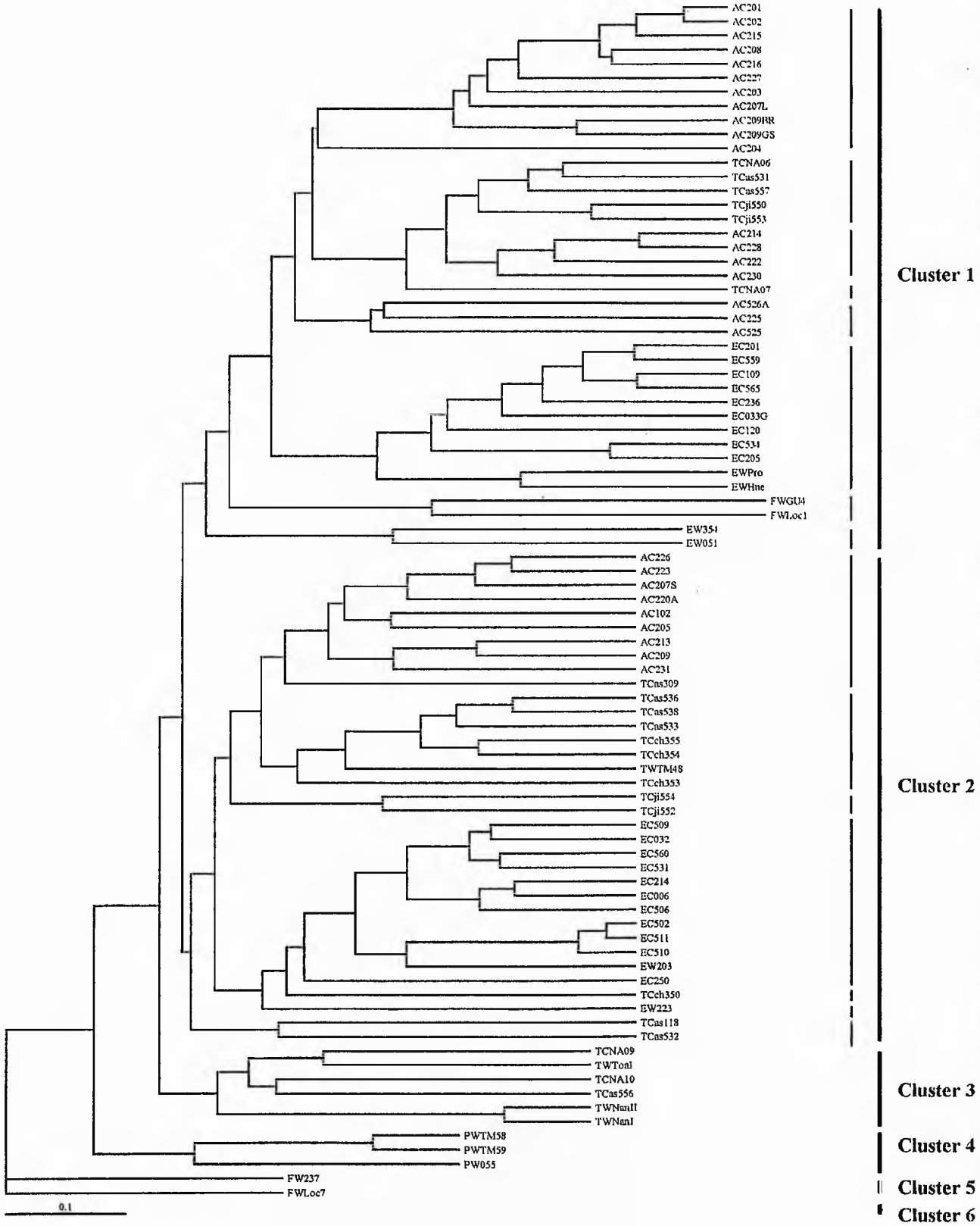


Figure 5.2. UPGMA cluster analysis of 1 - F values (based on Jaccard's coefficient) computed from pairwise comparisons of RAPD polymorphisms between 85 accessions of *Pachyrhizus*. Six clusters and their subgroups are indicated (coding of accessions as for Table 2.1; *P. tuberosus* cultigen types: as = ashipa; ch = chuin; ji = jíquima).

5.4. Discussion

5.4.1. RAPDs and molecular markers

This study has demonstrated that RAPDs are extremely useful markers for reliable identification of *Pachyrhizus* taxa. Interestingly, more RAPD bands were amplified with genomic DNA of cultivated materials (i.e. AC, EC and TC) than with DNA of wild taxa (i.e. EW, FW, PW and TW). In addition to resolving several accession-specific fragments among the five species of the genus, the study showed that unique fragments/species could be produced for *P. ahipa* and *P. ferrugineus* accessions when using particular primers (Tables 5.1 and 5.2). Overall, these differences in RAPD amplification might be a direct consequence of factors such as the genomic complexity of the plant species, the composition and size of the primers used, the number of accessions screened per species, and the experimental resolution and reproducibility. Reproducibility was most critical in ultimate inclusion for analysis.

Additional intraspecific sampling would be desirable to confirm the specificity of banding patterns, since considerable RAPD variation may be found within and between taxa (Chalmers *et al.*, 1992). Several studies have shown that strains and cultivars of particular species can be individually typed according to their RAPD profile (Demeke *et al.*, 1992; Virk *et al.*, 1995), and so it is likely that the same will be found true within *Pachyrhizus* species when larger numbers of accessions and/or primers are employed in future analyses. This would greatly aid taxon identification (even among relatively uncharacterised accessions), and would be of considerable importance in the management of germplasm for the breeding of *Pachyrhizus*.

5.4.2. Homology assessment

Most investigators assume comigrating bands as evidence of homology without actually testing for it. Moreover, only few published studies have tested this assumption in detail before constructing phenograms (e.g. Furman *et al.*, 1997). However, tests of homology should be routinely conducted in any RAPD study to

confirm the assumption and validate the derived genetic relationships between taxa (Smith *et al.*, 1996). In the present study, isolation of comigrating bands and their subsequent digestion with restriction enzymes proved to be a successful method for testing homology. RAPD fragments representing all species within the genus were chosen at random and, of these, 100% produced visible patterns after digestion with at least two of the three restriction enzymes used.

The level of homology between comigrating fragments examined within the 85 *Pachyrhizus* accessions was calculated to be 77.36%, validating the affinities resolved between them; however, the remaining 22.64% proved to correspond to non-homologous DNA fragments (i.e. of different sequence and origin) and, therefore, incorporated a certain degree of homoplasy in the data set. Intra- and interspecific pairwise comparisons showed homology levels of 74.2% and 77.5% among comigrating bands, respectively. Not surprisingly, the level of homology was higher when intraspecific comparisons were considered (60.9% of the homologous pairwise comparisons), in contrast to its low level when interspecific pairwise comparisons (39.1%) were examined. Overall, this suggested that the results should be treated with some caution, despite the fact that the phylogeny inferred from RAPDs was in general congruent with those produced in previous phylogenetic treatments (see earlier chapters).

5.4.3. Species relationships

This study showed that considerable RAPD variation was present within and between the *Pachyrhizus* taxa studied allowing an examination of genetic diversity and reconstruction of phylogenetic relationships. NJ and UPGMA analyses of a similarity matrix constructed from presence/absence of RAPD fragments produced phenograms that separated *Pachyrhizus* accessions into seven and six well resolved clusters, respectively. Overall, both phenograms were similar in structure and contained similar clusters of taxa; however, some important differences were evident between the NJ and UPGMA phenograms in the detailed topology of clusters and groups within (see further).

Several workers using RAPDs for phylogeny inference score presence/absence of bands, calculate pairwise genetic distances between taxa and use UPGMA for phenogram reconstruction (e.g. Graham & McNicol, 1995). However, the UPGMA method will only yield an accurate phylogeny when rates of evolutionary divergence are homogeneous (Swofford & Olsen, 1990); violation of this assumption may introduce errors. Given the arbitrary nature of RAPD fragments and the differing rates of evolutionary change in different regions of the genome (Soltis *et al.*, 1992; and references therein), some workers consider a rate-independent method of cluster analysis, such as Neighbour Joining, Distance Wagner or Fitch-Margoliash to be more appropriate (Avise, 1994). This study used both NJ and UPGMA approaches to examine intra- and interspecific relationships which appear, to a degree, to have generated similar findings in this regard. However, the main difference between them, i.e. UPGMA resolved accessions of all species into subclusters similar to those produced by NJ analysis, but placed in widely separated regions of the phenogram (except for *P. panamensis*, which remained basal in both methods), suggested that the NJ approach was more suitable in this instance. Therefore, in this discussion, species relationships will be examined with reference to the NJ generated tree.

The NJ consensus tree clearly showed that *P. ferrugineus* and *P. panamensis* are the most primitive species in the evolutionary history of yam beans, with *P. ferrugineus* (clusters 5 - 7; Figure 5.1) being ancestral to *P. panamensis* (cluster 4). Next, three lineages of *P. tuberosus* were evident. A first, basal, lineage of *P. tuberosus* comprised Ecuadorian accessions only from tropical and subtropical localities (cluster 3); it would appear that this lineage is the result of an early radiation. Two additional lineages were observed, one of which appears basal within the *P. erosus* cluster (TC350, a chuin; cluster 2), while the last lineage gave rise to the *P. tuberosus* complex and *P. ahipa* (cluster 1). Thus, RAPD data appear to indicate that *P. tuberosus* gave rise to different taxa and subgroups of taxa in different parts of its range; and, in the case of the cultigen types (i.e. ashipa, chuin and jíquima) and the highly derived species *P. ahipa*, this may have been aided by man through domestication.

Within *P. erosus*, there was an evident split of accessions into groups and subgroups (*E1* to *E4* in cluster 2; Figure 5.1), which were correlated with the geographic distribution of the accessions examined in this survey and the likely route of dispersal of this species after an evolutionary split from an early ancestor. For example, a first group (*E2*) is comprised of wild and cultivated *P. erosus* accessions from Central American origin only; two additional groups (*E1* and *E3*) were composed of *P. erosus* accessions (wild and cultivated status) from Central America and Mexico; and, a fourth group (*E4*) was composed exclusively of accessions of Mexican provenance. Thus, RAPD polymorphisms would suggest a split of *P. erosus* into: **(1)** a Mesoamerican pool (viz. a result of a first, early radiation); **(2)** a composite pool (viz. dispersal of populations to northern areas of Central America and southern Mexico); and, **(3)**, a Mexican pool (with further specialisation of primitive landraces and cultivars, including domestication, cultivation and man-made selection). This last pool reflects the close association within the cultivars/landraces from central Mexico and the Yucatan peninsula (also observed in a previous RAPD study; Estrella *et al.*, 1998), which might confirm the hypothesis of Sørensen (1996), in that they are derived from a restricted ancestral stock due probably to a limited germplasm introduction from southern Mexico, Guatemala and regions further south in Central America.

The history of the introduction of *P. erosus* to the countries of Southeast Asia and southern China is somewhat unclear, but there is little doubt that it must have been first introduced from the Philippines to the coastal regions of China and Vietnam; from Vietnam the crop was subsequently introduced to neighbouring countries (Sørensen, 1988 & 1996). In this study, Far Eastern cultivars of *P. erosus* (accessions EC109 and EC565) resolved together with Mexican materials at terminal nodes of cluster 2, supported by relatively high bootstrap values (Figures 5.1 and 5.2). This close linkage would indicate that these Asiatic cultivars are probably derived from introductions from the Mexican region, via old Spanish colonies.

An intriguing feature of the RAPD analysis concerns the resolution of accessions of the *P. tuberosus* complex. Polymorphisms produced in this study proved to be useful markers to resolve accessions of the same cultigen type in strongly supported terminal clusters, with bootstrap values ranging from 64% to 97% (see cluster 1, Figure 5.1). Nevertheless, at the most basal nodes the relationships they describe were ambiguous and were supported by relatively low bootstrap values; rather surprisingly, accessions belonging to the ashipa, chuin and jíquima cultigen types were distributed over four subgroups (*T1* - *T4*) positioned in different regions of cluster 1. This seems to suggest that after the ancestors of these cultigen types split from a more primitive phylogenetic entity within *P. tuberosus*, they were isolated by geographical barriers (and subsequently by cultivation and man-made selection), underwent rapid radiation in different tropical and subtropical regions of South America, and became progressively more specialised (e.g. changes in stem and leaf morphology, tolerance to different rainfall rates, humidity, competence with rapid weed growth). However, insufficient time has passed to accumulate enough mutations at RAPD loci to satisfactorily differentiate these cultivars.

Another interesting finding concerned the positioning of the wild accession of *P. tuberosus* TWTM48 from Ecuador (see *T4*; Figure 5.1). This accession was resolved basally to five cultivars of the ashipa and chuin types, supported by a moderate bootstrap value of 55%. Indeed, the close association of TWTM48 with the three cultigen types at basal levels of cluster 1 (instead of its placement with other TW taxa in cluster 3) would confirm the existence of different RAPD lineages within *P. tuberosus*. A reasonable hypothesis to explain this affinity between TWTM48 and the cultigen types is the present 'underdeveloped' status of *P. tuberosus* in cultivation. This species has traditionally been grown in shifting cultivation with minimal agricultural practices by indigenous people of the Amazonian region (Sørensen *et al.*, 1997). The present 'underdeveloped' agronomic status of the many local cultivars in the Amazon and the few introductions to areas outside its original distribution area (Sørensen, 1988 & 1990), would indicate that materials of *P. tuberosus* were selected according to elementary selection criteria, e.g. shade, variable soil features, high humidity, etc., to which the majority of the *P. tuberosus* complex (namely the ashipa

and chuin groups) are perfectly well suited. Additionally, Sørensen *et al.* (1997) have also pinpointed that, when the first plants were collected for domestication, the primary selection criteria that led to the development of the early original cultivars presumably included taste, yield, ease of cultivation and probably also tuber shape and their number per plant. Therefore, many cultivars resulted almost indistinguishable from known wild forms of this species.

In view of this and the current RAPD evidence, clustering of accession TWTM48 together with accessions belonging to the ashipa and chuin types (Figures 5.1 and 5.2) would be well-grounded. Furthermore, this wild accession would represent a relict between the more advanced, highly enhanced *P. tuberosus* cultivars (in cluster 1) with the more basal and primitive materials (in cluster 3), the latter ones representing an early ancestor which after rapid radiation generated the bulk of diversity now evident within the *P. tuberosus* complex. For comparative purposes, it should be noted that the ITS phylogeny did not resolve clearly the positioning of TWTM48, but placed this accession as a related entity to other wild species (Figures 4.3 and 4.4; chapter 4), whereas the cpDNA phylogeny placed it closely within the *P. tuberosus* complex (e.g. Figures 3.11, 3.13 and 3.14; chapter 3). Thus, the RAPD phylogenetic signal appears to have clarified relationships among these taxa and proved to be useful in providing additional information on the affinity between closely related species.

Placement of accession TC350 (a chuin) at the basal most node of cluster 2, which contained all *P. erosus* taxa, was also a surprising finding (Figure 5.1). The NJ consensus tree appears to suggest that *P. tuberosus* was not only present in the South American evolutionary branch, but was probably involved in the early parentage of *P. erosus*. If this is correct, then the separation of these two species must have occurred a long time ago, given the number of new mutations at RAPD loci which now separate them. Thus, RAPD evidence is showing a congruent pattern to that obtained in the phylogeny based on PCR-amplified cpDNA and the combined cpDNA data set (chapter 3), but also an inconsistency with total cpDNA and nuclear ITS sequence variation (chapter 4). This raises some concern on the usefulness of RAPD generated patterns for inferring species relationships.

In this regard, a potential problem in using the presence/absence of RAPD fragments as a basis for examining species relationships arises due to the occurrence of competition among RAPD primer sites. Primer annealing sites are scattered throughout the nuclear and cytoplasmic genomes in all classes of DNA from single-copy to multiple-copy DNA, and in coding and non-coding regions (Caetano-Anollés, 1993; Williams *et al.*, 1993). Sometimes the absence of a band may simply be due to the fact that there is competition between a number of the same priming sites for a given primer. Consequently, amplification does not occur though the appropriate DNA sequence for primer annealment is present, which in turn might have significant effects on the final phenetic analysis and its interpretation.

Phenetic analysis of RAPD data in *Pachyrhizus* demonstrated a close association between *P. ahipa* and *P. tuberosus* (cluster 1, Figure 5.1). Thus, the relationships derived using this approach were, in general, congruent with those from the cladistic analyses of cpDNA and ITS variation reported in previous chapters, and also with earlier studies indicating affinities between these two species (Døygaard & Sørensen, 1998; Ørting *et al.*, 1996; Sørensen *et al.*, 1997). A considerable amount of intraspecific RAPD variation was evident throughout cluster 1 and resulted in the grouping of accessions of *P. ahipa* and *P. tuberosus* into distinct subspecific groups, which tended to correlate with geographic distribution and morphology (e.g. cultigen types). For example, *P. ahipa* accessions were grouped into three distinct subclusters (*A1* - *A3*) and *P. tuberosus* accessions into four (*T1* - *T4*), suggestive of different lineages in the evolutionary pathways of these two species.

Although of high value in this respect, the exact affinity between *P. ahipa* and *P. tuberosus* suggested by the RAPD trees is somewhat difficult to explain. For example, the phylogenetic signal explaining the origin of *P. ahipa* is not conclusive in cluster 1 since the three subgroups were resolved either as a highly derived, independent lineage (i.e. *A1*), as a sister taxon to the *P. tuberosus* complex (i.e. *A2*) or as a phylogenetic entity closely within *P. tuberosus* (i.e. *A3*; Figure 5.1). Conversely, taking cluster 1 as a whole, current RAPD evidence would suggest that *P. ahipa* is

possibly derived from *P. tuberosus* with the former species resolving closely within the latter one; nevertheless this affinity should be treated with some caution, since bootstrap values supporting species relationships at basal nodes of cluster 1 were low.

CHAPTER 6

GENERAL DISCUSSION

Rapid radiations
J. Estrella

General Discussion

The research reported in this thesis demonstrates how molecular approaches can be used to elucidate species relationships and to detect differing levels of genetic diversity within *Pachyrhizus*, a tuberous legume of great potential and attractiveness for tropical and subtropical agriculture. No extensive research has been carried out previously on this crop using molecular techniques. This study therefore aimed to improve our understanding of the evolutionary history, systematics and genetic variation of the genus *Pachyrhizus* in three major ways: **(1)** by constructing a molecular phylogeny of the genus based on several assay methods that explore different components of the chloroplast and nuclear DNA genomes; **(2)** by determining levels of genetic diversity within *Pachyrhizus* species; and, **(3)** by identifying taxon-specific molecular markers and the nature of their evolution for use in germplasm identification and in marker-assisted breeding programmes. These objectives have been accomplished to a large extent as reported in chapters 3 to 5 of this thesis.

The analysis of both cpDNA and ITS variation (chapters 3 and 4) resulted in two independent data sets with 49 and 600 molecular characters, respectively, which complemented each other in establishing evolutionary relationships within *Pachyrhizus*. A third data set obtained from an analysis of 148 RAPD fragments provided additional information in regard to past evolutionary events and genomic affinity of the five species, and - most importantly - contributed substantially to an understanding of genetic relationships within the cultivated species (i.e. *P. ahipa*, *P. erosus* and the *P. tuberosus* complex) and their morphologically/geographically distinct landraces and primitive cultivar groups.

6.1. Molecular markers and DNA sequences in *Pachyrhizus*

The research identified a number of molecular markers and DNA sequences that were specific to particular *Pachyrhizus* taxa. These markers may be of use in future studies of germplasm characterisation, plant systematics, and in breeding activities, but

further intraspecific sampling would be desirable to confirm their specificity. Restriction fragment analysis of the total cpDNA genome and of PCR-amplified cpDNA regions (chapter 3) yielded a total of 45 markers that were specific to an accession within a species, a particular species, or even a species group or complex. A total of 820 restriction fragments were generated by both cpDNA surveys after FOA (fragment occurrence analysis). Of these, 84 (10%) were uniform across all taxa examined, 579 (71%) were synapomorphic, and 157 (19%) were unique to a particular accession (mainly those of *P. ferrugineus* and the operational outgroups). Overall, these genetic markers enabled formulation of explicit hypotheses concerning the evolution of *Pachyrhizus* species.

With respect to ITS variation, the sizes and divergence levels of the ITS regions in *Pachyrhizus* (chapter 4) were similar to those reported for other angiosperms (Baldwin *et al.*, 1995; Hamby & Zimmer, 1992). Nucleotide substitution was probably the main source of sequence divergence in the ITS, which might have evolved primarily by point mutations. This conclusion is reached from the reduced proportion of gaps that were required for alignment, the absence of evident length variants within the DNA accessions examined, and the moderately high levels of sequence divergence between and even within species. These indications of ITS length conservation, but nucleotide sequence variability, allowed DNA sequences of both spacers and the 5.8S subunit to be readily aligned across *Pachyrhizus* taxa for a comparison of nrDNA variation to be used in genealogical reconstruction at the species level.

In regard to RAPD analysis, more fragments were amplified with genomic DNA from cultivated accessions than with genomic DNA from wild accessions (chapter 5). In addition, unique fragments per species were only detected for *P. ferrugineus* and *P. ahipa*, i.e. both evolutionary 'extremes' observed in the phylogeny of the genus, which might be a direct effect of their genomic composition. However, RAPD analysis generated numerous polymorphisms that were useful in revealing the genetic structure, affinities and phylogeography within the five species and the subspecific groups detected.

In summary, molecular markers identified in this study have been useful in examining characteristic phylogenetic signatures within the genomes of *Pachyrhizus* species, and for estimating genetic similarities and differences within and between species which might have gone unnoticed in non-molecular appraisals.

6.2. Evolutionary relationships within the genus *Pachyrhizus*

The phylogeny based on cpDNA variation (chapter 3) revealed a consistent phylogeographical pattern of species relationships within *Pachyrhizus* and subdivided it into two evolutionary branches, i.e. a Mesoamerican and a South American branch. A first cladistic analysis of RFLP variation across the total cpDNA genome (Figure 3.7) indicated that *Pachyrhizus* is a monophyletic genus, i.e. derived from a single ancestral taxon, and was split into two distinct clades. *P. ferrugineus* was resolved in one of the clades as the most primitive species within the genus (although an additional lineage of this species was present in a different, more terminal clade), whereas accessions of wild and cultivated *P. erosus* made up the remaining of this clade. These two groups represent the Mesoamerican branch in the evolutionary history of the crop. With the exception of accessions FW237 and PW055, a second clade comprised materials of Andean and Amazonian distribution, namely *P. panamensis*, the *P. tuberosus* complex and the highly derived species *P. ahipa*, representing the South American evolutionary branch of the genus. Among these three species, *P. panamensis* was the most primitive taxon in the group.

A second cladistic analysis based on RFLP variation of PCR-amplified cpDNA regions (Figure 3.11) produced a phylogenetic tree which depicted an overall topology for the five species that was similar to that based on total cpDNA. However, three main contrasts were evident: **(1)** a lower level of resolution among accessions of the *P. tuberosus* complex was apparent, due to the presence of fewer synapomorphic characters; **(2)** *P. ahipa* was placed as an independent genealogical entity (instead of being resolved as a sister group to *P. tuberosus*); and, **(3)** an additional lineage of *P. tuberosus* (namely accession TC350) was present, basal to the *P. erosus* clade, and moderately supported by bootstrap values.

The combination and subsequent cladistic analysis of the Southern-RFLP and PCR-RFLP data matrices resulted in a highly resolved strict consensus tree (Figure 3.14; chapter 3). Again, two evolutionary branches were evident, with *P. ferrugineus* as the most ancestral species. Wild and cultivated *P. erosus* made up the Mesoamerican branch (but see further), while *P. panamensis*, *P. tuberosus* and *P. ahipa* defined the South American branch. Furthermore, several intriguing findings became evident: **(1)** both accessions of *P. ferrugineus* resolved at the basal-most level, although still appeared as separate lineages; **(2)** *P. tuberosus* (TC350, a chuin) was included in the Mesoamerican evolutionary group and was basal to all *P. erosus* accessions; and, **(3)** *P. ahipa* was resolved as a separate clade with *P. tuberosus* (namely TC553, a jíquima) as a sister taxon.

In the light of this cpDNA evidence, *P. tuberosus*, as a phylogenetic entity, appears to have played a significant role in the evolution of the genus. The consensus trees clearly suggest that: **(1)** *P. tuberosus*, although mainly present in the South American evolutionary branch, might also be ancestral to *P. erosus* as a separate lineage. If this is so, these two lineages of *P. tuberosus* must have diverged from each other long ago, given the number of cpDNA mutations that now separates them (Figures 3.11 and 3.14); **(2)** an early, wild ancestor of *P. tuberosus* was also closely related to other wild species, i.e. *P. panamensis* and *P. ferrugineus* (Figure 3.7); and, **(3)** some of the cultigen types of the *P. tuberosus* complex encompass the early ancestry of the highly derived species *P. ahipa* (e.g. accession TC553, a jíquima; Figures 3.7 and 3.14).

The phylogenetic consensus tree generated from ITS sequence variation (chapter 4) was not as well resolved as those constructed from cpDNA restriction site characters. Polytomies observed in the basal nodes of one of the two ITS clades (Figure 4.4) represented regions where clustering differences took place, due probably to low sequence divergence values among DNA accessions. The two evolutionary branches (*sensu* cpDNA) were not as clearly evident; nonetheless, the ITS phylogeny complemented the cpDNA phylogeny to a broad extent, especially in separating *Pachyrhizus* species into clades reflecting a congruent phylogeographical distribution.

The ITS trees (Figures 4.3 and 4.4) showed an interesting positioning of *P. panamensis* relative to wild taxa of *P. erosus* and *P. tuberosus*. A likely phylogenetic hypothesis might suggest that after an early ancestor of these wild taxa split from the most primitive phylogenetic entity within the genus (i.e. *P. ferrugineus*), it became isolated and radiated into neighbouring Mesoamerican ecosystems adjacent to the centre(s) of origin. In addition, these ancestral groups diverging early in the Neotropics may have remained diminished, while those splitting off later radiated rapidly and generated much of the diversity that is evident today. Speciation may have accompanied divergent adaptation to dissimilar ecological niches in two directions: to areas with conspicuous annual dry seasons and deciduous forests in central and northern areas of Mesoamerica (which in turn resulted in the origin of *P. erosus* and the specialisation of its several primitive landraces and cultivars, suggestive of a secondary centre of origin); and, to the tropical and subtropical rainforests and Andean valleys (giving rise to the different forms within *P. tuberosus* and subsequently to *P. ahipa*). In contrast to cpDNA data, the current ITS evidence points to an ashipa type within *P. tuberosus* (instead of a jíquima) as the sister taxon of the highly advanced species *P. ahipa*.

The close association observed in both cpDNA and ITS surveys of the wild accession TWTM48 of *P. tuberosus* with other wild taxa (i.e. PW and EW) at basal levels of a particular clade, was unexpected as it was considered to be more closely related to the *P. tuberosus* complex, placed at more terminal levels of the phylogeny of the genus (Figures 3.7 and 4.4). This apparent 'discordance' could stem from an alternative hypothesis of horizontal gene transfer (Smith *et al.*, 1992; Avise, 1994) within these wild taxa. It is feasible that TW, PW and EW taxa acquired genetic similarity during a recent phylogenetic stage via horizontal genetic transfer. Several reports of such lateral transfer of genetic elements across taxonomic boundaries have appeared in the literature in recent years (e.g. Calvi *et al.*, 1991; Flavell, 1992), indicating that it may be a frequent phenomenon in some genera. It should be emphasised, however, that several factors other than horizontal transfer might also lead to these apparent discordances. These include the shared retention of ancestral states by the taxa in

question, extreme molecular rate heterogeneities across lineages, convergent evolution to a shared molecular condition, and a mistaken assumption of orthology when the loci in question might truly be paralogous (Avise, 1994; and references therein). In the present study, the possibility of the horizontal transfer of particular genes added another important rationale to the inclusion of multiple lines of evidence in the phylogenetic reconstruction of *Pachyrhizus*. Further analysis, involving a larger number of wild taxa, would be worthwhile to investigate these instances in greater detail.

The phylogeny of *Pachyrhizus* inferred from a phenetic analysis of RAPD variation (chapter 5) was congruent to an extent with those obtained from analyses of cpDNA and ITS variation. NJ analysis of the RAPD data set partitioned *Pachyrhizus* taxa into seven well resolved clusters (Figure 5.1) and appears to confirm several aspects of the history of the genus indicated by the cpDNA and ITS trees. In the RAPD tree, *P. ferrugineus* was resolved as the most primitive species within the genus, with different accessions occurring in separated, but always basal nodes of the tree. Such variation between accessions of this species might reflect the wide geographic distribution of an early ancestor which originated during periods when populations were relatively isolated from each other allowing molecular differences to accumulate more easily. Two main evolutionary branches were also present in the NJ tree, although not in the strict sense as depicted by cpDNA or ITS data. The main differences concerned the resolution of *P. panamensis* as an independent cluster of early origin (a radiation after *P. ferrugineus*) and the presence of three *P. tuberosus* lineages in widely separated regions of the phylogeny (see further). Finally, the RAPD tree confirmed the affinity between *P. ahipa* and *P. tuberosus*, with the former species resolving closely within the latter one; however, several subspecific groups with intertwined topology were evident in each of this species and, thus, relationships should be treated with some caution.

In summary, the RAPD NJ tree clearly shows that *P. ferrugineus* and *P. panamensis* taxa are most primitive, with *P. ferrugineus* being ancestral to *P. panamensis*; in turn, this latter species is ancestral to a first lineage of *P. tuberosus* taxa (i.e. cluster 3;

Figure 5.1). Next, *P. tuberosus* contains two additional lineages, one of which appears basal within the *P. erosus* cluster (TC350, a chuin; cluster 2), while the remaining lineage is basal to the *P. tuberosus* complex and *P. ahipa* (cluster 1). Thus, RAPD data appear to indicate that *P. tuberosus* gave rise to different taxa and subgroups of taxa in different parts of its range, and, in the case of *P. ahipa,* this may have been aided by man through domestication.

RAPDs were also useful in partitioning *P. erosus* taxa into three genetic pools (the Mesoamerican, 'composite' and Mexican groups), which correlated with a likely direction of dispersal of the species to northernmost geographical localities, assuming a 'single' origin in strictly southern regions of Mesoamerica. However, considering the considerable range of geographical, morphological and genetic variation in wild and cultivated *P. erosus* that is evident at the present time, the proposal of multiple origins and domestication processes in this species attracts special attention. For example, wild populations are associated with climatically different areas, i.e. wet forests in Veracruz (Mexico), deciduous forests in Baja Verapaz (Guatemala), and dry savannahs in Huehuetenango (Guatemala) and Guanacaste (Costa Rica). In this regard, the three different lineages of wild *P. erosus* supported by RAPD evidence (i.e. EW accessions in subgroups *E1, E2* and *E4*; Figure 5.1), together with their considerable ecological, geographical and morphological variation, strongly support a hypothesis of multiple origins and domestication processes. Further analysis including a greater amount of wild germplasm of *P. erosus* would be of value (e.g. glabrous and strigose populations from Guatemala and Costa Rica vs. the northernmost populations from Veracruz, Mexico with a different morphological appearance; Sørensen, pers. comm.).

A combined analysis of molecular variation in *Pachyrhizus*

As a means of obtaining what might be the best estimate of a phylogeny of *Pachyrhizus* from the available information, the cpDNA, ITS and RAPD data were combined in different ways and subjected to phylogenetic analysis to determine whether the resolution of species relationships might be confirmed and/or improved.

Combined analysis of data sets was justified because the higher-order structure of the trees produced by each independent phylogenetic treatment was similar. In a first cladistic analysis of combined data, characters of restriction site variation across total cpDNA and ITS sequence variation were merged (i.e. a 630-character data matrix), while in a second cladistic analysis RAPD characters were also incorporated into this combined data set (i.e. a 778-character data matrix). Unfortunately, addition of cpDNA characters generated by the PCR-RFLP approach was not feasible since it hindered the incorporation of representative accessions of *P. erosus* and an operational outgroup.

Phylogenetic trees were generated from unordered character states (Fitch parsimony) using PAUP; the ACCTRAN option was chosen as a method for optimising unordered characters. Invariant sites and autapomorphous characters were ignored in the phylogenetic reconstruction (IGNORE UNINFORMATIVE CHARACTERS option), following the recommendation of Francisco-Ortega *et al.* (1997). Character state changes were weighted equally and gaps were treated as missing data. Heuristic searches were conducted using the options BRANCH SWAPPING, TBR and MULPARS; construction of strict consensus trees, bootstrap estimates and calculation of decay indices follow the same procedures from previous cladistic analyses.

CpDNA restriction site and ITS sequence variation data when combined were available for 13 *Pachyrhizus* taxa and one outgroup. Phylogenetic analysis of this first combined data set generated a single equally parsimonious tree of 269 steps (Figure 6.1) with CI and RI values of 0.677 and 0.62, respectively. Overall, bootstrap values for most nodes were high, with an average of 68%. Additionally, the high decay indices ≤ 283 and ≤ 294 for *P. ahipa* and *P. ferrugineus*, respectively, indicated that both 'extremes' in the phylogeny of the genus (i.e. the most primitive and the most derived phylogenetic entities) were strongly supported (Figures 6.1 and 6.2).

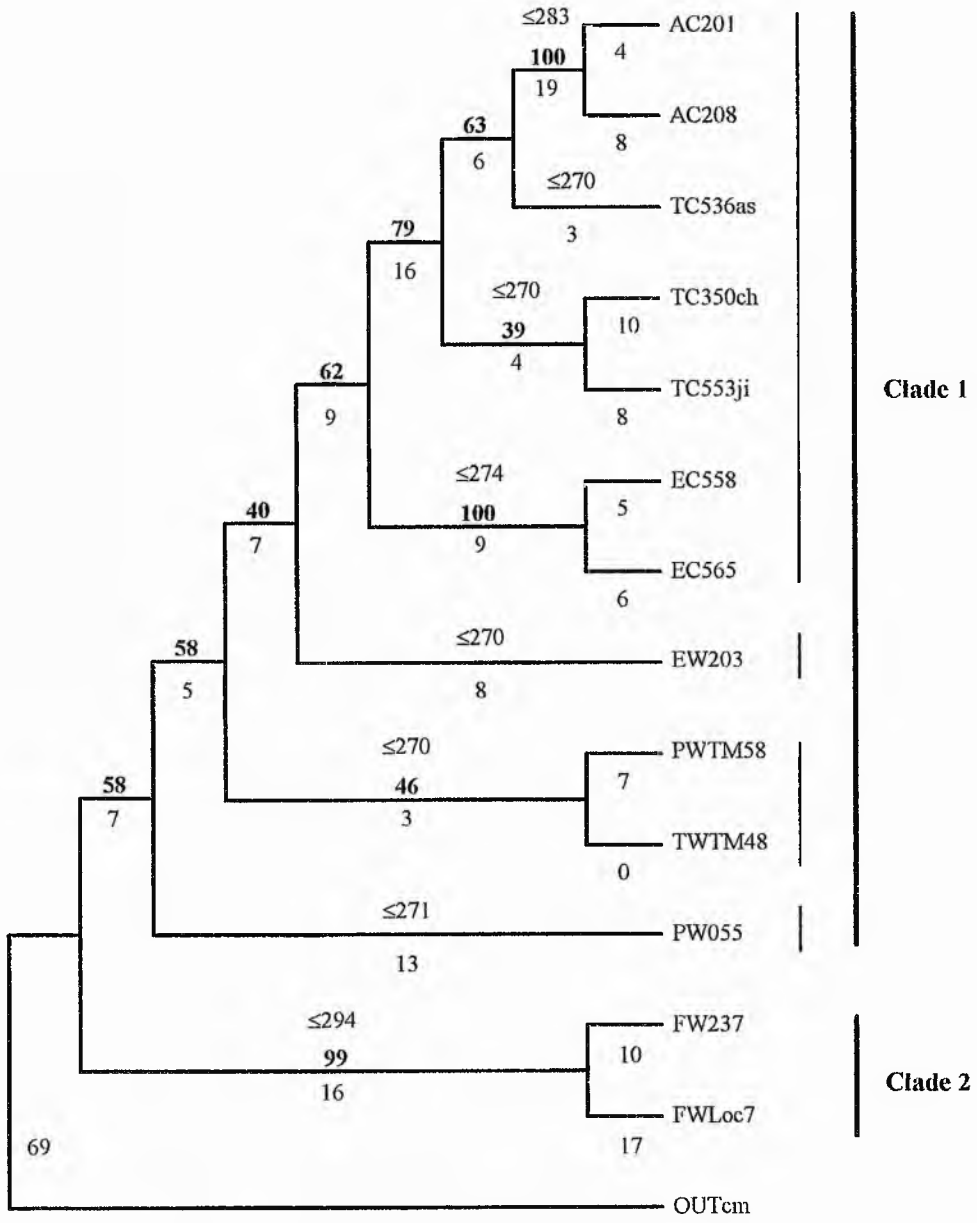
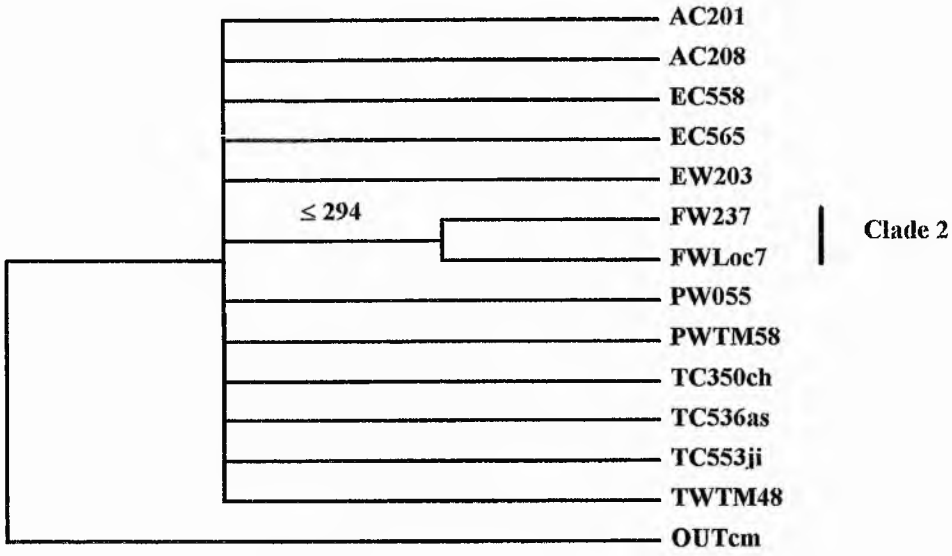


Figure 6.1. Single equally parsimonious Fitch tree derived from a combined analysis of cpDNA restriction site and ITS sequence variation in 13 *Pachyrhizus* taxa and one outgroup. Numbers above branches indicate bootstrap percentages (in boldface) and decay values; numbers below branches are synapomorphies (Tree length = 269 steps; CI = 0.677; HI = 0.323; RI = 0.620; see text for further details). Key for accessions as for Table 2.1; as = ashipa; ch = chuin; ji = jíquima cultigen type.

(A)



(B)

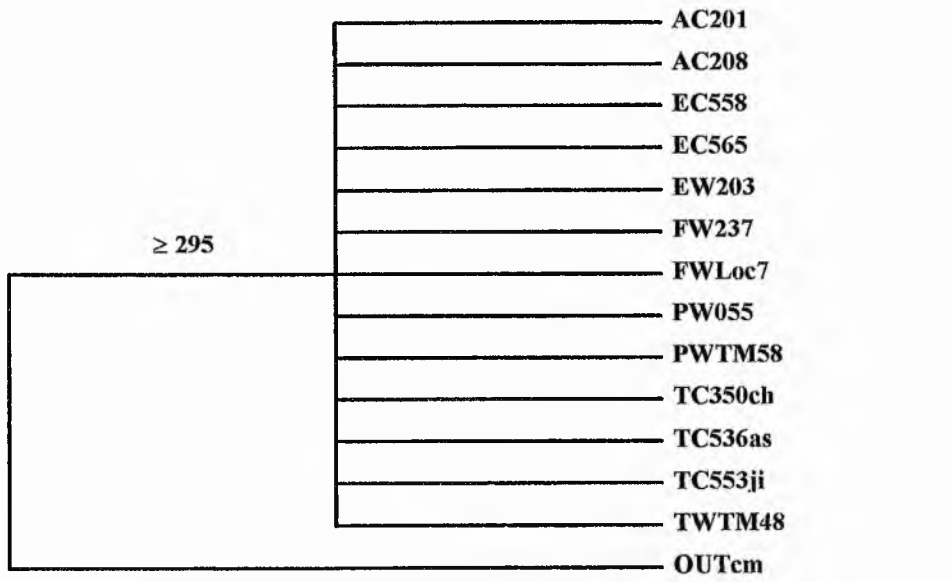


Figure 6.2. Trees showing the order of 'decay' of clades obtained after a combined analysis of cpDNA restriction site and ITS variation. **(A)** This tree is the strict consensus of trees whose length is 294 steps; note that all branches of clade 1 have collapsed. Clade 2, containing the two *P. ferrugineus* accessions examined, was still present at 294 steps. **(B)** Strict consensus of trees whose length is equal to 295 steps and over; the tree became an unresolved 'bush' at 26 steps away from the minimal tree (see text for further details).

The cpDNA-ITS combined tree is highly congruent with the ITS phylogeny, showing a very similar evolutionary history for the genus *Pachyrhizus*, but resolving satisfactorily the polytomy observed in the ITS tree for EW, PW and TW taxa (Figure 4.4; chapter 4). This polytomy was resolved such that a lineage of *P. panamensis* (PW055 collected in Panama) is basal to a second lineage of the same species (PWTM58 from Ecuador) and to a wild *P. tuberosus* accession (TWTM48, Ecuador), hence revealing a consistent pattern of dispersal from Central American to southern localities. EW203 was resolved as a separate, more terminal subclade. Further, *P. ahipa* was resolved within *P. tuberosus* (TC536, an ashipa), although this relationship was moderately supported by 63% bootstrap and a decay index ≤ 270, i.e. only one step longer than the original consensus tree.

For the second phylogenetic analysis that combined information from cpDNA, ITS and RAPD variation, data were available only for 12 *Pachyrhizus* taxa and, therefore, *P. ferrugineus* was selected as outgroup. A single equally parsimonious tree of 444 steps was generated (Figure 6.3) with CI and RI values of 0.581 and 0.554, respectively (lower than those obtained after cpDNA-ITS combined analysis alone). Bootstrap values for most nodes were high, with an average of 85%; additionally, high decay indices (e.g. ≤ 477 for *P. ahipa*) indicated a strong internal structure of the cladogram (Figure 6.4).

The cpDNA-ITS-RAPD combined tree clearly showed that *P. ferrugineus* and *P. panamensis* taxa are most basal, with *P. ferrugineus* ancestral to *P. panamensis*, and this latter species ancestral to a lineage of wild *P. tuberosus* (TWTM48). Next, two independent subclades were evident which gave rise to *P. erosus* (wild and cultivated taxa) and to the cultigen types of the *P. tuberosus* complex. Moreover, *P. ahipa* was resolved within *P. tuberosus* (TC553, a jíquima) and strongly supported by a bootstrap value of 94% and a decay index ≤ 451.

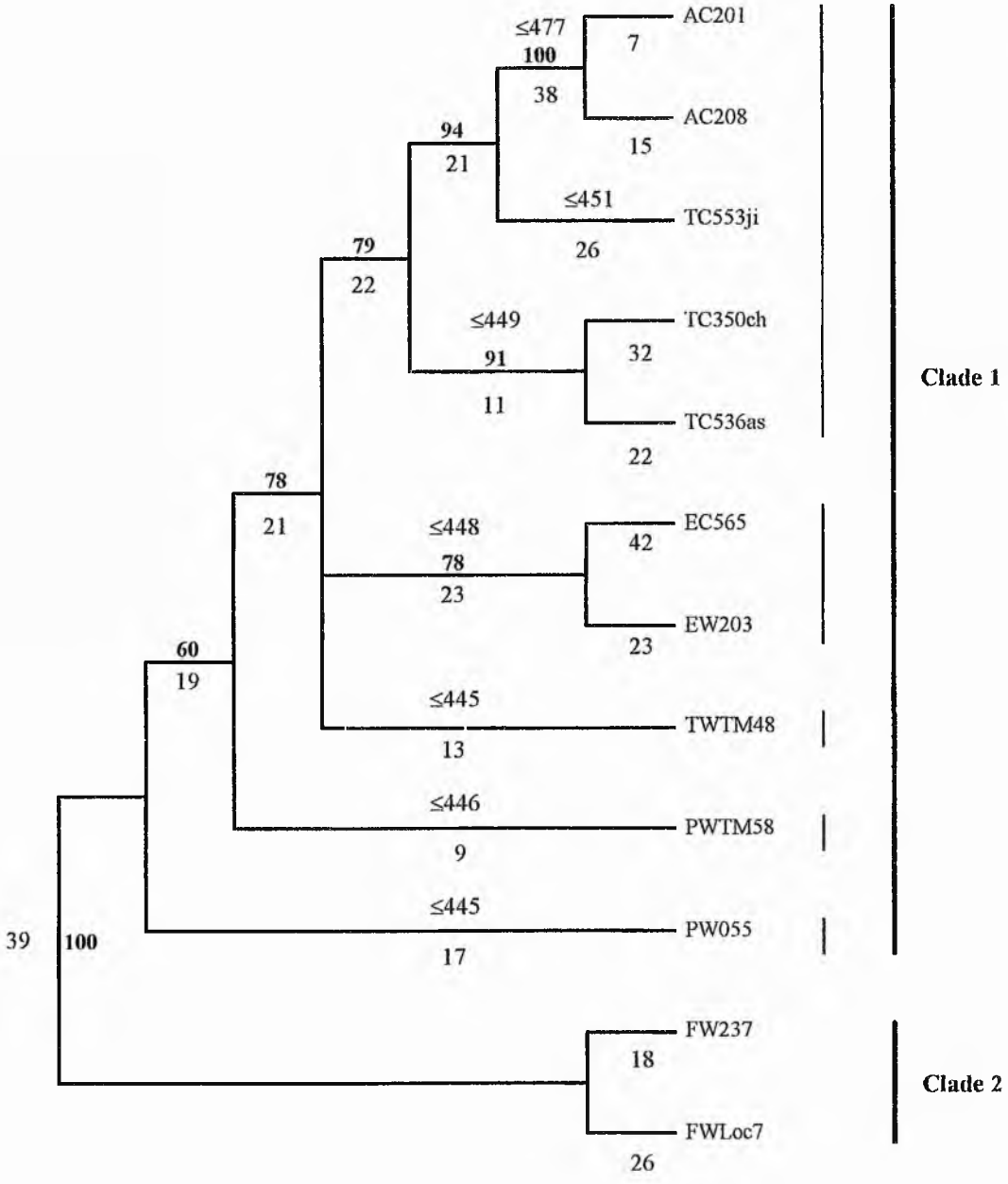


Figure 6.3. Single equally parsimonious Fitch tree derived from a combined analysis of characters obtained from cpDNA restriction site, ITS sequence and RAPD variation in 12 *Pachyrhizus* taxa. Numbers above branches indicate bootstrap percentages (in boldface) and decay values; numbers below branches are synapomorphies (Tree length = 444 steps; CI = 0.581; HI = 0.419; RI = 0.554). Key for accessions as for Table 2.1; as = ashipa; ch = chuin; ji = jíquima cultigen type.

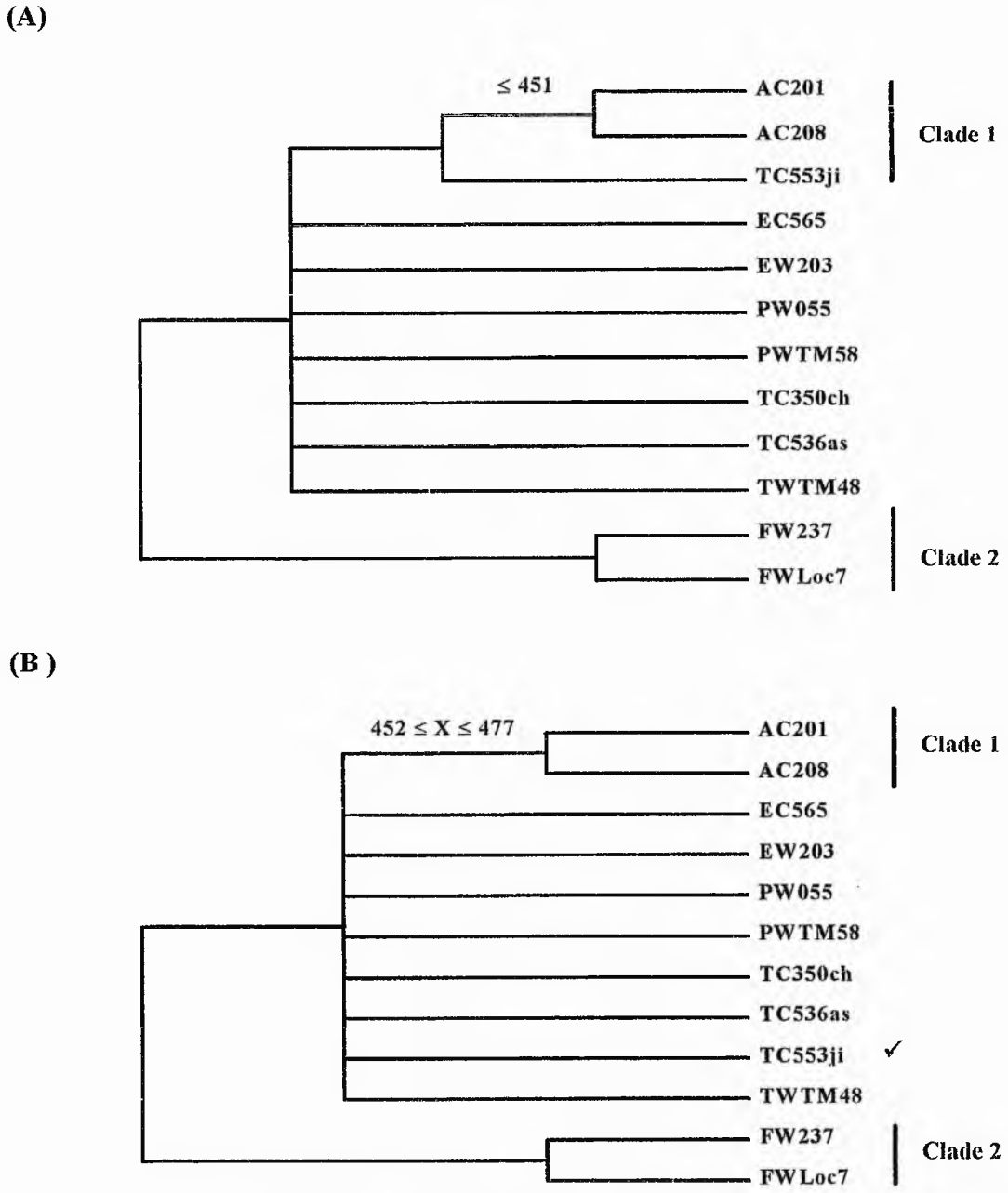


Figure 6.4. Trees showing the order of 'decay' of clades obtained after a combined analysis of cpDNA, ITS and RAPD variation. **(A)** Strict consensus of trees whose length is equal to 451 steps. All *P. ahipa* accessions examined were still resolving together with TC553 (a jíquima), whereas the remaining accessions of the *P. tuberosus* complex were no longer recovered in clade 1. **(B)** Strict consensus of trees whose length is equal to 452 steps; the branch containing TC553 (✓) collapsed and was no longer resolved together with *P. ahipa* accessions. The clade containing both *P. ahipa* accessions examined was still present up until 477 steps. The tree became an unresolved 'bush' at 34 steps away from the minimal tree, i.e. a length of 478 (see text for further details).

6.3. Concluding remarks and future research

This study has provided a substantial understanding of the evolutionary history of the yam beans from an array of molecular variation within the chloroplast and nuclear genomes. However, there are still gaps in our knowledge of relationships between particular taxa (e.g. the *P. erosus* and *P. tuberosus* complexes) and these will require further effort to achieve an in-depth understanding of the evolution of *Pachyrhizus*.

One particular taxon that requires additional analysis is *P. erosus* and its evident subspecific groups. It would be of interest to compare the molecular evidence presented in previous sections with the old taxonomic distinction between *P. erosus* of Mexican provenance and "*P. palmatilobus*" (now considered conspecific with *P. erosus*) of Central American distribution. The latter "species" was mainly distinguished from *P. erosus* on account of its deeply lobed leaflets (*P. erosus sensu stricto* having dentate leaflets; Sørensen, 1996). Nonetheless, this trait has also been found within Mexican material, particularly among the uniform landraces from the Yucatan peninsula, and, therefore, morphological evidence alone is not enough for the separation of these groups. In addition, the conspicuous linkage of the white flowered, red (or maroon)-seeded landraces from the Yucatán peninsula with an increased tolerance to high temperatures, but also with a longer production period, could be examined in the light of the molecular evidence presented here.

Additional investigation is also required into the *P. tuberosus* complex and its relationship to *P. ahipa*. It would certainly be of value to integrate the findings of the molecular analysis with information from fields such as botany, ecology, agronomy, archaeology, etc., in extending our knowledge of the evolution of this subset of material. For example, the fact that both species were known in Peru for their edible tuberous roots in the pre-agricultural period (12200 - 8500 BC), their present-day numerous non-related vernacular names and the several individually stable (mono- or multituberous) local landraces, strongly supports independent domestication processes at different locations by various Amerindian groups during the pre-Columbian period.

Future sampling should include detailed prospections of selected populations of *Pachyrhizus*, so that thorough research into population structure, outcrossing rates and intraspecific variation can be conducted, especially in regard to those species and subgroups considered to be endangered. This is of utmost importance as pressures on natural vegetation by encroaching agriculture, overgrazing, urban and industrial expansion, are causing a rapid decline in levels of genetic diversity within wild and cultivated *Pachyrhizus*. From this standpoint, the molecular techniques used throughout this study have proved to be valuable tools to investigate patterns of intra- and interspecific variation allowing the implementation (or continuation) of *in situ* and *ex situ* conservation strategies in Central and South America.

Despite the many unresolved questions, our current understanding of the evolutionary history of yam beans, which has been improved by the work reported in this thesis, has important implications for both germplasm management and enhancement. First, a wider genetic base can now be foreseen with considerable genetic diversity for future improvement residing in the wild forms. Once the implications of geographic origin and cytoplasmic interactions are fully understood, there is a real possibility of exploiting hybrid vigour more efficiently - with appropriate combinations of cytoplasm and nucleoplasm - thus enhancing genetic breeding in yam beans. Second, wild ancestral forms have provided evolutionary clues to a better understanding of genetic diversity patterns in the cultigens in locating the likely geographical origins of landrace groups. The next step will be to develop a deeper insight into the chronologies of domestication events (testing a molecular clock), and there again wild ancestral forms will play an important role.

Third, it is very likely that for some groups of landraces coevolution between the crop and its biotic (and abiotic) environment took place in the Neotropics for many thousand years. As a consequence, race formation can be expected to have occurred with respect to associated biota (e.g. *Rhizobium*). Therefore, more efficient selection for pest and disease resistance may be possible in the near future for *Pachyrhizus* accessions stored in germplasm banks. When trying to reduce expensive inputs such as nitrogen fertilisers for sustainable production, a better co-adaptation with associated

biota is worth establishing. Thus, the use of the wide range of potentially exploitable genetic resources identified in this study, may enable exciting advances to be made in the future development of yam beans.

REFERENCES

REFERENCES

Adams, R. and Demeke, T. 1993. Systematic relationships in *Juniperus* based on random amplified polymorphic DNAs (RAPDs). Taxon 42: 553 - 571.

Adjahossou, D.F. and Sogbenon, H. 1994. Effet insecticide des graines de *Pachyrhizus erosus* (L.) Urban sur quelques insectes parasites de *Vigna unguiculata* (L.) Walp. ssp. *unguiculata*. - In: Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI., 21-24 April 1992. Sørensen, M. (ed.) Jordbrugsforlaget, København. Pp. 199-214.

Appels, R. and Honeycutt, R. 1986. rDNA: evolution over a billion years. In: DNA Systematics - Vol. II: Plants. Palmer, J. (ed.). Pp. 81 - 135.

Appels, R. and Dvorak, J. 1982. The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. Theor. Appl. Genets. 63: 337 - 348.

Appels, R.; Gerlach, W.; Denis, E.; Swift, H. and Peacock, W. 1980. Molecular and chromosomal organization of DNA sequences coding for the ribosomal rRNAs in cereals. Chromosoma 78: 293 - 311.

Anonymous. 1904. The yam bean. Agricultural News - a fortnightly review of the Imperial Department of Agriculture for the West Indies. 3: 109 (Issued under the authority of the Comissioner of Agriculture of Barbados).

Armstrong, J.; Gibbs, A.; Peakall, R. and Weiller, G. 1994. The RAPDistance package. Software obtainable via anonymous FTP. Programmes and manual distributed by the authors or via www.ftp:/life.anu.edu/au/pub/RAPDistance.

Arnold, M.; Buckner, C. and Robinson, J. 1991. Pollen mediated introgression and hybrid speciation in Louisiana irises. Proc. Natl. Acad. Sci. USA 88: 1398 - 1402.

Arnheim, N. 1983. Concerted evolution in multigene families. Pp. 38 - 61. In: Evolution of Genes and Proteins. M. Nei and R. Koehn (eds.). Sinauer Associates, Sunderland, MA.

Avise, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall. London, UK. 511 p.

Ayliffe, M.; Lawrence, G.; Ellis, J and Pryor, A. 1994. Heteroduplex molecules formed between allelic sequences cause non-parental RAPD bands. Nucl. Acids Res. 22: 1632 - 1636.

Bachmann, K. 1994. Molecular markers in plant ecology. New Phytol. 126: 403 - 418.

Baldwin, B. G. 1992. Phylogenetic utility of the internal transcribed spacer of nuclear ribosomal DNA in plants: an example from the Compositae. Molecular Phylogenetics and Evolution 1(1), March: 3 - 16.

Baldwin, B. G. 1993. Molecular phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA: Chromosomal and morphological evolution reexamined. Am. J. Bot. 80 (2): 222 - 238.

Baldwin, B.; Sanderson, M.; Porter, J.; Wojciechowski, M.; Campbell, C. and Donoghue, M. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard.* 82: 247 - 277.

Ball, R.; Neigel, J. and Avise, J. 1990. Gene genealogies within the organismal pedigree of random-mating populations. Evolution 44: 360 - 370.

Banks, J. and Bieky, C. 1985. Chloroplast DNA diversity is low in a wild plant *Lupinus texensis*. Proc. Natl. Acad. Sci. USA 82: 6950 - 6954.

Barile, T. V. and Esguerra, E. B. 1984. Low temperature storage of yam beans. Postharvest Research Notes 1, 2: 23 - 25. Department of Horticulture, University of the Philippines, Los Baños.

Basudev, R. 1933. Studies in the development of the female gametophyte in some leguminous crop plants of India. - Indian Journal of Agricultural Science 3 (6): 1098 - 1107.

Baum, B. 1992. Combining trees as a way of combining data sets for phylogenetic inference and the desirability of combining gene trees. Taxon 41: 3 - 10.

Baum, D.; Sytsma, K. and Hoch, P. 1994. A phylogenetic analysis of *Epilobium* (Onagraceae) based on nuclear ribosomal DNA sequences. Syst. Bot. 19: 363 - 388.

Bayer, R.; Soltis, D. and Soltis, P. 1996. Phylogenetic inferences in *Antennaria* (Asteraceae: Gnaphalieae: Cassiniinae) based on the sequences from nuclear ribosomal DNA internal transcribed spacers (ITS). Am. J. Bot. 83: 516 - 527.

Bentham, G. 1865. Phaseoleae. In: Bentham, G. and Hooker, J. *Genera plantarum*. Vol. I, London, UK. Pp. 451 - 454.

Bertoni, M. S. 1910. Siembra del macucú (*Pachyrhizus tuberosus*). Revista Agron., Boletín Est. Agr. Pto. Bertoni. Paraguay. 4, 9 - 10: 47.

Black, I. 1993. PCR with arbitrary primers: approach with care. Insect Mol. Biol. 2: 1 - 6.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evol. 42: 795 - 803.

Bremer, B. 1991. Restriction data from chloroplast DNA for phylogenetic reconstruction: is there only one accurate way of scoring? Pl. Syst. Evol. 175: 39 - 54.

Brown, J. 1994. Probabilities of evolutionary trees. Syst. Biol. 43: 78 - 91.

Brown, D.; Wensink, P. and Jordan, E. 1972. Comparison on the ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. J. Mol. Biol. 63: 57 - 73.

Brunsfeld, S.; Soltis, P.; Soltis, D.; Gadek, P.; Quinn, C.; Strenge, D. and Ranker, T. 1994. Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: evidence from *rbc*L sequences. Syst. Bot. 19: 253: 262.

Bruneau, A.; Doyle, J. J. and Palmer, J. D. 1990. A chloroplast DNA inversion as a subtribal character in the Phaseoleae (Leguminosae). Syst. Bot. 15, 3: 378 - 386.

Bruneau, A. 1996. Phylogenetic and biogeographical patterns in *Erythrina* (Leguminosae: Phaseoleae) as inferred from morphological and chloroplast DNA characters. Syst. Bot. 21: 587 - 605.

Brücher, H. 1977. Tropische Nutzpflanzen: Ursprung, Evolution und Domestikation. Springer Verlag, Berlin, Germany. Pp. 1 - 24.

Brücher, H. 1989. Useful plants of Neotropical origin and their wild relatives. Springer Verlag, Leiden. Pp. 15 - 21.

Burkart, A. 1952. Las leguminosas argentinas silvestres y cultivadas. Acme Agency, Buenos Aires, Argentina. Segunda edición. Pp. 10.

Burkill, I. 1935. A dictionary of the economic products of the Malay Peninsula. Vol. 2. Crown Agents for the Colonies, London, UK. Pp. 15.

Caetano-Anollés, G. 1993. Amplifying DNA with arbitrary oligonucleotide primers. PCR Methods and Applications 3: 85 - 94.

Caetano-Anollés, G.; Bassam, G.; Gresshoff, P. 1991. DNA amplification fingerprinting: a strategy for genome analysis. Plant Mol. Biol. Rep. 9: 294 - 307.

Calvi, B.R. (*et al.*, snt). 1991. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: hobo, Activator, and Tam3. Cell 66: 465 - 471.

Camin, J. and Sokal, R. 1965. A method for deducing branching sequences in phylogeny. Evol. 19: 311 - 326.

Campbell, C.; Donoghue, M.; Baldwin, B. and Wojciechowski, M. 1995. Phylogenetic relationships in Maloideae (Rosaceae) - evidence from sequences of the internal transcribed spacers of nuclear ribosomal DNA and its congruence with morphology. Am.. J. Bot. 82: 903 - 918.

Cain, A. and Harrison, G. 1960. Phyletic weighting. Proc. Zool. Soc. Lond. 135: 1 - 31.

Carlson, T. & Chelm, B. 1986. Apparent eukaryotic origin of glutamine synthetase II from the bacterium *Bradyrhizobium japonicum*. Nature 322: 568 - 570.

Castellanos, J.; Zapata, F.; Peña-Cabriales, J.; Jensen, E. and Heredia. E. 1996. Symbiotic nitrogen fixation and yield of *Pachyrhizus erosus* (L.) Urban cultivars and *Pachyrhizus ahipa* (Wedd.) Parodi landraces as affected by flower pruning. Soil Biology and Biochemistry (in press). 19 p.

Castillo, R. 1995. Plant genetic resources in the Andes: impact, conservation and management. Crop Science 35 (2): 355 - 360.

Chalmers, K.; Waugh, R.; Sprent, J.; Simons, A. and Powell, W. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Heredity 69: 465 - 472.

Chan, P. and Wildman, S. 1972. Chloroplast DNA codes for the primary structure of the large subunit of fraction I protein. Biochem. Biophys. Acta 277: 677 - 680.

Chase, M. and Hills, H. 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon 40: 215 - 220.

Chase, M.; Soltis, D.; Olmstead, R.; Morgan, D.; Les, D.; Mishler, D.; Duvall, M.; Price, R.; Hillis, H.; Kiu, Y.; Kron, K.; Rettig, J.; Conti, E.; Palmer, J.; Manhart, J.; Systma, J.; Michaels, H.; Kress, W.; Karol, K.; Clark, W.; Edren, M.; Gaut, B.; Jansen, R.; Kim, K.; Wimpee, C.; Smith, J.; Furnier, G.; Strauss, S.; Xiang, Q.; Plunkett, G.; Soltis, P.; Swensen, S.; Williams, S.; Gadek, P.; Quinn, C.; Eguiarte, L.; Golenberg, E.; Learn, G.; Graham, S. Jr.; Barrett, S.; Dayanandan, S. and Albert, V. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 528 - 580.

Chen, Z. 1986. Suggestion on the classification of the genus *Paulownia*. J. Huazhong Agric. University 5: 261 - 265.

Cherfas, J. 1990. Genes unlimited. New Scientist 4: 29 - 33.

CIP. 1995. International Potato Center Programme Report 1993 - 1994. International Potato Center. Lima, Peru. 192 p.

Clausen, R. 1945. A botanical study of the yam beans (*Pachyrrhizus*). Mem. Cornell Univ. Agric. Exp. Stat. 264: 1 - 38.

Clegg, M. and Zurawski, G. 1992. Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: Molecular Systematics of Plants. Soltis, P.; Soltis, D. and Doyle, J. (eds.). Chapman & Hall. Pp. 1 - 13.

Cotter, D. J. and Gómez, R. E. 1979. Daylength effect on root development of jícama (*Pachyrrhizus erosus* Urban). HortScience 14 (6): 733 - 734.

Crawford, D. J. 1990. Plant molecular systematics: macromolecular approaches. John Wiley & Sons. New York, USA. 388 p.

Craxton, M. 1991. Linear amplification sequencing: a powerful method for sequencing DNA. Methods 3: 20 - 24.

Danforth, B. and Freeman-Gallant, C. 1996. DNA fingerprinting data and the problem of non-independence among pairwise comparisons. Molecular Ecology 5: 221 - 227.

Debouck, D. G. 1994. Introduction to the conservation of genetic resources of American tuber legumes (*Pachyrhizus*). In: Proceedings of the First International Symposium

on Tuberous Legumes; Guadeloupe, FWI, 21 - 24 April, 1992. Sørensen, M. (ed.) Jordbrugsforlaget, København. Pp. 5 - 7.

DeBry, R. and Slade, N. 1985. Cladistic analysis of restriction endonuclease cleavage maps within a maximum likelihood framework. Syst. Zool. 34: 21 - 34.

Demeke, T.; Adams, R. and Chibbar, R. 1992. Potential taxonomic use of random amplified polymorphic DNA (RAPD): a case to study in *Brassica*. Theor. Appl. Genets. 84: 990 - 994.

Demesure, B.; Sodzi, N. and Petit, R. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondria and chloroplast DNA in plants. Molecular Ecology 4: 129 - 131.

Demesure, B.; Comps, B. and Petit, R. 1996. Chloroplast DNA phylogeography on the common beech (*Fagus silvatica* L.) in Europe. Evolution 50(6): 2515 - 2520.

dePamphilis, C. and Palmer, J. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348: 337 - 339.

Doebley, J. and Wendel, 1989. Applications of RFLP to plant systematics. In: Development and application of molecular markers to problems in plant genetics. Helentjarus, T. and Barr, B. (eds.). Cold Spring Harbour Laboratories, New York, USA. Pp. 57 - 67.

Doebley, J. 1992. Molecular systematics and crop evolution. In: Molecular Systematics of Plants. Soltis, P.; Soltis, D. and Doyle, J. (eds.). Chapman Hall. New York. Pp. 202 - 222.

Donoghue, M. and Sanderson, M. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. In: Molecular Systematics of Plants. Soltis, P.; Soltis, D. and Doyle, J. (eds.). Chapman Hall. New York. Pp. 340 - 368.

Donoghue, M. and Systma, K. 1993. Phylogenetic analysis of *Viburnum* based on chloroplast DNA restriction site data. Am. J. Bot. 80 (supplement, abstract): 146.

Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. Nature 299.

Dover, G. and Flavell, R. 1984. Molecular coevolution: DNA divergence and the maintenance of function. Cell 38: 622 - 623.

Dowling, T.; DeMarais, B.; Minckley, W.; Douglas, M. and Marsh, P. 1992. Use of genetic characters in conservation biology. Conserv. Biol. 6: 7 - 8.

Dowling, T.; Moritz, C.; Palmer, J. and Rieseberg, L. 1996. Nucleic acids III. Analysis of fragments and restriction sites. In: Molecular systematics. Hillis, D.; Moritz, C. and Mable, B. (eds.) 1996. Second edition. Sinauer Associates, Inc., Sunderland, USA. Pp. 249 - 320.

Downie, S. and Katz-Downie, D. 1996. A molecular phylogeny of Apiaceae subfamily Apioideae: evidence from nuclear ribosomal DNA internal spacer sequences. Am. J. Bot. 83: 234 - 251.

Doyle, J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17: 144 - 163.

Doyle, J. and Dickson, E. 1987. Preservation of plant samples for DNA restriction endonuclease analyses. Taxon 36: 715 - 722.

Doyle, J. and Doyle, J. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf material. Phytochem. Bull. 19: 11 - 15.

Døygaard, S. and Sørensen, M. 1998. Principal component analysis of morphological characters in the genus *Pachyrhizus* DC. In: Proceedings of the Second International Symposium on Tuberous Legumes. Sørensen, M.; Estrella, J.; Hamann, O. and Ríos Ruíz, S.A. (eds.). Celaya, Gto. - Mexico, 5 - 8 August, 1996. MacKeenzie, Copenhagen, Denmark. Pp. 33 - 42.

Dyer, A. 1979. Investigating chromosomes. Edward Arnold Publishers Ltd. London, UK. 138 p.

Engel, F. 1984. Geografía humana prehistórica y agricultura precolombina de la quebrada de Chilca. Universidad Agraria, Lima, Perú. Pp. 110.

Estrella, J.; Phillips, S.; Abbott, R. J.; Gillies, A. C. & Sørensen, M. 1998. Genetic variation and relationships in agronomically important species of yam bean (*Pachyrhizus*) based on RAPD markers. In: Proceedings of the Second International Symposium on Tuberous Legumes. Sørensen, M.; Estrella, J.; Hamann, O. and Ríos Ruíz, S.A. (eds.). Celaya, Gto. - Mexico, 5 - 8 August, 1996. MacKeenzie, Copenhagen, Denmark. Pp. 43 - 59.

Evans, I. M., Boulter, D., Eaglesham, A. R. J. and Dart, P. J. 1977. Protein content and protein quality of tuberous roots of some legumes determined by chemical methods. Qual. Plant. Pl. Foods. Hum. Nutr. 27, 3 - 4: 275 - 285.

Ferguson, I. K. and Skvarla, J. J. 1981. The pollen morphology of the subfamily Papilionoideae (Leguminosae). In: Advances in Legume Systematics, part 2. Polhill, R. M. and Raven, P. H. (eds.). H. M. P. O., London, UK. Pp. 859 - 896.

Farris, J. 1989. The retention index and rescaled consistency index. Cladistics 5: 417 - 419.

Fiedler, P. L. and Jain, S. K. 1992. Conservation biology: the theory and practice of nature conservation, preservation and management. Chapman & Hall. London, UK. 514 p.

Flavell, A. 1992. *Ty1-copia* group retrotransposons and the evolution of retroelements in the eukaryotes. Genetica 86: 203 - 214.

Francisco-Ortega, J.; Jansen, R. and Santos-Guerra, A. 1996. Chloroplast DNA evidence of colonisation, adaptive radiation and hybridisation in the evolution of the Macaronesian flora. Proc. Natl. Acad. Sci. USA 93: 4085 - 4090.

Francisco-Ortega, J.; Santos-Guerra, A.; Hines, A. and Jansen, R. 1997. Molecular evidence for a Mediterranean origin of the Macaronesian endemic genus *Argyranthemum* (Asteraceae). Am. J. Bot. 84(11): 1595 - 1613.

Frankel, O. H. and Soulé, M. E. 1981. Conservation and evolution. Cambridge University Press. Cambridge, UK. 327 p.

Frankel, O. 1984. Genetic perspectives of plant germplasm conservation. In: Genetic Manipulation: Impact on Man and Society. Arber, W.; Llimensee, K.; Peacock, W. and Starlinger, P. (eds.). Cambridge University Press, Cambridge, UK. Pp. 161 - 170.

Fritsch, P. and Rieseberg, L. 1992. High outcrossing rates maintain male hermafrodite individuals in populations of the flowering plant *Datisca glomerata*. Nature 359: 633 - 636.

Felsenstein, J. 1983. Statistical inference of phylogenies. J. Roy. Statist. Soc. A. 146: 246 - 272.

Felsenstein, J. 1985. Confidence limit of phylogenies: an approach using bootstrap. Evol. 39: 783 - 791.

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics 22: 521 - 565.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5C. Software and manual distributed by the author. Department of Genetics, University of Washington, Seattle, USA.

Felsenstein, J. and Kishino, H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst. Biol. 42: 193 - 200.

Fitch, W. 1966. An improved method of testing for evolutionary homology. J. Mol. Biol. 16: 916.

Fitch, W. 1981. A non-sequential method for constructing trees and hierarchical classifications. J. Mol. Evol. 18: 30 -37.

Fritsch, P. and Rieseberg, L. 1992. High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata*. Nature 359: 633 - 636.

Furman, B.; Grattapaglia, D.; Dvorak, W. and O'Malley, D. 1997. Analysis of genetic relationships of Central American and Mexican pines using RAPD markers that distinguish species. Molecular Ecology 6: 321 - 331.

Gerbi, S. 1985. Evolution of ribosomal RNA. In: Molecular Evolutionary Genetics. MacIntyre Eds., Plenum Press, New York, USA. Pp. 419 - 518.

Gerbi, S. 1986. Evolution of ribosomal DNA. In: Molecular Evolutionary Genetics. MacIntyre Eds., Plenum Press. Pp. 419 - 517.

Gentry, A. 1982. Phytogeographic patterns as evidence for a Choco refuge. In: Biological diversification in the Tropics. Prance, G. (ed.) 1982. Columbia University Press, New York, USA. Pp. 112 - 136.

Gepts, P.; Stockton, T. and Sonnante, G. 1993. Use of hypervariable markers in genetic diversity studies. In: Application of RAPD Technology to Plant Breeding. Neff, M. (ed.), ASHS Publishers, St. Paul. Pp. 41 - 45.

Gillies, A. and Abbott, R. 1996. Phylogenetic relationships in the genus *Stylosanthes* (Leguminosae) based upon chloroplast DNA variation. Pl. Syst. Evol 200: 193 - 211.

Gillies, A. and Abbott, R. 1998. Evaluation of random amplified polymorphic DNA for species identification and phylogenetic analysis in *Stylosanthes* (Fabaceae). Pl. Syst. Evol. 211: 201 - 216.

Gillies, A.; Cornelius, J.; Newton, A.; Navarro, C.; Hernández, M. and Wilson, J. 1997. Genetic variation in Costa Rican populations of the tropical timber species *Cedrela odorata* L. assessed using RAPDs. Molecular Ecology 6: 1133 - 1145.

Girsel, L. 1994. Embryology of *Pachyrhizus ahipa* (Wedd.) Parodi (Diocleinae, Fabaceae): The ovary at the time of early ovule development. In: Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI, 21 - 24 April, 1992. Sørensen, M. (ed.). Jordbrugsforlaget, København. Pp. 145 - 150.

Goldblatt, P. 1981. Chromosome numbers in Legumes II. Ann. Missouri Bot. Gard. 68 (4): 551 - 557.

_____. 1981. Cytology and the phylogeny of Leguminosae. In: Advances in Legume Systematics. Part 2. Polhill, R. M. and Raven, P. H. (eds.). 1981. H. M. P. O., London, UK. Pp. 427 - 464.

Goldmann, W.; Goldberg, G.; Bowman, L.; Steinmetz, D. and Schlessinger, D. 1983. Mouse rDNA: sequences and evolutionary analysis of spacer and mature RNA regions. Mol. Cell Biol. 3: 1488 - 1500.

González, I.; Chambers, C.; Gorski, J.; Stambolian, D.; Schmickel, R. and Sylvester, J. 1990. Sequence and structure correlation of human ribosomal spacers. J. Mol. Biol. 212: 27 - 35.

Graham, J. and Nichol, R. 1995. An examination of the ability of RAPD markers to determine the relationships within and between *Rubus* species. Theor. Appl. Genet. 90: 1128 - 1132.

Grant, V. 1981. Plant speciation. Second edition. Columbia University Press, New York, USA.

Grum, M.; Stölen, O. and Sørensen, M. 1991. Yam bean (*Pachyrhizus* Rich. ex DC.) variety trials in Tonga, South Pacific: Fresh tuber yields, dry matter and nitrogen contents. In: Proceedings of the 26th annual meeting of the Caribbean Food Crops Society, July 29 to August 4, 1990. Mayagüez, Puerto Rico. 1991. Vol 26. - Published by the Caribbean Food Crops Society with the cooperation of the USDA-ARS-TARS, Mayagüez, Puerto Rico. Pp. 407-418.

Hadrys, H.; Balick, M. and Schierwater, B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Molecular Ecology 1: 55 - 63.

Halafihi, M. 1994. Effect of yam bean seed's (*Pachyrhizus* Rich. ex DC.) extract for controlling diamondback moth (*Plutella xylostella*) in head cabbage (*Brassica oleracea* var. KK-cross). In: Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI, 21 - 24 April 1992. Sørensen, M. (ed.). Jordbrugsforlaget, København. Pp. 191-198.

Halafihi, M.; Grum, M.; Stölen, O. and Sørensen, M. 1994. Biological nitrogen fixation in *Pachyrhizus*. In: Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI, 21 - 24 April 1992. Sørensen, M. (ed.). Jordbrugsforlaget, København. Pp. 215 - 225.

Halward, T.; Stalker, H.; LaRue, E. and Kochert, G. 1992. Use of a single primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). Plant. Mol. Biol. 18: 315 - 325.

Hamby, R. and Zimmer, E. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In: Molecular Systematics of Plants. Soltis, P. S., Soltis, D. E. and Doyle, J. J. (eds.). 1992. Chapman and Hall. London, UK. Pp. 50 - 91.

Hames, B. D. and Higgins, S. J. 1985. Nucleic acid hybridization: A practical approach. IRL Press: Oxford. 245 p.

Harris, D. 1973. The prehistory of tropical agriculture: an ethnoecological model. In: The Explanation of Culture Change: Models in Prehistory. Renfrew, C. (ed.). Duckworth, London, UK. Pp. 391 - 417.

Harris, S. A. Systematics and randomly amplified polymorphic DNA in the genus *Leucaena* Benth. (Mimosoideae; Leguminosae). Plant Syst. Evol. 197: 195 - 208.

Harris, S.; Hughes, C.; Ingram, R. and Abbott, R. 1994. A phylogenetic analysis of *Leucaena* (Leguminosae; Mimosoideae). Plant Syst. Evol. 191: 1 - 26.

Harris, S. and Ingram, R. 1992. Molecular systematics of the genus *Senecio* L. I: Hybridisation in a British polyploid complex. Heredity 69: 1 - 10.

Hawkes, J. G. 1986. The domestication of South American roots and tubers. In: Foraging and Farming. Harris, D. R. and Hillman, G. C. (eds.). Unwin Hyman, London, UK. Pp. 481 - 503.

Hemleben, V.; Ganal, M.; Gerstner, J.; Schiebel, K. and Torres, R. 1988. Organization and length heterogeneity of plant ribosomal RNA genes. In: Architecture of Eukaryotic genes. G. Kahl (ed.). VCH, Weimheim, Fed. Rep. Germany. Pp. 371 - 383.

Heredia G., E. 1994. Observación de materiales segregantes y evaluación de germoplasma de jícama (*Pachyrhizus* Rich. ex DC.) en México. In: Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI, 21-24 April 1992. Sørensen, M. (ed.). Jordbrugsforlaget, København. Pp. 273-282.

Heredia, Z. and Heredia, G. 1994. San Miguelito, San Juan y Vega de San Juan: nuevas variedades de jícama (*Pachyrhizus erosus* (L.) Urban) para el Bajío, Guanajuato, Mexico. In: Proceedings of the First International Symposium on Tuberous

Legumes; Guadeloupe, FWI, 21-24 April 1992. Sørensen, M. (ed.). Jordbrugsforlaget, København. Pp. 257 - 272.

Herdenberger, F.; Weil, J. and Steinmetz, A. 1988. Organisation and nucleotide sequence of the broad bean chloroplast genes *trn*L-UAG, *ndh*F and two unidentified open reading frames. Curr. Genet. 14: 609 - 615.

Herrera, F. L. 1942. Etnobotánica: Plantas endémicas domesticadas por los antiguos peruanos. Revista Mus. Nac. Lima. 11 (1): 25 - 30.

Hewitt, G.; Johnston, A.; and Young, J. 1991. Molecular techniques in taxonomy. Springer Verlag, NATO ASI Series. 407 p.

Hillis, D. and Davis, S. 1988. Ribosomal DNA: intraspecific polymorphisms, concerted evolution, and phylogeny reconstruction. Syst. Zool. 37: 63 - 66.

Hillis, D. and Bull, J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42: 182 - 192.

Hillis, D.; Moritz, C. and Mable, B. (eds.) 1996. Molecular systematics. Second edition. Sinauer Associates, Inc., Sunderland, USA. 655 p.

Hillis, D. and Dixon, M. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Quart. Rev. Biol. 66: 411 - 453.

Hoelzel, A. 1992. Molecular genetic analysis of populations: a practical approach. IRL Press, Oxford, England.

Hoisington, D.A.; CIMMYT Laboratories. 1994. Laboratory protocols. CIMMYT Applied Molecular Genetics Laboratory, Mexico, DF., CIMMYT.

Hood, L.; Campbell, J. and Elgin, S. 1975. The organization, expression and evolution of antibody genes and other multigene families. Ann. Rev. Genet. 9: 305 - 353.

Jorgensen, R. and Cluster, P. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Ann. Missouri Bot. Gard. 75: 1238 - 1247.

Ingham, J. L. 1979. Isoflavonoid phytoalexins of yam bean (*Pachyrrhizus erosus*). Z. Naturforsch. 34c: 683 - 688.

_____. 1990. Systematic aspects of phytoalexin formation within the tribe Phaseoleae of the Leguminosae (Subfamily Papilionoideae). Biochemical Systematics and Ecology 18 (5): 329 -343.

INIAP. 1992. Informe Anual 1991 del Departamento Nacional de Recursos Fitogenéticos, DENAREF. INIAP, Quito Ecuador. Pp. 1 - 50.

INIAP. 1995. Informe Anual 1994 del Departamento Nacional de Recursos Fitogenéticos, DENAREF. INIAP, Quito Ecuador. 98 p.

INIAP. 1996. Informe Anual 1995 del Departamento Nacional de Recursos Fitogenéticos, DENAREF. INIAP, Quito Ecuador. Pp. 3- 60.

Innis, M.; Gelfand, D.; Sninsky, J. and White, T. 1990. PCR protocols. Academic Press, New York, USA.

IPGRI. 1995. Annual Report 1994. International Plant Genetic Resources Institute. Rome, Italy. 129 p.

Iwabe, N.; Kuma, K.; Hasegawa, M.; Osawa, S. and Miyata, T. 1989. Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA 86: 9355 - 9359.

Jansen, R. and Palmer, J. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc. Natl. Sci. USA 84: 5818 - 5822.

Jansen, R. and Palmer, J. 1988. Phylogenetic implication of chloroplast DNA restriction site variation in the Mutisieae (Asteraceae). Am. J. Bot. 75: 753 - 766.

Jeffreys, A.; Royle, N.; Wilson, V. and Wong, Z. 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature 332: 278 - 281.

Jorgensen, R. and Cluster, P. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Ann. Missouri Bot. Gard. 75: 1238 - 1247.

Kaemmer, D.; Ramser, J.; Schön, M.; Weigand, F.; Saxena, M.; Driesel, A.; Kahl, G. and Weising, K. 1992. DNA fingerprinting of fungal genomes: a case study with *Ascochyta rabiei* in DNA polymorphisms in eukaryotic genomes. Hüthig Verlag, Heidelberg, Germany. Pp. 255 - 270.

Kavanagh, T. A. and Ferguson, I. K., 1981. Pollen morphology and taxonomy of the subtribe Diocleinae (Leguminosae: Papilionoideae: Phaseoleae). Rev. Palaeobot. Palynol. 32: 317 -367.

Kim, K. and Jansen, R. 1994. Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*): additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. Pl. Syst. Evol. 190: 157 - 185.

Kimura, M. 1983. The neutral theory of evolution. Cambridge Univ. Press, Cambridge, UK. Pp. 10.

Klekowski, E. 1988. Mutation, developmental selection, and plant evolution. Columbia, University Press, New York, USA.

Kluge, A. and Farris, J. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18: 1 - 32.

Koller, B.; Lehmann, A.; McDermott, J. and Gessler, C. 1993. Identification of apple cultivars using RAPD markers. Theor. Appl. Genet. 85: 901 - 904.

Kreitman, M. 1991. Nucleotide polymorphism at the *Adh* locus of *Drosophila melanogaster*. Nature 304: 412 - 417.

Krishnamurti, M. and Seshadri, T. R. 1966. Chemical components of yam beans: their evolution and interrelationship. Curr. Sci. 35 (7): 167 - 169.

Lackey, J. A. 1977. A revised classification of the tribe Phaseoleae (Leguminosae: Papilionoideae) and its relation to canavanine distribution. J. Linn. Soc. Bot. 74: 163 - 178.

_____. 1980. Chromosome numbers in the Phaseoleae (Fabaceae: Faboideae) and their relation to taxonomy. Am. J. Bot. 67, 4: 595 - 602.

_____. 1981. Phaseoleae. In: Advances in Legume Systematics. Part 1. Polhill, R. M. and Raven, P. H. (eds.). H. M. P. O., London, UK. Pp. 301 - 327.

Lackhan, N. 1994. Investigations on *in vitro* production of rotenone using *Pachyrhizus erosus* (L.) Urban. MPh Thesis, Dept. Plant Science, Fac. Nat. Sci., University of St. Augustin, Trinidad. 115 p.

León, J. 1969. Andean tuber and root crops: origin and variability. In: Proceedings of the International Symposium on Tropical Root Crops. 2 - 8 April 1967, St. Augustine (Trinidad). Vol. 1: 118 - 130.

_____. 1977. Origin, evolution and early dispersal of root and tuber crops. Proceedings of the Fourth Symposium of the International Society for Tropical Root Crops. Cali, Colombia. 1 - 7 August, 1976. Ottawa, Canada. Pp. 20 - 36.

_____. 1987. Botánica de los cultivos tropicales. Segunda edición. Servicio Editorial IICA, San José, Costa Rica. Pp. 1 - 210.

Li, W. and Graur, D. 1991. Fundamentals of molecular evolution. Sinnauer, Sunderland, MA, USA.

Lieckfeldt, E.; Meyer, W.; and Börner, T. 1993. Rapid identification and differentiation of yeasts by DNA and PCR fingerprinting. J. Basic Microbiol. 33: 413 - 426.

Liston, A. 1992. Variation in the chloroplast genes *rpo*C1 and *rpo*C2 of the genus *Astragalus* (Fabaceae): evidence from restriction site mapping of a PCR-amplified fragment. Am. J. Bot. 79: 953 - 961.

Linneus, C. von, 1753. Species plantarum. Vol. 2. Impensis Laurentii Salvii, Stockholm (reproduction).

Liu, Z. and Furnier, G. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theor. Appl. Genets. 87: 97 - 105.

Llaca, V.; Delgado, A. and Gepts, P. 1994. Chloroplast DNA as an evolutionary marker in the *Phaseolus vulgaris* complex. Theor. Appl. Genets. 88: 646 - 652.

Lundell, C. 1939. Plants probably utilized by the old empire Maya of Peten and adjacent lowlands. Pap. Mich. Acad. Sci. Arts & Lett, 24, 1: 37 - 56.

Maddison, D. and Maddison, W. 1996. The Tree of Life: a distributed Internet project containing information about phylogeny and biodiversity. Internet address: http://phylogeny.arizona.edu/tree/phylogeny.html

Manos, P. 1993. Cladistic analysis of sequences from the internal transcribed spacers (ITS) of nuclear ribosomal DNA of *Notofagus*. Am. J. Bot. 80 (supplement abstract): 163.

Márquez H. and Morera J. 1992. Caracterización sistemática, parámetros genéticos e índices de selección de la colección de jícama (*Pachyrhizus erosus* (L.) Urban) del CATIE. M. Sc. thesis, Centro Agronómico Tropical para la Investigación y Enseñanza, Programa de Enseñanza/Area de Posgrado. Turrialba, Costa Rica. Pp. 1 - 103.

Martínez, M. 1979. Catálogo de nombres vulgares y científicos de plantas mexicanas. Fondo de Cultura Económica, Mexico.

Maxam, A. and Gilbert, W. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74: 560 - 564.

Meyer, W.; Koch, A.; Niemann, C.; Beyermann, B.; Epplen, J.; and Börner, T. 1991. Differentiation of species and strains among filamentous fungi by DNA fingerprinting. Curr. Genet. 19: 239 - 242.

Millar, C. and Libby, W. 1991. Strategies for conserving clinal, ecotypic, and disjunct population diversity in wide spread species. In: Genetics and conservation of rare plants. Falk, D. and Holsinger, K. (eds.). Oxford University Press, New York. Pp. 149 - 170.

Mindell, D. and Honeycutt, R. 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic implications. Ann. Rev. Ecol. Syst. 21: 541 - 566.

Möller, M. and Cronk, Q. 1997. Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. Am. J. Bot. 84(7): 956 - 965.

Montenegro, P. de. 1740. Bacucú. In: Materia Médica Misionera. de Montenegro, P. 379 - 380, Buenos Aires, Argentina (1945, 2nd ed. *P. panamensis*. 422 - 423).

Morera, J.A. 1994. Rejuvenecimiento, caraterización, evaluación de jícama (*Pachyrhizus* Rich. ex DC.) en CATIE, Costa Rica. In: Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI, 21 - 24 April 1992. Sørensen, M. (ed.). Jordbrugsforlaget, København. Pp. 287-304.

Mueller, L. and Ayala, F. 1982. Estimation and interpretation of genetic distance in empirical studies. Genet. Res. 40: 127 - 137.

Muñoz, S. 1945. Nupe, un cultivo promisorio para Venezuela. Tercera Conferencia Interamericana de Agricultura. Caracas, Venezuela. 38: 5 - 34.

Murray, V. 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucl. Acids Res. 17: 8889.

Muse, S. and Weir, B. 1992. Testing for equality of evolutionary rates. Genetics 132: 269 - 276.

Nair, S. G. 1989. Improvement of tuber yield in yam bean by mutation induction. Mutation Breeding Newsletter 34: 20.

National Research Council. 1989. Lost crops of the Incas: little-known plants of the Andes with promise for world-wide cultivation. Washington DC, USA. National Academy Press. Pp. 1 - 45.

Neale, B.; Saghai-Maroof, M.; Allard, R.; Zhang, Q. and Jorgensen, R. 1988. Chloroplast DNA diversity in populations of wild and cultivated barley. Genetics 1105 - 1110.

Neigel, J. and Avise, J. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Evolutionary Processes and Theory. E. Nevo and S. Karlin (eds.). Academic Press, New York, USA. Pp. 515 - 534.

Neilsen, P.E. 1995. Summary of progress report from Tonga, October 1993 - September 1994. Pp. 153 - 184. In: Second Annual Progress Report, STD3 Contract No. ERBTS3-CT920115. M. Sørensen (ed.).

Nesbitt, K.; Potts, B.; Vaillancourt, R.; West, A. and Reid, J. 1995. Partitioning and distribution of RAPD variation in a forest tree species, *Eucalyptus globulus* (Myrtaceae). Heredity 74: 628 - 637.

Newton, C. and Graham, A. 1994. PCR. BIOS Scientific Publishers Limited, Oxford, UK. 161 p.

Nicholas, K. and Nicholas, H. 1997. Genedoc: a tool for editing and annotating multiple sequence alignments. Software and shading utility distributed by the authors. Version 2.4, available via FTP or Internet.

O'Neale, L. M. and T. W. Whitaker. 1947. .Embroideries of the early Nasca period and the crop plants depicted on them. Southw. J. Anthropol. 3: 294 - 321.

Olmstead, R. G. and Palmer, J. D. 1994. Chloroplast DNA systematics: a review of methods and data analysis. Am. J. Bot. 81(9): 1205 - 1224.

Ørting, B. 1996. *Pachyrhizus ahipa* (Wedd.) Parodi: a biometric test under greenhouse conditions. MSc thesis. Department of Botany, Royal Veterinary and Agricultural University, Copenhagen, Denmark. 41 p.

Ørting, B. 1996a. Collecting *Pachyrhizus ahipa* (Wedd.) Parodi in Argentina and Bolivia, May 5 - June 12, 1996. Field report, Department of Botany, Dendrology and Forest Genetics, RVAU, Copenhagen, Denmark. 33 p.

Ørting, B.; Grüneberg, W. and Sørensen, M. 1996. Ahipa (*Pachyrhizus ahipa* (Wedd.) Parodi) in Bolivia. Genet. Resour. Crop Evolution 43: 435 - 446.

Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesomo, K.; Shiki, Y.; Takeuchi, M.; Chang, Z.; Aota, S.; Inokuchi, H. and Ozeki, H. 1986. Chloroplast gene organisation deducted from complete sequence of liverworth (*Marchantia polymorpha*) chloroplast DNA. Nature 322: 572 - 574.

Page, R. 1996. TreeView version 1.2a. Software and manual distributed by the author. Available via FTP or Internet. Glasgow University, UK.

Panchen, A. 1992. Methods of classification: phenetics and cladistics. In: Classification, Evolution and the Nature of Biology. Panchen, A.L. Cambridge University Press. Pp. 132 - 168.

Palmer, J. 1985. Chloroplast DNA and molecular phylogeny. Bioassays 2: 263 - 267.

Palmer, J. 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19: 325 - 354.

Palmer, J. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am. Naturalist 130: 56 - 529.

Palmer, J.; Jansen, R.; Michaels, H.; Chase, M. and Manhart, J. 1988. Chloroplast DNA variation and plant phylogeny. Ann. Missouri Bot. Gard. 75: 1180 - 1206.

Palmer, J.; Osorio, B.; Aldrich, J. and Thompson, W. 1987. Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr. Genet. 11: 275 - 286.

Palmer, J.; Shields, C.; Cohen, D. and Orton, T. 1983. Chloroplast DNA and the origin of amphidiploid *Brassica* species. Theor. Appl. Genets. 65: 181 - 189.

Palmer, J. and Zamir, D. 1982. Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. Proc. Natl. Acad. Sci. USA 79: 5006 - 5010.

Pamilo, P. and Nei, M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5: 568 - 583.

Parent, J.; Fortin, M. and Pagé, D. 1993. Identification de cultivars de framboisier par l'analyse d'ADN polymorphe amplifié au hazard (RAPD). Can. J. Hortic. Sci. 73: 1115 - 1122.

Patiño, B. 1964. Plantas cultivadas y animales domésticos en América Equinoccial. Tomo 2. Plantas alimenticias. Impresa Departamental, Cali, Colombia. Pp. 364.

Pearsall, D. M. 1992. The origins of plant cultivation in South America. In: The origins of agriculture: an international perspective. Wesley, C. & Watson, P. (eds.). Smithsonian Institution Press. Washington, DC, USA. Pp. 65 - 83.

Petit, R.; Kremer, A. and Wagner, D. 1993. Finite island model for organelle and nuclear genes in plants. Heredity 71: 630 - 641.

Philips, S. 1994. Molecular marker variation and species relationships in the genus *Pachyrhizus*. BSc thesis. School of Biological and Medical Sciences, University of St. Andrews, Scotland. 63 p.

Quicke, D. 1993. Principles and techniques of contemporary taxonomy. Chapman & Hall. London, UK. 311 p.

Rafalski, A. and Tingey, S. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends in Genetics 9 (8).

Ragot, M. and Hoisington, D. 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. Theor. Appl. Genets 86: 975 - 984.

Ramamoorthy, T.; Bye, R.; Lot, A. and Fa, J. (eds.) 1993. Biological diversity of Mexico: origins and distribution. Oxford University Press, New York, USA. Pp. 459-511.

Riedy, M.; Hamilton, W. and Aquadro, C. 1992. Excess of non-parental bands in offspring from known primate pedigrees assayed using RAPD PCR. Nucl. Acids. Res. 20: 918.

Rieseberg, L. 1996. Homology among RAPD fragments in interspecific comparisons. Molecular Ecology 5: 99 - 105.

Rieseberg, L. and Soltis, D. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Pl. 5: 65 - 84.

Rogers, S. and Bendich, A. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Mol. Biol. 9: 509 - 520.

Roy, B. 1933. Studies in the development of the female gametophyte in some leguminous crop plants of India. Indian J. Agric. Sci. 3: 1098 - 1107.

Russell, J.; Hosein, F.; Johnson, E.; Waugh, R. and Powell, W. 1993. Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. Mol. Evol. 2: 89 - 97.

Sagai-Maroof, M.; Soliman, K.; Jorgensen, R. and Allard, R. 1984. Ribosomal DNA spacer-length in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. Sci. USA 81: 8014 - 8018.

Saitou, N. and Nei, M. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406 - 425.

Sales, A. M., Baldini, V. L. S., Cunha, M. F. and Henriques, E. A. 1990. Perfil eletroforético e composicao de aminoácidos de sementes de jacatupé (*Pachyrrhizus tuberosus* Spreng). Ciencia e Tecnología de Alimentos 10 (1): 87 - 108.

Salick, J. 1989. Ecological basis of Amuesha agriculture, Peruvian Upper Amazon. *In:* Posey, D. A. and Balée, W. (eds.) Resource management in Amazonia: Indigenous and folk strategies. Advances in Economic Botany 7: 189 - 212.

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. Molecular cloning: a laboratory manual. CSHL Press. New York, USA.

Sanderson, M. and Doyle, J. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Syst. Biol. 41: 4 - 17.

Sanger, F.; Nicklen, S. and Coulson, A. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463 - 5467.

Sauer, C. 1950. Cultivated plants of South and Central America. In: Handbook of South American Indians. J.H. Steward (ed.). Smithsonian Institution, Bureau of American Ethnology. Bull. 143. P.p. 487 - 543.

Schaal, B. and Learn, G. 1988. Ribosomal DNA variation within and among plant populations. Ann. Missouri Bot. Gard. 75: 1202 - 1216.

Schiebel, K. and Hemleben, V. 1989. Nucleotide sequence of the 18S-25S spacer region from rDNA of mung bean. Nucl. Acids Res. 17: 2852.

Schwarzbach, A. and Kadereit, J. 1995. Rapid radiation of Northern American desert genera of the Papaveraceae: evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. Plant Systematics and Evolution [supplement] 9: 159 - 170.

Scotlab. 1997. EASY Store software for Win32. Image capturing, gel image analysis and quantitative densitometry. Package distributed by Scotlab- Herolab. GMBH, Germany.

Sessions, S. and Kezer, J. 1987. Cytogenetic evolution in the plethodontid salamander genus *Aneides*. Chromosoma 95: 17 -30.

Sharon, D.; Hillel, J.; Vainstein, A. and Lavi, U. 1992. Application of DNA fingerprints for identification and genetic analysis of *Carica papaya* and other *Carica* species. Euphytica 62: 119 - 126.

Shinozaki, K.; Ohme, M.; Tanake, M.; Wakusugi, T.; Hayashida, N.; Matsubayashi, T.; Aita, N.; Chunwongse, J.; Obokata, J.; Yamaguchi-Shinozaki, K.; Ohton, C.; Torazawa, K.; Meng, B.; Sugita, M.; Deno, H.; Kamogashira, T.; Yamada, K.; Kusuda, J.; Takaiwa, F.; Kato, A.; Tohdoh, N.; Shimada, H. and Suguira, M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 5: 2043 - 2049.

Siebert, D. 1992. Tree statistics, trees and confidence, consensus trees, alternatives to parsimony, character weighting, character conflict and its resolution. In: Cladistic: a Practical Course in Systematics (P.L. Forey, C.J. Humphries, I.L. Kitching, R.W. Scotland, D.J. Siebert and D.M. Williams, eds.). Oxford Science Publications, London, UK. Pp. 72 - 88.

Sinha, R. P., Prakash, R. and Haque, M. F. 1977. Genetic variability in yam bean (*Pachyrhizus erosus* Urban). Trop. Grain Leg. Bull. 7: 21 - 23.

Slade, R.; Moritz, C. and Heidemann, A. 1994. Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. Mol. Biol. Evol. 11: 341 - 356.

Smith, G. 1974. Unequal cross over and the evolution of multigene families. Cold Spring Harbour. Symp. Quant. Biol. 38: 507 - 513.

Smith, J.; Burke, C. and Wagner, W. 1996. Interspecific hybridization in natural populations of the *Cyrtandra* (Gesneriaceae) on the Hawaiian islands: evidence from RAPD markers. Pl. Syst. Evol. 200: 61 - 77.

Smith, M.; Feng, D. and Doolittle, R. 1992. Evolution by acquisition: the case for horizontal gene transfers. Trends Biochem. Sci. 17: 489 - 493.

Sneath, P. and Sokal, R. 1973. Numerical taxonomy. W.H. Freeman and Co., San Francisco, USA.

Solbrig, O. 1970. Principles and methods of plant biosystematics. Macmillan, London, UK. Pp. 26 - 41.

Soltis, D. and Kuzoff, R. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evol. 49: 727 - 742.

Soltis, D.; Soltis, P. and Ness, B. 1989. High levels of chloroplast variation and multiple origins of autopolyploidy in *Heuchera micrantha* (Saxifragaceae). Evolution 43: 650 - 656.

Soltis, D.; Soltis, P. and Bothel, K. 1990. Chloroplast DNA evidence for the origins of the monotypic *Bensoniella* and *Conimitella* (Saxifragaceae). Systematic Botany 15: 349 - 362.

Soltis, D.; Collier, T. and Edgerton, M. 1991. Chloroplast DNA variation within and among genera of the *Heuchera* group (Saxifragaceae): evidence from chloroplast transfer and paraphyly. Am. J. Bot. 78: 1091- 1112.

Soltis, P. S., Soltis, D. E. and Doyle, J. J. (eds.). 1992. Molecular systematics of plants. Chapman and Hall. London, UK. 434 p.

Sørensen, M. 1988. A taxonomic revision of the genus *Pachyrhizus* Rich. ex DC. nom. cons. Nord. J. Bot. 8, 2: 167 - 192.

_____. (ed.) 1994. Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI, 21 - 24 April 1992. Jordbrugsforlaget, København. 328 p.

_____. 1989. Pollen morphology of species and interspecific hybrids in *Pachyrhizus* Rich. ex DC. (Fabaceae: Phaseoleae). Rev. Palaeobot. Palynol. 61: 319 - 339.

_____. 1990. Observations on distribution, ecology and cultivation of the tuber-bearing legume genus *Pachyrhizus* Rich. ex DC. (Fabaceae: Phaseoleae). Wageningen Papers 90-3: 1 - 38.

_____. (ed.) 1990. Second Annual Progress Report, STD2 Contract No. TS2-A-73-DK. 163 p.

_____. (ed.) 1991. Second Annual Progress Report, STD2 Contract No. TS2-A-73-DK. 364 p.

_____. 1991. Cross breeding experiments: Compatibility, pollen fertility and germination percentages of interspecific hybrids, F_1, F_2 & F_3 in the genus *Pachyrhizus* Rich. ex DC. (Fabaceae: Phaseoleae). In: Proceedings of the Caribbean Food Crops Society, Twenty fifth Annual Meeting, Guadeloupe. 1- 6 July 1989. Vol. 25. Degras, L. (ed.) 1991. Publication INRA Antilles-Guyane, Pointe-à-Pitre, Guadeloupe. Pp. 597 - 624.

_____. (ed.) 1993. The Yam Bean Project: Final scientific report. STD2 Contract No. TS2-A-73-DK. 414 p.

_____. 1994. Review of the yam bean project 1982 - 92. In: Proceedings of the First International Symposium on Tuberous Legumes; Guadeloupe, FWI, 21 - 24 April 1992. Sørensen, M. (ed.) 1994. Jordbrugsforlaget, København. Pp. 47-54.

_____. (ed.) 1995. First annual (second biannual) progress report. STD3 Contract No. ERBTS3-CT920115. 143 p.

_____. 1996. Yam bean (*Pachyrhizus* DC). Promoting the conservation and use of underutilized and neglected crops. 2. Institute of Plant Genetics and Crop Plant Research, Gatersleben / International Plant Genetic Resources Institute, Rome. 141 p.

_____; Grüneberg, W. and Ørting, B. 1997. Ahipa (*Pachyrhizus ahipa* (Wedd.) Parodi). Pp. 13 - 74. In: Andean roots and tubers: ahipa, arracacha, maca and yacón. Hermann, M. and Heller, J. (eds.). Promoting the conservation and use of underutilized and neglected crops. 21. Institute of Plant Genetics and Crop Plant Research, Gatersleben/ International Plant Genetic Resources Institute, Rome. 141 p.

_____; Grum, M., Paull, R. E., Vaillant, V., Venthou-Dumaine, A. and Zinsou, C. 1993. Yam bean (*Pachyrhizus* species). In: Underutilized Crops: Pulses and Vegetables. Williams, J. T. (ed.). Chapman & Hall, London, New York. Pp. 59 - 102.

_____; M., Døygaard, S., Estrella, J., Kvist, L. P. and Nielsen, P. E. 1997. Status of the South American tuberous legumes species: *Pachyrhizus tuberosus* (Lam.) Spreng. Biodiversity and Conservation 6: 1581 - 1625.

Sousa, S. and Delgado S. 1993. Mexican Leguminosae: Phytogeography, endemism and origins. In: Biological Diversity of Mexico: Origins and Distribution. Ramamoorthy, T., Bye, R., Lot, A. and Fa, J. (eds.). Oxford University Press, New York, Oxford. Pp. 459 -511.

Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Molec. Biol. 98: 503 - 517.

Stace, C. 1989. Plant taxonomy and biosystematics. Edward Arnold Inc. London, UK. 264 p.

Stebbins, G. 1970. Adaptive radiation in angiosperms. I. Pollination mechanisms. Annu. Rev. Ecol. Syst. 1: 307 - 326.

Stewart, C. and Excoffier, L. 1996. Assessing population genetic structure and variability with RAPD data: application to *Vaccinium macrocarpon* (American cranberry). J. Evol. Biol. 9: 153 - 171.

Stewart, M.; Hall, L. and Maden, B. 1983. Multiple heterogeneities in the transcribed spacers of ribosomal DNA from *Xenopus laevis*. Nucl. Acids Res. 10: 2851 - 2864.

Sugiura, M. 1989. The chloroplast chromosomes in land plants. Annual Review of Cell Biology 5: 51 - 70.

Suh, Y.; Thien, L. and Zimmer, E. 1992. Nucleotide sequences of the internal transcribed spacers and 5.8S rRNA gene in *Canella winterana* (Magnoliales: Canellaceae). Nucl. Acids Res. 20: 6101 - 6102.

Suh, Y,; Thien, L.; Reeve, H. and Zimmer, E. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. Amer. J. Bot. 80: 1042 - 1055.

Swofford, 1991. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1. Computer programme distributed by the Illinois Natural History Survey, Champaign, Ill., USA.

Swofford, D. 1993. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1.1 User manual and computer programme. Illinois Natural History Survey, Champaign, Ill., USA.

Swofford, D. and Olsen, G. 1990. Phylogenetic reconstruction. Pp. 411 - 501. In: Molecular Systematics. Hillis, D.; Moritz, C. and Mable, B. (eds.). Sinauer Associates, Inc., Sunderland, USA.

Swofford, D.; Olsen, G.; Waddell, P. and Hillis, D. 1996. Phylogenetic inference. Pp. 407 - 514. In: Molecular Systematics. Hillis, D.; Moritz, C. and Mable, B. (eds.) 1996. Second edition. Sinauer Associates, Inc., Sunderland, USA.

Sytsma, K. and Schaal, B. 1990. Ribosomal DNA variation within and among individuals of *Lisianthius* (Gentianaceae) populations. Pl. Syst. Evol. 170: 97 - 106.

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

Tadera, K., Tanguchi, T., Teremoto, M., Arima, M., Yagi, F., Kabayashi, A., Nagahama, T. and Ishihata, K. 1984. Protein and starch in tubers of winged bean, *Psophocarpus tetragonolobus* (L.) DC, and yam bean, *Pachyrhizus erosus* (L.) Urban. Mem. Fac. Agric., Kagashima University 20: 73 - 81.

Templeton, A. and Lawrence, J. 1988. FRAG-LENGTH III: algorithm for the calculation of molecular weights. Software distributed by the authors.

Thweatt, R. and Lee, J. 1990. Yeast precursor ribosomal RNA: molecular cloning and probing the higher-order structure of the internal transcribed spacer I by kethoxal and dimethylsulphate modification. J. Mol. Biol. 212: 305 - 320.

Tingey, S. and del Tufo, J. 1993. Genetic analysis with random amplified polymorphic DNA markers. Plant Physiology 101: 349 - 352.

Ugent, D., Pozorski, S. and Pozorski, T. 1986. Archaeological manioc (*Manihot*) from coastal Peru. Econ. Bot. 40 (1): 78 - 102.

_____. and Peterson, L. W. 1988. Archaeological Remains of Potato and Sweet Potato in Peru. CIP Circular 16,3: 1 - 10.

Urbatsch, L. and Baldwin, B. 1993. ITS DNA sequence data and chloroplast restriction site data in the phylogenetics of coneflowers (Asteraceae). Am. J. Bot. 80(2): 186.

Urbina, M. 1906. Raíces comestibles entre los antiguos mexicanos. Anales Mus. Nac. México 3 (Ser. 2): 117 - 190. (Xicama, *P. panamensis,* pp. 123 - 125).

Van Buren, R.; Harper, K; Andersen, W.; Stanton, D.; Seyoum, S. and England, J. 1994. Evaluating the relationships of autumn buttercup (*Ranunculus acriformis* var. *aestivalis*) to some close congeners using random amplified polymorphic DNA. Am. J. Bot. 81: 514 - 519.

Van Dijk, P. and Bakx-Schotman, T. 1997. Chloroplast DNA phylogeography and cytotype geography in autopolyploid *Plantago media*. Molecular Ecology 6: 345 - 352.

Van Dongen, S. 1995. How should we bootstrap allozyme data? Heredity 74: 445 -447.

Vaughn, A. and Jackson, M. 1994. The core as a guide to the whole collection. In: Proceedings of the International Workshop on Core Collections. Hodkin, I. and Brown, A. (eds.). IBPGR/CGN/CENARGEN. Workshop on Core Collections, Brasilia, Brasil. Pp. 23 - 28.

Venkateswarlu, K. and Nazar, R. 1991. A conserved core structure in the 18S-25S rRNA intergenic region from tobacco *Nicotiana rustica*. Pl. Molec. Biol. 17: 189 - 194.

Verdcourt, B. 1970. Studies in the Leguminosae-Papilionoideae for the Flora of Tropical East Africa. Vols. III and IV. Kew Bull. 24: 379 - 447 and 507 - 569.

Vidal, B. de C. and Pimentel, E. R. 1985. Caracteriçao bioquímica, topoquímica e estrutural dos corpos proteicos do feijao macuco ou jacatupé (*Pachyrhizus tuberosus* (Lam.) Spreng.). Revista Brasil. Bot. 8: 223 - 229.

Virk, P.; Ford-Lloyd, B.; Jackson, M. and Newbury, H. 1995. Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74: 170 - 179.

Wagner, D.; Furnier, G.; Saghai-Maroof, M.; Williams, S.; Dancik, B. and Allard, R. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. Proc. Natl. Acad. Sci. USA 84: 2097 - 2100.

Wang, W.; Pai, R.; Lai, C. and Lin, T. 1994. Molecular evidence for the hybrid origin of *Paulownia taiwaniana* based on RAPD markers and RFLP of chloroplast DNA. Theor. Appl. Genets. 89: 271 - 275.

Weatherhead, P. and Montgomerie, R. 1991. Good news and bad news about DNA fingerprinting. Trends in Ecology and Evolution 6: 173 - 174.

Weber, J. and May, P. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388 - 396.

Weising, K., Nybom, H., Wolff, K. and Meyer, W. 1994. DNA fingerprinting in plants and fungi. CRC Press. London, UK. 322 p.

Welsh, J. and McClelland, M. 1990. Fingerprinting genes using PCR with arbitrary primers. Nucl. Acids Res. 18: 7213 - 7218.

Wendel, J.; Schnabel, A. and Seelanan, T. 1995. Bi-directional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossipium*). Proc. Nat. Acad. Sci. USA 92: 280 - 284.

White, M. 1973. Animal cytology and evolution. Cambridge University Press. Third edition. Cambridge, UK.

White, T.; Bruns, T.; Lee, S. and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols: A guide to methods and applications (M. Innis, D. Gerald, J. Snisky and T. White, eds.). Academic Press, San Diego, California. Pp. 315 - 322.

Williams, S.; DeBry, R.; and Feder, J. 1988. A commentary on the use of ribosomal DNA in systematic studies. Syst. Zool. 37: 60 - 62.

Williams, J.; Kubelik, A.; Livak, K.; Rafalski, A. and Tingey, S. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research, Vol. 18 (No. 22).

Williams, J.; Hanafey, M.; Rafalski, A. and Tingey, S. 1993. Genetic analysis using random amplified polymorphic DNA markers. Methods in Enzymology 218: 704 - 740.

Wilde, J.; Waugh, R. and Powell, W. 1992. Genetic fingerprinting of *Theobroma* clones using random amplified polymorphic DNA markers. Theor. Appl. Genet. 83: 871 - 877.

Wilmer, P.; Gilbert, F.; Ghazoul, J.; Zalat, S. and Semida, F. 1994. A novel form of territoriality: daily paternal investment in an anthophorid bee. Animal Behaviour 48: 535 - 549.

Wojciechowski, M.; Sanderson, M.; Baldwin, B. and Donoghue, M. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): evidence from nuclear ribosomal DNA internal transcribed spacer sequences. Am. J. of Bot. 80(6): 711 - 722.

Wolfe, K.; Li, W.; and Sharp, P. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proc. Nat. Acad. Sci. USA 84: 9054 - 9058.

Wolfe, K. and Sharp, P. 1988. Identification of functional open reading frames in chloroplast genomes. Gene 66: 215 - 222.

Wolff, K.; Peters-Van Rijn, J. and Hofstra, H. 1994. RFLP analysis in chrysanthemum. I. probe and primer development. Theor. Appl. Genets. 88: 472 - 478.

Wolff, K.; El-Akkad, S. and Abbott, R. 1997. Population substructure in *Alkana orientalis* (Boraginaceae) in the Sinai desert, in relation to its pollinator behaviour. Molecular Ecology 6: 365 - 372.

Yacovleff, E. 1933. La jíquima, raíz comestible extinguida en el Perú. Revista Mus. Nac. Lima, Peru 2 (1): 51 - 66.

_____ and Herrera, F. L. 1934. El mundo vegetal de los antiguos peruanos. Botánica Etnológica. Revista Mus. Nac. Lima, Peru 3 (3): 241 - 322.

_____ and Muelle, J. C. 1933. Un fardo funerario de Paracas. Revista Mus. Nac. Lima, Peru 3 (3): 243 -153.

Yang, R.Z. and Tang, C.S. 1988. Plants used for pest control in China: a literature review. Econ. Bot. 42 (3): 376 - 406.

Yokota, Y.; Kawata, T.; Iida, Y.; Kato, A. and Tanifuji, S. 1989. Nucleotide sequences of the 5.8S rRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. J. Mol. Evol. 29: 294 - 301.

Yuan, Y.; Kupfer, P. and Doyle, J. 1996. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. Am. J. Bot. 83: 641 - 652.

Zietkiewicz, E.; Rafalski, A.; and Labuda, D. 1994. Genome fingerprinting by simple sequence repeats (SSR)-anchored PCR amplification. Genomics 20: 176 - 183.

Zimmer, E.; Martin, S.; Beverly, S.; Kan, Y. and Wilson, A. 1980. Rapid duplication and loss of genes coding for the alpha chains of haemoglobin. Proc. Natl. Acad. Sci. USA 77: 2158 - 2162.

Zuckerkandl, E. and Pauling, L. 1965. Evolutionary divergence and convergence in proteins. Pp. 97 - 166. In: Horizons in Biochemistry. M. Kasha and B. Pullman (eds.). Academic Press, New York, USA.

APPENDIX

Appendix 1. An optimised methodology for chromosome counting in *Pachyrhizus*.

The preparation of spreads of mitotic metaphase chromosomes involved four main steps: **(1)** selection of tissues with a high mitotic activity (and stimulation of such activity); **(2)** *in vitro* treatment with a mitotic arresting agent, **(3)** hypotonic treatment and fixing of tissues; and, **(4)** chromosome preparation and counting.

- Choose brittle, translucent roots with cream to white tips (for best results: place young *Pachyrhizus* plantlets in the dark for at least 10 h and expose them to 2000-lux light for 1 h to stimulate cell division).

- Immerse 1 cm root tips in freshly prepared prefixative solution (use either hydroxyquinoline 0.002 M, or colchicine 0.05%). Incubate at room temperature (18°C) for 2 - 5 h.

- Fix roots in freshly prepared solution of absolute alcohol : acetic acid (3:1) for 1 - 2 h. Store at 4°C (fridge). If necessary, roots can be stored in fixative for several months, provided that the solution does not dry out (alternatively, roots can be stored in 70% ethanol).

- Hydrolyse roots in 1N hydrochloric acid at 60°C for 10 min (tougher material may require up to 15 min) and rinse twice with distilled water.

- Staining: place roots in Feulgen solution (*Schiff's reagent*; S5133, SIGMA) in the dark for 30 min (material can be left overnight during this stage).

- Remove the root from the Feulgen solution, rinse in distilled water and place onto a clean slide. Cut off the stained part of the root (the root tip) and place on another clean slide.

- Add one drop of 2% aceto-orcein stain (or, alternatively, lacto-propionic orcein). Cut the stained portion into small pieces and cover with a clean coverslip (ensure that any glass chips or grit have been removed beforehand).

- Working under a low power microscope tap on the coverslip with a glass rod and/or spread with a cutting needle until the cells are distributed in a uniform monolayer. There should be enough stain to produce a thick suspension, but not enough to allow the material to escape from under the descending rod.

- Place filter paper over the coverslip and press down firmly with one thumb taking care not to move the coverslip. For best results: keep (parallel) preparations at 4°C while working on other root tips.

- The cells should be well separated into a monolayer and the chromosomes flat without the cell wall being broken. Identify the slide; seal the edges of the coverslip to the slide with nail polish (!). Examine under phase contrast microscopy; record data and file slides (short-term storage only).

Appendix 2. Restriction fragment data obtained by an FOA analysis of total cpDNA RFLP variation. Numbers in the header of each column refer to characters described in Appendix 3. The binary data represent presence (1) or absence (0) of a particular restriction fragment, while missing data are indicated by '?'.

Taxa	Characters
	1111111111222222222233333333334444444444555555555566666666667
	1234567890123456789012345678901234567890123456789012345678901234567890
AC201	11111101011111110000110111111001100000011011100000110101100001100
AC208	11111101?1??11100001101111110011000000110111000001101011000011000
AC215	11111101?11111100001101111110011000000110111000001101011000011000
AC220	11111101?11111100001101111110011000000110111000001101011000011000
AC222	111?11010??????00001101111110011000000110111000001101011000011000
AC226	11111101?????????00001101111110011000000110111000001101011000011000
AC231	111?11010????11100001101111110011000000110111000001101011000011000
EC032	11111101111111110000110111111001100011111011100000110101100001100
EC502	11111101011111110000110111111001100011111011100000110101100001100
EC509	11111101111111110000110111111001100011111011100000110101100001100
EC531	1111110107??1???00001101111110011000101110111000001101111000011000
EC558	11111101011111110000110111111001100010111011100000110101100001100
EC565	11111101011111110000110111111001100010111011100000110101100001100
EW051	??11110?011111110000010117?11100110001011101110000011010110000110001100
EW203	??11110?111111110000010117?11100110001011101110000011011110000110001100
FW237	01111101011001110000110117?10010110000101101110000011010110000110001100
FWLoc7	11111101017100111000011011?110101100001011011100000110101100001100
PW055	11111101011111110000110111111010110000001101110000011010110000110001100
PWTM58	????110?0??1????00001101111110101100000011011100000110111100001100
TC350ch	0???110101111???00001101111110011000000110111000001101111000011000
TC354ch	0?111101011111???00001101111110011000100110111000001101011000011000
TC536as	0?111101111111111000011011111100110001111011100000110111100001100
TC550ji	0?1?110101117????00001101111?00110000001101110000011011110000110001100
TC553ji	11111101017111110000110111110101100000011011100000110101100001100
TC556as	11111101011111110000110111111001100000011011100000110111100001100
TCNA10	01111101011111110000110111111001100011111011100000110101100001100
TW558	11101101011111110000110111111001100000011011100000110101100001100
TWNanII	01??11010??11???00001101111110011000000110111000001101011000011000
TWTM48	11101101111111110000110110110100001000011000010000110100010000000000
OUTcc	10010110100000001111107010011010001000001000?11100110101111001110011
OUTce	00011000100000100000000000010000010000000000110000001011001110011000
OUTcm	10010110000000001111107011011010017100001100007111001101011111011110011

Taxa	Characters
	11111111111111111111111111111111111111111
	77777777788888888889999999990000000000111111111122222222233333333334
	12345678901234567890123456789012345678901234567890123456789012345678901234567890
AC201	00101100000011100000100101011000111111111010101111100011110001101010111
AC208	00101100000011100000100101011000111111111010101111100011110001101010111
AC215	00101100000011100000100101011000111111111010101111100011110001101010111
AC220	01101100000011100000100101011000111111111010101111100011110001101010111
AC222	00101100000011100000100001011000111111111010101111100011101001101010111
AC226	00101100000011100000100101011000111111111010101111100011110001101010111
AC231	00101100000011100000100101011000111111111010101111100011110001101010111
EC032	00101100000011100000100101011000111111111010110110100011110010101010111
EC502	01101100000011100000100101011000111111111010110110100011110001101010111
EC509	00101100000011100000100101011000111111111010110110100011110001101010111
EC531	01101100000011100000100101111000111111111010110110100011110001011010111
EC558	00101100000011100000100101111000111111111010110110100011110001101010111
EC565	00101100000011100000100101111000111111111010110110100011110001101010111
EW051	01101100000011100000100101111000111111111010110110100011110001101010111
EW203	01101100000011100000100101111000111111111010110110100011110001101010111
FW237	10101100000001100111100000111000101111101101011101000011100010011010111
FWLoc7	10101100000011100000000010101100011111110110101110100011110001100111111
PW055	01101100000011100000100101011000101111101101011101000011110001101010111
PWTM58	01101100000011100000100101011000111111111010101111100011110001?0?0?0111
TC350ch	01101100000011100000?001?10110001111111101101101101000111101010010101111
TC354ch	01101100000011100000?001?10110001111111101010111010001111010110101011111
TC536as	01101100000011100000100101011000111111101101011101000111100011010101111
TC550ji	01101100000011100000?001?10110001111111101010111010001111000110101011111
TC553ji	01101100000011100000?001?10110001111111101010111110001111000110101011111
TC556as	01101100000011100000100101011000111111101101011101000111100011010101111
TCNA10	01101100000011100000?001?10110001111111101010111010001111000110101011111
TW558	11101100000011100000100101011000111111111111011101000111100011010101111
TWNanII	01101100000011100000100101011000111111111111011101000111100011010101111
TWTM48	01101100000011100000100101011000111111111111011101000111100011010101111
OUTcc	0?0100111101010110000111000011111100000000000001111110111000000001011 00
OUTce	00010000101010001011000000000010000000000000000000000000000010010100100 11
OUTcm	01010011?11010101000111100001?111100000000000001111110111000000001011 00

Appendix 2. Restriction fragment data obtained by FOA. Continued.

```
Taxa                              Characters
          1111111111111111111111111111111111111111111111111111111111122222222222
          4444444445555555555666666666677777777778888888888999999999900000000001
          1234567890123456789012345678901234567890123456789012345678901234567890
AC201     0000011101111111111110000001101111111100000111110011011000110111110 1100
AC208     0000011?01111111111100000011011111111000001111100110110001101111101100
AC215     0000011101111111111110000001101111111100000111110011011000110111110 1100
AC220     0000011101111111111110000001101111111100000111110011011000110111110 1100
AC222     0000011001111111111110000001101111111100000111110011011000110111110 1100
AC226     0000011101111111111110000001101111111?1000001111100110110001101111101100
AC231     0000011001111111111110000001101111111?1000001111100110110001101111101100
EC032     0000011010111111111110000001101010111010000011111001111100011111111101100
EC502     0000011010111111111110000001101010111010000011111001111100011111111101100
EC509     0000011010111111111110000001101010111010000011111001111100011111111101100
EC531     0000011010111111111110000001101010111010000011111001111100011111111101100
EC558     0000011010111111111110000001101010111010000011111001111100011111111101100
EC565     0000011010111111111110000001101010111010000011111001111100011111111101100
EW051     0000011010111111111110000001101010111010000011111001111100011111111101100
EW203     0000011001111111111110000001101010111010000011011001111100011111111101100
FW237     0000011001111111111110000001001001111010000011112001101100011011111101100
FWLoc7    0000011001111111111110000001101010111010000011111001101100011011111101100
PW055     0000011001111111111110000?1?0?1011101000001111100110110001101111101100
PWTM58    00000??001111111111110000???0??011101000001111?001101100011011111101100
TC350ch   0000011010111111111110000011101101110100000111110011011000110111110 1100
TC354ch   0000011010111111111110000011101101110100000111110011011000110111110 1100
TC536as   0000011010111111111110000011101101110110000111110011011000110111110 1100
TC550ji   0000011010111111111110000011101101110100000111110011011000110111110 1100
TC553ji   0000011001111111111110000001101111111110000111110011011000110111110 1100
TC556as   0000011010111111111110000011101101110110000111110011011000110111110 1100
TCNA10    0000011010111111111110000001?1011011101000001111100110110001101111101100
TW558     0000011010112111111110000011101101110100000111110011011000110111110 1100
TWNanII   0000011001111111111110000001101111111110000011111001101100011011111101100
TWTM48    0000011001110011111110000001100101110100000111110011011000110111110 1100
OUTcc     1111111101000000000111112010001010011111111000101001101100111010000111011
OUTce     1000000000010000001011000100010100100011?0100010000010000001010101010100
OUTcm     1111111001000000000111?10100010100111111101000111000011101010001110 11
```

```
Taxa                              Characters
          2222222222222222222222222222222222222222222222222222222222222222222222
          1111111112222222223333333333444444444455555555556666666666777777777778
          1234567890123456789012345678901234567890123456789012345678901234567890

AC201     0000111011000001111011010100111111111000111111000011?011000111001107011
AC208     0000111011000001111011010100111111111000111111000011001100011100110011
AC215     0000111011000001111011010100111111111007111111000011001100011100110011
AC220     0000111011000001111011010100111111111007111111000011001100011100110?011
AC222     0000111011000001111011010100111111111000111111000011001100011100110011
AC226     0000111011000001111011010100111111111007111111000011001100011100110011
AC231     0000111011000001111011010100111111111000111111000011001100011100110?011
EC032     0000111011000001001111101010011111111100011111100001100110001110011001 1
EC502     0000111011000010111011010100111111111007111111000011001100011100110011
EC509     0000111011000010111011010100111111111007111111000011001100011100110?011
EC531     0000111011000010111011010100111111111007111111000011001100011100110011
EC558     0000111011000010111011010100111111111007111111000011001100011100110011
EC565     0000111011000010111011010100111111111000111111000011001100011100110011
EW051     0000111011000010111011010100111111111000111111000011001100011100101011
EW203     0000111011000001111011010100111111111000111111000011001100011100110011
FW237     0000111011000001111011010100111111111000111111000011001100011100110111
FWLoc7    0000111?1000001111011010100111111111000111111000011001100111?01010011
PW055     0000111011000001111011010100111111111007111111000011001100011100110011
PWTM58    0000111011000001111011100100111111111007111111000011001100011100110011
TC350ch   0000111011000010111011010100111111111000111111000011001100011100110011
TC354ch   0000111011000001111011010100111111111000111111000011101100011100110011
TC536as   0000111011000001111011010100111111111000111111000011001100011100110011
TC550ji   0000111011000001111011010100111111111000111111000011001100011100110011
TC553ji   0000111011000001111011010100111111111007111111000011001100011100110011
TC556as   0000111011000001111011010100111111111007111111000011001100011100110011
TCNA10    0000111011000001111011010100111111111000111111000011001100011100110011
TW558     0000111011000001111011010100111111111000111111000011001100011100110011
TWNanII   0000111011000001111011010100111111111007111111000011001100011100110011
TWTM48    0000111011000001111011010100111111111007111111000011001100011100101011
OUTcc     ??111001001111000110111100111011?1011?111111??111101?01110001011100000
OUTce     0001010001000000000101001100000000001000000000010001010100000011001 00
OUTcm     ??111001001111000110101010111011?1011?1111111?111101?01110001?10100011
```

Appendix 2. Restriction fragment data obtained by FOA. Continued.

Taxa	Characters
	22222222222222222223333333333333333333333333333333333333333333333333333333
	88888888899999999990000000000111111111122222222223333333333444444444445
	12345678901234567890123456789012345678901234567890123456789012345678901234567890
AC201	10111100101010110101101101010110010110110110111?10101110011111110111111
AC208	101111001010101101011011010110010110010110111111110101110011111110111111
AC215	10111100101010110101101101010110010110010110111111110101110011111110111111
AC220	101111001010101101011011010110010110010110111111110101110011111110111111
AC222	101111001010101101011011010110010110010110111111110101110011111110111111
AC226	101111001010101101011011010110010110010110111111110101110011111110111111
AC231	101111001010101101011011010110010110010110111111110101110011111110111111
EC032	101111001010101101011011010110101001100101111100111010111001111110111111
EC502	10111100101010110101101101010110101010010111110?10101110011111110111111
EC509	101111001010101101011011010110101010010111110011101011100111111101111111
EC531	10111100101010110101101101010110101010010111110?10101110011111110111111
EC558	101111001010101101011011010110101010010111110?10101110011111110111111
EC565	101111001010101101011011010110101010010111110?10101110011111110111111
EW051	101111001010101101011011010010011010010111110011101011100111111101111111
EW203	101111001010101101011011010110100110010111110011101011100111111101111111
FW237	10111100101011101011010011000110010?1111111011111001111101111111
FWLoc7	10111100100010111001101101110001100101?111111101111100111111101111111
PW055	10111100101010110101101101011000?01100101101111?10101110011111110111111
PWTM58	101111001010101101011011010110?01100101101111?10101110011111110111111
TC350ch	10111100101010110101101101101110100101111100111010111001111110111111
TC354ch	101111001010101101011011010110100110010110111111?10101110011111110111011
TC536as	10111100101010110101101101010110100110010110111111110101110011111110111011
TC550ji	101111001010101101011011010110100110010110111111110101110011111110111011
TC553ji	10111100101010110101101101011000101101101101111111010111001111110111111
TC556as	101111001010101101011011010110100110110110111111110101110011111110111011
TCNA10	10111100101010110101101101010110?00?100101101?1?1?10101110011111110111?11
TW558	101111001010101101011011010110100110010110111111?10101110011111110111011
TWNanII	10111100101010110101101101011000?00110010110111?1?10101110011111110111?11
TWTM48	10111100101010110101101101011000101101101111111010111001111110111111
OUTcc	11111011?0101101010111010110000001101100111110101?10001111001111010011
OUTce	01000000000100?001000100000000000000000000000100000000000000010000100000
OUTcm	01111011?0101101010111010111000001101100111110101?10001111001111010011

Taxa	Characters
	33333333333333333333333333333333333333333333333333344444444444444444444444
	55555555566666666667777777777888888888899999999990000000000111111111112
	12345678901234567890123456789012345678901234567890123456789012345678901234567890
AC201	01110011111110001001000010111117?0?0101101100111011001100110011110101100
AC208	011100111111100010010000101111110000101101100111011001100110011110101100
AC215	0111001111111000100100001011111100001011011001110110011001100111101011
AC220	01110011111110001001000010111117?0?0101101100111011001100110011101?1100
AC222	011100111111100010010000101111110000101101100111011001100110011110101100
AC226	011100111111100010010000101111110000101101100111011001100110011110101100
AC231	011100111111100010010000101111110000101101100111011001100110011110101100
EC032	011100111111100010010000101111110000101101100111011001100110011101110100
EC502	011100111111100010010000101111110000101101100111011001100110101110101100
EC509	011100111111100010010000101111110000101101100111011001100110101110101100
EC531	011100111111100010010000101111110000101101100111011001100110101110101100
EC558	011100111111100010010000101111110000101101100111011001100110101110101100
EC565	011100111111100010010000101111110000101101100111011001100110101110101100
EW051	011100111111100010010000101111110000101101100111011001107110101110101100
EW203	011100101111100010010000101111110000101101100111011001100110101110101100
FW237	011100111111100010010100101111110000100111100111010101107110101110101100
FWLoc7	011100111111100010010000101111110000101101000111011001107111011100100
PW055	011100111111100010010000101111110000101101100111011001100110101110101100
PWTM58	011100111111100010010000101111110000101101100111011001100110101110101100
TC350ch	011100111111100010010000101111110000101101100111011001100110101110101100
TC354ch	011100111111100010010000101111110000101101100111011001100110101110101100
TC536as	011100111111100010010000101111???0101101100111011001100110101110101100
TC550ji	011100111111100010011000101111110000101101100111011001100110101110101101
TC553ji	011100111111100010010000101111??70101101100111011001100110101110101107
TC556as	011100111111100010010000101111??70101101100111011001100110101110101100
TCNA10	011100111111100010?000010111111000010110??001110110011001101011101101100
TW558	011100111111100010000000101111110000101101100111011001101110101110101100
TWNanII	011100111111100010000000101111110000101107?001110110011001101011101101100
TWTM48	011100111111100011000000101111110000101100110111011001101110101110101100
OUTcc	1100110?0011??100101111011111110001011010011000110010111111000001010010
OUTce	000000000000010?000000100000000001000007?000000010000000000000000000001
OUTcm	11001101?011??100100111011111?000101100100100011001011110100000101101 0

Appendix 2. Restriction fragment data obtained by FOA. Continued.

Taxa	Characters
	44444444444444444444444444444444444444444444444444444444444444444444444444444444
	22222222233333333334444444444455555555556666666666777777777788888888889
	12345678901234567890123456789012345678901234567890123456789012345678901234567890
AC201	10011110010101011111117111011111111111111111111110101011111111110110111
AC208	10011110010101011111117111011011111111111111111110101011111111110110111
AC215	10011110010101017111111?111011011111111111111111110101011111111110110111
AC220	10011110010101011111117111011011111111111111111110101011111111110110111
AC222	10011110010101011111117111011011111111111111111110101011111111110110111
AC226	10011110010101011111117111011011111111111111111110101011111111110110111
AC231	10011110010101011111117111011011111111111111111110101011111111110110111
EC032	10011110010101011111117111011011111111111111011101010111111111110110111
EC502	10011110010101011111117111011011111111111111011101010111111111110110111
EC509	10011110010101011111117111011011111111111111011101010111111111110110111
EC531	10011110010101011111117111011011111111111111011101010111111111110110111
EC558	10011110010101011111117111011011111111111111011101010111111111110110111
EC565	10011110010101011111117111011011111111111111011101010111111111110110111
EW051	10011110010101001111117111011011111111111111011101010111111111110110111
EW203	10011110010101011111117111011?11111111111111011101010111111111110110111
FW237	10011110010101011111117111011011111111111111011101010111111111110110111
FWLoc7	10011110010101011111007111011011111111111111111101010111111111111110111
PW055	10011110010101011111117111011011111111111111111101010111111111110110111
PWTM58	10011110010101011111117111011011111111111111111101010111111111110110111
TC350ch	11011110010101011111117111011111111111111111111101010111111111110110111
TC354ch	11011110010101017111117111011111111111111111111101010111111111110110111
TC536as	11011110010101011111117111011111111111111111111101010111111111111011?111
TC550ji	11011110010101011111117111011111111111111111111101010111111111110110111
TC553ji	11011110010101011111117111011111111111111111111101010111111111110110111
TC556as	11011110010101011111117111011111111111111111111101010111111111111011?111
TCNA10	10011110010101017111117111011?11111111111111111101010111111111110110111
TW558	10011110010101017111117111011011111111111111111101010111111111110110111
TWNanII	10011110010101017111117111011011111111111111111101010111111111110110111
TWTM48	10011110010101011111117111011011111111111111111101010111111111110110111
OUTcc	00?1100110101011101000000?1001111100001111110001111101110000101101000
OUTce	10000000001000000000000000000010000000000000000000000000000000000000000
OUTcm	00111001100010111010000000100011111000011111100010111011100000101101000

Taxa	Characters
	44444444455555555555555555555555555555
	99999999900000000011111111112222222222
	12345678901234567890123456789012345678
AC201	11100011011011111110000000110111111000111
AC208	11100011011011111110000000110111111000110
AC215	11100011011011111110000000110111111000111
AC220	11100011011011111110000000110111111000111
AC222	11100011011011111110000000110111111000110
AC226	11100011011011111110000000110111111000110
AC231	11100011011011111110000000110111111000110
EC032	11100011011011111110000000111111111000110
EC502	11100011011011111110000000111111111000110
EC509	11100011011011111110000000111111111000110
EC531	11100011011011111110000000111111111000110
EC558	11100011011011111110000000111111111000110
EC565	11100011011011111110000000111111111000110
EW051	11100011011011111110000000111111111000110
EW203	11100011011011111110000000111011111000110
FW237	11100010011011111110000000111011111000110
FWLoc7	11100011011011111110000000111011111000110
PW055	11100011111011111110000000110111111000110
PWTM58	11100011111011111110000000110111111000110
TC350ch	11100011011011111110000000117111111000110
TC354ch	11100011011011111110000000110111111000110
TC536as	11100011011011111110000000110111111000110
TC550ji	11100011011011111110000000110111111000110
TC553ji	11100011011011111110000000110111111000111
TC556as	11100011011011111110000000110111111000111
TCNA10	11100011011011111110000000110111111000110
TW558	11100011011011111110000000110111111000117
TWNanII	11100011011011111110000000110111111000110
TWTM48	11100011011011111110000000110111111000110
OUTcc	0001?1101111111??001111??000000000111000
OUTce	00000000000000000000000000000000000000000
OUTcm	0001?1001111111??001111??000000000111000

Appendix 3. Details of the 529 synapomorphic fragments produced by total cpDNA RFLP variation in *Pachyrhizus* taxa. These fragments have been scored as either present (1) or absent (0) in Appendix 2.

Enzyme: *Eco*RI

Character	Probe	Size (bp)	Character	Probe	Size (bp)
1	MB1	4778	25	MB2	2090
2	MB1	4254	26	MB2	1972
3	MB1	2129	27	MB2	1830
4	MB1	2096	28	MB2	1731
5	MB1	1920	29	MB2	1659
6	MB1	1876	30	MB2	1631
7	MB1	1820	31	MB2	1594
8	MB1	1681	32	MB2	1571
9	MB1	1619	33	MB2	1547
10	MB1	1555	34	MB2	1477
11	MB1	1540	35	MB2	1445
12	MB1	1482	36	MB7	3718
13	MB1	1466	37	MB7	1777
14	MB1	1446	38	MB7	1726
15	MB1	1425	39	MB7	1707
16	MB1	1400	40	MB7	1678
17	MB1	1389	41	MB7	1638
18	MB1	1359	42	MB7	1608
19	MB1	1326	43	MB7	1587
20	MB1	1295	44	MB7	1565
21	MB2	2462	45	MB7	1426
22	MB2	2389	46	MB7	1411
23	MB2	2301	47	MB7	1360
24	MB2	2252			

Enzyme: *Bam*HI

Character	Probe	Size (bp)	Character	Probe	Size (bp)
48	MB1	6256	63	MB2	1807
49	MB1	4339	64	MB2	1714
50	MB1	3119	65	MB2	1660
51	MB1	1812	66	MB2	1613
52	MB1	1770	67	MB2	1488
53	MB1	1709	68	MB2	1466
54	MB1	1657	69	MB2	1371
55	MB1	1638	70	MB9+10	2606
56	MB1	1523	71	MB9+10	2499
57	MB1	1475	72	MB9+10	2418
58	MB1	1469	73	MB9+10	2274
59	MB1	1397	74	MB9+10	2189
60	MB1	1350	75	MB9+10	1807
61	MB2	3485	76	MB9+10	1785
62	MB2	1846	77	MB9+10	1684

Enzyme: *Dra*I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
78	MB7	3433	85	MB7	1847
79	MB7	2986	86	MB7	1790
80	MB7	2653	87	MB7	1760
81	MB7	2561	88	MB7	1668
82	MB7	2129	89	MB7	1626
83	MB7	1925	90	MB7	1554
84	MB7	1876	91	MB7	1536

Appendix 3. Details of the 529 synapomorphic fragments. Continued.

Enzyme: *Cfo*I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
92	MB1	3243	110	MB7	3020
93	MB1	3007	111	MB7	2931
94	MB1	2507	112	MB7	2662
95	MB1	2186	113	MB7	2543
96	MB1	2143	114	MB7	2112
97	MB1	2102	115	MB7	1943
98	MB1	1547	116	MB7	1922
99	MB1	1464	117	MB7	1876
100	MB1	1333	118	MB7	1555
101	MB2	4648	119	MB7	1544
102	MB2	3433	120	MB7	1445
103	MB2	2241	121	MB7	1409
104	MB2	1851	122	MB7	1354
105	MB2	1599	123	MB9+10	3243
106	MB2	1558	124	MB9+10	2960
107	MB2	1484	125	MB9+10	1817
108	MB2	1441	126	MB9+10	1452
109	MB2	1400			

Enzyme: *Dde*I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
127	MB1	2389	170	MB5+6	1699
128	MB1	2312	171	MB5+6	1645
129	MB1	1805	172	MB5+6	1630
130	MB1	1796	173	MB5+6	1574
131	MB1	1788	174	MB5+6	1553
132	MB1	1759	175	MB5+6	1538
133	MB1	1741	176	MB5+6	1527
134	MB1	1730	177	MB5+6	1499
135	MB1	1703	178	MB5+6	1482
136	MB1	1690	179	MB5+6	1450
137	MB1	1669	180	MB5+6	1401
138	MB1	1627	181	MB5+6	1348
139	MB1	1585	182	MB7	2561
140	MB1	1505	183	MB7	2172
141	MB1	1423	184	MB7	2104
142	MB1	1411	185	MB7	1748
143	MB1	1389	186	MB7	1690
144	MB2	4648	187	MB7	1669
145	MB2	2814	188	MB7	1608
146	MB2	2610	189	MB7	1594
147	MB2	2337	190	MB7	1553
148	MB2	1944	191	MB7	1540
149	MB2	1822	192	MB7	1524
150	MB2	1818	193	MB7	1512
151	MB2	1733	194	MB7	1494
152	MB2	1711	195	MB7	1411
153	MB2	1646	196	MB7	1391
154	MB2	1620	197	MB9+10	3335
155	MB2	1567	198	MB9+10	3060
156	MB2	1549	199	MB9+10	1856
157	MB2	1533	200	MB9+10	1794
158	MB2	1519	201	MB9+10	1756
159	MB2	1501	202	MB9+10	1699
160	MB2	1487	203	MB9+10	1651
161	MB2	1457	204	MB9+10	1627
162	MB2	1433	205	MB9+10	1614
163	MB2	1409	206	MB9+10	1568
164	MB5+6	3007	207	MB9+10	1538
165	MB5+6	2089	208	MB9+10	1482
166	MB5+6	1805	209	MB9+10	1452
167	MB5+6	1753	210	MB9+10	1391
168	MB5+6	1711	211	MB9+10	1376
169	MB5+6	1706	212	MB9+10	1353

Appendix 3. Details of the 529 synapomorphic fragments. Continued.

Enzyme: *Eco*RV

Character	Probe	Size (bp)	Character	Probe	Size (bp)
213	MB1	3119	220	MB1	1679
214	MB1	2653	221	MB1	1665
215	MB1	2481	222	MB1	1347
216	MB1	2346	223	MB7	4764
217	MB1	1724	224	MB7	2006
218	MB1	1712	225	MB7	1925
219	MB1	1697	226	MB7	1913

Enzyme: *Alu*I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
227	MB5+6	2499	242	MB5+6	1360
228	MB5+6	2143	243	MB5+6	1341
229	MB5+6	1898	244	MB8	1829
230	MB5+6	1833	245	MB8	1703
231	MB5+6	1743	246	MB8	1668
232	MB5+6	1685	247	MB8	1619
233	MB5+6	1640	248	MB8	1590
234	MB5+6	1625	249	MB8	1545
235	MB5+6	1572	250	MB8	1494
236	MB5+6	1468	251	MB8	1451
237	MB5+6	1450	252	MB8	1420
238	MB5+6	1436	253	MB8	1403
239	MB5+6	1432	254	MB8	1373
240	MB5+6	1407	255	MB8	1352
241	MB5+6	1388	256	MB8	1316

Enzyme: *Bgl*II

Character	Probe	Size (bp)	Character	Probe	Size (bp)
257	MB1	3483	262	MB1	1513
258	MB1	2965	263	MB1	1498
259	MB1	1737	264	MB1	1454
260	MB1	1613	265	MB1	1382
261	MB1	1581			

Enzyme: *Cla*I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
266	MB9+10	3580	270	MB9+10	1707
267	MB9+10	2135	271	MB9+10	1598
268	MB9+10	2079	272	MB9+10	1537
269	MB9+10	2021	273	MB9+10	1461

Enzyme: *Hae*III

Character	Probe	Size (bp)	Character	Probe	Size (bp)
274	MB1	4218	302	MB2	1437
275	MB1	2458	303	MB2	1395
276	MB1	2454	304	MB2	1379
277	MB1	2311	305	MB7	3684
278	MB1	1817	306	MB7	2685
279	MB1	1723	307	MB7	2296
280	MB1	1675	308	MB7	1940
281	MB1	1650	309	MB7	1670
282	MB1	1625	310	MB7	1618
283	MB1	1449	311	MB7	1611
284	MB1	1420	312	MB7	1607
285	MB1	1384	313	MB7	1594
286	MB1	1354	314	MB7	1581
287	MB2	6292	315	MB7	1525
288	MB2	2365	316	MB7	1506
289	MB2	2264	317	MB7	1452
290	MB2	1976	318	MB7	1427
291	MB2	1728	319	MB9+10	2833
292	MB2	1699	320	MB9+10	2584
293	MB2	1687	321	MB9+10	2180
294	MB2	1664	322	MB9+10	1992

Appendix 3. Details of the 529 synapomorphic fragments. Continued.

Enzyme: *Hae*III. Continued.

295	MB2	1644	323	MB9+10	1859
296	MB2	1556	324	MB9+10	1743
297	MB2	1534	325	MB9+10	1685
298	MB2	1531	326	MB9+10	1516
299	MB2	1515	327	MB9+10	1468
300	MB2	1468	328	MB9+10	1387
301	MB2	1445			

Enzyme: *Hinf*I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
329	MB2	1987	347	MB7	1500
330	MB2	1919	348	MB7	1471
331	MB2	1812	349	MB7	1456
332	MB2	1646	350	MB7	1432
333	MB2	1619	351	MB7	1406
334	MB2	1583	352	MB7	1393
335	MB2	1565	353	MB9+10	1962
336	MB2	1542	354	MB9+10	1643
337	MB2	1534	355	MB9+10	1612
338	MB2	1517	356	MB9+10	1571
339	MB2	1500	357	MB9+10	1557
340	MB2	1478	358	MB9+10	1537
341	MB2	1468	359	MB9+10	1507
342	MB2	1422	360	MB9+10	1479
343	MB2	1401	361	MB9+10	1455
344	MB2	1390	362	MB9+10	1440
345	MB7	1591	363	MB9+10	1404
346	MB7	1539	364	MB9+10	1395

Enzyme: *Hpa*II

Character	Probe	Size (bp)	Character	Probe	Size (bp)
365	MB2	2022	406	MB7	1572
366	MB2	1982	407	MB7	1519
367	MB2	1899	408	MB7	1471
368	MB2	1720	409	MB7	1442
369	MB2	1712	410	MB7	1420
370	MB2	1603	411	MB7	1399
371	MB2	1549	412	MB7	1360
372	MB2	1542	413	MB7	1336
373	MB2	1520	414	MB8	1867
374	MB2	1508	415	MB8	1810
375	MB2	1491	416	MB8	1764
376	MB2	1465	417	MB8	1730
377	MB2	1442	418	MB8	1659
378	MB2	1426	419	MB8	1591
379	MB2	1403	420	MB8	1581
380	MB2	1392	421	MB8	1533
381	MB2	1373	422	MB8	1481
382	MB2	1345	423	MB8	1465
383	MB2	1324	424	MB8	1437
384	MB5+6	2199	425	MB8	1407
385	MB5+6	2149	426	MB8	1373
386	MB5+6	1976	427	MB8	1346
387	MB5+6	1915	428	MB9+10	3270
388	MB5+6	1860	429	MB9+10	2545
389	MB5+6	1842	430	MB9+10	2329
390	MB5+6	1785	431	MB9+10	2170
391	MB5+6	1723	432	MB9+10	2001
392	MB5+6	1549	433	MB9+10	1769
393	MB5+6	1506	434	MB9+10	1705
394	MB5+6	1484	435	MB9+10	1653
395	MB5+6	1434	436	MB9+10	1626
396	MB5+6	1392	437	MB9+10	1611
397	MB7	2731	438	MB9+10	1575
398	MB7	2283	439	MB9+10	1543

Appendix 3. Details of the 529 synapomorphic fragments. Continued.

Enzyme: *Hpa*II. Continued.

399	MB7	2070	440	MB9+10	1514
400	MB7	1991	441	MB9+10	1473
401	MB7	1861	442	MB9+10	1442
402	MB7	1835	443	MB9+10	1420
403	MB7	1740	444	MB9+10	1376
404	MB7	1645	445	MB9+10	1355
405	MB7	1618	446	MB9+10	1333

Enzyme: *Sau*3AI

Character	Probe	Size (bp)	Character	Probe	Size (bp)
447	MB1	1932	478	MB7	1367
448	MB1	1601	479	MB7	1347
449	MB1	1569	480	MB7	1328
450	MB1	1541	481	MB8	1766
451	MB1	1517	482	MB8	1631
452	MB1	1470	483	MB8	1593
453	MB1	1448	484	MB8	1579
454	MB1	1414	485	MB8	1491
455	MB1	1391	486	MB8	1470
456	MB1	1367	487	MB8	1466
457	MB1	1350	488	MB8	1424
458	MB1	1330	489	MB8	1405
459	MB1	1320	490	MB8	1391
460	MB5+6	1569	491	MB8	1369
461	MB5+6	1501	492	MB8	1334
462	MB5+6	1478	493	MB8	1313
463	MB5+6	1444	494	MB9+10	2170
464	MB5+6	1420	495	MB9+10	1912
465	MB5+6	1397	496	MB9+10	1855
466	MB5+6	1373	497	MB9+10	1772
467	MB5+6	1354	498	MB9+10	1717
468	MB5+6	1332	499	MB9+10	1559
469	MB7	1758	500	MB9+10	1526
470	MB7	1723	501	MB9+10	1481
471	MB7	1639	502	MB9+10	1472
472	MB7	1584	503	MB9+10	1453
473	MB7	1486	504	MB9+10	1440
474	MB7	1471	505	MB9+10	1401
475	MB7	1440	506	MB9+10	1380
476	MB7	1409	507	MB9+10	1364
477	MB7	1391	508	MB9+10	1335

Enzyme: *Tru*9I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
509	MB9+10	1534	517	MB9+10	1243
510	MB9+10	1504	518	MB9+10	1223
511	MB9+10	1468	519	MB9+10	1196
512	MB9+10	1450	520	MB9+10	1181
513	MB9+10	1434	521	MB9+10	1155
514	MB9+10	1395	522	MB9+10	1131
515	MB9+10	1319	523	MB9+10	1098/89
516	MB9+10	1254			

Enzyme: *Xba*I

Character	Probe	Size (bp)	Character	Probe	Size (bp)
524	MB9+10	4125	527	MB9+10	1942
525	MB9+10	2135	528	MB9+10	1899
526	MB9+10	2053	529	MB9+10	1446

Appendix 4. Restriction fragment data obtained by an FOA analysis of PCR-amplified cpDNA regions of *Pachyrhizus*. Numbers in the header of each column refer to characters described in Appendix 5. The binary data represent presence (1) or absence (0) of a particular restriction fragment, while missing data are indicated by '?'.

Taxa	Characters
	11111111112222222222333333333344444444445
	12345678901234567890123456789012345678901234567890
AC201	10001100?000000?000100111101011101100101111101000
AC220	10001100?000000?000100111101011101100101111101000
AC231	10001100?000000?000100111101011101100101?11101000
EC032	10001100?0000000?001001110101111011001 01?11101000
EC509	10001100?0000000?001001110101111011001 01?11101000
EC558	10001100?0000000?001011110101111011001010111010000
EW203	10001100?0000000?001001?101011110110010101111010000
FW237	00001100?100011001111101?100011110110010101111010000
FWLoc7	10000000?101111001111101?111011110000010101110011 00
PW055	10101101011011100111011?11010111111??1010111010000
PWTM58	10101101001000100000010011100111111001010111010000
TC556as	10101101001000100001011?110101110110010111110100 00
TC536as	10101101001000100001011?11010111011001 01?11101000
TC350ch	10011100?1?001100001101?10101111011001 01?11101000
TC354ch	10101111011001100001011?11010111011001 01?11101000
TC553ji	10101101001000100001011?110101110110010111110100 00
TC550ji	10101101001000100001011?110101110110010111110100 00
TWNanII	11101101001000100001011?110101110110010111111110 01?
TW558	11101111011001100111011?11010111111001?11111110011
TWTM48	11101101001000100001011?110101111111??1?11111110011
OUTcc	10111100001100000001011?000010001100000010000011 00

Appendix 5. Details of the 50 synapomorphic fragments produced by PCR-amplified cpDNA RFLP variation in *Pachyrhizus* taxa. These fragments have been scored as either present (1) or absent (0) in Appendix 4.

Character	Primer pair	Restriction enzyme	Size (bp)	Character	Primer pair	Restriction enzyme	Size (bp)
1	HK	*Hsp92*II	150	26	KK	*Rsa*I	701
2	KK	*Cfo*I	750	27	KK	*Rsa*I	572
3	KK	*Cfo*I	460	28	KK	*Rsa*I	556
4	KK	*Cfo*I	200	29	KK	*Rsa*I	540
5	KK	*Hae*III	1686	30	KK	*Rsa*I	362
6	KK	*Hae*III	676	31	KK	*Rsa*I	339
7	KK	*Hae*III	364	32	CD	*Hae*III	2058
8	KK	*Hae*III	267	33	CD	*Hind*III	507
9	KK	*Hae*III	150	34	CD	*Rsa*I	1244
10	KK	*Hind*III	435	35	CD	*Rsa*I	791
11	KK	*Hind*III	398	36	CS	*Cfo*I	422
12	KK	*Hind*III	370	37	CS	*Hind*III	416
13	KK	*Hpa*II	727	38	CS	*Hpa*II	442
14	KK	*Hpa*II	544	39	CS	*Hsp92*II	347
15	KK	*Hpa*II	442	40	CS	*Rsa*I	1287
16	KK	*Hpa*II	300	41	ML	*Cfo*I	251
17	KK	*Hpa*II	200	42	ML	*Hae*III	951
18	KK	*Hsp92*II	851	43	ML	*Hae*III	882
19	KK	*Hsp92*II	775	44	ML	*Hae*III	860
20	KK	*Hsp92*II	685	45	ML	*Hind*III	708
21	KK	*Hsp92*II	505	46	ML	*Hpa*II	1268
22	KK	*Hsp92*II	490	47	ML	*Hpa*II	636
23	KK	*Hsp92*II	346	48	ML	*Hpa*II	587
24	KK	*Hsp92*II	298	49	ML	*Hsp92*II	621
25	KK	*Rsa*I	742	50	ML	*Rsa*I	517

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

A2 (bp)	5400	5000	4800	4650	4450	4300	4200	4100	4072	4000	3800	3750	3700	3600	3500	3300	3200	3054	2700	2300	1800	1650	900	600
AC102	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	1
AC201	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC202	0	1	0	1	0	0	1	0	1	0	0	1	1	0	0	1	0	1	0	0	0	1	0	1
AC203	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC204	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1	0	1
AC205	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1
AC207L	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	1	0	0	1	1	0	1
AC207S	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	1	0	0	1	1	0	1
AC208	0	0	0	1	1	0	1	0	1	0	0	1	1	0	0	1	0	1	0	0	0	1	0	1
AC209	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	1	1
AC209BR	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC209GS	0	1	0	1	0	0	1	0	1	0	0	1	1	0	0	1	0	1	0	0	0	1	0	1
AC213	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	1	1	1	1
AC214	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC215	0	1	0	1	0	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC216	0	1	0	1	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC220A	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	0	1
AC222	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC223	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	1	1
AC225	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	1
AC226	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	0	1
AC227	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC228	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC230	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1
AC231	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	1	0	0	1	1	0	1
AC525	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	1
AC526A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1
EC006	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	0	1	1	1	0
EC032	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	1	0	0	0	1	0	0
EC033G	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0
EC109	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
EC120	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
EC201	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	1	0	0	0	1	0	0
EC205	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	1	0	0
EC214	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	1	1	0	0	1	1	1	0
EC236	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
EC250	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
EC502	1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	1	0	0	1	1	0	0
EC506	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	0	1	1	1	0
EC509	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
EC510	1	0	0	0	1	0	0	0	1	0	1	1	1	0	0	1	0	1	0	0	1	1	0	0
EC511	1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	1	0	0	1	1	0	0
EC531	1	0	0	0	1	0	1	0	1	0	0	1	0	0	0	1	0	1	0	0	1	1	1	0
EC534	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	1	0	0
EC559	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
EC560	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	1	0	0
EC565	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
EW051	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	1	0	0	1	1	0	0
EW203	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	1	1	0	0
EW223	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0	0
EW354	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
EWHue	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
EWPro	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
FW237	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0
FWGU4	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1	0
FWLoc1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	1	1	0
FWLoc7	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	0	0
PW055	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0
PWTM58	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0
PWTM59	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0
TCas118	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	1
TCas309	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	1	0	1
TCas531	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
TCas532	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0	1	1	0	1
TCas533	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	1	0	1
TCas536	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	1	0	1
TCas538	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0	1
TCas556	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	1	0	1	0	0	1	1	0	0
TCas557	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1
TCch350	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	1	1	0	0	1	1	1	0
TCch353	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	1	1	1	0	0	1	1	0	0
TCch354	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1	0	0	1	1	0	0
TCch355	1	0	0	0	1	0	0	0	0	0	1	1	0	1	0	1	1	1	0	0	1	1	1	0
TCji550	1	0	1	0	0	1	0	1	1	0	1	1	0	0	0	1	0	1	0	0	0	1	0	1
TCji552	1	0	1	0	1	1	0	0	0	0	1	0	1	1	0	1	0	1	0	0	1	1	0	0
TCji553	0	0	1	0	0	1	0	0	1	0	1	1	0	0	0	1	0	1	0	0	0	1	0	0
TCji554	1	0	1	0	1	1	0	0	1	0	1	0	1	0	0	1	0	1	0	0	0	1	0	0
TCNA06	1	0	1	0	0	0	0	1	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0
TCNA07	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0
TCNA09	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0
TCNA10	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0
TWNanI	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	1	1	0
TWNanII	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	1	1	0
TWTM48	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0	1
TWToaI	0	0	0	0	1	0	1	0	0	0	0	0	1	1	0	1	0	1	0	0	0	1	1	0

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

A13 (bp)	6300	6108	5500	5200	5100	4900	4800	4750	4500	4300	4200	4150	4100	4072	3700	3500	3450	3400	3300	3100	2036
AC102	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
AC201	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	1	0	0
AC202	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	1	0	0
AC203	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	1	0	0
AC204	0	1	0	0	0	0	1	0	0	1	0	0	1	1	0	1	0	1	1	0	1
AC205	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	0	0	0
AC207L	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	1	0	0
AC207S	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	0	0	0
AC208	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	1	0	0
AC209	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
AC209BR	0	1	0	1	0	0	1	0	0	1	0	0	1	1	0	1	0	1	0	0	0
AC209GS	0	1	0	1	0	0	1	0	0	1	0	0	1	1	0	1	0	1	0	0	0
AC213	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
AC214	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	1
AC215	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	0	0	0
AC216	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	0	0	0
AC220A	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	0	0	0
AC222	0	1	0	0	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	1	1
AC223	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
AC225	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0
AC226	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
AC227	0	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	0	1	0	0	0
AC228	0	1	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	1	1
AC230	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	1	1
AC231	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
AC525	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0
AC526A	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0
EC006	0	1	1	0	1	0	1	1	1	1	0	0	1	0	0	1	0	1	1	0	0
EC032	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1	1	0	0
EC033G	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
EC109	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
EC120	0	1	1	0	0	0	1	1	0	1	0	0	0	0	0	1	0	1	1	0	0
EC201	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
EC205	0	1	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0	1	0	0	0
EC214	1	1	1	1	1	0	1	1	0	1	0	0	1	0	0	1	0	1	1	0	0
EC236	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
EC250	1	0	0	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
EC502	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0
EC506	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1	1	0	0
EC509	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	1	0	1	1	0	0
EC510	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0
EC511	1	1	1	0	1	0	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0
EC531	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	1	0	1	1	0	0
EC534	1	1	1	1	0	0	1	1	0	1	0	1	0	1	0	1	0	1	0	0	0
EC559	1	1	1	1	0	0	1	1	0	1	0	0	0	1	0	1	0	1	0	0	0
EC560	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	1	0	1	1	0	0
EC565	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
EW051	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0
EW203	0	1	1	0	1	0	1	1	1	1	0	0	0	0	0	1	0	1	0	0	0
EW223	1	0	0	1	1	0	1	1	0	1	1	0	0	0	0	1	0	1	0	0	0
EW354	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
EWHue	1	0	1	1	0	0	1	1	0	1	0	0	0	1	0	1	0	1	0	0	0
EWPro	1	0	0	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
FW237	1	0	0	1	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
FWGU4	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
FWLoc1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
FWLoc7	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	1	0	0	0
PW055	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0	1	1	0	0
PWTM58	0	0	0	0	1	0	1	1	0	1	0	0	1	0	0	1	1	1	1	0	0
PWTM59	0	0	0	1	1	0	1	1	0	1	0	0	1	0	0	1	1	1	1	0	0
TCas118	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	0
TCas309	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0	1	1	1	0
TCas531	0	1	0	1	0	1	1	0	0	1	0	0	0	1	0	1	0	1	0	1	1
TCas532	0	1	0	0	1	1	1	0	0	1	1	0	0	0	0	1	1	1	1	1	0
TCas533	1	1	0	1	1	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	0
TCas536	1	1	0	1	1	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	0
TCas538	1	1	0	1	1	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	0
TCas556	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
TCas557	0	1	0	1	0	1	1	0	0	1	0	0	0	1	0	1	0	1	1	1	1
TCch350	0	1	0	0	1	0	1	0	1	1	0	0	1	0	0	1	1	1	1	0	0
TCch353	1	1	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0
TCch354	1	1	0	1	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
TCch355	1	1	0	1	1	0	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
TCji550	1	1	0	1	0	0	1	0	0	1	0	1	1	1	1	1	0	1	1	1	0
TCji552	1	1	0	0	1	0	1	0	0	1	0	1	1	1	0	1	1	1	0	0	0
TCji553	1	1	0	1	0	0	1	0	0	1	0	1	1	1	0	1	0	1	1	1	1
TCji554	1	1	0	1	1	0	1	0	0	1	0	1	1	1	0	1	1	1	0	0	0
TCNA06	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	1	0	1	1	1	1
TCNA07	0	1	0	0	0	1	1	0	0	1	0	1	0	0	0	1	0	1	1	1	1
TCNA09	0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0
TCNA10	0	0	0	0	1	1	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
TWNanI	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
TWNanII	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0
TWTM48	1	1	0	1	1	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	0
TWToaI	0	1	0	0	1	1	1	0	0	1	0	0	1	0	0	1	1	1	1	0	0

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

B7 (bp)	5100	5090	4900	4850	4500	4250	4150	4100	4072	3900	3850	3700	3650	3600	3400	3300	3200	3150	3100	3054	2500	2300	1800	1636	506
AC102	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	1	0	1	0	0	1	0	0
AC201	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
AC202	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
AC203	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
AC204	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
AC205	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
AC207L	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
AC207S	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1	0	0	1	0	0
AC208	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
AC209	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1	1	0
AC209BR	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	1	0	0	1	0	0
AC209GS	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	1	0	1	0	0
AC213	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1	1	0
AC214	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	0
AC215	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
AC216	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	0	0
AC220A	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1	0	0	1	0	0
AC222	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	0
AC223	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1	0	1	1	0	0
AC225	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	1	1	0
AC226	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	1	1	0	1	1	0	0
AC227	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0
AC228	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	0	1	0	0
AC230	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	0
AC231	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	1	1	1	0
AC525	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	1	1	0	0	1	0	0
AC526A	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	1	0	0
EC006	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	0
EC032	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	0	1	1	1	1	0	0	1	0	0
EC033G	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EC109	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0
EC120	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EC201	0	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EC205	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EC214	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	0
EC236	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EC250	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	1	1	1	1	1	0	0	1	0	1
EC502	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1	0	0	1	0	0
EC506	0	1	0	0	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	0	0	1	0	0
EC509	0	1	0	0	1	1	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	0
EC510	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1	0	0	1	0	0
EC511	0	1	0	0	1	0	0	0	1	0	1	1	0	0	0	1	1	1	0	1	0	0	1	0	0
EC531	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1	0	1	1	1	1	0	0	1	0	0
EC534	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EC559	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EC560	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	0
EC565	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
EW051	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1
EW203	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0
EW223	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	1	0	0	1	0	1
EW354	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1
EWHue	0	0	1	1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1
EWPro	0	0	1	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	1	0	0	1	0	1
FW237	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	1	0	1	0	1	1	0	1
FWGU4	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0	1	1	0	1
FWLoc1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0	1	0	1	1	0	1
FWLoc7	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	0	1
PW055	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	1	0	1	0	1	1	0	1
PWTM58	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1	0	1
PWTM59	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	1	0	1
TCas118	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	0
TCas309	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	1	1	0	0
TCas531	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	1	1	0	0
TCas532	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	1	0	0
TCas533	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	0
TCas536	0	0	0	0	0	1	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0
TCas538	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	0	1	1	0
TCas556	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	1	0	1
TCas557	1	0	0	1	0	1	0	0	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	0	0
TCch350	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	1	1	1	1	0	0	1	1	0
TCch353	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0	1	1	0	0
TCch354	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	0	1	1	0	0
TCch355	0	0	0	0	1	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	0	0
TCji550	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	1	1	1	1	0	0	1	1	0
TCji552	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	1	0	1	1	0
TCji553	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1	0	0
TCji554	0	1	0	0	1	1	0	0	1	0	0	1	0	1	0	1	0	1	0	1	0	0	1	0	0
TCNA06	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	1	1	0	0
TCNA07	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	1	1	0	0
TCNA09	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	0	1	1	1	0
TCNA10	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	1	1	1	0	1	1	1	1
TWNanI	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	1
TWNanII	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1	1	1
TWTM48	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	1	1	0	0
TWToaI	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	1	1	0	0

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

B8 (bp)	5090	4850	4700	4500	4300	4150	4072	3800	3750	3700	3600	3550	3500	3350	3250	3200	3150	3100	2700	2500	2300	1800	1636	1100	1018	800	506
AC102	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
AC201	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0
AC202	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0
AC203	0	1	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
AC204	0	1	0	0	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0
AC205	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
AC207L	0	1	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
AC207S	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1
AC208	0	1	0	1	1	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0
AC209	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
AC209BR	0	1	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0
AC209GS	0	1	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	1	0	0	0	0
AC213	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
AC214	0	1	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0
AC215	0	1	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0
AC216	0	1	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
AC220A	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0
AC222	0	1	1	1	1	0	0	0	0	0	1	0	1	1	0	1	0	1	0	0	1	1	1	0	0	1	0
AC223	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1	0
AC225	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
AC226	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0
AC227	0	1	0	1	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0
AC228	0	1	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0	0
AC230	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0	0
AC231	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0	0
AC525	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
AC526A	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0
EC006	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0
EC032	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0
EC033G	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
EC109	0	1	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0
EC120	0	1	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0
EC201	0	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
EC205	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
EC214	0	1	0	0	0	1	0	0	0	0	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0
EC236	0	1	1	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
EC250	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0
EC502	0	1	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0
EC506	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0
EC509	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0
EC510	0	1	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0
EC511	0	1	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0
EC531	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0
EC534	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
EC559	0	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0
EC560	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0
EC565	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0
EW051	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0
EW203	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0	0	0	0
EW223	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0
EW354	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0
EWHue	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0
EWPro	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0
FW237	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0
FWGU4	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0
FWLoc1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0
FWLoc7	0	1	0	0	0	1	1	0	0	0	1	1	0	1	0	0	0	1	1	0	1	0	0	1	0	0	0
PW055	0	1	0	0	0	1	1	0	0	1	1	1	0	1	0	0	1	1	0	0	1	0	0	0	1	0	0
PWTM58	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	0	0
PWTM59	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	0	0	1	0	0
TCas118	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
TCas309	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0
TCas531	0	1	1	1	0	1	0	1	1	0	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0
TCas532	0	1	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
TCas533	0	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0
TCas536	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	0
TCas538	0	1	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	0
TCas556	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
TCas557	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	1	0	0	0	0	0
TCch350	0	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
TCch353	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	0
TCch354	0	1	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	0
TCch355	0	1	1	0	0	0	1	0	0	1	1	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	0
TCji550	0	1	0	1	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0
TCji552	0	1	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0
TCji553	0	1	0	1	1	0	0	1	1	0	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0
TCji554	0	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0
TCNA06	1	1	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0
TCNA07	0	1	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0	0
TCNA09	0	1	0	0	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0
TCNA10	0	1	1	0	0	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0
TWNanI	0	1	0	0	0	0	1	1	0	1	1	1	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0
TWNanII	0	1	0	0	0	0	1	1	0	1	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0	0	0
TWTM48	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
TWToaI	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

H2 (bp)	5700	5500	4800	4600	4200	4100	4072	3800	3700	3600	3500	3300	3200	3150	3100	3054	2500	2400	2054	1915
AC102	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
AC201	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
AC202	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
AC203	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
AC204	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1
AC205	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1
AC207L	0	0	1	1	0	1	1	0	1	0	1	1	0	1	0	0	0	0	1	1
AC207S	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
AC208	0	0	1	1	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	1
AC209	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1
AC209BR	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	1	0	0	0	1
AC209GS	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	1	0	0	0	1
AC213	1	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	1	0	0	1
AC214	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
AC215	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
AC216	0	0	1	1	0	1	1	0	1	0	1	1	0	1	0	0	0	0	0	1
AC220A	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
AC222	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
AC223	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
AC225	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1
AC226	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
AC227	0	0	1	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1
AC228	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
AC230	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1
AC231	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1
AC525	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1
AC526A	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1
EC006	1	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0
EC032	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0
EC033G	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
EC109	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
EC120	0	0	1	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
EC201	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
EC205	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
EC214	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EC236	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
EC250	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0
EC502	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EC506	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EC509	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0
EC510	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EC511	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EC531	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
EC534	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
EC559	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
EC560	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0
EC565	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0
EW051	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EW203	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EW223	1	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0
EW354	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EWHue	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
EWPro	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
FW237	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
FWGU4	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
FWLoc1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0
FWLoc7	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
PW055	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	1
PWTM58	1	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	1	0	1
PWTM59	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0	1
TCas118	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1
TCas309	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1
TCas531	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
TCas532	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1
TCas533	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1
TCas536	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1
TCas538	1	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	1
TCas556	1	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	1	0	1
TCas557	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
TCch350	1	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1
TCch353	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
TCch354	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1
TCch355	1	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1
TCji550	0	0	1	1	1	0	0	0	1	0	1	0	1	1	1	0	0	0	0	1
TCji552	1	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	1
TCji553	0	0	1	1	1	0	0	0	1	0	1	0	0	1	1	0	0	0	0	1
TCji554	1	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	1
TCNA06	0	0	1	1	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1
TCNA07	0	0	1	1	0	0	0	0	1	0	1	0	1	0	1	0	1	1	0	1
TCNA09	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
TCNA10	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0	1	0	1
TWNanI	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1
TWNanII	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1
TWTM48	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
TWTomI	1	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	1

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

H3 (bp)	4900	4700	4600	4500	4077	3900	3750	3650	3450	3200	3150	3100	2954	2500	2036	1800	1636	1300	1018
AC102	0	0	1	1	1	1	0	0	1	1	0	0	0	1	0	0	1	0	1
AC201	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
AC202	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
AC203	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1
AC204	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1
AC205	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0	1
AC207L	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
AC207S	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1
AC208	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
AC209	0	0	0	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0	1
AC209BR	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	1	1
AC209GS	0	0	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	1
AC213	0	0	0	1	1	0	0	0	1	1	0	1	0	1	0	0	0	0	1
AC214	0	0	1	1	1	1	0	0	1	1	1	0	0	0	0	0	1	0	1
AC215	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
AC216	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
AC220A	0	0	0	1	1	1	0	0	1	1	0	0	0	1	0	0	1	0	1
AC222	0	0	1	1	1	1	0	0	1	1	1	0	1	0	0	0	1	0	1
AC223	0	0	0	1	1	1	0	0	1	1	0	0	0	1	0	0	1	0	1
AC225	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1
AC226	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1
AC227	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	1	1
AC228	0	0	1	1	1	1	0	0	1	1	1	0	0	0	0	0	1	0	1
AC230	0	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	1	0	1
AC231	0	0	0	1	1	1	0	1	1	1	0	1	0	1	0	0	1	0	1
AC525	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0	1
AC526A	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1
EC006	0	1	0	1	1	1	1	1	1	0	0	1	0	1	0	0	1	0	1
EC032	0	1	0	1	1	0	1	1	1	0	0	1	0	1	0	0	1	0	1
EC033G	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0	0	1
EC109	0	1	0	1	1	1	1	0	1	1	0	1	0	0	0	0	0	0	1
EC120	0	1	0	1	1	1	1	0	1	1	0	0	1	0	0	1	1	0	1
EC201	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	1
EC205	0	1	0	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	1
EC214	0	1	0	1	1	1	1	1	1	0	0	1	0	1	0	0	1	0	1
EC236	0	1	0	1	1	0	1	0	1	1	0	1	0	0	0	0	0	0	1
EC250	0	1	0	0	1	0	0	1	1	0	0	1	0	0	0	1	0	0	1
EC502	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1
EC506	0	1	0	1	1	1	1	1	1	0	0	1	0	0	0	0	1	0	1
EC509	0	1	0	1	1	0	1	1	1	0	0	0	0	0	0	0	1	0	1
EC510	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1
EC511	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1
EC531	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	0	1	0	1
EC534	0	1	0	0	1	0	1	0	1	1	0	0	1	0	0	0	0	0	1
EC559	0	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	1
EC560	0	1	0	1	1	1	1	1	1	0	0	1	0	1	0	0	1	0	1
EC565	0	1	0	0	1	1	1	0	1	1	0	1	0	0	0	0	1	0	1
EW051	0	1	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	1
EW203	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1
EW223	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	1	1	0	1
EW354	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	0	1
EWHue	0	1	0	0	1	0	0	0	1	1	0	1	0	0	0	1	0	0	1
EWPro	0	1	0	0	1	0	0	0	1	1	0	1	0	0	0	1	1	0	1
FW237	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1
FWGU4	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1
FWLoc1	0	1	0	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1
FWLoc7	0	1	0	0	1	0	0	1	1	1	0	0	0	1	0	0	0	0	1
PW055	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1
PWTM58	0	0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	1	0	1
PWTM59	0	0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	1	0	1
TCas118	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1
TCas309	0	0	0	1	1	1	0	0	1	0	0	1	0	1	0	0	1	0	1
TCas531	1	0	1	1	1	1	0	0	1	1	0	1	1	0	1	0	0	0	1
TCas532	0	0	0	1	1	1	0	1	1	0	0	0	0	1	0	0	1	0	1
TCas533	0	0	1	1	1	1	0	1	1	1	0	1	0	1	0	0	1	0	1
TCas536	0	0	1	0	1	1	0	1	1	0	0	1	0	1	0	0	1	0	1
TCas538	0	0	1	1	1	1	0	1	1	0	0	1	0	1	0	0	1	0	1
TCas556	0	0	1	1	1	1	0	1	1	1	0	0	0	1	0	0	0	0	1
TCas557	1	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1
TCch350	0	0	1	0	1	0	1	1	1	0	0	1	0	1	0	0	0	0	1
TCch353	0	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	1
TCch354	0	0	1	0	1	1	0	1	1	0	0	1	0	1	0	0	1	0	1
TCch355	0	0	1	0	1	1	0	1	1	0	0	1	0	1	0	0	1	0	1
TCji550	1	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1
TCji552	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	1
TCji553	1	0	1	1	1	1	0	0	1	1	1	0	0	0	0	0	1	0	1
TCji554	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	1
TCNA06	0	0	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	1
TCNA07	0	0	0	1	1	1	0	0	1	1	0	0	0	0	1	0	0	0	1
TCNA09	0	0	0	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1
TCNA10	0	0	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	1
TWNanI	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
TWNanII	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
TWTM48	0	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	1
TWToaI	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

H5 (bp)	6300	5700	4800	4500	4300	4150	4072	3900	3750	3600	3400	3250	3200	3100	2954	2700	2016	1850	1618
AC102	1	1	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC201	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	0	0	1	0
AC202	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0
AC203	1	1	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0	1	0
AC204	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	1	0
AC205	1	1	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0	1	0
AC207L	1	1	1	0	0	1	1	1	0	0	1	1	0	0	0	0	0	1	0
AC207S	1	1	1	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC208	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0
AC209	1	1	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC209BR	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0
AC209GS	1	1	1	1	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0
AC213	1	1	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC214	1	1	0	1	0	1	1	1	1	1	1	1	1	0	1	0	1	1	0
AC215	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	0	0	1	0
AC216	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0	0	0	1	0
AC220A	1	1	1	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC222	1	1	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	0
AC223	1	1	1	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC225	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0
AC226	1	1	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC227	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	1	0
AC228	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	1	1	0
AC230	1	1	0	0	0	1	1	0	1	1	1	1	0	0	0	1	1	1	0
AC231	1	1	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	0
AC525	1	1	0	0	0	1	1	1	0	0	1	1	0	1	0	0	0	1	0
AC526A	1	1	0	0	0	1	1	0	1	1	1	0	0	1	0	1	1	1	0
EC006	0	0	0	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
EC032	0	0	0	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
EC033G	0	0	0	1	0	1	1	0	0	0	1	1	1	0	0	0	1	0	0
EC109	0	0	0	1	0	1	1	1	1	0	1	1	1	1	0	0	1	0	0
EC120	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1	0	0
EC201	0	0	0	1	0	1	1	1	0	0	1	1	1	0	0	0	1	0	0
EC205	0	0	0	1	0	1	1	1	0	0	1	1	1	0	0	0	1	0	0
EC214	0	0	0	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
EC236	0	0	0	1	0	1	1	1	1	0	1	1	1	1	0	0	1	0	0
EC250	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EC502	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EC506	0	0	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
EC509	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EC510	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EC511	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EC531	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EC534	0	0	0	1	0	1	1	0	0	0	1	1	1	0	0	0	1	0	0
EC559	0	0	0	1	0	1	1	1	1	0	1	1	1	1	0	0	1	0	0
EC560	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EC565	0	0	0	1	0	1	1	1	1	0	1	1	1	1	0	0	1	0	0
EW051	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EW203	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EW223	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0
EW354	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
EWHue	0	0	0	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0
EWPro	0	0	0	0	0	1	1	0	1	0	1	1	1	1	0	0	0	0	0
FW237	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1
FWGU4	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	1
FWLoc1	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1
FWLoc7	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1
PW055	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	1
PWTM58	0	0	0	0	1	1	1	0	0	0	1	0	0	1	0	1	0	1	1
PWTM59	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	1
TCas118	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TCas309	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TCas531	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	1	1	1	0
TCas532	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TCas533	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0
TCas536	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0
TCas538	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0
TCas556	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	0
TCas557	0	0	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	0	0
TCch350	0	0	0	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
TCch353	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TCch354	0	0	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	1	0
TCch355	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0
TCJi550	0	0	0	1	0	1	1	1	1	1	1	1	1	1	0	1	1	0	0
TCJi552	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TCJi553	0	0	0	1	0	1	1	1	0	1	1	1	1	1	0	0	1	1	0
TCJi554	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TCNA06	0	0	0	0	1	1	1	1	0	1	1	1	1	1	0	0	1	1	0
TCNA07	0	0	0	0	0	1	1	0	0	1	1	1	1	1	0	0	1	1	0
TCNA09	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	0
TCNA10	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	1	0
TWNanI	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TWNanII	0	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0
TWTM48	0	0	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
TWToaI	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	1	0

Appendix 6. Binary data matrix showing presence (1) and absence (0) of the RAPD fragments identified in this study. Primers and band size of each RAPD are indicated.

H14 (bp)	4500	4300	4150	4100	4072	3800	3700	3650	3600	3550	3400	3300	3150	3100	2400	2036
AC102	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0
AC201	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
AC202	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
AC203	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0
AC204	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
AC205	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
AC207L	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
AC207S	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0
AC208	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
AC209	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
AC209BR	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0
AC209GS	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0
AC213	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
AC214	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0	0
AC215	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
AC216	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
AC220A	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
AC222	1	1	0	0	1	1	0	1	0	1	1	0	0	0	0	0
AC223	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0
AC225	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
AC226	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0
AC227	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0
AC228	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0	0
AC230	0	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0
AC231	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
AC525	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
AC526A	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
EC006	0	1	1	0	0	1	1	0	1	0	1	1	1	0	0	0
EC032	0	1	0	0	0	1	1	0	1	0	1	0	0	0	0	0
EC033G	0	1	0	0	0	1	1	0	1	0	1	0	0	0	1	1
EC109	0	1	0	0	0	1	1	0	1	0	1	0	0	0	0	0
EC120	0	1	0	0	0	1	0	0	1	0	1	0	0	0	1	1
EC201	0	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0
EC205	0	1	1	0	1	1	1	0	1	0	1	0	0	0	1	0
EC214	0	1	1	0	0	1	1	0	1	0	1	0	1	0	0	0
EC236	0	1	0	0	0	1	1	0	1	1	1	1	0	0	1	0
EC250	0	1	1	0	0	1	1	0	1	0	1	1	0	0	0	0
EC502	0	1	0	0	1	1	1	0	1	0	1	0	0	0	1	0
EC506	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0	0
EC509	0	1	0	0	0	1	1	0	1	0	1	1	1	0	0	0
EC510	0	1	0	0	1	1	1	0	1	0	1	0	0	0	1	0
EC511	0	1	0	0	1	1	1	0	1	0	1	0	1	0	1	0
EC531	0	1	1	0	0	1	1	0	1	0	1	1	1	0	0	0
EC534	0	1	1	0	1	1	1	0	1	0	1	0	0	0	1	0
EC559	0	1	0	0	0	1	1	0	1	0	1	0	0	0	1	0
EC560	0	1	1	0	0	1	1	0	1	0	1	0	0	0	1	0
EC565	0	1	0	0	0	1	1	0	1	0	1	0	0	0	0	0
EW051	0	1	0	0	0	1	0	0	1	0	1	1	0	0	1	0
EW203	0	1	0	1	0	1	1	0	1	0	1	1	0	0	0	0
EW223	0	1	0	0	0	1	0	0	1	0	1	1	0	0	1	0
EW354	0	1	0	0	0	1	0	0	1	0	1	1	0	0	0	0
EWHue	0	1	0	0	0	1	1	0	1	0	1	1	0	0	1	0
EWPro	0	1	0	0	0	1	0	0	1	0	1	1	0	0	0	0
FW237	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1
FWGU4	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1
FWLoc1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1
FWLoc7	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	1
PW055	1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	1
PWTM58	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	1
PWTM59	1	1	0	1	0	1	1	0	0	0	1	1	0	0	0	1
TCas118	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0
TCas309	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0
TCas531	0	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0
TCas532	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0
TCas533	1	1	1	0	0	1	0	0	0	0	1	1	0	0	0	0
TCas536	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0
TCas538	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0
TCas556	0	1	0	0	0	1	1	0	0	0	1	1	0	0	1	0
TCas557	1	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0
TCch350	0	1	1	0	1	1	1	0	0	0	1	1	0	0	1	0
TCch353	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0
TCch354	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0
TCch355	1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0
TCji550	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0	0
TCji552	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0
TCji553	1	1	0	0	1	1	1	0	0	1	1	0	0	0	0	0
TCji554	1	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0
TCNA06	1	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0
TCNA07	0	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0
TCNA09	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
TCNA10	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0
TWNanI	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0
TWNanII	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0
TWTM48	1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0
TWToaI	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0

Appendix 7. Pairwise genetic distances for 85 *Pachyrhizus* taxa.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	AC102	0													
2	AC201	0.775862	0												
3	AC202	0.786885	0.071429	0											
4	AC203	0.76	0.357143	0.325581	0										
5	AC204	0.744186	0.673913	0.666667	0.555556	0									
6	AC205	0.40625	0.816326	0.826923	0.775	0.71875	0								
7	AC207L	0.762712	0.431373	0.433962	0.55102	0.714286	0.8	0							
8	AC207S	0.466667	0.774194	0.765625	0.781818	0.82	0.526316	0.677966	0						
9	AC208	0.796875	0.212766	0.148936	0.369565	0.686275	0.836364	0.346154	0.776119	0					
10	AC209	0.414634	0.779661	0.770492	0.764706	0.75	0.514286	0.745763	0.510638	0.78125	0				
11	AC209BR	0.727273	0.428571	0.431373	0.583333	0.723404	0.8125	0.537037	0.75	0.403846	0.775862	0			
12	AC209GS	0.732143	0.44	0.411765	0.591837	0.755102	0.84	0.545455	0.733333	0.384615	0.736842	0.25	0		
13	AC213	0.488372	0.8	0.790323	0.764706	0.75	0.514286	0.745763	0.571429	0.761905	0.263158	0.816667	0.758621	0	
14	AC214	0.660714	0.561404	0.559322	0.622642	0.72549	0.734694	0.666667	0.688525	0.603175	0.711864	0.65	0.698413	0.711864	0
15	AC215	0.79661	0.142857	0.159091	0.395349	0.644444	0.816326	0.431373	0.754098	0.25	0.8	0.46	0.44	0.8	0.561404
16	AC216	0.8	0.244444	0.255319	0.444444	0.708333	0.82	0.346939	0.777778	0.191489	0.803279	0.44	0.42	0.783333	0.616667
17	AC220A	0.487179	0.777778	0.767857	0.787234	0.775	0.466667	0.740741	0.315789	0.779661	0.5	0.773585	0.754717	0.571429	0.703704
18	AC222	0.716418	0.569231	0.567164	0.645161	0.711864	0.803279	0.680556	0.77027	0.585714	0.757143	0.626866	0.671429	0.774648	0.298246
19	AC223	0.409091	0.758065	0.75	0.763636	0.77551	0.538462	0.725806	0.214286	0.761194	0.456522	0.754098	0.737705	0.520833	0.672131
20	AC225	0.611111	0.76087	0.77551	0.736842	0.666667	0.576923	0.717391	0.658537	0.788462	0.542857	0.755556	0.787234	0.583333	0.702128
21	AC226	0.439024	0.775862	0.766667	0.76	0.744186	0.542857	0.741379	0.317073	0.777778	0.452381	0.75	0.732143	0.555556	0.684211
22	AC227	0.775862	0.340426	0.3125	0.5	0.702128	0.816326	0.461538	0.774194	0.352941	0.779661	0.326087	0.44	0.819672	0.61017
23	AC228	0.7	0.559322	0.557377	0.592593	0.716981	0.75	0.661538	0.703125	0.6	0.746032	0.645161	0.692308	0.746032	0.145833
24	AC230	0.692308	0.636364	0.631579	0.625	0.651163	0.75	0.737705	0.741379	0.672131	0.698113	0.745763	0.790323	0.722222	0.387755
25	AC231	0.541667	0.694915	0.688525	0.740741	0.75	0.575	0.746032	0.529412	0.723077	0.348837	0.774194	0.737705	0.456522	0.693548
26	AC525	0.674419	0.666667	0.686275	0.625	0.694444	0.666667	0.653061	0.708333	0.703704	0.711111	0.6875	0.72	0.73913	0.583333
27	AC526A	0.714286	0.755102	0.745098	0.7	0.705882	0.677419	0.764706	0.717391	0.759259	0.72093	0.75	0.78	0.75	0.645833
28	EC006	0.642857	0.847222	0.853333	0.861538	0.9	0.811321	0.851351	0.672131	0.873418	0.6	0.861111	0.830986	0.694915	0.808219
29	EC032	0.779661	0.855072	0.861111	0.852459	0.912281	0.865385	0.875	0.758065	0.881579	0.648148	0.885714	0.855072	0.696429	0.797101
30	EC033G	0.916667	0.714286	0.728814	0.74	0.854167	0.938776	0.786885	0.870968	0.761905	0.9	0.816667	0.83871	0.918033	0.733333
31	EC109	0.873016	0.7	0.693548	0.722222	0.890909	0.925926	0.787879	0.830769	0.727273	0.875	0.777778	0.761905	0.857143	0.633333
32	EC120	0.872727	0.745455	0.758621	0.723404	0.844444	0.911111	0.79661	0.864407	0.770492	0.875	0.827586	0.85	0.894737	0.696429
33	EC201	0.919355	0.678571	0.694915	0.72549	0.86	0.941176	0.754098	0.875	0.730159	0.903226	0.783333	0.786885	0.903226	0.677966
34	EC205	0.933333	0.754386	0.745763	0.807692	0.851064	0.9375	0.822581	0.904762	0.777778	0.916667	0.813559	0.816667	0.916667	0.75
35	EC214	0.677966	0.851351	0.842105	0.865672	0.903226	0.818182	0.824324	0.68254	0.8625	0.589286	0.864865	0.819444	0.683333	0.813333
36	EC236	0.888889	0.694915	0.688525	0.692308	0.846154	0.924528	0.765625	0.863636	0.703125	0.890625	0.774194	0.777778	0.854839	0.693548
37	EC250	0.833333	0.888889	0.893939	0.910714	0.980769	0.913043	0.925373	0.807018	0.914286	0.745098	0.921875	0.870968	0.745098	0.878788
38	EC502	0.736842	0.887324	0.876712	0.923077	0.931035	0.82	0.859155	0.649123	0.881579	0.719298	0.885714	0.838235	0.762712	0.814286
39	EC506	0.701754	0.859155	0.849315	0.892308	0.915254	0.826923	0.814286	0.661017	0.87013	0.611111	0.873239	0.826087	0.728814	0.851351
40	EC509	0.741379	0.84058	0.847222	0.836066	0.913793	0.846154	0.876712	0.741935	0.853333	0.678571	0.855072	0.84058	0.724138	0.8
41	EC510	0.754386	0.869565	0.859155	0.904762	0.929825	0.84	0.857143	0.689655	0.864865	0.758621	0.867647	0.818182	0.8	0.794118
42	EC511	0.745763	0.890411	0.88	0.925373	0.933333	0.826923	0.863014	0.637931	0.884615	0.728814	0.888889	0.842857	0.770492	0.819444
43	EC531	0.694915	0.849315	0.84	0.828125	0.919355	0.814815	0.868421	0.738461	0.831169	0.607143	0.863014	0.833333	0.7	0.810811
44	EC534	0.951613	0.758621	0.75	0.788462	0.854167	0.96	0.806452	0.90625	0.78125	0.935484	0.816667	0.819672	0.935484	0.793651
45	EC559	0.904762	0.666667	0.661017	0.686275	0.843137	0.943396	0.761905	0.861538	0.698413	0.888889	0.75	0.754098	0.888889	0.644068
46	EC560	0.701754	0.859155	0.864865	0.857143	0.933333	0.826923	0.878378	0.725806	0.855263	0.584906	0.857143	0.826087	0.660714	0.802817
47	EC565	0.883333	0.745763	0.737705	0.75	0.924528	0.92	0.793651	0.83871	0.75	0.885246	0.803279	0.786885	0.866667	0.631579
48	EW051	0.75	0.912281	0.898305	0.895833	0.977273	0.8	0.915254	0.77551	0.903226	0.808511	0.929825	0.872727	0.808511	0.881356
49	EW203	0.72	0.873016	0.861538	0.912281	0.941176	0.785714	0.876923	0.62	0.884058	0.72549	0.888889	0.836066	0.773585	0.809524
50	EW223	0.811321	0.888889	0.876923	0.87037	0.960784	0.837209	0.892308	0.785714	0.882353	0.792453	0.904762	0.888889	0.836364	0.878788
51	EW354	0.825	0.941176	0.944444	0.953488	0.972222	0.866667	0.962963	0.844444	0.965517	0.857143	0.960784	0.92	0.883721	0.90566
52	EWHue	0.949153	0.810345	0.8	0.77551	0.938776	0.979167	0.890625	0.901639	0.809524	0.932203	0.847458	0.85	0.913793	0.822581
53	EWPro	0.888889	0.824561	0.813559	0.791667	0.913043	0.955556	0.887097	0.859649	0.84127	0.910714	0.862069	0.844828	0.890909	0.79661
54	FW237	0.836735	0.95082	0.953125	0.942308	1	0.871795	0.935484	0.830189	0.939394	0.884615	0.95	0.933333	0.90566	0.969697
55	FWGU4	0.860465	0.943396	0.946429	0.931818	1	0.941176	0.964286	0.961538	0.949153	0.93617	0.981481	0.962963	0.913043	0.947368
56	FWLoc1	0.897959	0.929825	0.915254	0.916667	0.926829	0.891892	0.913793	0.886792	0.901639	0.854167	0.965517	0.929825	0.804348	0.951613
57	FWLoc7	0.836364	0.907692	0.895522	0.912281	0.941176	0.914894	0.876923	0.810345	0.867647	0.773585	0.923077	0.890625	0.796296	0.928571
58	PW055	0.792453	0.854839	0.861538	0.851852	0.875	0.866667	0.84127	0.789474	0.833333	0.796296	0.870968	0.854839	0.818182	0.897059
59	PWTM58	0.660377	0.835821	0.842857	0.830508	0.890909	0.73913	0.80597	0.642857	0.833333	0.588235	0.850746	0.852941	0.641509	0.776119
60	PWTM59	0.685185	0.835821	0.842857	0.830508	0.910714	0.765957	0.823529	0.689655	0.849315	0.641509	0.867647	0.852941	0.690909	0.794118
61	TC118as	0.702703	0.9	0.90566	0.878049	0.848485	0.793103	0.882353	0.795455	0.910714	0.710526	0.875	0.854167	0.775	0.886792
62	TC309as	0.627907	0.862069	0.85	0.86	0.860465	0.722222	0.827586	0.577778	0.857143	0.571429	0.859649	0.8	0.636364	0.77193
63	TC531as	0.808824	0.686567	0.7	0.725806	0.816667	0.866667	0.75	0.805556	0.712329	0.794118	0.757143	0.760563	0.811594	0.466667
64	TC532as	0.512195	0.847458	0.836066	0.843137	0.866667	0.702703	0.813559	0.653061	0.84375	0.622222	0.844828	0.807018	0.708333	0.8
65	TC533as	0.54902	0.753846	0.764706	0.8	0.857143	0.702128	0.761194	0.589286	0.757143	0.557692	0.730159	0.693548	0.557692	0.712121
66	TC536as	0.618182	0.797101	0.805556	0.825397	0.9	0.788462	0.819444	0.627119	0.797297	0.574074	0.776119	0.742424	0.649123	0.720588
67	TC538as	0.509804	0.797101	0.805556	0.825397	0.881356	0.74	0.802817	0.578947	0.813333	0.574074	0.738461	0.723077	0.649123	0.720588
68	TC556as	0.586957	0.819672	0.790323	0.833333	0.854167	0.761905	0.806452	0.679245	0.78125	0.595745	0.836066	0.779661	0.625	0.793651
69	TC557as	0.78125	0.629032	0.646154	0.689655	0.785714	0.87931	0.73913	0.779412	0.681159	0.80303	0.727273	0.712121	0.80303	0.438596
70	TC350ch	0.615385	0.855072	0.84507	0.870968	0.912281	0.795918	0.826087	0.649123	0.851351	0.510204	0.852941	0.784615	0.596154	0.814286
71	TC353ch	0.590909	0.85	0.819672	0.823529	0.893617	0.710526	0.816667	0.574468	0.809524	0.6875	0.847458	0.767857	0.714286	0.783333
72	TC354ch	0.54902	0.808824	0.816901	0.83871	0.877193	0.755102	0.779412	0.537037	0.808219	0.557692	0.769231	0.753846	0.636364	0.75
73	TC355ch	0.54717	0.833333	0.84	0.863636	0.901639	0.769231	0.788732	0.61017	0.815789	0.607143	0.779412	0.764706	0.655172	0.777778
74	TC550jl	0.780822	0.647887	0.643836	0.701493	0.838235	0.882353	0.710526	0.763158	0.657895	0.783784	0.716216	0.72	0.8	0.4375
75	TC552jl	0.647059	0.830769	0.820896	0.803571	0.886792	0.697674	0.78125	0.62963	0.811594	0.727273	0.880597	0.830769	0.727273	0.75
76	TC553jl	0.768116	0.606061	0.623188	0.703125	0.809524	0.875	0.676056	0.767123	0.638889	0.805556	0.71831	0.722222	0.821918	0.372881
77	TC554jl	0.654545	0.772727	0.764706	0.758621	0.857143	0.729167	0.723077	0.537037	0.73913	0.728814	0.823529	0.772727	0.728814	0.712121
78	TCNA06	0.8	0.692308	0.705882	0.711864	0.807018	0.87931	0.774648	0.779412	0.736111	0.765625	0.8	0.785714	0.784615	0.438596
79	TCNA07	0.806452	0.772727	0.764706	0.758621	0.836364	0.890909	0.814286	0.784615	0.774648	0.770492	0.84058	0.826087	0.770492	0.583333
80	TCNA09	0.595238	0.862069	0.830508	0.8125	0.74359	0.722222	0.827586	0.693878	0.8	0.571429	0.839286	0.8	0.636364	0.793103
81	TCNA10	0.666667	0.854839	0.84375	0.851852	0.826087	0.785714	0.876923	0.745455	0.833333	0.7	0.870968	0.854839	0.72549	0.846154
82	TWNanI	0.782609	0.894737	0.881356	0.875	0.878049	0.868421	0.898305	0.849057	0.868852	0.8125	0.931035	0.894737	0.787234	0.936508
83	TWNanII	0.723404	0.901639	0.888889	0.884615	0.888889	0.825	0.904762	0.818182	0.859375	0.78	0.935484	0.901639	0.78	0.907692
84	TWTM48	0.622222	0.722222	0.736842	0.77551	0.844444	0.74359	0.732143	0.66	0.728814	0.6	0.716981	0.698113	0.6875	0.762712
85	TWTeal	0.613636	0.903226	0.873016	0.865385	0.840909	0.769231	0.870968	0.705882	0.84375	0.652174	0.864407	0.847458	0.622222	0.857143
86		1	2	3	4	5	6	7	8	9	10	11	12	13	14

Appendix 7. Pairwise genetic distances for 85 *Pachyrhizus* taxa.

	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15	0														
16	0.162791	0													
17	0.754717	0.781818	0												
18	0.590909	0.657143	0.772727	0											
19	0.758065	0.78125	0.333333	0.739726	0										
20	0.787234	0.791667	0.636364	0.758621	0.666667	0									
21	0.79661	0.8	0.487179	0.753623	0.205128	0.571429	0								
22	0.340426	0.387755	0.8	0.590909	0.777778	0.733333	0.775862	0							
23	0.559322	0.634921	0.719298	0.275862	0.6875	0.72	0.7	0.606557	0						
24	0.636364	0.711864	0.714286	0.363636	0.724138	0.65	0.716981	0.684211	0.392157	0					
25	0.694915	0.761905	0.488372	0.722222	0.48	0.634146	0.510638	0.716667	0.6875	0.62963	0				
26	0.666667	0.673469	0.666667	0.661017	0.714286	0.612903	0.674419	0.666667	0.58	0.613636	0.764706	0			
27	0.729167	0.76	0.638889	0.689655	0.723404	0.571429	0.682927	0.755102	0.64	0.55	0.723404	0.575758	0		
28	0.863014	0.88	0.754386	0.833333	0.655738	0.830189	0.689655	0.863014	0.831169	0.80597	0.633333	0.85	0.881356	0	
29	0.855072	0.888889	0.824561	0.825	0.761905	0.862745	0.779661	0.871429	0.805556	0.793651	0.631579	0.859649	0.872727	0.444444	0
30	0.736842	0.783333	0.909091	0.774648	0.890625	0.891304	0.898305	0.758621	0.725806	0.789474	0.890625	0.791667	0.829787	0.777778	0.741379
31	0.7	0.746032	0.842105	0.708333	0.833333	0.903846	0.873016	0.741935	0.650794	0.75	0.833333	0.769231	0.803922	0.742424	0.637931
32	0.745455	0.793103	0.86	0.707692	0.847458	0.829268	0.872727	0.789474	0.689655	0.705882	0.827586	0.777778	0.790698	0.786885	0.727273
33	0.678571	0.728814	0.912281	0.728571	0.893939	0.895833	0.901639	0.724138	0.672131	0.754386	0.859375	0.77551	0.86	0.765625	0.636364
34	0.775862	0.819672	0.927273	0.753623	0.90625	0.913043	0.896552	0.754386	0.741935	0.847458	0.90625	0.8125	0.92	0.793651	0.779661
35	0.866667	0.883117	0.741379	0.837209	0.666667	0.836364	0.7	0.866667	0.820513	0.811594	0.622951	0.836066	0.866667	0.2	0.346154
36	0.694915	0.741935	0.898305	0.722222	0.865672	0.923077	0.888889	0.716667	0.6875	0.766667	0.865672	0.764706	0.846154	0.757576	0.655172
37	0.888889	0.907692	0.884615	0.924051	0.830508	0.934783	0.811321	0.90625	0.884058	0.901639	0.789474	0.923077	0.94	0.584906	0.456522
38	0.871429	0.888889	0.735849	0.853659	0.7	0.884615	0.736842	0.887324	0.837838	0.865672	0.721311	0.898305	0.892857	0.472727	0.461538
39	0.875	0.891892	0.722222	0.857143	0.666667	0.846154	0.701754	0.875	0.857143	0.835821	0.688525	0.844828	0.877193	0.313725	0.423077
40	0.857143	0.859155	0.827586	0.841463	0.746032	0.865385	0.719298	0.857143	0.808219	0.815385	0.683333	0.862069	0.894737	0.396226	0.23913
41	0.852941	0.871429	0.730769	0.8375	0.737705	0.882353	0.775862	0.869565	0.819444	0.846154	0.737705	0.896552	0.890909	0.517857	0.480769
42	0.875	0.891892	0.722222	0.857143	0.688525	0.888889	0.724138	0.890411	0.842105	0.869565	0.730159	0.901639	0.896552	0.464286	0.481481
43	0.849315	0.835616	0.8	0.848837	0.703125	0.833333	0.716667	0.849315	0.818182	0.826087	0.66129	0.870968	0.883333	0.320755	0.365385
44	0.779661	0.822581	0.947368	0.791667	0.924242	0.9375	0.916667	0.758621	0.765625	0.868852	0.924242	0.84	0.921569	0.796875	0.762712
45	0.642857	0.716667	0.877193	0.7	0.863636	0.9	0.887097	0.689655	0.639344	0.719298	0.828125	0.76	0.795918	0.772727	0.649123
46	0.875	0.861111	0.767857	0.843373	0.709677	0.823529	0.701754	0.875	0.810811	0.818182	0.62069	0.864407	0.877193	0.377358	0.255319
47	0.724138	0.75	0.851852	0.710145	0.84127	0.895833	0.883333	0.766667	0.65	0.754386	0.84127	0.75	0.787234	0.765625	0.636364
48	0.912281	0.913793	0.837209	0.915493	0.78	0.891892	0.75	0.912281	0.887097	0.886792	0.755102	0.883721	0.928571	0.686275	0.74
49	0.854839	0.892308	0.688889	0.868421	0.653846	0.863636	0.666667	0.873016	0.818182	0.847458	0.679245	0.882353	0.875	0.509804	0.530612
50	0.852459	0.873016	0.84	0.896104	0.810345	0.886364	0.833333	0.870968	0.884058	0.864407	0.789474	0.901961	0.918367	0.660714	0.58
51	0.941176	0.962264	0.894737	0.938462	0.847826	0.935484	0.825	0.92	0.910714	0.914894	0.847826	0.888889	0.942857	0.744681	0.75
52	0.810345	0.833333	0.944444	0.864865	0.903226	0.956522	0.931035	0.830508	0.830769	0.862069	0.903226	0.921569	0.916667	0.825397	0.727273
53	0.803571	0.847458	0.877551	0.828571	0.862069	0.930233	0.888889	0.844828	0.786885	0.792453	0.842105	0.847826	0.837209	0.779661	0.692308
54	0.95082	0.934426	0.869565	0.974026	0.811321	0.925	0.787234	0.95082	0.955882	0.966102	0.854545	0.913043	0.906977	0.810345	0.796296
55	0.943396	0.944444	0.952381	0.955882	0.9	0.970588	0.934783	0.962963	0.967213	0.94	0.9	0.97561	0.945946	0.803922	0.882353
56	0.910714	0.912281	0.860465	0.958904	0.823529	0.947368	0.875	0.948276	0.953846	0.945455	0.846154	0.931818	0.926829	0.842105	0.830189
57	0.890625	0.892308	0.795918	0.925	0.77193	0.913043	0.792453	0.924242	0.931507	0.903226	0.793103	0.903846	0.92	0.754098	0.8
58	0.854839	0.857143	0.865385	0.883117	0.75	0.837209	0.72	0.854839	0.885714	0.866667	0.793103	0.882353	0.851064	0.8125	0.83871
59	0.835821	0.838235	0.705882	0.807692	0.625	0.76087	0.634615	0.852941	0.802817	0.770492	0.649123	0.814815	0.755102	0.66129	0.661017
60	0.835821	0.855072	0.754717	0.822785	0.649123	0.787234	0.660377	0.852941	0.802817	0.790323	0.672414	0.814815	0.755102	0.68254	0.661017
61	0.921569	0.923077	0.805556	0.888889	0.772727	0.785714	0.736842	0.9	0.872727	0.869565	0.8	0.861111	0.882353	0.867925	0.88
62	0.881356	0.883333	0.682927	0.808824	0.452381	0.714286	0.361111	0.862069	0.762712	0.788462	0.617021	0.790698	0.804878	0.685185	0.735849
63	0.686567	0.728571	0.8125	0.485294	0.791667	0.844828	0.808824	0.724638	0.467742	0.559322	0.774648	0.701754	0.754386	0.8375	0.780822
64	0.866667	0.868852	0.634146	0.814286	0.66	0.794872	0.673913	0.866667	0.809524	0.796296	0.632653	0.851064	0.840909	0.696429	0.767857
65	0.734375	0.757576	0.647059	0.736842	0.571429	0.723404	0.603774	0.753846	0.705882	0.721311	0.545455	0.781818	0.769231	0.590164	0.586207
66	0.797101	0.8	0.685185	0.74359	0.586207	0.784314	0.618182	0.814286	0.714286	0.730159	0.561404	0.810345	0.821429	0.625	0.526316
67	0.797101	0.816901	0.660377	0.74359	0.561404	0.734694	0.592593	0.761194	0.714286	0.75	0.61017	0.789474	0.8	0.646154	0.576271
68	0.819672	0.803279	0.75	0.855263	0.634615	0.756098	0.617021	0.819672	0.820896	0.830508	0.660377	0.862745	0.854167	0.737705	0.803279
69	0.629032	0.69697	0.803279	0.484848	0.746269	0.836364	0.78125	0.692308	0.440678	0.535714	0.764706	0.660377	0.763636	0.815789	0.771429
70	0.871429	0.873239	0.735849	0.839506	0.631579	0.84	0.666667	0.871429	0.837838	0.848485	0.631579	0.859649	0.912281	0.352941	0.644068
71	0.810345	0.813559	0.704545	0.816901	0.612245	0.857143	0.622222	0.85	0.793651	0.842105	0.666667	0.854167	0.818182	0.701754	0.727273
72	0.791045	0.811594	0.647059	0.78481	0.518519	0.77551	0.576923	0.808824	0.760563	0.8	0.571429	0.803571	0.814815	0.542373	0.561404
73	0.816901	0.819444	0.714286	0.792683	0.59322	0.811321	0.649123	0.816901	0.786667	0.808824	0.616667	0.833333	0.844828	0.564516	0.583333
74	0.666667	0.706667	0.816901	0.43662	0.766234	0.790323	0.780822	0.666667	0.415385	0.590909	0.766234	0.677419	0.725806	0.811765	0.74026
75	0.8125	0.815385	0.72	0.786667	0.684211	0.854167	0.698113	0.830769	0.761194	0.84127	0.706897	0.807692	0.843137	0.808824	0.815385
76	0.626866	0.671429	0.823529	0.378788	0.77027	0.816667	0.768116	0.666667	0.377049	0.516667	0.77027	0.677966	0.75	0.817073	0.76
77	0.734375	0.738461	0.62	0.736842	0.596491	0.867925	0.654545	0.826087	0.686567	0.78125	0.688525	0.735849	0.792453	0.735294	0.71875
78	0.712121	0.753623	0.803279	0.507463	0.746269	0.836364	0.761905	0.768116	0.440678	0.561404	0.746269	0.709091	0.785714	0.8	0.753623
79	0.791045	0.828571	0.810345	0.608696	0.75	0.8	0.745763	0.826087	0.533333	0.607143	0.75	0.735849	0.792453	0.853333	0.794118
80	0.842105	0.824561	0.714286	0.826087	0.645833	0.714286	0.560976	0.842105	0.822581	0.788462	0.673469	0.790698	0.775	0.754386	0.803571
81	0.854839	0.857143	0.770833	0.853333	0.703704	0.809524	0.666667	0.836066	0.852941	0.847458	0.703704	0.882353	0.851064	0.793651	0.875
82	0.913793	0.933333	0.866667	0.945946	0.784314	0.894737	0.782609	0.913793	0.923077	0.928571	0.784314	0.934783	0.930233	0.885246	0.967742
83	0.919355	0.920635	0.854167	0.921053	0.754717	0.904762	0.75	0.919355	0.895522	0.896552	0.754717	0.895833	0.93617	0.857143	0.9375
84	0.745455	0.727273	0.704545	0.782609	0.64	0.702703	0.590909	0.722222	0.754098	0.754717	0.666667	0.777778	0.761905	0.745763	0.727273
85	0.885246	0.868852	0.808511	0.893333	0.66	0.794872	0.613636	0.866667	0.880597	0.859649	0.711538	0.851064	0.866667	0.696429	0.767857
86	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29

Appendix 7. Pairwise genetic distances for 85 *Pachyrhizus* taxa.

	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30	0														
31	0.4	0													
32	0.384615	0.5	0												
33	0.3	0.27907	0.452381	0											
34	0.452381	0.520833	0.52381	0.404762	0										
35	0.725806	0.671875	0.774194	0.693548	0.761905	0									
36	0.386364	0.288889	0.521739	0.302326	0.444444	0.707692	0								
37	0.769231	0.727273	0.849057	0.754717	0.811321	0.54717	0.722222	0							
38	0.741379	0.661017	0.75	0.660714	0.714286	0.436364	0.7	0.607843	0						
39	0.728814	0.693548	0.758621	0.716667	0.786885	0.204082	0.709677	0.596154	0.423077	0					
40	0.745763	0.666667	0.754386	0.666667	0.783333	0.358491	0.661017	0.434783	0.411765	0.403846	0				
41	0.736842	0.631579	0.722222	0.62963	0.709091	0.482143	0.672414	0.627451	0.073171	0.471698	0.431373	0			
42	0.75	0.672131	0.758621	0.672414	0.724138	0.428571	0.709677	0.622642	0.047619	0.384615	0.403846	0.116279	0		
43	0.761905	0.727273	0.75	0.709677	0.796875	0.345455	0.723077	0.538462	0.454545	0.418182	0.244898	0.472727	0.446429	0	
44	0.390244	0.56	0.568182	0.454545	0.194444	0.725806	0.456522	0.769231	0.719298	0.75	0.766667	0.714286	0.728814	0.78125	0
45	0.333333	0.190476	0.44186	0.153846	0.466667	0.703125	0.295455	0.763636	0.672414	0.725806	0.677966	0.642857	0.683333	0.71875	0.478261
46	0.728814	0.672131	0.714286	0.672414	0.766667	0.307692	0.688525	0.446809	0.423077	0.415094	0.234043	0.442308	0.444444	0.22449	0.75
47	0.454545	0.15	0.488372	0.292683	0.574468	0.693548	0.340909	0.754717	0.684211	0.716667	0.689655	0.654545	0.694915	0.709677	0.583333
48	0.833333	0.830189	0.767442	0.816326	0.829787	0.698113	0.803922	0.682927	0.659575	0.72549	0.666667	0.652174	0.673469	0.666667	0.857143
49	0.75	0.660377	0.734694	0.735849	0.745098	0.528302	0.727273	0.6	0.363636	0.54902	0.479167	0.386364	0.391304	0.519231	0.773585
50	0.769231	0.793103	0.729167	0.754717	0.833333	0.672414	0.745455	0.55814	0.634615	0.596154	0.588235	0.627451	0.622642	0.509804	0.814815
51	0.829268	0.826087	0.815789	0.863636	0.853659	0.78	0.795455	0.606061	0.777778	0.76087	0.727273	0.772727	0.787234	0.77551	0.857143
52	0.534884	0.466667	0.636364	0.555556	0.680851	0.793651	0.553191	0.673913	0.793103	0.819672	0.754386	0.767857	0.8	0.770492	0.630435
53	0.613636	0.477273	0.651163	0.6	0.77551	0.745763	0.595745	0.595238	0.763636	0.75	0.698113	0.735849	0.77193	0.762712	0.755102
54	0.884615	0.877193	0.9	0.909091	0.944444	0.79661	0.875	0.727273	0.773585	0.735849	0.730769	0.792453	0.759259	0.727273	0.90566
55	0.93617	0.943396	0.906977	0.938776	0.957447	0.854545	0.942308	0.857143	0.903846	0.886792	0.884615	0.901961	0.907407	0.851852	0.958333
56	0.9	0.910714	0.893617	0.903846	0.897959	0.807018	0.909091	0.822222	0.851852	0.857143	0.875	0.849057	0.857143	0.864407	0.921569
57	0.87931	0.890625	0.872727	0.901639	0.896552	0.721311	0.852459	0.74	0.8	0.724138	0.741379	0.79661	0.766667	0.737705	0.898305
58	0.898305	0.924242	0.872727	0.919355	0.915254	0.818182	0.870968	0.833333	0.83871	0.806452	0.803279	0.836066	0.825397	0.777778	0.934426
59	0.875	0.869565	0.830508	0.895522	0.907692	0.692308	0.867647	0.703704	0.746032	0.714286	0.688525	0.761905	0.753846	0.645161	0.925373
60	0.857143	0.852941	0.830508	0.878788	0.907692	0.692308	0.850746	0.703704	0.765625	0.714286	0.688525	0.78125	0.772727	0.645161	0.909091
61	0.934783	0.942308	0.904762	0.9375	0.933333	0.872727	0.92	0.906977	0.833333	0.862745	0.86	0.829787	0.84	0.849057	0.934783
62	0.909091	0.881356	0.882353	0.912281	0.907407	0.696429	0.87931	0.816326	0.759259	0.698113	0.740741	0.8	0.745455	0.690909	0.928571
63	0.698413	0.603175	0.725806	0.666667	0.753846	0.797468	0.661538	0.84058	0.844156	0.833333	0.815789	0.826667	0.848101	0.825	0.757576
64	0.912281	0.885246	0.865385	0.915254	0.929825	0.706897	0.883333	0.867925	0.722222	0.685185	0.75	0.692308	0.732143	0.654545	0.931035
65	0.897059	0.791045	0.857143	0.850746	0.911765	0.557377	0.80597	0.696429	0.698413	0.645161	0.616667	0.714286	0.707692	0.573771	0.897059
66	0.884058	0.797101	0.861538	0.855072	0.944444	0.548387	0.861111	0.684211	0.6875	0.612903	0.583333	0.703125	0.676923	0.564516	0.915493
67	0.884058	0.797101	0.861538	0.871429	0.929577	0.59375	0.828571	0.706897	0.707692	0.634921	0.629032	0.703125	0.69697	0.609375	0.915493
68	0.9	0.892308	0.894737	0.885246	0.898305	0.765625	0.890625	0.769231	0.719298	0.809524	0.766667	0.736842	0.728814	0.677966	0.918033
69	0.661017	0.583333	0.689655	0.603448	0.741935	0.789474	0.622951	0.850746	0.821918	0.810811	0.791667	0.802817	0.826667	0.802632	0.746032
70	0.893939	0.871429	0.90625	0.880597	0.857143	0.464286	0.852941	0.685185	0.596491	0.535714	0.627119	0.637931	0.61017	0.559322	0.84127
71	0.913793	0.868852	0.928571	0.898305	0.949153	0.666667	0.885246	0.78	0.627451	0.690909	0.685185	0.673077	0.641509	0.636364	0.894737
72	0.897059	0.826087	0.892308	0.884058	0.942857	0.533333	0.823529	0.672727	0.655738	0.551724	0.616667	0.672131	0.666667	0.596774	0.897059
73	0.901408	0.833333	0.880597	0.873239	0.945205	0.555556	0.830986	0.733333	0.650794	0.525424	0.612903	0.666667	0.640625	0.548387	0.901408
74	0.695652	0.588235	0.720588	0.647059	0.746479	0.77381	0.680556	0.842105	0.802469	0.807229	0.775	0.78481	0.807229	0.785714	0.732394
75	0.888889	0.848485	0.864407	0.875	0.887097	0.797101	0.846154	0.847458	0.716667	0.784615	0.8	0.733333	0.725806	0.776119	0.852459
76	0.69697	0.606061	0.723077	0.646154	0.75	0.792683	0.681159	0.864865	0.822785	0.82716	0.794872	0.805195	0.82716	0.804878	0.735294
77	0.828125	0.734375	0.819672	0.758065	0.84375	0.666667	0.75	0.741379	0.633333	0.707692	0.68254	0.65	0.645161	0.701493	0.809524
78	0.704918	0.583333	0.711864	0.672131	0.741935	0.789474	0.645161	0.833333	0.821918	0.810811	0.791667	0.802817	0.826667	0.802632	0.765625
79	0.706897	0.714286	0.690909	0.716667	0.701754	0.826667	0.709677	0.859375	0.84507	0.849315	0.814286	0.842857	0.849315	0.824324	0.728814
80	0.928571	0.9	0.903846	0.912281	0.907407	0.741379	0.898305	0.84	0.759259	0.745455	0.740741	0.777778	0.767857	0.736842	0.928571
81	0.951613	0.940298	0.931035	0.953125	0.967742	0.818182	0.923077	0.833333	0.83871	0.825397	0.822581	0.836066	0.825397	0.796875	0.984375
82	0.963636	0.967213	0.961538	0.982759	0.943396	0.90625	0.912281	0.921569	0.933333	0.9	0.934426	0.932203	0.935484	0.904762	0.944444
83	0.966102	0.969231	0.964286	0.983871	0.947368	0.878788	0.918033	0.907407	0.903226	0.870968	0.887097	0.919355	0.90625	0.859375	0.948276
84	0.854545	0.830508	0.846154	0.839286	0.912281	0.711864	0.827586	0.755102	0.75	0.690909	0.685185	0.745455	0.758621	0.660714	0.894737
85	0.949153	0.920635	0.946429	0.933333	0.910714	0.706897	0.883333	0.75	0.789474	0.754386	0.703704	0.807018	0.79661	0.701754	0.912281
86	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44

Appendix 7. Pairwise genetic distances for 85 *Pachyrhizus* taxa.

	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															
31															
32															
33															
34															
35															
36															
37															
38															
39															
40															
41															
42															
43															
44															
45	0														
46	0.683333	0													
47	0.2	0.672414	0												
48	0.8	0.673469	0.816326	0											
49	0.698113	0.52	0.686275	0.589744	0										
50	0.763636	0.568627	0.777778	0.578947	0.6875	0									
51	0.844444	0.76087	0.809524	0.48	0.657143	0.685714	0								
52	0.477273	0.736842	0.488372	0.738095	0.784314	0.673913	0.75	0							
53	0.522727	0.727273	0.463415	0.692308	0.723404	0.659091	0.612903	0.342857	0						
54	0.892857	0.735849	0.867925	0.736842	0.787234	0.634146	0.75	0.854167	0.767442	0					
55	0.941176	0.886792	0.916667	0.787879	0.886364	0.829268	0.814815	0.880952	0.846154	0.645161	0				
56	0.907407	0.857143	0.924528	0.783784	0.826087	0.895833	0.878788	0.893617	0.888889	0.72973	0.555556	0			
57	0.887097	0.745763	0.901639	0.777778	0.792453	0.74	0.825	0.872727	0.823529	0.459459	0.675676	0.578947	0		
58	0.904762	0.806452	0.936508	0.777778	0.814815	0.74	0.825	0.912281	0.888889	0.575	0.833333	0.772727	0.545455	0	
59	0.865672	0.672131	0.861538	0.784314	0.732143	0.653846	0.826087	0.85	0.824561	0.666667	0.808511	0.803922	0.634615	0.456522	0
60	0.848485	0.672131	0.84375	0.784314	0.732143	0.627451	0.826087	0.830508	0.803571	0.608696	0.782609	0.803922	0.607843	0.422222	0.142857
61	0.918367	0.816326	0.9375	0.852941	0.857143	0.906977	0.892857	0.954545	0.952381	0.921053	1	1	0.909091	0.8	0.829787
62	0.896552	0.673077	0.872727	0.717949	0.688889	0.816326	0.764706	0.924528	0.9	0.790698	0.9	0.886364	0.795918	0.744681	0.705882
63	0.634921	0.802632	0.645161	0.909091	0.826087	0.873239	0.933333	0.765625	0.777778	0.942857	0.952381	0.907692	0.890411	0.890411	0.810811
64	0.9	0.709091	0.877193	0.761905	0.702128	0.75	0.871795	0.946429	0.903846	0.851064	0.878049	0.9375	0.826923	0.78	0.716981
65	0.838235	0.526316	0.796875	0.773585	0.701754	0.696429	0.86	0.857143	0.793103	0.803571	0.907407	0.877193	0.766667	0.745763	0.627119
66	0.826087	0.491228	0.784615	0.824561	0.711864	0.706897	0.886792	0.84375	0.8	0.745455	0.910714	0.9	0.733333	0.733333	0.639344
67	0.842857	0.566667	0.80303	0.803571	0.689655	0.684211	0.865385	0.861538	0.819672	0.767857	0.929825	0.918033	0.774194	0.711864	0.616667
68	0.870968	0.706897	0.903226	0.727273	0.773585	0.745098	0.857143	0.875	0.890909	0.765957	0.863636	0.829787	0.72549	0.617021	0.530612
69	0.54386	0.794521	0.603448	0.904762	0.818182	0.867647	0.929825	0.774194	0.766667	0.924242	0.95	0.920635	0.852941	0.852941	0.802817
70	0.884058	0.561404	0.880597	0.764706	0.690909	0.754386	0.877551	0.888889	0.885246	0.859649	0.86	0.872727	0.779661	0.819672	0.661017
71	0.883333	0.666667	0.839286	0.767442	0.680851	0.729167	0.846154	0.909091	0.862745	0.75	0.880952	0.916667	0.830189	0.784314	0.767857
72	0.871429	0.551724	0.815385	0.818182	0.678571	0.719298	0.86	0.875	0.833333	0.781818	0.886792	0.877193	0.766667	0.766667	0.603448
73	0.876712	0.55	0.823529	0.827586	0.737705	0.666667	0.867925	0.880597	0.822581	0.703704	0.872727	0.919355	0.737705	0.694915	0.645161
74	0.597015	0.777778	0.626866	0.90411	0.844156	0.842105	0.925373	0.73913	0.75	0.906667	0.957747	0.946667	0.8875	0.844156	0.769231
75	0.84375	0.80303	0.819672	0.8	0.767857	0.785714	0.869565	0.901639	0.859649	0.76	0.92	0.946429	0.810345	0.767857	0.689655
76	0.615385	0.797468	0.646154	0.898551	0.851351	0.864865	0.920635	0.761194	0.772727	0.931507	0.955224	0.958333	0.924051	0.881579	0.789474
77	0.746032	0.6875	0.737705	0.818182	0.701754	0.783333	0.882353	0.857143	0.793103	0.803571	0.946429	0.857143	0.806452	0.806452	0.734375
78	0.616667	0.777778	0.627119	0.904762	0.8	0.884058	0.910714	0.774194	0.766667	0.940298	0.95	0.920635	0.885714	0.869565	0.768116
79	0.683333	0.782609	0.737705	0.87931	0.825397	0.876923	0.944444	0.819672	0.852459	0.919355	0.964912	0.896552	0.84375	0.806452	0.772727
80	0.915254	0.722222	0.931035	0.837209	0.82	0.862745	0.894737	0.982143	0.942308	0.818182	0.952381	0.860465	0.744681	0.744681	0.754717
81	0.938462	0.825397	0.953125	0.829787	0.836364	0.811321	0.880952	0.966667	0.947368	0.8125	0.833333	0.875	0.72	0.608696	0.685185
82	0.966102	0.952381	0.982759	0.853659	0.923077	0.942308	0.885714	0.941176	0.938776	0.804878	0.794118	0.820513	0.782609	0.634146	0.764706
83	0.968254	0.90625	0.983871	0.813954	0.909091	0.907407	0.868421	0.945455	0.943396	0.738095	0.815789	0.837209	0.75	0.604651	0.735849
84	0.824561	0.666667	0.839286	0.847826	0.784314	0.755102	0.928571	0.867925	0.84	0.804348	0.931818	0.916667	0.807692	0.734694	0.647059
85	0.935484	0.732143	0.933333	0.844444	0.826923	0.823529	0.9	0.946429	0.924528	0.8	0.878049	0.891304	0.755102	0.729167	0.692308
86	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59

Appendix 7. Pairwise genetic distances for 85 *Pachyrhizus* taxa.

	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV	BW
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17															
18															
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															
31															
32															
33															
34															
35															
36															
37															
38															
39															
40															
41															
42															
43															
44															
45															
46															
47															
48															
49															
50															
51															
52															
53															
54															
55															
56															
57															
58															
59															
60	0														
61	0.829787	0													
62	0.68	0.69697	0												
63	0.810811	0.898305	0.793651	0											
64	0.716981	0.59375	0.6	0.818182	0										
65	0.603448	0.765957	0.5625	0.71831	0.607843	0									
66	0.59322	0.75	0.55102	0.690141	0.648148	0.313725	0								
67	0.59322	0.723404	0.55102	0.708333	0.54	0.28	0.204082	0							
68	0.56	0.675676	0.666667	0.84507	0.652174	0.636364	0.672414	0.649123	0						
69	0.768116	0.851852	0.741379	0.333333	0.770492	0.686567	0.656716	0.676471	0.820896	0					
70	0.683333	0.782609	0.686275	0.844156	0.673077	0.586207	0.6	0.622951	0.648148	0.821918	0				
71	0.722222	0.820513	0.642857	0.838235	0.627907	0.56	0.576923	0.603774	0.659575	0.793651	0.653846	0			
72	0.603448	0.816326	0.62	0.77027	0.634615	0.352941	0.377358	0.377358	0.684211	0.742857	0.481481	0.5	0		
73	0.6	0.826923	0.641509	0.794872	0.603774	0.358491	0.320755	0.351852	0.7	0.753425	0.534483	0.5	0.26	0	
74	0.753247	0.876923	0.8	0.441176	0.805556	0.697368	0.671053	0.653333	0.815789	0.415385	0.802469	0.791667	0.746835	0.740741	0
75	0.689655	0.822222	0.769231	0.805556	0.653061	0.704918	0.734375	0.672131	0.703704	0.797101	0.737705	0.574468	0.704918	0.633333	0.676056
76	0.789474	0.887097	0.80597	0.40625	0.828571	0.716216	0.706667	0.706667	0.805556	0.403226	0.822785	0.779412	0.75	0.759494	0.225806
77	0.734375	0.884615	0.767857	0.736111	0.732143	0.6	0.676923	0.65625	0.75	0.705882	0.71875	0.56	0.645161	0.619048	0.625
78	0.785714	0.872727	0.741379	0.272727	0.790323	0.724638	0.695652	0.714286	0.820896	0.327273	0.821918	0.830769	0.742857	0.77027	0.439394
79	0.772727	0.816326	0.745455	0.483333	0.79661	0.782609	0.735294	0.753623	0.728814	0.508475	0.828571	0.83871	0.816901	0.824324	0.565217
80	0.754717	0.735294	0.65	0.830769	0.697674	0.673077	0.709091	0.685185	0.571429	0.803279	0.711538	0.575	0.673077	0.690909	0.816901
81	0.685185	0.769231	0.717391	0.859155	0.673913	0.724138	0.733333	0.666667	0.522727	0.835821	0.758621	0.708333	0.724138	0.716667	0.873418
82	0.74	0.75	0.813954	0.942029	0.738095	0.862069	0.885246	0.847458	0.642857	0.90625	0.792453	0.826087	0.842105	0.810345	0.905405
83	0.711538	0.777778	0.777778	0.915493	0.733333	0.833333	0.819672	0.8	0.613636	0.878788	0.763636	0.765957	0.813559	0.741379	0.881579
84	0.62	0.685714	0.609756	0.725806	0.595238	0.434783	0.458333	0.458333	0.630435	0.666667	0.703704	0.636364	0.530612	0.529412	0.720588
85	0.692308	0.75	0.697674	0.869565	0.73913	0.634615	0.648148	0.648148	0.55814	0.846154	0.62	0.659091	0.58	0.576923	0.853333
86	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74

Appendix 7. Pairwise genetic distances for 85 *Pachyrhizus* taxa.

	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CH
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35											
36											
37											
38											
39											
40											
41											
42											
43											
44											
45											
46											
47											
48											
49											
50											
51											
52											
53											
54											
55											
56											
57											
58											
59											
60											
61											
62											
63											
64											
65											
66											
67											
68											
69											
70											
71											
72											
73											
74											
75	0										
76	0.695652	0									
77	0.42	0.642857	0								
78	0.742424	0.377049	0.686567	0							
79	0.784615	0.515625	0.727273	0.4	0						
80	0.693878	0.80597	0.673077	0.783333	0.745455	0					
81	0.722222	0.866667	0.786885	0.835821	0.806452	0.560976	0				
82	0.755102	0.915493	0.842105	0.90625	0.821429	0.692308	0.564103	0			
83	0.7	0.890411	0.77193	0.878788	0.793103	0.625	0.571429	0.142857	0		
84	0.711538	0.742424	0.690909	0.711864	0.779661	0.674419	0.680851	0.826087	0.791667	0	
85	0.705882	0.84507	0.732143	0.828125	0.775862	0.444444	0.581395	0.641026	0.609756	0.659091	0
86	75	76	77	78	79	80	81	82	83	84	85