

Chapter 1

General Background

1.1 Introduction to molecular systematics

Living organisms are all connected by ancestor-to-descendent relationships. Phylogeny, defined as the historical relationships among lineages of organisms, attempts to reveal the ancestral and evolutionary relationships of species or higher taxonomic groups of organisms (Williams and Ebach 2008). In the past, taxonomists have grouped organisms based mainly on morphological characters. Nowadays, DNA sequences are widely used to infer phylogenies of organisms. Morphological (or anatomical) and DNA sequence data are used to classify biological living organisms through phenetics or cladistics approaches. Phenetics advocates grouping by overall or raw phenotypic, anatomical or molecular similarities. Thus, phenetics recognizes taxa on the basis of percentage similarity, for the most part ignoring other phylogenetic considerations. This approach is used in my study and shown in Chapter 5. Whereas cladistics, which is among others employed in Chapter 3, advocates grouping by shared derived characters (Williams and Ebach 2008).

Amino acid or nucleotide sequences are the most widely and frequently used markers of making phylogenetic inferences, called molecular phylogenetics (also called molecular systematics). The internal transcribed space (ITS) region of nuclear ribosomal DNA is one of the most important molecular markers that have been used in numerous systematic studies, including in Chapter 3 of this dissertation. It consists of two internal spacers, ITS1 and ITS2 (Figure 1.1), located between genes encoding the 18S, 5.8S and 26S nuclear ribosomal DNA (nrDNA) subunits (Baldwin 1992). DNA sequences of the ITS region are preferred in phylogenetic analysis due to several reasons. The ITS region is highly repeated in plant nuclear genomes and its high copy number and high degree of homogeneity facilitates easy amplification and sequencing of the nrDNA from small quantities of DNA. It is also relatively small and is flanked by highly conserved sequences, the 18s and 26s nrDNA genes which aid the use of universal primers to amplify and sequence the ITS regions in most plant families (Baldwin et al. 1995). The ITS region within individuals of a single species mostly reveals little or no sequence divergence. On the other hand, alignments from different species, and although not always, between closely related species show good level of sequence divergence (Baldwin et al. 1992, Baldwin et al. 1995). These characteristics, especially the low-intra species sequence variability and inter-species sequence divergence make ITS one of the candidate genes for DNA barcoding, an

approach to determine species by specific DNA sequences (Stoeckle 2003, Kress et al. 2005, Chen et al. 2010).

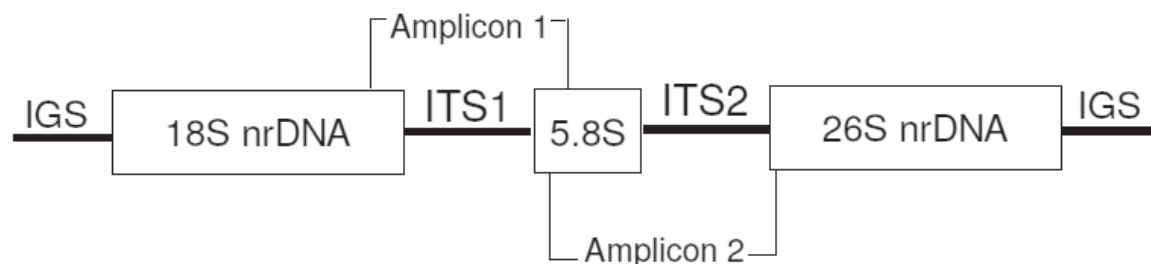


Figure 1.1. The internal transcribed spacer region of the nuclear ribosomal DNA.

1.2 Population genetics and diversity studies in plants

Population genetics is a branch of genetics which deals with the frequency of alleles responsible for genetic relationships or variations within populations or at species level. It considers the factors that determine the population size, the evolution of a population, such as natural selection, genetic drift, mutation, recombination, and gene flow (Halliburton 2004). Through population genetics tools, it is possible to study the naturally occurring genetic differences among organisms. Studying the level and patterns of genetic diversity is useful in several ways. At the most basic level, genetic diversity in a species is the raw stuff on which evolutionary processes are acting (May 2000). From a biodiversity point of view, genetic diversity studies help to generate information on origin and maintenance of the global set of species. Knowledge of the amount and distribution of genetic diversity is among the major criteria in locating sites for genetic reserves (Iriondo et al. 2008). Genetic variability in these reserves can also contribute in identification of genetic traits that foster tolerance to biotic or abiotic stresses.

In the field of population genetics, the level of genetic diversity of a given organism is usually investigated using molecular or morphological approaches. One of the methods frequently used in the molecular approaches, used in my study in Chapter 5, are amplified fragment length polymorphisms (AFLP) (Vos et al. 1995), which are used in several research areas, including the inference of phylogenetic relationships, investigation of genome wide genetic variation, individual ancestry, population structure, and in identifying hybrid individuals (Bensch and Akesson 2005, Falush et al. 2007, Pleines and Blattner 2008).

Morphological characterization is also used to investigate diversity of germplasm collections and contributes valuable information for agronomic and breeding programs (Chapter 5). In the field of plant systematics, morphometric analysis is also used to investigate morphological evolution, which is useful for

addressing species-level taxonomic questions. An understanding of evolutionary processes at the infraspecific level, and ultimately of speciation, requires the accurate characterization of patterns of variation within species (Gould and Johnston 1972, Thorpe 1976). One way of doing this is through multivariate morphometric techniques, which have the power to reveal complex patterns of population differentiation related to geographic divergence and evolutionary processes such as gene flow, selection and genetic drift (Gould and Johnston 1972).

1.3 Genome size variation in plants

The amount of DNA, including the non-coding DNA present within one copy of a genome of an organism is known as genome size (Gregory 2002). It is usually expressed in base pairs (bp) after converting the nuclear DNA content measured in picograms (pg). Nowadays, flow cytometry is the most widely used method of estimating nuclear DNA content because it is convenient, fast and reliable (Doležel et al. 2007). Genome size differences among species has been a puzzle in genetics for many decades, as relatively simple organisms could have 1000 times more nuclear DNA than more complex multicellular organisms. This phenomenon is known as the C-value paradox or C-value enigma (Gregory 2002), and could be explained by large fractions of non-coding DNA, which contribute to the total DNA content without increasing gene numbers. However, probable functional properties of non-coding DNA in living organisms are still an open question.

Estimation of nuclear DNA content can help to investigate ploidy level, which is useful in population biology, breeding, and quality control in seed production (Suda et al. 2007). Plants exhibit remarkable variation in genome size (Bennett et al. 2000, Bennett and Leitch 2005, Greilhuber et al. 2007). As a result, genome size has also been found to be a useful taxonomic marker and is employed in the field of systematics of various plant groups (Dimitrova et al. 1999, Mishiba et al. 2000, Zonneveld 2001, Doležel et al. 2007). This dissertation includes results of flow cytometric investigation of genome size in jute species, and its association with phenotypic traits and elevation is discussed in Chapter 4.

1.4 Systematics, distribution, and economic importance of jute

The genus *Corchorus* was first described by Linnaeus (1753) who recognized four species, namely *C. olitorius*, *C. capsularis*, *C. siliquosus*, and *C. hirsutus*. Later, several species belonging to this genus were identified and described by various authors. The genus was formerly classified under the traditional family Tiliaceae. However, this family is no longer recognized by the Angiosperm Phylogeny Group since 1998 (APG I 1998, APG II 20003, APG III 2009). Based on molecular evidence of chloroplast genes, *Corchorus* was included in the subfamily Grewioideae, which belongs to the family Malvaceae s.l. of the order Malvales (Alverson et al. 1999, Bayer et al. 1999, Whitlock et al. 2003). A reclassification of the broader Malvaceae was proposed by Heywood et al. (2007) and *Corchorus* was described within the family Sparrmanniaceae. However, this classification is not recognized by the Angiosperm Phylogeny Group (APG III 2009), thus the genus *Corchorus* at present maintains its phylogenetic position within the subfamily Grewioideae of the family Malvaceae.

The genus *Corchorus*, commonly known as jute, consists of shrubs, subshrubs, and annual and perennial herb species. Average plant height varies from ca. 0.09 m to over 4 m, the smallest recorded in *C. depressus* and *C. pumilio*, whereas the tallest are in the two cultivated species *C. olitorius* and *C. capsularis*. The leaves, mostly with palmate veins and serrated margins are simple, alternate and the shape varies from lanceolate, elliptical, ovate to palmate. Most species are morphologically similar and identification at vegetative phase is extremely difficult. Although shared by several species, a careful examination of the presence or absence, and the color of stipules are few of the morphological characters that help identify species at vegetative stage. The presence or absence of setae, a tail-like structure at the base of the blade, can also aid in species identification during the vegetative or reproductive stage (Figure 2.4). Stems are well developed with abundant fibres in the phloem tissue, which make several species a source of natural fibre, jute. Flowers (3-5-merous) are yellow usually with 4-5 free, acuminate sepals. Identification of *Corchorus* species is mainly possible during the reproductive stage, especially using capsule and seed morphology. Fruit (except in *C. capsularis*) is a straight or slightly curved, cylindric or spherical capsule that terminates in a beak with 2-7 valves (Figure 1.2). Seeds are mainly angular or ellipsoid in shape and the color varies from black, brown to gray (Figure 1.3). Most *Corchorus* species are diploids with $2n = 14$, while few species, namely *C. cunninghamii*, *C. hirtus*, *C. junodii*, *C. orinocensis*, and *C. siliquosus* are tetraploids (Coleman 1982, Halford 1995, Islam et al. 1975, Rao and Datta 1953). Hexaploid ($2n = 42$) and aneuploid ($2n = 13$) species were also

reported (Bayer and Kubitzki 2003). The ploidy levels of several other species (for example *C. africanus*, *C. angolensis*, *C. brevicornutus*, *C. cinerascens*, *C. pilosus*, *C. urticifolius*, etc.) remain unknown.

The first and preliminary sectional classification of *Corchorus* was proposed by De Candolle (1824), and it included about 19 species within five sections (*Coreta*, *Coretoides*, *Seratocoreta*, *Janya*, and *Guazumoides*). However, several species were left undescribed and *C. depressus* was treated under the former genus *Antichorus* as *A. depressus* L. (Anonymous 1767). Moreover, of the 19 species that were included under the five sections, several were presently identified as synonymies. As a result, the actual numbers of species described under the five sections are at present less than nineteen (*C. aestuans*, *C. asplenifolius*, *C. capsularis*, *C. fascicularis*, *C. hirsutus*, *C. hirtus*, *C. olitorius*, *C. orinocensis*, *C. pilosus*, *C. siliquosus*, *C. tridens*, and *C. trilocularis*). Currently, the number of *Corchorus* species listed varies from 40 to 100 (Wild 1984, Edmonds 1990, Heywood 1993, Bayer and Kubitzki 2003, Kubitzki and Chase 2003). However, only 68 species are listed with accepted names, leaving the rest with synonymy (Hinsley 2004). Africa accounts for the majority (30 to 40) of world *Corchorus* species. Out of these, 15 are reported in South Africa, 13 in Ethiopia and Tanzania, each (Edmonds 1990, Benor et al. 2010). Thus, much of the species distribution is in the horn/eastern to southern parts of Africa. About 26 to 35 species are also reported in Australia (Anonymous 2004, Halford 2004). The number of species in Asia and tropical America is not more than 10 (Howard 1989, Mahapatra et al. 2003). Thus, the genus is widely distributed throughout the tropics, subtropics and warm temperate regions of the World with the majority of the species found in Africa and Australia. Detailed information on main geographical distribution and area of endemism of *Corchorus* species is shown in Table 3.1.

Jute was originally used as a pot herb and medicinal plant especially in Africa, Middle East, and East Bengal (Kundu 1951, Burkill 2004). Ancient Greeks used to call a pot herb as “*Korkhoros*”, from which the generic name *Corchorus* is derived (Royle 1855). Young shoots of several annual herb species are used as vegetable crop, and are sources of household income generation in several African countries (Edmonds 1990, Krebs 2001, Grubben and Denton 2004, Kinabo et al. 2006, Dansi et al. 2008, Benor et al. 2010). Later, the increased demand of fibres by the European markets led to a discovery of jute as an inexpensive source of natural fibre, which changed its principal use from food to fibre in most Asian countries (Kundu 1951, The Natural History Museum 2007).



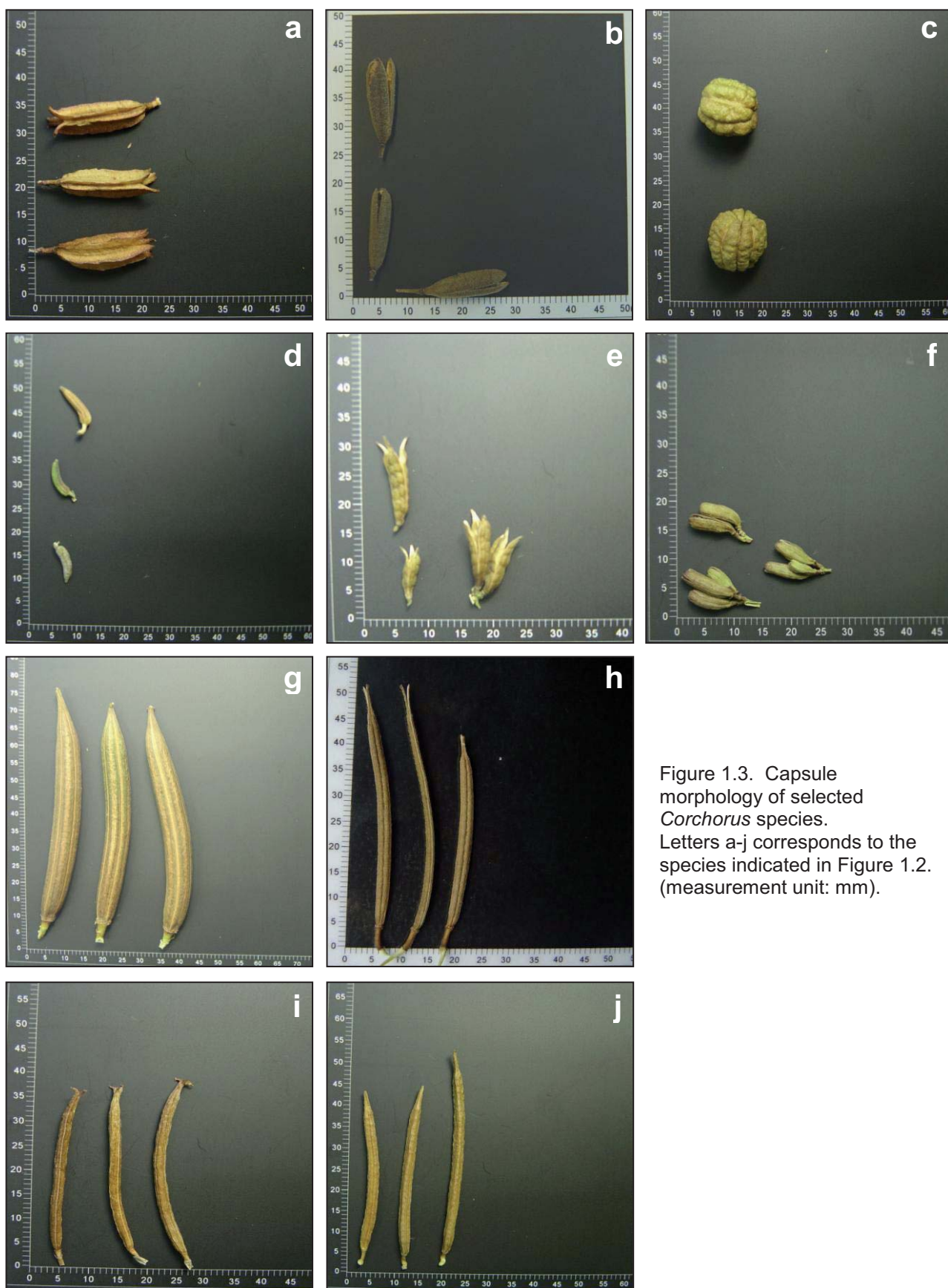


Figure 1.3. Capsule morphology of selected *Corchorus* species. Letters a-j corresponds to the species indicated in Figure 1.2. (measurement unit: mm).

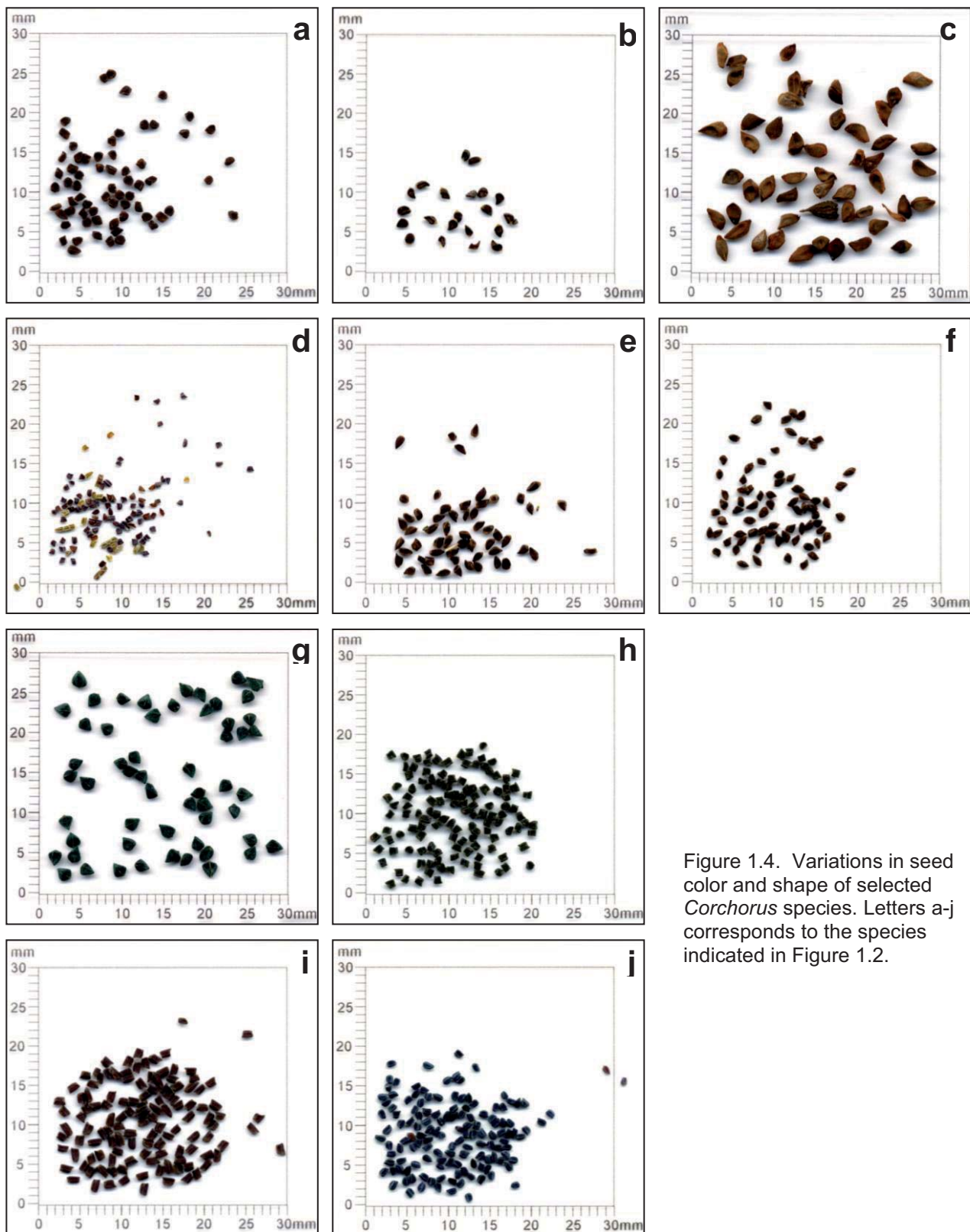


Figure 1.4. Variations in seed color and shape of selected *Corchorus* species. Letters a-j corresponds to the species indicated in Figure 1.2.

As a result, jute was introduced to Europe early in the 18th century, grown in Brazil since 1932, and several fibre industries were established in the Americas in the mid 20th century (The Natural History Museum 2007). Currently, *C. olitorius* and *C. capsularis* rank next to cotton as the most important natural fibre in Asia (Roy et al. 2006, Taliaferro et al. 2006). The use of *Corchorus* species as leafy vegetable, source of fibre, and the potential in other industrial sectors is more broadly discussed in Chapter 2.

1.5 Rationales and objectives of the study

Nowadays, biodiversity loss and species extinction in response to climate change and human impact is becoming a hot topic in scientific-related political discussions. Thus, research in the area of population genetics, particularly assessment of genetic variability will help in planning conservation measures for species threatened by biotic or abiotic factors. This is paramount, especially in tropical taxa, like jute, which are left with little research attention. Conserved genetic resources can be used as a source of genetic variation in breeding programs. Characterization and further studies on the genetic variability of germplasms also help to avoid storage of duplicate samples in gene banks. From evolutionary point of view, population genetics and phylogenetic investigations help to answer various questions about origin, gene flow, speciation, and domestication of living organisms. Such studies can also generate useful information on genes that are associated with tolerance to biotic and abiotic stresses, which in turn can play a major role in agricultural, pharmaceutical and other industrial sectors.

Although the systematic position of the genus *Corchorus* at family level is settled (APG I 1998, APG II 2003, APG III 2009), its historical and evolutionary relationship within the genera of the subfamily Grewioideae is poorly understood. Moreover, the evolutionary relationships among species of the genus *Corchorus*, and the relationship of the cultivated species to the wild *Corchorus* taxa are scarcely known. Several *Corchorus* species are reported in Africa in general and in Ethiopia in particular, the latter being the second in the total number of species in the continent (Benor et al. 2010). Despite its abundance in Ethiopia, there is little investigation on ecological distribution, indigenous knowledge, and status of species utilization in the country prior to this study.

Thus, the rationale and overall goal of the PhD project is to investigate population genetics, genomics, and systematics research gaps in jute. The general objectives were to (1) study the adaptation and ecological distribution of the species in Ethiopia, (2) collect, characterize, and assess the indigenous knowledge of

farmers with respect to the species use, (3) investigate the genetic diversity and population structure of *C. olitorius* at morphological and molecular levels, (4) investigate the origin and routes of dispersal of *C. olitorius*, (5) estimate genome size variation in *C. olitorius* and investigate its correlation with phenotypic traits, elevation, and interpret the results with morphological and molecular diversity studies, and (6) infer the phylogeny of the genus *Corchorus* using all available species from its pantropical distribution.