Molecular Analysis of Genetic Diversity in Domesticated Pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild relatives

A thesis submitted in the fulfilment of the requirements for the degree of

Doctor of Philosophy
of

Rhodes University

By
Mulualem Tamiru Kassa

July 2011

Dedication

This thesis is dedicated to my late father Capt. Tamiru Kassa and my late brother Zena T. Kassa

Abstract

Cajanus cajan (L.) Millsp. (Pigeonpea) belongs to the Leguminosae genus *Cajanus* which is composed of 34 species. Pigeonpea is the only cultivated member of the genus, while the remaining species are wild relatives belonging mainly to the secondary gene pool.

DNA sequence data from the nuclear *ITS* region and the chloroplast *trnL-F* spacer were utilized to investigate the phylogenetic relationships between *Cajanus* and five other allied genera in the subtribe Cajaninae. This study revealed the non-monophyly of *Cajanus* and *Rhynchosia* and supported the monophyly of *Eriosema* and *Flemingia*, but more sampling ,especially from the large genera of *Rhynchosia* and *Eriosema*, is recommend to adequately test the hypothesis of generic monophyly. The phylogenetic relationships within the genus *Cajanus* resolved *Cajanus scarabaeoides* (L.) Thouars as the most basal species in the *Cajanus* clade. The study also utilized Single Nucleotide Polymorphism (SNP) markers derived from low copy orthologous genes and genotyped using the high throughput SNP-OPA Illumina golden gate assay. The aim was to understand phylogenetic and domestication history, genetic structure, patterns of genetic diversity, gene flow and historical hybridization between *Cajanus cajan* (pigeonpea) and wild relatives. The neighbor-joining tree resolved well-supported clusters, which reflect the distinctiveness of species and congruence with their geographical origin. It supported the *ITS* based phylogeny and resolved *C. scarabaeoides* as basal to the *Cajanus* clade. The phylogenetic signal and genetic signatures revealed insights into the domestication history of pigeonpea. Our results supported *Cajanus cajanifolius* as the presumed progenitor of pigeonpea and we speculate that for pigeonpea there was a single major domestication event in India. Genetic admixture and historical hybridization were evident between pigeonpea and wild relatives. Abundant allelic variation and genetic diversity was found in the wild relatives, with the exception of wild species from Australia, as compared to the domesticated pigeonpea. There was a reduction of about 75% in genetic polymorphism in domesticated pigeonpea as compared to the wild relatives, indicating a severe "domestication bottleneck" during pigeonpea domestication.

We discovered SNP markers associated with disease resistance (NBS-LRR) loci. The SNPs were mined in a comparison of BAC-end sequences (BES) of *C. cajan* and amplicons of the wild species, *C. scarabaeoides*. A total of ~3000 SNPs were identified from 304 BES. These SNPs could potentially be used in constructing a genetic map and for marker assisted breeding.

Table of contents

CHAPTER 2

Phylogenetic reconstruction of *Cajanus* and allied genera

CHAPTER 3

Genetic Structure and Patterns of Genetic Diversity in Domesticated Pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild relatives

CHAPTER 4

Mining of Single nucleotide polymorphism (SNP) markers associated with disease resistance (NBS-LRR) genomic regions in pigeonpea (*Cajanus cajan* (L.) Millsp.) BAC-end sequences

List of Figures

Chapter 1: Introduction: The Pigeonpea

List of Tables

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisor, Professor Nigel Barker (Rhodes University) for his intellectual guidance, meticulous and persistent supervision and detailed constructive comments. I would like to extend my great appreciation and thanks to my supervisor Professor Doug Cook (University of California, Davis) without him this work would have been impossible. I gained tremendous knowledge from Doug's intellectual and logical thinking and supervision in areas of genomics and molecular genetics, which had a remarkable influence on my future career in the fields of molecular genetics. My gratitude also goes to my co-supervisor Professor Jos van der Maesen (Wageningen University, The Netherlands), who conceived the work on *Cajanus*, for his important comments and critical advice and also providing some of the plant material.

Many thanks goes to my family and especially to my wife Amsale T. Asfaw for her unfailing love, support, encouragement and understanding during many challenges of this journey, thanks "Amsi".

I cherish the scientific and social cohesiveness of members of the Cook lab and I would like to thank all but the contributions of Dr. Varma Penmetsa, Ben Rosen and Noelia Carrasquilla were huge and need to be mentioned. I thank Dr. Eric von Wettberg (Florida International University) for his important insights and help in genetic structure and diversity analyses. I also thank Andrew Farmer (National Center for Genome Resources, Santa Fe, New Mexico) for his considerate support in computational component of the R-gene SNP discovery.

I thank Dr. Syd Ramdahani for his critical help on my work in Nigel's lab and I appreciate Dr Muthama Muasya (University of Cape Town) for his kind donation of the *Eriosema* samples. I appreciate ICRISAT and Western Australian herbarium, Perth for providing plant materials used in this work.

I am grateful to the Department of Agricultural Research of Botswana which created a conducive environment to start my PhD work. I thank many of my friends in Botswana, South Africa, USA and Ethiopia for their moral support especially the input of my friend Dr. Berhanu F. Alemaw.

This work was mainly supported from grant of the National Science Foundation (NSF) for Legume Genome Evolution (awarded to Doug Cook) and I greatly express my appreciation to NSF.

DECLARATION

I hereby declare that this thesis is my original work and has not been submitted in any other university for degree or award. Where other sources of information have been used, they have been acknowledged.

.....

MULUALEM TAMIRU KASSA

Preface

Despite the economic and ecological importance of *Cajanus* DC. and allied genera in the subtribe Cajaninae, there is little information on molecular phylogenetic relationships among taxa in the subtribe. Even though there are various studies and treatments on morphological, alpha-taxonomic and some biochemical studies in these taxa, there is a dearth of phylogenetic studies using DNA sequence data. One of the objectives of this study was to independently test chloroplast and nuclear DNA sequences to explain phylogenetic relationships of *Cajanus* and allied genera in the sub tribe Cajaninae.

This study included example taxa from most of the genera in the subtribe and more focus was given to the genus *Cajanus* in which the cultivated pigeonpea (*C. cajan*) belongs. Given that there are about 500 species in the subtribe, detailed future study is recommended to include more taxa especially from large genera such as *Rhynchosia* and *Eriosema* to understand the phylogenetic relationships and biogeography of the subtribe.

Similarly no detailed information exists on genetic and genomic diversity studies, genetic structure, domestication history, tracking historical hybridization and admixture between wild and domesticated pigeonpea. This study uses high throughput Single Nucleotide Polymorphism (SNP) genotyping technology to elucidate the genetic and genomic profile of domesticated pigeonpea (*Cajanus cajan*) and wild relatives. It also aims to reinforce the phylogenetic relationships and evolutionary history of *Cajanus* using the genomic data set. The genotypes included in this study have been collected from various parts of Asia, Africa and Australia and an attempt has been made to give insight into the correlation between geography and phylogenetic structure of the taxa. This genetic and genomic analysis also provides insights into the domestication history of pigeonpea.

The other realm of this study is molecular marker discovery, which may have potential application in marker assisted breeding in pigeonpea improvement. This study aims to discover SNP markers associated with disease resistance loci. The SNP markers have been mined from BAC-end sequences containing the NBS-LRR (disease resistance) regions of the genome developed by the University of California, Davis.

Main aims of the present study

There are three major objectives of this study, which are detailed in the respective chapters of this thesis.

1. To understand the evolutionary history and phylogenetic relationship of *Cajanus* DC. and allied genera in subtribe Cajaninae. We used sequence data from the nuclear ribosomal Internal Transcribed Spacer (*ITS*) and the cpDNA (*trnL-F* spacer) to infer inter-generic and inter-specific relationships of *Cajanus* and allied genera (Chapter 2).

2. To understand evolutionary history, genetic diversity, genetic structure and genetic polymorphism of domesticated and wild relatives of pigeonpea (*Cajanus cajan*) using high throughput Single Nucleotide Polymorphisms (SNPs). This will provide molecular insights into the domestication history of pigeonpea and to track historical genetic admixture, hybridization and successive gene flow during domestication of pigeonpea from wild progenitor (s) (Chapter 3).

3. To discover Single nucleotide polymorphism (SNP) markers associated with disease resistance loci, which can potentially be used for marker-assisted breeding in pigeonpea. The SNP markers are mined from BAC-end sequences of the Nucleotide Binding site-Leucine Rich Repeat (NBL-RR) regions of the genome (Chapter 4).

The first chapter (Chapter 1) is a review of topics spanning from Leguminosae-Papilionoideae-Cajaninae systematics to genomics of pigeonpea (*Cajanus cajan*). It reviews systematics of Cajaninae, gene pool and genetic resources of pigeonpea and wild relatives. Use of wild relatives in breeding and improvement programs of pigeonpea including the use of wild species to develop hybrid cultivars is documented. It also highlights the current status of molecular and genomic resources of pigeonpea.

The final chapter (Chapter 5) is a synthesis and summary of the above chapters and articulation of thoughts of the findings of the study and recommendations for future work.

Chapter 1

Introduction: The Pigeonpea

The Legume Family

The *Leguminosae* is the third largest angiosperm family behind only daisies (family *Asteraceae*) and orchids (family *Orchidaceae)* with more than 19,325 species in about 727 genera (Lewis et al., 2005). Legumes are only second, next to the family *Poaceae* (grasses), in their magnitude of economic, ecological and agricultural significance and many other benefits (Wojciechowski et al., 2004). Legumes have agricultural importance and are used for food (grain legumes) as well as feed (forage and pasture legumes). Grain legumes (pulses) and leguminous oilseeds provide about one-third of all dietary protein and one-third of processed vegetable oil for human consumption (Graham and Vance, 2003). Legumes also have industrial, pharmaceutical and medicinal values such as secondary metabolites, fertilizers, aesthetic value, textiles, cover crops and enriching soil fertility. They also play a crucial role in the terrestrial nitrogen cycle.

From the perspective of morphology, legumes are the most diverse plant group with adaptations that made them nearly ubiquitous throughout temperate and tropical regions of the world. They are found in diverse ecological settings, ranging from rainforests to deserts, lowlands to alpine habitats as well as aquatic environments (Rundel, 1989). Legumes occur both as giant trees that form the major overhead canopy of tropical rainforests as well as annual and perennial herbaceous species in the temperate biome (Lewis et al., 2005). Legumes are particularly abundant in arid and semi-arid habitats and McKey (1994) suggests that such environments, where leaves can be produced economically and efficiently, might be optimal to meet the metabolic demand for nitrogen. Alternatively the comparative advantage of nitrogen fixation might make legumes fitter in the marginal environments.

The distinctive legume fruit has a variety of sizes and shapes ranging from small single-seeded types to very long woody pods and includes typical dehiscent pod forms, indehiscent wind-dispersed winged pods and articulated loments that are adapted for dispersal by adhering to the bodies of animals (Doyle and Luckow, 2003). The unique attributes of many legumes, such as nodulation and the butterfly shaped flower are not found in all members of the family. For example, the family includes many non-nodulating lineages and other taxa exhibiting

diverse forms of flower morphology, pollination biology, genomic organization and breeding systems (Tucker, 2003; Doyle and Luckow, 2003).

According to global paleontological evidence, the first distinctive and identifiable legume originated during the Late Cretaceous or Early Tertiary during the Late Paleocene, about 56 million years ago (Mya) and subsequently most Legume groups including members of the three subfamilies and large clades diversified soon afterward, beginning around 50 to 55 Mya (Herendeen, 1992; Herendeen, 2001; Magallon et al. 1999; Schrire et al. 2005). The fossil records indicate that diversification of lineages commenced immediately after the first legume appeared and that speciation was rapid. The result is that there is little difference between the estimated age of the origin (stem clade) of legumes and of the subsequent diversification (crown clade) of the major extant lineages (Herendeen and Wing, 2001). Evolutionary rate and age estimation of lineages of legumes based on molecular phylogenies calibrated with fossils (Lavin et al., 2005) strongly supported the paleontological evidence for the early explosion hypothesis of legume radiation and diversification (Cronk et al., 2006).

Schrire et al. (2005) suggested that the seasonally dry margins of the Tethys seaway was the ancestral area for legumes, implying that all South American taxa must be considered immigrants. The classification of all legumes as immigrants was disputed by some authors (e.g. Raven & Polhill 1981; Morley 2000, 2003), and the putative Madrean–Tethyan origin hypothesized by Schrire et al. (2005) was not inferred by a clear optimization of an ancestral area but rather from inference of historical ecological preferences. Legumes, however, dominate neo-tropical rainforests, dry forests and woody savannahs, and long-distance trans-oceanic dispersal has been instrumental in shaping the distribution and historical assembly of these biomes (Lewis et al., 2005).

For legumes, Lavin et al. (2004) showed that out of 59 trans-oceanic crown clades, only eight are older than 25 Myr ago, which implied that most have expanded their distributions by long-distance dispersal. Schrire et al. (2005) suggested that the legume phylogeny is more explicable by ecological setting, and Lavin et al. (2004) further bolstered the hypothesis that the frequency of long-distance dispersal between habitats is the principal determinant of legume phylogenetic patterns. However, recent analyses based on a sample of taxa from the temperate Southern Hemisphere (Crisp et al., 2009) confidently demonstrated that shift of phylogenetic biome has not been prevalent during the radiation of plant lineages, both within continents and in transoceanic colonizations.

Reconstructing evolutionary histories and phylogenetic relationships of this diverse, ecologically and economically important member of the angiosperms is of paramount importance if we are to understand the major biological process associated with the origin and diversification of this major family (Wojciechowski, 2003).

Systematics of *Leguminosae-Papilionoideae*

Legumes have been classified with the ill-defined, heterogeneous and phylogenetically challenging group of eudicots called "rosids" (Soltis et al., 2000; Hilu et al., 2003). Rosids comprise about 140 families (70,000 species), which represent more than one-fourth of the species of angiosperms (Soltis and Soltis, 2004; Wang et al., 2009). A better understanding of rosid phylogeny based on a combined data set of nuclear and plastid genes (10 plastid genes, 2 nuclear and plastid inverted repeat (IR)) resolved Vitaceae as sister to all other rosids (Wang et al., 2009). This study further resolved the rosid clade into two large sub-clades with strong bootstrap support; (*i*) eurosids I (Fabidae) include the nitrogen-fixing clade (*Cucurbitales, Fabales, Fagales* and *Rosales*), Celastrales, Huaceae, Zygophyllales, Malpighiales, and Oxalidales; and (*ii*) eurosids II (Malvidae) include Tapisciaceae, Brassicales, Malvales, Sapindales, Geraniales, Myrtales, Crossosomatales, and Picramniaceae. Previous molecular phylogenetic analyses have also resolved these two subclades of rosids (Savolainen et al., 2000; Soltis et al., 2000; Ravi et al., 2007). Various phylogenetic hypotheses have been advanced to explain relationships within the nitrogen-fixing clade (Soltis et al., 1995; Sytsma et al., 2002; Hilu et al., 2003;).

Traditionally Legumes were erroneously grouped with the Connaraceae, Chrysobalanaceae, Crossosomataceae, Krameriaceae and Sapindaceae (Dickison, 1981). Phylogenetic analyses with molecular and combined datasets have refuted these hypotheses and instead indicate a monophyletic order the Fabales (Eurosid I) which contains the monophyletic Leguminosae which is closely related to a clade comprising the families Polygalaceae, Surianaceae and the genus *Quillaja* (the Chilean soap tree) (Lewis et al., 2005). This same work further resolved Surianaceae as the most closely related family to the Leguminosae (Soltis et al., 2000; Persson, 2001). Phylogenetic Relationships among the first branching lineages of the legumes are not well supported, with Cercideae, Detarieae, and the genus *Duparquetia* alternatively resolved as sister group to all of the legumes (Bruneau et al., 2008).

The legume family has been divided into three major subfamilies of unequal size on the basis of major morphological characters, particularly floral structure. These subfamilies are the Caesalpinioideae, the Mimosoideae and the Papilionoideae, with the latter being the largest and most diverse of the three (Polhill, 1981; 1994). The subfamily Caesalpinioideae contains four tribes namely Cercideae (12 genera), Detarieae (82 genera), Cassieae (21 genera) and Caesalpinieae (56 genera). The subfamily Mimosoideae contains four tribes namely Mimoseae (40 genera), Ingeae (36 genera), Mimozygantheae (temporarely monogeneric) and Acacieae (monogeneric) with a total of 78 genera and 3271 species and includes such large and common groups as the genus *Acacia* (Lewis et al., 2005).

The subfamily Papilionoideae is the largest group comprising 478 genera and 13,805 species and includes most of the familiar "beans and peas", to which all of the model taxa belong (Lewis et al., 2005). The subfamily Papilionoideae is easily distinguished from the other sub-families by traits that are now considered to be morphological synapomorphies. Examples include floral, vegetative and fruiting characters (Polhill 1981; Tucker, 2002) and especially orientation of the seed hilum and unidirectional initiation of sepals (Doyle et al., 2000).

Papilionoideae dominate diverse agro-ecological conditions from tropical rainforests to deserts and alpine tundra and play important roles in global biological and geochemical process. Most members of the subfamily form a root-nodule symbiosis with bacteria and fix atmospheric nitrogen (Sprent, 1994; 2001). Most of the economically important cultivated legume crops are Papilionoideae species, which occur in two major clades, informally called the "tropical" (phaseoloid) legumes and temperate (galegoid) legumes. The former includes genera such as *Phaseolus, Vigna, Glycine, Cajanus* etc. and the latter includes genera such as *Trifolium, Medicago, Pisum, Vicia, Lotus, Cicer, Lens* and others (Young et al., 2003).

Numerous recent molecular phylogenetic studies and other datasets suggest a need for revision of the generic delimitation outlined by Polhill (1994). In the modern phylogenetic context (e.g. Wojciechowski et al., 2004), papilionoids can be broadly divided into the following main subclades and groups: the Cladrastis clade, the Genistoid clade, the Dalbergioid s.lat. (s.l.) clade, Mirbelioid s.l. clade, the Millettioid s.l. Clade and Hologalegina. However, the papilionoids also include paraphyletic lineages (such as the old tribes Sophorea and Swartzieae) that diverged before the main radiation of the papilionoid clade. These paraphyletic lineages also occur among "higher" groups (notably genistoids) (Pennington et al., 2001).

The most recent and robust molecular phylogeny of Leguminosae, which is based on monophyly (natural groupings) of clades, redefined the generic and tribal limits of traditional morphological systematics of Polhill (1981, 1994). These molecular phylogenetic studies resolved relationships of many taxa that were poorly understood by morphological systematics.

Molecular phylogeny clearly resolved the subfamilies Papilionoideae and Mimosoideae as fully or nearly monophyletic and nested within a paraphyletic assemblage of Caesalpinioideae (Doyle et al., 2000; Kajita et al., 2001; Wojciechowski, 2003; Wojciechowski et al., 2004). Caesalpinioideae contains a diverse assemblage of unrelated "caesalpinioid" lineages mostly diverging early in the history of the family (Wojciechowski, 2003). Plastid gene phylogeny (*matK* gene, the *trnL* and 3′-*trnK* introns) resolved the monophyly of several major groups within the caesalpinioid legumes: the Cercideae, Detarieae, Detarieae s. str., *Prioria*, Amherstieae, Dialiinae, *Cassia*, *Caesalpinia*, *Peltophorum*, and *Tachigali* clades (Bruneau et al., 2008).

The detailed *matK* sequence analysis of Wojciechowski et al. (2004) has resolved a robust legume phylogeny with strong statistical support for the monophyly of Papilionoideae, unlike the *rbcL* study (Kajita et al., 2001), which only weakly resolved their monophyly. This *matK* phylogeny recognized and informally named seven strongly supported clades within Papilionoideae (Wojciechowski et al., 2004).

The seven resolved subclades are the *Cladrastis* clade, the Genistoid s.l., the Dalbergioid s.l., the Mirbelioids, the Millettioids, the Robinioids, and the Inverted-Repeat Lacking Clade (IRLC). The last two clades are further combined into the Hologalegina clade (Wojciechowski et al., 2004). This matK phylogeny also resolved a super-clade associated with the evolution and accumulation of canavanine (a non-protein amino acid). The canavanine accumulating clade contains the mirbelioids, millettioids and the Hologalegina and is consistent with the early hypothesis that canavanine evolved only once in the evolution of legumes (Bell, 1981).

The Millettioid clade contains the tribes of Millettieae, Phaseoleae, Abreae, Psoraleeae and subtribes Lespedezinae and Desmodiinae of the tribe Desmodieae (Hu et al., 2000; Kajita et al., 2001; Wojciechowski et al., 2004). The Millettioid clade is characterized by the morphological synamorphy of the unique pseudoracemose inflorescence type (Tucker, 1987).

The tribe Phaseoleae, commonly known as beans, is the largest tribe of the legume family in terms of the number of genera. The current phylogenetic treatment accepts 89 genera and about 1567 species belonging to this tribe (Lewis et al., 2005). The core-Phaseoleae

comprises subtribes Kennediinae, Cajaninae, Glycininae, Phaseolinae and Erythrininae and it includes most of the important food, forage and ornamental species such as *Phaseolus, Vigna, Glycine, Cajanus, Lablab, Erythrina* and others (Lewis et al., 2005).

Subtribe Cajaninae (*Leguminosae-Papilionoideae*)

The subtribe Cajaninae Benth. is one of the economically important subtribes because it includes the agriculturally important cultivated crop, pigeonpea (*Cajanus cajan* (L.) Millspaugh, 2n=22) (Lackey 1977, 1981; van der Maesen, 2003). The subtribe Cajaninae Benth is a monophyletic basal clade of the core-Phaseoleae and is sister to a large clade containing most of the warm season crop legumes of tribes Phaseolinae and Glycinae (Bruneau et al., 1995; Doyle et al., 2000; Kajita et al., 2001). The recent study on the phylogenetic relationships of the phaseoloid Legumes based on eight chloroplast regions confirmed that Cajaninae is sister to the above-mentioned groups (Stefanovic et al., 2009; Fig. 1.1 of this thesis).

According to the recent taxonomic review of the Cajaninae, 10 genera and about 495 species are included within the subtribe (Lewis et al., 2005). The ten genera included in current taxonomic treatment are *Cajanus* (34 spp.), *Dunbaria* (20 spp.), *Bolusafra* (1 sp.), *Flemingia* (30-35 spp.), *Eriosema* (150 spp.), *Rhynchosia* (230 spp.), *Chrysoscias* (3-4 spp.), *Carrissoa* (1 sp.), *Adenodolichos* (15-20 spp.) and *Paracalyx* (6 spp.) (Table 1.1).

Endomallus Gagnep. (Lackey, 1981) and *Atylosia* Wight & Arn. (Lackey 1981; Reynolds and Pedley 1981) have been formally relegated and treated synonymous of the genus *Cajanus* and before the merger of these genera, *Cajanus* was considered as a monotypic genus (van der Maesen, 1986).

The subtribe Cajaninae is the most distinctive and internally consistent natural group of the tribe Phaseoleae and the key distinctive and peculiar morphological synamorphic characters of these taxa (Lackey 1978, 1981; van der Maesen, 2003) include:

- Presence of bulbous-based hairs (usually on the calyx and pods).
- Presence of yellowish vesicular pearl bodies (mainly on calyx, pods and lower surface of leaflets) and turning black or orange upon drying. These glands “consist of a squat head of cells contained within a shallow depression of the epidermis” (Moteetee and van Wyk, 2006).
- Non-nodose but straight inflorescences.

- Absolute absence of bracteoles and absent (or inconspicuous) stipules (except *Adenodolichos*),
- The style is slender at base, thickened distally and does not have a beard on the style (except *Adenodolichos*),
- Absence of the non-protein amino acid (canavanine) from seeds (except *Adenodolichos*).

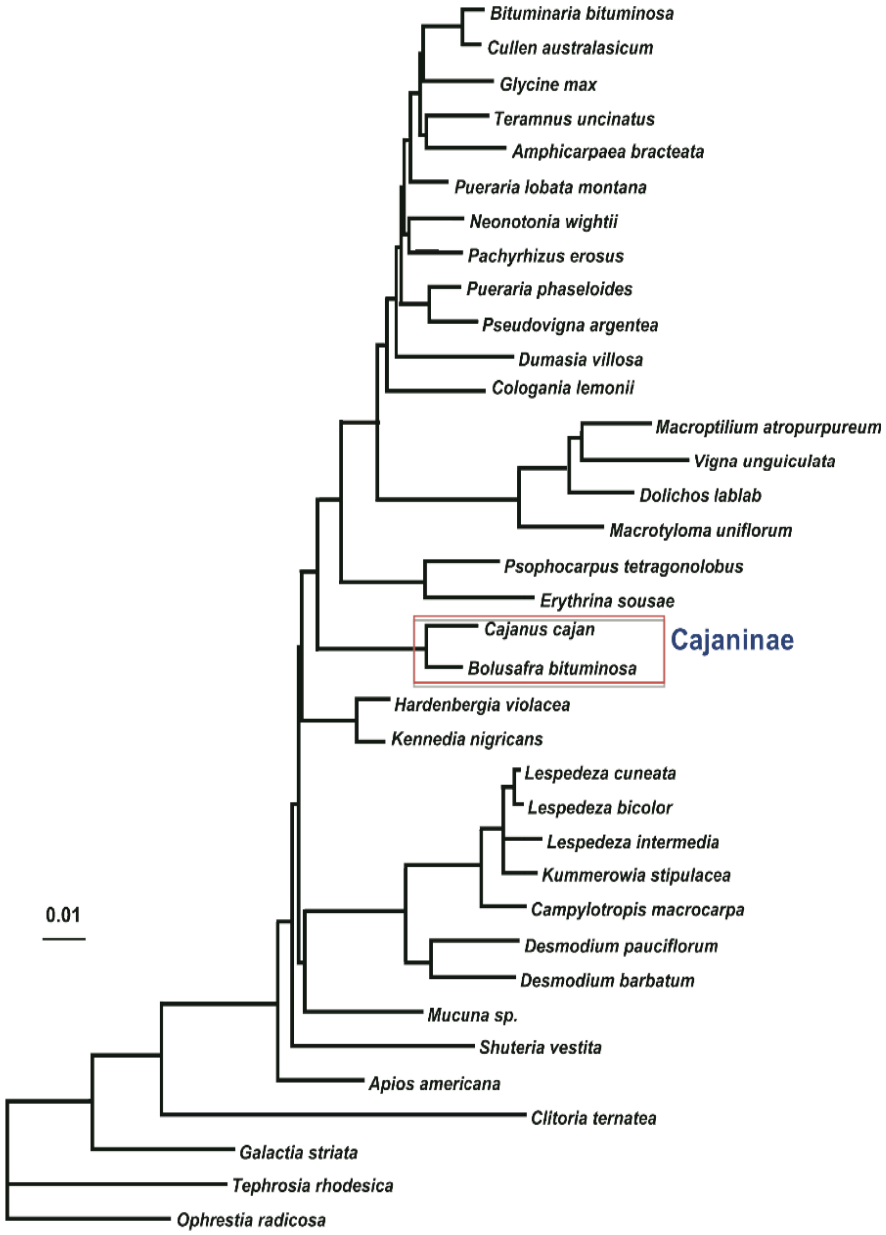


Fig. 1.1. Phylogenetic position of Cajaninae within the phaseoloid group based on sequences of seven concatenated chloroplast regions (*trnL-F, rbcL, atpB, trnK/matK, rpl2, clpP & rps1)* (taken from Stefanovic et al., 2009).

Even though the tropical African genus *Adenodolichos* shows distinct characters that spoil the common features of the subtribe such as the presence of bracteoles, canavanine accumulation and a bearded style, the presence of glandular hairs and the overall resemblance in general morphology with the Cajaninae results in the provisional placement of the genus in the subtribe Cajaninae (Lackey, 1981). Furthermore, chloroplast restriction mapping phylogeny (Doyle and Doyle, 1993) and *rbcL* phylogeny (Kajita et al., 2001) supported the monophyly of the subtribe including *Adenodolichos*.

There was little resolution on generic relationships of the subtribe from this plastid mapping phylogeny except the resolution of the sister relationship between *Rhynchosia* and *Eriosema*. This was consistent with the morphological classification of the subtribe (Baudet, 1978) in which both *Rhynchosia* and *Eriosema* posses 2 ovules in each ovary.

The subtribe Cajaninae is believed to have originated in Africa with a biogeographical distribution of predominately paleotropical and warm old world temperate regions (about 400 spp.). Cajaninae occur less frequently in the Neotropics and Subtropics (only about 95 spp.) and the latter is mainly represented by the two pantropical genera of *Eriosema* (DC.) Rchb. and *Rhynchosia* Lour. (Fortunato, 2000; van der Maesen, 2003; Lewis et al., 2005).

Superimposition of taxon-biome relationships derived from the *matK* phylogeny super-tree (Wojciechowski et al., (2004) illustrated that 255 species are associated with a Grass (G) biome; 236 species associated with a Succulent (S) biome and 4 species associated with a Temperate Southern Hemisphere (TS) biome (Fig. 13 of Lewis et al., 2005).

Species of Cajaninae reflect a range of growth habits, with individual species occurring as herbs, vines, lianas, subshrubs or shrubs and sometimes climbing herbs or subshrubs. They are adapted to a range of ecological situations from the seasonally dry tropics to the subtropics from open forests to grasslands in forest margins, woodland thickets, scrubs, wooded and open grasslands. They also occur on diverse substrates, including open rocky areas or sometimes near swamps and streams and also in disturbed and old cultivated lands or waste areas (Lewis et al., 2005; Table 1.1 of this thesis).

Table 1.1. Biogeographical distribution and growth habit of genera Cajaninae (Modified from Lewis et al., 2005)

Genera	No of species	Geographical distribution	Habit
Rhynchosia Lour.	230	Pantropical [Africa- Madagascar (c 140 spp.); tropical and subtropical America (c 55 spp.; North and Central America (28 spp.); South America (20 spp.) and both (6 spp.)]; c 30 – 35 spp. Warm temperate to tropical Asia to N Australia (2 endemic spp.).	Herbs, vines or subshrubs; seasonally dry forest, forest margins, woodland, thicket, wooded grassland, shrubland and grassland.
Eriosema (DC.) Rchb.	150	Pantropical [c 100- 110 spp. (African-Madagascar); 40 spp. (North and South America and most are South American (c30 spp.)); 2 spp. (SE Asia & Australia).	Herbs or subshrubs; seasonally dry tropical to subtropical forest margins, woodland, thicket and wooded grassland.
Flemingia Roxb.	30 - 35	26-29 spp. (SE Asia & Indo-China); 3-4 spp. (Australia); 1 sp (*F. grahamiana*, widespread in Africa and Asia).	Herbs or shrubs; seasonally dry tropical forest, woodland, wooded grassland and grassland.
Cajanus DC.	34	SE Asia, Indo-china Pacific (16 spp.); N Australia (15 spp.); W (Sudanian) Africa (1 sp.); widespread in old world including *Cajanus cajan* (L.) Millsp. (2 spp.).	Herbs or shrubs; seasonally dry tropical open forest to grassland.
Dunbaria Wight & Arn.	20	Predominantly centered in SE Asia and Indo-China; 1 sp E Asia; 2 spp. in N Australia. 1 sp. recently in Africa	Climbing herbs or subshrubs; seasonally dry tropical forest, thicket, scrub and grassland.
Adenodolichos Harms	15 - 20	Predominantly SC Africa (Zambezian region); 1 sp to Sudanian (WC Africa).	Herbs or shrubs; seasonally dry tropical woodland, scrub, wooded grassland and grassland.
Paracalyx Ali	6	Predominantly NE Africa; 1 sp. Indo-China.	Herbs; seasonally dry tropical forest, woodland, thicket, bushland and scrub.
Chrysoscias E. Mey.	3 – 4	Endemic to South Africa (confined to south of western cape region).	Herbs or subshrubs, Mediterranean montane sclerophyllous shrubland (fynbos).
Bolusafra Kuntze	1	Endemic to South Africa (confined to Western cape region).	Herbs or subshrubs; Mediterranean montane sclerophyllous shrubland (fynbos).
Carrissoa Baker f.	1	Endemic to Angola * There is speculation that *C. angolensis* Baker f. may be *Rhynchosia* (q.v.).	Herbs or subshrubs; Mediterranean montane sclerophyllous shrubland (fynbos).

Previous systematic studies of *Cajanus* and allied genera

Morphology and alpha-taxonomy

Numerous morphological and alpha taxonomic studies of *Cajanus* and related genera in the subtribe Cajaninae have been undertaken (Grear, 1978; Lackey, 1978; Stirton, 1981; Pundir, 1985a, 1985b; van der Maesen, 1986, 1990, 1998, 2003; Satyanararayana, 1993; Fortunato, 2000).

Biosystematics and alpha-taxonomic revisions of *Cajanus* and *Dunbaria* focus on accessions obtained from Asia and/or Australia (van der Maesen 1985, 1998, 2003). Several treatments are available for the large genera *Rhynchosia* (Grear, 1970; 1978; Fortunato, 2000) and *Eriosema* (Stirton, 1975) of new world and African origin respectively.

The majority of the species of *Rhynchosia* and *Eriosema* are found in Africa and the Americas, with less frequent occurrence in Asia and Australia. Most of the species of the genus *Cajanus* are endemic to either Southern/South-eastern Asia or Australia (Fortunato, 2000). Among these, 16 *Cajanus* species occur in Asia (8 of which are endemic to India), 15 species in Australia (of which 13 are endemic), one species of *Cajanus* is confined to West Africa and 2 species (including *Cajanus cajan*) are widespread throughout the old world (Table 1.1).

Using morphological and ecological characters such as habit, leaf structure, hairiness, pod size, strophiole characters and other traits, van der Maesen (1986) grouped the genus *Cajanus* into six sections vis-à-vis Cajanus (2 species), Atylosia (7 species), Fruticosa (9 species), Cantharospermum (5 species), Volubilis (6 species) and Rhynchosoides (3 species). Species in sections Cajanus, Atylosia and Fruticosa have erect growth habit, Cantharospermum, Volubilis and Rhynchosoides are climbing and creeping species and Rhynchosoides are trailing species. Three *Cajanus* species have been further subdivided into botanical varieties; *C. scarabaeoides* into var. *pedunculatus* and var. *scarabaeoides*, *C. reticulatus* into var. *grandifolius*, var. *reticulatus*, and *var. maritimus*, and *C. volubilis* into *var. burmanicus* and *var. volubilis* (van der Maesen, 1986).

Assessment of diversity using seed protein markers (Panigrahi et al., 2007) reaffirmed the morphology-based groupings, such as section Cajanus (*C. cajanifolius* and *C. cajan*). Morphologically, *C. cajanifolius* resembles *C. cajan* (pigeonpea) in all traits except that it has a prominent strophiole (van der Maesen, 1990); section Atylosia (*C. lineatus* and *C. sericeus*); section Cantharospermum (*C. albicans, C. goensis* and *C. scarabaeoides*) and section Fruticosa (*C. acutifolius*, *C. lanceolatus* and *C. reticulatus*) and these proteins suggested a

close relationship between *C. sericeus* (section Atylia) and *C. volubilis* (sec. Volubilis) but not to *C. lineatus* in contrast to the classical taxonomy. This protein analysis placed *C. platycarpus* (sect. Rhynchosoides) as sister to the remaining taxa of the group.

Dunbaria Wight & Arn. is restricted to Asia and north Australia (van der Maesen, 2003). Infra-generic studies of the neotropic *Rhynchosia* species identified three major sectional groupings: Section *Copismsa*, Section *Rhynchosia* and Section *Archiphyllum* (Fortunato, 2000). Previous morphological, anatomical and chemotaxonomy studies have demonstrated a close sister relationship between *Rhynchosia* and *Eriosema* (Lackey, 1981) and early chloroplast DNA and morphological analyses have also suggested sister relationships of the two genera (Doyle and Doyle, 1993; Bruneau et al., 1995).

The genus *Chrysoscias* E.Mey. consists of three or four species and is endemic in South Africa and is usually included in *Rhynchosia* (Germishuizen, 2000). The monotypic genus *Carissoa* Baker f. (*C. angolensis*) is endemic to Angola, but Stirton (1981) speculated the genus might belong to *Eriosema* (Stirton in Lackey 1981). More recently, Vercourt in Mackinder et al., (2001) subsumed *C. angolensis* Baker f. into *Rhynchosia*. The monotypic genus *Baukea* (endemic to Madagascar) has recently been relegated and sunk into the genus *Rhynchosia* and in current taxonomic treatments it is included as *Rhynchosia baukea* (Du Puy et al., 2002).

All species of *Paracalyx* were originally included in the genus *Cylista* which is currently accepted as synonym of *Rhynchosia* (Ali, 1968). The genus *Flemingia* Roxb, which occurs predominantly in South East Asia and Indo-China, is currently under revision (van der Maesen, personal communication.). The monotypic genus *Bolusafra* Kuntze (formerly *Fagelia* DC.), consisting of the only species *Bolusafra bituminosa* (L.) Kuntze, is endemic and confined to a small area in the Western Cape of South Africa. *Bolusafra bituminosa* is recognized by its twining perennial habit with sticky branches, pleasant herbal odor and with morphological traits that resemble *Rhynchosia* Lour. In fact, *B. bituminosa* was previously described as "viscid *Rhynchosia*-like vine" (Lackey, 1981) but distinct from *Rhynchosia* in its conspicuous seed arils and turgid fruits.

A chloroplast restriction mapping phylogeny of the tribe *Phaseoleae* (Doyle and Doyle, 1993) did not include the genus *Bolusafra* within the *Rhynchosia-Eriosema* clade. The recent alpha-taxonomic revision of *Bolusafra* (Moteetee and Van Wyk, 2006) based on morphological features provided additional botanical information on this little known monotypic legume species since the description of the genus by Kuntze in 1891.

Biochemical and Molecular studies

Many genetic and taxonomic studies focusing on the relationships of the cultivated *Cajanus cajan* species and the wild relatives have been conducted. These use different genetic and biochemical methodologies including isozyme analysis (Krishna and Reddy, 1982), seed proteins (Jha and Ohri, 1996; Panigrahi et al., 2007), Randomly Amplified Polymorphic DNA (RAPD) studies (Nadimpalli et al., 1993), Restriction Fragment Length Polymorphism (RFLP) data (Ratnaparkhe et al., 1995; Lakshmi et al., 2000), Karyotype (Ohri and Singh, 2002), Amplified Fragment Length Polymorphism (AFLP) (Panguluri et al., 2006), Single Sequence Repeats (SSR) (Odeny et al., 2007) and Diversity Arrays Technology (DArT) (Yang et al., 2006).

Restriction Fragment Length Polymorphism (RFLP) analysis of four chloroplast gene regions and 15 restriction enzymes on five genera of Cajaninae (*Cajanus, Rhynchosia, Dunbaria, Flemingia* and *Paracalyx*) revealed very poor resolution of inter-generic relationships among these taxa, suggesting close relationships (Lakshmi et al., 2000). In contrast, another RFLP study (Sivaramakrishnan et al., 2002) detected variation at both inter-generic and infrageneric taxa of *Cajanus* and *Rhynchosia*. This study indicated variation in mitochondrial DNA hybridization patterns and found discernable rearrangements of this genome among *Cajanus* species.

Plant domestication as a model for recent evolution

Crop domestication, a process involving both biological and cultural evolution, is described as a continuum of increasing mutual dependence between humans and plants they rely upon (Zeder, 2006; Aguilar-Meléndez et al., 2009). The research on crop domestication is a fascinating interdisciplinary field of study spanning archeology to molecular biology (Gross and Olsen, 2010). Recent advances in molecular and genomic tools have provided insights into the crop evolution and domestication histories of important food crops (e.g Hyten et al., 2006).

One common feature of domesticated genomes is the reduction of genetic diversity (usually referred as a "genetic bottleneck") in crops relative to their wild relatives. There are two forces that cause the reduction in genetic diversity during the process of domestication. The first is the initial small population size (population genetic bottleneck) that occurs during the initial founding of new crop lineage and the second factor that exacerbates the reduction in genetic diversity is the 'selective sweep', or directional or artificial selection (e.g. plant breeding) for local genomic regions or genes associated with domestication traits and

agronomically desirable traits (Tanksley and McCouch, 1997).

Fundamental questions that should be addressed in the study of crop domestication include when, where, how often and how quickly did crop domestication occur, recurrence of domestication (single or multiple times), pinpointing the putative wild progenitor and investigation of the overall changes in genetic architecture during domestication and possible identification of genomic regions and genes that were subjected to selection during the evolution of crops (Burger et al., 2008; Aguilar-Meléndez et al., 2009). The phenotypic and morphological changes that occur during crop evolution and domestication are collectively referred as the "domestication syndrome" (Harlan, 1992).

Although the domestication syndrome is associated with dramatic changes in phenotypic and morphological features in the domesticated lineage, recent genetic analysis has shown that traits related to domestication are highly heritable, governed by relatively few major genes or loci with huge phenotypic effects and are often clustered within the genome. This suggests that the domestication process may have happened quite rapidly for at least some crops (Frary and Doúanlar, 2003). Recent archeological and genetic studies have suggested that the duration of domestication can range from a rapid to a multimillennial process depending on the crop species (Li et al., 2010). Allelic patterns of the Single Nucleotide Polymorphisms (SNPs) and genetic structure between pigeonpea (*Cajanus cajan*) and its putative wild progenitor (*Cajanus cajanifolius*) observed in the present study provide insights on the domestication history of pigeonpea (see Chapter 3 of this thesis).

Pigeonpea domestication and constraints imposed by genetic bottlenecks

The domestication bottleneck is the most ancient genetic constraint in the history of crop evolution and represents the first intervention of human selection. In addition to a decrease in genetic diversity, the bottleneck effects also alter allele frequencies, increases linkage disequilibrium (LD), and eliminates rare alleles in the resulting domesticated population (Halliburton, 2004). The magnitude of these effects will depend on the variety of evolutionary factors such as the effective population size, the strength of selection intensity, duration of the genetic bottleneck and rate of gene flow (Le Thierry d'Ennequin et al., 1999).

Even though there are no detailed studies on the evolutionary history of domestication in pigeonpea, morphological and alpha-taxonomic studies attempted to give insight on the centers of origin, diversity and possible domestication process of the crop. Asia (the Indian

subcontinent), Africa and Australia are considered as important centers of diversity for pigeonpea. The name pigeonpea was first coined in 1692 from Barbados where it was used to feed pigeons (van der Maesen, 1986).

In general, India (van der Maesen, 1980) and Africa (Purseglove, 1968; Tindall, 1988) were indicated as possible centers of origin for pigeonpea. The Indian center was strongly supported based primarily on the presence of several wild species, the recognition of an endemic center for the putative progenitor, existence of high morphological diversity, presence of diverse sets of primitive (or feral) varieties, ample linguistic evidence and a variety of use in the daily cuisine (van der Maesen, 1990).

Pigeonpea exhibited high levels of phenotypic and morphological diversity in terms of vegetative, floral, days to maturity, photoperiod sensitivity, growth habit and other phenotypic and agronomic traits. Pigeonpea germplasm comprises diverse sets of landraces and heterogenous feral forms with extensive morphological diversity. There are determinate, semi-determinate and indeterminate genotypes that are adapted to various agro-ecological settings.

Regardless of this extensive morphological and phenotypic diversity, recent genetic and genomic analyses using molecular markers such as SSR, DArT and SNPs in pigeonpea and wild relatives revealed that there is very low genetic diversity in the domesticated gene pool as compared to the wild groups (Yang et al., 2006; Odeny et al., 2007; Kassa et al., in preparation). This striking low polymorphism within the domesticated accessions (including the landraces) signals the severity of the "genetic bottleneck" which happened during pigeonpea domestication.

To broaden the genetic base of this highly constrained and narrowed genetic diversity in the domesticated gene pool there is a need to utilize the high genetic diversity present in the wild gene pool. The SSR and DArT studies recognized that a vast amount of genetic diversity within pigeonpea is untapped in wild germplasm as a result of domestication. The wild relatives are a potential source of novel alleles that can be exploited in breeding and improvement programs in pigeonpea and have indeed been used already.

Economic importance and cultivation of pigeonpea in the world

Cajanus cajan (L.) Millsp. or pigeonpea (2n = 22 and genome size of 808 Mbp) is a hardy, widely adapted and drought tolerant pulse crop cultivated on 4.92 million hectares of dry and marginal lands of primarily semi arid tropics and subtropics in south Asia (mainly on the Indian-subcontinent), Africa, Caribbean and Latin America with 3.65 million tonnes of

global annual production with average productivity of 898 kg/hectare (http://faostat.fao.org). Asia contributes about 89% in global area and 87% in global production. India is the major pigeonpea growing country with 3.56 million hectare (76% of global area) and 2.31 million tonnes 70% of global production) (www.faostat.org). In Asia, India (3.58 Mha), Myanmar (560,000 ha), China (150,000 ha), and Nepal (20,703) are major pigeonpea growing countries. In Africa the major pigeonpea growing countries are Kenya (196,261 ha), Malawi (123,000 ha), Uganda (86,000 ha) and Tanzania (68,000 ha).

Pigeonpea is a short-lived perennial shrub (van der Maesen et al. 1980) with wide difference in days to maturity (90 – 300 days) and high level of morphological and phenotypic diversity. It is a highly adaptable grain legume for promoting food security in rain-fed agriculture because it tolerates drought, requires very minimal inputs to give sustainable yield and it tolerates very harsh biotic and abiotic stresses. As a legume crop, it plays a major role in fixing atmospheric nitrogen through symbiotic nitrogen fixation with soil bacteria and improves the nutrient status of the soil. In addition to its main use as de-hulled split peas ("dhal") which is the primary source of dietary protein (20-24%) for millions of resource-poor people around the world. Pigeonpea also has other uses: its immature seeds and pods are consumed fresh as green vegetables, stems are used as domestic fuel wood and for making huts and leaves are used as quality fodder, (Saxena et al. 2006).

In spite of having typical Papilionoideae flowers, pigeonpea exhibited considerable variation (20-70 %) in natural insect-aided out-crossing rate and is a partially cross-pollinated species (Saxena et al., 1990). This considerable out-crossing rate may have two major impacts on pigeonpea agronomy and breeding. It creates a problem for maintaining genetic purity in cultivar development, but on the other hand it has been used effectively in developing elite hybrid varieties through hybridization (Saxena, 2008). Recently, Saxena and Kumar (2010b) observed that insect species also regularly visited the wild relatives of pigeonpea and noted a few naturally out-crossed plants with distinct traits.

The gene pool structure of pigeonpea and wild relatives

According to the gene pool concept of Harlan and de Wet (1971), which is based on the basis of genetic crossability and exchange of genetic material, within intra-generic taxa, there are three major gene pools:

1- *Primary gene pool*- all of cultivated germplasm of a species.

2- *Secondary gene pool*- wild species that can be crossed easily with cultivated germplasm.

3- *Tertiary gene pool*- wild relatives which cannot be crossed with the cultivated germplasm unless with special techniques such as hormonal treatment or embryo rescue.

In pigeonpea, the allied wild *Cajanus* species belong to the secondary genepool while most Cajaninae genera belong to the tertiary genepool (van der Maesen, 2003). Ramanandan (1990) classified 10 *Cajanus* species into the secondary genepool while 20 *Cajanus* species were placed in the tertiary genepool.

The gene bank at International Crop Research Institute for the Semi Arid Tropics (ICRISAT) possesses 13,632 pigeonpea accessions of which 13,077 accessions belong to the primary genepool of domesticated genotypes and 555 accessions are wild relatives represented by 57 species belonging to secondary and tertiary genepool (Upadhyaya et al., 2007). Species in the secondary gene pool such as *Cajanus albicans, C. lineatus, C. scarabaeoides*, *C. sericeus* have genes for high seed protein and *Cajanus sericeus* has genes for resistance to sterility mosaic virus and P2 race of *Phytophthora* blight disease. Similarly, *Cajanus platycarpus* from the tertiary genepool has shown to be the only source of resistance to the P3 race of *Phytophthora* blight disease (Saxena, 2008).

Wild species in secondary and tertiary gene pools also posses useful genes for extra-early flowering and maturity, photoperiod insensitivity, good flowering and pod setting, true annuality, rapid seedling growth (Mallikarjuna and Moss 1995). These wild species are also gene sources for salinity tolerance (Subbarao 1988, Srivastava et al., 2006), drought tolerance, resistance to sterility mosaic virus, *Phytophthora* blight disease (Reddy et al. 1996, Mallikarjuna et al. 2005, 2006) and tolerance to pod borers (*Helicoverpa armigera* and *Maruca testulalis*) and pod fly (*Melanagromyza obtusa*) (Saxena, 2008).

Wild relatives and Cytoplasmic Male Sterility (CMS) breeding system in pigeonpea improvement

In contrast to the extensive morphological variation exhibited in the cultivated species, SSR (Odeny et al., 2007), AFLP (Panguluri et al., 2006), and DArT (Yang et al., 2006) studies revealed very low molecular polymorphism within cultivated species compared to the wild relatives, emphasizing the need to broaden the genetic base of pigeonpea through cross breeding with wild relatives. Recent high-throughput Single Nucleotide Polymorphism (SNP) diversity studies using Illumina Golden Gate assay (Kassa et al., in preparation; Chapter 3 of this thesis) also confirmed the low genetic diversity of the cultivated genepool.

All wild species of pigeonpea have the same number of chromosomes (2n=22) and similar karyotype and their inter-specific hybrids showed chromosomal homology and complete chromosomal pairing. These wild species are widely utilized as desirable gene donors in pigeonpea breeding and improvement programs (Ariyanayagam et al., 1995; Saxena et al., 1996; Tikka et al., 1997; Wanjari et al*.,* 2000; Saxena & Kumar, 2003; Saxena et al., 2005a). Wild species of pigeonpea have also been utilized in breeding programs to develop cleistogamous lines (Saxena, 1992a), genetic dwarfs (Saxena and Sharma, 1995) and cytoplasmic male sterile (CMS) lines (Saxena, 2006).

Breeders at ICRISAT have been utilizing the partial out crossing nature of pigeonpea without use of cytoplasmic male sterile (CMS) system, to develop a high yielding hybrid cultivar (Reddy et al., 1978). Saxena et al. (1992b) developed a pigeonpea hybrid ICPH 8, which showed increased yield gain (30.5%) over the best performing pure-line control. Despite the success of releasing high yielding ICPH8 to farmers, it was not adopted effectively due to the high cost of seed production of the hybrid. This led to the development of the CMS breeding system, which was more efficient in large-scale hybrid seed production.

Cytoplasmic male sterility (CMS) is a phenotypic expression of incompatibility between nuclear and cytoplasmic genomes, and is a maternally inherited trait that has been successfully used as an efficient pollination control system in developing hybrid seed production (Havey, 2004). In most cases, CMS is caused by the interaction between the recessive nuclear genes and specific genetic factors housed in the mitochondrial genome which cause disfunctionality of the anthers and result in male-sterility. Fertility can be restored if dominant genes substitute the recessive nuclear genes or fertility-inducing factors arise in the mitochondrial genome. Three parents are required to maintain the CMS based breeding system: A male sterile line (A-line), Maintainer line (B-line) and Fertility restorer line (R-line). CMS systems can be caused by spontaneous mutation, intra-specific, inter-specific or inter-generic crosses. About 75 % of the CMS systems are a result of inter-specific and inter - generic crosses (Kaul, 1988).

The absence of CMS lines within pigeonpea germplasm led to the synthesis of CMS lines through interspecific crosses between cultivated pigeonpea and wild species using the cytoplasm genome of the wild parent and nuclear genome of the cultivated parent (Saxena, 2008). To date, seven CMS breeding lines have been developing by crossing wild parental lines with the cultivated pigeonpea parent.

- **A1 CMS** (*Cajanus sericeus* cytoplasm; Ariyanayagam et al., 1993): high yielding hybrid, but due to presence of pollen shedders within in female parent commercial seed production was not pursued.
- **A2 CMS** (*Cajanus scarabaeoides* cytoplasm; Saxena and Kumar, 2003): High yielding hybrid but the poor seed setting of the hybrid did not attract commercialization of the variety
- **A3 CMS** (*Cajanus volubilis* cytoplasm; Wanjari et al., 1991): lacks quality fertility restoration for hybrid seed production.
- **A4 CMS** (*Cajanus cajanifolius* cytoplasm; Saxena et al., 2005b): High yielding and stable and is being used extensively by breeders to develop commercial pigeonpea hybrids.
- **A5 CMS** (*Cajanus cajan* cytoplasm; Mallikarjuna and Saxena, 2005): a unique CMS system with the wild species, *Cajanus acutifolius,* used as male parent. The CMS system is maintained only by wild male parent and all the cultivated types restored the male fertility. Further investigations are being conducted to identify CMS lines in the cultivated types.
- **A6 CMS** (*Cajanus lineatus* cytoplasm): the naturally out-crossed partial male- sterile plant was observed in an open-pollinated population of *C. lineatus* (K. B. Saxena, unpublished data) and hybrids with unique phenotypes were observed. Currently this CMS source is in BC5F1 stage with perfect male-sterility maintenance system available (Saxena et al., 2010a).
- **A7 CMS** (*Cajanus platycarpus* cytoplasm; Mallikarjuna et al. 2006): Hormone-aided crossing coupled with embryo rescue techniques were implemented in crosses between *Cajanus platycarpus* and cultivated parental line. The studies on this CMS system are on-going.

Development of elite and high yielding pigeonpea hybrid cultivars

After decades of intense research, the world's first commercial CMS based legume (pigeonpea) ("magic pea hybrid") was developed by ICRISAT. This recent breakthrough was achieved through a CMS system of using the wild progenitor of pigeonpea, *Cajanus cajanifolius*, as a parental line (Saxena, 2009; Saxena et al., 2009; New Scientist, 2009). Unlike previous attempts of hybrid development (ICPH 8) in pigeonpea, this hybrid, named ICPH 2671 (Pushkal), is stable across diverse environments and has an excellent male fertility

restorer system. This hybrid gives about 30-40% yield increase over the best controls and showed resistance to wilt and sterility mosaic diseases (Saxena, 2009).

Disease and pest resistance as a major challenge in pigeonpea breeding

Wild species have coexisted with pests and pathogens on an evolutionary time scale and they have developed alleles conferring pest and pathogen resistance (Acosta-Gallegos et al., 1998). These natural defense mechanisms for diseases and pests have been lost during domestication and intense selection for agriculturally desirable traits such as high yield, improved nutritional quality and other desirable agronomic traits. Abiotic (e.g. drought, salinity) and biotic (e.g. diseases and pests) stresses constrain and adversely affect pigeonpea production and cause huge economic damage. The major diseases affecting pigeonpea production are Fusarium wilt (FW), sterility mosaic, *Phytophthora* blight disease and major pests causing severe damage are pod borer (*Helicoverpa armigera* and *Maruca vitrata*) and pod fly (*Melanagromyza obtusa*) (Minja et al. 2000).

Breeding strategies to tackle these problems in pigeonpea have been attempted by various researchers (Reviewed by Saxena, 2008). The breeding programs for developing disease resistant cultivars using resistance gene sources from cultivated pigeonpea germplasm did not succeed in controlling devastating pests (e.g. pod borer). The cultivated gene pool has low genetic polymorphism and lack resistance alleles. Alternative approaches of utilizing wild species as a source of resistance have showed promising results, as there is wider genetic diversity and presence of resistance genes in the wild gene pool. Most wild species have unique traits (e.g. presence of trichomes) that confer resistance to these diseases and pests (Aruna et al., 2005).

Wild relatives of pigeonpea such as *Cajanus scarabaeoides*, *C. sericeus, C. acutifolius, C. albicans, Rhynchosia aurea*, *R. bracteata* and *Flemingia bracteata* have shown high resistance to pod borer. *Cajanus platycarpus* has shown resistance to the most virulent race of phytoptora blight disease and some of the wild relatives of pigeonpea have shown high level of resistance to pod fly (*Melanagromyza obtusa*) and pod wasp (*Tanaostigmodes cajaninae*) (Sharma et al. 2003).

Advances in pigeonpea genomics

Next Generation Sequencing Technologies

Using the classical Sanger sequencing technology model plant species and major food crops (e.g. *Arabidopsis*, *Medicago*, *Lotus*, rice, wheat, maize, soybean etc) have been fully/partially sequenced and annotated and molecular markers have been accessed for these species. However in several crop species and particularly for "orphan" crops such as pigeonpea this was not achieved due to high cost of genome sequencing. With the advent of next generation sequencing/re-sequencing technologies, it has been achieved or potentially possible to enrich the genomic resources of neglected crop species including pigeonpea (Varshney et al. 2009b). Currently, there are three main commercially available next generation sequencing technologies (a) 454/FLX sequencing, (b) Solexa/ Illumina 1 GB SBS (sequencing-by synthesis) technology, and (c) AB SOLiD (Sequencing by Oligo nucleotide Ligation and Detection) technology (Varshney et al. 2009b).

Single Nucleotide Polymorphisms (SNPs) and GoldenGate Assay

Molecular genetic markers have been demonstrated to be powerful tools for genetic and genomic studies e.g. population genetic structure, genetic diversity studies, marker assisted plant breeding, phylogenetic and genome evolution studies. Various molecular marker technologies have been developed over the years such as Restriction Fragment Length Polymorphisms (RFLP) (Botstein et al. 1980), Random Amplified Polymorphic DNA (RAPD) (Williams et al. 1990), Simple Sequence Repeats (SSR) (Weber and May 1989), Amplified Fragment Length Polymorphisms (AFLP) (Mackill et al. 1995; Vos et al. 1995), Single Nucleotide Polymorphisms (SNPs) (Wang et al. 1998) and Diversity Arrays Technology (DArT) (Jaccoud et al. 2001).

Most of these markers suffer from constraints associated with low throughput and high cost of mining the marker. Sequence-based markers and high-throughput genotyping methods, such as SSRs and SNPs are now the most widely used technologies in modern plant genetic and genomic analysis, including anonymous marker systems like AFLPs and DArT (Appleby et al., 2009).

SNP markers are the most abundant and evenly distributed across the genomes of most plant species and have become an ideal marker system for genetic and genomic research in many crop species (reviewed by Rafalski 2002). SNPs are amenable to several high throughput genotyping platforms that are rapid and capable of genotyping up to a million SNP markers

(e.g GoldenGate assay). With the reducing cost of DNA sequencing and increasing availability of large sequence genomic resources, SNP markers are now the marker of choice for many genetic and genomic studies (Varneshey, 2009a; see Chapter 3 of this thesis).

Currently there are more than 30 different genotyping assays available (Gupta et al., 2008). The Illumina Company now provides two types of large scale, cost effective, fast and high throughput genotyping platform; the GoldenGate array for medium-density genotyping that contains 96, 192, 384, 768 or 1,536 SNPs per array and the Infinium array for large scale and high-density genotyping of up to 1 million SNPs per array (Fan et al. 2006; Varshney, 2009a).

The GoldenGate technology is now being used for genetic analysis in several crop species, for example a custom oligo pool assay (OPA) containing 1,524 SNPs per assay has been developed to estimate linkage disequilibrium (LD) and marker-trait association studies in barley (Rostoks et al. 2006). In soybean research, a custom OPA containing 384 SNPs per assay was used for construction of a high-density consensus linkage map and to genotype diverse germplasm panels (Hyten et al. 2006). A custom GoldenGate assay containing 1,536 SNPs has also been recently used to genotype two recombinant inbred line (RIL) populations and a panel of 154 diverse inbred lines in maize (Yan et al., 2009). A custom GoldenGate OPA assay containing 768 SNPs (Conserved Orthologous Sequences) has also been used to genotype diverse panels of domesticated pigeonpea and wild relatives and progress is under way to construct genetic linkage map of pigeonpea using these SNPs (discussed below)

Current status of pigeonpea genetic and genomic resources

Pigeonpea is among the less studied and previously neglected legume crops and often considered among "orphan" food crops in terms of genomic and molecular research. Through concerted effort of various partners, international legume researchers have recently advanced rigorous efforts in applying molecular and modern genomic technologies to enrich genomic resources in pigeonpea (Varshney et al., 2009b). The development of a genomic data set has potentially opened new possibilities in the application of marker/genomic-assisted selection in pigeonpea breeding and improvement programs and can also be utilized for genetic and genomic analyses.

A great advance has been made in enriching molecular and genomic resources in pigeonpea in the last few years through the Pigeonpea Genomics Initiative programs and the most intesive effort has been made through India-USA (University of California, Davis)

collaboration projects (Varshney et al., 2010). According to recently published information (Varshney et al., 2009b), Pigeonpea transcriptome resources have been impressively enriched with the generation of about 10,000 expressed sequence tags (ESTs) from Sanger sequencing and about 2 million short ESTs by 454/FLX sequencing technology. Progress is under way to use the next generation sequencing (NGS) of Solexa technology to sequence transcriptomes of ten pigeonpea genotypes that are parents of six mapping populations.

These genomic resources will accelerate candidate gene discovery of agronomic important traits and assist molecular marker development for breeding programs. An 11X-genome coverage bacterial artificial chromosome (BAC) library was generated that comprises about 87,590 BACend sequences (BESs). From BESs data set it was possible to mine molecular markers e.g. about 21,000 simple sequence repeat (SSR) were identified of which 6,698 SSRs are under analysis. Single nucleotide polymorphism (SNP) markers have been identified from 768 orthologous gene sequences (COS) using Oligo Pool Assay (OPA) in an Illumina GoldenGate platform and similarly 15,000 features of a diversity array technology (DArT) has been developed (Varshney et al., 2009b).

These SNP markers have been used for genetic and genomic diversity studies (see chapter 3 of this thesis) and progress is underway to anchor these SNPs for developing genetic linkage map of pigeonpea (University of California, Davis).

It has also been possible to enrich the genomic resources of disease resistance (R genes) across legume crop species. Using the Nucleotide Binding site- Leucine Rich Repeat (NBS-LRR) disease resistance homologs from the reference sequence of the legume model species *Medicago truncatula,* a set of 544 degenerate primers have been developed and used in pigeonpea. Hybridization using probes from representative NBS clades in pigeonpea BAC library identified 960 BAC clones, which were used for the construction of a physical map (Rosen et al., unpublished). These genomic resources form the basis for SNPs mining from NBS-LRR contigs, and potentially can be utilized as markers in molecular breeding of disease resistance (see Chapter 4 of this thesis). This fast paced field of genomics and molecular biology will in the near future have potential to make possible the application of genome wide selection in crops such as pigeonpea.

Chapter 2

Phylogenetic reconstruction of *Cajanus* and allied genera

Introduction

Plant systematics has been revolutionized in the last decade of the 20th century by the application of molecular biology and bioinformatics to elucidate evolutionary patterns and processes. Molecular plant systematics emerged as a science that involves the analysis of variation in organelle and nuclear DNA to construct plant phylogeny at all taxonomic levels (Soltis and Soltis, 1998).

Choice of a gene (DNA) for reconstructing phylogenies is based primarily on characteristics such as copy number, ability to resolve relationships, character congruence (low homoplasy), suitable number of parsimony-informative characters, and rate of evolution relative to the taxonomic group regardless of function of the gene (Mort et al., 2007). The gene should show enough variation (variable nucleotides) to mark the important cladogenic events, but should not be so variable that multiple changes in each site have erased the evolutionary information and give noise without phylogenetic signal (Doyle and Luckow, 2003).

Identifying an appropriate DNA marker for a given taxonomic level is a prerequisite for phylogenetic reconstruction (Kelchner, 2000). For example, (Shaw et al., 2005, 2007) profiled the relative rates of evolution of the non-coding regions of the single copy portion of the chloroplast genome between broad taxonomic lineages. The study found previously unexplored regions with better variability (phylogenetic utility) than regions indentified in earlier studies. This study also noted that the phylogenetic utility of a given genetic marker could vary significantly depending on the lineage under study (Shaw et al., 2007).

Chloroplast DNA sequences as well as nuclear ribosomal Internal Transcribed Spacer (*ITS*) sequences are very popular genomic regions for inferring plant phylogenies at different taxonomic levels due to their mode and tempo of molecular evolution (Olmstead and Palmer, 1994; Soltis and Soltis, 1998; Alvarez and Wendell, 2003; Shaw et al., 2005). The other genomic sources are mitochondrial DNA and nuclear DNA (other than ribosomal DNA) and the former is not suited for plant systematic studies due to a slow rate of sequence evolution and fast rate of structural changes (Palmer et al., 2000). The nuclear genome is also unsuited for molecular phylogenetic studies due to intrinsic problems associated with complex genomic architecture, high copy number, repetitive elements (e.g transposable elements), polyploidy,

hybridization and mostly nuclear genes which exist as gene families. These factors may cause difficulty in distinguishing sequences related by descent (orthologous) from paralogous (non-orthologous) elements (Small et al., 2004). Low copy nuclear genes, on the other hand, have shown potential for phylogenetic studies as they are relatively stable and less prone to the problem of paralogy (Small et al., 2004.). However, developing phylogenetically informative Low Copy Nuclear Genes for non-model organisms requires an investment in time and resources and the use of *ITS* sequences was strongly advocated for studies of species-level phylogeny or when use non-anonymous nuclear markers are required (Feliner and Rossello, 2007).

The chloroplast genome is usually a circular structure found in multiple copies in the chloroplast of a plant cell and ranges in size from 120 to 170 kilobase (kb) and usually contains about 120-130 genes. It exhibits a high degree of conservation in size, structure, gene content, linear order (collinearity) of the genes, and is mostly uniparentally inherited and non-recombinant (Downie and Palmer, 1992; Jansen et al., 2008).

The chloroplast genome of land plants, with few exceptions (e.g. Inverted Repeat Lacking Clade of Leguminosae species), contains two inverted repeats (IR) (c 25 kb each) that are mirror images of one another in terms of gene complement. The IR regions separate the large single-copy (LSC) and the small single-copy (SSC) regions of the genome (Shaw et al., 2007). The IR regions have low phylogenetic utility due to low mutational rate as compared to the single-copy regions (Wolfe, 1991; Gaut, 1998). Perry and Wolfe (2002) showed that the nucleotide substitution rate in single-copy regions is 2.3 times higher than that of the inverted repeats. Most of the popular and widely used (e.g *trnL-F* and *trnK/matK*) and first adopted (*rbcL*) regions are found in the single copy regions of the chloroplast genome (Shaw et al., 2007).

The chloroplast genome has three major functional units: 1) protein-coding genes 2) introns and 3) intergenic spacers. The latter two do not encode any protein and are collectively referred to as non-coding regions. Based on the reference chloroplast map of *Nicotiana* (Wakasugi et al., 1998) it was found that 43% of the single copy regions are non-coding regions. The non-coding region posesses fifteen introns and 92 intergenic spacers of which the introns comprise 10.6% while the intergenic spacers comprise the remaining part of the non-coding region respectively (Shaw et al., 2007).

The use of a non-coding region of the chloroplast genome for inferring plant phylogenies began with the seminal publication of Taberlet et al., (1991) and these are

currently the most widely used sources of sequence data for plant phylogenetics, phylogeographic, and population biology studies (Shaw et al., 2007). Many primers for both coding and non-coding regions of the chloroplast genome have been developed (e.g. Taberlet *et al*., 1991; Sang *et al*., 1997; Shaw *et al.,* 2005, 2007) which has greatly facilitated the use of chloroplast genome in phylogenetic studies. Molecular phylogenetic studies rely on rapidly evolving DNA regions (such as introns and intergenic spacers) to infer relationships between recently diverged lineages (Taberlet, 1996; Hale et al., 2004) whereas slowly evolving DNA regions are used to resolve phylogenetic relationships among distant lineages (Graham and Olmstead, 2000; Soltis et al., 2000). Moreover, the mosaic structure of introns, which consists of a conserved (helical structure) and a more relaxed hotspots (loop elements), make this region amenable for use at different taxonomic levels (Kelchner, 2002).

Empirical studies on the phylogenetic utility of rapidly evolving coding genes, non-coding introns and spacers and slowly evolving coding genes have challenged the widely accepted notion of molecular phylogenetic hypothesis which assumes a negative relationship between the breadth of the evolutionary history to be inferred and the overall mutational rate of the DNA region used (Borsch et al., 2003; Muller et al., 2006). These studies strongly suggested that it is not only the overall mutational rate of a given DNA region that determines the phylogenetic utility of the marker but, rather importantly, patterns of mutational dynamics and molecular evolution of individual sites (characters) of the given locus (Borsch et al., 2003; Muller et al., 2006).

DNA regions used in Legume Phylogeny

Early detailed phylogenetic studies of Leguminosae started with the plastid gene *rbcL* (Doyle, 1995; Kass and Wink, 1995, 1996; Doyle et al., 1997) following the pioneering use of this gene for phylogenetic studies of land plants (Chase et al., 1993). Subsequent legume molecular systematics and evolutionary biology studies have also used chloroplast genes at different taxonomic levels, e.g. *rbcL* gene (Doyle et al, 2000; Crisp et al., 2000; Kajita et al, 2001; Haston et al., 2005), the *matK* gene and flanking *trnK* intron (Hu et al., 2000; Lavin et al., 2001, 2003; Miller and Bayer, 2001; Luckow et al., 2003; Steele and Wojciechowski, 2003; Thulin *et al*, 2004; Wojciechowski *et al*., 2004; Wojciechowski, 2005), *trnL-F* region (Wojciechowski et al., 1999; Ireland et al., 2000; Luckow et al., 2000, Lavin et al., 2001, 2003; Bruneau et al., 2001; Pennington et al., 2001; Luckow et al., 2003; Haston et al., 2005); *ndhf* , rpS16 (Haston et al., 2005), rpl16 genes (Davis et al., 2002), *psbA-trnH* intergenic spacer

(Chandler et al., 2001) and rpoC1/rpoC2 genes (Liston and Wheeler, 1994). More recently a combined set of eight plastid regions (*trnL-F, rbcL, atpB, trnK/matK, rpl2, clpP, rps16, and ycf4*) were used to study the phylogenetic relationship of Phaseoloid legumes (Stefanovic´ et al., 2009). Similarly several legume phylogenetic and evolutionary studies have utilized the *ITS* region of the nuclear ribosomal genome (e.g. Kass and Wink, 1997a; Wojciechowski et al., 1999; Thompson et al., 2001; Hu et al., 2002; Ellison et al., 2006 and others).

DNA regions used in this study

To gain better understanding and insights on evolutionary history and phylogenetic relationships of the Cajaninae, DNA sequences from both chloroplast and nuclear genomes were included in this study. Three chloroplast regions (*trnL-F* spacer *matK* and *psbA-trnH* spacer) and one nuclear region (*ITS*) were initially tested for phylogenetic utility on the Cajaninae samples.

The *ITS* region

The structure and molecular evolution of Internal Transcribed Spacer (*ITS*) of nuclear ribosomal DNA (rDNA) facilitates the utility of this region for phylogenetic inferences at lower taxonomic levels in plants (Alvarez and Wendel, 2003). In plants there are two distinct tandem arrays of rDNA at one or more chromosomal loci, 5S rDNA and the 18S–5.8S–26S nuclear ribosomal cistron (Small et al., 2004). Each repeating unit of 18S-5.8S-26S cistron contains the *18S* gene, an internal transcribed spacer (*ITS 1*), the *5.8 S* gene; a second internal transcribed spacer (*ITS 2*) and the *26S* gene. While the mature rRNA coding regions of the gene are highly conserved, the two internal transcribed spacers (*ITS1* and *ITS2*) evolve rapidly making them suitable for infrageneric phylogenetic studies (Wendel *et al*., 1995).

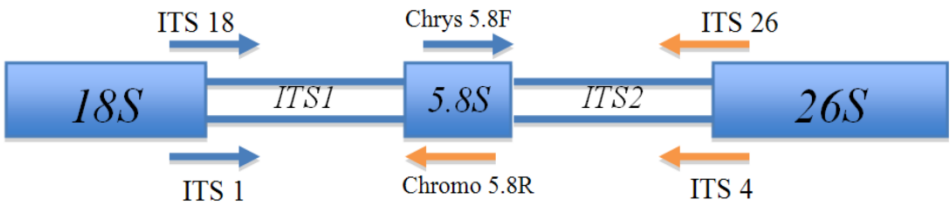


Fig. 2.1. Schematic diagram of the nuclear ribosomal DNA internal transcribed spacer region (ITS 1, *ITS4, ITS 18, ITS 26* are universal primers while *Chrys 5.8F* and *Chromo 5.8R* are internal primers used in this study.

The major properties of the *ITS* region that facilitated its wider usage for phylogenetic studies include bi-parental inheritance, high copy number and availability of universal primers. The *ITS* region can easily be amplified using the conserved regions of 18S, 5.8S and 26S genes (Fig. 2.1). The *ITS* region has intra-genomic uniformity (due to concerted evolution), inter-genomic variability and low functional constraints and usually has higher sequence variation than plastid genes (Alvarez and Wendel, 2003 Chen *et al*., 2004). High frequency crossing over or gene conversions have been implicated as a cause of concerted evolution (Baldwin *et al*., 1995). The major drawbacks of the *ITS* region that may hamper its use for phylogenetic studies include the problem of paralogy (due to duplication gene events and incomplete concerted evolution) and existence of pseudogenes (Alvarez and Wendel, 2003; Small et al., 2004).

Paralogy of the ribosomal repeats has not been known to cause problems in phylogenetic studies in papilionoid legumes (e.g., Bailey et al. 2003; Hughes et al. 2006), and *ITS* sequences have been successfully utilized in several legume phylogenetic studies, including *Lupinus* (Kass and Wink, 1997b; Ainouche and Bayer, 1999); *Lotus* (Allan and Porter, 2000); *Lens* (Mayer and Bagga, 2002); *Lathyrus* (Kenicer et al., 2005); *Caragana* (Zhang et al., 2009b); *Wajira* (Thulin et al., 2004); *Astragalus* (Wojciechowski et al., 1999)*; Brongniartieae* (Thompson et al., 2001); Millettieae (Hu et al., 2002); *Trifolium* (Ellison et al., 2006); *Phaseolus* (Delgado-Salinas et al., 2006).

The trnL-F region

The *trnL-F* region, which is a section of the *trnT-F* region, comprises the *trnL* intron and *trnL-F* spacer and is one of the most widely used chloroplast markers for phylogenetic studies in plants (Quandt et al., 2004). The *trnT-trnF* region is located in the large single copy of the chloroplast genome about 8 Kb downstream from *rbcL* gene (Borsch et al., 2003). There are three highly conserved tRNA genes coding for threonine (UGU), leucine (UAA), and phenylalanine (GAA) respectively. These tRNA genes are found in tandem and separated by two spacers (*TrnT-L* and *TrnL-F* spacers) and one group I intron (*TrnL* intron). The *TrnL* intron is found between the 5' exon (35bp) and 3' exon (50bp) of tRNA gene for leucine (Taberlet et al. 1991; Fig. 2.2 of this thesis).

Due to high rate of nucleotide variation in the two spacers and the group I intron, absence of gene rearrangements at tRNA genes and the ease of amplification using primers at the conserved tRNA genes, this region has been widely used in phylogenetic studies at

different taxonomic levels (e.g. Taberlet et al., 1991; Bakker et al. 1999; Richardson et al. 2000; Borsch et al. 2003).

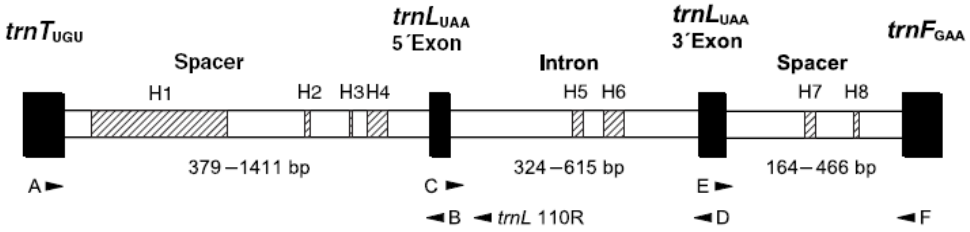


Fig. 2.2. Typical Structure of the *trnT-trnF* region in cpDNA of plants 9taken from Borsch et al., 2003). *tRNA* genes (*trnT* and *trnF* are each 73 bp long) and exons (*trnL-5'*is 35 bp and 3'is 50 bp) are represented by black boxes. The spacers and the Group I intron are illustrated by an empty bar with mutational hotspots are indicated in hatched boxes and letters A, B, C, D, E and F are primers.

Absence of conserved promoter regions and lack of secondary structural elements in the *trnL-F* spacer suggests that this region is less conserved and under less functional constraint than the *trnL* intron (Bakker et al., 2000).

Several legume studies have used the *trnL-F* region to infer phylogenetic relationships at inter-specific and inter-generic levels (e.g., Hurr et al., 1999; Wojciechowski et al., 1999; Ireland et al., 2000; Bruneau et al., 2001; Lavin et al., 2001, 2003; Brouat et al., 2001; Cubas et al., 2002; Luckow et al., 2000; 2003; Haston et al., 2005; Ellison et al., 2006; Torke and Schaal, 2008; Stefanovic´ et al., 2009)

The *psbA-trnH* spacer

The *psbA-trnH* intergenic spacer is located in the inverted repeat (IR) of the chloroplast genome and is among the most variable regions in the chloroplast genome (Sang et al., 1997; Hamilton, 1999; Miller et al., 2003). It is a popular region for plant population genetics and species level phylogenetics and has been shown to be more variable than *matK* and *trnL-F* regions and has been proposed as suitable for DNA barcoding studies (Kress et al., 2005). Although the *psbA-trnH* region contains a very high percentage of variable characters, it is comparatively short and may not yield enough parsimony-informative characters to infer a well-resolved phylogeny (Hamilton et al., 2003). However, this spacer is easily amplified using universal primers at the two flanking genes (*psbA* and *trnH* genes) and is usually coupled with other regions to infer well-resolved phylogenetic trees (Azuma et al., 2001; Hamilton et al., 2003). Short insertions/deletions (indels) are mostly found in the middle portion of this

spacer (Aldrich et al., 1988). This region was screened in this study but showed very little sequence variation among the Cajaninae taxa and further utilization of this region was not further continued.

The trnK/matK region

The *matK* gene is located within domain IV of the group II intron for the transfer RNA gene for Lysine (trnK $_{uuu}$) in the chloroplast genome and has a length of about 1500-1600 bp. It is ubiquitous in nearly all plants species (Neuhaus and Link, 1987; Ems et al., 1995). The *matK* gene and the flanking non-coding introns are co-transcribed together and *matK* protein is expressed (Chieba et al., 1996). The *matK* gene has been extensively used in several legume phylogenetic studies varying from family to intra-species phylogenetic relationships (Hu et al., 2000, 2003; Lavin et al., 2001, 2003; Steele and Wojciechowski, 2003, Wojciechowski et al., 2004 Riley-Hulting et al. 2004; Thulin et al. 2004; Delgado-Salinas et al. 2006). An evolutionary rates analysis of the family Leguminosae (Lavin et al. 2005) has shown that the core Phaseoleae lineages have the fastest rates of nucleotide substitution for the *matK* gene (Delgado-Salinas et al., 2006).

The above attributes of *matK* in resolving phylogenetic relationships in legumes and its fast evolutionary rates in Phaseoleae made it the prefered candidate region in this study. However, regardless of extensive attempts at sequencing this region using different pairs of internal primers, there was little nucleotide variation detected and the region failed to give resolution in the Cajaninae. This region was thus not utilized in this study. The low sequence variation of *matK* was also observed in phylogenetic studies of the American Cajaninae (Espert, personal comm.). It is difficult to speculate why the *matk* gene is conserved (or lacks variation) in the subtribe Cajaninae, because in the sister group (subtribe Phaseolinae), the *matk* gene has the fastest substitution rates (Lavin et al. 2005) thus rendering it very informative for infrageneric phylogenetic reconstruction.

Materials and Methods

Taxon sampling

Species from six genera of the subtribe Cajaninae collected from Africa, Asia (particularly from the Indian sub-continent) and Australia were included in our analyses. The genera included in the study are *Cajanus, Eriosema, Rhynchosia, Flemingia, Dunbaria* and *Bolusafra*. Twenty species of *Cajanus*, ten species of *Eriosema*, seven species of *Rhynchosia*,

five species of *Flemingia*, one species of *Dunbaria*, one species of *Bolusafra* and two *ITS* sequences were included from the genbank (*Rhynchosia hauthalii*-EU499367 and *Flemingia macrophylla*- FJ980288). Species of *Glycine* and *Vigna* were used as outgroups based on previous phylogenetic information (Table 2.1). Most of the *Cajanus* species were collected from Australia and Asia, as these regions are endemic sites for 13 and 18 *Cajanus* species respectively (van der Maesen, 1980).

PCR amplification and Sequencing of *ITS* and *trnL-F* spacer

Total genomic DNA from leaf material was extracted using a modified hot CTAB protocol of Doyle and Doyle (1987). Polymerase Chain Reaction (PCR) amplifications were carried out in 50 µL volumes using ThermoHybaid PCR Thermal Cycler using the following PCR conditions; 95˚C for 45 seconds, 53-59˚C for 45 seconds and 72˚C for three minutes repeated for 30-35 cycles and a 10 minutes 72˚C extension was included at the end of the PCR reaction. (Annealing temperature and number of cycles were adjusted depending on the template and primers).

The *trnL-F* spacer region was amplified using primers 'e' and 'f' as described by Taberlet et al., 1991. The *ITS* region was amplified by PCR using primers '*ITS* 1' and '*ITS* 4' (White *et al.* (1990), and internal forward primer 'Chrys 5.8F' [5'GACTCTCGGCAACGGATATC3'] (Barker *et al*., 2009) and reverse primer 'Chromo5.8R' [5'GATTCTGCAATTCAC3'] (Barker *et al*., 2005).

The PCR products were checked by electrophoresis in a 1% agarose gel and visualized by ethidium bromide on a UV trans-illuminator. PCR products were cleaned with Promega Magic PCR PrepsTM, QIAGEN© QIAquickTM or Promega Wizard ® kits following the standard protocols of the manufacturer.

Cleaned PCR products were directly sequenced using an ABI Prism BigDye Terminator v3.1 Ready Reaction Cycle Sequencing Kit (Applied Biosystems) following the manufacturers Protocol. Sequencing reactions were done in 20µl reaction volume containing: 2µl sequence mix, 3µl sequence buffer, 0.5µl primer, 2µl purified DNA template and 12.5µl nuclease free water. The amounts of template and water were adjusted depending on the quality of the DNA template with the final reaction volume of 20µl.

*Table 2.1. List of taxa used in this study with voucher and locality details. [Key to collectors: VDM=Van der Maesen, M & S= Muasya & Stirton An * indicates sequences retrieved from the Genebank]*

	Taxon	**Voucher**	**Locality**
1	*Bolusafra bituminosa* Kuntze	-	South Africa: Western cape
2	*Cajanus acutifolius* (Benth) Maesen	VDM 7834	India ex Australia
3	*Cajanus acutifolius* (F.Muell.) Maesen	T. Handasyde & J. Russell-Smith THO1 180	West Australia: Bachsten Creek area
4	*Cajanus acutifolius* (F.Muell.) Maesen	T. Handasyde TH 00 379	West Australia: Buccaneer Archipelago
5	*Cajanus albicans* (Wight & Arn) Maesen	VDM 7835	India
6	*Cajanus cajan* (L) Millsp	A.A. Mitchell 6350	West Australia: Frank Wise Research Institute
7	*Cajanus cajan* (L) Millsp.	VDM 7737	West Africa: Benin
8	*Cajanus cajanifolius* (Haines) Maesen	VDM 7847	India
9.	*Cajanus cinereus* (F. Muell.) F. Muell.	VDM 7833	India
10	*Cajanus cinereus* (F. Muell.) F. Muell.	J.E. Wajon 451	West Australia: Hearson Cove
11	*Cajanus cinereus* (F. Muell.) F. Muell.	M.E. Trudgen	West Australia: Fortescue Botanical district
12	*Cajanus crassicaulis* Maesen	J. Grimes, D. Murphy & CAJANUS Hohnen JG 3562	West Australia: Red Rock Creek, Hall District
13	*Cajanus crassicaulis* Maesen	J. Solomon 799	West Australia: Bungle National Park
14	*Cajanus hirtopilosus* Maesen	CAJANUS Brockway CB 63	West Australia: E Kimberley

	Taxon	Voucher	Locality
15	*Cajanus lanceolatus* (W.Fitzg.) Maesen	R.O. Makinson 1676	West Australia: W Kimberley, Gardner District
16	*Cajanus latisepalus* (Reynolds & Pedley) Maesen	K.F. Kenneally K 11864	West Australia: Berkeley River
17	*Cajanus latisepalus* (Reynolds & Pedley) Maesen	T. Handasyde TH99 024	West Australia: IBRA region
18	*Cajanus lineatus* (Wight & Arn) Maesen	VDM 7839	India
19	*Cajanus marmoratus* (Benth.) F.Muell.	G. Byrne 1668	West Australia: Mt Phyre on Anna Plains station
20	*Cajanus mollis* (Benth) Maesen	VDM 7838	India
21	*Cajanus pubescens* (Ewart & Morrison) Maesen	G. Byrne 1133	West Australia: Snake Creek
22	*Cajanus reticulatus* (Dryand.) F. Muell var. *grandifolius* (F. Muell) Maesen	VDM 7837	India ex Australia
23	*Cajanus reticulatus var*. grandifolius (Benth) Maesen	A.A. Mitchell 6703	West Australia: Molly springs, E Kimberley
24	*Cajanus scarabaeoides* (L.) Thouars var. *scarabaeoides*	VDM 7845	India
25	*Cajanus sericeus* (Benth. ex Baker) Maesen	VDM 7846	India
26	*Cajanus viscidus* Maesen	A.N. Start 1568	West Australia: IBRA region
27	*Dunbaria ferruginea* Wight & Arn	VDM 7843	India
28	*Eriosema burkei* Benth. ex Harv. & Sond.	M & S 3781	South Africa
29	*Eriosema griseum* Baker	VDM 7696	Benin
30	*Eriosema kraussianum* Meisn.	M & S 3708	South Africa

	Taxon	**Voucher**	**Locality**
31	*Eriosema lucipetum* C.H.Stirt.	M & S 3776	South Africa
32	*Eriosema psoraleoides* (Lam.) G.Don	VDM 7844	India
33	*Eriosema rossii* C.H.Stirt.	M & S 3685	South Africa
34	*Eriosema salignum* E.May.	M & S 8841.6	South Africa
35	*Eriosema simulans*	M & S 3778	South Africa
36	*Eriosema squarrosum* Walp.	M & S 3818	South Africa
37	*Eriosema umtamvunense* C.H.Stirt.	M & S 8841.4	South Africa
38	*Flemingia lineata* (L.) W.T.Aiton	I. Cowie & Stewart IC 4212	W. Australia: Kimberley region
39	*Flemingia parviflora* Benth	A.A. Mitchell 3388	W. Australia: Mitchell Plateau
40	*Flemingia stricta* (L.) *W.T. Aiton*	VDM 7840	India
41	*Flemingia strobilifera* (L.) W.T. Aiton	VDM 7841	India
42	*Flemingia trifoliastrum* Domin	A.A. Mitchell 5458	W. Australia: North Kimberley
43	*Flemingia macrophylla** (Willd.) Merr.	-	Gene Bank (FJ980288)
44	*Flemingia glutinosa**	-	Gene Bank (FJ980289)
45	*Rhynchosia aurea* (L.) DC.	VDM 7933	India
46	*Rhynchosia australis* Benth.	K.F Kenneally K 11824	W. Australia: Low Rocks Nature Reserve
47	*Rhynchosia bungarensis* Maesen	M. Maier & K. McCreery 551	W. Australia: Roy Hill Road
48	*Rhynchosia minima* (L.) DC.	VDM 7896	India
49	*Rhynchosia orthobotrya* Harms	VDM 7674	West Africa: Benin

	Taxon	**Voucher**	**Locality**
50	*Rhynchosia verdcourtii* Thulin	VDM 7842	India
51	*Rhynchosia sp*	VDM 7691	West Africa: Benin
52	*Rhynchosia hauthalii** Harms	-	Gene Bank (EU499367)
53	*Glycine microphylla** *(Benth.) Tindale	-	Gene Bank (EF517918)
54	*Glycine max** (L.) Merr.	-	Gene Bank (FJ980442)
55	*Glycine tomentella** Hayata	-	Gene Bank (AY433889)
56	*Glycine pullenii** B.E.Pfeil, Tindale & Craven	-	Gene Bank (AY433928)
57	*Glycine stenophita** B.E.Pfeil & Tindale	-	Gene Bank (AY433934)
58	*Vigna luteola** (Jacq.) Benth.	-	Gene Bank (AY583519)

Sequencing reactions were carried out using a ThermoHybaid PCR Sprint Temperature Cycling System using the following conditions: 95°C for 45 seconds, 50°C for 45 seconds and 60°C for three minutes repeated for 30-35 cycles.

The sequencing reaction was precipitated as follows; adding the 20µl sequence product to 50µl 100% ethanol, 2µl 125mM EDTA solution and 2µl 3M sodium acetate solution. The mixture was votexed and left to stand for 15 minutes at room temperature and then centrifuged at 10,000 rpm for 15 minutes. The supernatant was carefully discarded leaving the pellet undisturbed. One hundred and fifty microlitres of 70% ethanol was added and samples were centrifuged for additional ten minutes. The supernatant was pipetted off from the pellet. The pellet was left to air-dry, and re-suspended in Hi-Dye formamide. Sequencing was done on an ABI 3100 automated sequencer at Rhodes University's Sequencing Unit.

Sequence editing and alignment

Sequence data were curated, edited, and contigs assembled using Sequencher™ v.4.2 (Gene Codes Corporation). Sequence alignments were done by Multiple Alignment Program, MAFFT version 6 (Katoh et al., 2005) and manually edited using MacClade version 4.06 (Maddison and Maddison, 2000). As explained by authors (Katoh et al., 2005) "MAFFT is one of the fastest methods among the currently available multiple alignment tools, and used in several projects, such as Pfam, ASTRAL and MEROPS". The alignment was constructed initially by the progressive method and then refined by the iterative refinement methods of MAFFT. The most accurate strategy (FFT-NS-i) option of MAFFT was selected. The FFT-NS-i method optimizes the weighted sum-of-pairs (WSP) score using an approximate group-to-group alignment algorithm and the tree-dependent restricted partitioning technique. The default parameters of MAFFT such as gap penalties and scoring matrix were applied. The alignment was confirmed using manual alignment in MacClade and was easy to do "by eye". The full alignment of the *ITS* and *trnT-F* spacer sequences are provided in Appendices 1 and 2 respectively.

Phylogenetic Analyses

Three data sets were analyzed to infer the phylogenetic relationships of Cajaninae: (1) an *ITS* sequence data set of 50 taxa with *Glycine* and *Vigna* as out-groups and (2) the plastid *trnL-F* spacer sequence data set of 36 taxa with *Glycine* as an out-group and (3) the combined *ITS* and *trnL-F* spacer sequence data set of 28 taxa with *Glycine* as an out-group. All the gaps in

the alignments were treated as missing data. The difference in number of samples used for *ITS* and *trnL-F* spacer regions is due to differences in PCR and sequencing success rates for the two regions. The reason we excluded gaps is that their positions are often difficult to determine, especially when analyzing a broad range of taxa and/or highly diverged sequences (Kawakita et al., 2003), such as our *ITS* data. However the *trnL-F* data are not that divergent and code-able gaps could potentially be used.

Parsimony analysis

Parsimony analyses were performed with PAUP* 4.0b10 (Swofford, 2002). Parsimony analyses weighted all characters equally and treated states as unordered. Gaps were treated as missing data. A full Heuristic search was performed with 1000 random addition sequence replicates using tree bisection reconnection (TBR) branch swapping, MulTrees in effect, and with Fitch parsimony (Fitch, 1971). Clade Support was evaluated through bootstrapping (Felsenstein, 1985) with 500 replicates using TBR branch swapping. Incongruence among data partitions (*ITS* and *trnL-F* spacer) was evaluated with the partition homogeneity test of Farris et al. (1994) implemented in PAUP* 4.0b10 (Swofford, 2002). The partition homogeneity test used 1000 re-samplings under the parsimony criterion with only variable characters included. For *ITS* data set *Glycine* and *Vigna* were used as an out-group while for *trnL-F* and combined data sets only *Glycine* was used as an out-group.

Bayesian analysis

As the Bayesian inference method is based on explicit models of DNA evolution, the program MrModelTest (Nylander, 2004) was used to determine the model of DNA substitution that best fit the data. Bayesian inference of phylogeny using Markov Chain Monte Carlo (MCMC) method was computed with MrBayes 3.1.1 (Huelsenbeck and Ronquist 2001). The analysis was conducted with four Monte Carlo Markov Chains with three heated and one cold. The analysis was run for 1,000,000 generations and sampled every 100th generation. The stabilization of the log likelihood values was examined to determine the "burn-in". These "burn-in" generations were excluded when constructing the Bayesian Inference trees. Posterior probabilities (PP) were estimated by constructing a 50% majority rule consensus tree in PAUP* 4.0b10 (Swofford, 2002). Bayesian Inference trees were visualized with Figtree v1.3.1.

Results

Sequence Information

Data set statistics for the three data sets (*ITS*, *trnL-F* spacer and the combined *ITS* and *trnL-F* spacer) as well as tree statistics of this study are shown in Table 2.2. The ITS matrix contained 662 aligned characters, whereas the *trnL-F* spacer data set contained 373 aligned characters. The Incongruence Length Difference (ILD) test did not detect any incongruence between the chloroplast and nuclear in this study (P=0.17), so the two data sets were combined. The combined *ITS* and *trnL-F* spacer data set contained 1035 aligned characters.

The *ITS* region contained more informative sites (271 sites; 41%); the *trnL-F* spacer sequences contained only 8.3% (31 sites) parsimony-informative characters. The combined data set of *ITS* and *trnL-F* spacer has 29% (300) parsimony- informative sites. MrModel Test identified the best DNA substitution model for the *ITS* data as the GTR model. For the *trnL-F* spacer data, the best DNA substitution model was identified as GTR+G.

Table 2.2. Sequence characteristics and tree statistics of ITS, trnL-F spacer sequences and the combined data set of ITS and trnL-F spacer sequences

DNA region	***ITS***	***trnL-F spacer***	***ITS + trnL-F spacer***
No. of taxa	50	36	50
No. of characters	662	373	1035
No. of parsimony Uninformative variable characters	80	42	124
No. of parsimony Informative characters	271	31	300
% Informative characters	41%	8.3%	29%
No of equally most parsimonious trees	94	276	46
Tree length	84	46	125
CI*	0.65	0.82	0.69
RI**	0.89	0.92	0.90

CI*= consistency index; RI**=retention index.

Phylogenetic analysis of the nuclear *ITS* data

For the taxa examined in this study, both parsimony and bayesian phylogenies of the *ITS* data set supported the monophyly of the subtribe Cajaninae (100% bootstrap support and 1.00 posterior probability). The strict consensus tree of the 94 most parsimonious trees of the maximum parsimony analyses and majority rule consensus tree of the Bayesian inference method are shown in Figures 2.3 and 2.4. Both methods showed similar phylogenetic relationships with no major topological differences. Parsimony analysis and Bayesian inferences phylogenies resolved six major clades in the ingroup with strong bootstrap support (BS) and Bayesian Posterior probability (PP) for these clades. The *Cajanus* Clade (100% BS and 1.0 PP) comprises all except two *Cajanus* species (*Cajanus mollis* and *Cajanus marmoratus*) which showed sister relationships with other genera. *Cajanus* is thus not monophyletic.

The *ITS* phylogeny (Fig. 2.4.) resolved *Cajanus scarabaeoides* as the earliest branching or basal lineage within this *Cajanus* clade. The second Clade (100% BS and 1.0 PP) comprises the *Cajanus-Dunbaria* complex; the third Cade (100% and 1.0 PP) corresponds to the *Cajanus-Rhynchosia* complex; the fourth Clade (94% BS and 1.0 PP) corresponds to the core *Rhynchosia* clade, which excludes *Rhynchosia orthobotrya* and *Rhynchosia aurea*. The core *Rhynchosia* clade shows sister relationship with the *Eriosema* clade (75% BS and 0.98 PP). The fifth Clade (100% and 1.0 PP) corresponds to the monophyletic *Eriosema* group and the sixth Clade (100% and 1.0 PP) corresponds to the monophyletic *Flemingia* clade. In bayesian analyses, the only species of the South African monotypic genus *Bolusafra* was resolved as sister group to the *Rhynchosia* and *Eriosema* clades with weak nodal support (55% BS and 0.67 PP). The *Flemingia* clade is shown to be the most basal group in the subtribe Cajaninae.

The Bayesian analyses resolved *C. cajan* (the domesticated pigeonpea species from Australia) to be sister to the putative progenitor *C. cajanifolius* from India with modest nodal support (0.55 PP), but parsimony did not resolve this sister relationship. Both parsimony and Bayesian trees have resolved *C. viscidus* as sister group to *C. hirtopilosus* with strong support (87 % BS and 1.0 PP). Bayesian analyses resolved the sister relationship between *C. latisepalus* and *C. crassicaulis* (0.81 PP) and another accession of *C. latisepalus* shows close relationships with this sub-clade (0.65 PP) (all these three samples are Australian species). The two Australian samples of *C. cinereus* are not resolved as monophyletic.

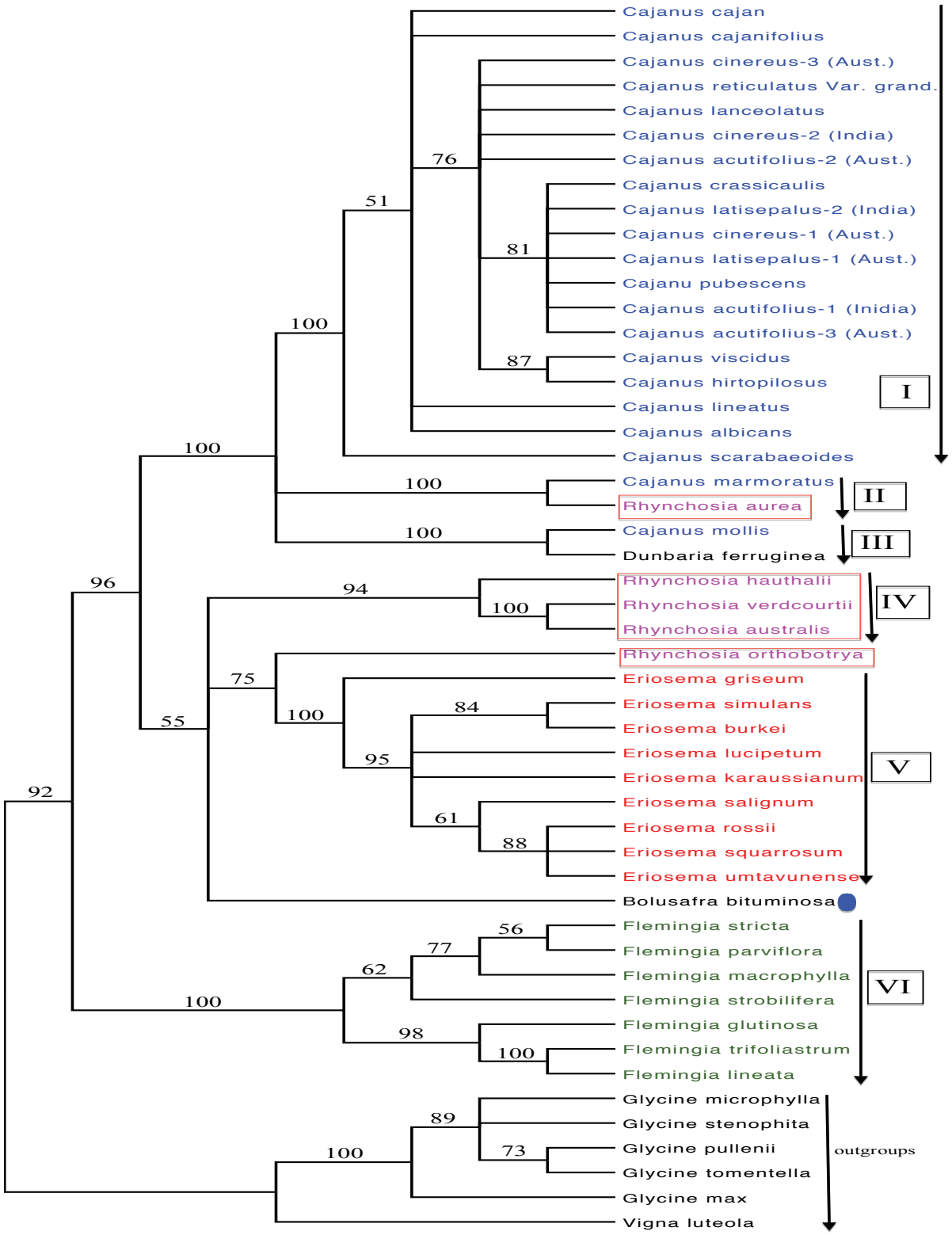


Fig. 2.3. Strict consensus tree of the 94 most parsimonious trees inferred from *ITS* sequences for 50 taxa. Numbers above branches indicate bootstrap support (>50%). The red boxes indicate *Rhynchosia* species, the monotypic *Bolusafra* is indicated by a blue dot and other clades are indicated by vertical arrows (I- Core *Cajanus*; II-*Cajanus-Rhynchosia*; III- *Cajanus-Dunbaria*; IV- core-*Rhynchosia*; V- *Eriosema*; VI-*Flemingia*).

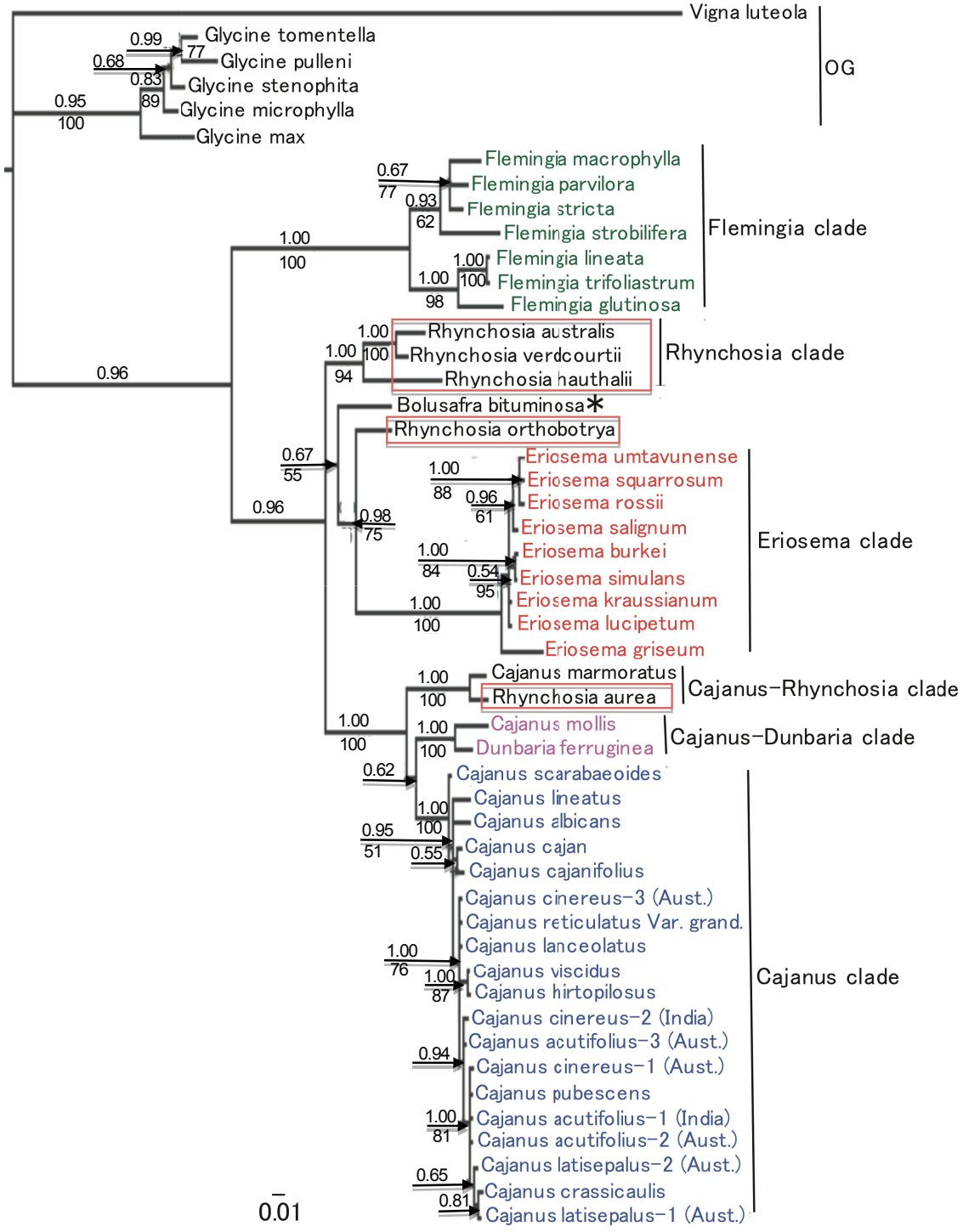


Fig. 2.4. Bayesian consensus tree from analysis of the nuclear *ITS* sequences of Cajaninae samples. Numbers above the branches indicate posterior probability and numbers below the branches are parsimony bootstrap values. The red boxes indicate *Rhynchosia* species, the monotypic *Bolusafra* species is indicated by star and other clades are indicated by vertical bar. Vigna and Glycine spp. were used as outgroups (OG).

Similar distant relationships between multiple samples of *C. acutifolius* are also observed. Among the five allied genera of *Cajanus* included in this study, *Dunbaria* and some species of *Rhynchosia* showed the closest sister relationships with the core *Cajanus* clade. The *Cajanus-Dunbaria* clade (100% BS and 1.00 PP) comprises the two Indian species *Cajanus mollis* and *Dunbaria ferruginea*.

The *Cajanus-Rhynchosia* clade (100% BS and 1.00 PP) comprises the Australian species *Cajanus marmoratus* and the Indian species *Rhynchosia aurea*. Core *Rhynchosia* clade (94% BS and 1.00 PP) comprises the Indian *Rhynchosia verdcourtii*, the Australian *Rhynchosia australis* and the gene bank accession *Rhynchosia hauthalii* (EU499367) sourced from South America. The Indian species of *Rhynchosia orthobotrya* was not part of the core *Rhynchosia* clade but rather basal to the *Eriosema* clade (75% BS and 0.98 PP). The South African *Bolusafra bituminosa* showed the closest relationships (55% BS and 0.67 PP) with the *Rhynchosia-Eriosema* clade.

The *Eriosema* clade is a strongly supported monophyletic group (100 PS and 1.0 PP) of all South African and one West African *Eriosema* species. The West African *Eriosema griseumbaker* is the most basal species in this clade. *Eriosema simulans* is sister group to *Eriosema burkei* (84% BS and 1.0 PP). Both parsimony and Bayesian analyses grouped *Eriosema rossii*, *Eriosema umtammvunense* and *E. squarrosum* as a sub clade (88% BS and 1.0 PP) and *E. salignum* is sister to these taxa (61% and 0.96 PP).

The monophyletic *Flemingia* clade (100% PS and 1.0 PP) was the most basal clade among the Cajaninae studied here and parsimony and Bayesian analysis resolved similar relationships in this clade. *Flemingia trifoliastrum* was sister group to *F. stricta* (100% BS and 1.0 PP) and the GeneBank accession of *F. glutinosa* was most closely related to this sub clade (98 % BS and 1.0 PP). Similarly both analyses grouped *F. lineata*, *F. strobilifera* and the gene bank accession *F. macrophylla* together (77% BS and 0.67 PP) with *F. parviflora* as the closest species to this subclade (62% BS and 0.93 PP). Parsimony further resolved the sister relationship of *F. lineata* and *F. strobilifera* with a modest bootstrap support (56%).

Phylogenetic analysis of the chloroplast *trnL-F spacer* data

Both parsimony and Bayesian inference methods showed similar phylogenetic relationships without major topological differences. The Bayesian inference tree showing nodes retained by both parsimony and Bayesian inference is presented in Fig. 2.5 with both posterior probability (PP) and parsimony Bootstrap Support (BS) values.

Despite fewer parsimony-informative characters, the chloroplast *trnL-F* sequence data resolved 4 major clades of the ingroup (Fig. 2.5). The first clade (90% BS and 0.90 PP) corresponds to the genus *Flemingia*. The second clade (98% BS and 1.00 PP) corresponds to *Eriosema*. The third clade (98% BS and 1.00 PP) comprises the core *Rhynchosia* clade, excluding *Rhynchosia aurea* and the West-Africa *Rhynchosia* sp. from Benin. The latter two *Rhynchosia* species were embedded within *Cajanus*. The fourth clade (77% BS and 0.99 PP) comprises a paraphyletic *Cajanus,* with *Dunbaria* embedded within. Unlike the *ITS* data set, the chloroplast DNA data lacks sufficient variation to resolve infraspecific relationship within *Cajanus*. It only resolved one sub-clade comprising nine *Cajanus* samples with very weak (0.76 PP) nodal support. Interestingly it showed sequence divergence of two samples of *Cajanus reticulatus* var. *grandifolius* collected from India (originally came from Australia) and Australia respectively, which suggests genetic divergence between the two regions.

As retrieved by *ITS* data, the South African *Bolusafra bituminosa* is placed sister to the (*Rhynchosia-Eriosema*) clade (67% and 0.74 PP). The West-African *Eriosema griseum*, which was revealed as most basal within this clade in the *ITS* data set, is shown to be sister species to the Indian *Eriosema psoraleoides* (66 BS and 0.96 PP). There is also a sister relationship (72% BS and 0.98 PP) between the two South African species, *E. salignum* and *E. rosii*.

Phylogenetic analysis of the Combined *ITS* and *trnL-F spacer* data

The combined analysis resolved a similar phylogenetic relationship to that of the *ITS* phylogeny. There are few instances of differences in the level of nodal support (BS and PP) between the ITS and the combined phylogenetic trees (Fig. 2.6). The only difference observed was the lack of resolution of the sister relationship of *C. cajan* and *C. cajanifolius* in the combined data set while this relationship was retained in Bayesian analysis of the *ITS* phylogeny (Fig. 2.4). Both parsimony and Bayesian inference methods showed similar phylogenetic relationships without major topological differences The Bayesian inference consensus tree based on combined data is presented in Fig. 2.6 and the nodes retained by both parsimony and Bayesian inference are indicated with both posterior probability (PP) and parsimony Bootstrap Support (BS) values.

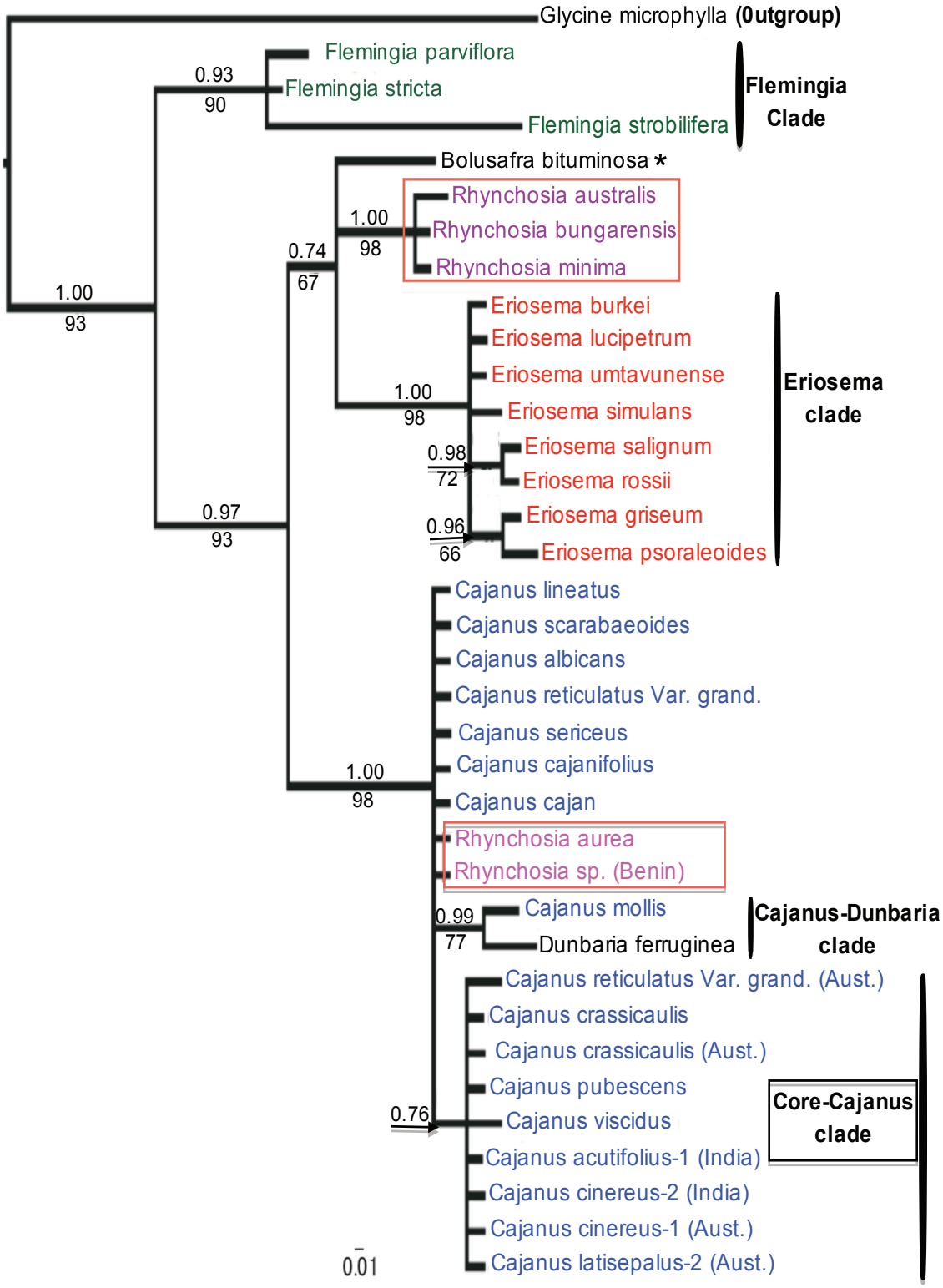


Fig. 2.5. Bayesian consensus tree from analysis of the chloroplast *trnL-F spacer* sequences of Cajaninae samples. Numbers above the branches indicate posterior probability and numbers below the branches are parsimony bootstrap values. The red boxes indicate *Rhynchosia* species, the monotypic *Bolusafra* is indicated by star and other clades are indicated by vertical bars. *Glycine microphylla* was used as outgroup.

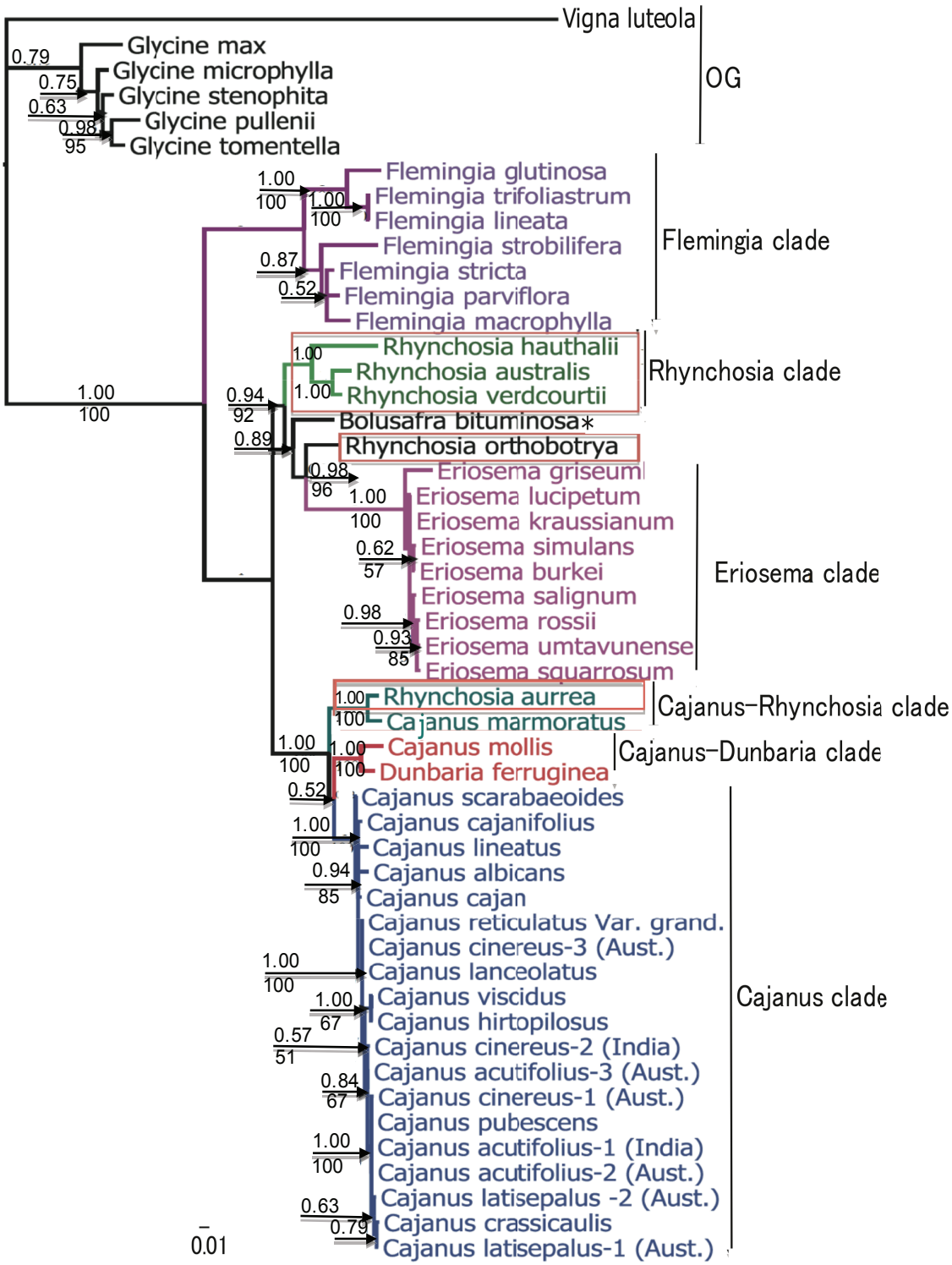


Fig. 2.6. Bayesian consensus tree from analysis of the nuclear and chloroplast combined sequences of Cajaninae samples. Numbers above the branches indicate posterior probabilitity and numbers below the branches are parsimony bootstrap values. The red boxes indicate *Rhynchosia* species, the single *Bolusafra* species is indicated by star and other clades are indicated by vertical bars. *Vigna* and *Glycine* spp. were used as outgroups (OG).

Discussion

Monophyly of subtribe Cajaninae

Sampling of Cajaninae in previous family-wide molecular phylogenetic studies of legumes was limited and little is known about the systematic relationships of the subtribe. The comprehensive plasid *matK* phylogeny of legumes (Wojciechowski et al., 2004) did not include any representative of the subtribe Cajaninae and similarly the comprehensive analyses of legumes based on the chloroplast *rbcL* region (Kajita et al. 2001) sampled only seven taxa from this large group. This study addresses this limitation by presenting results of a well-sampled cpDNA and nrDNA phylogeny.

Both individual and combined data from the plastid and nuclear regions supported the monophyly of the subtribe Cajaninae. This study has also shown the non-monophyly of the genera *Cajanus* and *Rhynchosia* while supporting the monophyly of *Eriosema* and *Flemingia*. Among the five allied genera of *Cajanus* included in this study, the single sample of *Dunbaria* and one *Rhynchosia* species are nested within *Cajanus* (Fig. 2.4).

Given the lack of sampling of *Dunbaria* and the large size of *Rhynchosia* in this study, it is not possible to speculate whether other species of these two genera will nest within *Cajanus*. Sister relationships were evident between the Indian species *Cajanus mollis* and *Dunbaria ferruginea* and also the Australian *Cajanus marmoratus* and the Indian *Rhynchosia aurea*. The core *Rhynchosia* Clade (94% BS and 1.00 PP) comprises the Indian *Rhynchosia verdcourtii*, the Australian *Rhynchosia australis* and the GeneBank accession *Rhynchosia hauthalii* from South America, while the African species (*Rhynchosia orthobotrya*) was placed out of the Rhynchosia clade and showed a close relationship to the African *Eriosema* clade (75% BS and 0.98 PP).

The close relationship of South African *Bolusafra bituminosa* with the *Rhynchosia-Eriosema* clade is concordant with the description of *Bolusafra* as "viscid *Rhynchosia*-like vine" (Lackey, 1981; Moteetee and Van Wyk, 2006). *Eriosema* was resolved as monophyletic group with strong branch support (100 PS and 1.0 PP).

The *ITS* data set resolved all the South African species as a sub-clade while the West-African *Eriosema griseum* was the most basal species in this clade. Chloroplast DNA resolved sister relationship between this West-African *Eriosema* species and the Indian *E. psoraleoides*. The *Flemingia* clade (100% PS and 1.0 PP) was also monophyletic and is the most basal clade.

Comparing molecular phylogeny with morphology

This study corroborates with the previous morphological, anatomical, chemotaxonomic and DNA fragment mapping studies (Lackey, 1981; Doyle and Doyle, 1993) which revealed the sister relationship of *Rhynchosia* and *Eriosema.* The monophyly of *Eriosema*, which was suggested by early DNA studies, (Doyle and Doyle, 1993; Bruneau et al., 1995) was also supported by this study but in contrast to previous molecular studies, the present study did not support the monophyly of *Rhynchosia*.

This study also clearly indicates that *Cajanus* is not monophyletic as currently circumscribed. *Dunbaria* and elements of *Rhynchosia* are embedded within *Cajanus*, which is in agreement to previous PCR-RFLP study (Lakshmi et al., 2000).

Similarly this study corroborates Lakshmi et al. (2000), who also showed the distant relationship of *Flemingia* to other genera of Cajaninae. In the taxonomic revision of *Dunbaria,* van der Maesen (1998) found that some species of *Dunbaria* showed the facies of climbing *Cajanus* and suggested the possibility of placing *Dunbaria* in a secondary or tertiary genepool of cultivated pigeonpea.

The overall phylogenetic relationship within the genus *Cajanus* in this study partially supports the sectional classification of the genus, based on morphology, as proposed by van der Maesen (1986). The Bayesian analysis of the *ITS* data placed the species belonging to section *Cajanus* (*C. cajan* and *C. cajanifolius*) together. Section *Fruticosa* (*C. lanceolatus, C. reticulatus* and *C. acutifolius*) were also placed together in a larger clade with other *Cajanus* species. However, this study is incongruent with seed protein analyses (Panigrahi et al., 2007), which indicated the genetic divergence between C. *lineatus* of section Atylia and *C. cajan* of section Cajanus.

Conclusions

Our results from individual data sets of nrDNA and cpDNA as well as the combined (nuclear and chloroplast genomes) data set resolved the subtribe Cajaninae as monophyletic (Figures 2.3-2.6). However, only six genera were sampled and the inclusion of other genera could provide novel relationships. This supports previous findings from morphological, alpha taxonomic and few molecular studies (Lackey 1978, 1981; van der Maesen, 2003; Doyle et al., 2000; Kajita et al., 2001).

The genus *Cajanus* needs to be revised as *Dunbaria* and some *Rhynchosia* species appear to be embedded within it. This hypothesis has been suggested by previous taxonomic works (e.g. van der Maesen, 1998). Taxonomic revision and generic delimitation has paramount importance especially in hunting for potential gene sources from close wild gene pools for pigeonpea improvement.

This study corroborates previous morphological studies by resolving the sister relationship between *Rhynchosia* and *Eriosema* and also the close relationship of the monotypic South African genus *Bolusafra* to these two genera. *Flemingia* was sister to the other Cajaninae species included in this study. The taxa included in this study were collected from diverse ecological regions of Africa, Asia and Australia and this study has shed a little light on the phylogenetic relationships of *Cajanus* across geographical regions particularly in India and Australia which are considered as important centers of diversity and endemism for *Cajanus*.

Both *ITS* and *trnL-F* were better suited to resolving intergeneric relationships and not relationships between or within *Cajanus* species. However, the *ITS* data set unambiguously resolved *Cajanus scaraeboides* as the most basal (ancestral) species in core *Cajanus* clade. This information has important implications for the gene pool classification of pigeonpea (*Cajanus cajan*) and its wild relatives and may have also important application in pigeonpea breeding programs.

Low-resolution of inter- and intra-specific relationships was observed, as evident from short branch lengths in the phylogenetic trees. The relationship within *Cajanus* was rather better resolved from the SNP based Neighbor-Joining tree (Chapter 3), which reinforced the phylogenetic framework revealed in this chapter.

Even though wider taxon sampling is waranted, particularly from the large genera of *Rhynchosia* and *Eriosema*, this study shows the polyphyly of *Rhynchosia* and monophyly of *Eriosema* as presently circumscribed.

Chapter 3

Genetic Structure and Patterns of Genetic Diversity in Domesticated Pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild relatives

Introduction

Genetic diversity and domestication

The study of genetic diversity and structure between a crop and its wild relatives may yield important insights into evolutionary history and the process of crop domestication. For example, population structure and genetic diversity can be evaluated before and after the domestication bottleneck by means of comparison to extant wild populations (Doebley et al., 2006), informing us about genetic differentiation and revealing patterns of historical gene flow and recent introgression between a crop and its wild relatives (Garris et al., 2005). In many cases wild relatives are a good representation of the crop species, except for their greater genetic diversity, while in other cases wild relatives are genetically distinct and may harbor numerous idiosyncratic adaptations (Doebley et al., 2006). Both increased genetic diversity and lineage-specific adaptations in wild relatives have potential applications to improvement of the domesticated material. Thus wild ancestors of cultivated crop species not only harbor valuable sources of genetic diversity for crop improvement, they also provide important signals of the evolutionary history of domestication.

Recent advances in molecular genetics and genomics provide powerful tools with which to address fundamental questions regarding when and where domestication occurred, estimate the minimum number of independent domestication events, establish the identity of wild progenitor species, highlight specific molecular changes underlying domestication traits, and document the impact of artificial selection during domestication and subsequent crop improvement on genome content (Aguilar-Meléndez et al., 2009; Gross and Olsen, 2010).

Studying domestication and crop evolution began with morphological studies, followed by cytogenetic and protein analyses, but the greatest advancement in the study of domestication came with the advent of DNA-based molecular markers beginning in the 1980s (Burger et al., 2008). Until recently, most molecular studies focused on neutral polymorphisms and candidate genes using markers such as SNPs and SSRs. Although early studies involved only sparsely distributed markers, such as in the case of barley (Morrell & Clegg, 2007), they nevertheless yielded a variety of fascinating insights about the evolution of crops from their wild ancestors

(Matsuoka et al., 2002; Morrell & Clegg, 2007; Zhang et al., 2009a; Aguilar-Melendez, 2009; Li et al., 2010). In the past few years, with rapid advances in DNA sequencing technologies and parallel genotyping platforms, it is now possible to compare entire genomes of numerous accessions of the same and related species to the base pair level of resolution. Such detailed analysis permits scientists to dissect genome-wide features, including patterns of linkage disequilibrium and genetic differentiation, as well as spatial patterns of nucleotide diversity.

Crop domestication, which began within the past 12,000 years, is sufficiently recent that most crop species have accumulated limited neutral mutations since their domestication, and the majority of genetic diversity in a crop is a subset of the diversity prevailing in the progenitor wild populations (Olsen and Gross, 2008; Gross and Olsen, 2010). However, novel mutations do arise in domesticated lineages, and such mutations can contribute desirable characters if their frequency is elevated by breeding selection. Alternatively, lineage-specific mutations can reduce fitness when undesirable alleles persist due to factors such as linkage tradeoffs or selection relaxation (Cruz et al., 2008). The major factor contributing to reduced genetic diversity during domestication is the "domestication bottleneck", a special instance of "founder effect" in which the subset of individuals selected for domestication are reproductively isolated from the original population by agricultural practices and breeding. Although domestication bottlenecks are driven, in part, by strong human selection on a few key domestication traits (such as loss of pod shattering or altered apical dominance exhibited in domesticated crop species), they also shift allele frequency at non-selected sites in the genome resulting in randomly derived allele content in the domesticated population, a concept known as "genetic drift". New bottlenecks and genetic drift associated with secondary domestication events may mask the true domestication history of a species (Olsen and Gross, 2008). Irrespective of specific domestication history, the process of human selection on crop germplasm results in reduced allelic diversity in the crop compared to wild progenitor populations (Gross and Olsen, 2010).

The classic model of a domestication bottleneck assumes that genetic drift occurs only during initial periods of domestication (Eyre-Walker et al., 1998). The effect and severity of genetic drift during this bottleneck is determined by the size of the bottlenecked population (Nb) and the bottleneck's duration (d generations) (Olsen and Gross, 2008). Bottleneck severity (k') can be quantified as the ratio of the size of the bottlenecked population (Nb) to the duration of the bottleneck (d) in generations (k'= Nb/d) and k' is inversely proportional to the reduction in neutral genetic diversity (Wright et al., 2005). Using simulation studies (Allaby et

al., 2008) proposed a second model known as "the protracted model" which considers the compound effect of genetic drift not only during an initial domestication bottleneck, but also when the bottleneck is continued for many subsequent generations.

All of the proposed models predict a reduction of genetic diversity in crop species as compared to the wild progenitors (Doebley et al., 2006). However, it is difficult to distinguish between reduction of genetic diversity resulting from an archaic domestication bottleneck and reduction due to intense human selection (breeding bottlenecks) during the last few hundred years (Kilian et al., 2006). Moreover, domestication of crop species can occur more than once and recent genetic and archaeological analyses have supported the existence of multiple domestication events during crop evolution. Multiple domestication events have been reported for crop species such as barley (*Hordeum vulgare*) (Morrell & Clegg, 2007), Asian rice (*Oryza sativa*) (Londo et al., 2006) and common beans (*Phaseolus vulgaris*) (Sonnante et al., 1994).

Previous genetic diversity studies on pigeonpea and wild relatives

Cajanus cajan (L.) Millsp. (Pigeonpea) belongs to the genus *Cajanus* DC. which is composed of 34 species. Pigeonpea is the only cultivated member of the genus, while the remaining species are wild relatives belonging mainly to secondary gene pools. Most *Cajanus* species are endemic and confined either to Southern/South-eastern Asia or Australia (Fortunato, 2000). Among these, 16 *Cajanus* species occur in Asia (eight are endemic to India), 15 species are distributed in Australia (13 are endemic to Australia), one species of *Cajanus* is confined to West Africa, and two species (including *Cajanus cajan*) are widespread in the old world (Lewis et al., 2005). Morphological evidence suggests that *Cajanus cajanifolius*, which is native to the Indian subcontinent, is the putative progenitor of pigeonpea (van der Maesen, 1980; 1990).

Despite its economic and cultural importance, there is very limited molecular understanding of the domestication history of pigeonpea, including its genetic diversity and relationship with wild species. Understanding phylogenetic relationships among *Cajanus* species, determining the genetic structure and diversity within cultivated and wild gene pools, and identifying the most related progenitor species are important to the proper management and efficient use of the *Cajanus* gene pool for pigeonpea improvement and breeding programs.

Multi-locus genotyping has important applications in plant genetics and population biology, as well as in the applied field of plant breeding (Haudry et al., 2007; Morrell and Clegg, 2007; Chen et al., 2009, Varshney, 2009a). The most abundant types of polymorphism

are single nucleotide polymorphisms (SNPs), which due to their ease of assay are popular molecular markers for high throughput applications. Most SNP positions in a given genome are biallelic, and SNP genotyping requires prior knowledge of allelic variation at the assay position. Analysis of allele frequencies between domesticated accessions and wild relatives derived from natural environments should indicate which wild populations are ancestral to the crop. In turn SNP genotyping can facilitate the discovery and utilization of novel traits in wild populations that may be used for wide cross introgression in crop breeding and improvement programs.

In the present study, the genetic diversity in cultivated pigeonpea (*Cajanus cajan*) and related wild species collected from diverse environments in Asia, Australia, Africa and the Caribbean is investigated using the high throughput SNP-OPA assay of the Illumina GoldenGate platform. Genotyping data were correlated with geographical distribution of the analyzed species, as well as with known history of breeding for cultivated accessions.

Materials and Methods

Plant materials

Table 3.1 lists 110 accessions of *Cajanus cajan* and allied species that were analyzed for allele content. With the exception of 12 wild accessions (which were acquired from Western Australia Herbarium and were originally collected from the field in Australia.) all the other genotypes were derived from the gene bank of International Crop Research Institute for the Semi-Arid Tropics (ICRISAT). The cultivated genotypes also include core and mini-core accessions identified by ICRISAT based on various morphological descriptors and agro-morphology traits (Reddy et al., 2005; Upadhyaya et al., 2006).

Molecular Methods

Genotyping was based on a set of 768 single nucleotide polymphisms discovered in a comparison of 1440 amplicons between *C. cajan* accession ICP28 and *C. scarabaeoides* accession ICPW94. Together these SNP assays survey biallelic states at 670 distinct genes. For purposes of genotyping, DNAs were extracted using the Qiagen DNeAsy protocol using a Retch mixer mill, according to manufacturer's instruction, and delivered to the UC Davis Genome Center genotyping core facility for analysis. Allele calls were curated using the Illumina Beadstudio software package (Illumina, San Diego, CA, USA).

*Table 3.1. List and country of origin of domesticated and wild genotypes used in this study (*Genotype no- refers to the code used for STRUCTURE analyses in Fig. 3.6)*

Genotype no*	Genotype name	Genetic status	Source of material	Origin	Accession	Species name
1	ICP28-1	cultivar	ICRISAT	India	Pusa Ageti	*C. cajan*
2	ICP28-2	cultivar	ICRISAT	India		*C. cajan*
3	ICPW94-1	Wild	ICRISAT	India		*C. scarabaeoides*
4	ICPW94-2	Wild	ICRISAT	India		*C. scarabaeoides*
5	C. scarabaeoides-C0651	Wild	PERTH	Australia	03340651	*C. scarabaeoides*
6	C. lanceolatus-C7698	Wild	PERTH	Australia	05937698	*C. lanceoltus*
7	C. reticulatus C7960	Wild	PERTH	Australia	06867960	*C. reticulatus*
8	C. cinereus-C8603	Wild	PERTH	Australia	06698603	*C. cinereus*
9	C. latisepalus C8898	Wild	PERTH	Australia	06508898	*C. latisepalus*
10	C. cajan-C6364	Natural	PERTH	Australia	05146364	*C. cajan*
11	C. cajan-C5225	Natural	PERTH	Australia	06225252	*C. cajan*
12	C. hirtopilosus -C0828	Wild	PERTH	Australia	05430828	*C. hirtopilosus*
13	C. pubescens -C0666	Wild	PERTH	Australia	07380666	*C. pubescens*
14	C. reticulatus C7837	Wild	vd Maesen	Australia	VDM 7837	*C. reticulatus*
15	C.cajanifolius-C7847	Wild	vd Maesen	India	VDM 7847	*C. cajanifolius*
16	C. crassus-C7832	Wild	vd Maesen	India	VDM 7832	*C. crassus*
17	ICP15756	Wild	ICRISAT	Indonesia	KLM 786	*C. scarabaeoides*
18	ICP15644	Wild	ICRISAT	India	NKR	*C. lineatus*
19	ICP7035	Cultivar	ICRISAT	India	Mosaic Res.	*C. cajan*
20	ICP8863	Cultivar	ICRISAT	India	ICWR-6	*C. cajan*
21	ICP12039	Cultivar	ICRISAT	Unknown		*C. cajan*
22	ICP12043	Cultivar	ICRISAT	Unknown		*C. cajan*
23	ICP22049	Cultivar	ICRISAT	Unknown		*C. cajan*
24	ICPR2438	R-Line	ICRISAT	Unknown		*C. cajan*
25	ICPR2447	R-Line	ICRISAT	Unknown		*C. cajan*
26	ICPR2463	R-Line	ICRISAT	Unknown		*C. cajan*
27	ICPR2671	R-Line	ICRISAT	Unknown		*C. cajan*
28	ICPL2	Landrace	ICRISAT	Unknown		*C. cajan*
29	ICPL332	Landrace	ICRISAT	Unknown		*C. cajan*
30	ICPL84023	Landrace	ICRISAT	Unknown		*C. cajan*
31	ICPL85010	Landrace	ICRISAT	Unknown		*C. cajan*
32	ICPL85030	Landrace	ICRISAT	Unknown		*C. cajan*
33	ICPL87091	Landrace	ICRISAT	India	Early maturing	*C. cajan*
34	ICPL87119	Landrace	ICRISAT	Asha		*C. cajan*
35	ICPL88034	Landrace	ICRISAT	Unknown		*C. cajan*
36	ICPL88039	Landrace	ICRISAT	Unknown		*C. cajan*
37	ICPL99050	Landrace	ICRISAT	Unknown		*C. cajan*
38	ICPL20096	Landrace	ICRISAT	Unknown		*C. cajan*
39	ICPL20097	Landrace	ICRISAT	Unknown		*C. cajan*
40	ICPL20102	Landrace	ICRISAT	Unknown		*C. cajan*
41	ICPL20108	Landrace	ICRISAT	Unknown		*C. cajan*
42	TTB7	cultivar	ICRISAT	Unknown		*C. cajan*
43	AKT8811	cultivar	ICRISAT	Unknown		*C. cajan*
44	TAT10	cultivar	ICRISAT	Unknown		*C. cajan*
45	BSMR736	cultivar	ICRISAT	Unknown		*C. cajan*
46	T.VISHAKA	cultivar	ICRISAT	Unknown		*C. cajan*
47	C11	cultivar	ICRISAT	Unknown		*C. cajan*
48	G T 288	cultivar	ICRISAT	Unknown		*C. cajan*
49	GULLYL White	cultivar	ICRISAT	Unknown		*C. cajan*
50	GULLYAL Red	cultivar	ICRISAT	Unknown		*C. cajan*

Table 3.1. continued……

51	GS1	cultivar	ICRISAT	Unknown		*C. cajan*
52	ICP2376		ICRISAT	India	RG 102; P 3888	*C. cajan*
53	BSMR 736	cultivar	ICRISAT	Unknown		*C. cajan*
54	C. sericeus	Wild	ICRISAT	Unknown		*C. Sericeus*
55	ICPW12	Wild	ICRISAT	Unknown		*C. latisepalus*
56	ICPW46	Wild	ICRISAT	Unknown		*C. lineatus*
57	ICPW29	Wild	ICRISAT	Unknown		*C. cajanifolius*
58	ICPW68	Wild	ICRISAT	Unknown		*C. paltycarpus*
59	ICPW69	Wild	ICRISAT	Unknown		*C. paltycarpus*
60	ICPW94-3	Wild	ICRISAT	India		*C. scarabaeoides*
61	ICPW130	Wild	ICRISAT	Unknown		*C. scarabaeoides*
62	ICP49	PI 394548	ICRISAT	India		*C. cajan*
63	ICP1071	P 4578	ICRISAT	India		*C. cajan*
64	ICP1535	Core	ICRISAT	India	P 3816/1-1	*C. cajan*
65	ICP4266	Core	ICRISAT	India	P 631	*C. cajan*
66	ICP6523	Core	ICRISAT	India		*C. cajan*
67	ICP6933	Core	ICRISAT	Trinidad	Code No. 3	*C. cajan*
68	ICP7337	Reference	ICRISAT	India	ANM 16	*C. cajan*
69	ICP7409		ICRISAT	India	ANM 79	*C. cajan*
70	ICP7782	Core	ICRISAT	India		*C. cajan*
71	ICP7941	Core	ICRISAT	India		*C. cajan*
72	ICP8095		ICRISAT	India	ANM 450	*C. cajan*
73	ICP8242	Reference	ICRISAT	India	PLA 332	*C. cajan*
74	ICP8265		ICRISAT	Unknown		*C. cajan*
75	ICP8817	Reference	ICRISAT	India	Kuselghat 1	*C. cajan*
76	ICP9236	Reference	ICRISAT	India	PI 394816	*C. cajan*
77	ICP10240	PI 394590	ICRISAT	India		*C. cajan*
78	ICP10531	PI 396259	ICRISAT	India		*C. cajan*
79	ICP10880	PI 275	ICRISAT	Philippines		*C. cajan*
80	ICP10922	T-2 (III-5)	ICRISAT	Australia		*C. cajan*
81	ICP10963	Core	ICRISAT	India	RPSP 580	*C. cajan*
82	ICP11246	Core	ICRISAT	Unknown		*C. cajan*
83	ICP11543		ICRISAT	Unknown	ICPL 87	*C. cajan*
84	ICP11754	Reference	ICRISAT	Unknown	ICPL 304	*C. cajan*
85	ICP11975	Reference	ICRISAT	Unknown	D-0 Type	*C. cajan*
86	ICP12079	Core	ICRISAT	Tanzania		*C. cajan*
87	ICP12094	Core	ICRISAT	Tanzania	PR 5474	*C. cajan*
88	ICP12765	Reference	ICRISAT	Philippines	PR 5302-4	*C. cajan*
89	ICP12977	Core	ICRISAT	India	PR 6088-1	*C. cajan*
90	ICP13004	Reference	ICRISAT	India	PR 6109-1	*C. cajan*
91	ICP13799	Core	ICRISAT	Trinidad	PR 6499	*C. cajan*
92	ICP14126	Reference	ICRISAT	Jamaica	PR 6692	*C. cajan*
93	ICP14153	Core	ICRISAT	Brazil	EEI 85137	*C. cajan*
94	ICP14389	Core	ICRISAT	C. Africa	AK 288	*C. cajan*
95	ICP14444	Minicore	ICRISAT	Unknown	ICPL 85021	*C. cajan*
96	ICP14471	Minicore	ICRISAT	Unknown	ICPL 86014	*C. cajan*
97	ICP14524	Core	ICRISAT	India	MRP 146	*C. cajan*
98	ICP14770	Reference	ICRISAT	Unknown	ICPL 332	*C. cajan*
99	ICP15454	Core	ICRISAT	Nigeria	PRAN 21	*C. cajan*
100	ICP16198	Core	ICRISAT	Unknown		*C. cajan*
101	ICP16235	Reference	ICRISAT	Unknown	ICPL 92042	*C. cajan*
102	ICP15614	Wild	ICRISAT	India	JM 2337	*C. albicans*
103	ICP15627	Wild	ICRISAT	India	ICPW 026	*C. albicans*
104	ICP15629	Wild	ICRISAT	India	ICPW 028	*C. cajanifolius*
105	ICP15632	Wild	ICRISAT	India	ICPW 031	*C. cajanifolius*

Table 3.1. continued……

106	ICP15762	Wild	ICRISAT	Unknown		*C. sericeus*
107	ICP15882	Wild	ICRISAT	India	ICPW 281	*C. scarabaeoides*
108	ICP15665	Wild	ICRISAT	India	ICPW 064	*C. paltycarpus*
109	ICP15747	Wild	ICRISAT	India	ICPW 146	*C. scarabaeoides*
110	ICPW94-4	Wild	ICRISAT	India		*C. scarabaeoides*

Data Analysis

To deduce overall similarity among accessions and resolve relationships between individual genotypes, pairwise dissimilarity was calculated by simple matching according to the formula:

$$\boldsymbol{d_{ij} = 1 - \frac{1}{L}\sum_{i}^{L}\frac{m_i}{\pi}}$$

Where $\boldsymbol{d_{ij}}$ is the dissimilarity between *i* and *j*, L is the number of loci, π is the ploidy and $\boldsymbol{m_i}$ is the number of common alleles between i and j for locus l.

Based on the dissimilarity matrix a Neighbor Joining tree (Saitou and Nei, 1987) with 1000 bootstraps was constructed to gain insight into the genetic similarity between individual genotypes. Specifically a "Weighted Neighbor Joining'' analysis was employed which uses likelihood-based criterion that models the distances as random variables obeying a Gaussian distribution (Bruno et al., 2000). Weighted Neighbor joining tree was choosen because it is relatively immune to the "long branches attract" and "long branch distracts" drawbacks observed with Neighbor Joining (Bruno et al., 2000). The analysis was carried out with DARwin5 software (Perrier and Jacquemoud-Collet 2006).

Genetic structure was analyzed using the program STRUCTURE 2.1 (Pritchard et al., 2000; Falush et al., 2003). Structure assigns individual genotypes to a specified number of groups "K" based on membership coefficients calculated from the genotype data. The analysis was run from K 1 to 10 using a burn-in period of 50,000 steps followed by 500,000 MCMC (Monte Carlo Markov Chain) replicates (with 2 iterations), assuming an admixture model and correlated allele frequencies. Optimal K, which is essentially the number of sub-populations from which the analyzed accessions derive, was determined using an *ad hoc* static ΔK based on the rate of change in the natural log probability of data between successive K values as described by Evanno et al. (2005). At optimal K, individual sub-populations were extracted and analyzed separately using STRUCTURE 2.1 to resolve additional genetic relationships; when combined with information on phylogenetic history, geography, output from PCoA and breeding history, 6 subgroups were defined.

GenAlEx 6.3 (Peakall and Smouse, 2006) was used to calculate attributes of allelic diversity, including the average number of alleles per locus (Na), effective number of alleles (Ne), the observed heterozygosity (Ho) and the expected heterozygosity (He) or gene diversity, population polymorphism, number of less common alleles and private alleles for each locus and each population. Polymorphism Information Content (PIC) for each marker was also calculated (Botstein et al., 1980).

Hierarchical Analysis of Molecular Variance (AMOVA) was conducted to assess the level of variation among wild and domesticated groups and populations. Genetic variation within groups (Fct), variation within populations (Fst) and variation between populations within a group (Fsc) were analyzed. Nei's Genetic distance and gene flow (Nm) were also analyzed using GenAlEx v.6.3 (Peakall and Smouse, 2006) and Arlequin (Excoffier et al., 2005). A principal coordinate analysis (PCoA) using GenALEx v.6.3 (Peakall and Smouse, 2006) was conducted to complement the output of the phenetic analyses, which is more sensitive to closely related individuals, whereas PCoA is more informative regarding distances among major groups (Hauser and Crovello, 1982).

Results

Single nucleotide polymorphisms were assayed in a total of 110 accessions representing cultivated *Cajanus cajan* (L.) Millsp. (79 accessions) and its wild relative relatives (31 accessions), all of which belong to the genus *Cajanus* DC. The wild accessions represent 13 species of *Cajanus* DC, while the cultivated group includes semi-domesticated naturally occuring (found in the wild and not having been planted or harvested by farmers) perennial pigeonpea accessions, landraces, breeding lines and elite cultivars (mainly, from India) (Table 3.1). These genotypes originated from widespread geographical regions, representing both tropical and subtropical environments of Africa, Asia, Latin America, the Caribbean, the Indian sub-continent and Australia. Individual SNPs were identified based on comparisons between *Cajanus cajan* accession ICP28 and *C. scarabaeoides* accession ICPW94 (R. Varma Penmetsa, unpublished data). Together these two species span a wider evolutionary distance that the proposed domestication gradient, from cultivated *C. cajan* to its presumed progenitor *C. cajanifolius*.

Relationships between wild and domesticated groups

In total, 752 loci reported allele status in 110 genotypes sampled. Based on the combined data set, which represents two distinct allele calls. The dissimilarity matrix forms the basis of a weighted Neighbor Joining tree (Fig. 3.1.) that was rooted with *C. scarabaeoides*. The choice of *C. scarabaeoides* as the root was based on combined analysis with nuclear ITS and chloroplast *trnL-F* spacer sequences that reveal *C. scarabaeoides* as the most basal member of the *Cajanus* clade, albeit with modest bootstrap support (Chapter 2).

Three well-resolved clusters were evident from this analysis, including a basal set of *C. scarabaeoides* of Indian origin, and two sister clades representing wild species of Australian origin and a more diverse but well-supported clade containing the remaining wild *Cajanus* species which were exclusively of Indian origin.

Thus, the topology reflects both the distinctiveness of species and the geographical origin of species' groups. The domesticated accessions formed a monophyletic clade that was internal to, and significantly less diverse than, the group of non-*scarabaeoides* wild species of Indian origin. *Cajanus cajanifolius* is the presumed progenitor of domesticated pigeonpea and indeed certain accessions of *C. cajanifolius* appear as the most similar outgroup to the domesticated clade, with strong bootstrap support.

A single genotype from the Philippines (ICP12765) is the only domesticated genotype placed sister to the main group of domesticated *Cajanus* in the NJ tree, and it is part of a paraphyletic assemblage that also contains two *C. cajanifolius* accessions (ICPW29 and ICP15629) and a single accession of *C. lineatus* (ICPW46). A cluster containing the two genuine (non-admixed) accessions of *C. cajanifolius* accessions, namely *C. cajanifolius* C7847 and *C. cajanifolius* ICP15632, is sister to the large group of domesticated genotypes but also more similar to the wild non-*scarabaeoides* species of Indian origin.

Thus these later *C. cajanifolius* accessions, which we speculate are true wild representatives, have the expected affinities with coherent sets of *C. platycarpus* and *C. sericeus* genotypes, as well as to individual representatives of *C. albicans* and *C. crassus*. Taken together, these data are consistent with the possibility that *C. cajanifolius* accessions ICPW29 and ICP15629, as well as *C. cajan* ICP12765 and *C. lineatus* ICPW46, are hybrids between wild and domesticated *Cajanus* material.

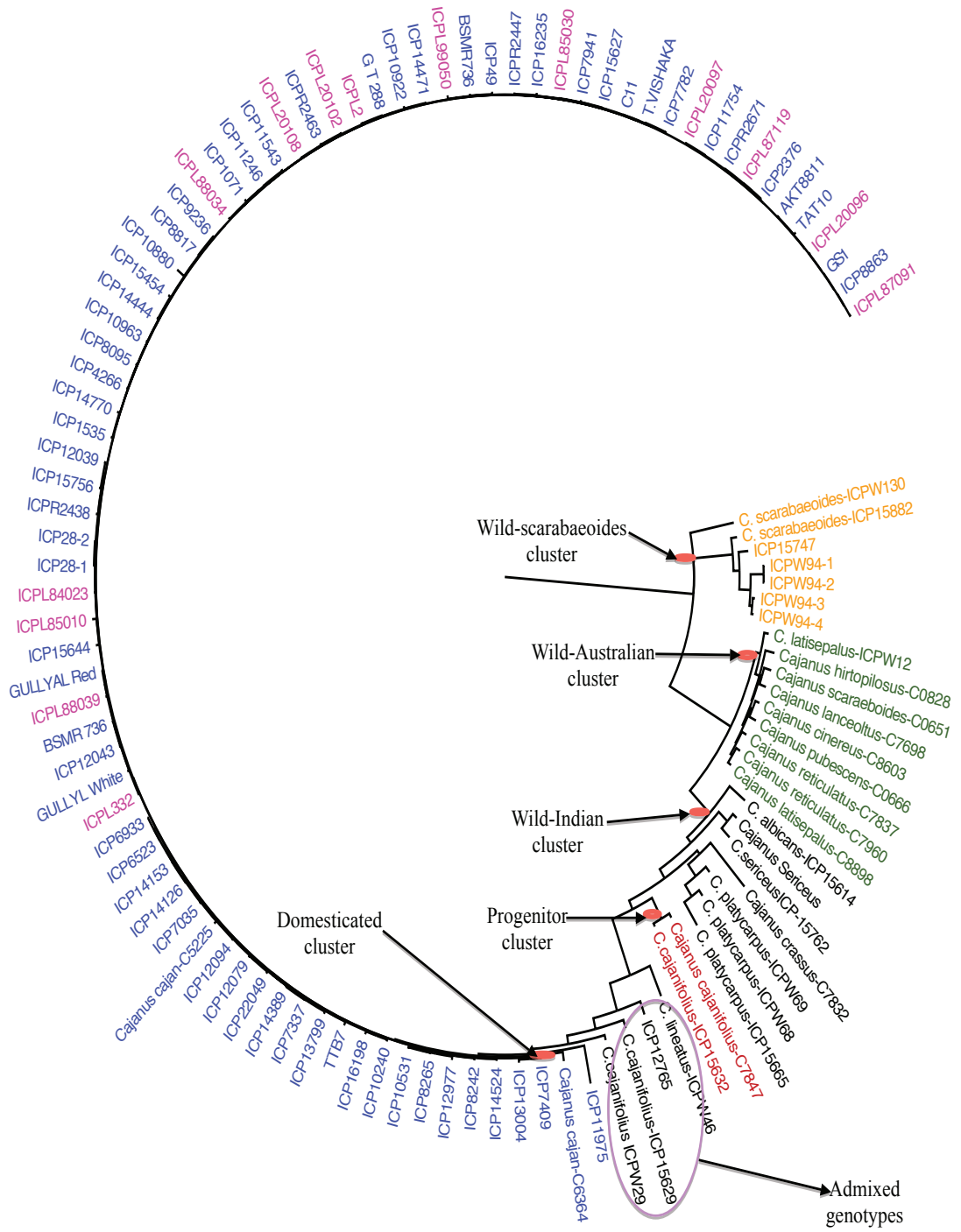


Fig. 3.1. Weighed Neighbor-Joining tree of wild and domesticated pigeonpea (*Cajanus cajan*) genotypes constructed using SNP-OPA markers (Key: Blue- domesticated cultivars and gene bank accessions; Pink-landraces; Red-wild Indian; Green-Wild Australian; orange-Wild *scarabaeoides* accessions and admixed genotypes are shown in a circle).

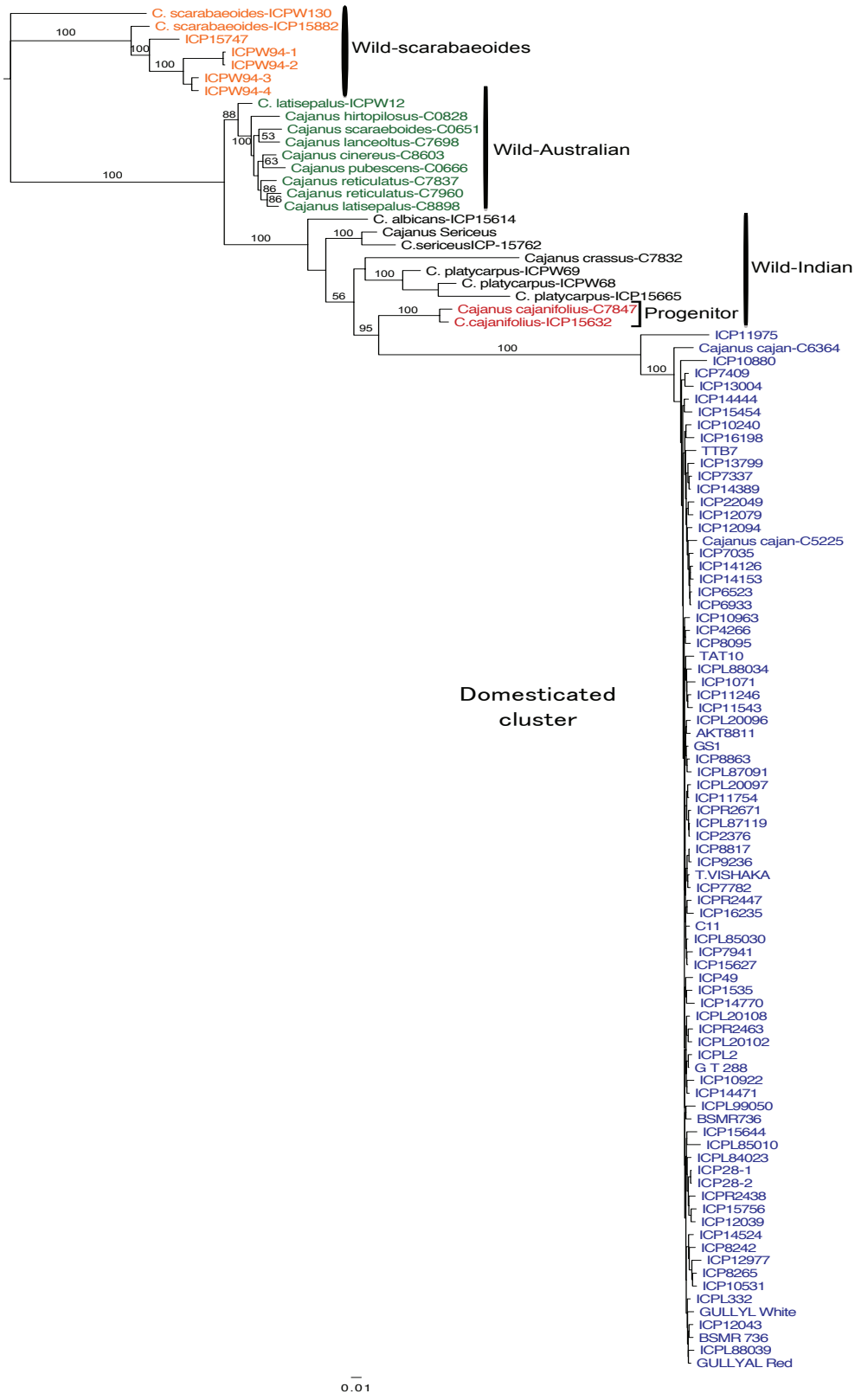


Fig. 3.2. A phenogram of wild and domesticated pigeonpea (excluding the admixed genotypes) using SNP-OPA markers (Blue-domesticated Black-wild-Indian; Green-Wild-Australian; Orange-Wild *scarabaeoides* accessions and red- putative progenitor accessions) (Bootstrap value > 50% Shown above the branches).

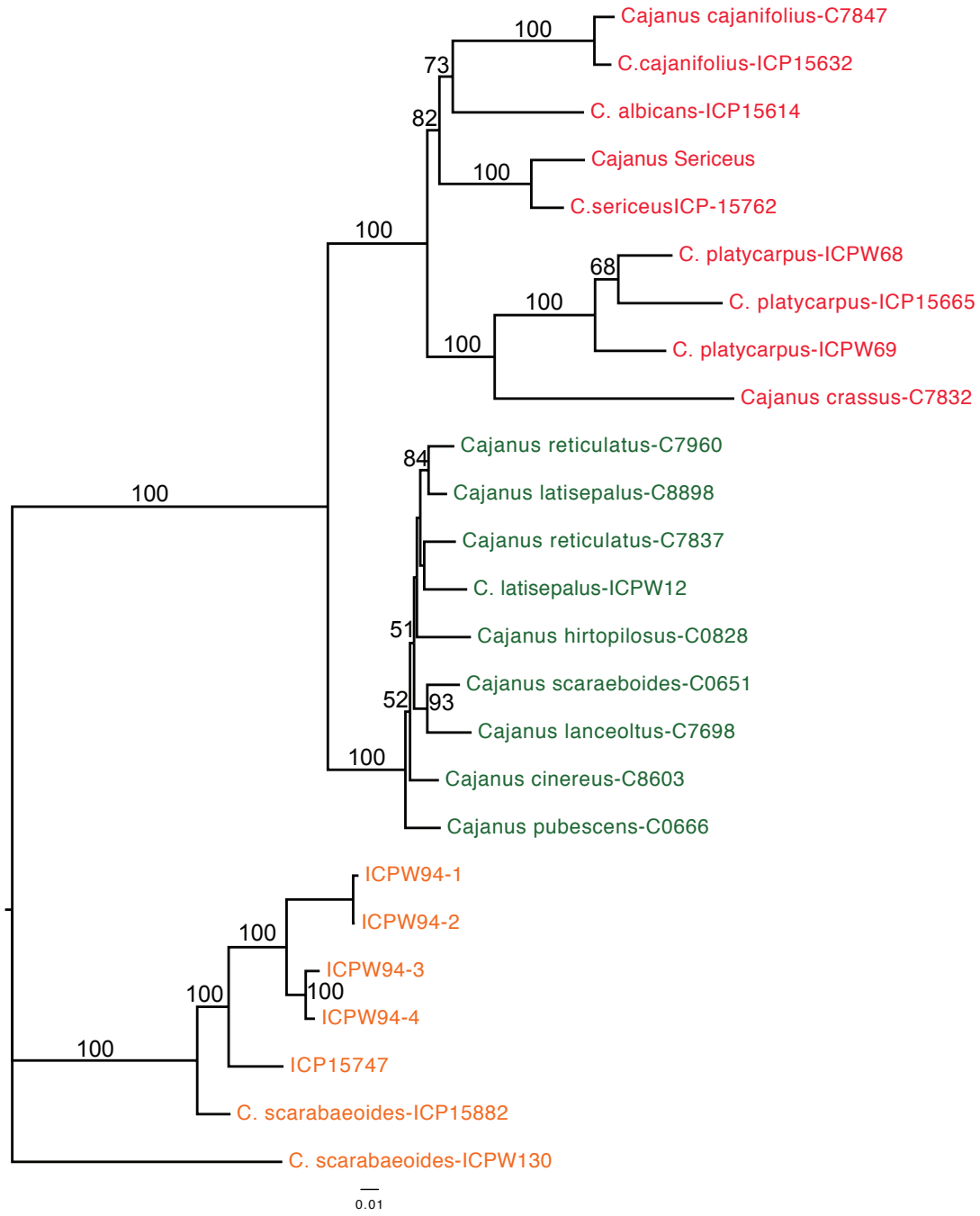


Fig. 3.3. A phenogram of wild *Cajanus* species using SNP-OPA markers (Bootstrap value > 50% shown above the branches), Red- wild Indian; Green-wild Australia; Orange-Wild scarabaeoides.

To avoid the potentially confounding effects of breeding admixture within the cultigen *C. cajan* and the possible hybrid origin of ICPW29, ICP15629, ICP12765 and ICPW46, we conducted a second Neighbor Joining analysis from which these individuals were excluded. The resulting phenogram (Fig. 3.2.), rooted with *C. scarabaeoides*, reveals patterns that are

largely consistent with, but more resolved than, our prior *ITS* and chloroplast *trnL-F* spacer analysis.

Thus, Australian species of *Cajanus* are distinct and well resolved from a sister cluster that is entirely of Indian origin and in which *C. cajanifolius* and *C. albicans* have close affinities (Figs. 3.2 and 3.3.). Three species that were not part of the previous gene-based phylogeny are also well-supported members of this Indian cluster, namely *C. sericeus*, *C. platycarpus* and *C. crassus*. The Neighbor Joining analysis placed three accessions that are annotated as wild non-*cajan* species (ICP15627, *C. albicans*; ICP15756, *C. scarabaeoides*, and ICP15644, *C. lineatus*) within the domesticated clade. There were no distinctly resolved sub-clades within the *C. cajan* complex, which consists of landraces, improved cultivars and naturally occurring perennial pigeonpea genotypes.

Within this *Cajanus cajan* complex, *Cajanus cajan* 6364 and ICP11975 (a reference genotype from ICRISAT) were the most basal genotypes. Interestingly, *C. cajan* 6364 is annotated as a naturally occurring, semi-domesticated and rarely found Australian woody herbaceous pigeonpea. However, the distant relationship of this accession to other *Cajanus* spp. of Australian origin (see below), and its close affinity with domesticated *C. cajan*, is perhaps more consistent with the origin of *C. cajan* 6364 as a feral genotype. Nine accessions, representing seven species, all of which were derived from Australia, form a homogenous cluster that is designated "wild Australia" in Figure 3.1.

Although few taxa within this cluster were sampled more than once, two accessions of *C. latisepalus* stand out because they are more similar to other wild Australian species than to one another. Moreover, within the Australian cluster one accession was designated as *C. scarabaeoides* 0651 but is unambiguously part of the Australian genetic unit and dissimilar to the Indian *C. scarabaeoides* accessions.

Principal Coordinate Analysis (PCoA) distinguished three groups of individuals (I, II and III) along discriminate axes 1 and 2, which accounted for 85.81% and 8.02% of the total molecular variability respectively (Fig. 3.5). Along the first axis, wild accessions were resolved from domesticated accessions, while the second axis resolved the Indian *Cajanus scarabaeoides* group (group I) from the remaining wild accessions of both Australian and Indian origin (group II). Within group II the Australian set forms a homogenous subgroup and the Indian genotypes form a more diverse assemblage, consistent with the previous Neighbor Joining analysis.

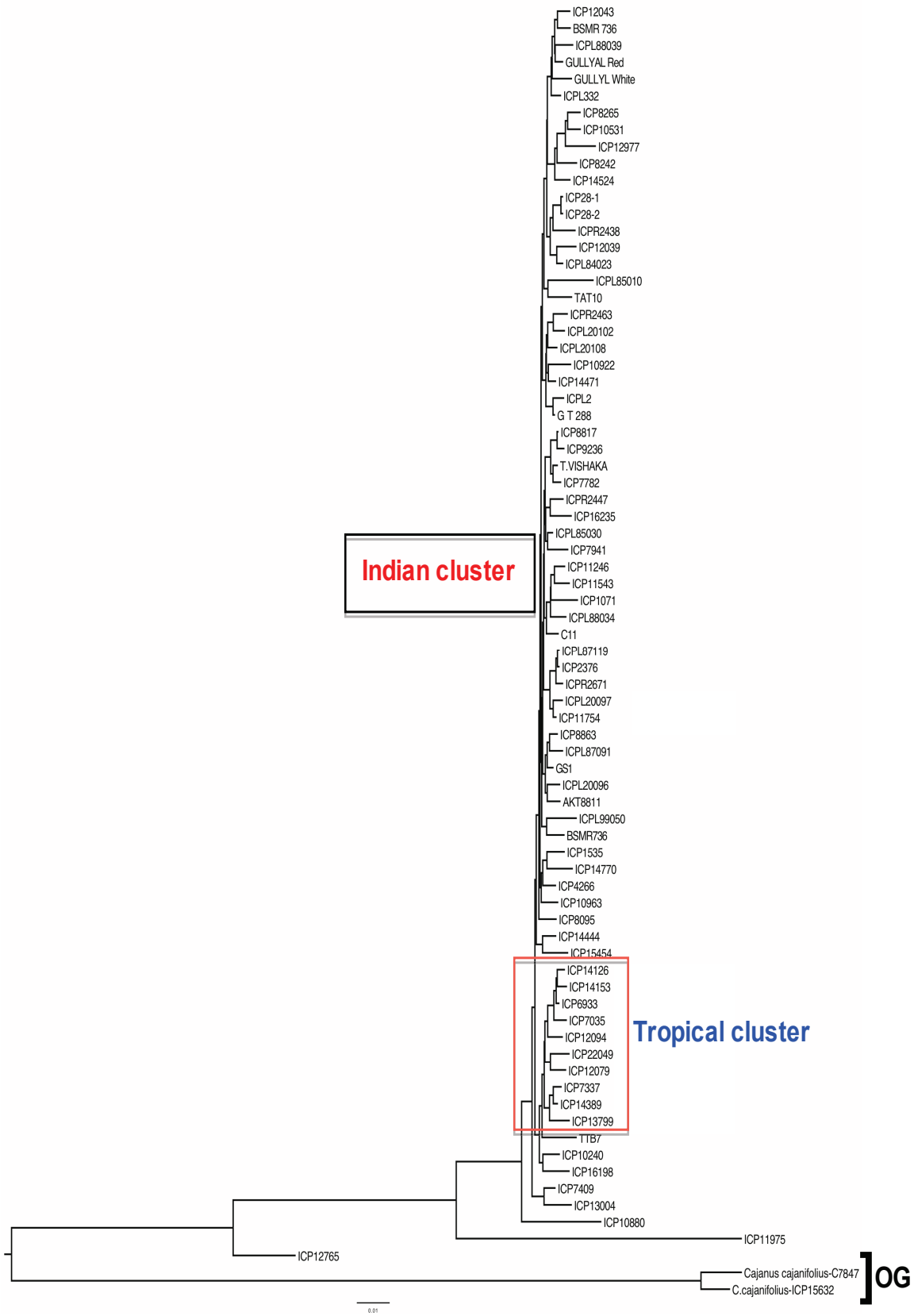


Fig. 3.4. A weighed Neighbor-Joining tree of domesticated pigeonpea (*Cajanus cajan*) constructed using SNP-OPA markers (the two *C. cajanifolius* genotypes were used as outgroups).

Group III contained the domesticated *C. cajan* cluster. The low level of variation of the domesticated cluster is reflected in the tight clustering of these genotypes, with the exception of the more basal genotypes from Fig. 3.1, including those that we nominated (above) as hybrid accessions, and also the basal *C. cajan* genotype ICP11975. Similar to the Neighbor Joining analysis, PCoA also grouped three "wild" species (*C. albicans* ICP15627, *C. scarabaeoides* ICP15756 and *C. lineatus* ICP15644) within the domesticated group, suggesting that additional identification or curation and/or renaming of these accessions is warranted.

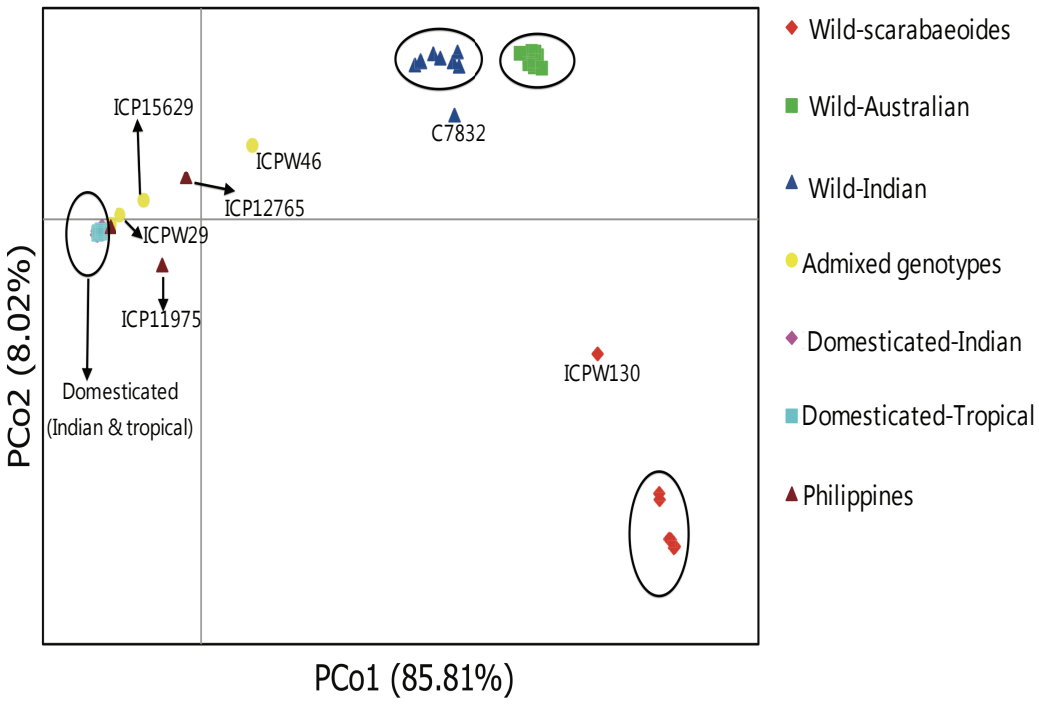


Fig. 3.5. A plot of the scores for the two principal coordinates analysis of domesticated pigeonpea and wild relatives surveyed for SNP variation. The percentage of total variation accounted for by PCo1 and PCo2 were 85.81% and 8.02% respectively.

Genetic structure of wild and domesticated pigeonpea

The Bayesian approach, STRUCTURE (Pritchard et al. 2000; Falush et al., 2003), was applied to investigate genetic structure and estimate the number of subpopulations, "K", among wild and domesticated accessions. STRUCTURE is a model-based algorithm that uses allele frequencies to derive subsets from a set of sampled individuals that approximate Hardy-Weinberg equilibrium, and thus represent subpopulations in the genetic sense. In the current

study the taxonomic divisions are species level distinctions, but the use of allele frequency data and the analogy to natural populations are equally relevant and STRUCTURE has been used widely in similar situations to deduce genetic relationships (Morrell & Clegg, 2007; Zhang et al., 2009a; Aguilar-Meléndez, 2009; Li et al., 2010).

In principle, the optimum K value should indicate the number of distinct genetic subgroups within the analyzed set of accessions. Optimal K=3 (Fig. 3.6 and Fig. 3.8) was estimated according to the method of Evanno et al. (2005), where the point of inflection in maximum *ln* probability is interpreted as optimal K (Table 3.2. and Fig. 3.8). The Schematic representation of inferring population structure (e.g. K=2-4) using STRUCTURE is depicted in Fig. 3.7. At K=3, most (93.6%) accessions were assigned to one of three subgroups, and these subgroups were biologically sensible and congruent with previous Neighbor Joining and Principal Coordinate Analysis (PCoA; Figures 3.1 and 3.5). When one has knowledge of the biology and history of a set of accessions, it can also be informative to analyze the partitioning of accessions to subgroups at a range of K values.

When applied to the *Cajanus* data set, STRUCTURE analysis (Fig. 3.6.) delineated the wild and domesticated groups at K=2. At K=3 the wild cluster further resolved into two groups, corresponding to *Cajanus scarabaeoides* and wild *Cajanus* species of Australian origin. The wild species of Indian origin, which based on Neighbor Joining analysis are a coherent but diverse unit, share a high membership coefficient with the wild Australia group and to a lesser but consistent extent with domesticated accessions. Even though they did not form a separate cluster until K=5, this hybrid constitution involving the Australian and domesticated groups readily distinguished the wild Indian species from all other accessions even at K=2.

Genetic admixture was evident between wild and domesticated groups for seven accessions, including those four genotypes nominated previously as admixed based on Neighbor Joining and PCoA analyses, two additional *C. cajan* accessions (C-6364 and ICP11975) that were basal to the domesticated group, and one accession (ICP10880) that was placed internal to the domesticated group in Neighbor Joining analysis but with a relatively long branch length.

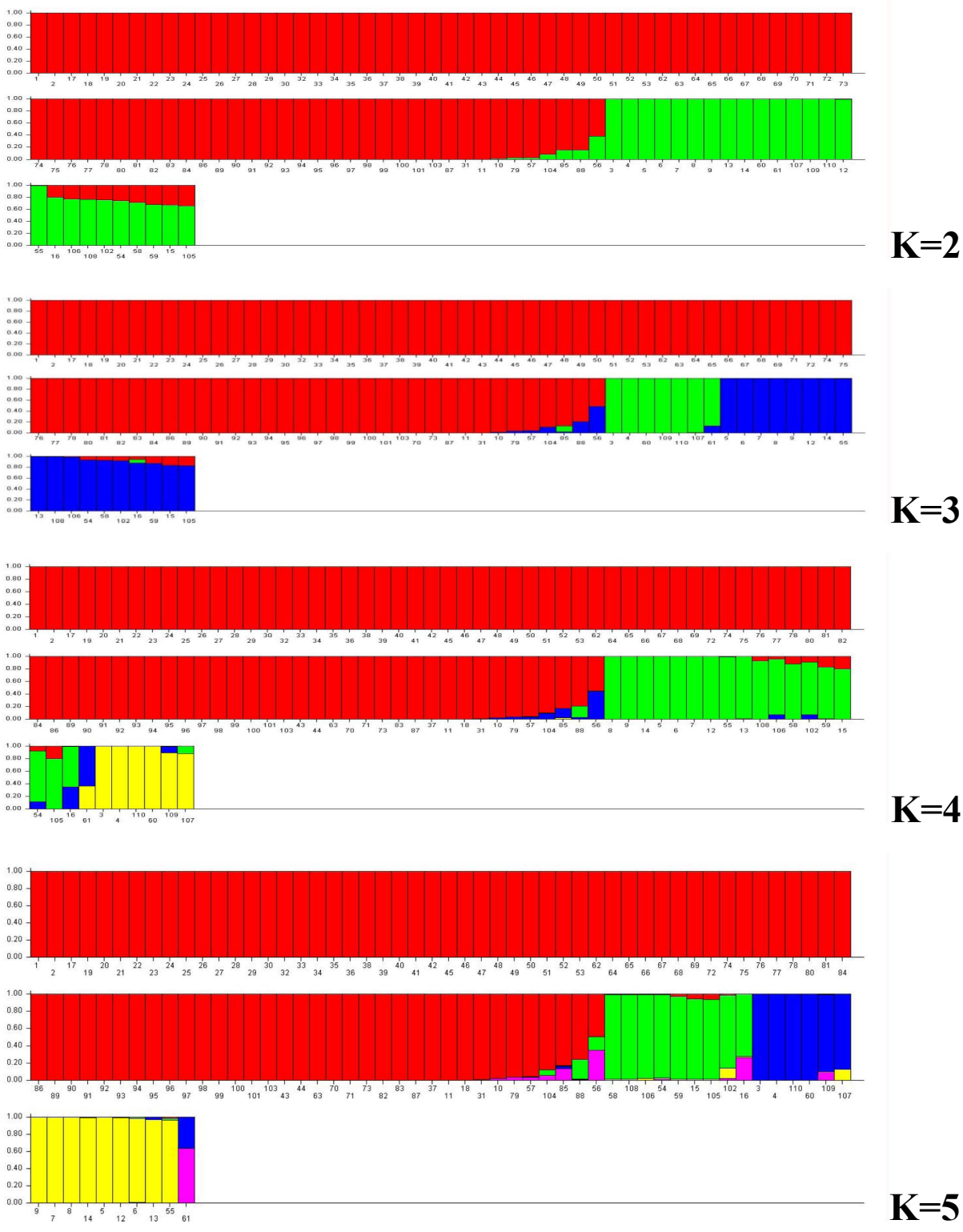


Fig. 3.6. Population structure of the wild and domesticated genotypes of *Cajanus* by STRUCTURE program. Each individual is shown as a vertical bar partitioned into K colored components, representing inferred membership in K genetic clusters (see Table 3.1. and the text for detailed information for each genotype).

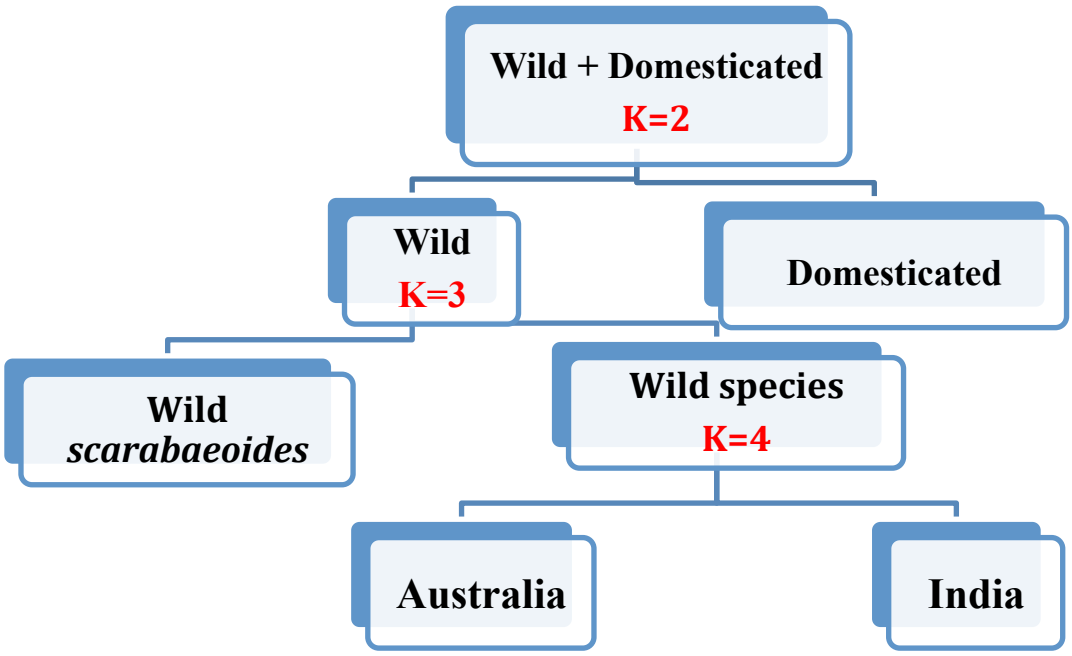


Fig. 3.7. Schematic clustering procedure during inferring population structure using STRUCTURE.

In all cases admixture involved predominant membership with domesticated accessions, typically involving minor shared membership with species in the wild Australia cluster. Only in the case of *C. cajan* ICP11975 was there evidence of genetic contribution from the *C. scarabaeoides* gene pool to a domesticated genetic background. ICP11975 was also unusual as an outlier in the PCoA analysis (Fig. 3.5.), reflecting its unique genetic constitution.

Cajanus cajanifolius ICPW29 and ICP15629 had >95% of the genome of the domesticated cluster and less than 5% of the wild background, which is essentially the inverse constitution of the wild *Cajanus* accessions of Indian origin, including *C. cajanifolius* C7847 and ICP15632. These features support the previous proposal that *C. cajanifolius* ICPW29 and ICP15629 are either feral hybrids, or breeding admixtures between domesticated *C. cajan* and wild *C. cajanifolius*. *Cajanus lineatus* (ICPW46) was admixed in nearly equal proportions from both wild and domesticated backgrounds, and had been previously proposed as an admixed genotype by van der Maesen (personal communication). Among domesticated accessions, the two Philippines accessions (ICP12765 and ICP10880) and a reference ICRISAT accession (ICP11975) showed 1-15% of the wild genome.

We analyzed the cultivated accessions alone to determine if genetic structure could be detected without the confounding effect of far greater differences among the wild taxa.

Table 3.2. Part of the Summary of Simulations (K from 2-5 with two iterations) to show Change of ln Probability)

Run Name	K	Ln P(D)	Var [LnP(D)]	a1	Fst_1	Fst_2	Fst_3	Fst_4	Fst_5
Run_2	2	-11906.3	672.2	0.0377	0.3664	0.9631	-	-	-
Run_1	2	-11899.7	668.3	0.0350	0.9629	0.3672	-	-	-
Run_4	3	-8244.8	1016.4	0.0258	0.9628	0.7155	0.6206	-	-
Run_3	3	-8238.7	996.9	0.0268	0.7119	0.9634	0.6210	-	-
Run_6	4	-7893.2	974.9	0.0259	0.9676	0.7509	0.0069	0.9303	-
Run_5	4	-8084.2	1340.4	0.0268	0.0063	0.9303	0.9685	0.7520	-
Run_8	5	-20771.5	29295.8	0.0243	0.6858	0.0074	0.9673	0.9019	0.9308
Run_7	5	-23583.7	34897.9	0.0239	0.9671	0.6959	0.9308	0.9016	0.0090

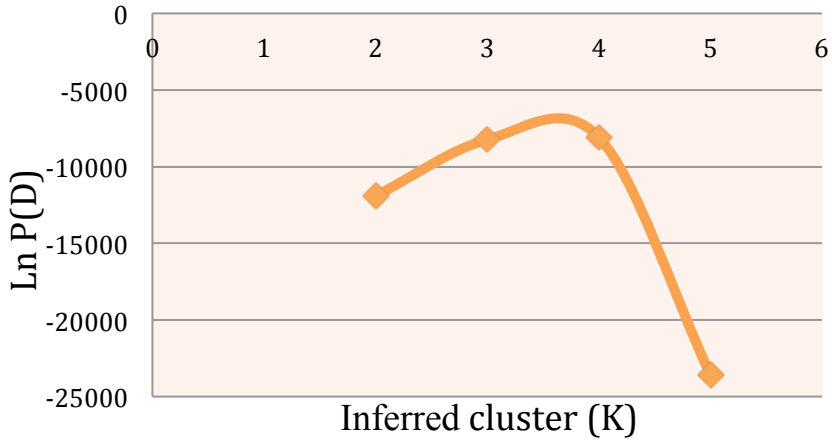


Fig. 3.8. Estimated value of ln probability for inferred cluster (K).

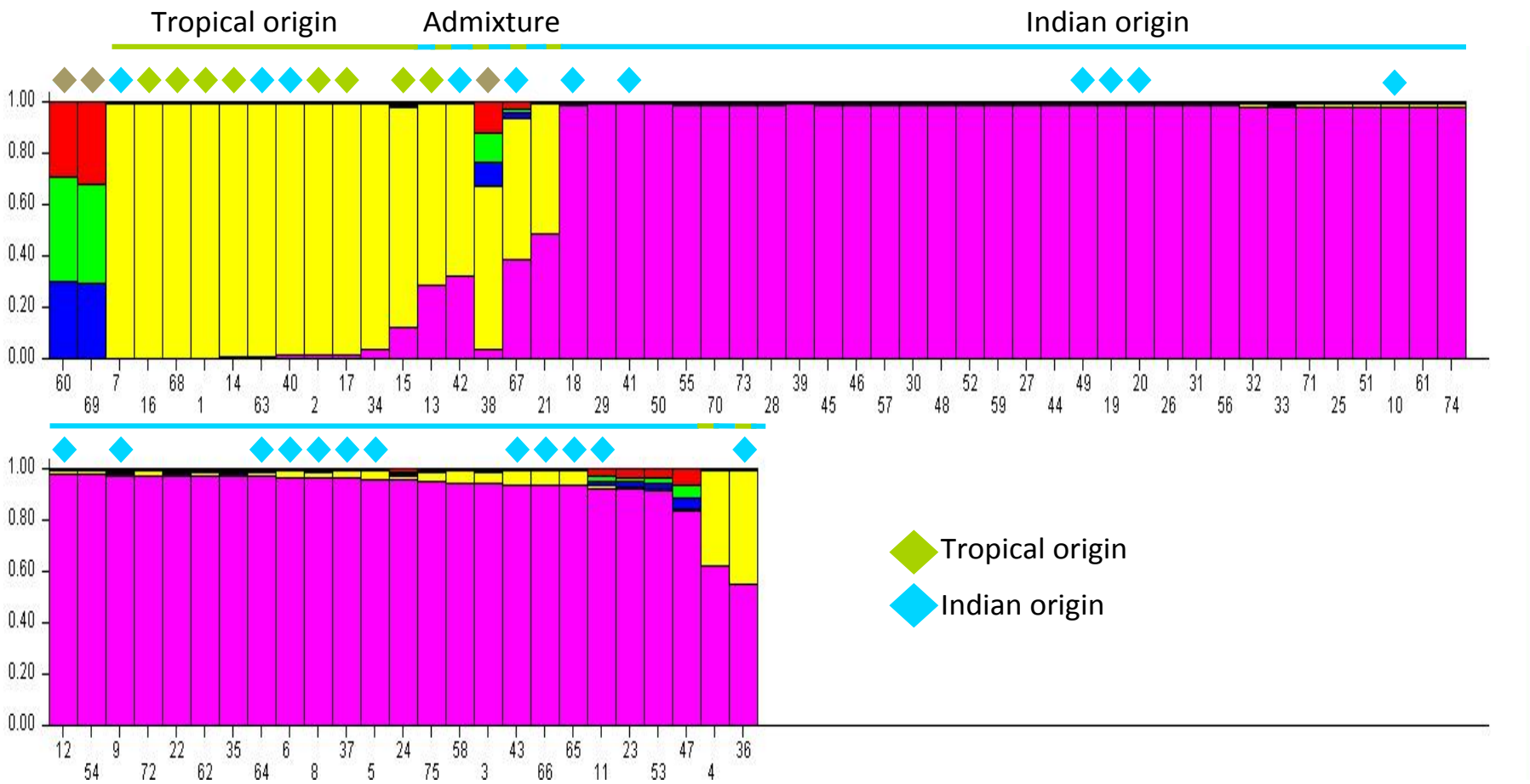


Fig. 3.9. Population structure of the 75 domesticated genotypes by STRUCTURE program. Each individual is shown as a vertical bar partitioned into K colored components, representing inferred membership in K genetic clusters (see Table 3.3 for detailed information for the genotypes.

*Table 3.3. List and country of origin of domesticated genotypes used in this study. *Genotype no-refers to the code used for STRUCTURE analyses in Fig. 3.9)*

Genotype no	Genotype name	Genetic status	Source of material	Origin	Accession	Species name
1	ICP14153	Core	ICRISAT	Brazil	EEI 85137	*C. cajan*
2	ICP14389	Core	ICRISAT	C. Africa	AK 288	*C. cajan*
3	ICP11246	Core	ICRISAT	Unknown		*C. cajan*
4	ICP16198	Core	ICRISAT	Unknown		*C. cajan*
5	ICP1535	Core	ICRISAT	India	P 3816/1-1; T 74	*C. cajan*
6	ICP4266	Core	ICRISAT	India	P 631; PI 394475	*C. cajan*
7	ICP6523	Core	ICRISAT	India		*C. cajan*
8	ICP7782	Core	ICRISAT	India		*C. cajan*
9	ICP7941	Core	ICRISAT	India		*C. cajan*
10	ICP10963	Core	ICRISAT	India	RPSP 580	*C. cajan*
11	ICP12977	Core	ICRISAT	India	PR 6088-1	*C. cajan*
12	ICP14524	Core	ICRISAT	India	MRP 146	*C. cajan*
13	ICP15454	Core	ICRISAT	Nigeria	PRAN 21	*C. cajan*
14	ICP12079	Core	ICRISAT	Tanzania		*C. cajan*
15	ICP12094	Core	ICRISAT	Tanzania	PR 5474	*C. cajan*
16	ICP6933	Core	ICRISAT	Trinidad	Code No. 3	*C. cajan*
17	ICP13799	Core	ICRISAT	Trinidad	PR 6499	*C. cajan*
18	ICP8863	Cultivar	ICRISAT	India	ICWR-6	*C. cajan*
19	ICP28-1	Cultivar	ICRISAT	India	Pusa Ageti	*C. cajan*
20	ICP28-2	Cultivar	ICRISAT	India		*C. cajan*
21	TTB7	Cultivar	ICRISAT	Unknown		*C. cajan*
22	AKT8811	Cultivar	ICRISAT	Unknown		*C. cajan*
23	TAT10	Cultivar	ICRISAT	Unknown	early maturing	*C. cajan*
24	BSMR736	Cultivar	ICRISAT	Unknown		*C. cajan*
25	T.VISHAKA	Cultivar	ICRISAT	Unknown		*C. cajan*
26	C11	Cultivar	ICRISAT	Unknown		*C. cajan*
27	G T 288	Cultivar	ICRISAT	Unknown		*C. cajan*
28	GULLYAL White	Cultivar	ICRISAT	Unknown		*C. cajan*
29	GULLYAL Red	Cultivar	ICRISAT	Unknown		*C. cajan*
30	GS1	Cultivar	ICRISAT	Unknown		*C. cajan*
31	BSMR 736	Cultivar	ICRISAT	Unknown		*C. cajan*
32	ICP12039	Cultivar	ICRISAT	Unknown		*C. cajan*
33	ICP12043	Cultivar	ICRISAT	Unknown		*C. cajan*
34	ICP22049	Cultivar	ICRISAT	Unknown		*C. cajan*
35	ICP8265		ICRISAT	Unknown		*C. cajan*
36	ICP10240	PI 394590	ICRISAT	India		*C. cajan*
37	ICP10531	PI 396259	ICRISAT	India		*C. cajan*
38	ICP10880	PI 275	ICRISAT	Philippines		*C. cajan*
39	ICP10922	T-2 (III-5)	ICRISAT	Australia		*C. cajan*
40	ICP7035	Cultivar	ICRISAT	India	A Sterility Res.	*C. cajan*
41	ICP2376		ICRISAT	India	RG 102; P 3888	*C. cajan*
42	ICP7409		ICRISAT	India	ANM 79	*C. cajan*
43	ICP8095		ICRISAT	India	ANM 450	*C. cajan*
44	ICPL2	Landrace	ICRISAT	Unknown		*C. cajan*
45	ICPL332	Landrace	ICRISAT	Unknown		*C. cajan*
46	ICPL84023	Landrace	ICRISAT	Unknown		*C. cajan*
47	ICPL85010	Landrace	ICRISAT	Unknown		*C. cajan*
48	ICPL85030	Landrace	ICRISAT	Unknown		*C. cajan*
49	ICPL87091	Landrace	ICRISAT	India	Early maturing	*C. cajan*

Table 3.2. continued…….

50	ICPL87119	Landrace	ICRISAT	Asha		*C. cajan*
51	ICPL88034	Landrace	ICRISAT	Unknown		*C. cajan*
52	ICPL88039	Landrace	ICRISAT	Unknown		*C. cajan*
53	ICPL99050	Landrace	ICRISAT	Unknown		*C. cajan*
54	ICPL20096	Landrace	ICRISAT	Unknown		*C. cajan*
55	ICPL20097	Landrace	ICRISAT	Unknown		*C. cajan*
56	ICPL20102	Landrace	ICRISAT	Unknown		*C. cajan*
57	ICPL20108	Landrace	ICRISAT	Unknown		*C. cajan*
58	ICP14444	Minicore	ICRISAT	Unknown	ICPL 85021	*C. cajan*
59	ICP14471	Minicore	ICRISAT	Unknown	ICPL 86014	*C. cajan*
60	ICP11975	Reference	ICRISAT	Unknown	D-0 Type	*C. cajan*
61	ICP16235	Reference	ICRISAT	Unknown	ICPL 92042	*C. cajan*
62	ICP14770	Reference	ICRISAT	Unknown	ICPL 332Abhaya	*C. cajan*
63	ICP7337	Reference	ICRISAT	India	ANM 16	*C. cajan*
64	ICP8242	Reference	ICRISAT	India	PLA 332	*C. cajan*
65	ICP8817	Reference	ICRISAT	India	Kuselghat 1	*C. cajan*
66	ICP9236	Reference	ICRISAT	India	PI 394816	*C. cajan*
67	ICP13004	Reference	ICRISAT	India	PR 6109-1	*C. cajan*
68	ICP14126	Reference	ICRISAT	Jamaica	PR 6692	*C. cajan*
69	ICP12765	Reference	ICRISAT	Philippines	PR 5302-4	*C. cajan*
70	ICP11754	Reference	ICRISAT	Unknown	ICPL 304	*C. cajan*
71	ICPR2438	R-Line	ICRISAT	Unknown		*C. cajan*
72	ICPR2447	R-Line	ICRISAT	Unknown		*C. cajan*
73	ICPR2463	R-Line	ICRISAT	Unknown		*C. cajan*
74	ICPR2671	R-Line	ICRISAT	Unknown		*C. cajan*
75	ICP11543		ICRISAT	Unknown	ICPL 87	*C. cajan*

At K=5, the majority of domesticated *Cajanus cajan* accessions resolved into two primary subpopulations and two outlying genotypes of Philippine origin (Fig. 3.9), in addition to numerous apparently admixed genotypes. The primary genetic subdivision mirrors the eco-geographic history of the associated genotypes, with one cluster containing accessions mainly from tropical regions in Africa, the Caribbean, and Latin America, and the second cluster containing genotypes of Indian origin. We noted previously that the two Philippine genotypes were admixed. In the current analysis, in addition to being distinct from the two primary *C. cajan* divisions, they also share partial genetic membership with other admixed individuals (Fig. 3.9.). Thus, two classes of admixed *C. cajan* accessions were evident: those sharing admixed membership with the Philippine accessions and those with apparently simple admixture between the tropical and Indian groups of *C. cajan*.

To provide an independent assessment of relationships within the domesticated *C. cajan* group, we used Darwin5 to construct a weighted Neighbor Joining tree (Fig. 3.4) that was rooted using two wild accessions of *C. cajanifolius*. As was the case with STRUCTURE, most accessions fell into two primary groups: one composed mostly of the tropical accessions and a second larger group of exclusively Indian origin. Moreover, genotypes revealed as admixed by

STRUCTURE (Fig. 3.9) were always the most basal members of each group, with an especially large number of admixed accessions at the base of the tropical group.

Genetic differentiation between wild and domesticated accessions

Analysis of Molecular Variance (AMOVA) was performed using ARLEQUIN 3.1 and GenAlExv.6.3. Variance was partitioned among hierarchical sets of individual genotypes (three groups and six populations) that were circumscribed by a combination of phylogenetic analysis, geographical origin, breeding history, and the outputs of PCoA and STRUCTURE. To avoid the potentially confounding effects of admixture (revealed by STRUCTURE), we removed all admixed genotypes prior to analysis with the exception of the Philippines genotypes because the Philippines accessions are of great interest for breeders. Knowing levels of genetic differentiation and gene flow with wild species is very important on future efforts of improvement programs.

The principle subdivisions for AMOVA were as follows:

Group I (2 populations): Wild-Australia cluster and Wild-India

Group II (1 population): Wild- *C. scarabaeoides*

Group III (3 populations): Domesticated-India, Domesticated-Tropical & Domesticated Philippines

Genetic variation was assessed at three hierarchical levels: among groups, among populations within groups, and within populations. In Table 3.4, total variation is presented as both the percentage of variation and in terms of Phi (Φ)-statistics, which are the AMOVA analogues of F-statistics. In particular, PhiRT (analogous to F_{CT}) refers the variation among groups, PhiPR (analogous to F_{SC}) is the variation among populations within groups, and PhiPT (analogous to F_{ST}) is the variation among populations. Statistical significance was determined by random permutation (always set to 1000).

Most genetic variation was attributed to differences among the three groups (89%) (Table 3.4). Nevertheless, genetic differentiation was evident at all levels of analysis, with fixation indices (PhiRT, PhiPT and PhiPR) ranging from 0.521 within populations to 0.949 within populations among groups (Table 3.4). These broad patterns of genetic differentiation reflect patterns established in previous phylogenetic, PCoA and STRUCTURE analyses.

Table 3.4. Summary results of AMOVA analyses within and among populations of domesticated Pigeonpea and wild relatives

Source	df	SS	MS	Est. Var.	% Variation	Statistics	Value	P
Among groups	2	18135.809	9067.905	492.714	**89%**	**PhiRT**	**0.895**	**0.010**
Among pops within groups	3	1107.072	369.024	30.265	**5%**	**PhiPR**	**0.521**	**0.010**
Within pops	89	2476.214	27.823	27.823	**5%**	**PhiPT**	**0.949**	**0.010**
Total	94	21719.095		550.801	100%			

d.f.: Degrees of freedom; SS, sum of squared observations; MS, mean of squared observations; Est. var., estimated variance % Var., percentage of total variance; PhiRT, proportion of the total genetic variance between groups; PhiPR, proportion of the total genetic variance among populations within a group; PhiPT, proportion of the total genetic variance among individuals within populations.

Genetic diversity estimates in wild and domesticated accessions

As shown in Table 3.5, genetic polymorphism (as well as genetic diversity [He], Table 3.6) was highest within wild populations of Indian origin (both wild-Indian and wild-*scarabaeoides*), with rates ~3-fold higher than within the wild-Australian population (~37% polymorphism compared to ~12%). In contrast, the lowest rates of polymorphism were documented for the domesticated-Indian and domesticated-tropical populations, with the domesticated-tropical population having ~60% the polymorphism of the domesticated-India population. Polymorphism was strikingly high with the domesticated-Philippines population, consistent with a complex process of admixture and selection in these geographically isolated genotypes.

The propensity for any given population to exhibit private (unique) alleles is highly biased in this assay, because only two genotypes (one *C. scarabaeoides* and one *C. cajan*) were sequenced for SNP discovery and thus only their respective populations have the possibility of having unique (private) alleles. With this caveat in mind, private alleles were abundant in the wild *scarabaeoides* populations (174 alleles), compared to two private alleles recorded in the domesticated-India subpopulation (Table 3.6). This reflects not only their phylogenetic distance, but also the recent derived status of the domesticated germplasm from the wild-India population. One private allele each was identified for the wild-Indian and the wild-Australian populations; we assume that these alleles represent false discovery because as explained above private alleles should be confined to *C. scarabaeoides* and *C. cajanus*.

An alternative measure of allele distribution is to enumerate the percentage of alleles in a particular population that are found in less than 50% of the remaining populations. In the current situation, where each locus is defined by two alleles, this equates to >50% of the remaining populations being fixed for the alternate allele. By this measure, each of the six populations were relatively well differentiated from at least 50% of the remaining populations, with number of locally common alleles found in 50% or fewer populations ranging from 0.197 for wild-tropical to 0.503 for wild-*scarabaeoides*.

Table 3.5. Percentage of Polymorphic loci in wild and domesticated groups

Sub-groups	Number of accessions	Genetic Status	Polymorphic loci (%)
Wild *scarabaeoides*	4	Wild	36.7%
Wild Australian	9	Wild	11.84%
Wild Indian	9	Wild	37.37%
Domesticated Indian	58	Domesticated	8.64%
Domesticated Tropical	12	Domesticated	5.45%
Philippines	3	Domesticated	23.94%
Mean			**20.66%**
SE			**5.78%**

Genetic distance and gene flow between wild and domesticated accessions

Pairwise genetic differentiation and genetic distances among wild and domesticated populations were estimated by genetic differentiation (F_{ST}) and genetic distance (Nei's genetic distance) indices using GeneAlEx6.3 software. The most distant relationships and the highest genetic differentiation were found between wild and domesticated populations (Tables 3.7 and 3.8). There was little genetic differentiation and the lowest genetic distance was between domesticated-India and domesticated-Tropical populations, while the Philippines population showed relatively high genetic differentiation with these domesticated groups. The Philippines population had closer affinity to the wild-Indian population than with either wild-Australian or wild-*scarabaeoides*, and this was reflected in higher estimated gene flow between these populations (Table 3.8). Because wild Philippines was less differentiated from wild-Indian

than were either of the other domesticated populations, these results support wild-Indian as the primary source of admixture for the Philippines accessions.

Table 3.6. Allelic Patterns across wild and domesticated clusters

	Wild *scarabaeoides*	Wild Australian	Wild Indian	Domesticated Indian	Domesticated Tropical	Philippines
Na	1.364	1.117	1.374	1.086	1.055	1.239
Na Freq. >= 5%	1.364	1.117	1.374	1.060	1.052	1.239
Ne	1.156	1.053	1.186	1.044	1.040	1.182
No of Private Alleles	0.231	0.001	0.001	0.003	0.000	0.000
No. LComm Alleles (<=50%)	0.503	0.379	0.390	0.207	0.197	0.242
He	0.105	0.033	0.114	0.024	0.021	0.102

Na = Number of Different Alleles
Na (Freq >= 5%) = Number of Different Alleles with a Frequency >= 5%
Ne = Number of Effective Alleles = 1 / (Sum pi^2)
No. Private Alleles = Number of Alleles Unique to a Single Population
No. LComm Alleles (<=50%) = Number of Locally Common Alleles (Freq. >= 5%) found in 50% or Fewer Populations
He = Expected Heterozygosity = 1 - Sum pi^2

Table 3.7. Pairwise estimates of FST among wild and domesticated groups

	Wild *scarabaeoides*	Wild Australian	Wild Indian	Domesticated Indian	Domesticated Tropical	Philippines
Wild *scarabaeoides*						
Wild Australian	0.533					
Wild Indian	0.496	0.290				
Domesticated Indian	0.808	0.812	0.565			
Domesticated Tropical	0.812	0.829	0.573	0.050		
Philippines	0.712	0.667	0.448	0.165	0.179	

Interestingly, wild-Indian cluster was less differentiated from wild-*scarabaeoides* than was wild-Australian cluster, despite the relatively low level of genetic differentiation between the wild-Indian and wild-Australian populations. The limited differentiation between wild-

Indian and wild-Australian is consistent with previous phylogenetic and allele frequency analyses that establish these populations as sister groups. Thus, the lower differentiation of wild-India from wild-*scarabaeoides* (compared to wild-Australia from wild-*scarabaeoides*) is presumably a consequence of prolonged gene flow (Table 3.8) between these divergent yet co-resident Indian populations, a situation that would not have been possible between the geographically isolated wild-Indian (*scarabaeoides* and non-*scarabaeoides* clades) and wild-Australia populations.

Domesticated populations were considerably less differentiated from the wild-Indian population compared to either wild-Australian or wild-*scarabaeoides*, and this was mirrored by higher relative rates of gene flow between wild-Indian and domesticated populations.

Table 3.8. Pairwise estimates of Nei's genetic distance (lower diagonal) and gene flow (Nm) values (upper diagonal) among wild and domesticated groups

	Wild *scarabaeoides*	**Wild Australian**	**Wild Indian**	**Domesticated Indian**	**Domesticated Tropical**	**Philippines**
Wild *scarabaeoides*	****	0.219	0.254	0.059	0.058	0.101
Wild Australian	0.349	****	0.611	0.058	0.052	0.125
Wild Indian	0.476	0.135	****	0.192	0.186	0.309
Domesticated Indian	1.578	0.661	0.414	****	4.705	1.264
Domesticated Tropical	1.574	0.662	0.416	0.004	****	1.148
Philippines	1.254	0.549	0.333	0.028	0.029	****

In contrast, and despite the high rates of gene flow (the highest observed) and low genetic distance between wild-India and wild-Australia, wild-Australia had very low rates of gene flow with the primary domesticated populations (domesticated-Indian and domesticated-tropical). These results suggest that gene flow between wild-Indian and wild-Australian was archaic relative to domestication of *C. cajanifolius* from wild-India, or that domestication occurred from a somewhat isolated subpopulation within the wild-India population.

Utility and Polymorphic Information Content (PIC) of the COS markers

We calculated the polymorphism information content (PIC) of each marker across the data set of 95 genotypes (excluding admixed genotypes except the Philippines accessions), according to Botstein et al. (1980) using the formula:

$$PIC = 1 - \left[\sum_{i=1}^{n} P_i^2\right] - \left[\sum_{j=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 P_j^2\right]$$

Where P_i and P_j are the allelic frequencies of the ith and (i+1)th alleles. For biallelic SNP marker systems this formula can be modified as described by Petkov et al. (2004) using the formula PIC = 2pq, where *p* and *q* are the SNP allele frequencies and PIC can range from 0.0 to 0.5 (Botstein et al. 1980; Anderson et al. 1993). The average PIC value of the markers for the data set of 95 genotypes was 0.226. Among these markers, ~40% were estimated to be highly informative (with PIC > 0.25), 39% were informative (0.1 < PIC < 0.25) and 15% (113 markers) were slightly informative (with PIC < 0.01). To test the utility of the SNP markers in detecting polymorphism within the cultivated populations we analyzed the PIC values among these genotypes (with the exceptions of the Philippines accessions) and found only 5% (about 38 of the SNP markers) that were highly informative among the domesticated group (Fig. 3.10). This result is not surprising given the low rates of polymorphism within the domesticated groups.

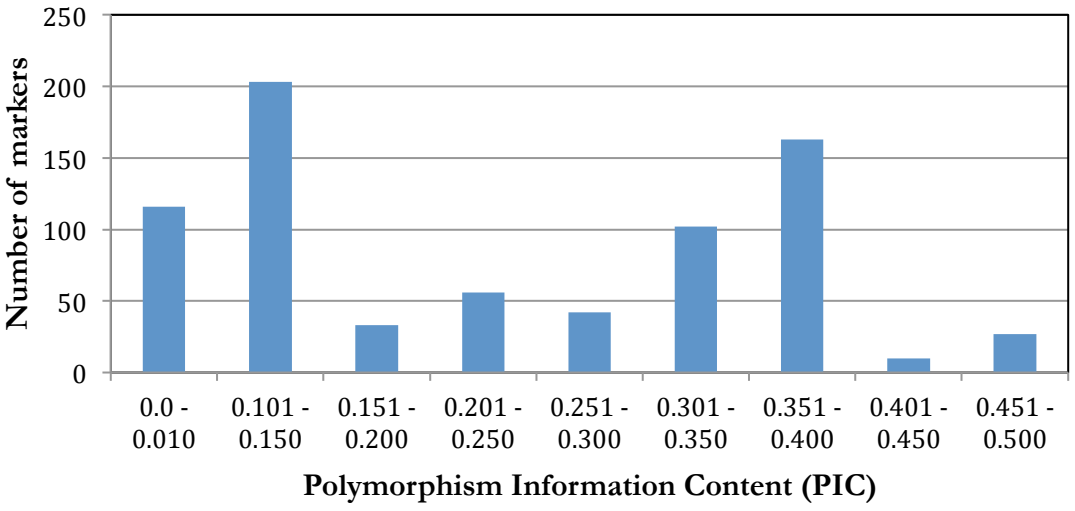


Fig 3.10. Distribution of Polymorphic Information Content (PIC) of the SNP-COS markers among 95 genotypes.

Discussion

Here we have investigated the genetic diversity and population structure of domesticated pigeonpea and its wild relatives in the genus *Cajanus*. The SNP markers employed in this study had an average PIC of 0.226 and 79% were either highly informative (40% with PIC > 0.25) or informative (39% with 0.1 < PIC < 0.25; Fig. 3.10.). The data were sufficient to derive relationships that were simultaneously congruent with and more detailed than previous single gene phylogenies. When combined with patterns of population structure inferred from allele diversity, it was possible to derive estimates of genetic diversity and differentiation within and among groups at various levels of organization. As described below, this permitted assignment of the most probable progenitor species, and allowed us to infer the origin of modern cultivated pigeonpea from nested domestication events, the older of which is focused in India and a second more recent event focused in the tropics.

In the present study, abundant allelic variation and genetic diversity was observed in wild *Cajanus* accessions, as compared to limited genetic diversity in domesticated pigeonpea. In particular, we estimate that ~75% of ancestral allele diversity was lost during domestication. Reduced genetic diversity for *C. cajan* has been reported in previous molecular studies using Diversity Array Technology (DArT) markers (Yang et al., 2006) and Simple Sequence repeat (SSR) markers (Odeny et al., 2007). Despite the genetically narrow base of pigeonpea, the cultigen is noted for high levels of morphological diversity. Thus, different genotypes are adapted for acceptable agronomic yield in both tropical and semi-arid regions of the world, as reflected in the eco-geographical variation in collection sites for accessions used in this study.

Although there was no clear distinction between landraces and modern cultivars, domesticated genotypes were resolved into two sections based on weighted Neighbor Joining analysis. This same division was evident when domesticated accessions were subdivided using allele frequency data (STRUCTURE). The subdivision reflects the geographical origin of the respective genotypes, further supporting the validity of the groups, with one clade of Indian origin containing approximately twice the genetic diversity of the second clade of tropical origin. Both of these populations are depauperate of genetic diversity, with low genetic differentiation and low genetic distance between them. The phylogenetic and genetic signatures of *C. cajan* shown in Figures 3.2 and 3.6, respectively, are consistent with a single major domestication event in India. Furthermore, we suggest that the genetic distinctiveness of the domesticated-tropical clade (Fig. 3.4) results from a more recent secondary domestication event in tropical regions. The distinctiveness of the tropical and Indian subgroups within *C.*

cajan may derive either from differing parental contribution during breeding for the tropical vs semi-arid environments, or alternatively repeated and independent selection for common adaptive alleles in the tropical environments. The presence of 30% Indian accessions in the "tropical" group is more consistent with the former suggestion, because it indicates that the respective haplotypes pre-existed in the Indian situation.

The only exception to high genetic diversity among wild accessions was observed in wild accessions collected from Australia. This situation is curious, because the wild-Australia clade contains seven distinct taxonomic species, yet possesses less than one-third the genetic diversity of the taxonomically homogeneous *C. scarabaeoides* clade of Indian origin. The majority of these Australian species are endemic to Australia and possess similar morphological characters (e.g. leaf shape, leaf and flower color and the growth habit; Yang et al., 2006), probably due to the narrowly distributed endemics of recent evolutionary origin. In fact, Australia is designated as an important center of species diversity for *Cajanus* (Nene and Sheila, 1990), but our results argue against this conclusion because genetic diversity was quite low among the seven species used in this analysis. Nevertheless, the expansive geography of Australia combined with limited efforts to collect *Cajanus* from this region suggest that additional Australian *Cajanus* species are likely to be found (van der Maesen, personal communication). In any case, it seems appropriate to consider revising the taxonomic status of these Australian species using combined data set of molecular and morphological markers.

The low genetic differentiation and high estimated gene flow between wild *Cajanus* in Australia and the non-*scarabaeoides* wild *Cajanus* in India suggests that these species were recently part of a common population. Moreover, the observation that the genetic diversity of *Cajanus* in Australia is one-third of that found in related Indian species, with polymorphism rates comparable to modern domesticated pigeonpea, is consistent with a bottleneck from India to Australia. Could human migration have been a factor? It is likely that migrating humans carried seed for nutrition, if not planting; if so, then genetic drift would have had pronounced effects on the characteristics of even casually collected seed stocks. Interestingly, genetic evidence suggests that humans may have colonized Australia by migration from the Indian subcontinent.

A proposed but controversial early migration route includes movement from the Indian sub-continent to Australia in the late Pleistocene, i.e., >10,000 years ago (Macaulay et al., 2005), while a proposed more recent event corresponds to changes to the anthropological record in Australia around 5,000 to 3,000 years ago (Redd et al., 2002). Although we have no

evidence that humans either cultivated or carried pigeonpea along this migration route, the apparent origin of related *Cajanus* spp in India and the presence of a narrow genetic base of derived *Cajanus* species in Australia are consistent with this possibility. If true, then this "Australian-focused *Cajanus* bottleneck" was entirely independent of the recognized Indian-focused domestication, because the modern cultivated pigeonpea is genetically distinct from its Australian relatives.

Genetic Structure and tracking historical admixture between domesticated and wild groups

van der Maesen (1980; 1990) proposed India as the center of origin for pigeonpea based on the presence of several wild species, the existence of high morphological diversity, the presence of diverse sets of primitive (or feral) varieties, and ample linguistic and cultural evidence. Regions containing wild populations that are phylogenetically close to cultivars are often proposed as the domestication site of crops (Matsuoka et al., 2002; Spooner et al., 2005). Thus our results support van der Maesen's hypothesis, because India harbors the highest level of genetic diversity of related wild species. This study could not confirm the alternative hypothesis of African origin (Purseglove, 1968; Tindall, 1988) due to the lack of sampling wild *Cajanus* species from Africa. However, an AFLP study (Wasike et al., 2005) as well as our current data demonstrates that African pigeonpea cultivars are less genetically diverse than, and apparently derived from, Indian cultivars.

Cajanus cajanifolius has been proposed as the direct progenitor of pigeonpea, and in this study *C. cajanifolius* was sister to the domesticated cluster in Neighbor-Joining analyses. This is consistent with the high morphological resemblance between *C. cajan* and *C. cajanifolius*, differing mainly in the prominence of the cultigen's strophiole. Gene flow subsequent to initial domestication is also likely to have contributed to the character of cultivated pigeonpea, and the haplotypes of several accessions provided evidence of recent genetic admixture between wild and cultivated gene pools.

The potential for gene flow is significant, as insect aided natural out-crossing for pigeonpea is in the range of 20-70% (Bhatia et al. 1981; Ratnaparkhe et al., 1995; Souframanien et al., 2003) and recently up to 17 % natural out-crossing has been observed for wild species (Saxena and Kumar, 2010b). The highest out-crossing rate (17%) among wild species was recorded for *Cajanus lineatus* (Saxena and Kumar 2003). van der Maesen (personal communication) notes that multiplication of *C. lineatus* in experimental gardens is

common practice in India and therefore the occurrence of spontaneous hybrids involving *C. lineatus* and nearby cultivated *C. cajan* should be expected and has been observed. Here we identified one highly admixed genome described as *C. lineatus*.

High rates of gene flow from the wild-India population were estimated for two Philippines accessions (ICP12765 and ICP10880), ICRISAT reference accession (ICP11975), and two accessions of *C. cajanifolius* (C-7847 and ICPW32) collected from the field. Two accessions of *C. cajan* (ICP15629 and ICPW29) possess genomes that are predominantly domesticated, but with 5-10% membership of the wild-Indian group. The latter accessions are of interest because they served as parental lines to develop a stable cytoplasmic male sterility (CMS) system in pigeonpea. The CMS accession, ICP2039A, was derived from an inter-specific hybrid of ICPW29 and cultivar ICP 11501 (Saxena et al., 2005); we speculate that this intentional hybridization and repeated backcrossing may have contaminated the genome of these parental accessions.

Genetic bottleneck severity in pigeonpea domestication

Crop domestication is accompanied by genome-wide reductions to genetic diversity (Tanksley and McCouch 1997). This reduction in genetic diversity derives from a population bottleneck imposed during the founding of a new crop lineage (Eyre-Walker et al. 1998) and subsequently due to selection on specific loci that confer agronomically important traits (Tenaillon et al., 2004). Bottleneck severity varies among crop species depending on the duration of domestication and number of domestication events. For example several grasses have about two-thirds of the genetic diversity found in their wild relatives (Buckler et al., 2001), and simulations reveal a more severe bottleneck for rice than maize (Tenaillon et al., 2004; Zhu et al., 2007). Previous studies using Diversity Arrays Technology (DArT) (Yang et al., 2006) and Single Sequence Repeats (SSR) (Odeny et al., 2007) detected a reduction in level of genetic diversity in domesticated pigeonpea compared to wild relatives. These previous studies however were not able to quantify the bottleneck effect.

This study was able to quantify the level of reduction in genetic diversity using genetic diversity indices, indicating that ~75% of ancestral polymorphism was lost during domestication. It is noteworthy that landraces (primitive cultivars) and improved (elite) cultivars that comprise the domesticated portion of our genotype panel (Table 3.1) contained similar levels of polymorphic SNPs, indicating that much of the diversity survived through the incipient stages of domestication can be found in the improved cultivars. Similar genetic

bottleneck effects have also been observed in other crop species such as soybean (Guo et al., 2010; Hyten et al., 2006), sunflower (Liu and Burke, 2006), and lima beans (Motta-Aldana et al., 2010).

Out of 752 SNP markers used for genotyping, only 62 markers detected variation among the domesticated *Cajanus cajan* group (excluding the Philippines accessions). In contrast, 283 SNP markers were polymorphic among the progenitor wild-India accessions. Similarly low levels of polymorphic loci were reported for domesticated pigeonpea using DArT markers (Yang et al., 2006), with only 64 of 700 markers being polymorphic among the domesticated accessions.

Conclusions

The SNP (Single Nucleotide Polymorphism) markers employed in this study reported important information on genetic structure and diversity of domesticated and wild relatives of pigeonpea (*Cajanus cajan*). Cluster and genotype assignment analyses resolved two major groups consisting of the domesticated *Cajanus cajan* complex and a second group consisting of wild species. No distinct genetic differentiation exists between landraces and elite cultivars within the domesticated group. However, the domesticated group could be broadly divided into two sub-groups reflecting the eco-geographic situation in which they are cultivated. We interpret this result as owing to a single major domestication event in India, and a subsequent minor domestication event to adapt germplasm for cultivation in tropical regions. The wild group was also further differentiated into distinct sub-groups, one that contained the majority of wild accessions of Indian origin, a second cluster consisting endemic wild species of Australian accessions, and a third divergent group that contained wild accessions of *Cajanus scarabaeoides* primarily of Indian origin.

In contrast to previous morphological studies (e.g. van der Maesen, 1990), *Cajanus platycarpus* resolved to have more close affinity with the domesticated group, and instead placed *Cajanus scarabaeoides* as the most distant from the domesticated *Cajanus cajan* group. Thus, we propose a revision of the gene pool classification of *Cajanus*. This result is consistent with the recent plastid and *ITS* phylogenetic study of *Cajanus* and allied genera in which *Cajanus scarabaeoides* was resolved as the most basal and ancestral species among the core *Cajanus* group (Chapter 2 of this thesis). There was massive loss of genetic diversity during pigeonpea domestication. There was a reduction of about 75% in allelic diversity in domesticated pigeonpea as compared to the wild relatives (with the exception of wild species

from Australia), indicating a severe "domestication bottleneck" during pigeonpea domestication.

This study confirmed the long held alpha-taxonomic hypothesis that *Cajanus cajanifolius* is the most recent progenitor of cultivated pigeonpea and identified India as the most likely center of origin for pigeonpea domestication. However, crop domestication is most certainly a progressive process, with a hierarchy of selection events potentially spanning thousands of years and adapting germplasm to diverse eco-geographical conditions, as well as the intervention of intentional and accidental genetic admixture to impact allele content in the cultivated gene pool. Genetic structure and Hierarchical Analysis of Molecular Variance Analysis (AMOVA) analyses revealed genetic admixture between wild and cultivated genomes, suggesting the involvement of successive rounds of gene flow between wild and domesticated species. The wild gene pool of *Cajanus* contains not only high genetic diversity but also unique and rare alleles that might be linked to agronomically important traits (e.g. trichomes of *C. scarabaeoides* for pod borer resistance); thus pigeonpea breeding and improvement programs need to continue to utilize this bounty of genetic resources prevailing in the wild gene pool.

Chapter 4

Mining of Single nucleotide polymorphism (SNP) markers associated with disease resistance (NBS-LRR) genomic regions in pigeonpea (*Cajanus cajan* (L.) Millsp.) BAC-end sequences

Introduction

Plants have evolved a variety of defense mechanisms to protect themselves from a range of environmental pathogenic organisms and pests. One well-characterized defense mechanism is controlled by a diverse group of disease resistance genes that encode NBS-LRR proteins and represent one of the largest plant gene families (Martin et al. 2003). NBS-LRR proteins contain a centrally-located Nucleotide Binding Site (NBS) and a C-terminal Leucine-Rich Repeat (LRR) domain (Hulbert et al., 2001). NBS domains bind and hydrolyze ATP and GTP (Meyers et al. 1999), while the LRR motif is involved in protein–protein interactions and pathogen specificity (Dangl and Jones 2001).

Most NBS-LRR proteins are further divided into two distinct classes known as TIR and non-TIR types based on the identity of the sequences at their N-terminus (Meyers et al. 2003). The TIR-NBS-LRR class has an N-terminal Toll-interleukin-like receptor (TIR) domain, named based on homology to the *Drosophila* Toll and mammalian Interleukin-1 (IL-1) receptors. The N- terminal region of non-TIR-NBS-LRR class is less conserved, but often contains a coiled-coil (CC) domain (DeYoung and Innes 2006). In general, the non-TIR (CC) types are widely found in both monocot and dicot species whereas no TIR types have been identified in monocots (Tarr and Alexander, 2009). The NBS-LRR resistance gene family has been studied in numerous plant species, including the model organisms *Arabdopsis thaliana* (Tan et al. 2007) and *Oryza sativa* (Monosi et al. 2004; Zhou et al. 2004; Yang et al. 2008), as well as in several legume species including *Medicago truncatula* (Ameline-Torregrosa et al., 2008), *Vicia faba* L. (Palomino et al. 2006), and *Cicer arietinum* L. (Palomino et al. 2006). These studies demonstrate that the NBS-LRR gene family is one of the largest gene families in plants, accounting for 0.6–1.8% of the total genes in their respective genomes (Mun et al., 2009). Most NBS-LRR genes are not evenly distributed across the genome, but rather are clustered in specific chromosomal regions (Ameline-Torregrosa et al. 2008).

To gain insight into the genome evolution of the NBS-LRR domains and understand the evolution of disease resistance genes in legume species, University of California, Davis

designed degenerate primers based on the conserved motifs of the NBS homologs from the reference legumes *Medicago truncatula* and soybean (*Glycine max*) (Varshney e al., 2009b). The compiled NBS domains were used to extract the conserved P-loop and GLPL motifs, from which a set of 544 degenerate primer pairs were developed for use in PCR reactions (Rosen et al., unpublished data). These degenerate primers have been tested in the pigeonpea reference genotype Asha (ICPL87119) and more than 600 unique nucleotide binding site sequences were cloned and sequenced, yielding large sets of TIR and non-TIR NBS domains (Rosen et al., unpublished data). A representative subset of these NBS domains were hybridized to a pigeonpea bacterial artificial chromosome (BAC) library, yielding resistance gene-containing BAC clones that were fingerprinted to develop physical map contigs and end sequenced to yield resistance gene-associated BAC end sequences (BES).

Many cultivars of pigeonpea have little genetic polymorphism and lack sources of resistance genes to pests and diseases, while wild relatives possess high genetic diversity and are potential sources of disease resistance traits. Resistance Gene homolog-containing BAC end-sequences are an ideal genomic source from which to mine SNP markers, with the goal of developing a suite of genetic markers linked to disease resistance genes. Developing SNP markers for pigeonpea breeding, especially in the vicinity of agronomically important candidate resistance genes, should facilitate the development of elite and more robust pigeonpea varieties.

The main objective of this research is to mine BES-based SNP markers associated with disease resistance loci that may potentially be used for marker-assisted breeding and for integration into the pigeonpea genetic map.

Materials and Methods

Plant material, DNA extraction, PCR amplification

The BAC end sequences from the reference pigeonpea genotype ICPL 87119 (Asha) were used to design oligonucleotide primers, which in turn were used to obtain PCR products from the wild species ICPW94 (*Cajanus scarabaeoides*). Total DNA was extracted from fresh leaves using the Qiagen DNeAsy protocol, according to manufacturer's instruction. PCR amplifications were carried out in 10 µl volumes, containing 10 ng genomic DNA, 1 µl of 10X *Ex Taq* PCR buffer, 0.8 µl of *ExTaq* dNTPs, 0.05 µl of Taq DNA polymerase and 1.4 µl of forward and reverse primers mix (5 µM each). For PCR, a DNA engine Tetrad2 peltier thermal

cycler (Bio-RAD) was programmed for an initial denaturation at 95 Cº for 3 min, followed by 40 cycles consisting of denaturation at 94 Cº for 20 s, annealing at 56 Cº for 20 s, and extension at 68 Cº for 1 min 30 s, with a final extension at 68 Cº for 5min. PCR products were prepared for sequencing by treatment with ExoSAP to remove unincorporated primers. PCR products were sequenced using Applied Biosystems (ABI) BigDye terminator chemistry on an ABI 3730xl DNA Analyzer at the University of California-Davis Genomics Facility.

Source of BAC-end sequences (BES)

Resistance gene homolog-containing BAC clones were identified by filter hybridization with ^{32}d-CTP-labeled pigeonpea NBS derived probes. Most of the BAC clones yielded data when subjected to high information content fingerprinting (HICF) that was analyzed by means of fingerprint contig (FPC) software (Rosen et al., unpublished data). The same set of BAC clones were end sequenced at the J. Craig Venter Institute and submitted to the National Center for Biotechnology Information (NCBI) genome survey sequence database.

Sequence alignment and SNP validation interface

High quality sequences (PHRED score >20) from *C. cajanus* Asha and *C. scarabaeoides* ICPW94 were aligned and analyzed by an automated SNP-prediction tool developed at the National Center for Genome Resources in collaboration with UC Davis (Andrew Farmer, personal communication). Manual validation of the predicted SNPs was performed using the interface for manual verification of the predicted SNPs (Fig. 4.1.).

Identification of Open reading frames (ORFs)

All pigeonpea BAC end sequences were annotated using a combination of BLAST-based analyses to NCBI and tools such as the InterPro database (Hunter et al., 2009) and the Gene Ontology Consortium (Gene Ontology Consortium, 2010), as described in Bohra et al., (2011). SNP-containing BAC end sequences that were annotated as either Resistance Gene Homologs (RGHs), anonymous genes, or reteroelements were surveyed for open reading frame (ORF) content to identify potential coding sequences. ORFs were identified by means of BLASTX search of the non-redundant protein database of the NCBI and using the gene-finding program FGENESH (http://www.softberry.com).

Evolutionary Selection pressure analysis

To estimate evolutionary selection pressure and molecular adaptation among the different gene classes of the BAC-end sequences, an analysis was conducted to quantify the rates of non-synonymous substitutions per non-synonymous site (dN) and synonymous substitutions per synonymous site (dS) of each open reading frame (ORF). The overall dN/dS ratio or ω was computed for all ORFs derived from BAC end sequences annotated as anonymous genes, RGHs or reteroelement containing sequences. All analyses were conducted using the Nei-Gojobori model in MEGA4 (Tamura et al. 2007). The Analyses were based on pairwise analysis of two sequences and standard error was estimated by a bootstrap procedure (500 replicates). All gaps and missing data in the alignment were eliminated using the pairwise deletion option.

Results

BAC end sequence analyses

A total of 2271 non-redundant BAC-end sequences were derived from 1471 pigeonpea BAC clones that also hybridized to NBS probes. Clustering of these BAC-end sequences based on 95% sequence identity and 100 bp overlap yielded a set 1757 sequence clusters. In parallel, 1235 of the 1471 BAC clones yielded high information content fingerprint data and these 1235 clones were condensed into 71 contigs and 224 non-contiged singletons. Thus, >82% of BAC clones comprised 71 contigs, with contigs varying in size from 2 clones (120 Kb) to 363 clones (>2 Mbp). One thousand nine hundred and eighty one (1981) BAC-end sequences are associated with these fingerprinted BAC clones, while 290 BAC end sequences are associated with clones that failed fingerprinting.

The BAC-end sequences were analyzed by means of BLAST and InterPro (Hunter et al., 2009) to assess protein coding potential and classified into four major categories, as described in Bohra et al., (in review). (1) Resistance gene containing (RGH); (2) Gene containing (gene); (3) reteroelement (RE) containing, and (4) not annotated (NA). Non-annotated (NA) sequences represent the majority of RGH-associated BAC end sequences, accounting for 51% of all the non-redundant singletons and clusters. Equal proportions of BAC end sequences were annotated as genes (23%) or retroelements (RE) (23%), while 3% of RGH-hybriziding BAC end sequences were annotated as resistance gene containing (RGH) sequences. The total sequence length of the combined BES set is 1.9 Mbp (Table 4.1).

Table 4.1. RGH containing BAC end sequence characteristic

Annotation	Genes	RGH	RE	NA	Totals
Total ends	559	99	536	1,256	2,450
Total unique clusters[1]	489	70	499	1,118	2,176
Total unique sequence[2]	451,896	62,943	573,144	840,075	1,928,057

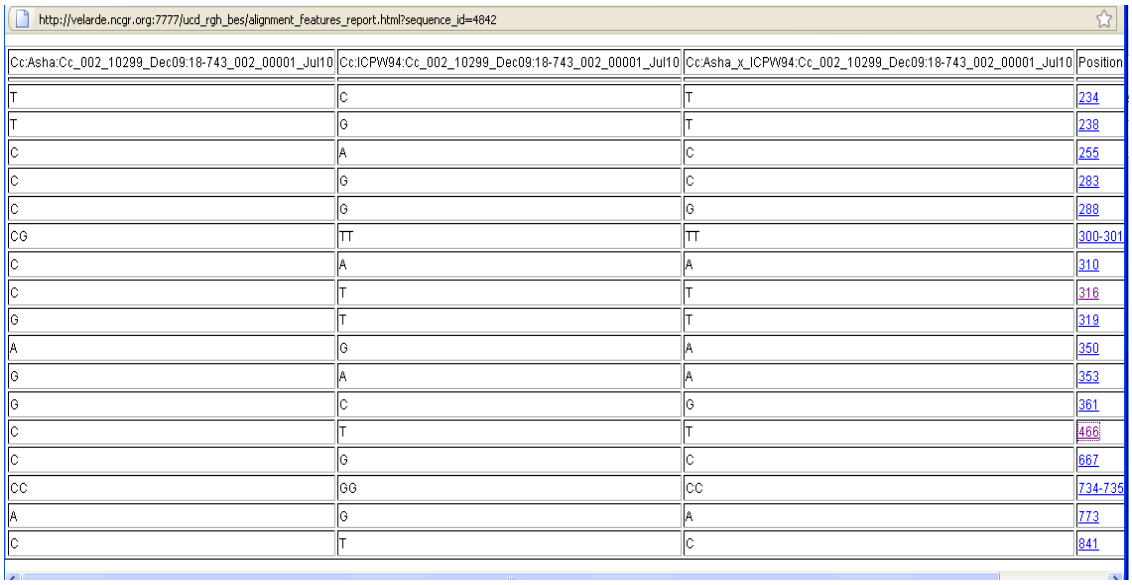
[1]Total unique clusters represents the total number of sequence clusters plus the number of singleton (non-clustered) sequences.
[2]Total unique sequence represents the sum of the nucleotide length of all unique sequence clusters.
RGH-Resistance gene homologs.
RE-Reteroelements.
NA- Not annotated.

Mining of BES-SNP markers

For purposes of SNP discovery, we first filtered BAC end sequences to exclude those sequences that were annotated as chloroplast, mitochondrial or ribosomal DNA, as well as those that were predicted to be high copy number based on sequence cluster depth; this later filter eliminated many retroelement-like sequences. With the goal of identifying SNP in as many singleton BAC clones and BAC fingerprint contigs as possible, BES were prioritized by selecting all ends from singleton and non-fingerprinted BAC clones, and up to five BES for each BAC contigs with the exception of 30 of the large contigs surveyed at >5 BES). This strategy resulted in 768 BES as templates for SNP discovery, including all of the BAC ends annotated as RGH-containing. Primer pairs were designed for each of these 768 BAC end sequences and used to amplify and obtain sequence data from *C. scarabaeoides* ICPW94.

Low quality bases in the ICPW94 sequences were masked and the resulting sequences used for alignment with the corresponding *C. cajan* vs. Asha BAC end sequence, and from this alignment a list of mismatched bases was generated using software produced as a collaboration between NCGR and UC Davis. Manual validation of the computationally predicted SNP calls was performed using the interface for manual verification (Fig. 4.1.). Predicted SNP calls with ambiguity including double peaks, mixed product, and divergent flanking sequences were culled from the analyses, retaining only well-supported nucleotide differences.

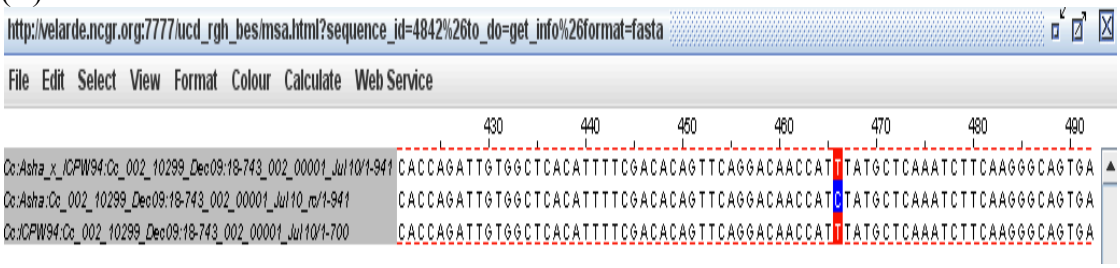
(A)



http://velarde.ncgr.org:7777/ucd_rgh_bes/alignment_features_report.html?sequence_id=4842

Cc:Asha:Cc_002_10299_Dec09:18-743_002_00001_Jul10	Cc:ICPW94:Cc_002_10299_Dec09:18-743_002_00001_Jul10	Cc:Asha_x_ICPW94:Cc_002_10299_Dec09:18-743_002_00001_Jul10	Position
T	C	T	234
T	G	T	238
C	A	C	255
C	G	C	283
C	G	G	288
CG	TT	TT	300-301
C	A	A	310
C	T	T	316
G	T	T	319
A	G	A	350
G	A	A	353
G	C	G	361
C	T	T	466
C	G	C	667
CC	GG	CC	734-735
A	G	A	773
C	T	C	841

(B)



(C)

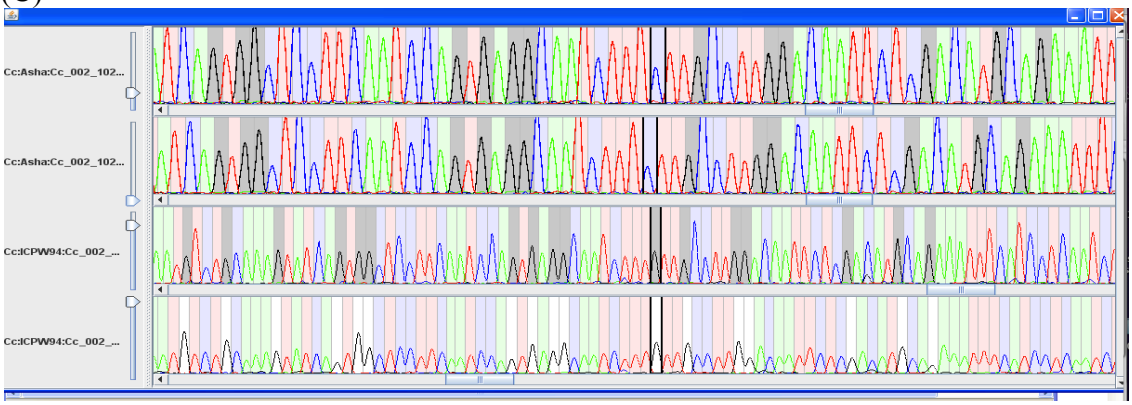


Fig. 4.1. User interface for manual validation of computationally predicted SNPs in BES in pigeonpea. Example of SNP position 466 in BES (Cc_002_10299_Dec09) between pigeonpea genotype ICPL 87119 ("Asha") and *C. scarabaeoides* (ICPW94). (A) SNP table of position of the computationally predicted SNPs (B) Multiple sequence alignment (MSA) of the region surrounding SNP pos 466. (C) Chromatogram window of region flanking the predicted SNP selected in panel (A). MSA and chromatograms are automatically adjusted and the corresponding SNP highlighted when a new SNP is selected in panel A. (Varshney et al., 2010).

Validated SNPs across genomic regions

Among the 768 BAC ends which were tested for PCR and sequencing, we selected 304 highly clean sequences for mining the SNPs. In total 2985 SNPs were identified among 304 of the *C. cajan* and *C. scarabaeoides* sequence alignments. The proportion of SNPs identified from each annotation category is shown in Fig. 4.2. and Table 4.2. The highest number of SNPs were associated with non- annotated sequences (55%), which was also the most frequent annotation category. The sequence characteristics of the SNP containing BES are shown in Table 4.2 and Appendix 3.

Average SNP frequency (SNP per 100 bp) among the BAC-end sequences of the four annotation categories was highest in non-annotated (1.37 SNPs/100bp) and reteroelement containing sequences (1.34 SNPs/100bp), and lowest in BES annotated as either RGH-containing (1.09 SNP/100 bp) or gene-containing (1.0 SNP/100 bp) (Table 4.2).

Single Nucleotide Polymorphism (SNP) frequency

Most of the BAC-end sequences in the four annotation groups have SNP frequencies of less than 1 SNPs/100bp, with the highest number of SNPs/100bp observed in unannotated sequences (6.1. SNPs/100bP) For BAC end sequences annotated as genes the highest SNP frequency was 5.4 SNPs/100bp, followed by RGH sequences (4.5 SNPs/100bp) and reteroelement containing sequences (4 SNPs/100bp) (Fig. 4.3).

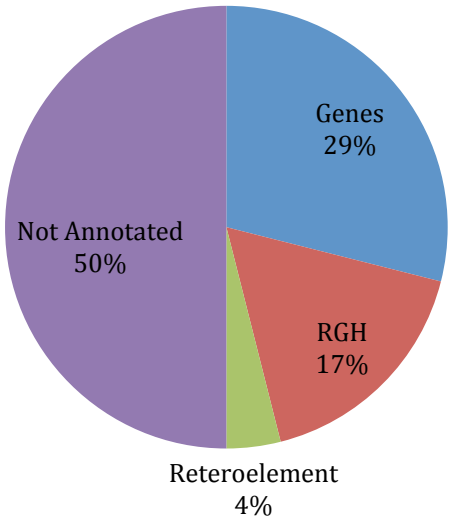


Fig 4.2. Polymorphic SNPS in each genomic category.

Coverage of SNP discovery in clusters of the BAC-end sequences

Among the 304 SNP containing BAC-end sequence, 47 were associated with singleton BAC clones (total of 463 SNPs), 172 were identified associated with BAC fingerprint contigs, and 85 SNP containing sequences (total of 745 SNPs) were associated with BAC-end sequences for which the BAC clones failed fingerprinting. All BAC clones were positive for the hybridization with NBS domain probes. Greater than 80% of all BAC clones were associated with 71 contigs, of which 64 contigs (90%) contained validated SNPs. Two unusually large contigs, namely contig 7 and contig 136, yielded 79 and 98 SNPs respectively. The distribution of SNP among these 64 contigs is shown in Fig. 4.4 and Appendix 4.1. Note that there is little if any correlation between SNP discovery and contig size, because we limited the number of BES surveyed for SNPs to five (with the exception of very large contigs).

Table 4.2. Sequence characteristics of SNP containing BAC-end sequences

	RGH	**Genes**	**Reteroelement**	**Unannotated**
No of Amplicons with SNPs	52	88	12	152
Average length (bp)	843	829	856	784
Average depth	2.17	1.5	2.16	1.5
No of singletons	1	18	0	28
No of contigs	20	27	11	44
No of unclustered BES	17	27	0	41
Total SNPs	478	736	138	1633
SNPs/100bp	1.09	1	1.34	1.37

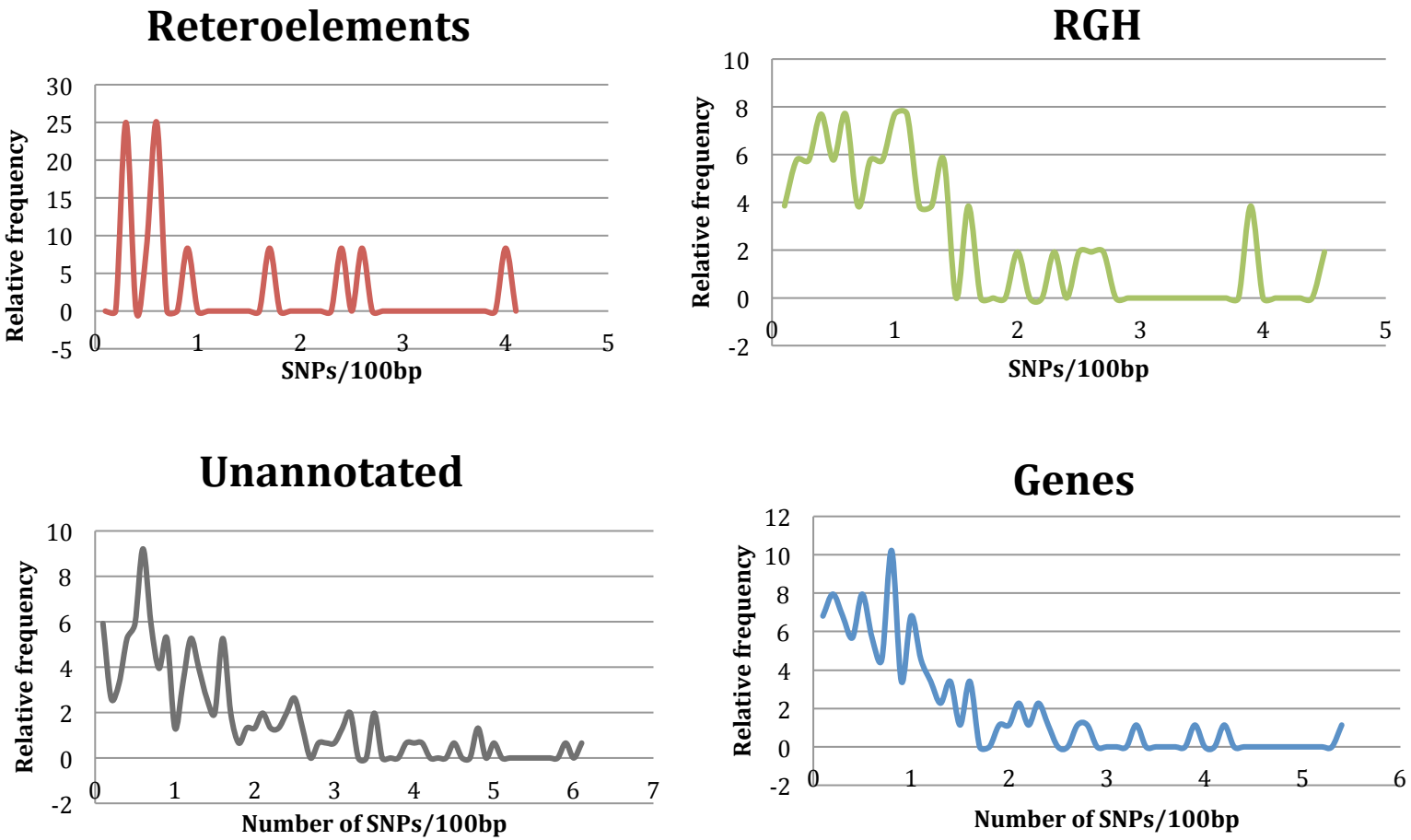


Fig 4.4. SNP frequency across the four genomic categories.

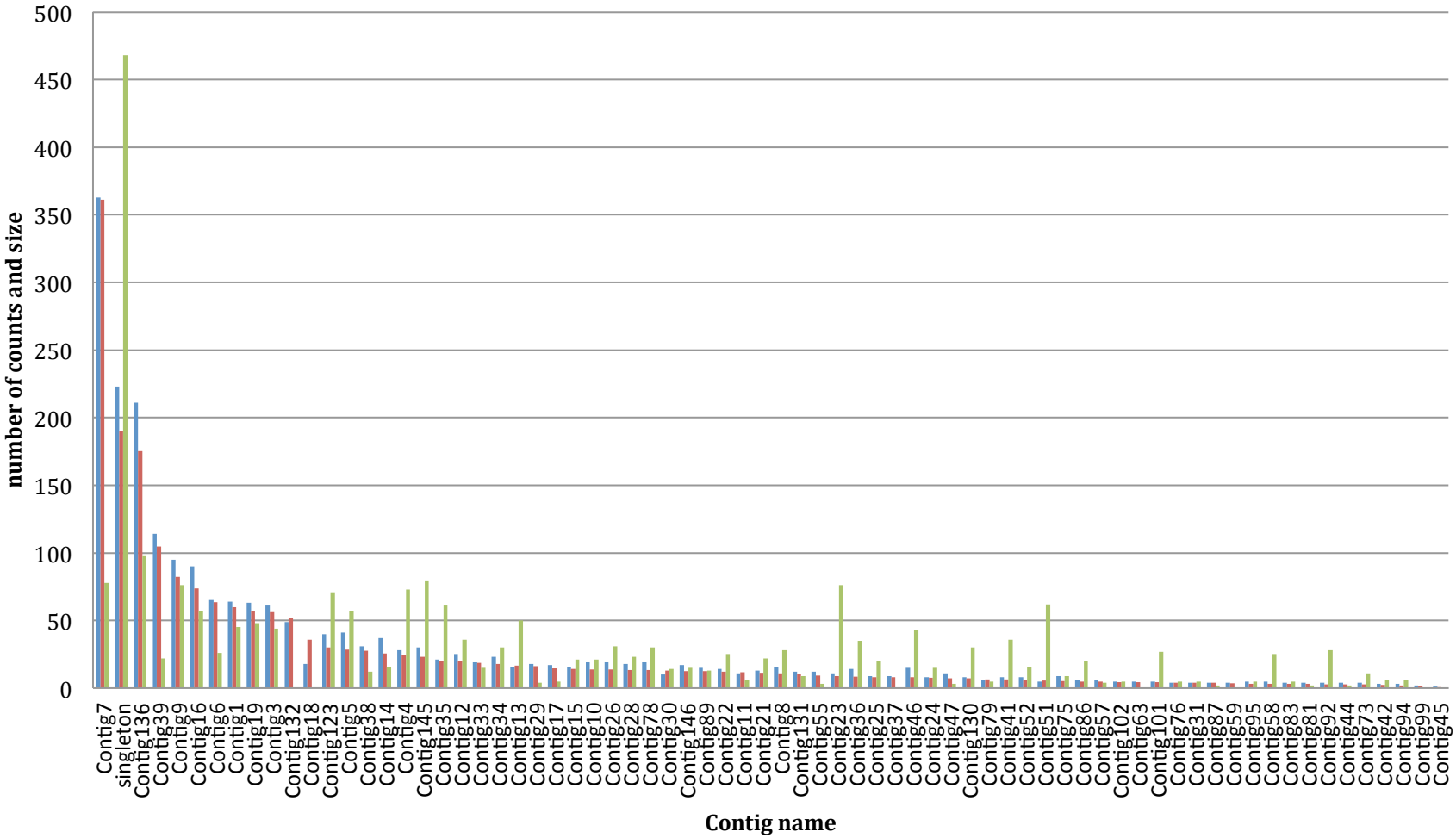


Fig. 4.4. Number of clones, contig size and number of valid SNPs identified in singletons and contigs of the RGH associated BES.

Assessing evolutionary selection pressure on coding sequences of the BES

To identify potential coding sequences within BES annotated as "genes", "RGH" or "retroelement", open reading frames were predicted using a combination of BLASTx to query the NCBI non-redundant protein database and the *ab initio* FGENESH HMM based gene prediction tool (http://www.softberry.com). Among the 152 annotated sequences (genes, RGH and reteroelements), 120 BAC-end sequences contained identifiable open reading frames. Of these, 58 open reading frames were identified from BAC sequences annotated as genes, 52 open reading frames were from BAC sequences annotated as RGH, and 10 open reading frames were identified in BAC sequences annotated as reteroelements.

The occurrence of synonymous (dS) and non-synonymous (dN) substitutions is one means to infer the history of selection pressure on sets of genes, although there is considerable debate about the utility of such measures (Kryazhimskiy and Plotkin, 2008). Sudhir Kumar (personal communication) recommends using the difference (dN-dS) rather than the ratio (dN/dS) because statistical tests are better suited for differences than ratios. Here we didn't test statistical significances on the value of the ratio (dN/dS), rather our interest was to classify the three types of selection pressures on the different genomic regions. We used the MEGA4 software package and the Nei-Gojobori model to evaluate dN and dS in each of the identified ORFs. The overall dN/dS ratios for BES annotated as RGH and genes was highly variable, although on average dN/dS ratios were almost twice as large for RGH (0.79 +/- 0.58) compared to geneic (non-RGH) loci (0.40 +/- 0.41).

These dN/dS values are consistent with a history of stronger purifying selection on non-RGH genes compared to RGH genes, and thus relaxed or in some cases possibly diversifying selection on RGH genes. The dN/dS ratio for reteroelement containing sequences was highest at 1.026, which could represent relaxed selection, but also highly variable. There was one ORF of the reteroelement sequence ("Cc_002_10460") that had an exceedingly high dN/dS ratio of 5.7. The number of non-synonymous and synonymous SNPs and the rate of non-synonymous and synonymous substitution for each ORF of the three genomic categories are shown in Tables 4.3-4.5

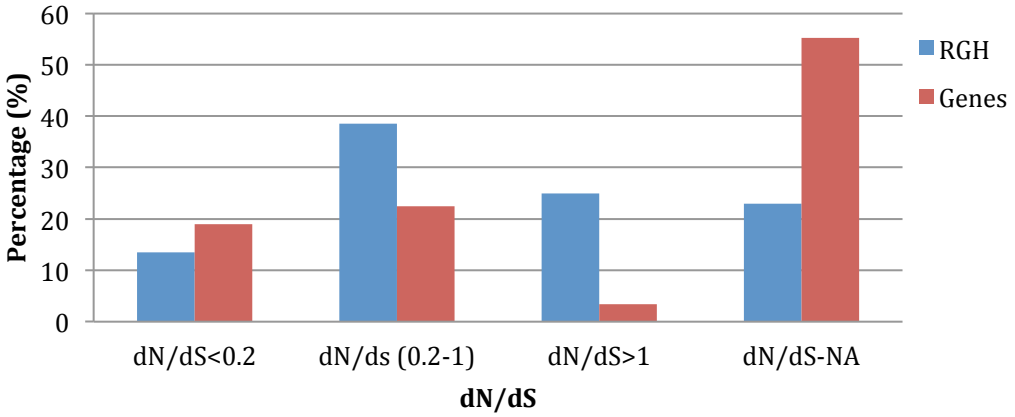


Fig 4.5. Percentage of ORF of RGH and anonymous genes with the respective dN/dS values.

Despite the overlap in average dN/dS ratios between sequences annotated as RGH compared to genes, the set of RGH-annotated ORFs possessed a higher proportion of sequences with high dN/dS ratio. Thus, 25% of RGH ORF sequences have dN/dS ratios greater than 1, compared to only 3% of the anonymous genes with dN/dS ratio greater than 1 (Fig. 4.5.). In fact, seven RGH ORFs possessed dN/dS between 1.2 and 2.0, consistent with the possibility of diversifying selection. By contrast, the set of ORFs annotated as genes contained a greater percentage of low dN/dS values (19 %) (i.e., dN/dS less than 0.2), compared to 13.5 % ORFs annotated as RGH. Our ability to calculate dN/dS was limited for BES ORFs that were annotated as genes, because >50% of such ORFs lacked SNPs at either synonymous or non-synonymous sites, which precluded calculating a value for dN/dS.

Table 4.3. The dN/dS ratio calculated for each ORF of the RGH region.

Sequence Name (ORF)	nSNPs[1]	Potential nSNPs	dN	sSNPs[2]	Potential sSNPs	dS	dN/dS
Cc_002_00375_Dec09	4	664.5	0.006	1	178.5	0.0056	1.0714
Cc_002_00516_Dec09	5	999.0833	0.005	0	293.9167	0	NA
Cc_002_00920_Dec09	15	664.0833	0.0226	9	190.9167	0.0471	0.4798
Cc_002_01912_Dec09	1	780.6667	0.0013	1	248.3333	0.004	0.3250
Cc_002_02074_Dec09	1	635.6667	0.0016	1	195.3333	0.0051	0.3137
Cc_002_02295_Dec09	1	665.6667	0.0015	0	174.3333	0	NA
Cc_002_02547_Dec09	1	811.5	0.0012	1	220.5	0.0045	0.2667
Cc_002_02644_Dec09	3	599.8333	0.005	0	165.1667	0	NA
Cc_002_02801_Dec09	9	914.25	0.0098	4	261.75	0.0153	0.6405
Cc_002_02959_Dec09	6	635.5	0.0094	1	195.5	0.0051	1.8431
Cc_002_03223_Dec09	4	734	0.0054	0	193	0	NA
Cc_002_03340_Dec09	3	679.0833	0.0044	4	178.9167	0.0224	0.1964
Cc_002_03447_Dec09	4	676.9167	0.0059	2	172.0833	0.0116	0.5086
Cc_002_04272_Dec09	7	933.9167	0.0075	3	296.0833	0.0101	0.7426
Cc_002_04584_Dec09	11	651.4167	0.0169	3	203.5833	0.0147	1.1497
Cc_002_06381_Dec09	11	664.1667	0.0166	1	175.8333	0.0057	2.9123
Cc_002_06507_Dec09	4	662.3333	0.006	1	198.6667	0.005	1.2000
Cc_002_07431_Dec09	9	670.9167	0.0134	4	196.0833	0.0204	0.6569
Cc_002_08247_Dec09	2	803.6667	0.0025	1	222.3333	0.0045	0.5556
Cc_002_09764_Dec09	1	636.3333	0.0016	2	182.6667	0.0109	0.1468
Cc_002_09767_Dec09	7	559.6667	0.0125	0	163.3333	0	NA
Cc_002_09916_Dec09	5	657.5	0.0076	0	176.5	0	NA
Cc_002_10156_Dec09	21.5	597.75	0.036	5.5	173.25	0.0317	1.1356
Cc_002_10277_Dec09	5	660.1667	0.0076	0	167.8333	0	NA
Cc_002_10278_Dec09	8	622.25	0.0129	3	172.75	0.0174	0.7414
Cc_002_10280_Dec09	0	649.5	0	3	190.5	0.0157	0
Cc_002_10299_Dec09	3.5	698.1667	0.005	1.5	195.8333	0.0077	0.6494
Cc_002_10312_Dec09	23	641.4167	0.0359	10	186.5833	0.0536	0.6698
Cc_002_10334_Dec09	1	687.5	0.0015	3	188.5	0.0159	0.0943
Cc_002_10373_Dec09	1	671.8333	0.0015	2	216.1667	0.0093	0.1613
Cc_002_10398_Dec09	5.5	691.8333	0.0079	1.5	184.1667	0.0081	0.9753
Cc_002_10403_Dec09	14	658.4167	0.0213	2	178.5833	0.0112	1.9018

Table 4.3. Continued…….

Cc_003_08690_Dec09	1	562.3333	0.0018	0	169.6667	0	NA
Cc_003_08861_Dec09	1	929.6667	0.0011	2	246.3333	0.0081	0.1358
Cc_003_10179_Dec09	3	857.3333	0.0035	1	228.6667	0.0044	0.7955
Cc_003_10480_Dec09	7	693.75	0.0101	2	194.25	0.0103	0.9806
Cc_003_10524_Dec09	5	686.5	0.0073	1	171.5	0.0058	1.2586
Cc_004_03654_Dec09	3	715.75	0.0042	3	208.25	0.0144	0.2917
Cc_004_10225_Dec09	20	864.0833	0.0231	4	227.9167	0.0176	1.3125
Cc_004_10249_Dec09	4	727.8333	0.0055	0	208.1667	0	NA
Cc_005_01748_Dec09	3	670.3333	0.0045	1	184.6667	0.0054	0.8333
Cc_005_06482_Dec09	5	688.4167	0.0073	0	196.5833	0	NA
Cc_005_08191_Dec09	0	793.6667	0	0	220.3333	0	NA
CccaBA_B_Rearray_1c5_002-A6TR	3	769.1667	0.0039	6	241.8333	0.0248	0.1573
CccaBA_B_Rearray_1c5_002-G6TR	3	801	0.0037	0	213	0	NA
CccaBA_B-Rearray-1c5-003-C10JF	19	731	0.0262	10.8333	217	0.0499	0.5251
CccaBA_B-Rearray-1c5-003-E24JF	5	629.3333	0.0079	1	168.6667	0.0059	1.3390
CccaBA_B-Rearray-1c5-003-O2JF	5	657	0.0076	1	189	0.0053	1.4340
cccaBa-Rearray-4-3F01TR	8	835.1667	0.0096	2	229.8333	0.0087	1.1034
CccaBa039-B6TV	6	790.5	0.0076	3	214.5	0.014	0.5429
CccaBa052-F18TR	8	711.75	0.0112	3	203.25	0.0148	0.7568
CccaBa064-D7TR	8	761.6667	0.0105	2	213.3333	0.0094	1.1170
Total	**313.6667**	**37155.7502**	**0.0084**	**113.3333**	**10484.2498**	**0.0108**	**0.79**

nSNPs[1]- number of non-synonymous SNPs.
sSNPs[2]- number of synonymous SNPs.
NA- not available.

Table 4.4. The dN/dS ratio calculated for each ORF of the anonymous genes.

Sequence Name (ORF)	nSNPs[1]	Potential nSNPs	dN	sSNPs[2]	Potential sSNPs	dS	dN/dS
Cc_002_02618_Dec09	11	568.6667	0.0193	2	184.3333	0.0108	1.7870
Cc_002_04374_Dec09	0	679.6667	0	0	172.3333	0	NA
Cc_002_04484_Dec09	1	699	0.0014	0	177	0	NA
Cc_002_05103_Dec09	1	664.1667	0.0015	0	178.8333	0	NA
Cc_002_05563_Dec09	1	686.5	0.0015	2	198.5	0.0101	0.1485
Cc_002_08818_Dec09	12.5	631.8333	0.0198	3.5	190.1667	0.0184	1.0761
Cc_002_09892_Dec09	2.5	769	0.0033	4.5	224	0.0201	0.1642
Cc_002_09986_Dec09	6	647.8333	0.0093	0	168.1667	0	NA
Cc_002_10214_Dec09	2	676.5	0.003	0	223.5	0	NA
Cc_002_10226_Dec09	2	729	0.0027	0	210	0	NA
Cc_002_10302_Dec09	1	606	0.0017	0	180	0	NA
Cc_002_10316_Dec09	0	648.5	0	0	194.5	0	NA
Cc_002_10329_Dec09	0	593.1667	0	0	180.8333	0	NA
Cc_002_10381_Dec09	1	603.3333	0.0017	4	161.6667	0.0247	0.0688
Cc_002_10400_Dec09	2	656.5	0.003	0	195.5	0	NA
Cc_002_10437_Dec09	0	638.1667	0	2	183.8333	0.0109	0.0000
Cc_002_10448_Dec09	1	675.5	0.0015	0	187.5	0	NA
Cc_002_10456_Dec09	29	623.3333	0.0465	13	195.6667	0.0664	0.7003
Cc_002_10472_Dec09	2	679.8333	0.0029	2	199.1667	0.01	0.2900
Cc_002_10479_Dec09	1	650.6667	0.0015	0	186.3333	0	NA
Cc_002_10482_Dec09	1	648.8333	0.0015	0	185.1667	0	NA
Cc_002_10574_Dec09	5	668.8333	0.0075	0	174.1667	0	NA
Cc_002_10585_Dec09	1	574.5	0.0017	0	163.5	0	NA
Cc_002_10644_Dec09	1	634.1667	0.0016	1	163.8333	0.0061	0.2623
Cc_003_02793_Dec09	4.5	788	0.0057	1.5	223	0.0067	0.8507
Cc_003_02932_Dec09	2	681	0.0029	0	165	0	NA
Cc_003_08205_Dec09	0	696.5	0	1	194.5	0.0051	0.0000
Cc_003_09985_Dec09	0	742.6667	0	0	208.3333	0	NA
Cc_003_10177_Dec09	1	916.6667	0.0011	1	274.3333	0.0036	0.3056
Cc_003_10184_Dec09	5	669.6667	0.0075	2	185.3333	0.0108	0.6944

Table 4.4. Continued…….

Cc_003_10307_Dec09	5	743.5	0.0067	5	204.5	0.0244	0.2746
Cc_003_10503_Dec09	0	688.3333	0	2	184.1667	0.0109	0.0000
Cc_003_10514_Dec09	2	676.5	0.003	0	178.5	0	NA
Cc_003_10621_Dec09	3	649.0833	0.0046	0	178.9167	0	NA
Cc_003_10677_Dec09	2	689	0.0029	3	196	0.0153	0.1895
Cc_003_10688_Dec09	2	627.3333	0.0032	1	191.6667	0.0052	0.6154
Cc_003_10692_Dec09	8.5	661.9167	0.0128	2.5	184.0833	0.0136	0.9412
Cc_003_10707_Dec09	2	674.8333	0.003	0	189.1667	0	NA
Cc_004_05777_Dec09	2	612	0.0033	0	195	0	NA
Cc_004_07311_Dec09	1	750.0833	0.0013	1	206.9167	0.0048	0.2708
Cc_004_10148_Dec09	2	1350.6667	0.0015	1	347.3333	0.0029	0.5172
Cc_004_10586_Dec09	3	683.0833	0.0044	0	198.9167	0	NA
Cc_005_00413_Dec09	0	938.5	0	0	246.5	0	NA
Cc_005_06174_Dec09	2	682.3333	0.0029	0	187.6667	0	NA
CccaBA_B_Rearray_1c5_001-P23TR	1	734	0.0014	0	187	0	NA
CccaBA_B-Rearray-1c5-003-E22JF	8	665.25	0.012	0	207.75	0	NA
CccaBA_B-Rearray-1c5-003-I20JF	2	710.1667	0.0028	0	201.8333	0	NA
CccaBA_B-Rearray-1c5-003-J13JR	2	732.3333	0.0027	1	200.6667	0.005	0.5400
CccaBA_B-Rearray-1c5-003-K14JR	3	773.3333	0.0039	0	225.6667	0	NA
CccaBA_B-Rearray-1c5-003-L24JF	2	672.1667	0.003	0	194.8333	0	NA
CccaBA_B-Rearray-1c5-003-M4JR	1	776.3333	0.0013	2	213.6667	0.0094	0.1383
cccaBa-Rearray-4-2A04TR	1	864.9167	0.0012	1	230.0833	0.0043	0.2791
cccaBa-Rearray-4-2A08TV	0	744.1667	0	0	203.8333	0.0049	0.0000
cccaBa-Rearray-4-3D08TR	1	874.6667	0.0011	0	220.3333	0	NA
CccaBb014-O13JR	4	692	0.0058	8	220	0.0364	0.1593
Cc_003_05008_Dec09	2	670.3333	0.003	0	193.6667	0	NA
Cc_003_04848_Dec09	1	602	0.0017	1	190	0.0053	0.3208
Cc_002_08253_Dec09	2	701.8333	0.0028	3	219.1667	0.0137	0.2044
Total	**161**	**40788.6666**	**0.0039**	**71**	**11336.8387**	**0.0063**	**0.40**

nSNPs[1]- number of non-synonymous SNPs.
sSNPs[2]- number of synonymous SNPs.
NA- not available.

Table 4.5. The dN/dS ratio calculated for each ORF of the reteroelements.

Sequence Name (ORF)	nSNPs[1]	Potential nSNPs	dN	sSNPs[2]	Potential sSNPs	dS	dN/dS
Cc_002_09394_Dec09	12	637.0833	0.0188	3	184.9167	0.0162	1.1605
Cc_002_10040_Dec09	8	720	0.0111	11	207	0.0531	0.2090
Cc_002_10258_Dec09	5	888	0.0056	1	264	0.0038	1.4737
Cc_002_10352_Dec09	3	582.1667	0.0052	0	197.8333	0	NA
Cc_002_10460_Dec09	20	664.6667	0.0301	1	190.3333	0.0053	5.6792
Cc_002_10619_Dec09	4	637	0.0063	0	179	0	NA
Cc_003_01602_Dec09	4	645.5	0.0062	1	227.5	0.0044	1.4091
Cc_003_06696_Dec09	1	633.3333	0.0016	1	161.6667	0.0062	0.2581
CccaBA_B-Rearray-1c5-003-M18JR	1	803.8333	0.0012	0	213.1667	0	NA
cccaBa-Rearray-4-3G09TR	8.8333	817.3333	0.0108	2.1667	205.6667	0.0105	1.0286
Total	**66.8333**	**7028.9166**	**0.01080**	**20.1667**	**2031.0834**	**0.0105**	**1.0259**

nSNPs[1]- number of non-synonymous SNPs
sSNPs[2]- number of synonymous SNPs
NA- not available

Discussion

The BAC-end sequences associated with RGH in Pigeonpea

Low genetic polymorphism in cultivated pigeonpea poses a challenge to development of molecular markers for use in breeding programs. Currently there are about 156 simple sequence repeat (SSR) markers published for pigeonpea (Odeny et al. 2007, 2009; Saxena et al. 2010c), although recently there are efforts to develop large scale SSR markers from pigeonpea BAC-end sequences (Bohra et al., 2011) and EST sequences (Dutta et al, 2011). BAC end sequences can be a useful source of molecular information. In other legume species BAC end sequences have been used to mine molecular markers to construct genetic maps as in the examples of soybean (*Glycine max*) (Shultz et al., 2007) and *Medicago truncatula* (Mun et al., 2006), and they can facilitate genetic studies of agronomically important traits as in the example of pea (*Pisum sativum*) (Coyne et al., 2007).

Low genetic diversity also makes pigeonpea cultivars vulnerable to wide array of disease and pests, because they lack allelic diversity and corresponding phenotypic variation that may confer resistance to these biotic stresses. The development of molecular markers associated with disease resistance genes has potential value for marker assisted breeding as well as for gene discovery and characterization. Here we mined BAC end sequences from resistance gene-containing BAC clones of pigeonpea for single nucleotide polymorphisms and discovered a suite of disease resistance gene-associated molecular markers.

Annotation of these BES for similarity to known genes and proteins indicated that 51% of these RGH-associate BES did not contain genes. Only 3% of these BES appear to encode portions of NBS-LRR disease resistance genes, while roughly equally proportions (i.e., 23% each) were annotated as either genes (non-RGH) or reteroelements. This distribution of annotation categories was similar to that observed in a larger set of 88,860 BES (including those BES analyzed here) where non-annotated sequences accounted for 53% of total BES, and genes (non-RGH) and reteroelements represented 21% and 22% respectively (Bohra et al., 2011).

SNP discovery in the BAC end sequences

The identification of ~3000 SNPs in 304 BAC-end sequences should be a valuable tool with which to integrate RGH loci into the developing pigeonpea genetic map. However, as we demonstrated in Chapter 3, only a limited number of SNP discovered in comparisons between *C. cajan* and *C. scarabaeoides* are also polymorphic in cultivated accession. The cause of this

attrition is the domestication bottleneck, but in any case it suggests that additional SNP discovery efforts focused specifically on cultivated material will be necessary. The many BES that were not investigated for polymorphism in this study could be used to expand the search for cultivar-level polymorphisms.

Almost 50% of the BES with SNPs were annotated as anonymous genes, reteroelements or resistance genes (RGH) and as such many BES sequences may be good sources of SNP markers linked with genic regions, in this case disease resistance loci. Plant NBS-LRR genes are typically non-randomly distributed with genomes, occurring in clusters that can contain up to dozens of distinct RGH loci and often spanning Mbp intervals. Our BAC fingerprinting and SNP discovery takes advantage of this feature of clustered RGH loci, by condensing several thousand BAC clones into a manageable number of BAC contigs that become targets ("mega loci") for SNP discovery. As consequence, this strategy enables mapping of large regions of contiguous RGH genome space with a limited number of SNP markers.

RGH clusters are also often rich in retrolements. Although on average these RGH associated BAC clones did not contain higher than average numbers of retroelements in their end sequences compared to the larger BES data set, it is also true that retroelements are themselves are unevenly distributed within plant genomes. Without further analysis, it is difficult to conclude whether the RGH containing BAC clone set is enriched in retroelements relative to other gene-containing regions of the genome, as opposed to retroelement-rich heterochromatin. The average cluster depth (in our sequence assemblies) of these SNP containing genes is about 2, which is consistent with most SNPs being derived from low copy sequences in the vicinity of RGH loci. In fact, we pre-filtered BES based on copy number with the explicit goal to avoid high copy loci that might confound the interpretation and discovery of genetic polymorphisms. The SNP coverage among the contigs is about 90% and this shows that most contigs have been mined for SNPs. The calculated average SNP frequency of 1 SNP per 100 bp is similar to SNP frequency observed in comparisons of other related plant species (Nasu et al., 2002; Ching et al., 2002).

Evolutionary selection pressure

Understanding the impact of evolutionary selection on genetic loci undergoing adaptation is a core objective of evolutionary biology. Using statistical models and rates of mutation in nucleotide sequences it is possible quantify selection pressure acting on protein-coding regions

(Kryazhimskiy and Plotkin, 2008). One of these methods of measuring selection pressure is the computation of the ratio known as dN/dS. Where dN is rate of substitutions at non-synonymous sites whereas dS is the rate of substitutions at non-synonymous sites (Kryazhimskiy and Plotkin, 2008). Elevated values of dN/dS ratio may be due to either positive selection that favors change or it may arise when the negative selection (also known as "purifying selection") against change has been relaxed (Wyckoff et al., 2000).

According to Ridely (2004) the three zones of dN/dS ratio associated with evolutionary impacts are 1) dN/dS <0.2 implies that there is no evidence that selection is driving the change in amino acids, also interpreted as purifying selection to remove changes that would othewise alter protein function; 2) dN/dS between 0.2 and 1 implies either positive selection on changed amino acids or that selection has been relaxed; 3) dN/dS >1is often interpreted as positive selection acting to fix amino acid changes in an allele. In this study we measured the dN/dS ratio of the individual open reading frames (ORF) of the three coding sequence classes: resistance genes (RGH), anonymous genes, and reteroelements.

The comparison between dN/dS ratio between resistance genes and anonymous genes revealed elevated dN/dS for RGH loci (Fig. 4.5). This agrees with a recent study showing strong positive selection on NBS-LRR disease resistance genes in *Arabidopsis*, especially in the LRR region of the NBS-LRR domain (Chen et al., 2010). However the current data set for pigeonpea is small and further detailed study will be required to properly assess the evolutionary pressure acting on resistance genes and genes in the vicinity of resistance genes in pigeonpea.

For future work, the SNPs identified in this chapter could be potentially a great addition for the useful SNPs utilized for genotyping and diversity studies in Chapter 3.

Conclusions

Low genetic diversity makes pigeonpea vulnerable to diseases and pests due to lack of allelic diversity that may confer resistance to these biotic stresses. The Development of molecular markers associated with diseses resistance genes has a potential value in molecular breeding programs. Resistance gene associated BAC end sequences (BES) are potential sources of molecular markers associated with disease resistance loci. Clustering of these BAC-end sequences based on sequence identity helps to condense the sequences into contigs and non-contiged singletons. BLAST analyses of the BES generated equal proportions of

annotated and non-annotated classes of the BAC end sequences. The annotated class comprises resistance genes, restroelaments and anonymous genes. Testing of the BAC end sequences through PCR and sequencing of *C. cajan* and amplicons of the wild species (*C. scarabaeoides*) generated a total of about 3000 SNP markers associated with disease resistance (NBS-LRR) genes which may potential be used for marker assisted breeding.

An evolutionary selection pressure was measured using a test of synonymous (dS) to non-synonymous (dN) substitution ratio. Comparision of the dN/dS ratio between resistance genes (R-genes) and anonymous genes showed an elevated dN/dS for the resistance genes that indicates strong positive selection on R-genes and this is consistent with recent study on other plant species such as *Arabidopsis* (Chen et al., 2010).

Chapter 5

Summary and future directions

Summary and key findings

This chapter summarizes findings of the studies detailed in Chapters 2-4. It also provides insights and directions for future research work. The thesis covers wide range of topics including evolutionary history of *Cajanus* and allied genera, history of domestication of pigeonpea, and the genetic diversity and genetic structure of wild and domesticated pigeonpea. It also attempts to track gene flow and historical hybridization between wild and domesticated species. The study also identified molecular markers (Single Nucleotide Polymorphisms, SNPs) associated with disease resistance loci in the genome that may potentially have applications in pigeonpea improvement and breeding. The study utilized high-throughput genomic technologies to dissect the genetics and evolutionary biology of pigeonpea and wild relatives, revealing important new findings some of which may have practical implications. Below is the summary of the key findings of this study.

Phylogeny of *Cajanus* and allied genera

There was a deficit of sampling of Cajaninae in previous comprehensive family-wide phylogeny studies of legumes (Kajita et al. 2001; Wojciechowski et al., 2004). This study reveals the molecular phylogeny based on sequence data, with improved taxon sampling, using plastid and nuclear markers, of the subtribe Cajaninae and focusing particularly on the relationships on the economically important genus *Cajanus*. This study found nr*ITS* region to be an excellent region in resolving the inter-generic phylogeny of Cajaninae and the inter-specific relationships in the genus *Cajanus*. The key findings are:

- Based on the presented data set, both individual and combined evidence from the plastid (*trnL-F* spacer) and nuclear (*ITS*) sequences supported the monophyly of the subtribe Cajaninae, but more taxon sampling of the subtribe is needed to confirm this hypothesis.
- This study revealed the non-monophyly of *Cajanus* and *Rhynchosia* and supported the monophyly of *Eriosema* and *Flemingia*. More sampling is needed from *Rhynchosia* and *Eriosema* to further investigate their relationships
- This study found *Dunbaria* and *Rhynchosia aurea* to be the closest taxa to species of *Cajanus*.

- This study supported the hypothesis from previous morphological, anatomical, chemotaxonomic and DNA fragment mapping studies (Lackey, 1981; Doyle and Doyle, 1993) that *Rhynchosia* and *Eriosema* are sister taxa.
- This study resolved the close relationship of South African monotypic *Bolusafra bituminosa* with *Rhynchosia-Eriosema* clade.
- This study partially supports the sectional classification of the genus *Cajanus*, based on morphology as proposed by van der Maesen (1986).
- This study resolved the sister relationship of *Cajanus cajan* (domesticated pigeonpea) and *Cajanus cajanifolius* (putative progenitor) and also resolved *Cajanus scarabaeoides* as the most basal species of the *Cajanus* clade.

Reinforcing Phylogenetic knowledge through genomics

One of the exciting findings of this study is the utility of Single Nucleotide Polymorphism (SNP) derived from low copy orthologous genes in resolving phylogenetic relationships of *Cajanus*. The SNPs were discovered in a comparison of amplicons between *cultivated* and *wild* accession of *Cajanus* and genotyped using the high throughput SNP-OPA Illumina golden gate assay. The results document the utility of SNPs to bolster the phylogenetic knowledge of the genus *Cajanus*.

- This study clearly resolved well-supported clusters, which reflect the distinctiveness of species and congruence with their geographical origin. There were three wild clusters and one domesticated cluster. The major clusters of wild species are the basal cluster of *C. scarabaeoides*, wild species of Australian origin and wild species of Indian origin.
- The domesticated accessions formed a cluster that was internal to, and significantly less diverse than, the group of non-scarabaeoides wild species of Indian origin.
- This study corroborates the *ITS* phylogeny and placed the putative progenitor, *Cajanus cajanifolius,* as the most closely related out-group to the domesticated clade, with strong bootstrap support (Fig. 5.1).
- This study found a close relationship between *C. cajanifolius* accessions and also between *C. platycarpus* and *C. sericeus* (Fig. 3.2). This indicates that *C. platycarpus* may not be genetically distant from pigeonpea as presently assumed (currently it is placed in the tertiary gene pool of pigeonpea).

- In the broad analysis, there were no distinctly resolved sub-clades within the *C. cajan* complex which consists of landraces, improved cultivars and naturally occurring perennial pigeonpea genotypes. This reflects low genetic polymorphism within domesticated pigeonpea.

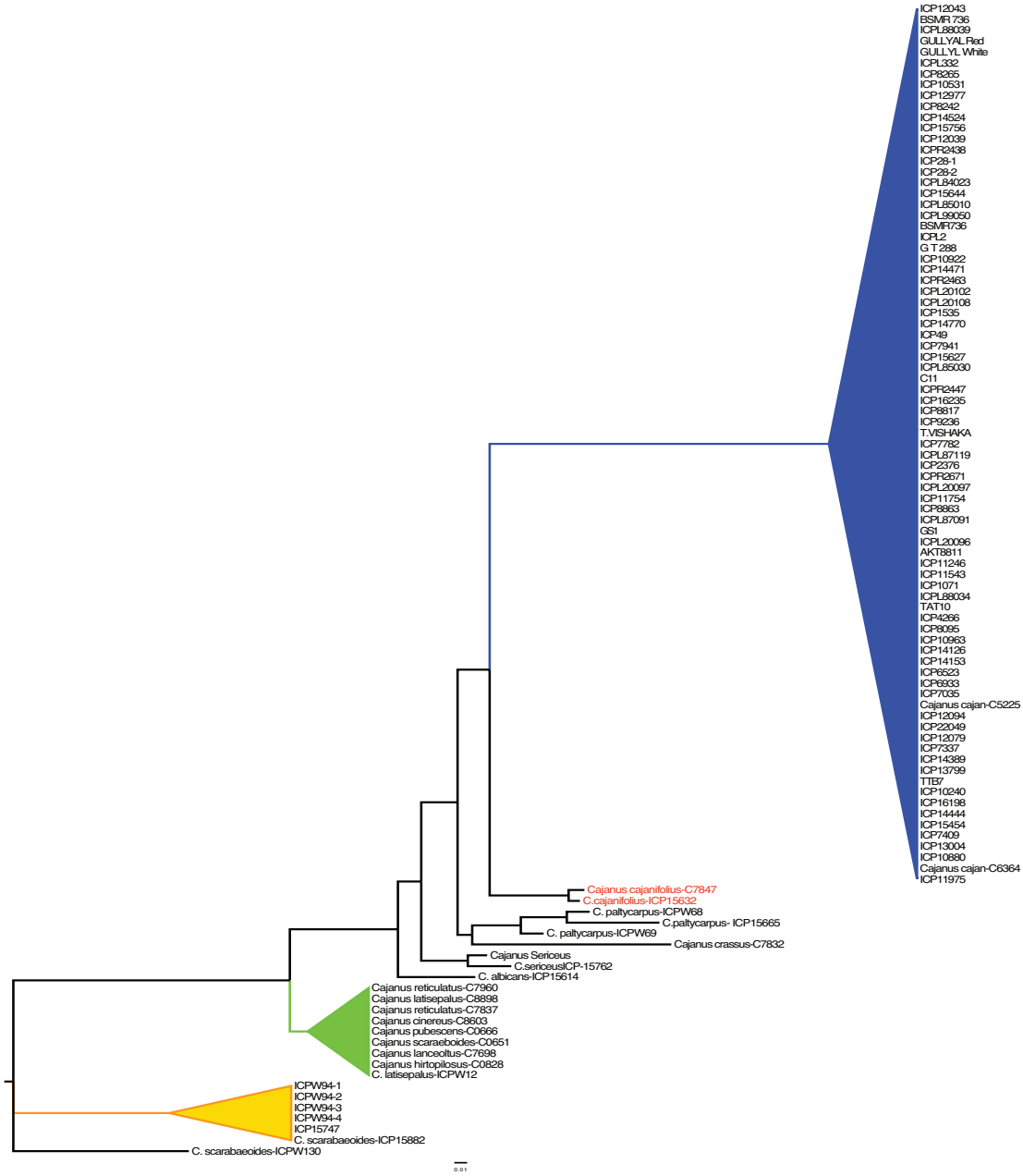


Fig. 5.1. Schematic representation of the Neighbor Joining tree of wild and domesticated genotypes based on SNP markers from the orthologous markers. (Key: Orange clade: Wild-*scarabaeoides*; Green clade: Wild Australian; Blue clade: domesticated group; red font: putative progenitor).

Phylogenetic signal and genetic signatures- insights into the domestication of pigeonpea

The SNP markers provided important information on the genetic structure of wild and domesticated accessions and gave insights into the domestication history of pigeonpea.

- STRUCTURE analysis (Pritchard et al., 2000; Falush et al., 2003) resolved major groups including a clear separation between wild and domesticated genotypes at k=2. The wild group further divided into two groups, corresponding to *Cajanus scarabaeoides* and the remaining wild *Cajanus* species of Australian and India origin.
- Genetic admixture was evident between wild and domesticated groups as revealed by STRUCTURE and PCoA (Fig. 3.4 and 3.8). Similarly, admixed genotypes were detected in the Neighbor Joining tree (Fig. 3.1). Seven (three wild and four domesticated) accessions were predicted to be either feral hybrids or breeding admixtures of wild and domesticated groups. More of interest is the observation of presumed natural spontaneous hybridization in the case of *C. lineatus* (ICPW46) and the possible breeding admixture of the two *C. cajanifolius* accessions (ICPW29 and ICP15629).
- The phylogenetic and genetic signatures (Figiures 3.4 and 5.1) of *C. cajan* indicate *Cajanus cajanifolius* as the presumed progenitor of pigeonpea. We speculate that for pigeonpea there was a single major domestication event in India.
- Within the domesticated group, similar genetic subdivisions were obtained using both weighted Neighbor Joining and allele frequency analyses, and these subdivisions mirrored the geographical distribution of individual accessions. One subdivision contained accessions mainly from tropical regions in Africa, the Caribbean, and Latin America, and the second subdivision contained genotypes of Indian origin (Fig. 3.5).

Impact of domestication in genetic diversity of pigeonpea

SNP data was also used to estimate genetic diversity and molecular differentiation between wild and domesticated pigeonpea. Based on these analyses, the low level of molecular diversity in domesticated pigeonpea as compared to the wild groups was confirmed, which has also been reported in previous studies (Yang et al., 2006; Odeny et al., 2007).

- This study found an abundant allelic variation and genetic diversity in the wild groups, with the exception of wild species from Australia, as compared to the domesticated pigeonpea.

- This study found a reduction of about 75% in genetic polymorphism in domesticated-India as compared to the wild-India samples establishes that there was a severe "domestication bottleneck" during pigeonpea domestication.

Molecular marker development for pigeonpea improvement and breeding

In the last data chapter of this thesis, Single Nucleotide Polymorphisms (SNPs) were mined and analyzed in the vicinity of pigeonpea disease resistance gene loci.

- SNPs were discovered in a comparison of BAC-end sequences (BES) of *C. cajan* and amplicons of the wild species, *C. scarabaeoides*. The identified Single nucleotide polymorphism (SNP) markers were associated with disease resistance (NBS-LRR) genomic regions.
- A total of ~3000 SNPs were derived from 304 BES of which 50% were annotated as genes. Of particular interest was that ~25% of the BES with SNPs were resistance genes. These SNP markers may potentially be used in marker assisted breeding, especially for markers mapped in vicinity of agronomically important candidate resistance genes.
- Many cultivars of pigeonpea have little genetic polymorphism as evidenced from previous chapters and lack sources of resistance genes to pests and diseases. Wild relatives in contrast, possess high genetic diversity and possess potential sources of disease resistance genes, and thus SNPs linked to disease resistance genes may have value in breeding applications.
- This study also attempted to infer the trend of evolutionary selection pressure on coding sequences of the resistance genes and other anonymous genes derived from the BES.

Future directions

This thesis attempts to provide a framework for future research work. The major recommendations for future work are:

1. Wider taxon sampling

To gain detailed phylogenetic information and phylogeography of the subtribe Cajaninae, there is a need to include wider taxon sampling especially from the two large genera of *Rhynchosia* and *Eriosema* (the two genera represent about 75% of the Cajaninae

species). The coverage of *Rhynchosia* species particularly from Africa was scarce in this study and adding more of these taxa will refine the phylogenetic framework presented in this study. If possible, inclusion of rare and scarce, hard to find species, such as *Cajanus kerstingii, Carrissoa sp., Chyrsoscias sp.* and *Paracalyx* sp. is also recommended for future work. This study demonstrated the utility of SNPs derived from orthologous genes for phylogenetic analysis and further exploration and test of the low copy orthologous nuclear genes for phylogenetic analysis is recommended.

2. Phylogeography and Ecological studies

This study generates exciting information on population genetic structure and patterns of genetic diversity of domesticated pigeonpea and wild relatives collected from diverse geographical areas. However, the lack of climatic and ecological data limits further investigation to link genetic diversity with ecology and geographical variables. We recommend future work on ecological genetics that integrates geographical variables with genetic data to further explore the underlying phylogeography of the taxa.

3. Mapping of the SNP markers

The mapping of the identified SNPs associated with resistance genes is recomended. The SNP markers can also be potentially useful for integrating the genetic map of pigeonpea. Further analysis on evolutionary selection of resistance genes is also recommended. The identified SNPs can also be useful for future genotyping and genetic diversity studies.

References

Acosta-Gallegos JA, Quintero C, Vargas J, Toro O, Tohme J, Cardona, C. 1998. A new variant of arcelin in wild common bean, *Phaseolus vulgaris* L. from southern Mexico. *Genetic Resources and Crop Evolution*. **45**: 235–242.

Aguilar-Melendez A, Morrell PL, Roose MK, Kim, SC. 2009. Genetic diversity and structure in semiwild and domesticated chiles (*Capiscum annuum*; Solanaceae) from Mexico. *American Journal of Botany* **96** : 1190-1202.

Ainouche A, Bayer RJ. 1999. Phylogenetic relationships in *Lupinus* (Fabaceae, Papilionoideae) based on internal transcribed spacer sequences (*ITS*) of nuclear ribosomal DNA. *American Journal of Botany* **86**: 590–607.

Aldrich J, Cherney BW, Christopherson L. 1988. The role of insertions/deletions in the evolution of the intergenic region between *psbA* and *trnH* in the chloroplast genome. *Current Genetics* **14**: 137–146.

Ali SI. 1968. *Paracalyx* Ali, a new papilionaceous genus. *Univ. stud. Karachi* **5**: 93-97.

Allaby RG, Fuller DQ, Brown TA. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences of the USA* **105**: 13982–13986.

Allan GJ, Porter JM. 2000. Tribal delimitation and phylogenetic relationships of Loteae and Coronilleae (Faboideae: Fabaceae) with special reference to *Lotus*: evidence from nuclear ribosomal *ITS* sequences. *American Journal of Botany* **87**: 1871-1881.

Álvarez I, Wendel JF. 2003. Ribosomal *ITS* sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**: 417-434.

Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB. 2008. Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiology* **146**: 5-21.

Anderson JA, Churchill GA., Autrique JE, Tanksley SD, Sorrells M.E. 1993. Optimizing parental selection for genetic linkage maps. *Genome Research* **36**: 181–186.

Appleby N, Edwards D, Batley J. 2009. New technologies for ultra-high throughput genotyping in plants. *Methods in Molecular Biology* **513**:19–39.

Ariyanayagam RP, Rao AN, Zaveri PP. 1993. Gene-cytoplasmic male-sterility in pigeonpea. *Int. Pigeonpea News letter* **18**: 7—11.

Ariyanayagam RP, Rao AN, Zaveri PP. 1995. Cytoplasmic-genic male- sterility in interspecific matings of *Cajanus*. *Crop Science* **35**: 981–985.

Aruna R, Manohar RD, Reddy LJ, Upadhyaya HD, Sharma HC. 2005. Inheritance of Trichomes and Resistance to Pod borer (*Helicoverpa armigera*) and their Association in Interspecific Crosses between Cultivated Pigeonpea (*Cajanus cajan*) and its wild relative C. *scarabaeoides*. *Euphytica* **145**: 247-257.

Azuma H, García-Franco JG, Rico-Gray V, Thien LB. 2001. Molecular phylogeny of the Magnoliaceae: the biogeography of tropical and temperate disjunctions. *American Journal of Botany* **88**: 2275–2285.

Bailey CD, Carr TG, Harris SA, Hughes CE. 2003. Characterization of angiosperm *nrDNA* polymorphism, paralogy and pseudogenes. *Molecular Phylogeny and Evolution* **29**: 435–455.

Bakker FT, Culham A, Daugherty LC, Gibby M. 1999. *A trnL-F* based phylogeny for species of *Pelargonium* (Geraniaceaea) with small chromosomes. *Plant Systematics and Evolution* **216**: 309–324.

Bakker FT, Culham A, Gomez-Martinez R, Carvalho J, Compton J, Dawtrey R, Gibby M. 2000. Patterns of nucleotide substitution in angiosperm *cpDNA trnL* (UAA)-*trnF* (GAA) regions. *Molecular Biology and Evolution* **17**: 1146–1155.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ. 1995. The *ITS* region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82**: 247–277.

Barker NP, Von Senger I, Howis S, Zachariades C, Ripley BS. 2005. Plant phylogeography based on nrDNA ITS sequence data: two examples from the Asteraceae. In: Bakker FT, Chatrou LW, Gravendeel B, Pelser PB. eds. *Plant species level systematics: new perspectives on pattern and process*, ARG Gantner Verlag, Ruggell: Chapter 11: 217-244.

Barker NP, Howis S, Nordenstam B, Kallersjo M, Eldenas P, Griffioen C, Linder HP. 2009. Nuclear and chloroplast DNA-based phylogenies of *Chrysanthemoides* Tourn. ex Medik (Calenduleae; Asteraceae) reveal extensive incongruence and generic paraphyly, but support the recognition of infraspecific taxa in *C. monilifera*. *South African Journal of Botany* **75**: 560-572.

Baudet JC. 1978. Prodrome d'une classification generique des Papilionaceae-Phaseoleae. *Bulletin du Jardin Botanique National du Belgique* **48**: 183–220.

Bell EA. 1981. Non-protein amino acids in the Leguminosae. In: Polhill RM, Raven PH. eds. *Advances in legume systematics 2*. Royal Botanic Gardens, Kew, UK: 489–499.

Bhatia GK, Gupta SC, Green JM, Sharma D. 1981. Estimates of natural cross-pollination in *Cajanus cajan* (L.) Millsp, several experimental approaches. In: *Proceedings of International Workshop on Pigeonpeas*. International Crops Research Institute for the Semi-Arid Tropics: 129-136.

Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poornima1 KN, Kumar1 N, Farmer AD, Srivani G, Upadhyaya1 HD, Gothalwal R, Ramesh S, Singh D, Saxena K, Kishor PBK, Singh NK, Town CD, May GD, Cook DR, Varshney RK. **2011.** Analysis of BAC-end

sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). *BMC plant biology* **11**: 56.

Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W. 2003. Noncoding plastid *trnT-trnF* sequences reveal a well-resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* **16**: 558–576.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of genetic linkage map in man using restriction length polymorphisms. *The American Journal of Human Genetics* 32: 314–331.

Brouat C, Gielly L, McKey D. 2001. Phylogenetic relationships in the genus *Leonardoxa* (Leguminosae: Caesalpinioideae) inferred from chloroplast *trnL intron* and *trnL-trnF* intergenic spacer sequences. *American Journal of Botany* **88**:143–149.

Bruneau A, Doyle JL, Doyle JJ. 1995. Phylogenetic evidence in Phaseoleae: evidence from chloroplast restriction site characters. In: Crisp MD, Doyle JJ. eds. *Advances in legume systematics 7, Phylogeny*. Royal Botanic Gardens, Kew, UK: 309-330.

Bruneau A, Forest F, Herendeen PS, Klitgaard BB, Lewis GP. 2001. Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trnL intron* sequences. *Systematic Botany* **26**: 487–514.

Bruneau A, Mercure M, Lewis GP, Herendeen PS. 2008. Phylogenetic patterns and diversification in the caesalpinioid legumes. *Canadian Journal of Botany* **86**: 697–718.

Bruno WJ, Socci ND, Halpern AL. 2000. Weighted Neighbor Joining: A Likelihood-Based Approach to Distance-Based Phylogeny Reconstruction. *Molecular Biology and Evolution* **17**: 189-197.

Buckler ES, Thornsberry JM, Kresovich S, Buckler ES, Thornsberry JM, Kresovich S. 2001. Molecular diversity, structure and domestication of grasses. *Genetic Research* **77**: 213–218.

Burger JC, Chapman MA, Burke JM. 2008. Molecular insights into the evolution of crop plants. *American Journal of Botany* **95**: 113–122.

Chandler GT, Bayer RJ, Crisp MD. 2001. A molecular phylogeny of the endemic Australian genus *Gastrolobium* (Fabaceae: *Mirbelieae*) and allied genera using chloroplast and nuclear markers. *American Journal of Botany* **88**: 1675– 1687.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q-Y, Plunkett GM, Soltis PS, Swensen S, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn Jr GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* **80**: 528–580.

Chen CA, Chang CC, Wei NV, Chen CH, Lein YT, Lin HE, Dai CF, Wallace CC. 2004. Secondary structure and phylogenetic utility of the ribosomal internal transcribed spacer 2 (*ITS2*) in scleractinian corals. *Zoological Studies* **43**: 759-771.

Chen HF, Morrell PL, Ashworth V, Cruz, MDL, Clegg MT. 2009. Tracing the geographic origins of major avocado cultivars. *Journal of Heredity* **100**: 56 – 65.

Chen Q, Han Z, Jiang H, Tian D, Yang S. 2010. Strong positive selection drives rapid diversification of R-genes in *Arabidopsis* relatives. Journal of Molecular *Evolution* **70**: 137–148.

Chieba T, Harada T, Goto S, Ishikawa R, Niizeki M. 1996. Transcription of *tRNA* genes from large-scale plastid DNA deletion clearly reveals the action of nuclear-encoded RNA polymerase in the plastid. *Journal of Plant Physiology* **148**: 652–656.

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

Coyne CJ, McClendon MT, Walling JG, Timmerman-Vaughan GM, Murray S, Meksem K, Lightfoot DA, Shultz JL, Keller KE, Martin RR, Inglis DA, Rajesh PN, McPhee KE, Weeden NF, Grusak MA, Li CM, Storlie EW. 2007. Construction and characterization of two bacterial artificial chromosome libraries of pea (*Pisum sativum* L.) for the isolation of economically important genes. *Genome* **50**: 871–875.

Crisp MD, Arroyo MTK, Cook LG, Gandolfo MA, Jordan GJ, McGlone MS, Weston PH, Westoby M, Wilf P, Linder HP. 2009. Phylogenetic biome conservatism on a global scale. *Nature* **458**: 754–756.

Crisp MD, Gilmore S, van Wyk B-E. 2000. Molecular phylogeny of the genistoid tribes of papilionoid legumes. In: Herendeen PS, Bruneau A. eds. *Advances in legume systematics 9*. Royal Botanic Gardens, Kew, UK: 249-276.

Cronk Q, Ojeda I, Pennington RT. 2006. Legume comparative genomics: progress in phylogenetics and phylogenomics. *Current Opinion in Plant Biology* **9**: 99–103.

Cruz F, Carles V, Matthew TW. 2008. The legacy of domestication: Accumulation of deleterious mutations in the dog genome. *Molecular Biology and Evolution* **25**: 2331–2336.

Cubas P, Pardo C, Tahiri H. 2002. Molecular approach to the phylogeny and systematics of *Cytisus* (Leguminosae) and related genera based on nucleotide sequence of *nrDNA* (*ITS* region) and *cpDNA* (*trnL-trnF* intergenic spacer). *Plant Systematics and Evolution* **233**: 223–242.

Dangl JL, Jones JDG. 2001. Plant pathogens and integrated defense responses to infection. *Nature* **411**: 826-833.

Davis CC, Fritsch PW, Li J, Donoghue MJ. 2002. Phylogeny and biogeography of *Cercis* (Fabaceae): Evidence from nuclear ribosomal *ITS* and chloroplast *ndhF* sequences. Systematic Botany **27**: 289-302.

De DN. 1974. Pigeonpea. In: Hutchinson JB. ed. *Evolutionary studies on world crops*. Cambridge University Press, Cambridge: 79-87.

Delgado-Salinas A, Bibler R, Lavin M. 2006. Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Systematic Botany* **31**: 779–791.

DeYoung BJ, Innes RW. 2006. Plant NBS-LRR proteins in pathogen sensing and host-defense. *Nature Immunology* **7**: 1243-1249.

Dickison WC. 1981. The evolutionary relationships of the Leguminosae. In: Polhill RM, Raven PH. eds. *Advances in legume systematics, part 1*. Royal Botanic Gardens, Kew: 35-54.

Doebley J, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* **127**: 1309–1321.

Downie SR, Palmer JD. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ. eds. *Molecular systematics of plants*, eds. New York: Chapman and Hall: 14-35.

Doyle JJ. 1995. DNA data and legume phylogeny: a progress report. In: Crisp MD, Doyle JJ. eds. *Advances in legume systematics 7, Phylogeny.* Royal Botanic Gardens, Kew, UK: 11-30.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11-15.

Doyle JJ, Doyle JL. 1993. Chloroplast DNA phylogeny of the Papilionoid legume tribe Phaseoleae. *Systematic Botany* **18**: 309–327.

Doyle JJ, Doyle JL, Ballenger JA, Dickson EE, Kajita T, Ohashi H. 1997. A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: Taxonomic correlations and insights into the evolution of nodulation. *American Journal of Botany* **84**: 541–554.

Doyle JJ, Chappill JA, Bailey CD, Kajita T. 2000. Towards a comprehensive phylogeny of legumes: evidence from *rbcL* sequences and non- molecular data. In: Herendeen P, Bruneau, A. eds. *Advances in Legume Systematics 9*. Royal Botanic Gardens, Kew, UK: 1-20.

Doyle JJ, Luckow MA. 2003. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiology* **131**: 900–910.

Du Puy DJ, Labat J-N, Rabevohitra R, Villiers J-F, Bosser J, Moat J. 2002. The Leguminosae of Madagascar. Royal Botanic Gardens, Kew, UK: 737.

Dutta S, **Kumawat G**, **Singh BP**, **Gupta DK**, **Singh S**, **Dogra V**, **Gaikwad K**, **Sharma TR**, **Raje RS**, **Bandhopadhya TK**, **Datta S**, **Singh MN**, **Bashasab F**, **Kulwal P**, **Wanjari KB**, **Varshney RK**, **Cook DR**, **Singh NK**. **2011**. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biology* **11**:17.

Ellison NW, Liston A, Steiner JJ, Williams WM, Taylor NL. 2006. Molecular phylogenetics of the clover genus *Trifolium* (Leguminosae). *Molecular Phylogenetics and Evolution* **39**: 688-705.

Ems SC, Morden CW, Dixon CK, Wolfe KH, dePamphilis CW, Palmer JD. 1995. Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Molecular Biology* **29**: 721–733.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* Online **1**: 47-50.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the National Academy of Sciences of the USA* **95**: 4441–4446.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Fan JB, Chee MS, Gunderson KL. 2006. Highly parallel genomic assays. *Nat Rev Genet* 7:632–644.

Farris JS, Kallersjo M, Kluge AG, Bult C. 1994. Testing significance of congruence. *Cladistics* **10**: 315-319.

Feliner GN, Rossello JA. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA *ITS* in species-level evolutionary studies in plants. *Molecular phylogenetics and Evolution* **44**: 911-919.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783-791.

Fitch WM. 1971. Toward defining the course of evolution: minimal change for a specific tree topology. *Systematic Zoology* **20**: 406–416.

Fortunato RH. 2000. Systematic relationships in *Rhynchosia* (Cajaninae Phaseoleae–Papilionoideae–Fabaceae) from the neotropics. In: Herendeen PS, Bruneau A. eds. *Advances in legume systematics 9.* The Royal Botanic Gardens, Kew, UK: 339-354.

Frary A, Doganlar S. 2003. Comparative genetics of crop plant domestication and evolution. *Turkish Journal of Agriculture and Forestry* **27**: 59–69.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**: 1631–1638.

Gaut BS. 1998. Molecular clocks and nucleotide substitution rates in higher plants. In: Hech MK, MacIntyre RJ, Clegg MT. eds. *Evolutionary biology 30*. Plenum Press, New York, USA: 93-120.

Gene Ontology Consortium. 2010. The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res* **38** (Database issue): D331-D335.

Germishuizen G. 2000. Fabaceae (Leguminosae). In: Leistner OA. ed. *Seed plants of Southern Africa: families and genera*. Strelitzia 10: 262-303.

Gielly L, Taberlet P. 1996. Chloroplast DNA sequencing to resolve plant phylogenies between closely related taxa. In: *Molecular Genetic Approaches in Conservation.* Oxford University Press, Oxford: 143-153.

Graham PH, Vance CP. 2003. Legumes: Importance and constraints to grater use. *Plant Physiology* **131**: 872-877.

Graham SW, Olmstead RG. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* **87**: 1712-1730.

Grear JW. 1970. A revision of the America species of *Eriosema* (Leguminosae-Lotoideae). *Mem. New York Bot. Gard*. **20**: 1-98.

Grear JW. 1978. A revision of the New World species of *Rhynchosia* (Leguminosae - Faboideae). *Mem. New York Bot. Gard*. 31: 76.

Gross BL, Olsen KM. 2010. Genetic perspectives on crop domestication. *Trends in Plant Science* **15**: 529-537.

Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H, Wang Y. 2010. A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Annals of Botany* **106**: 505–514.

Gupta PK, Rustgi S, Mir RR. 2008. Array-based high- throughput DNA markers for crop improvement. *Heredity* **101**: 5–18.

Hale ML, Borland AM, Gustafsson MHG, Wolff. K. 2004. Causes of size homoplasy among chloroplast microsatellites in closely related *Clusia* species. *Journal of Molecular Evolution* **58**: 182–190.

Halliburton R. 2004. Introduction to population genetics. Upper Saddle River, NJ: Pearson-Prentice-Hall.

Hamilton MB. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8**: 513–525.

Hamilton MB, Braverman JM, Soria-Hernanz DF. 2003. Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in New World species of the Lecythidaceae. *Molecular Biology and Evolution* **20**: 1710–1721.

Harlan JR. 1992. Crops and man. 2nd ed. Am. Soc. Agronomy, Madison, WI.

Harlan JR, de Wet JMJ. 1971. Toward a rational classification of cultivated plants. *Taxon* **20**: 509–517.

Haston EM, Lewis GP, Hawkins JA. 2005. A phylogenetic reappraisal of the *Peltophorum* group (Caesalpinieae: Leguminosae) based on the chloroplast *trnL-F*, *rbcL*, and *rps16* sequence data. *American Journal of Botany* **92**: 1359-1371.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I. 2007. Grinding up Wheat: A massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution* **24**:1506 – 1517.

Hauser LA, Crovello TJ. 1982. Numerical analysis of genetic relationships in *Thelypodieae* (Brassicaceae). *Systematic Botany* **7**: 249–268.

Havey MJ. 2004. The use of cytoplasmic male sterility for hybrid seed production. In: Daniell H, Chase C. eds. *Molecular biology and biotechnology of plant organelles*. Springer, Dordrecht, The Netherlands: 617-628.

Herendeen PS. 1992. The fossil history of Leguminosae from the Eocene of southeastern North America. In: Herendeen, PS, Dilcher DL. eds. *Advances in Legume Systematics 4, The Fossil Record.* Royal Botanic Gardens, Kew, UK: 85-160.

Herendeen PS. 2001. The fossil record of the Leguminosae: recent advances. *In Legumes Down Under: the Fourth International Legume conference, Abstracts.* Australian National University, Canberra, Australia:34-35

Herendeen PS, Wing S. 2001. Papilionoid legume fruits and leaves from the Paleocene of northwestern Wyoming. *Botany 2001 Abstracts*, published by Botanical Society of America (http://www.botany2001.org/).

Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, Savo-Lainen V, Chase MW, Powell MP, Alice LA, Evans R, Sau-Quet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW. 2003. Angiosperm phylogeny based on *matk* sequence information. *American Journal of Botany* **90**: 1758–1776.

Hu J-M, Lavin M, Wojciechowski MF, Sanderson MJ. 2000. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trnK/matK* sequences and its implications for evolutionary patterns in Papilionoideae. *American Journal of Botany* **87**: 418–430.

Huelsenbeck JP, Ronquist F. 2001. MR BAYES: Bayesian inference of phylogenetic trees. *Bioinfomatics* **17**: 754-755.

Hughes CE, Eastwood RJ, Bailey, C.D. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society, London B* **361**: 211–255.

Hulbert SH, Webb CA, Smith SM, Sun Q. 2001. Resistance gene complexes: Evolution and utilization. *Annual Review of Phytopathology* **39**: 285–312.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Research* **37**: D211–D215.

Hurr KA, Lockhart PJ, Heenan PB, Penny D. 1999. Evidence for the recent dispersal of *Sophora* (Leguminosae) around the southern oceans: molecular data. *Journal of Biogeography* **26**: 565–577.

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. *Proceedings of the National Academy of Sciences of the USA* **103**: 16666–16671.

Ireland H, Pennington RT, Preston J. 2000. Molecular systematics of the Swartzieae. In: Herendeen PS, Bruneau A. eds. *Advances in legume systematics 9*. Royal Botanic Gardens, Kew, UK: 217-231.

Jaccoud D, Peng K, Feinstein D, Kilian A. 2001. Diversity arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Research* **29**: e25.

Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. 2008. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP intron* losses among legumes (Leguminosae). *Molecular Phylogenetics and Evolution* **48**: 1204–1217.

Jha SS, Ohri D. 1996. Phylogenetic relationships of *Cajanus cajan* (L.) Millsp. (Pigeonpea) and its wild relatives based on seed protein profiles. *Genetic Resources and Crop Evolution* **43**: 275–281.

Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ. 2001. *rbcL* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and Allies. *Systematic Botany* **26**: 515–536.

Kass E, Wink M. 1995. Molecular phylogeny of the Papilionoideae (family Leguminosae): *rbcL* gene sequences versus chemical taxonomy. *Botanica Acta* **108**: 149–162.

Kass E, Wink M. 1996. Molecular evolution of the Leguminosae: phylogeny of the three subfamilies based on rbcL-sequences. *Biochemical Systematics and Ecology* **24**: 365–378.

Kass E, Wink M. 1997a. Phylogenetic relationships in the Papilionoideae (Family Leguminosae) based on nucleotide sequences of *cpDNA (rbcL)* and *ncDNA* (*ITS1* and *2*). *Molecular Phylogenetics and Evolution* **8**: 65–88.

Kass E, Wink M. 1997b. Molecular phylogeny and phylogeography of *Lupinus* (Leguminosae) inferred from nucleotide sequences of the rbcL gene and *ITS 1 + 2* regions of *rDNA*. *Plant Systematics and Evolution* **208**: 139–167.

Katoh K, Kuma K, Toh H, Miyata T. 2005. Improvement in accuracy of multiple sequence alignment. MAFFT version 5. *Nucleic Acids Research* **33**: 511–518.

Kaul M. 1988. Male sterility in higher plants. In: Frankel R, Grossman M, Maliga P. eds. *Monographs on Theoretical and Applied Genetics 10*. Springer-Verlag, Berlin: 775-795.

Kawakita A, Sota T, Ascher JS, Ito M, Tanaka H, Kato M. 2003. Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (*Bombus*). *Molecular Biology and Evolution* **20**: 87–92.

Kelchner SA. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* **87**: 482–498.

Kenicer GJ, Kajita T, Pennington RT, Murata J. 2005. Systematics and biogeography of *Lathyrus* (Leguminosae) based on *Internal Transcribed Spacer* and *cpDNA* sequence data. *American Journal of Botany* **92**: 1199–1209.

Kilian B, Ozkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F. 2006. Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. *Molecular Genetics and Genomics* **276**: 230–241.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the USA* **102**: 8369–8374.

Krishna TG, Reddy LJ. 1982. Species affinities between *Cajanus cajan* and some *Atylosia* species based on esterase isoenzymes. *Euphytica* **31**: 709–713.

Kryazhimskiy S, Plotkin JB. 2008. The population genetics of dN/dS. *PLoS Genetics* **4**: e1000304.

Kuntze CEO. 1891. *Revisio Generum Plantarum Part 1*: 162–163.

Lackey JA. 1977. A revised classification of the tribe Phaseoleae (Leguminosae, Papilionoideae), and its relation to canavanine distribution. *Botanical Journal of the Linnean Society* **74**: 163-178.

Lackey JA. 1978. Leaflet anatomy of Phaseoleae (Leguminosae: Papilionoideae) and its relation to taxonomy. *Botanical Gazette* **139**: 436–446.

Lackey JA. 1981. Phaseoleae. In: Polhill, RM, Raven PH. eds. *Advances in Legume Systematics 1*, Royal Botanic Gardens, Kew, UK: 301-327.

Ladizinsky G, Hamel A. 1980. Seed protein profiles of pigeonpea (*Cajanus cajan*) and some Atylosia species. *Euphytica* **29**: 313–317.

Lakshmi M, Senthilkumar P, Parani M, Jithesh MN, Parida A. 2000. PCR-RFLP analysis of chloroplast gene regions in *Cajanus* (Leguminosae) and allied genera. *Euphytica* **116**: 243–250.

Lavin M, Pennington RT, Klitgaard BB, Sprent JI, de Lima HC, Gasson PE. 2001. The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. *American Journal of Botany* **88**: 503–533.

Lavin M, Wojciechowski MF, Gasson P, Hughes CE, Wheeler E. 2003. Phylogeny of robinioid legumes (Fabaceae) revisited: *Coursetia* and *Gliricidia* recircumscribed, and a biogeographical appraisal of the Caribbean endemics. *Systematic Botany* **28**: 387–409.

Lavin M, Schrire BD, Lewis G, Pennington RT, Delgado-Salinas A, Thulin, M, Hughes CE, Beyra-Matos A, Wojciechowski MF. 2004. Metacommunity processes rather than continental tectonic history better explain geographically structured phylogenies in legumes. *Philosophical transactions of the Royal Society of London*. Series **B 359**: 1509–22.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* **54**: 530-549.

Le Thierry D' Ennequin M, Toupance B, Robert T, Godelle B, Gouyon PH. 1999. Plant domestication: A model for studying the selection of linkage. *Journal of Evolutionary Biology* **12**: 1138–1147.

Lewis G, Schrire B, MacKinder B, Lock M. 2005. Legumes of the world. Royal Botanic Gardens, Kew, UK.

Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ. 2010. Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytologist* **188**: 242-253.

Liston A, Wheeler JA. 1994. The phylogenetic position of the genus *Astragalus* (Fabaceae): evidence from the chloroplast genes *rpoC1* and *rpoC2*. *Biochemical Systematics and Ecology* **22**: 377–388.

Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* **173**: 321-330.

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences of the USA* **103**: 9578–9583.

Luckow M, White PJ, Bruneau A. 2000. Relationships among the basal genera of Mimosoid legumes. In: Herendeen PS, Bruneau A. eds. *Advances in legume systematics 9*. Royal Botanic Gardens, Kew, UK: 165-180.

Luckow M, Miller JT, Murphy DJ, Livshultz T. 2003. A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. In: Klitgaard BB, Bruneau A. eds. *Advances in Legumes Systematics 10, higher level systematics*, eds. Royal Botanic Gardens, Kew, UK: 197-220.

Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z, Goodwin W, Bul-beck D, Bandelt HJ, Oppenheimer S., Torroni A, Richards M. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308:** 1034–1036.

Mackill DJ, Zhang E, RedonRa ED, Colowit PM. 1996. Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* **39**: 969–977.

Mackinder B, Pasquet R, Polhill RM, Verdcourt B. 2001. In: Pope GV, Polhill RM. eds. *Flora Zambesiaca 3(5), Phaseoleae*. Royal Botanic Gardens, Kew, UK: 1-261.

Maddison WP, Maddison DR. 2000. Analysis of phylogeny and character evolution. *MacClade 4*: Sinauer Associates, Sunderland.

Magallon S, Crane PR, Herendeen PS. 1999. Phylogenetic pattern, diversity, and diversification of eudicots. *Annals of the Missouri Botanical Garden* **86**: 297–372.

Mallikarjuna N, Moss JP. 1995. Production of hybrids between *Cajanus platycarpus* and *Cajanus cajan*. *Euphytica* **83**: 43–46.

Mallikarjuna N, Deepak J, Reddy MV, Usharani DT. 2005. Introgression of phytophthora blight disease resistance from *Cajanus platycarpus* into short duration pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Indian Journal of Genetics and Plant Breeding* **65**: 261–263.

Mallikarjuna N, Saxena KB. 2005. A new cytoplasmic nuclear male-sterility system derived from cultivated pigeonpea cytoplasm. *Euphytica* **142**: 143-148.

Mallikarjuna N, Jadhav D, Reddy P. 2006. Introgression of *Cajanus platycarpus* genome into cultivated pigeonpea, *C. cajan*. *Euphytica* **149**: 161–167.

Martin GB, Bogdanove AJ, Sessa G. 2003. Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology* **54**: 23-61.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez JG, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the USA* **99**: 6080–6084.

Mayer MS, Bagga SK. 2002. The phylogeny of lens (Leguminosae): new insight from *ITS* sequence analysis. *Plant Systematics and Evolution* **232**: 145–154.

McKey D. 1994. Legumes and nitrogen: the evolutionary ecology of a nitrogen-demanding lifestyle. In: Sprent JI, McKey D. eds. *Advances in Legume Systematics 5, the nitrogen factor*. Royal Botanic Gardens, Kew, UK: 211-228.

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *The Plant Journal* **20**: 317–332.

Miller JT, Bayer RJ. 2001. Molecular phylogenetics of *Acacia* (Fabaceae: Mimosoideae) based on the chloroplast *matK* coding sequence and flanking *trnK intron* spacer regions. *American Journal of Botany* **88**: 697–705.

Miller JT, Grimes JW, Murphy DJ, Bayer RJ, Ladiges PY. 2003 A phylogenetic analysis of the Acacieae and Ingeae (Mimosoideae: Fabaceae) based on *trnK, matK, psbA-trnH*, and *trnL/trnF* sequence data. *Systematic Botany* **28**: 558-566.

Minja EM, Shanower TG, Silim SN, Karuru O. 2000. Efficacy of different insecticides for pigeonpea pest management in Kenya. *ICPN* 7: 30-43.

Monosi B, Wisser RJ, Pennill L, Hulbert SH. 2004. Full-genome analysis of resistance gene homologues in rice. *Theoretica and Applied Genetics* **109**: 1434–1447.

Morley RJ. 2000. *Origin and Evolution of Tropical Rain Forests*. John Wiley & Sons: 378.

Morley RJ, Dick CW. 2003. Missing fossils, molecular clocks, and the origin of the Melastomataceae. *American Journal of Botany* **90**: 1638–1645.

Morrell PL, Clegg MT. 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences of the USA* **104**: 3289 – 3294.

Mort ME, Archibald JK, Randle CP, Levsen ND, O'Leary TR, Topalov K, Wiegand CM, Crawford DJ. 2007. Inferring phylogeny at low taxonomic levels: utility of rapidly evolving *cpDNA* and nuclear *ITS* loci. *American Journal of Botany* **94**: 173–183.

Moteetee A, Van Wyk B-E. 2006. A revision of the genus *Bolusafra* (tribe Phaseoleae, Fabaceae). *South African Journal of Botany* **72**: 604–608.

Motta-Aldana JR, Serrano ML, Hernández-Torres J, Castillo-Villamizar G, Debouck DG, Chacón-S MI. 2010. Multiple origins of lima bean landraces in the Americas: Evidence from chloroplast and nuclerar DNA polymorphisms. *Crop Science* **50**: 1773-1787.

Muller K, Borsch T, Hilu KW. 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT–F*, and *rbcL* in basal angiosperms. *Molecular Phylogenetics and Evolution* **41**: 99–117.

Mun JH, Kim DJ, Choi HK, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez AM, Kiss GB, Roe B, Young ND, Cook DR. 2006. Distribution of microsatellites in the genome of *Medicago truncatula*: A resource of genetic markers that integrate genetic and physical maps. *Genetics* **172**: 2541-2555.

Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, Yu, HJ, Kim DS, Kim N, Lim KB, Lee SI, Hahn JH, Lim YP, Bancroft I, Park BS. 2009. Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biology* **10**: R111.

Nadimpalli RG, Jarret RL, Phatak SC, Kochert G. 1993. Phylogenetic relationships of the pigeonpea (*Cajanus cajan*) based on nuclear restriction fragment length polymorphism. *Genome* **36**: 216-223.

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y. 2002. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Research* **9**: 163–171.

Nene YL, Sheila VK. 1990. Pigeonpea: geography and importance. In: Nene YL, Hall SH, Sheila VK. eds. *The pigeonpea*. CAB International, Wellingford, UK: 1-14

Neuhaus H, Link G. 1987. The chloroplast tRNALys (UUU) gene from mustard (*Sinapsis alba*) contains a class II intron poten- tially coding for a maturase-related polypeptide. *Current Genetics* **11**: 251–257.

New Scientist. 2009. Magic pea hybrid could help feed the world. Issue No **2699**: 6-7.

Nylander JAA. 2004. MrModeltest 2.0. Program distributed by the author. Evolutionary Biology Centre, Uppsala University. http://www.ebc.uu.se/systzoo/staff/nylander.html.

Odeny DA, Jayashree B, Ferguson M, Hoisington D, Cry LJ, Gebhardt, C. 2007. Development, characterization and utilization of microsatellite markers in pigeonpea. *Plant Breeding* **126**: 130–136.

Odeny DA, Jayashree B, Gebhardt C, Crouch J. 2009. New microsatellite markers for pigeonpea (*Cajanus cajan* (L.) millsp.). *BMC Research Notes* 2:35.

Ohri D, Jha SS, Kumar S. 1994. Variability in nuclear DNA content within pigeonpea. *Plant Systematics and Evolution* **189**: 211–216.

Ohri D, Singh SP. 2002. Karyotypic and genome size variation in *Cajanus cajan* (L.) Millsp. (pigeonpea) and some wild relatives. *Genetic Resources and Crop Evolution* **49**: 1–10.

Olmstead RG, Palmer JD. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* **81**: 1205-1224.

Olsen KM, Gross BL. 2008. Detecting multiple origins of domesticated crops. *Proceedings of the National Academy of Sciences of the USA* **105**: 13701–13702.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the USA* **97**: 6960–6966.

Palomino C, Satovic Z, Cubero JI, Torres AM. 2006. Identification and characterization of NBS-LRR class resistance gene analogs in faba bean (*Vicia faba* L.) and chickpea (*Cicer arietinum* L.). *Genome* **49**: 1227–1237.

Panguluri SK, Janaiah K, Govil JN, Kumar PA, Sharma PC. 2006. AFLP fingerprinting in pigeonpea (*Cajanus cajan* (L.) Millsp.) and its wild relatives. *Genetic Resources and Crop Evolution* **53**: 523—531.

Panigrahi J, Kumar DR, Mishra M, Mishra RP, Jena P. 2007. Genomic relationships among 11 species in the genus *Cajanus* as revealed by seed protein (albumin and globulin) polymorphisms. *Plant Biotechnology Reports* **1**: 109–116.

Peakall R, Smouse PE. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288–295.

Pennington RT, Lavin M, Ireland H, Klitgaard B, Preston J, Hu J-M. 2001. Phylogenetic relationships of basal papilionoid legumes based upon sequences of the chloroplast *trnL* intron. *Systematic Botany* **26**: 537–556.

Perrier X, Jacquemoud-Collet JP. 2006. DARwin s. http://darwin.cirad.fr/darwin.

Perry AS, Wolfe KH. 2002. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution* **55**: 501–508.

Persson C. 2001. Phylogenetic relationships in Polygalaceae based on plastid DNA sequences from the *trnL-F* region. *Taxon* **50**: 763–779.

Petkov PM, Ding Y, Cassell MA, Zhang W, Wagner G, Sargent EE, Asquith S, Crew V, Johnson KA, Robinson P, Scott VE, Wiles MV. 2004. An efficient SNP system for mouse genome scanning and elucidating strain relationships. Genome Research **14**: 1806–1811.

Polhill RM. 1981. Papilionoideae. In: Polhill RM, Raven PH. eds. *Advances in legume systematics 1*. Royal Botanic Gardens, Kew, U.K: 191-208.

Polhill RM. 1994. Classification of the Leguminosae. In: Bisby FA, Buckingham J, Harborne JB. eds. *Pages in Phytochemical Dictionary of the Leguminosae*. Chapman and Hall, New York: xxxv–xlviii.

Pritchard JK, Stephens P, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pundir RPS, Singh RB. 1985a. Biosystematic relationships among *Cajanus*, *Atylosia* and *Rhynchosia* species and evolution of pigeonpea. *Theoretical and Applied Genetics* **69**: 531–534.

Pundir RPS, Singh RB. 1985b. Crossability relationships among *Cajanus, Atylosia* and *Rhynchosia* species and detection of crossing barriers. *Euphytica* **34**: 303–308.

Purseglove JW. 1968. Tropical Crops: Dicotyledons. Longman, UK.

Quandt D, Muller K, Stech M, Hilu KW, Frey W, Frahm J-P, Borsch T. 2004. Molecular evolution of the chloroplast *trnL-F* region in land plants. In: Goffinet B, Hollowell V, Magill R. eds. *Molecular Systematics of Bryophytes*, *Monographs in Systematic Botany 98*. Missouri Botanical Garden Press, St. Louis, Missouri, USA: 13-37.

Rafalski A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* **5**: 94–100.

Ramanadam P. 1990. Pigeonpea: genetic resources. In: Nene YL, Hall SD, Shelia VK. eds. *The pigeonpea*. CAB International, Wallingford, Oxon, UK: 349-374.

Ratnaparkhe MB, Gupta VS, Venmurthy MR, Ranjekar PK. 1995. Genetic fingerprinting of pigeonpea (*Cajanus cajan* (L.) Millsp.) and its wild relatives using random amplified polymorphic DNA markers. *Theoretical and Applied Genetics* **91**: 893-898.

Raven PH, Polhill RM. 1981 Biogeography of the Leguminosae. In: Polhill RM, Raven PH. eds. *Advances in legume systematics 1*. Royal Botanic Gardens, Kew, UK: 27-34

Ravi V, Khurana JP, Tyagi AK, Khurana P. 2007. Rosales sister to Fabales: towards resolving the rosid puzzle. *Molecular Phylogenetics and Evolution* **44**: 488-493.

Redd AJ, Roberts-Thomson J, Karafet T, Bamshad M, Jorde LB, Naidu JM, Walsh B, Hammer MF. 2002. Gene Flow from the Indian Subcontinent to Australia: Evidence from the Y Chromosome. *Current Biology* **12**: 673–677.

Reddy BVS, Green JM, Bisen SS. 1978. Genetic male-sterility in pigeonpea. *Crop Science* **18**: 362–364.

Reddy MV, Raju TN, Sheila V.K. 1996. Phytophthora blight disease in wild pigeonpea. *International Chickpea and Pigeonpea Newsletter* **3**: 52–53.

Reddy LJ, Upadhyaya HD, Gowda CLL, Singh S. 2005. Development of core collection in pigeonpea [*Cajanus cajan* (L.) Millsp.] Using geographic and qualitative morphological descriptors. *Genetic Resources and Crop Evolution* **52**:1049–1056.

Reynolds ST, Pedley L. 1981. A revision of *Atylosia* (Leguminosae) in Australia. *Austrobaileya* **1**: 420–428.

Richardson JE, Fay MF, Cronk QCB, Bowman D, Chase MW. 2000. A phylogenetic analysis of Rhamnaceae using *rbcL* and *trnL-F* plastid DNA sequences. *American Journal of Botany* **87**: 1309-1324.

Ridley M. 2004. Evolution. Blackwell publishing. 3rd edition.

Riley-Hulting E, Delgado-Salinas A, Lavin M. 2004. Phylogenetic systematics of *Strophostyles* (Fabaceae): a North American temperate genus within a neotropical diversification. *Systematic Botany* **29**: 627–653.

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF. 2006. Recent history of artificial outcrossing facilitates whole genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences of the USA* **103**:18656-18661.

Rundel PW. 1989. Ecological success in relation to plant form and function in the woody legumes. In: Stirton CH, Zarucchi JL. eds. *Advances in legume biology*. Monographs in Systematic Botany from the Missouri Botanical Gardens 29: 377-398.

Saitou N, Nei M . 1987. The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4:** 406–425.

Sang T, Crawford DJ, Stuessy TF. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* **84**: 1120– 1136.

Satyanararayana P. 1993. A taxonomic revision of the tribe Cajaneae (Fabaceae) in India. *PhD thesis*, Calcutta University, India.

Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank M, Cameron KM, Johnson SA, Lledó MD, Pintaud J-C, Powell M, Sheahan MC, Soltis DE, Soltis PS, Weston P, Whitten WM, Wurdack KJ, Chase MW. 2000. Phylogeny of eudicots: a nearly complete familial analysis based on *rbcL* gene sequences. *Kew Bulletin* **55**: 257-309.

Saxena KB, Singh L, Gupta MD. 1990. Variation for natural outcrossing in pigeonpea. *Euphytica* **46**:143–148.

Saxena KB, Singh L, Ariyanayagam RP. 1992a. Role of partial cleistogamy in maintaining genetic purity of pigeonpea. *Euphytica* **66**: 225-229.

Saxena KB, Chauhan YS, Johansen C, Singh L. 1992b. Recent developments in hybrid pigeonpea research. In: *New Frontiers in Pulses Research and Development*. Indian Institute of Pulses Research, Kanpur, India: 58-69

Saxena KB, Sharma, D. 1995. Sources of dwarfism in pigeonpea. *Indian Journal of Pulses Research* **8**:1–6.

Saxena KB, Rao AN, Singh U, Remnanadan P. 1996. Interspecies variation in *Cajanus platicarpus* for some agronomic traits and crossability. *Intel Chickpea and Pigeonpea Newsletter* **3**: 49–51.

Saxena KB, Kumar RV. 2003. Development of a cytoplasmic- nuclear male-sterility system in pigeonpea using *C. scarabaeoides* (L.) Thouars. *Indian Journal of Genetics* **63**: 225—229.

Saxena KB, Srivastava DP, Chauhan YS, Ali M. 2005a. Hybrid pigeonpea. In: Ali M, Kumar S. eds. *Advances in pigeonpea research*. IIPR Kanpur, India: 96-133.

Saxena KB, Kumar RV, Srivastava N, Bao S. 2005b. A cytoplasmic-nuclear male-sterility system derived from a cross between *Cajanus cajanifolius* and *Cajanus cajan*. *Euphytica* **145**: 289-294.

Saxena KB. 2006. Hybrid pigeonpea seed production manual. *International Crops research Institute for the Semi Arid Tropics*. Patancheru, India. Bull **74**: 27.

Saxena KB, Kumar RV, Madhavilatha K, Dalvi VA. 2006. Commercial pigeonpea hybrids are just a few steps away. *Indian Journal of Pulses Research* **19**: 7–16.

Saxena KB. 2008. Genetic Improvement of Pigeonpea- A Review. *Tropical Plant Biology* **1**:159–178.

Saxena KB. 2009. Evolution of hybrid breeding technology in pigeonpea. In: Ali M, Kumar S. eds. *Milestones in Food Legume Research.* Indian Institute of Pulses Research, Kanpur, India: 82-114.

Saxena KB, Sultana R, Mallikarjuna N, Saxena RK, Kumar RV, Sawargaonkar SL, Varshney RK. 2010a. Male-sterility systems in pigeonpea and their role in enhancing yield. *Plant Breeding* **129**: 125-134.

Saxena KB, Kumar RV. 2010b. Insect-aided natural out-crossing in four wild relatives of pigeonpea. *Euphytica* **173**: 329–335.

Saxena RK, Prathima C, Saxena K, Hoisington DA, Singh NK, Varshney RK. 2010c. Novel SSR Markers for Polymorphism Detection in Pigeonpea (*Cajanus* spp.). *Plant Breeding* **129**: 142–148.

Schrire BD, Lavin M, Lewis GP. 2005. Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Biologiske Skrifter* 55: 375–422.

Sharma HC, Pampapathy G, Reddy LJ. 2003. Wild relatives of pigeonpea as a source of resistance to the pod fly (*Melanagromyza obtusa* Malloch) and pod wasp (*Tanaostigmodes cajaninae* La Salle). *Genetic Resources and Crop Evolution* ***50***: 817–824.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare. II: Relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* ***92***: 142–166.

Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**: 275–288.

Shultz JL, Kazi S, Bashir R, Afzal JA, Lightfoot DA. 2007. The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. *Theoretical and Applied Genetics* **114**:1081–1090.

Sivaramakrishnan S, Seetha K, Reddy LJ. 2002. Diversity in selected wild and cultivated species of pigeonpea using *RFLP* of *mtDNA*. *Euphytica* **125**: 21-28.

Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17**: 145-170.

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JD, Martin PG. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic

nitrogen fixation in angiosperms. *Proceedings of the National Academy of Sciences USA* **92**: 2647–2651.

Soltis DE, Soltis PS. 1998. Choosing an approach and an appropriate gene for phylogenic analysis. In: Soltis, DE, Soltis PE, Doyle JJ. eds. *Molecular Systematics of Plants II: DNA Sequencing*. Kluwer Academic, Dordrecht: 1-42.

Soltis PS, Soltis DE. 2004. The origin and diversification of angiosperms. *American Journal of Botany* **91**: 1614-1626.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Michael, Fay MF, Axtell M, Swensen SM, Prince LM, W.John Kress WJ, Nixson KC, Farris, JS. 2000. Angiosperm phylogeny inferred from a combined data set of *18s rDNA*, rbcl and *atpb* Sequences. *Botanical Journal of The Linnean Society* **133**: 381–461.

Sonnante G, Stockton T, Nodari RO, Becerra Vela´squez VL, Gepts P. 1994. Evolution of genetic diversity during the domestication of common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* **89**: 629–635.

Souframanien J, Manjaya JG, Krishna, TG, Pawar ME. 2003. Random amplified polymorphic DNA analyses of cytoplasmic male sterile and male fertile pigeon pea (*Cajanus cajan* (L.) Millsp. *Euphytica* **129**: 293-299.

Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ. 2005. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences of the USA* **102**: 14694–14699.

Sprent JI. 1994. Evolution and diversity in the legume-rhizobium symbiosis: chaos theory? *Plant and Soil* **161**: 1-10.

Sprent JI. 2001. Nodulation in legumes. Royal Botanic Gardens, Kew, UK.

Srivastava N, Vadez V, Upadhyaya HD, Saxena KB. 2006. Screening for intra and inter specific variability for salinity tolerance in pigeonpea (*Cajanus cajan* (L.) Millsp.) and its related wild species. *E-Journal of SAT Agriculture Research Crop Improvement* **2**: 1.

Steele KP, Wojciechowski MF. 2003. Phylogenetic analyses of tribes Trifolieae and Vicieae, based on sequences of the plastid gene *matK* (Papilionoideae: Leguminosae). In: *Advances in legume systematics 10*, higher level systematics, eds. Klitgaard, B.B. & Bruneau, A., pp. 355–370. Royal Botanic Garden, Kew, UK.

Stefanovic S, Pfeil BE, Palmer JD, Doyle JJ. 2009. Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. *Systematic Botany* **34**: 115-128.

Stirton CH. 1975. A contribution to knowledge of the genus *Eriosema* (Leguminosae-Lotoideae) in Southern Africa (excluding Mocambique and Rhodesia): a thesis. University of Natal, Pitermaritzburg.

Stirton CH. 1981. Studies in the Leguminosae-Papilionoideae of Southern Africa. *Bothalia* **13**: 317-325.

Subbarao GV. 1988. Salinity tolerance in pigeonpea (*Cajanus cajan* (L.) Millsp.) and its wild relatives. PhD thesis, Indian Inst Technol, Kharagpur, India

Swofford DL. 2002. PAUP*- phylogenetic analysis using parsimony (*and other methods), Version 4.0b10. Sinauer Associates, Sunderland.

Sytsma KJ, Morawetz J, Pires JC, Nepokroeff M, Conti E, Zjhra M, Hall JC, Chase MW. 2002. Urticalean Rosids: Circum- Scription, Rosid ancestry, and phylogenetics based on *rbcl*, *trnlf*, and *ndhf* Sequences. *American Journal of Botany* **89**: 1531–1546.

Taberlet P, Geilly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105-1109.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**:1596–1599.

Tan X, Meyers BC, Kozik A, West MAL, Morgante M, St. Clair DA, Bent AF, Michelmore, RW. 2007. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biology* **7**: 56.

Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**: 1063–1666.

Tarr DE, Alexander HM. 2009. TIR–NBS–LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Research Notes* **2**:197.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. *Molecular Biology and Evolution* **21**: 1214–1225.

Thompson IR, Ladiges PY, Ross JH. 2001. Phylogenetic studies of the tribe *Brongniartieae* (Fabaceae) using nuclear DNA (*ITS-1*) and morphological data. *Systematic Botany* **26**: 557–570.

Thulin M, Lavin M, Pasquet R, Delgado-Salinas A. 2004. Phylogeny and biogeography of *Wajira* (Leguminosae): a monophyletic segregate of *Vigna* centered in the horn of Africa region. *Systematic Botany* **29**: 903–920.

Tikka SBS, Parmar LD, Chauhan RM. 1997. First record of cytoplasmic-genic male-sterility in pigeonpea (*Cajanus cajan* (L.) Millsp.) through wide hybridization. *GAU Res J.* **22**: 160–162

Tindall HD. 1988. *Vegetables in the Tropics*. Macmillan International, UK.

Torke BM, Schaal BA. 2008. Molecularphylogenetics of the species-rich neotropical genus *Swartzia* (Leguminosae, Papilionoideae) and related genera of the swartzoid clade. *American Journal of Botany* **95**: 215–228.

Tucker SC. 1987. Pseudoracemes in papilionoid legumes: their nature, development and variation. *Botanical Journal of the Linnean Society* **95**: 181–206.

Tucker SC. 2002. Floral ontogeny in *Sophoreae* (Leguminosae: Papilionoideae). III. Radial symmetry and random petal aestivation in *Cadia purpurea*. *American Journal of Botany* **89**: 748–757.

Tucker SC. 2003. Floral development in legumes. *Plant Physiology* **131**: 911–926.

Upadhyaya HD, Reddy LJ, Gowda CLL, Reddy KN, Singh S. 2006. Development of a mini-core subset for enhanced and diversified utilization of pigeonpea germplasm resources. *Crop Science* **46**: 2127–2132.

Upadhyaya HD, Reddy KN, Gowda CLL, Singh S. 2007. Phenotypic diversity in the pigeonpea (*Cajanus cajan*) core collection *Genetic Resources and Crop Evolution* **54**: 1167–1184.

Van der Maesen LJG. 1980. India is the native home of the pigeonpea. *Liber gratulatorius in honorem HCD de Wit, Misc. Papers 19.* Landbouwhogesch, Wageningen, the Netherlands: 257-262.

Van der Maesen LJG. 1983. World distribution of pigeonpea. *ICRISAT Inf. Bull. 14*. Patanchem, AP India.

Van der Maesen LJG. 1986. *Cajanus* DC. and *Atylosia* W. & A. (Leguminosae). Agricultural University Wageningen papers 85-4: 1–225.

Van der Maesen LJG. 1990. Pigeonpea origin, history, evolution, and taxonomy. In: Nene YL, Halls D, Sheila VK. eds. *The Pigeonpea*. CAB International, Wallingford, Oxon, UK: 15-46.

Van der Maesen LJG. 1998. Revision of the genus *Dunbaria* Wight & Arn. (Leguminosae-Papilionoideae). *Wageningen Agricultural University Papers* 98:1.

Van der Maesen LJG. 2003. Cajaninae of Australia (Leguminosae: Papilionoideae). *Australian Systematic Botany* **16**: 219-227.

Varshney RK. 2009a. Gene-based marker systems in plants: high throughput approaches for discovery and genotyping. In: Jain SM, Brar DS. eds. *Molecular techniques in crop improvement volume 2*. Springer, The Netherlands: doi: 10.1007/978-90-481-2967-6-5.

Varshney RK, Nayak SN, May GD, Jackson SA. 2009b. Next generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* **27**:522-530.

Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Datta S, Sharma TR, Rosen B, Carrasquilla- Garcia N, Farmer AD, Dubey A, Saxena KB, Gao J, Fakrudin B, Singh MN, Singh BP, Wanjari KB, Yuan M, Srivastava RK, Kilian A, Upadhyaya HD, Mallikarjuna N, Town, CD, Bruening GE, He G, May GD, McCombie R, Jackson SA, Singh NK, Cook, DR. 2010. Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Molecular Breeding* **26**: 393-408.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407- 4414.

Wakasugi T, Sugita M, Tsudzuki T, Sugiura M. 1998. Updated gene map of tobacco chloroplast DNA. *Plant Molecular Biology Reporter* **16**: 231–241.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES, Lipshutz R, Chee M. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**:1077–1082.

Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences* USA **106**: 3853-3858.

Wanjari KB, Patil AN, Manapure P, Manjaya JG, Manish P. 1999. Cytoplasmic male-sterility in pigeonpea with cytoplasm from *Cajanus volubilis*. *Annual Review of Plant Physiology* **13**: 170—174.

Wanjari KB, Patil AN, Patel MC, Manjaya JG. 2000. Male sterility derived from *Cajanus sericeus* and *Cajanus cajan*: Confusion of cytoplasmic male sterility with dominant genic male-sterility. *Euphytica* **115**: 59–64.

Wasike S, Okori P, Rubaihayo PR. 2005. Genetic variability and relatedness of the Asian and African pigeonpea as revealed by AFLP. *African Journal of Biotechnology* **4**: 1228-1233.

Weber JL, May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *The American Journal of Human Genetics* **44**: 388 396.

Wendel JF, Schnabel A, Seelanan T. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences of the USA* **92**: 280–284.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ.

eds. *PCR protocols: a guide to methods and applications.* Academic Press, San Diego: 315-324

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531-6535.

Wojciechowski MF. 2003. Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective. In: Klitgaard BB, Bruneau A. eds. *Advances in legume systematics 10, Higher level systematics*. Royal Botanic Garden, Kew, UK: 5-35.

Wojciechowski, MF. 2005. *Astragalus* (Fabaceae): a molecular phylogenetic perspective. *Brittonia* **57**: 382-396.

Wojciechowski MF, Sanderson MJ, Hu J-M. 1999. Evidence on the monophyly of *Astragalus* (Fabaceae) and its major subgroups based on nuclear ribosomal DNA *ITS* and chloroplast DNA *trnL intron* data. *Systematic Botany* **24**: 409–437.

Wojciechowski MF, Lavin M, Sanderson MJ. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well- supported subclades within the family. *American Journal of Botany* **91**: 1846-1862.

Wolfe KH. 1991. Protein-coding genes in chloroplast DNA: compilation of nucleotide sequences, data base entries, and rates of molecular evolution. In: Bogorad L, Vasil IK. eds. *Cell culture and somatic cell genetics of plants 7B*. Academic Press, New York, USA: 467-482.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. *Science* **308**:1310–1314.

Wyckoff GJ, Wang W, Wu CI. 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**: 304-309.

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J. 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* **4**: e8451.

Yang S, Ash G, Harper J, Varling J, Wenzl P, Huttner E. 2006. Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. *Theoretical and Applied Genetics* **113**: 585–595.

Yang Z, Gu S, Wang X, Li W, Tang Z, Xu C. 2008. Molecular evolution of the CPP-like gene family in plants: Insights from comparative genomics of *Arabidopsis* and rice. *Journal of Molecular Evolution* **67**: 266-277.

Young ND, Mudge J, Ellis THN. 2003. Legumes genomes: more than peas in a pod. *Current Opinion in Plant Biology* **6**: 199–204.

Zeder MA. 2006. Central questions in the domestication of plants and animals. *Evol. Anthropology* **15**:105–117.

Zhang LB, Zhu Q, Wu Z, Ross-Ibarra J, Gaut BS, Ge S, Sang T. 2009a. Selection on grain shattering genes and rates of rice domestication. *New Phytologist* **184**: 708–720.

Zhang ML, Fritsch PW, Cruz BC. 2009b. Phylogeny of *Caragana* (Fabaceae) based on DNA sequence data from *rbcL, trnS-trnG*, and *ITS*. *Molecular Phylogenetics and Evolution 50: 547–559.*

Zhou JH, Wang JL, Xu JC, Lei CL, Ling ZZ. 2004. Identification and mapping of a rice blast resistance gene Pi-g(t) in the cultivar Guangchangzhan. *Plant Pathology* **53**: 191-196.

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: Severe bottleneck during domestication of rice. *Molecular Biology and Evolution* **24**: 875–882.

Appendix 1: Sequence alignments of the ITS region

```
[ 10                              20         30          40                  50]
Cajanus cajan           CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus cajanifolius    CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus scarabaeoides   CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus cinereus-1      CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus reticulatus     CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus lanceolatus     CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus cinereus-2      CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus acutifolius-2   CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus crassicaulis    CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus latisepalus-1   CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus cinereus-1      CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus latisepalus-2   CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus pubescens       CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus acutifolius-1   CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus acutifolius-3   CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus viscidus        CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus hirtopilosus    CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTGA
Cajanus lineatus        CCGGTGAATT TGTTTATCTA ----CTCAGG AT--TGGCTT GAGGTGTTGA
Cajanus albicans        CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GATGTGTTGA
Cajanus marmoratus      CCGGTGAACT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTGTTCG
Rhynchosia aurea        CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GTGGTGTTCA
Cajanus mollis          CCGGTGAATT TGTTTATCTA ----CTCGGG AT--TGGCTT GAGGTATTCA
Dunbaria ferruginea     CCGGTGAATT TGTTTATCTA ----CTCGGG ATTGTGGCTT GAGGTGTTAA
Rhynchosia hauthalii    CCAGTGAATT TGTTTATCTA ----CTTGGG AT--TGGCAT GACTTGTTGA
Rhynchosia verdcourtii CCGGTGAATT TGTTTATCTA ----CTTGGG AT--TGGCAT GACTTGTTGA
Rhynchosia australis    CCGGTGAATT TGTTTATTTA ----CTTGGG AT--GGGCAT GACTTGTTGA
Rhynchosia orthobotrya CTGGTGAATT TGTTTATCTA ----CTTGGG AT--TGGCAT GAGATGTTCA
Bolusafra bituminosa    CCGGCGAATT TGTTTATCTA ----CTTGGG AT--TGGCAT GAGGTGCTTA
Eriosema griseum        CTGGTGAATT TGTTATTCAA ----CTTAGG AT--TGGCAT GAGGTGTTTA
Eriosema simulans       CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Eriosema burkei         CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Eriosema lucipetum      CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Eriosema kraussianum    CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Eriosema salignum       CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Eriosema rossii         CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Eriosema squarrosum     CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Eriosema umtamvunense   CTGGTGAATT TGTTATTCAA ----CTTGGG AT--TGGCAT GAGGTGTTCA
Flemingia stricta       CCGGCGAACC TGTTCGTCTA TACCCATAGG GT---TGGAG GAGGCGTTCA
Flemingia parviflora    CCGGCGAACC TGTTTGTCTA TACCCATTGG GT--CTGGAG GAGGCGTTCA
Flemingia macrophylla   CCGGCGAACC TGTTTGTCTA TACCCATTGG GT--CTGGAG GAGGCGTTCA
Flemingia strobilifera CCGGCGAACT CGTTTGTCTA --CCCGTTGG GTCTTCGCAG GAGGCGTTCA
Flemingia glutinosa     CCGGCGAACC TGTTTGTCTA --CCCGTCGG GT--CTGGAG GAGGCGTTCG
Flemingia trifolias     CCGGCGAACT TGTTTGTCTA --CCTGTCGG GT--CTGGAG G-GACGTTCG
Flemingia lineata       CCGGCGAACT TGTTTGTCTA --CCTGTCGG GT--CTGGAG G-GACGTTCG
Glycine microphylla     CCCGCGAATT TGTTTATCCA ---------- ---------- --------TC
Glycine stenophita      CCCGCGAATT TGTTTATCCA ---------- --------TC TACCGTCG--
Glycine pullenii        CCGGTGGATT TGTTTATACA ---------- --------TC TATCGTCG--
Glycine tomentella      CCCGCGAATT TGTTTATACA ---------- --------TC TACCGTCG--
Glycine max             CCCGCGAACT TGTTTATTCA ---------- --------TC TACCGTCG--
Vigna luteola           CCAGTGAATT TGTTCATGCC --C-T--CTA AAT------- ---GTCG---

[                                60         70         80         90          100]
Cajanus cajan             -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus cajanifolius      -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus scarabaeoides     -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus cinereus-1        -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus reticulatus       -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus lanceolatus       -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus cinereus-2        -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus acutifolius-2     -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus crassicaulis      -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus latisepalus-1     -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus cinereus-1        -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus latisepalus-2     -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus pubescens         -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus acutifolius-1     -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus acutifolius-3     -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus viscidus          -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus hirtopilosus      -TTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus lineatus          -CTAACATCT CATCCTTCCT CGTGTTT--G GAGGGAGTG- G---------
Cajanus albicans          -TTAACATCT CATCCTTCCT CGTGTTT--A GAGGGAGTG- G---------
```

```
Cajanus marmoratus        -TGAACATCT CATCCAGTCC TGTGTTT--G GAGGGAGTT- G---------
Rhynchosia aurea          -TGAACATCT CATCCAGCCC TGTGTTT--G TAGGGAGTT- G---------
Cajanus mollis            -TCAATATCT CATCCTCCCC CATGTTT--G GGGGGAGTG- G---------
Dunbaria ferruginea       -TCAATATCT CATCCTCCCC CGTGTTT--- GGAGGAGTG- G---------
Rhynchosia hauthalii      -ACAACATTT CATCCTTTCT TTTGTTG--G GAGGGAGTT- G---------
Rhynchosia verdcourtii    -ACAACATTT CATCCTTTCC TTTGTTG--- GGGGGAGGT- G---------
Rhynchosia australis      -AAAACATCT CATCCTTTCC TTCTTTGTTG GGGGAAGGT- G---------
Rhynchosia orthobotryasa -ACAACATTA CATCCTTTCC TCTGTTG--G GAGGGTGCT- G---------
Bolusafra bituminosa      -ACGGCATCT CATACTCTCC TTTGTTG--G GAGGGAGGT- G---------
Eriosema griseum          -ACACCATCT CATCCTTTTC TTGTTTG--T GAGGTAGTTG G---------
Eriosema simulans         -ACAACATCT CATCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Eriosema burkei           -ACAACATCT CATCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Eriosema lucipetum        -ACAACATCT CATCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Eriosema kraussianum      -ACAACATCT CATCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Eriosema salignum         -ACAACATGT CATCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Eriosema rossii           -ACAACATGT CGTCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Eriosema squarrosum       -ACAACATGT CGTCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Eriosema umtamvunense     -ACAACATGT CGTCCTTTTC TTGTTTG--G GAGGGAGTT- G---------
Flemingia stricta         -ACGACGCCC CCTCCTCTCC CGAGTCG--G GGCGGAGGC- G---------
Flemingia parviflora      -ACGACGCCC CCTCCTCTCC CGAGTCG--G GGCGGAGGC- G---------
Flemingia macrophylla     -ACGACGCCC CCTTTTCTCC CGAGTCG--G GGCGGAGAC- G---------
Flemingia strobilifera    -ACAACGCCC TCTCCTCTCC CGAGTCG--G GGCGGAGGC- G---------
Flemingia glutinosa       -AGGACGCCC TTTCCTCTCC CGAGTTG--G GAAGGAGGC- ----------
Flemingia trifoliastrum  AAGGACGCCC TTTCATCTCC CGAGTTG--G GAAGGAGGC- A---------
Flemingia lineata         -AGGACGCCC TTTCATCTCC CGAGTTG--G GAAGGAGGC- A---------
Glycine microphylla       TACCGTCG-- ---------- ---------G GAGGGAGGG- A--GGAAGAC
Glycine pullenii          ---------- ---------- ---------G GAGGGAGGG- G--GGAAGAC
Glycine pullenii          ---------- ---------- ---------G GAGGGGAGG- GGTGGAAGAC
Glycine tomentella        ---------- ---------- ---------G GAGGAGGGG- G--GGAAGAC
Glycine max               ---------- ---------- ---------G GAGGG-A-G- G--GGATGAC
Vigna luteola             ---------- ---------- ---------T GAGAGATGG- G---------

[                                  110        120        130        140        150]
Cajanus cajan             ---------- -GCTGTGCCA TGCTCGGCTG ACTTTCTTTG ---AACAAAA
Cajanus cajanifolius      ---------- -GCTGTGCCA TGCTCGGCTG ACTTTCTTTG ---AACAAAA
Cajanus scarabaeoides     ---------- -GCTGTGCCA TGCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus cinereus-1        ---------- -GCTGGGCCA TGCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus reticulatus       ---------- -GCTGGGCCA TGCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus lanceolatus       ---------- -GCTGGGCCA TGCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus cinereus-2        ---------- -GCTGGGCCA CTCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus acutifolius-2     ---------- -GCTGGGCCA TTCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus crassicaulis      ---------- -GCTGGGCCA TTCTTGGCCG ACTTTCTTTG ---AACAAAA
Cajanus latisepalus-1     ---------- -GCTGGGCCA TTCTTGGCCG ACTTTCTTTG ---AACAAAA
Cajanus cinereus-1        ---------- -GCTGGGCCA TTCTTGGCCG ACTTTCTTTG ---AACAAAA
Cajanus latisepalus-2     ---------- -GCTGGGCCA TTCTTGGCCG ACTTTCTTTG ---AACAAAA
Cajanus pubescens         ---------- -GCTGGGCCA TTCTTGGCCG ACTTTCTTTG ---AACAAAA
Cajanus acutifolius-1     ---------- -GCTGGGCCA TTCTTGGCCG ACTTTCTTTG ---AACAAAA
Cajanus acutifolius-3     ---------- -GCTGGGCCA TTCTTGGCCG ACTTTCTTTG ---AACAAAA
Cajanus viscidus          ---------- -GCTGGGCCA CGCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus hirtopilosus      ---------- -GCTGGGCCA CGCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus lineatus          ---------- -GCTGTGCCA TGCTTCGCCG ACTTTCTTTG ---AACAAAA
Cajanus albicans          ---------- -GCTGTGCCA TGCTCGGCCG ACTTTCTTTG ---AACAAAA
Cajanus marmoratus        ---------- -GCTGTGCCA TGCTCGGTCG ACTTTCTATG ---AACAAAA
Rhynchosia aurea          ---------- -GCTGTGCCA TGCTCGGTCG ACTTTCTATG ---AACAAAA
Cajanus mollis            ---------- -GTTGTGCCA TGCTCGGTCC ACTCTCTTTG ---AACAAAA
Dunbaria ferruginea       ---------- -GCTGTGCCA TGCTCGGTCG ACTCTCTTTG ---AACAAAA
Rhynchosia hauthalii      ---------- -GTTGTGCCA TGCTCAGTTG GCTTTCTCTC AACAATAAAA
Rhynchosia verdcourtii    ---------- -GCTGTGCCA TGCTCGGTTG GCTTTCTCTC ---AACAAAA
Rhynchosia australis      ---------- -ACCGTGCCA TGCTCAGTTG GCTTTCTCTC ---AACAAAA
Rhynchosia orthobotryasa  ---------- -GCTGGGCCA TGCTCGGTTG GCTCTCTCTC A--AACAAAA
Bolusafra bituminosa      ---------- -GCTGTGCCC TGCTCGGTCG GCTCTCTCTC A--AACAAAA
Eriosema griseum          ---------- -GATGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema simulans         ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema burkei           ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema lucipetum        ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema kraussianum      ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema salignum         ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema rossii           ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema squarrosum       ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Eriosema umtamvunense     ---------- -GTTGGGTTG TGCTCAATTG TCTATCTCTC A--AACTAAA
Flemingia stricta         ---------- -GTCGTGCCT TGCTCGGTTG GCTCCTTTCC C--GACAAAA
Flemingia parviflora      ---------- -GTCGTGCCT TGCTCGGCTG GCTCCTTTCC C--GACAAAA
Flemingia macrophylla     ---------- -GTCGTGCCT TGCTCGGCCG GCTCCTTTCC C--GACAAAA
Flemingia strobilifera    ---------- -GTTGTGCCT AGCTTGGCTG GCTCCTTTCC C--GACAAAA
Flemingia glutinosa       ---------- ---------- ---------G GCTCC-TTCC C--AGCAAAA
Flemingia trifoliastrum   ---------- -GTCGTGCCT TGCTCGGCTG GCTCTCTTCC C--AGCAAAA
Flemingia lineata         ---------- -GTCGTGCCT TGCTCGGCTG GCTCTCTTCC C--AGCAAAA
```

```
Glycine microphylla      CAGGTCGCCC CGTGCGGCTG GGCTCATCCT CCTCGTCCTC A--CGGCAAA
Glycine stenophita       CAGGTCGCCT TGTGCGGCTG GGCTCCTGCT CCTCGTCCTC A--CGACAAA
Glycine pullenii         CAAGTCGCCC CGTGCGGCTG GCCTCCTCAT CCTCGTCCTC A--CGACAAA
Glycine tomentella       aCAAGTCGCCC CGTGCGGCTG GGCTCCTCCT CCTCGTCCTC A--CGACAAA
Glycine max              CACGGCGCCC CGTGC----- -GCCCGGCCT CCTCGTCCTC G--CGACAAA
Vigna luteola            ---------- --------AT AATGCGTCCT -GTCTTTTTC G--GCAAAAT

[                                  160        170        180        190        200]
Cajanus cajan            CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus cajanifolius     CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus scarabaeoides    CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus cinereus-1       CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus reticulatus      CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus lanceolatus      CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus cinereus-2       CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus acutifolius-2    CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus crassicaulis     CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus latisepalus-1    CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus cinereus-1       CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus latisepalus-2    CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus pubescens        CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus acutifolius-1    CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus acutifolius-3    CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus viscidus         CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus hirtopilosus     CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus lineatus         CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus albicans         CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTTCAGTGCA
Cajanus marmoratus       CTCAAACCCC GGCGCTTCGT GTGCCAAGGA ATAAAAAATT GTTAAGTGCA
Rhynchosia aurea         CTCAAACCCC GGCGCTTCGT GTGCCAAGGA ATAAAAAATT GTTAAGTGCA
Cajanus mollis           CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTCCAGTGCA
Dunbaria ferruginea      CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATAAAAAATT GTCCAGTGCA
Rhynchosia hauthalii     CTCTAACCCC GGCGCTTCGT GTGCCAAGGA ATCCAAAATT GTTTAGTGCA
Rhynchosia verdcourtii   CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATCCAAATAT GTTTAGTGCA
Rhynchosia australis     CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATCCAAATAT GTTTAGTGCG
Rhynchosia orthobotryasa CTAAAACCCC GGCGCTTCGT GTGCCAAGGA ATCCAAAATT GTTTAGTGCA
Bolusafra bituminosa     CTCAAACCCC GGCGCTTCGT GCGCCAAGGA ATCCAAAATT GTTTAGTGCA
Eriosema griseum         TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTG
Eriosema simulans        TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Eriosema burkei          TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Eriosema lucipetum       TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Eriosema kraussianum     TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Eriosema salignum        TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Eriosema rossii          TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Eriosema squarrosum      TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Eriosema umtamvunense    TTCAAACCCC GGCACTTTGT GTGCCAAGGA ATCCATTATT GTTTAGTGTA
Flemingia stricta        CACAAACCCC GGCGCTTCGT GCGCCAAGGA ATTCAAAACT GTTAAGCGCA
Flemingia parviflora     CACAAACCCC GGCGCTTCGT GCGCCAAGGA ATTCAAAACT GTTAAGCGCA
Flemingia macrophylla    CACAAACCTC GGCGCTTCGT GCGCCAAGGA ATTCAAAAAT GTTAAGCGCA
Flemingia strobilifera   CACAAACCCC GGCGCTTCGT GCGCCAAGGA ATTCAAAACT GTTAAGCGCA
Flemingia glutinosa      CACAAACCCC GGCGCTTCGT GCGCCAAGGA ATTCATAATT GTTAAGCGCA
Flemingia trifoliastrum  CAACAACCCC GGCGCTTCGT GCGCCAAGGA ATTCAGAACT GTTAAGCGCA
Flemingia lineata        CAACAACCCC GGCGCTTCGT GCGCCAAGGA ATTCAGAACT GTTAAGCGCA
Glycine microphylla      CACAAACCCC GGCGCTTCGT GCGCCAAGGA ACTCAAATCT GTTAAGTGCG
Glycine stenophita       CACAAACCCC GGCGCTTCGT GCGCCAAGGA ACTCAAATCT GTTAAGTGCG
Glycine pullenii         CACAAACCCC GGCGCTTCGT GCGCCAAGGA ACTCAAATCT GTTAAGTGCG
Glycine tomentella       CACAAACCCC GGCGCTTCGT GCGCCAAGGA ACTCAAATCT GTTAAGTGCG
Glycine max              CACAAACCCC GGCGCTTCGT GCGCCAAGGA ACTCAAATCT GTTAAGTGCG
Vigna luteola            CACAAACTCC GGCACTTTAT GTGCCAAGGA TCCTAAACGT TTTT-GTG-G

[                                  210        220        230        240        250]
Cajanus cajan            A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus cajanifolius     A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus scarabaeoides    A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus cinereus-1       A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus reticulatus      A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus lanceolatus      A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus cinereus-2       A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus acutifolius-2    A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus crassicaulis     A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus latisepalus-1    A-TGTCGTGG ACCCGGAGAC GGCGTTCTAC G-GCAATGTC TCAACACATG
Cajanus cinereus-1       A-TGTCGTGG ACCCGGAGAC GGCGTTCTAC G-GCAATGTC TCAACACATG
Cajanus latisepalus-2    A-TGTCGTGG ACCCGGAGAC GGCGTTCTAC G-GCAATGTC TCAACACATG
Cajanus pubescens        A-TGTCGTGG ACCCGGAGAC GGCGTTCTAC G-GCAATGTC TCAACACATG
Cajanus acutifolius-1    A-TGTCGTGG ACCCGGAGAC GGCGTTCTAC G-GCAATGTC TCAACACATG
Cajanus acutifolius-3    A-TGTCGTGG ACCCGGAGAC GGCGTTCTAC G-GCAATGTC TCAACACATG
Cajanus viscidus         A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus hirtopilosus     A-TGTCGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
```

```
Cajanus lineatus          A-TGTTGTGG ACCCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus albicans          A-TGTCGTGG AACCGGAGAC GGTGTTCTAC G-GCAATGTC TCAACACATG
Cajanus marmoratus        A-CCTCGTGG GCCCGGATTC GGTGTCCTGC G-GTTTTGTC ACGACACATG
Rhynchosia aurea          A-CCTCATGG GCCCGTATTC GGTGTCCTGC G-GTGTTGTC ACGACACATG
Cajanus mollis            A-TCTCGTGG ACCCGGAGAC GGTGTTCTAC G-GTTTTGTT GGAACACATG
Dunbaria ferruginea       A-TCTCGTGG ACCCGGAGAC GGTGTTCTAC G-GTGTTGTC CCGACACAC-
Rhynchosia hauthalii      A-TCTCGGGG ACCTGGAGAC AGTGTTTCYT GGGTGTTGTC ATGACACACA
Rhynchosia verdcourtii    A-TCTCAGGG ACCCGGAGAC GGTGTTTCCT GGGTGTTGTC ACGACACACG
Rhynchosia australis      A-TCTCGGGG ACCCGGAGAC GGTGTTTCCT GGGTGTTGTC ACGACACACG
Rhynchosia orthobotryasa  A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGTC ACGACACATG
Bolusafra bituminosa      A-TCTCGGGG ACCCGGAGAC GGTGTCTCCC GAGTGTTGTC ACGACACACG
Eriosema griseum          A-TCTCGAGG ACCCGGAGAC GGTGTTTCGC GAGTGCTGCC ATAACACATG
Cajanus marmoratus        A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Eriosema burkei           A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Eriosema lucipetum        A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Eriosema kraussianum      A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Eriosema salignum         A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Eriosema rossii           A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Eriosema squarrosum       A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Eriosema umtamvunense     A-TCTCGAGG ACCCGGAGAC GGTGTTTCAC GAGTGTTGCC ATAACACATG
Flemingia stricta         G-CACGGCGG ACCCGGAGAC GGCGTTCCGC GGGTGTTGTG GTGACACGAG
Flemingia parviflora      G-CACGGCGG ACCCGGAGAC GGTGTTCCGC GGGTGTTGTG CCGACAAGAG
Flemingia macrophylla     G-CACGGCGG ACCCGGATAC GGCGTTCCGC GGGTGTTGTG GTGACACGAG
Flemingia strobilifera    G-C------- --------AC GGCGTTCCGC GGGTGTTGTG GCTACACGAG
Flemingia glutinosa       G-CACCGTGG ACCCGGAGAC GGCGTTCT-- -GGCGTTGTT GCGACACGAG
Flemingia trifoliastrum   G-CACTGTGG ACCCGGAGAC GGCGTTCTGC GGGTGTTGTT GCGACACGAG
Flemingia lineata         G-CACTGTGG ACCCGGAGAC GGCGTTCTGC GGGTGTTGTT GCGACACGAG
Glycine microphylla       ACTCCCGGGG GCCCGGAGAC GGTGTCCCGC GGGCGTCGTC ACGACACATC
Glycine stenophita        ACTCCCGGGG GCCCGGAGAC GGTGTCCCGT GGGCGTCGTC ACGACACATC
Glycine pullenii          ACTCTCGAGG GCCCGGAGAC GGTGTCCCGC GGGCGTCGTC ACGACAAATC
Glycine tomentella        ACTCCCGGGG GCCCGGAGAC GGTGTCCCTT GGGCGTCGTC ACGACAAATC
Glycine max               ACTCCCGGGG GCCCGGAGAC GGTGTCCCGC GGGAGTCGTC ACGACACAAC
Vigna luteola             AC----G-AT TCTCAAGGAC TGTGTCCTAC GGGAATCGTC TCG-----T-

[                                  260        270        280        290        300]
Cajanus cajan             ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus cajanifolius      ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus scarabaeoides     ----ATA-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus cinereus-1        ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus reticulatus       ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus lanceolatus       ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus cinereus-2        ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus acutifolius-2     ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus crassicaulis      ----ATA-TG AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus latisepalus-1     ----ATA-TG AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus cinereus-1        ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus latisepalus-2     ----ATA-TG AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus pubescens         ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus acutifolius-1     ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus acutifolius-3     ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus viscidus          ----GTA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus hirtopilosus      ----GTA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus lineatus          ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus albicans          ----ATA-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus marmoratus        ----ATA-CA GAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Rhynchosia aurea          ----ATA-CA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Cajanus mollis            ----ATT-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Dunbaria ferruginea       ----ATA-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Rhynchosia hauthalii      ----ATATTA TAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Rhynchosia verdcourtii    ----ATATTA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Rhynchosia australis      ----ATATTA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Rhynchosia orthobotryasa  ----ATA-CA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Bolusafra bituminosa      ----ATA-CA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema griseum          ----ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema simulans         ----ATC-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema burkei           ----ATC-TA AAAGGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema lucipetum        ----ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema kraussianum      ----ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema salignum         ----ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema rossii           ----ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema squarrosum       ----ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Eriosema umtamvunense     ----ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Flemingia stricta         T-T-ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Flemingia parviflora      T-T-ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Flemingia macrophylla     T-T-ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Flemingia strobilifera    T-T-ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Flemingia glutinosa       T-T-ATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
```

```
Flemingia trifoliastrum  T-TAATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Flemingia lineata        T-TAATC-TA AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Glycine microphylla      A-TTACA-TA CAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Glycine stenophita       A-TTACA-TA CAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Glycine pullenii         A-TTACA-TA CAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Glycine tomentella       A-TTACA-TA CAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Glycine max              ATTTACA-TA CAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA
Vigna luteola            ATTGTCA-TG AAATGACTCT CGGCAACGGA TATCTCGGCT CTTGCATCGA

[                                 310        320        330        340        350]
Cajanus cajan            TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus cajanifolius     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus scarabaeoides    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus cinereus-1       TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus reticulatus      TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus lanceolatus      TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus cinereus-2       TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus acutifolius-2    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus crassicaulis     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus latisepalus-1    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus cinereus-1       TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus latisepalus-2    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus pubescens        TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus acutifolius-1    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus acutifolius-3    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus viscidus         TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus hirtopilosus     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus lineatus         TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus albicans         TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus marmoratus       TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Rhynchosia aurea         TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Cajanus mollis           TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Dunbaria ferruginea      TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Rhynchosia hauthalii     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Rhynchosia verdcourtii   TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Rhynchosia australis     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Rhynchosia orthobotryasa TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Bolusafra bituminosa     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema griseum         TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema simulans        TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema burkei          TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema lucipetum       TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema kraussianum     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema salignum        TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema rossii          TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema squarrosum      TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Eriosema umtamvunense    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Flemingia stricta        TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Flemingia parviflora     TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Flemingia macrophylla    TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Flemingia strobilifera   TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Flemingia glutinosa      TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Flemingia trifoliastrum  TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Flemingia lineata        TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Glycine microphylla      TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Glycine stenophita       TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Glycine pullenii         TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Glycine tomentella       TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Glycine max              TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG
Vigna luteola            TGAAGAACGT AGCGAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG

[                                 360        370        380        390        400]
Cajanus cajan            AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Cajanus cajanifolius     AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Cajanus scarabaeoides    AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Cajanus cinereus-1       AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus reticulatus      AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus lanceolatus      AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus cinereus-2       AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus acutifolius-2    AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus crassicaulis     AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus latisepalus-1    AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus cinereus-1       AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus latisepalus-2    AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus pubescens        AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus acutifolius-1    AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus acutifolius-3    AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
```

```
Cajanus viscidus          AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus hirtopilosus      AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTTAGG
Cajanus lineatus          AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Cajanus albicans          AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Cajanus marmoratus        AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Rhynchosia aurea          AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Cajanus mollis            AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGTTGAGG
Dunbaria ferruginea       AACCATCGAG TTTTTGAACG CAAGTTGCGC CTAAAGCCAT TAGGCTGAGG
Rhynchosia hauthalii      AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Rhynchosia verdcourtii    AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGCTGAGG
Rhynchosia australis      AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGCTGAGG
Rhynchosia orthobotryasa  AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCTGAGG
Bolusafra bituminosa      AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema griseum          AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema simulans         AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema burkei           AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema lucipetum        AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema kraussianum      AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema salignum         AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema rossii           AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema squarrosum       AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Eriosema umtamvunense     AACCATCGAG TCTTTGAACG CAAGTTGCGC CTGAAGCCAT TAGGTTGAGG
Flemingia stricta         AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAC TAGGCCGAGG
Flemingia parviflora      AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAC TAGGCCGAGG
Flemingia macrophylla     AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAC TAGGCCGAGG
Flemingia strobilifera    AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCGC TAGGCCGAGG
Flemingia glutinosa       AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Flemingia trifoliastrum   AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Flemingia lineata         AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Glycine microphylla       AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Glycine stenophita        AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Glycine pullenii          AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Glycine tomentella        AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Glycine max               AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGCCGAGG
Vigna luteola             AACCATCGAG TCTTTGAACG CAAGTTGCGC CCGAAGCCAT TAGGTTGAGG

[                                  410        420        430        440        450]
Cajanus cajan             GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTTG
Cajanus cajanifolius      GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCA
Cajanus scarabaeoides     GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus cinereus-1        GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus reticulatus       GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus lanceolatus       GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus cinereus-2        GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus acutifolius-2     GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus crassicaulis      GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus latisepalus-1     GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus cinereus-1        GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus latisepalus-2     GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus pubescens         GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus acutifolius-1     GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus acutifolius-3     GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCG
Cajanus viscidus          GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCA
Cajanus hirtopilosus      GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCA
Cajanus lineatus          GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTTG
Cajanus albicans          GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGAGC --ATGTCTCG
Cajanus marmoratus        GCACGCCTGC CTGGGTGTCA CAAAACGTTA CCCCTTGTGC --ATGTCTTG
Rhynchosia aurea          GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTTG
Cajanus mollis            GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCA
Dunbaria ferruginea       GCACGCCTGC CTGGGTGTCA CAAATCGTTA CCCCTTGTGC --ATGTCTCA
Rhynchosia hauthalii      GCACGCCTGC CTGGGTGTCA CACATCGTTA CCCCTCATGC ACATGTCTCA
Rhynchosia verdcourtii    GCACGCCTGC CTGGGTGTCA CACATCGTTA CCTCTTATGC ACATGTCTCG
Rhynchosia australis      GCACGCCTGC CTGGGTGTCA CACATCGTTA CCTCTTATGC ACATGTCTCG
Rhynchosia orthobotryasa  GCACGCCTGC CTGGGTGTCA CACATCGTTA CCCCCTATGC ATATGTCTTG
Bolusafra bituminosa      GCACGCCTGC CTGGGTGTCA CACGTCGTTA CCCCTTATGC ATATGCCTAA
Eriosema griseum          GCACGCCTGC CTGGGTGTCA CATATCGTTA CCCATTATGC ATATGTCTTA
Eriosema simulans         GCACGCCTGC CTGGGTGTCA CATGTCGTTA CCCATTATGC ATACGACTTA
Eriosema burkei           GCACGCCTGC CTGGGTGTCA CATGTCGTTA CCCATTATGC ATACGACTTA
Eriosema lucipetum        GCACGCCTGC CTGGGTGTCA CATATCGTTA CCCATTATGC ATACGACTTA
Eriosema kraussianum      GCACGCCTGC CTGGGTGTCA CATATCGTTA CCCATTATGC ATACGACTTA
Eriosema salignum         GCACGCCTGC CTGGGTGTCA CATATCGTTA CCCATTATGC ATACGACTTA
Eriosema rossii           GCACGCCTGC CTGGGTGTCA CATATCGTTA CCCATTATGC ATACGACTTA
Eriosema squarrosum       GCACGCCTGC CTGGGTGTCA CATATCGTTA CCCATTATGC ATACGACTTA
Eriosema umtamvunense     GCACGCCTGC CTGGGTGTCA CATATCGTTA CCCATTATGC ATACGACTTA
Flemingia stricta         GCACGCCTGC CTGGGTGTCA CGCGTCGTTA CCCCCCA-GC A-AC------
Flemingia parviflora      GCACGCCTGC CTGGGTGTCA CGCGTCGTTA CCCCCCA-GC A-GC------
Flemingia macrophylla     GCACGGCTGC CTGGGTGTCA CGCGTCGTTA CCCCCCA-GC A-ACG-----
```

```
Flemingia strobilifera   GCACGCCTGC CTGGGTGTCA CGCGTCGTTA CCCCCCA-GC A-TCG-----
Flemingia glutinosa      GCACGCCTGC CTGGGTGTCA CGCGCCGTTA CCCCC----C A-GC-A----
Flemingia trifoliastrum  GCACGCCTGC CTGGGTGTCA CGCGCCGTTA CCCCC--T-C A-GC-A----
Flemingia lineata        GCACGCCTGC CTGGGTGTCA CGCGCCGTTA CCCCC--T-C A-GC-A----
Glycine microphylla      GCACGCCTGC CTGGGTGTCA CACATCGTTA CCCCCAACGC --------TA
Glycine stenophita       GCACGCCTGC CTGGGTGTCA CACATCGTTA CCCCCAACGC --------TA
Glycine pullenii         GCACGCCTGC CTGGGTGTCA CACATCGTTA CTCCCAACAC --------TA
Glycine tomentella       GCACGCCTGC CTGGGTGTCA CACATCGTTA CCCCCAACGC --------TA
Glycine max              GCACGCCTGC CTGGGTGTCA CACATCGTTT CCCCAA-CGC AAAC-ATGTA
Vigna luteola            GCACGCCTGC CTGGGTGTCA CATGTGGTCA CCACTAATGC AA--------

[                                460       470        480        490        500]
Cajanus cajan            TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus cajanifolius     TTTTAGATAT TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus scarabaeoides    TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus cinereus-1       TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus reticulatus      TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus lanceolatus      TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus cinereus-2       TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus acutifolius-2    TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus crassicaulis     TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus latisepalus-1    TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus cinereus-1       TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus latisepalus-2    TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus pubescens        TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus acutifolius-1    TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus acutifolius-3    TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus viscidus         TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus hirtopilosus     TTTTAGATAC TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus lineatus         TTTTAGATAT TG----TGTT GGGGGTGCAT GATGACTTCC CATGAGCA-T
Cajanus albicans         TTTTAGAAAC TG----TGTT GGGGGTGCAT GATGGCTTCC CATGAGCA-T
Cajanus marmoratus       TTTTGGATAC TG----TGCT -GGGGTGCAT GATGACTTCC CACGAGCA-T
Rhynchosia aurea         TTTTGGATAC TG----TGCT -GGGGTGCAT GATGACTTCC CACGAGCA-T
Cajanus mollis           TTTTAGACAG TG----TGTT -GGGGTGCAT GATGGCTTCC CACGAGCA-A
Dunbaria ferruginea      TTTTAGACAG TGCCACTGTT -GGGGTGCAT GATGACTTCC CACGAGCA-A
Rhynchosia hauthalii     TGTTA-ATAG TG----TGTT -GGGGTGCAT GATGACTTCC CGCGAGTG-T
Rhynchosia verdcourtii   TGCTAGATAC TG----TGTT -GGGGTGCAT GATGGCTTCC CGTGAGTG-T
Rhynchosia australis     TGCTAGATAC TG----TGTT -GGGGTGCAT GATGGCTTCC CGTGAGTG-T
Rhynchosia orthobotryasa TGTTAGATAA TG----TGTT -TGGGTGCAT GATGACTTCC CACAAGCG-T
Bolusafra bituminosa     TGCTAGACAA TG----TGTA -GGGGTGCAT GATGACTTCC CACGAGCG-T
Eriosema griseum         TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema simulans        TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema burkei          TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema lucipetum       TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema kraussianum     TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema salignum        TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema rossii          TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema squarrosum      TATTAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Eriosema umtamvunense    TAATAGACAA AG----TGTT -GGGGTGCAT GATGACTTCC CATGAGCA-T
Flemingia stricta        ---TAGACGC TG----CGT- -GGGGTGCAT GATGGCCTCC CGCGAGCG-T
Flemingia parviflora     ---TAGACGC TG----CGT- -GGGGTGCAT GATGGCCTCC CGCGAGCG-T
Flemingia macrophylla    ----AGACGT TG----CGT- -GGGGTGCAT GATGGCCTCC CGCGAGCG-T
Flemingia strobilifera   ----ATACGT TG----TGT- -GGGGTGCAT GATGGCCTCC CGCGAGCG-T
Flemingia glutinosa      ----AGAAGC TG----TGT- -AGGGTGCAC GATGGCCTCC TGCGAGCG-T
Flemingia trifoliastrum  ----AGACGC TG----CGT- -GGGGTGCAC GATGGCCTCC CGCGAGCG-T
Flemingia lineata        ----AGACGC TG----CGT- -GGGGTGCAC GATGGCCTCC CGCGAGCG-T
Glycine microphylla      ACCTAG-TGT TG----CGC- -GGGGTGTAT GCTGACCTCC CGCGAGCACC
Glycine stenophita       ACCTAG-TGT TG----CGC- -GGGGTGTAT GCTGACCTCC CGCGAGCACC
Glycine pullenii         ACCTAG-TGT TG----CGT- -GGGGTGTAT GCTGACCTCC CGCGAGCACC
Glycine tomentella       ACCTAG-TGT TG----CGC- -GGGGTGTAT GCTGACCTCC CGCGAGCACC
Glycine max              ACAATGTTGC TG----CGC- -GGGGTGTAT GCTGACCTCC CGCGAGCACC
Vigna luteola            ------TTG- TG----CATT -GAGGTGTAA GTTGGCTTCC CTCGAGAA-A

[                                510       520        530        540        550]
Cajanus cajan            TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus cajanifolius     TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus scarabaeoides    TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus cinereus-1       TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus reticulatus      TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus lanceolatus      TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus cinereus-2       TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus acutifolius-2    TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus crassicaulis     TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus latisepalus-1    TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus cinereus-1       TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus latisepalus-2    TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus pubescens        TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
```

```
Cajanus acutifolius-1     TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus acutifolius-3     TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus viscidus          TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus hirtopilosus      TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus lineatus          TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus albicans          TGTCTCGTGG TTGGTTG-AA AATTGGGTTC ACGGTTGAGC GTGCCATGGT
Cajanus marmoratus        TGTCTCGTGG TTGGTTG-AA ATTTGAGTTC AGGGTTGAGC GTGCCATGGT
Rhynchosia aurea          CGTCTCGTGG TTGGTTG-AA AATTGAGTTC AGGGTTGAGC GTGCCATGGT
Cajanus mollis            TGTCTCGTGG TTGGTTG-AA AATAGAGTTC A-GGGCGAGC GTGCCATGGT
Dunbaria ferruginea       TGTCTCGTGG TTGGTTG-AA AATAGAGTTC A-GGGCGAGC GTGCCATGGT
Rhynchosia hauthalii      TGCCTCGTGG TTGGTTG-AA ATTTGAGTTC AAGGTCAATT GTGCCATGGT
Rhynchosia verdcourtii    TGCCTCGCGG TTGGTTG-AA AATCGAGTTC ACGGTCGAGC GTGCCATGGT
Rhynchosia australis      TGCCTCGCGG TTGGTTG-AA AATCGAGTTC ACGGTCGAAC GTGCCATGGT
Rhynchosia orthobotryasa  TGTCTTGTGG TTGGTTG-AA AAATGAGTTC ACGGTCGAGT GTGCCATGGT
Bolusafra bituminosa      CGTCTCGTGG TTGGTTG-AA AATTGAGTTC ACGGTCGAGC GTGCCATGAT
Eriosema griseum          TGTCTTGTGG TTGGTTGAAA AGTTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema simulans         TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema burkei           TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema lucipetum        TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema kraussianum      TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema salignum         TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema rossii           TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema squarrosum       TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Eriosema umtamvunense     TGTCTCGTGG TTGGTTGAAA AATTGTGTTC ATGGTCAAGT GTGCCATGGT
Flemingia stricta         TGTCTCGTGG TTGGCTG-AA AACTGAGTTC ACGGCGGAGC GCGCGGTGGC
Flemingia parviflora      TGTCTCGTGG TTGGCTG-AA AACTGAGTTC ACGGCCGAGC GCGCGGTGGC
Flemingia macrophylla     TGTCTCTTGG TTGGTTG-AA AACTGAGTTC ACGGCCGAGC GTGCGGTGGC
Flemingia strobilifera    TGTCTTGTGG TTGGTTG-AA AACTGAGTTC ACGGCCAAGC GTGCGGTGGC
Flemingia glutinosa       TGCCTTGTGG TTGGCTG-AA AACTGAGTTC ATGGCCGAGT GTGCTGTGGT
Flemingia trifoliastrum   TGCCTCGTGG TTGGCTG-AA AACTGAGTTC GCGGCGGAGC GTGCTGTGGT
Flemingia lineata         TGCCTCGTGG TTGGCTG-AA AACTGAGTTC GCGGCGGAGC GTGCTGTGGT
Glycine microphylla       TGTCTCGTGG TTGGTTG-AA ATCTGGGTTC ATGGCCGACT TCGCCGTGAC
Glycine stenophita        CGTCTCGTGG TTGGTTA-AA ATCTGGGTTC ATGGCCGACT TCGCCGTGAC
Glycine pullenii          CGTCTCGTGG TTGGTTG-AA ATCTGGGTTC ATGGCCGACT TCGCCGTGAC
Glycine tomentella        CGTCTCGTGG TTGGTTG-AA ATCTGGGTTC ATGGCCGACT TCGCCGTGAC
Glycine max               CGCCTCGTGG TTGGTTG-AA ATCTGGGTTC ATGGCCGACT TCGCCGTGAT
Vigna luteola             GTCCTCGTGG TTGGTTG-AA AACTCAGTTA ACGATACAGT CACCTGCGAT

[                                    560        570        580        590        600]
Cajanus cajan             ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus cajanifolius      ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus scarabaeoides     ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus cinereus-1        ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus reticulatus       ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus lanceolatus       ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus cinereus-2        ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus acutifolius-2     ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus crassicaulis      ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus latisepalus-1     ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus cinereus-1        ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus latisepalus-2     ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus pubescens         ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus acutifolius-1     ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus acutifolius-3     ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus viscidus          ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus hirtopilosus      ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus lineatus          ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTCTTT
Cajanus albicans          ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCGT---GT GTGCGTCTTT
Cajanus marmoratus        AAAATGGTGG ATGAGTAAGG CTCGAGACCA ATCAT---GT GTGCGTCTCT
Rhynchosia aurea          AAAATGGTGG ATGAGTAAGG CTCGAGACCA ACCAT---GC GTGCGTCTCT
Cajanus mollis            AAAATGGTGG ATGAGTAATG CTCGATACCA ATCAT---GT GTGCGTCTCT
Dunbaria ferruginea       AAAATGGTGG ATGAGCAATG CTCGATACCG ATCAT---GT GTGCGTCTCT
Rhynchosia hauthalii      ATAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GTGCGTGTTT
Rhynchosia verdcourtii    AAAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GCGCGTCTCA
Rhynchosia australis      AAAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GCGCGTCTTA
Rhynchosia orthobotryasa  AAAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GCACGTCTTT
Bolusafra bituminosa      AAAATGGTGG ATGAGTAATG CTCGAGACCA ATCAT---GT GCGCGTCTTT
Eriosema griseum          AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTTCATCTTT
Eriosema simulans         AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTGCATCTTT
Eriosema burkei           AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTGCATCTTT
Eriosema lucipetum        AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTGCATCTTT
Eriosema kraussianum      AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTGCATCTTT
Eriosema salignum         AAAATGGTGG AAGAGTGATC CTCGGGACCA ACCAT---GT GTGCATCTTT
Eriosema rossii           AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTGCATCCTT
Eriosema squarrosum       AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTGCATCCTT
Eriosema umtamvunense     AAAATGGTGG AAGAGTGATC CTCGAGACCA ACCAT---GT GTGCATCCTT
Flemingia stricta         CAAACGGTGG ATGAGTAATG CTCGATGCCG GCCAC---GT GCACGTCTGT
```

```
Flemingia parviflora      CAAACGGTGG ATGAGTAATG CTCGATGCCG GCCAC---GT GCACTTCTGT
Flemingia macrophylla     CAAACGGTGG ATGAGTAATG CTCGATGCCG GCCAC---GT GCACCTCTGT
Flemingia strobilifera    CAAACGGTGG ATGAGTAATG CTCGATGCCG GCCAC---GT GCACGTATGT
Flemingia glutinosa       CAAACGGTGG ATGAGCAAAG CTCGATGCCG GCCAC---GT GCACGTTCGT
Flemingia trifoliastrum   CAAACGGTGG ATGAGCAAAG CTCGATGCCG GCCAC---GT GCACGTCTGT
Flemingia lineata         CAAACGGTGG ATGAGCAAAG CTCGATGCCG GCCAC---GT GCACGTCTGT
Glycine microphylla       AAAATGGTGG ATGAGCAACG CTCGAGACCA ATCAC---GT GCGAGCCGGT
Glycine stenophita        AAAATGGTGG ATGAGCAACG CTCGAGACCA ATCAC---GT GCGAGCCGGT
Glycine pullenii          AAAATGGTGG ATGAGCAACG CTCGAGACCA ATCAC---GT GCGAGTCGGT
Glycine tomentella        AAAATGGTGG ATGAGCAACG CTCGAGACCA GTCAC---GC GCGAGCCGGT
Glycine max               AAAATGGTGG ATGAGCCACG CTCGAGACCA ATCAC---GT GCGAGCCGGT
Vigna luteola             -TAATGGTGG ACGAGAAACT CTCGAGACCA GTCATTGCTT GTGGTACTCT

[                                    610         620        630         640         650]
Cajanus cajan             G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus cajanifolius      G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus scarabaeoides     G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus cinereus-1        G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus reticulatus       G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus lanceolatus       G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus cinereus-2        G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus acutifolius-2     G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus crassicaulis      G-CAGGTTTG GACTC-T--- TGGCCCAAT- ---TTGT-CT TG--TATTTT
Cajanus latisepalus-1     G-CAGGTTTG GACTC-T--- TGGCCCAAT- ---TTGT-CT TG--TATTTT
Cajanus cinereus-1        G-CAGGTTTG GACTC-T--- TGGCCCAAT- ---TTGT-CT TG--TATTTT
Cajanus latisepalus-2     G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus pubescens         G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus acutifolius-1     G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus acutifolius-3     G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus viscidus          G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus hirtopilosus      G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus lineatus          G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus albicans          G-CAGGTTTG GACTC-T--- TGACCCAAT- ---TTGT-CT TG--TATTTT
Cajanus marmoratus        G-TAGGCTTG GACTC-T--- TGACCCAAT- ---TTGT-CT CG--TATTAT
Rhynchosia aurea          G-TAGGCTTG GACTC-T--- TGACCCAAT- ---TTGT-CT CG--TATTAT
Cajanus mollis            A-CAAGTTTG GACTT-T--- TGGCCCAAT- ---TTGT-GT TG--TATTTT
Dunbaria ferruginea       A-CAAGTTTG GACTT-T--- TGACCCAAT- ----TGT-CT TG--TATCTT
Rhynchosia hauthalii      GACTGTTGTA GACCC-T--- TAACCTAGT- ---GCGT-CT CA--TAT-GT
Rhynchosia verdcourtii    GACTGGTTTG GACCT----- TGACCCGTT- ---GCGT-CT TC--TAT-GT
Rhynchosia australis      GACTGGTTTG GACCC-T--- TGACCCGGT- ---GCGT-CT TC--TAT-GT
Rhynchosia orthobotryasa  GGCAAGTTTG GACCC-T--- TGACCCAAT- ---GCGT-CT GGTTTAT-TT
Bolusafra bituminosa      GGCCGGTTTG GACCCTT--- TGACCCAAT- ---GCGT-CC GG--TAT-CT
Eriosema griseum          GGCAATTATG GACCC-T--- TGACCCTAT- ---GTGT-TT CT--CAT-AT
Eriosema simulans         GGCAATTATG GACCC-T--- TGACCCTAT- ---GTGT-TT GT--TAT-AT
Eriosema burkei           GGCAATTATG GACCC-T--- TGACCCTAT- ---GTGT-TT GT--TAT-AT
Eriosema lucipetum        GGCAATTATG GACCC-T--- TGACCCTAT- ---GTGT-TT GT--TAT-AT
Eriosema kraussianum      GGCAATTATG GACCC-T--- TGACCCTAT- ---GTGT-TT GT--TAT-AT
Eriosema salignum         GGCAATTATG GACCC-T--- TGACCCTAT- ---GTGT-TT GT--TAT-AT
Eriosema rossii           GGCAATTATG GACCC-T--- TGGCCCTAT- ---GTGT-TT GT--TAT-AT
Eriosema squarrosum       GGCAATTATG GACCC-T--- TGACCCTAT- ---ATGT-TT GT--TAT-AT
Eriosema umtamvunense     GGCAATTATG GACCC-T--- TGACCCTAT- ---GTGT-TT GT--TAT-AT
Flemingia stricta         GTCGGATTTG GACCC-T--- TGACCTCGT- ---GCGT-CT TT--TAC-TT
Flemingia parviflora      GTCGGATTTG GACCC-T--- TGACCTCGT- ---ACGT-CT TC--TAC-TT
Flemingia macrophylla     GTCGGATTTG GACCC-T--- TGACCTCGT- ---GCGT-CT TT--TAC-TT
Flemingia strobilifera    GCCGGATATG GACCC-T--- TGACCTTGC- ---GCGT-CT TT--TAC-TT
Flemingia glutinosa       GCCGTATTTG GACCC-T--- GGACCTCGT- ---GCGTGAT TT--TAC-AC
Flemingia trifoliastrum   GCCGGATTTG GACCC-T--- CGACCTCGT- ---GCGTGAT TT--TAC-TC
Flemingia lineata         GCCGGATTTG GACCC-T--- CGACCTCGT- ---GCGTGAT TT--TAC-TC
Glycine microphylla       --CAGTTATG GACTCACAGA CGACCCTGT- ---GCGTG-- ------C-AC
Glycine stenophita        --CAGTTATG GACTCACGGA CGACCCTAT- ---GCGTG-- ------C-AC
Glycine pullenii          --CAGTTATG GACTCACGGA CGACCCTGT- ---GCGTG-- ------C-AC
Glycine tomentella        --CAGTTACG GACTCACGGA CGACCCTAT- ---GCGTG-- ------C-AC
Glycine max               --CAGTTCTG GACCCATCGA CGACCCTTT- ---GCGTG-- ------C-AC
Vigna luteola             GGCCGTTCTA AC--TGATTG ACTCTACTGT GTCGTTTG-C AGATAAC-AA

[                                   660]
Cajanus cajan             TGACATCTCT AA
Cajanus cajanifolius      TGACATCTCT AA
Cajanus scarabaeoides     TGACATCTCT AA
Cajanus cinereus-1        TGACATCTCT AA
Cajanus reticulatus       TGACATCTCT AA
Cajanus lanceolatus       TGACATCTCT AA
Cajanus cinereus-2        TGACATCTCT AA
Cajanus acutifolius-2     TGACATCTCT AA
Cajanus crassicaulis      TGACATCTCT AA
Cajanus latisepalus-1     TGACATCTCT AA
Cajanus cinereus-1        TGACATCTCT AA
```

```
Cajanus latisepalus-2     TGACATCTCT AA
Cajanus pubescens         TGACATCTCT AA
Cajanus acutifolius-1     TGACATCTCT AA
Cajanus acutifolius-3     TGACATCTCT AA
Cajanus viscidus          TGACATCTCT AA
Cajanus hirtopilosus      TGACATCTCT AA
Cajanus lineatus          TGACATCTCT AA
Cajanus albicans          TGACATCTCT AA
Cajanus marmoratus        TGACATATCT AA
Rhynchosia aurea          TGACATATCT AA
Cajanus mollis            TGACATCTCT AA
Dunbaria ferruginea       TGACATCTCT AA
Rhynchosia hauthalii      GGACGTCTT- AA
Rhynchosia verdcourtii    GGACGTCTTT AA
Rhynchosia australis      GGACGTCTTT AA
Rhynchosia orthobotryasa  GGACGCTTTT AA
Bolusafra bituminosa      GGACGCCTT- AA
Eriosema griseum          CGAAGCCTTT AG
Eriosema simulans         CAATGCCTTT AA
Eriosema burkei           CAATGCCTTT AA
Eriosema lucipetum        CAATGCCTTT AA
Eriosema kraussianum      CAATGCCTTT AA
Eriosema salignum         CAATGCCTTT AA
Eriosema rossii           CAATGCCTTT AA
Eriosema squarrosum       CAATGCCTTT AA
Eriosema umtamvunense     CAATGCCTTT AA
Flemingia stricta         GGACGCCTCT AA
Flemingia parviflora      GGACGCCTCT AA
Flemingia macrophylla     GGACGCCTCT AA
Flemingia strobilifera    GGATGCCTCT AA
Flemingia glutinosa       GGATGCCTCT AA
Flemingia trifoliastrum   GGACGCCTCT AA
Flemingia lineata         GGACGCCTCT AA
Glycine microphylla       GCACGCTCCC AA
Glycine stenophita        GCACGCTCCC AA
Glycine pullenii          GCACGCTCCC AA
Glycine tomentella        GCACGCTCCC AA
Glycine max               GCACGCTCCC AA
Vigna luteola             TGAGGCACTT AC
```

Appendix 2: Sequence alignments of the trnL-F spacer.

```
[                              10         20         30         40         50]
Cajanus reticulatus-1  ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus lineatus       ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus crassicaulis-1 ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus crassicaulis-2 ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus scarabaeoides  ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus pubescens      ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus viscidus       ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus mollis         ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus acutifolius    ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus albicans       ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus cinereus-1     ATCCCCAACC AAAGGGCCTC TTT-ACTTTC AAATTCTTTT TCCTATATCC
Cajanus cinereus-2     ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus reticulatus-2  ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus sericeus       ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus cajanifolius   ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Cajanus cajan          ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTC-TTT TCCTATATCC
Cajanus latisepalus    ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Erisoma burkei         ATCCCCAACC AAAGGGCCTG TTTCACTTTC GAATTCTTTT TCCTATATCC
Erisoma lucipetum      ATCCCCAACC AAAGGGCCTG TTTCACTTTC GAATTCTTTT TCCTATATCC
Erisoma umtamvunense   ATCCCCAACC AAAGGGCCTG TTTCACTTTC GAATTCTTTT TCCTATATCC
Erisoma salignum       ATCCCCAACC AAAGGGCCTG TTTCACTTTC GAATTCTTTT TCCTATATCC
Erisoma simulans       TTCCCCAACC AAAGGGCCTG TTTCACTTTC GAATTCTTTT TCCTATATCC
Erisoma griseum        ATCCCCAACC AAAGGGCCTC TTTCACTTTC GAATTCTTTT TCCTATATCC
Erisoma psoraleoides   ATCCCCAACC AAAGGGCCTC TTTCACTTTC GAATTCTTTT TCCTATATCC
Erisoma rossii         ATCCCCAACC AAAGGGCCTG TTTCACTTTC GAATTCTTTT TCCTATATCC
Rhynchosia australis   ATCCCCAACC AAAGGGCCTG TTTCATTTTC GAATTCTTTT TCCTATATCC
Rhynchosia bungarensis ATCCCCAACC AAAGGGCCTG TTTCATTTTC GAATTCTTTT TCCTATATCC
Rhynchosia minima      ATCCCCAACC AAAGGGCCTG TTTCATTTTC GAATTCTTTT TCCTATATCC
Rhynchosia aurea       ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Rhynchosia sp          ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Flemingia parviflora   ATCCCCAACC AAAGGGCCTG TTTAATTTTC GAATTCTTTT TCCTATATCC
Flemingia stricta      ATCCCCAACC AAAGGGCCTG TTTAACTTTC GAATTCTTTT TCCTATATCC
Flemingia strobilifera CTCCCCAACC AAGGGGCGGG CATA--CTTT CGATTCTTTT TCCTATATCC
Bolusafra bituminosa   ATCCCCAACC GAAGGGCCTG TTTCACTTTC TAATTCTTTT TCCTATATCC
Dunbaria ferruginea    ATCCCCAACC AAAGGGCCTC TTTAACTTTC AAATTCTTTT TCCTATATCC
Glycine microphylla    ATCC----CC AAAATGCCTG TTTAACTTTC TAA-----TT TCTCATACCC

[                            60         70        80             90         100]
Cajanus reticulatus-1  TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus lineatus       TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus crassicaulis-1 TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus crassicaulis-2 TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus scarabaeoides  TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus pubescens      TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus viscidus       TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus mollis         TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus actifolius     TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus albicans       TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus cinereus-1     TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus cinereus-2     TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus reticulatus-2  TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus sericeus       TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus cajanifolius   TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus cajan          TCTCTGTC-T TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Cajanus latisepalus    TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Erisoma buurkei        TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Erisoma lucipetum      TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Erisoma umtamvunense   TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Erisoma salignum       TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Erisoma simulans       TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Erisoma griseum        TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Erisoma psooraleoides  TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Erisoma rossii         TCTCTGTCTT TAAGTCGTTC TTTTTTT--- ---GTTT--- --------TA
Rhynchosia australis   TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---ATTT--- --------TA
Rhynchosia bungarensis TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Rhynchosia minima      TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Rhynchosia aurea       TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Rhynchosia sp          TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Flemingia parviflora   TCTCTGTCTT GAAGTCGTTA TTTATGTGTT TTAGTTT--- --------TA
Flemingia stricta      TCTCTGTCTT GAAGTCGTTA TTTATGTGTT TTAGTTT--- --------TA
Flemingia stobilifera  TCTCTGTCTT GAAGTCGTTA TTTATGTGTT TTAGTTT--- --------TA
Bolusafra bituminosa   TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
```

```
Dunbaria ferruginea    TCTCTGTCTT TAAGTCGTTA TTTATGT--- ---GTTT--- --------TA
Glycine microphylla    TCTCTATCTT TAAGTCGTTA TTTATGT--- ---GCTTTAT TCAGTTTATA

[                              110        120        130        140        150]
Cajanus reticulatus-1  TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTATT-----
Cajanus lineatus       TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Cajanus crassicaulis-1 TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Cajanus crassicaulis-2 TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Cajanus scarabaeoides  TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Cajanus pubescens      TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Cajanus viscidus       TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Cajanus mollis         TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTATT-----
Cajanus actifolius     TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Cajanus albicans       TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Cajanus cinereus-1     TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Cajanus cinereus-2     TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Cajanus reticulatus-2  TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Cajanus sericeus       TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Cajanus cajanifolius   TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Cajanus cajan          TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Cajanus latisepalus    TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma buurkei        TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma lucipetum      TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma umtamvunense   TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma salignum       TTCCAATTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma simulans       TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma griseum        TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma psooraleoides  TTCCATTTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Erisoma rossii         TTCCAATTAT TCT------T TCACAAAGAA ATTGTAATTT TTCTT-----
Rhynchosia australis   TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Rhynchosia bungarensis TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Rhynchosia minima      TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Rhynchosia aurea       TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Rhynchosia sp          TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Flemingia parviflora   TTCAATTTAT TCTTTCACAT TCACAAATAA ATTGGAATTT TTCTT-----
Flemingia stricta      TTCAATTTAT TCTTTCACAT TCACAAATAA ATTGGAATTT TTCTT-----
Flemingia stobilifera  TTCAATTTAT TCTTTCACAT TCAAAAATAA ATTGGAATTT TTCTT-----
Bolusafra bituminosa   TTCCTTTTAT TCT------T TCACAAAGAA ATTGGAATTT TTCTT-----
Dunbaria ferruginea    TTCCATTTAT TCT------T TCACAAAGAA ATTGGAATTT TTATT-----
Glycine microphylla    TTCAGTTTAT TCT------T TCACAACTAA ATTGGAATTT TTCTTTTTAT

[                             160        170        180        190        200]
Cajanus reticulatus-1  ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus lineatus       ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus crassicaulis-1 ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus crassicaulis-2 ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus scarabaeoides  ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus pubescens      ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus viscidus       ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus mollis         ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus actifolius     ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus albicans       ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus cinereus-1     ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus cinereus-2     ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus reticulatus-2  ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus sericeus       ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus cajanifolius   ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus cajan          ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Cajanus latisepalus    ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Erisoma buurkei        ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Erisoma lucipetum      ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Erisoma umtamvunense   ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Erisoma salignum       ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Erisoma simulans       ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Erisoma griseum        ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Erisoma psooraleoides  ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Erisoma rossii         ---------T ------TCAC AATTACGAGT CACAAGTTTT GGAATATATA
Rhynchosia australis   ---------T TTCTTATCAT AATTACGAGT CACAAGTTTT GGAATATATA
Rhynchosia bungarensis ---------T TTCTTATCAT AATTACGAGT CACAAGTTTT GGAATATATA
Rhynchosia minima      ---------T TTCTTATCAT AATTACGAGT CACAAGTTTT GGAATATATA
Rhynchosia aurea       ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Rhynchosia sp          ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
Flemingia parviflora   ---------T TTCTTATCAG AATTGCGAGT CACAAGTTTT GGAATATATA
Flemingia stricta      ---------T TTCTTATCAG AATTGCGAGT CACAAGTTTT GGAATATATA
Flemingia stobilifera  ---------T TTCTTATCAG AATAGCGAGT CACAAGTTTT GGAATATATA
Bolusafra bituminosa   ---------T ------TCAT AATTACGAGT CACAAGTTTT GGAATATATA
Dunbaria ferruginea    ---------T TTCTTATCAT AATTAGGAGT CACAAGTTTT GGAATATATA
```

```
Glycine microphylla    TTTCAAAAAT TTCTTATCAT AATTACAAGT CACAAGTTTT GAAATATATA

[                              210        220        230        240        250]
Cajanus reticulatus-1  GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus lineatus       GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus crassicaulis-1 GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus crassicaulis-2 GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus scarabaeoides  GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus pubescens      GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus viscidus       GAGCATCTTT GAAATATGTA ATTATTTGGG TGAAACA-CG TATAA-TTTT
Cajanus mollis         GAACATCTTT GAAATATTTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus actifolius     GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus albicans       GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus cinereus-1     GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus cinereus-2     GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus reticulatus-2  GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus sericeus       GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus cajanifolius   GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus cajan          GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Cajanus latisepalus    GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Erisoma buurkei        GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Erisoma lucipetum      GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Erisoma umtamvunense   GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Erisoma salignum       GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Erisoma simulans       GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Erisoma griseum        GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Erisoma psooraleoides  GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Erisoma rossii         GAACATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Rhynchosia australis   GAGCATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Rhynchosia bungarensis GAGCATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Rhynchosia minima      GAGCATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Rhynchosia aurea       GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Rhynchosia sp          GAGCATCTTT GAAATATGTA ATTATTTGTG TGAAACA-CG TATAA-TTTT
Flemingia parviflora   GAGCATCTTT TGAATATTGA ATTATTTGTG TGAAACA-CA TATAA-TTTT
Flemingia stricta      GAGCATCTTT TGAATATTGA ATTATTTGTG TGAAACA-CA TATAA-TTTT
Flemingia stobilifera  GAGCATCTTT TGAATATTGA ATTATTTGTG TGAAACA-CA TATAA-TTTT
Bolusafra bituminosa   GAGCATCTTT GGAATATTTA ATTATTTGTG TGAAACA-CG TATCA-TTTT
Dunbaria ferruginea    GAACATCTTT GAAATATGTA ATTATTTGTG TGAAACACCG TATAA-TTTT
Glycine microphylla    ---------- ---------- --------TG TGAAACA-CA TAGAATTTTT

[                              260        270        280        290        300]
Cajanus reticulatus-1  TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus lineatus       TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus crassicaulis-1 TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus crassicaulis-2 TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus scarabaeoides  TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus pubescens      TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus viscidus       TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus mollis         TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus actifolius     TTTTATGATA AT-AAAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus albicans       TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus cinereus-1     TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus cinereus-2     TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus reticulatus-2  TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus sericeus        TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus cajanifolius   TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus cajan          TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Cajanus latisepalus    TTTTATGATA ATAAAAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Erisoma buurkei        TTTTATGATA AC--CAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Erisoma lucipetum      TTTTATGATA AC--CAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Erisoma umtamvunense   TTTTATGATA AC--CAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Erisoma salignum       TTTTATGATA AC--CAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Erisoma simulans       TTTTATGATA AC--CAAAAA AATGAATATC TAATATCTTA TTTTTGAGCA
Erisoma griseum        TTTTATGATA AC--CAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Erisoma psooraleoides  TTTTATGATA AC--CCAAAA AATG------ -AATATCTTA TTTTTGAGCA
Erisoma rossii         TTTTATGATA AC--CAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Rhynchosia australis   TTTGATGATA AC--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Rhynchosia bungarensis TTTGATGATA AC--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Rhynchosia minima      TTTGATGATA AC--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Rhynchosia aurea       TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Rhynchosia sp          TTTTATGATA AT--AAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Flemingia parviflora   TTTTATGATA AC--CCAAAA AATG------ -AATATCTTA TTTTGGAGCA
Flemingia stricta      TTTTATGATA AC--CCAAAA AATG------ -AATATCTTA TTTTGGAGCA
Flemingia stobilifera  TTTTATGATA AC--CGAAAA AAAG------ -AATATCTTA TTTTGGAGCA
Bolusafra bituminosa   TTTTATGATA AC--CAAAAA AATG------ -AATATCTTA TTTTTGAGCA
Dunbaria ferruginea    TTTTATGATA AT--AAAAAA AATG------ -AATATCTTT TTTTTGAGCA
Glycine microphylla    TTTTATGATA AA--CGTACA AATG------ -AAGATTTTC TTTTTGAGCA
```

```
[                                    310        320        330        340         350]
Cajanus reticulatus-1                AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus lineatus                     AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTACTGA
Cajanus crassicaulis-1               AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus crassicaulis-2               AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus scarabaeoides                AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus pubescens                    AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus viscidus                     AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus mollis                       AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus actifolius                   AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus albicans                     AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus cinereus-1                   AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus cinereus-2                   AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus reticulatus-2                AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus sericeus                     AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus cajanifolius                 AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus cajan                        AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Cajanus latisepalus                  AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Erisoma buurkei                      AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Erisoma lucipetum                    AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Erisoma umtamvunense                 AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Erisoma salignum                     AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Erisoma simulans                     AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Erisoma griseum                      AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Erisoma psooraleoides                AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Erisoma rossii                       AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Rhynchosia australis                 AGGAATCCTC ATATGCATGA TTAATGATAC AAAATAATTA CTATTAC---
Rhynchosia bungarensis               AGGAATCCTC ATATGCATGA TTAATGATAC AAAATAATTA CTATTAC---
Rhynchosia minima                    AGGAATCCTC ATATGCATGA TTAATGATAC AAAATAATTA CTATTAC---
Rhynchosia aurea                     AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Rhynchosia sp                        AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Flemingia parviflora                 AGGAATCCTC ATATGCATGA TTAACGATAC AAAAGAATTA CTATTAC---
Flemingia stricta                    AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Flemingia stobilifera                AGGAATCCTC ATATGCATGA TTAACGATAC AAAATAATTA CTATTAC---
Bolusafra bituminosa                 AGGAATCCTC ATATGCATGA TTAACGATAC AAAG-AATTC CTATTAC---
Dunbaria ferruginea                  AGGAATCCTC ATATGCCTGA TTAACGATAC AAAATAATTA CTATTAC---
Glycine microphylla                  AGGAATTCTC ATATGCGTGA TTAACAAAAC AATACAA--- ----------

[                                      360        370  ]
Cajanus reticulatus-1                ---------- ------TGAA ACT
Cajanus lineatus                     AACTTACTTA TAATAATGAA ACT
Cajanus crassicaulis-1               ---------- ------TGAA ACT
Cajanus crassicaulis-2               ---------- ------TGAA ACT
Cajanus scarabaeoides                ---------- ------TGAA ACT
Cajanus pubescens                    ---------- ------TGAA ACT
Cajanus viscidus                     ---------- ------TGAA ACT
Cajanus mollis                       ---------- ------TGAA ACT
Cajanus actifolius                   ---------- ------TGAA ACT
Cajanus albicans                     ---------- ------TGAA ACT
Cajanus cinereus-1                   ---------- ------TGAA ACT
Cajanus cinereus-2                   ---------- ------TGAA ACT
Cajanus reticulatus-2                ---------- ------TGAA ACT
Cajanus sericeus                     ---------- ------TGAA ACT
Cajanus cajanifolius                 ---------- ------TGAA ACT
Cajanus cajan                        ---------- ------TGAA ACT
Cajanus latisepalus                  ---------- ------TGAA ACT
Erisoma buurkei                      ---------- ------TGAA ACT
Erisoma lucipetum                    ---------- ------TGAA ACT
Erisoma umtamvunense                 ---------- ------TGAA ACT
Erisoma salignum                     ---------- ------TGAA ACT
Erisoma simulans                     ---------- ------TGAA ACT
Erisoma griseum                      ---------- ------TGAA ACT
Erisoma psoraleoides                 ---------- ------TGAA ACT
Erisoma rossii                       ---------- ------TGAA ACT
Rhynchosia australis                 ---------- ------TGAA ACT
Rhynchosia bungarensis               ---------- ------TGAA ACT
Rhynchosia minima                    ---------- ------TGAA ACT
Rhynchosia aurea                     ---------- ------TGAA ACT
Rhynchosia sp                        ---------- ------TGAA ACT
Flemingia parviflora                 ---------- ------GGAA ACT
Flemingia stricta                    ---------- ------GGAA ACT
Flemingia stobilifera                ---------- ------GGAA ACT
Bolusafra bituminosa                 ---------- ------TGAA ACT
Dunbaria ferruginea                  ---------- ------TGAA ACT
Glycine microphylla                  ---------- ------AATA ATT
```

Appendix 3: Detailed sequence information of SNP containing BAC-end sequences

FPC Contig	FPC clone	Depth	SEQ name	# SNPs	Length	Primere left SEQ	Primer right SEQ	Left Pos.	Right Pos.	Left TM	Right TM	Product size
FPC_Contig0	CccABa-62K5	2	Cc_002_03223_Dec09	4	873	CCAACCCTCCAGTCAAGAAA	TCCAGCCCTAAAAATGTCCA	113	828	60.081	60.439	716
FPC_Contig0	CccABa-15O02	2	Cc_002_03974_Dec09	9	787	TGTGAGAGAAGAAAGTGGCG	AAAATCAGCCTTCAAAAACAATG	136	770	59.161	59.549	635
FPC_Contig0	CccABb-18O09	2	Cc_002_04484_Dec09	11	902	CATTTCTTCCAGCATCACCC	GTGACAACTGTGGAATGAGCA	16	788	60.461	59.742	773
FPC_Contig0	CccABa-10P04	2	Cc_002_05044_Dec09	27	536	CCAGGATCAATCTTCGCAAC	GCTTATCTAACCTCGCGCTC	19	507	60.603	59.232	489
FPC_Contig0	CccABa-26G15	2	Cc_002_05103_Dec09	20	873	AAGTTCCCAACAAGTCACCG	GGCCCATGCTTATTACCAAA	138	864	60.005	59.795	727
FPC_Contig0	CccABa-15O7	2	Cc_002_07065_Dec09	19	806	CATCCATTTGCATCACTTTTG	GTTCATCCAAAAATGAAAGGG	48	791	59.027	58.405	744
FPC_Contig0	CccABa-27A18	2	Cc_002_07093_Dec09	1	730	GCATGCAAGCTTCATTACAGTC	GCACACATCCCTATTCCTGG	20	603	59.917	60.34	584
FPC_Contig0	CccABa-08H15	2	Cc_002_09700_Dec09	15	830	CATGCAAGCTTTGGAATCAA	AACCATGCACAAATGTCTCG	20	772	59.809	59.572	753
FPC_Contig0	CccABa-27K8	2	Cc_002_10204_Dec09	6	944	TGGCAATTAGAAACGTGGAA	TTTCTTTAATCCGTGTTATGAGTGA	88	853	59.157	59.14	766
FPC_Contig0	CccABa-16O19	2	Cc_002_10216_Dec09	25	799	AAGTCCGATGATTATCAACTTATTTTT	TTTCTTTTCCGACATGAGGC	40	739	58.941	60.192	700
FPC_Contig0	CccABa-11K1	2	Cc_002_10431_Dec09	9	786	AGGCATGCAAGCTTTTAGGA	TTCAACCCATTAGCAGCACA	16	674	59.982	60.257	659
FPC_Contig0	CccABa-01L03	2	Cc_002_10585_Dec09	3	534	CCATTGGCTTAGTGATGGGT	TTTCCTTTGAAGAAAAAGGGC	82	528	59.813	59.703	447
FPC_Contig0	CccABa-15O02	2	Cc_002_10620_Dec09	8	683	CAGTGGCGATATGAAGCAAA	TGGAAAAACCTGATCAAGCA	5	663	59.833	59.247	659
FPC_Contig0	CccABa-56B04	2	Cc_002_10645_Dec09	15	658	TCGATGACATGTCCAGCTTT	GCATGACAACATGTCAAAGTGA	13	631	59.245	59.621	619
FPC_Contig0	CccABb-48D16	2	Cc_002_10651_Dec09	18	840	TGTTCATTGCTCAAACATTGC	ATGGCGGGTCATTACAAAAC	1	681	59.727	59.691	681
FPC_Contig0	CccABa-62K5	3	Cc_003_00889_Dec09	6	1007	GGAACCAAGGAAAGGTTGAAG	GCCAAGCCCAAGGTATTCTT	26	870	59.963	60.451	845
FPC_Contig0	CccABb-48N21	3	Cc_003_02819_Dec09	1	863	TCTCCTAGGCCCAATTTGAGT	TGGTGTGTTGAAGAGGGTCA	1	808	60.082	60.129	808
FPC_Contig0	CccABb-15H10	3	Cc_003_04848_Dec09	2	865	TCCTGATAAGGCTCGAGGAA	CTTCCCTCAGCTGAAAATGG	111	817	59.91	59.807	707
FPC_Contig0	CccABb-17N16	3	Cc_003_04990_Dec09	2	836	GGAGGAAGATGGACAAGGTG	AGAGGCTACTGAGGCATCCA	90	799	59.505	59.973	710
FPC_Contig0	CccABb-48A18	3	Cc_003_05008_Dec09	7	925	TTGTGACATCTATCTCCGATGC	TCGAATCATTGAAGAAAACCCT	4	743	60.103	59.945	740
FPC_Contig0	CccABa-02A15	3	Cc_003_06882_Dec09	10	820	AGGCATGCAAGCTTTGACTT	ACAACATAAGGGGGAGAGCA	18	807	60.022	59.55	790
FPC_Contig0	CccABb-48A21	3	Cc_003_10503_Dec09	2	910	CATGACATGCTTTGCCAAATA	AGCAGGCTGCATAAAGGAAA	61	809	59.576	59.982	749
FPC_Contig0	CccABb-48L19	3	Cc_003_10505_Dec09	5	756	TGCAAGCTACCTGATAGCCA	TGGCAGCAAAATTGTTATTGA	53	677	59.59	59.203	625
FPC_Contig0	CccABb-56D07	3	Cc_003_10514_Dec09	2	868	CCTCATCAGCACCATCATTG	AAGCAAATCTTCCCATGCTTT	81	865	60.072	60.089	785
FPC_Contig0	CccABa-22P02	3	Cc_003_10527_Dec09	1	701	CTTGAGGAACCAAGGCAAAG	AAAAAGGGGTGGCATTTAGC	2	605	59.846	60.312	604
FPC_Contig0	CccABb-41D19	3	Cc_003_10530_Dec09	4	725	TCATTAATGGACCTTGCTTGG	GGATTGGCAAAATGCAAAAT	118	709	59.946	59.779	592
FPC_Contig0	CccABb-48O04	3	Cc_003_10614_Dec09	10	645	TGGAGCCAAGAGAAACCATT	GACACATTTGATGGACCGTG	88	637	59.67	59.812	550
FPC_Contig0	CccABb-48P12	3	Cc_003_10621_Dec09	3	909	TGGTGGTGAATTCTGCTGTT	TCCACCAGGATCCATACCTC	79	850	59.139	59.737	772
FPC_Contig0	CccABb-48P12	3	Cc_003_10634_Dec09	10	845	GTTGAGTGGCATCTAGCGTG	AGGTTGGTTCCCAAGCTTTT	2	814	59.47	59.976	813
FPC_Contig0	CccABb-48L20	3	Cc_003_10677_Dec09	18	818	CTTCGACAGTGCACCAGAAA	TCCAAACCAGGCATTTCTTC	16	775	60.025	60.051	760
FPC_Contig0	CccABb-48H21	3	Cc_003_10683_Dec09	13	746	GCAAATTGAATGCTAAGTACCAA	AGCAAGGTTGTCATGTCACG	58	702	58.39	59.751	645
FPC_Contig0	CccABb-48K22	4	Cc_004_00203_Dec09	1	772	TTGTTCGATTGGAAGAAGGG	AACAAGATGAGGGGAGAGGG	83	728	60.044	60.447	646
FPC_Contig0	CccABb-48O19	4	Cc_004_10509_Dec09	25	874	ACCCTATGTGGTAACCTAGGCA	TGCAATGGCTTAATAGGGATG	88	866	59.783	59.936	779
FPC_Contig0	CccABb-02E24	4	Cc_004_10547_Dec09	11	803	TGTGAATGATTTACTGCACCAAT	GGTTCACCTAGGAACCCACA	10	763	59.397	59.82	754
FPC_Contig0	CccABb-02E24	4	Cc_004_10586_Dec09	8	928	CAGCAGGTCATAGCCAGTCA	TCGGATCTTGCTTTCCAAAC	48	811	60.008	60.192	764
FPC_Contig0	CccABb-48E23	4	Cc_004_10690_Dec09	9	1390	GGGTGGCAATGCTAGAAAAA	ACCTTTGTTGGGGTTTGATG	86	1202	60.074	59.688	1117
FPC_Contig0	CccABb-17M03	5	Cc_005_01671_Dec09	14	894	TGGCTTACAAACTTGGGACC	TGCAATATTTGCTCAATATTTATTTCT	25	884	59.971	58.744	860

FPC_Contig0	CccABb-48E23	5	Cc_005_06174_Dec09	8	900	CCAGAGCACATGGAATGAGA	CAGACCACTGAGCCAAAGG	117	899	59.787	59.399	783
FPC_Contig0	CccABb-80P05	5	Cc_005_09567_Dec09	6	804	TGATGGAATCATGCTTTGGA	AGTAAGCACGGTTTTGGCAC	102	761	60.009	60.176	660
FPC_Contig0	CccABa-25F20	1	CccaBA_B_Rearray_1c5_001-K6TR	24	692	AGGCCATGGCAATTTGATAA	TGCATCATTCACCACATTCA	22	620	60.289	59.469	599
FPC_Contig0	CccABa-70C16	1	CccaBA_B_Rearray_1c5_002-G3TR	20	788	AAGCTTGCATTCCCTCACAT	TCCTCATAAGGTACACTTTCCACA	19	670	59.7	59.927	652
FPC_Contig0	CccABa-26G15	1	cccaBa-Rearray-4-2A08TV	9	878	TCGTGGTTACTTCATCATTTGG	TCAATTGCGGTTAAGAAGCC	36	790	59.861	60.209	755
FPC_Contig0	CccABa-34G16	1	cccaBa-Rearray-4-2C10TV	12	489	ACTTGGCCAAAGAAAGAGCA	CCTTCTCTTTAAGGAGGAAAATGA	7	460	59.993	59.32	454
FPC_Contig0	CccABa-68G14	1	cccaBa-Rearray-4-3D07TR	4	553	CAAAAGATTTCCCCTTCGGT	ATCAGCCAGAAGGACTTCCA	67	549	60.291	59.803	483
FPC_Contig0	CccABa-50F13	1	CccaBa050-F13TR	9	855	CAACAAGTGGCTGAGGATCA	CCCAACTCTGGTGTAGTGGC	143	851	59.831	60.567	709
FPC_Contig0	CccABa-53M24	1	CccaBa053-M24TV	17	830	ATGCAAGCTTCTCCACCACT	GAAGATGACATGCCCAATGA	22	819	59.874	59.456	798
FPC_Contig0	CccABb-69H12	1	cccaBb-Rearray-5-2G08JR	5	815	TCTAGTGGCCTCCCTAAGCA	CGTCAGGCGTGGATACTCTT	93	744	59.971	60.277	652
FPC_Contig1	CccABa-02B01	1	cccaBa-002-B1TV	6	814	TACAACTGTGCCTGCTTTGC	CTTCACGTGGGTTTGTCCTT	61	720	60.058	60.005	660
FPC_Contig1	CccABa-20A04	1	CccaBa020-A4TR	31	884	GCTTGGATGCCTTGTTCATT	GAAAATGCCTCCAAAGCAAG	20	793	60.081	59.823	774
FPC_Contig1	CccABa-22D18	1	CccaBa022-D18TR	1	866	GAATACTCAAGCTTGGAGGGAG	ATTGGAGCTCTTGGTTGAGC	10	770	59.371	59.434	761
FPC_Contig1	CccABb-89A18	1	cccaBb-Rearray-5-3A10JF	7	770	CCAAAAAGGGTGAAGGTTGA	AGCACAAGCTTCCCAAGTGT	16	660	59.942	59.914	645
FPC_Contig10	CccABa-85F10	2	Cc_002_03739_Dec09	4	736	CTCGGGCCCACTCTTAGAC	TTTTACCCGCAGATACCCAC	140	734	59.812	59.823	595
FPC_Contig10	CccABa-20J7	2	Cc_002_08106_Dec09	6	741	TAGAGTCGACCTGCAGGCAT	GCTTGCTCTTCTTCAACCTCA	4	617	60.966	59.749	614
FPC_Contig10	CccABa-22O2	2	Cc_002_10472_Dec09	8	810	TGCAAAGATGGAACGTGAAG	GATTCCATCAAGAGACAACAACA	23	808	59.84	59.091	786
FPC_Contig10	CccABa-18E14	3	Cc_003_06922_Dec09	3	748	AATCCAGAAACACTTTCGCAG	GGGTCTTGAGCTGGTTGAGT	94	727	59.369	59.305	634
FPC_Contig101	CccABa-12F19	3	Cc_003_03206_Dec09	27	889	TGGTTTCCGGATGAGAAGAC	GGTTGGGGGTACATCTCCTT	80	843	60.05	60.052	764
FPC_Contig102	CccABa-23C12	1	CccaBA_B_Rearray_1c5_001-G22TR	5	712	CTAAAATTGTCCGGGTCGG	GCTCGCTCCTTCCTAAAACA	87	695	60.308	59.592	609
FPC_Contig11	CccABa-39J1	4	Cc_004_03654_Dec09	6	951	ATGCCTTCAGGCAATGACTT	CCATGTTGCCTCTCTGGACT	143	928	59.7	60.261	786
FPC_Contig12	CccABb-41C03	2	Cc_002_10574_Dec09	5	625	GGGTGGTAGGGACTTATGGG	CCATTATTTCAAAGCCTTGCTC	3	502	60.434	60.095	500
FPC_Contig12	CccABb-16P22	2	Cc_002_10654_Dec09	12	788	AAGTTGCTTCTCACATGGGC	CCAATTAGGTTCAGGTCATGC	69	730	60.263	59.443	662
FPC_Contig12	CccABa-03D03	3	Cc_003_04016_Dec09	2	655	AGTGGAAGTGGCTTGGAAAT	TTTTTAAGTACCTCTTGTTGTCTCCA	86	613	58.646	59.666	528
FPC_Contig12	CccABb-31H20	4	Cc_004_07311_Dec09	7	905	TCATCTTCAACATCAAAGGCA	TGTCTCCTACATGCCCACCT	20	788	59.262	60.533	769
FPC_Contig12	CccABb-47H08	4	Cc_004_10696_Dec09	10	813	TCCGTTAATTTGGGAAATGC	GAAAGCATAAAAACGGCCAA	24	703	59.773	60.075	680
FPC_Contig123	CccABa-59H6	2	Cc_002_03376_Dec09	11	871	CCACATGACCATTGCAGTTC	CAGCACTTGGAACCACTGAG	87	791	59.967	59.461	705
FPC_Contig123	CccABa-25L3	2	Cc_002_05974_Dec09	20	820	CAAGCTTTGTGCGTTTGCTA	CATTTATGCGAATTTTATGGACAA	25	688	60.191	60.095	664
FPC_Contig123	CccABb-42C20	2	Cc_002_10364_Dec09	4	688	TTCTCTCTCCCTCCCATTCA	GCATTCAAATGCTCCAAAAAT	15	587	59.731	59.071	573
FPC_Contig123	CccABa-59O12	2	Cc_002_10381_Dec09	13	817	GCTTTCATGGGTATTGAGGG	TTCATTTTCCAACAGGGGAG	23	787	59.387	59.903	765
FPC_Contig123	CccABa-64D07	1	CccaBa064-D7TR	8	880	AGGCAACTTTTGGAGAGCAC	CTGCAGCAACAGCACTAGGA	86	843	59.478	60.347	758
FPC_Contig123	CccABb-14O13	1	CccaBb014-O13JR	11	733	TCCCAAAAAGGTTCAAGGTG	GTCTTCTGGGAGGAGGCTTT	53	712	59.942	59.817	660
FPC_Contig123	CccABb-62N15	1	CccaBb062-N15JR	4	821	CCCCGAGTCCACTGTTTTTA	TTCCAATTTTGTCTGTGGCA	38	814	59.964	60.088	777
FPC_Contig13	CccABa-48A24	2	Cc_002_04584_Dec09	20	741	GTGGTAGGATCATTGGGTGC	AGCAGTCAAGCGTTAGAGTCG	27	689	60.203	59.834	663
FPC_Contig13	CccABb-56A19	2	Cc_002_10226_Dec09	4	788	AAAAATACCCCTTCCCTCCC	GGAGACCAAGGATGATCCAA	81	765	60.358	59.862	685
FPC_Contig13	CccABa-48A24	2	Cc_002_10401_Dec09	22	797	GAAAAAGCCCAGGAGTCAGA	TTAATACGTACAAAATAATTGTCCAAA	106	742	59.405	57.466	637

FPC_Contig13	CccABb-78J10	1	CccaBA_B-Rearray-1c5-003-I20JF	4	785	CAATGACGCTGTCTTTGTTGA	CATGATGAATACATATTTACACCCAA	0	630	59.896	59.021	631
FPC_Contig130	CccABa-50O1	2	Cc_002_10278_Dec09	8	670	TGGTTCATGGAAGATTGAAAGA	TTCAGCCATCCACAACAGAA	76	644	59.546	60.24	569
FPC_Contig130	CccABa-38G11	2	Cc_002_10460_Dec09	22	842	TCTTGTTGCTCTCCCCAAAT	AAGGCCTGGGTTTGAAAGAG	44	779	59.67	60.601	736
FPC_Contig131	CccABa-36L10	2	Cc_002_10175_Dec09	8	901	TTTTCTACATTCCCACCCCA	TCAGGAGACCCTGACATCAA	108	865	60.162	59.18	758
FPC_Contig131	CccABb-37F9	3	Cc_003_01991_Dec09	1	924	TCTGCAAGATGAAGAGAATCAAA	ATGCTCAAACCCATCAAAGG	92	897	59.101	59.933	806
FPC_Contig136	CccABa-45J5	2	Cc_002_10156_Dec09	29	739	CCGTTGTGGGAGAAGAATGT	CTACCTTCAAAACGACCCCA	31	706	59.966	59.964	676
FPC_Contig136	CccABa-45F14	2	Cc_002_10299_Dec09	14	875	CAAGCTTGTGCAGTCTCCAA	CCAGACATTTTTGAGTGGATTG	18	743	60.175	59.471	726
FPC_Contig136	CccABa-57G13	2	Cc_002_10334_Dec09	4	681	GGATTGCTGGTTAGTGTTTGC	TGGAGCAACTTCAAAATGGA	3	678	59.629	59.247	676
FPC_Contig136	CccABa-77L13	1	CccaBA_B_Rearray_1c5_002-A6TR	9	732	GCTTCTCAGGTGTTGGGTGT	ACCCACATCTTGAATGGGAA	21	634	60.159	60.173	614
FPC_Contig136	CccABb-36P17	1	CccaBA_B-Rearray-1c5-003-L24JF	3	768	GATTGTGATACCCCTGGTGC	CCTCGCTGTTGGATGTTCTT	96	738	60.203	60.255	643
FPC_Contig136	CccABb-84E14	1	CccaBA_B-Rearray-1c5-003-M2JF	1	872	CTTCCGACTGATGATTGACCT	TTGCAATGTAATTGTGGTTTCA	0	730	59.151	58.982	731
FPC_Contig136	CccABa-18E7	1	CccaBa018-E7TR	7	717	TTGATGTTTGGTGCATGGAT	CCATAAACACATGAGTCAAGGAA	111	714	59.781	59.007	604
FPC_Contig136	CccABa-45E14	1	CccaBa045-E14TR	24	878	TCAAGCTTTACGAGTGGCATT	TGAGGGTACTTAGGGGCTGA	16	721	59.897	59.688	706
FPC_Contig136	CccABb-10L09	1	cccaBb-Rearray-5-3D09JR	7	744	TCCAACCAATTGAAGTAACCG	AACTGATTGCTTTGCTTGGG	53	703	59.847	60.249	651
FPC_Contig14	CccABb-16F13	2	Cc_002_02618_Dec09	14	681	TCGACAAAACGTTGAAGCTG	GGAATCCTTCGGACCTTCAT	101	649	60.027	60.272	549
FPC_Contig14	CccABa-16E11	2	Cc_002_04541_Dec09	2	1025	CCTCGTCAAACGAGCAATTT	GCAGTGAACTCAATACGGGC	150	983	60.249	60.667	834
FPC_Contig145	CccABb-56C6	2	Cc_002_10272_Dec09	16	854	GGGAGACAAACATGGGAGAC	TGAGTGTTTCGTGAAAATTTGG	78	812	59.359	60.011	735
FPC_Contig145	CccABa-50M13	2	Cc_002_10414_Dec09	39	820	TACCTTGGCTTCTCCCACAC	TCTTTTATGATGGCTTCGGG	62	790	60.111	60.031	729
FPC_Contig145	CccABb-56C6	2	Cc_002_10448_Dec09	8	849	ATTGGAACTGGCTGGATGAC	CCCGTGAAATAGACTGGCTC	31	817	59.934	59.694	787
FPC_Contig145	CccABb-86I04	3	Cc_003_03258_Dec09	16	753	CAAAGCCAAGTCATTGTTGC	CCACTGCCCAGTATCGTTTT	15	656	59.322	59.993	642
FPC_Contig146	CccABa-63O24	2	Cc_002_00073_Dec09	15	607	TGCATCAATAAGTTGGCATGA	CATCGCTATCCATGTCTCCA	17	606	60.088	59.629	590
FPC_Contig15	CccABb-9I8	2	Cc_002_10040_Dec09	21	891	ACCTGAAGGCATTCCCTTTT	GGTGCTTGATTTTGCTGGTT	52	834	59.94	60.118	783
FPC_Contig16	CccABa-54F4	2	Cc_002_07431_Dec09	13	791	CCTCCTATGTGCTTTGGCAT	GTTGTGGATGGCAGAAGGTT	72	737	60.096	59.973	666
FPC_Contig16	CccABb-38O6	2	Cc_002_10280_Dec09	5	711	AGGCTGAAATTGAGCTTCCA	GCTCCCTCCATCTTCTCCTT	78	686	59.955	59.778	609
FPC_Contig16	CccABa-11I3	2	Cc_002_10312_Dec09	29	744	GCAACAATTTTGGCATCTGA	AGATCATGCATTCTCGGACC	48	733	59.67	60.042	686
FPC_Contig16	CccABb-83C16	5	Cc_005_01748_Dec09	4	883	GCCAACTACATCCTTGGTGG	AGCAAAACAGCGCTTCAGAT	13	817	60.375	60.162	805
FPC_Contig16	CccABa-81M14	1	CccaBA_B_Rearray_1c5_002-E6TR	6	658	AAGCTTAGCGTGGCTCTCAC	CCCTAATCTCTTACAAGTTGGCA	2	566	59.786	59.675	565
FPC_Contig17	CccABb-36O17	2	Cc_002_10706_Dec09	5	670	CTCAGGAGCCCTCTCCTTTC	TCTGCTCATCTGGCTGTTTG	51	635	60.468	60.136	585
FPC_Contig19	CccABa-29I03	2	Cc_002_07064_Dec09	5	797	CCGATTTTGACTTCGAGGAA	CCGGAGTCGACAGAATAAGC	74	767	60.184	59.836	694
FPC_Contig19	CccABb-35O5	2	Cc_002_10316_Dec09	4	841	TCAATGAACACTGCACAAAGC	GTCCCAGAGCACACAGATGA	120	839	59.903	59.827	720
FPC_Contig19	CccABa-1J14	2	Cc_002_10432_Dec09	3	727	GCATTTTAAAAGAGTGAAGAGGTTG	ATGGGAACTTCGACTCATTTTC	5	585	59.742	59.465	581
FPC_Contig19	CccABa-29I03	3	Cc_003_04222_Dec09	2	907	GCGTACAGGTTAATTTCCCG	TGATGGATGGTTGTGGAATG	107	833	59.474	60.177	727
FPC_Contig19	CccABb-48O14	3	Cc_003_10707_Dec09	2	847	GAGAATGTTGGCGGGTTTTA	TCTGACAGCACACAAGGACC	126	834	59.938	59.872	709
FPC_Contig19	CccABb-72J14	1	CccaBA_B-Rearray-1c5-003-E22JF	11	707	ATCTTCTCCCGATCTCCGAT	AAGAGGCATGAATTGAAGCAA	7	664	59.999	59.837	658
FPC_Contig19	CccABb-72K21	1	CccaBA_B-Rearray-1c5-003-E24JF	9	835	TACGTCATCATTCCCATCCA	ATTGTCCCCAAAGTGTTTGC	71	792	59.732	59.836	722
FPC_Contig19	CccABa-14L23	1	CccaBa014-L23TV	12	754	TGGAGATTGCCAATGTTGAA	TGCTCTTATTTTTGGGCAGC	53	713	60.049	60.343	661

FPC_Contig21	CccABb-47K20	2	Cc_002_02959_Dec09	7	785	TGACGGGTTCTGAATATCTCG	TCCAAGGACATTTGTTTCAGG	39	674	60.081	59.956	636
FPC_Contig21	CccABb-64E14	3	Cc_003_10480_Dec09	9	844	ACTTCATGAGCCCCACAATC	TGTGAGAGCATGGCTTCTTG	0	742	59.934	60.136	743
FPC_Contig21	CccABb-20O7	5	Cc_005_06482_Dec09	6	882	ACAAAAGTGGCTTCTGCCAT	CAATCCCTTAGCAGTGGCAT	10	731	59.74	60.096	722
FPC_Contig22	CccABb-26A7	2	Cc_002_01238_Dec09	17	794	GTTCCAGCAGCATTCACAAA	CCAGATGAGAGGGATCCAAA	15	662	59.847	60.003	648
FPC_Contig22	CccABa-31P4	1	CccaBa031-P4TR	8	756	CCTTGTGCCAAACATCGTAA	CAATTCCTGAATCACAACATGG	88	724	59.585	60.227	637
FPC_Contig23	CccABa-37J3	2	Cc_002_03340_Dec09	8	854	TGATGTCCAAATGATGAGGG	TGCAAGAAATGTGGCTTTTG	29	819	59.296	59.849	791
FPC_Contig23	CccABa-37J3	3	Cc_003_05873_Dec09	8	585	CAGGCATGCAAGCTTATTGA	TTTTATTTTGTATTTCGCCATACTTG	16	491	59.976	59.746	476
FPC_Contig23	CccABa-51P13	4	Cc_004_10250_Dec09	23	732	CAAAAGAAATAAACAAGAAAAAGGCT	TTTTGCGGGTGTCTATAGGG	53	686	59.674	59.953	634
FPC_Contig23	CccABb-86J14	1	CccaBA_B-Rearray-1c5-003-O2JF	9	895	CTCTCGACGAACAATCTCCC	TGGAGTCGGTAAGACAACCC	27	884	59.803	59.966	858
FPC_Contig23	CccABb-86J14	1	CccaBA_B-Rearray-1c5-003-O2JR	28	584	TGGGAATATCAGTTGATAAAATTCA	TCAATCAAAATAAAAGATAAAGCACAA	6	476	58.848	59.54	471
FPC_Contig24	CccABb-16L02	2	Cc_002_10644_Dec09	5	880	CCCCTCCATGCTAGAATCAAT	GCCAACTCTGGAGGCTGTAG	7	759	60.293	60.012	753
FPC_Contig24	CccABa-56K13	5	Cc_005_00018_Dec09	4	1064	CATCCAAGCTCAAGCATCAA	ATCCTTTTCAAGAGGGAGGG	64	934	59.948	59.508	871
FPC_Contig24	CccABa-56K13		CccaBa056-K13TV	6	739	ATATTTTGGTGCACGGTGTG	CAACACAAGGGGAAACATCA	101	705	59.322	59.389	605
FPC_Contig25	CccABb-31J22	2	Cc_002_04910_Dec09	10	638	TCAATTCGGTCCCTTACTTTG	TCTCCATTGCTCGAGTTTGA	30	549	59.074	59.522	520
FPC_Contig25	CccABb-58G7	3	Cc_003_10307_Dec09	10	1012	GCATCCCTTATTTTTGGTGC	GGAGAAGGAACATTGGGAGA	22	872	59.415	59.065	851
FPC_Contig26	CccABa-53G02	2	Cc_002_10529_Dec09	22	678	GCCTGGACTCGTCTACAACC	AGCTGAAATTGTCCATATCTTTTG	55	645	59.727	58.759	591
FPC_Contig26	CccABb-09F10	3	Cc_003_01602_Dec09	4	792	TTCTCGGTTTAGCATCAGCA	ATGCATGGGTGGTCTCTTTT	41	783	59.566	59.41	743
FPC_Contig26	CccABa-03O03	3	Cc_003_09298_Dec09	5	742	GCAGCAAATGGGAGAGACAT	TGAATCGACAACTAAACTGGGA	26	726	60.226	59.609	701
FPC_Contig28	CccABa-57D06	1	cccaBa-Rearray-4-3A03TR	12	513	TAGAAACAAATGCTTCGGATGA	GCAACGTAACAAAAGATTGGC	52	461	59.72	59.644	410
FPC_Contig28	CccABa-57D06	1	CccaBa057-D6TV	11	684	ATGCTTCCTCAAAATGGTGC	CTTTGATGAGCAACAATTGCAT	56	683	60.081	60.138	628
FPC_Contig29	CccABb-52O11	4	Cc_004_08627_Dec09	4	824	CCAACCTCTCAGCACGGTAT	TCCACTGGGGAAGTCTAGAGAG	88	803	60.134	59.872	716
FPC_Contig3	CccABa-13C23	1	CccaBA_B_Rearray_1c5_001-K15TR	32	823	CTCCCATAGTTGCCACATCA	TTGGTGAGAAGGCGAGAGAT	36	707	59.522	59.95	660
FPC_Contig3	CccABb-82O11	1	cccaBb-Rearray-5-3A01JF	12	761	TCCCATATCCAACAACACGA	CAGCTTTCCTTTCTCCTTTCC	89	721	59.774	59.469	633
FPC_Contig30	CccABb-65G8	4	Cc_004_10249_Dec09	4	913	CGACACAGAAGTGTGGCATC	CTAGAGGCAACAAAGCCAGG	16	805	60.319	60.008	790
FPC_Contig30	CccABb-64N16	5	Cc_005_10266_Dec09	10	815	TTTCGTGCACCACAGAGTTT	CTCTTTCGCACATCCCAACT	129	808	59.339	60.255	680
FPC_Contig31	CccABb-51O13	3	Cc_003_02932_Dec09	5	855	AGTGGTCTAATCTGGGGCAA	TGCATGCATGGTCTGGTAAT	22	723	59.55	59.955	702
FPC_Contig33	CccABa-50G10	2	Cc_002_10329_Dec09	15	771	GTTGAGACATTTCCCCGAGA	TTGCATAGTCCCCTTTTTGG	50	686	60.05	59.931	637
FPC_Contig34	CccABb-06L16	3	Cc_003_01846_Dec09	21	839	TGGAAGTATGGCAAGGAAGG	AGGGCTATGGCAAGAAGGAT	99	796	60.066	60.06	698
FPC_Contig34	CccABb-40G05	3	Cc_003_06164_Dec09	9	775	TGAACGAGGAGGATAGCACA	AACCAAAAAGGATGAAGACAACA	24	643	59.394	59.9	620
FPC_Contig35	CccABa-63H15	3	Cc_003_00042_Dec09	26	1012	CGGCCGTATTGAAGTTTCTG	GCCAAAGGTTGAACCGAGTA	49	877	60.628	60.11	829
FPC_Contig35	CccABb-39H03	3	Cc_003_10692_Dec09	30	765	ATCACTCCGACCACTATCGC	TCCAGAAGTTCACCAGCCTT	2	638	60.104	59.844	637
FPC_Contig35	CccABa-62I15	1	CccaBa062-I15TR	5	734	GCCAAGAACAAAGGCAAAAG	ATTCCCTTTGCAGGCTATGA	33	669	59.861	59.668	637
FPC_Contig36	CccABa-11K15	3	Cc_003_04040_Dec09	7	788	AGCACAAGCTTCCCAAGTGT	AATTTGCCTTGTTTGGAACG	29	678	59.914	59.975	650
FPC_Contig36	CccABa-11K15	3	Cc_003_08690_Dec09	1	664	AGATGAGGTTGTGGGCTTTG	TGCCTTGTTACAGAACAATCTGA	58	598	60.111	59.799	541
FPC_Contig36	CccABb-60N11	3	Cc_003_10525_Dec09	17	915	TGCTTTTGGAAGCCATCTCT	TCCACGAGTTTGTGAAACCA	68	832	59.955	60.128	765
FPC_Contig36	CccABb-77F12	1	cccaBb-Rearray-5-2H07JF	10	676	ATCGAGGATCGTGAGGATTG	GCGCAGCTCATGTAAGTCAA	36	642	60.034	60.165	607
FPC_Contig38	CccABa-87H17	2	Cc_002_00945_Dec09	6	777	TTGTTCTTATGCTTTGCCCC	AGGTCCAAACACCACCAAAA	20	754	60.074	60.246	735

FPC_Contig38	CccABb-48I14	4	Cc_004_04976_Dec09	6	818	GGCTCCTCTCCTAAGCATAACC	GTCAGGTTTGAGCCGAAAAT	59	775	60.579	59.174	717
FPC_Contig39	CccABa-45N05	2	Cc_002_03606_Dec09	8	743	GCATGCAAGCTTGTAGGAGA	CAGGAAGAGTTGGACCTTGTG	20	668	59.178	59.748	649
FPC_Contig39	CccABa-21I17	2	Cc_002_05027_Dec09	8	648	ACCAACATGTCCCCAAGAAC	AGAGGCTCACAACAAGGGTG	54	643	59.679	60.298	590
FPC_Contig39	CccABa-63D22	2	Cc_002_10653_Dec09	2	612	GCCAATGTGAACATAAGGCA	GGCTTCTACAGAGGAGGTGC	42	587	59.548	59.043	546
FPC_Contig39	CccABa-83G12	1	CccaBA_B_Rearray_1c5_002-G6TR	4	788	GTAGGCCATGGCAATTTGAT	TGGTGATGTTGTTGATGGCT	40	710	59.791	59.967	671
FPC_Contig4	CccABa-41C2	2	Cc_002_04557_Dec09	13	831	GAGCTTATTCACGGAAAGCG	TTTGAATCCATCCCGAAGAG	54	750	59.982	60.006	697
FPC_Contig4	CccABb-37H7	2	Cc_002_06507_Dec09	4	704	GCGCTAAATCTCTCTGGCTG	AAAGAAAGAAAGGATGGAGTCGT	87	652	60.257	59.664	566
FPC_Contig4	CccABb-80I12	2	Cc_002_10403_Dec09	22	897	CCTAAGACGTGGGACTCCTCT	GTGGACCAGACTTTGGGATG	32	793	59.748	60.363	762
FPC_Contig4	CccABa-41C2	2	Cc_002_10437_Dec09	8	782	AGGCATGCAAGCTTCTCAAT	GGCTTGAAACTGAGGAACCA	18	699	59.985	60.232	682
FPC_Contig4	CccABb-67M15	1	CccaBA_B-Rearray-1c5-003-C10JF	26	575	GGATGCAAAGCCAATTTCTC	TTTTCAGGCCACAACATGAA	23	494	59.651	60.088	472
FPC_Contig41	CccABa-58J16	2	Cc_002_00375_Dec09	5	866	TCATCGAATTGTGCTCCAAA	TTGGATGATGTGTGGGAGAA	76	811	60.197	59.893	736
FPC_Contig41	CccABa-25P11	2	Cc_002_09986_Dec09	17	804	TGACCCTTTTGAGCTTTGTCA	GCTCATCGATCGTGGGAATA	89	765	60.791	60.973	677
FPC_Contig41	CccABa-25P11	3	Cc_003_05976_Dec09	10	759	GCCTTAGTTTGCAATTTCCC	AACAACAAGAATCAAAAACATGC	56	672	58.696	58.225	617
FPC_Contig41	CccABb-89K12	1	cccaBb-Rearray-5-3B01JR	4	859	ATTGCCTTCCAGAAGAAGCA	AATGGAGATGCTGGAACAGG	105	825	59.955	60.073	721
FPC_Contig42	CccABa-16E22	2	Cc_002_10352_Dec09	6	635	AAGCTTGCATCCATACTCCG	CTTCTCCGCACTCTCCACTC	26	563	60.235	60.135	538
FPC_Contig44	CccABa-63I23	1	CccaBa063-123TV	2	652	TTCAAGTGGTTCCCTGACAA	CTGTCCCACAACAGCTTCAC	118	638	59.109	59.31	521
FPC_Contig46	CccABa-03C11	3	Cc_003_09896_Dec09	11	885	ACAAGCTTGCCCTCAAGTGT	CGTCGATTCGTTGTGCTCT	77	798	59.914	60	722
FPC_Contig46	CccABa-69H08	1	cccaBa-Rearray-4-3D08TR	2	561	GCAACACCATCTTCCAAGTGT	TGGGGATGAATCTGAAGAAGA	76	552	60.028	59.614	477
FPC_Contig46	CccABa-69H08	1	cccaBa-Rearray-4-3D08TV	30	671	GGTTACCGGAGATCGGTCAT	TGTGCATTTAATTATGATAACCATTTT	107	651	61.106	59.13	545
FPC_Contig47	CccABa-88O22	4	Cc_004_08708_Dec09	3	922	CGGTTGTATTGGGCTTCATT	AAGCAGCATTGTGGACTGTG	48	915	59.823	59.905	868
FPC_Contig5	CccABa-24J18	2	Cc_002_09764_Dec09	7	707	AAGCTTGCTACAACCATCAGC	TGAGCATGTGTATTGTCTGTGC	19	647	59.536	59.794	629
FPC_Contig5	CccABb-09C14	3	Cc_003_10697_Dec09	5	798	AAAGCCTGACAAGAAACCCA	CGGTGCCTTGATTTGTTCTT	44	776	59.711	60.11	733
FPC_Contig5	CccABb-09C14	5	Cc_005_08191_Dec09	10	889	ATATTCCGTGGGATCCCTTT	TGGATGCTGATGGTTGTAGC	15	736	59.492	59.679	722
FPC_Contig5	CccABa-58O24	1	cccaBa-Rearray-4-3A09TR	25	627	TCTCCATGAATTTGTGCTCG	GCGTGAGAAAAGAGGAGAGG	82	624	59.799	59.162	543
FPC_Contig5	CccABa-75M21	1	cccaBa-Rearray-4-3F01TR	10	741	TCCACCTATCAAATTGACCTCTC	TGGGCACAAATTTACCATCA	9	603	59.473	59.786	595
FPC_Contig51	CccABa-52A20	2	Cc_002_09394_Dec09	14	832	ATTTCTCCACACTTGCCACC	TTTTCAACTCCAAAATGCCC	52	796	59.973	59.916	745
FPC_Contig51	CccABa-64F16	2	Cc_002_10258_Dec09	48	1186	GGCATGCAAGCTTGTTCTCT	GCTCACAAATTGGTGCCTAAA	19	1006	60.547	60.124	988
FPC_Contig52	CccABa-5N6	2	Cc_002_09916_Dec09	11	797	GCATGCAAGCTTAGTGCAAC	TGTACCAATCAGAAATCATTCCA	19	739	59.644	59.329	721
FPC_Contig52	CccABa-37I14	2	Cc_002_10254_Dec09	5	911	TTCCCAAGGTTCAAGCTCAG	AAGCACTGTAAAGAGGGCGA	83	887	60.366	60.015	805
FPC_Contig55	CccABa-14P12	3	Cc_003_03764_Dec09	3	718	TAAATTTTGAGCTGGGGGAA	ATCAGCTGGCAAAGACAGGT	76	683	59.52	59.874	608
FPC_Contig57	CccABa-08P19	3	Cc_003_06696_Dec09	4	1152	GCATGCAAGCTTGTTCAAAA	GGGGCGAAAGTCTCATCATA	20	1068	59.999	60.036	1049
FPC_Contig58	CccABa-41H14	1	CccaBa041-H14TR	25	712	GGCTTGCTTTTGCTCACTTT	ACCGCAACACATCCTTTTTC	28	690	59.637	59.978	663
FPC_Contig6	CccABa-15P20	2	Cc_002_09960_Dec09	3	770	GCATGCAAGCTTACACTCCA	TGGAAAAACGTCCAAAGGTC	20	721	60.019	59.948	702
FPC_Contig6	CccABb-5P14	2	Cc_002_10251_Dec09	2	701	TTTTTCTTCGTTCATTGTATTAGCA	CCTGACTTGACTACGATTGGTC	46	664	59.286	58.721	619
FPC_Contig6	CccABb-16H11	3	Cc_003_10524_Dec09	10	928	CCTCCGACAAATACTGACAAAA	GTGATGCAAGCACGACAATC	185	926	59.131	60.281	742
FPC_Contig6	CccABa-75M13	1	CccaBA_B_Rearray_1c5_002-M9TV	11	876	GCAAGCTTTGAAGGACCAAG	TGCACAAACTAGAGCGTGAAA	24	734	59.993	59.665	711

FPC_Contig7	CccABa-62A12	2	Cc_002_08818_Dec09	23	834	TTTCCCAATCATTTGAAGCC	TGTCCAACTCAACCCATTCA	30	728	59.878	59.935	699
FPC_Contig7	CccABa-14B01	3	Cc_003_07626_Dec09	9	759	AGCATGCAAAGGAAAGCACT	TGCATCATTCCATTTTAAGCC	35	645	60.022	59.921	611
FPC_Contig7	CccABa-31F13	3	Cc_003_09985_Dec09	12	833	AGCTTTAGAAACCGCCTTCC	TCTTATTTTTGTGCAGCCCC	28	731	59.856	60.074	704
FPC_Contig7	CccABa-64C24	3	Cc_003_10147_Dec09	15	755	ACTTGTGGTGGTTGGTCCTC	GCTTTGCAACAAGCAATGAA	105	726	59.859	59.999	622
FPC_Contig7	CccABb-48P23	3	Cc_003_10580_Dec09	1	1289	TTGAGTTGCCTAATTTCTTCCC	ACTTAATGGCCGAATGAGACA	100	1209	59.609	59.583	1110
FPC_Contig7	CccABa-17D21	5	Cc_005_00951_Dec09	3	854	GCACAAGCTTCTCCTTAGCAA	TCAAGAAAAGCAAGACAAGGTG	36	735	59.783	59.534	700
FPC_Contig7	CccABa-20O24	1	CccaBA_B_Rearray_1c5_001-E10TR	7	880	CACGAAAAACTTAGTCTTGTGCC	GAAAAATAATCCCCGCAGTTT	98	802	60.208	59.348	705
FPC_Contig7	CccABa-58H20	1	CccaBA_B_Rearray_1c5_001-J6TR	6	792	GTGAACCTCACCATGGTTCC	CATCTCCTAGGCCCATTTGA	50	788	60.223	60.029	739
FPC_Contig7	CccABa-34E16	1	CccaBa034-E16TR	2	888	CAAACACCCTTGGTGGAAGT	TGATGACCATGCACACTACTCTC	30	771	59.861	60.194	742
FPC_Contig73	CccABa-51I19	2	Cc_002_10516_Dec09	11	877	AATTCAAGCGTGGCTAATGG	GCCGTCAGAATTGGCATTAT	46	770	60.096	59.929	725
FPC_Contig75	CccABa-10H13	2	Cc_002_09892_Dec09	7	877	AGCTTTAGAAACCGCCTTCC	TTTTGTGGGATTCTTAGCGG	27	759	59.856	60.067	733
FPC_Contig75	CccABa-13P11	4	Cc_004_10571_Dec09	2	730	CCTCCAATTCCTCACCTTCA	CAGGTGGTGATTATGGACTGG	76	672	60.042	60.239	597
FPC_Contig76	CccABa-83F08	1	cccaBa-Rearray-4-3G09TR	5	833	CTGAAAACCCACTTGGTTCC	GATTTAAGCTTTCGAGGAGCAA	5	749	59.425	59.99	745
FPC_Contig78	CccABa-06N21	3	Cc_003_08861_Dec09	4	1097	GTTATTGATTGGAAGCCCGA	TGCCAGGGAACTAGTGCAG	133	1024	59.901	59.992	892
FPC_Contig78	CccABa-06N21	4	Cc_004_10506_Dec09	11	753	CCATACTCAACCCCAAATGAA	AAATTAGATAAATTGAGGAGGGATTT	12	674	59.671	58.179	663
FPC_Contig78	CccABa-65M12	1	CccaBa065-M12TR	11	829	ATTCCATATCATACCCTTAAAAACAC	GTTCGGGTCACCCAATGTC	49	711	57.935	61.205	663
FPC_Contig78	CccABb-04H17	1	CccaBb004-H17JF	4	833	TCTAAAACAAGCTTCAGTGTTTGG	TCGCTGTGTAAACCCATTTTT	6	671	59.862	59.505	666
FPC_Contig79	CccABa-23N20	2	Cc_002_08253_Dec09	5	891	ACTCATGACCGTGGGTGAAT	CCTGTGCACAAAGATGGAGA	110	829	60.246	59.831	720
FPC_Contig8	CccABa-49G19	2	Cc_002_00414_Dec09	10	704	GCTCAAGGGAAAGGAGGAAA	AGGCCAAGCAAAACAACAAC	53	701	60.683	60.154	649
FPC_Contig8	CccABb-55I10	1	CccaBA_B-Rearray-1c5-003-I23JR	18	816	TCCATGTGGAAGAGAGGCTT	CACTCCGGTGTTGTTTGTTG	19	751	59.803	60.041	733
FPC_Contig81	CccABa-70P14	4	Cc_004_05777_Dec09	2	797	TATAAATTCGCCATGGCCTC	TTGTGAGTTAGCCATGCAGC	94	733	59.892	60.019	640
FPC_Contig83	CccABb-39B4	3	Cc_003_09616_Dec09	5	786	TAGAAGGACCACGGGATTTG	CATCCAATGTGAATGAAACTTAGC	77	723	59.926	59.904	647
FPC_Contig86	CccABa-14B17	2	Cc_002_00783_Dec09	1	778	CTCGTTTTGGTTGCAAGTGA	TCGGTAAAAATGATTACAAGGATG	60	714	59.881	59.314	655
FPC_Contig86	CccABa-14B17	3	Cc_003_10177_Dec09	10	980	ACCGTTCACTTCAAACCAGG	CGAGACGACCGAGATAAAGC	31	949	60.005	59.978	919
FPC_Contig86	CccABa-39B6	1	CccaBa039-B6TV	9	897	GCAATCTGTGCAGTCCTTGA	TGCAACACCCCCTTAAACTC	47	768	59.992	59.971	722
FPC_Contig87	CccABb-86C16	1	CccaBA_B-Rearray-1c5-003-M18JR	2	692	GGATAGTGGATCTTGGTGCAA	TTCACCTCAAGAGGCCAAAC	62	655	59.947	60.232	594
FPC_Contig89	CccABb-48O13	3	Cc_003_10631_Dec09	5	768	TCAGAAAAAGGCAAGCTTCAA	AGCCGGCCCACTACTAGTTT	78	752	60.117	60.151	675
FPC_Contig89	CccABb-84J15	1	CccaBA_B-Rearray-1c5-003-M4JR	8	759	CCTCCCAGTGCTTAAATTGG	TCGCTTGAATAACATCACGC	35	712	59.564	59.839	678
FPC_Contig9	CccABa-23P18	2	Cc_002_07105_Dec09	29	711	GCTCCTTAAGCTACCTCTCCC	AAAACGCACAGGGACAATG	92	677	58.637	59.556	586
FPC_Contig9	CccABa-28I02	3	Cc_003_10617_Dec09	1	798	ACTAGTGGAAGTGGCTCGGA	ATTGGAGCTCTTGGTTGAGC	66	747	59.867	59.434	682
FPC_Contig9	CccABa-50M09	4	Cc_004_10225_Dec09	24	1036	TGGAATATGGGGTATGGGTG	TCGACTTCGTCTTCCTGGAT	82	958	60.266	59.803	877
FPC_Contig9	CccABa-52F18	1	CccaBa052-F18TR	13	654	GGCATATTTGGAATCTTTTGGA	GGCATAGGAAAGACCACCCT	80	653	60.148	60.328	574
FPC_Contig9	CccABa-52F18	1	CccaBa052-F18TV	4	686	CAGGTCAAATTCATGCACCA	TTTGCTGTTCAGCCATTCTG	42	639	60.517	59.988	598
FPC_Contig9	CccABb-28J16	1	CccaBb028-J16JR	5	825	AGCTTCCGAGAATGGTCAAA	CTGAGGGTCACTCGGTCAAT	34	743	59.813	60.112	710
FPC_Contig92	CccABa-81D06	1	cccaBa-Rearray-4-3G03TR	28	473	TTATGGAACTCAAGGGCCTC	TTCCTTGTGCATCCATGTTG	26	419	59.126	60.517	394
FPC_Contig94	CccABa-46M2	2	Cc_002_10398_Dec09	6	800	TTTGTTGGCAAGGAAAAAGG	TCGGAATCATACCCTTCACC	73	719	60.081	59.75	647
FPC_Contig95	CccABa-04C13	2	Cc_002_10619_Dec09	5	792	TTTGACCAAAGGAGATAGTTGGA	ATTTCGTCCCCGAAATTACC	40	673	59.987	60.02	634

-	-	2	Cc_002_00516_Dec09	5	1183	GGATGGGACTTCCAAACTCA	ACTGGTTGCCTCAAATCGAC	106	1082	59.903	60.119	977
-	-	2	Cc_002_00908_Dec09	3	858	TGAAGGAGTTGCCACAGCTA	TCCAACAAGATTCCTCAAGGTT	7	700	59.591	59.976	694
-	-	2	Cc_002_00920_Dec09	23	889	TGTCTCGTCCCAAGCTAACA	ATTGAACCCTCCGCTGTATG	48	774	59.44	59.955	727
-	-	2	Cc_002_01912_Dec09	3	972	TGATGAAGGGATCTTCCCAG	CAGTCACCACCACTGCAACT	26	926	60.003	59.782	901
-	-	2	Cc_002_02074_Dec09	2	840	TGGAATGAAAGAACTTCGGG	TGGGAAATGCTTTCTCAGGA	18	811	60.044	60.713	794
-	-	2	Cc_002_02295_Dec09	1	797	CGGGTGTCAATTTTTGTCAG	AAAACATGTTCGACCATTAATTACAA	19	700	59.021	59.953	682
-	-	2	Cc_002_02547_Dec09	2	979	TTTGTGCTGAAAGAGATGGCT	TCTGAGATAAAAGACCGCCG	13	795	60.008	60.344	783
-	-	2	Cc_002_02644_Dec09	3	872	TGCACATCAAAAGTTGCTCC	TTTCAATGTTCTAGGGATGGAAAT	14	836	59.847	60.076	823
-	-	2	Cc_002_02801_Dec09	17	1244	TTGCCTTTGTTGTTGTGGAA	GCCCATATTGAAGGCACCTA	4	1097	60.127	59.923	1094
-	-	2	Cc_002_03447_Dec09	8	822	CCTGGAATTCCTACAACGGA	TGCACTCACCACAAACACAA	148	806	59.926	59.751	659
-	-	2	Cc_002_04272_Dec09	10	1293	GGAATGGAAGATGGCTATGC	CATCTTCGTCCAAAGCATCA	84	1124	59.491	59.799	1041
-	-	2	Cc_002_04374_Dec09	1	712	CTAGTGTGCAACCGCTTCAA	CGAAGTCTGCATGCTGAGAA	113	696	60.05	60.287	584
-	-	2	Cc_002_04693_Dec09	4	749	GAGAGAGAGGAAGGAAGGTGG	TTTAAGCTAGATGCCCACATCA	111	709	59.425	59.742	599
-	-	2	Cc_002_04838_Dec09	1	707	TCGGTTGAATGCATGAAAAG	AAACAGGCCAAACCAACCTA	114	695	59.664	59.471	582
-	-	2	Cc_002_05156_Dec09	6	890	AAAAATGGCCCTTGTATTTGG	TGGTTACTCAAGGGTTAGTGAGC	41	864	60.062	59.703	824
-	-	2	Cc_002_05563_Dec09	7	859	GAATGAGCCCGAACTTCAAT	AACTGCCAAGTTTTGAAGGC	18	806	59.129	59.359	789
-	-	2	Cc_002_05882_Dec09	1	915	GTTGAAAAAGCCGAAGCATT	CGGCCCACTGCTAGTTTTAG	38	782	59.343	59.898	745
-	-	2	Cc_002_06381_Dec09	12	839	AGTGTCCTTACGGTTGGTGC	TGCATGCGTTTAGGCAATAA	112	835	60.035	60.232	724
-	-	2	Cc_002_06573_Dec09	17	648	TCTTGCCTAAAGTACCCTAAATGG	AGGGGACTCCACTAAGGCAT	89	613	59.925	59.957	525
-	-	2	Cc_002_07672_Dec09	5	833	ACTCAAGCTTAGCGTGGCAT	TTGTTGGGGTTTGACTTCACT	14	682	60.044	59.482	669
-	-	2	Cc_002_08247_Dec09	4	1020	GCAATTTGTGCCTATCCCAC	AGTTGGCTTATGGTGCAAGC	64	956	60.339	60.278	893
-	-	2	Cc_002_09471_Dec09	11	841	ATGCAAGCTTCTACGGCAAG	GAATCACACCTAAACATTCTCATACG	22	698	60.543	60.182	677
-	-	2	Cc_002_09767_Dec09	9	671	AGATGAGGTTGTGGGCTTTG	TGCCTTGTTACAGAACAATCTGA	58	596	60.111	59.799	539
-	-	2	Cc_002_10155_Dec09	12	919	TGCAAGATGAAGAGAATTAGACAAA	GATGCTCAAAACCATCAAAGG	97	902	59.465	59.56	806
-	-	2	Cc_002_10162_Dec09	25	773	CTGTGATTGCCGTTTGTCTC	CTCCCTCCCAACCCAATAAT	72	713	59.293	60.011	642
-	-	2	Cc_002_10176_Dec09	10	638	TGAAACAATGCTAAAGAACAAGGA	ATTAAGGGGAGGGAAGCTCA	74	598	60.16	60.032	525
-	-	2	Cc_002_10198_Dec09	7	1230	AAAGGCACCAACATCAATCC	GCCACAGGTTGATTATGGAAA	48	1038	59.797	59.815	991
-	-	2	Cc_002_10211_Dec09	8	828	GTGTGGGGAGAAGGGTGTAA	CCCAAAAATTACAAATACATGGC	2	691	59.82	59.529	690
-	-	2	Cc_002_10212_Dec09	4	702	GACTCCTTTTGAAGTGTTTGCTTT	ACTCCCAAAAAGGCCTTCAC	54	636	60.197	60.476	583
-	-	2	Cc_002_10214_Dec09	7	842	CCGTCGTCTTCCTCTACTGG	TAACACCACCGTTCCCTTTC	19	731	59.861	59.83	713
-	-	2	Cc_002_10233_Dec09	11	800	TCTTGAAAATTCTTGTTGTTTCATTT	TGGTGATACAGCAGGGATGA	34	761	59.516	60.072	728
-	-	2	Cc_002_10239_Dec09	7	866	AGGCATGCAAGCTTGTTGTA	TAAAGGCATGATGGTGGTGA	18	753	59.493	59.924	736
-	-	2	Cc_002_10261_Dec09	48	792	TTTTCGGAATCTGTGCCTTT	CTCCAAGGTCGATTGCATTT	23	750	59.685	60.074	728
-	-	2	Cc_002_10277_Dec09	5	606	CGCTTCCCTCAATATTTGCT	ATTGTCGGCAATGTCATTCA	41	549	59.32	59.931	509
-	-	2	Cc_002_10302_Dec09	4	808	CTTTTGATTGTGATGCCCCT	AAACAAGTCAACCATATGCACG	74	779	59.933	59.926	706
-	-	2	Cc_002_10308_Dec09	19	840	AGAGCATCATGGTCCTGACA	TCATGTTACTCAGCACATCACACT	24	812	59.212	59.731	789
-	-	2	Cc_002_10317_Dec09	4	779	GGCCAGTGGACTCTGTCAAT	GGCGTCAGAATTGGCATTAT	44	755	60.12	59.929	712
-	-	2	Cc_002_10338_Dec09	12	847	TATGCCATTGCTTCTGTTGG	TTCTTAAACATGCCACACGC	79	801	59.688	59.735	723
-	-	2	Cc_002_10373_Dec09	3	812	GCATGCAAGCTTCAAAACAA	CTTCAATGTGTCTCCGTCCC	20	740	59.999	60.51	721

-	-	2	Cc_002_10377_Dec09	13	917	CAGGCATGCAAGCTTAGAGTT	TTGCATATCCCATTGTGTGG	16	820	59.668	60.197	805
-	-	2	Cc_002_10400_Dec09	8	831	CCATCGATAGCTTGCTCACA	ATGCTGTGAAAGACGCATAAA	122	830	59.972	58.452	709
-	-	2	Cc_002_10416_Dec09	6	736	CTTGACCCAAAATTGCCCTA	CCCTAGCTCAATACCTGCACA	89	686	59.931	60.277	598
-	-	2	Cc_002_10456_Dec09	40	740	CAACTCCCCGTTCACAATCT	TGAGTGCATGTCTCGGTGAT	0	661	59.966	60.279	662
-	-	2	Cc_002_10458_Dec09	15	863	TGTATTGGTTTGTGGGGACA	GTCGCGTTTAGTGAGGCATT	94	854	59.667	60.278	761
-	-	2	Cc_002_10470_Dec09	26	783	ATGTGAAAATCTACCCACCAAA	CCCCCTCCATGGATACAAAT	70	705	58.353	60.767	636
-	-	2	Cc_002_10479_Dec09	2	959	CAGGAAGTTGCACCTTGTCA	TGGAATTAGGAGCCTTGCAT	82	901	59.873	59.668	820
-	-	2	Cc_002_10482_Dec09	3	752	TCCTAGATGCAAATGAGGCA	TAATGGGTCAAGTCAAGGGG	120	731	59.375	59.784	612
-	-	2	Cc_002_10518_Dec09	3	569	GCACGTAGCATGTGAGATCC	GAAACAGAATGTGGCCAAGC	6	531	59.276	60.646	526
-	-	3	Cc_003_02793_Dec09	8	732	TGCAAGACTCTTCTTGTGCAAT	GAGGCTTCGATGTTAGTCCG	2	644	60.059	59.836	643
-	-	3	Cc_003_03028_Dec09	22	685	GCAAGCTTCCATAGGGATCA	ACACTGCTCAAAAACCCCAG	24	607	60.177	60.149	584
-	-	3	Cc_003_03134_Dec09	7	919	TTACACCTGGCCTCCCTATG	CATTTGCAGGTGCACATTTT	86	902	59.948	59.59	817
-	-	3	Cc_003_08205_Dec09	1	888	CCAACCAATTGCCTCATTCT	AGTGACTCAGTGGGGTGGAC	119	860	59.933	60.006	742
-	-	3	Cc_003_10179_Dec09	5	1016	GGGAATTTCACACCTCCAAA	AGTTGGGACAAATGAGTGGG	91	1005	59.767	59.82	915
-	-	3	Cc_003_10184_Dec09	36	860	GCGATTGAAAAACTTTTGGG	TTTGGCGCCATAAAGAAAAT	70	836	59.564	59.565	767
-	-	3	Cc_003_10659_Dec09	3	902	CATAGGGTGCACTTTCCACA	AACACACTTCTAGGTGGGCG	30	824	59.566	60.171	795
-	-	3	Cc_003_10688_Dec09	3	702	ACACTCCCTTTGGAGCAATG	CCCGCTACAATTCCTCTTCA	14	589	60.111	60.206	576
-	-	4	Cc_004_00991_Dec09	7	806	CCAACCTCTCAGCACGGTAT	GGACATGAGTGGAAAGCCAT	76	778	60.134	59.934	703
-	-	4	Cc_004_03020_Dec09	1	659	GGCATGCAAGCTTCTTCTTC	AGCTGCAAGTTTCTGCCACT	19	620	60.103	60.201	602
-	-	4	Cc_004_10148_Dec09	5	1688	CATCAATGTTGCAAATGAAATTAG	CTTTTCAAACTTTTGCGATGC	94	1463	58.614	59.882	1370
-	-	4	Cc_004_10347_Dec09	9	1220	CCGGTTCTTCTGTTTTCCCT	AAGGATGCAATTCATCGGAG	156	1184	60.467	60.036	1029
-	-	5	Cc_005_00413_Dec09	1	1264	ACCAATGTACACCTCTCGGC	TGGTTAGGAGGAGCTAAGGAGA	252	1263	59.997	59.499	1012
-	-	5	Cc_005_07563_Dec09	5	914	TAGCTTGCTCTCCAAGCCTC	GGAGTCCAAAGGGAGATAAAAGA	67	827	59.859	59.958	761
-	-	1	CccaBA_B_Rearray_1c5_001-B17TR	17	716	AAGCTTGATCTTGTGGGGG	GTTGTCTACTTCAATGCAACCCT	0	691	60.058	59.578	692
-	-	1	CccaBA_B_Rearray_1c5_001-G7TV	9	792	TACTGCCGTTCCCCACTAAG	GGATGAGTGAATTGTGTGTGC	43	789	60.125	59.002	747
-	-	1	CccaBA_B_Rearray_1c5_001-L21TV	2	675	TAGAGTCGACCTGCAGGCAT	TCAAGCCCAAGGTATTCTTCA	3	631	60.966	59.693	629
-	-	1	CccaBA_B_Rearray_1c5_001-M6TR	1	771	AGCTTGATGCGGTAATGGAG	AACCTTTCTTCCCCTTTGGA	20	686	60.235	59.912	667
-	-	1	CccaBA_B_Rearray_1c5_001-P23TR	14	710	AAAAGCAACCATGCAGACATC	CTGATAAGCGCACCAATGAC	93	696	60.133	59.301	604
-	-	1	CccaBA_B_Rearray_1c5_002-C2TR	10	593	CAAGCTTCGTGCATAAGTGC	TGATCATCAAGACGGAGGTG	5	567	59.638	59.631	563
-	-	1	CccaBA_B_Rearray_1c5_002-G15TV	15	913	TAGAGTCGACCTGCAGGCAT	TGGAAATCTAATCCCCAAGGA	4	745	60.966	60.619	742
-	-	1	CccaBA_B_Rearray_1c5_002-G18TR	1	612	TTTTGTGTGTGAGGGTTGGA	TTCTTAGGCGCAGCAAAAAT	42	537	59.976	59.985	496
-	-	1	CccaBA_B_Rearray_1c5_002-G18TV	4	715	TGGATGCTTCCATCCATAGTC	AAAGCAAAAACGTACCAAACAAA	31	602	59.909	59.963	572
-	-	1	CccaBA_B_Rearray_1c5_002-G2TV	14	632	TGTTTGATGAATTGGTCGTGA	TTGTCCCAAAGTCCACAAGTC	60	591	59.956	59.997	532
-	-	1	CccaBA_B_Rearray_1c5_002-O20TR	2	718	TGGACACCTAATCACCATTTGA	CAATCAATCTCTTGCCACCA	63	666	60.227	59.648	604

-	-	1	CccaBA_B-Rearray-1c5-003-G1JR	3	838	CGGTCCGCGAAATAAATAGA	GGGGTGGGTATATGGACTTG	104	826	60.054	58.981	723
-	-	1	CccaBA_B-Rearray-1c5-003-J11JR	7	773	GGATGGGCAAAATGGAAGTA	TGGCTAAAATTGTCACCGAA	36	693	59.762	59.157	658
-	-	1	CccaBA_B-Rearray-1c5-003-J13JR	5	763	CCTGACAGATGGAATCCAGAA	TTTGGTATGTTTTTCCCATGC	136	747	60.058	59.69	612
-	-	1	CccaBA_B-Rearray-1c5-003-K14JR	6	767	CCAACGTTGTGGAAGAGGTT	AATGACCTAACATGCAGCCA	63	762	60.005	59.152	700
-	-	1	CccaBA_B-Rearray-1c5-003-L4JR	5	776	CAAGTGTTTTTGCAGAAGGC	AGGCCTAACTCAACCTTATAAAACA	39	691	58.554	58.802	653
-	-	1	CccaBA_B-Rearray-1c5-003-M14JF	8	888	CCCAAACGTTTTCTTCTCCA	GAGGAATGTTACGCATGTTTTG	26	822	60.081	59.519	797
-	-	1	CccaBA_B-Rearray-1c5-003-M14JR	4	880	AGATGGCTTGTGCTCCTTGT	ATTTTAGGCATGGCAATCGT	44	825	59.874	59.438	782
-	-	1	CccaBA_B-Rearray-1c5-003-M8JR	11	831	AAGAGGAGAAAGAGAAACGATTAGC	TGAATTTTTCCTTAACGCGG	2	750	59.945	60.068	749
-	-	1	CccaBA_B-Rearray-1c5-003-N23JR	20	838	CCTTTCCCTGGTGTTTTGTC	AGTTTGCCCTGAAGGCATTA	46	817	59.425	59.708	772
-	-	1	CccaBA_B-Rearray-1c5-003-O8JR	3	822	CTAAGGGGTTAGCCCACACA	TCCCAGAGCATCAAGAATCC	51	778	59.986	60.158	728
-	-	1	cccaBa-Rearray-4-2A04TR	5	472	TATTGAAACTTTGTGCCCCA	TCCATGATGCATTCAAGATTT	4	432	59.013	58.023	429
-	-	1	cccaBa-Rearray-4-3H03TV	9	732	TCCAAGTCACAACTGACCCA	CCCTATTCTCTAGTAGAAGCCCAAC	113	705	60.129	60.053	593