2003

Polyploidy and the Evolutionary History of Cotton

Jonathan F. Wendel
Iowa State University, jfw@iastate.edu

Richard Clark Cronn
United States Department of Agriculture

Polyploidy and the Evolutionary History of Cotton

Abstract
The cotton genus (Gossypium) includes approximately 50 species distributed in arid to semi-arid regions of the tropic and subtropics. Included are four species that have independently been domesticated for their fiber, two each in Africa–Asia and the Americas. Gossypium species exhibit extraordinary morphological variation, ranging from herbaceous perennials to small trees with a diverse array of reproductive and vegetative characteristics. A parallel level of cytogenetic and genomic diversity has arisen during the global radiation of the genus, leading to the evolution of eight groups of diploid (n ¼ 13) species (genome groups A–G, and K). The evolutionary history of the genus included multiple episodes of trans-oceanic dispersal, invasion of new ecological niches, and a surprisingly high frequency of natural interspecific hybridization among lineages that are presently both geographically isolated and intersterile. Recent investigations have clarified many aspects of this history, including relationships within and among the eight genome groups, the domestication history of each of the four cultivated species, and the origin of the allopolyploid cottons. Data implicate an origin for Gossypium 5–15 million years ago (mya) and a rapid early diversification of the major genome groups. Allopolyploid cottons appear to have arisen within the last million years, as a consequence of trans-oceanic dispersal of an A-genome taxon to the New World followed by hybridization with an indigenous D-genome diploid. Subsequent to formation, allopolyploids radiated into three modern lineages, including those containing the commercially important species G. hirsutum and G. barbadense. Genome doubling has led to an array of molecular genetic interactions, including inter-locus concerted evolution, differential rates of genomic evolution, inter-genomic genetic transfer, and probable alterations in gene expression. The myriad underlying mechanisms are also suggested to have contributed to both ecological success and agronomic potential.

Keywords
Gossypium, diversification, allopolyploids, genome duplication

Disciplines
Botany | Genetics | Plant Breeding and Genetics

Comments
This article is from *Advnces in Agronomy* 78(2003): 139.

POLYPLOIDY AND THE EVOLUTIONARY HISTORY OF COTTON

Jonathan F. Wendel[1] and Richard C. Cronn[2]

[1]Department of Botany, Iowa State University, Ames, Iowa 50011, USA
[2]Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, Oregon 97331, USA

The cotton genus (*Gossypium*) includes approximately 50 species distributed in arid to semi-arid regions of the tropic and subtropics. Included are four species that have independently been domesticated for their fiber, two each in Africa–Asia and the Americas. *Gossypium* species exhibit extraordinary morphological variation, ranging from herbaceous perennials to small trees with a diverse array of reproductive and vegetative characteristics. A parallel level of cytogenetic and genomic diversity has arisen during the global radiation of the genus, leading to the evolution of eight groups of diploid ($n = 13$) species (genome groups A–G, and K). The evolutionary history of the genus included multiple episodes of trans-

Advances in Agronomy, Volume 78

0065-2113/02$35.00

oceanic dispersal, invasion of new ecological niches, and a surprisingly high frequency of natural interspecific hybridization among lineages that are presently both geographically isolated and intersterile. Recent investigations have clarified many aspects of this history, including relationships within and among the eight genome groups, the domestication history of each of the four cultivated species, and the origin of the allopolyploid cottons. Data implicate an origin for *Gossypium* 5–15 million years ago (mya) and a rapid early diversification of the major genome groups. Allopolyploid cottons appear to have arisen within the last million years, as a consequence of trans-oceanic dispersal of an A-genome taxon to the New World followed by hybridization with an indigenous D-genome diploid. Subsequent to formation, allopolyploids radiated into three modern lineages, including those containing the commercially important species *G. hirsutum* and *G. barbadense*. Genome doubling has led to an array of molecular genetic interactions, including inter-locus concerted evolution, differential rates of genomic evolution, inter-genomic genetic transfer, and probable alterations in gene expression. The myriad underlying mechanisms are also suggested to have contributed to both ecological success and agronomic potential.

I. INTRODUCTION

One of the most remarkable stories in the annals of crop domestication is the origin of cultivated cotton. Perhaps the most striking aspect of this history is that it is global in scope, involving ancient human cultures in both the Old and New Worlds and a convergent or parallel plant domestication process from divergent and geographically isolated wild ancestors. Indeed, cotton is unique among crop plants in that four separate species were independently domesticated (Brubaker *et al.*, 1999a; Brubaker and Wendel, 1994; Percy and Wendel, 1990; Wendel, 1989; Wendel *et al.*, 1992: Wendel *et al.*, 1999) for the specialized single-celled trichomes, or fibers, that occur on the epidermis of the seeds. This parallel domestication process involved four species, two from Americas, *Gossypium hirsutum* and *G. barbadense*, and two from Africa–Asia, namely *G. arboreum* and *G. herbaceum*. In each of these four cases, aboriginal people discovered several thousand years ago that the unique properties of cotton fibers made them useful for ropes, textiles and other applications. As a consequence, cotton cultivation became increasingly widespread, such that over the millennia cotton became firmly established as the world's most important fiber crop and an important source of seed oil and protein meal.

Each of the four domesticated *Gossypium* species has its own unique history of domestication, diversification, and utilization. Many aspects of this history have been detailed elsewhere (Brubaker *et al.*, 1999a; Hutchinson, 1951, 1954, 1959; Hutchinson *et al.*, 1947; Wendel, 1995; Wendel *et al.*, 1999), including the various stages in the domestication process, the origin of present patterns of

genetic diversity, the shape and severity of genetic bottlenecks that accompanied the development of landraces and cultivars, and the influence of recent human history on geographic patterns of cultivation. Although all four cotton species spread far beyond their ancestral homes during the last several millennia, one species, *G. hirsutum*, recently has come to dominate world cotton commerce, having supplanted the vast majority of cultivation of the other three species. *Gossypium hirsutum* presently is responsible for over 90% of the annual cotton crop internationally, having spread from its original home in Mesoamerica to over 50 countries in both hemispheres. As the world's leading textile fiber plant, cotton forms a vital part of global agriculture and is a mainstay of the economy of the United States, as underscored by the fact that the $100 billion/year cotton agriculture and textile industry employs 300,000 Americans and ranks among the largest contributors to the US gross national product. Cotton is grown on about 12 million acres in the United States, more than all crops except maize, wheat, or soybean (Anonymous, 1997).

Because of its economic importance, *G. hirsutum* has attracted considerable scientific interest, not only among plant breeders and agricultural scientists, but also among taxonomists, geneticists, and evolutionary biologists. Concomitant with these efforts has been the recent explosion in our understanding of the organization and structure of eukaryotic genomes. This work, representing the combined efforts of innumerable investigators in diverse disciplines during the last century, has increasingly clarified the evolutionary history of cotton and its genome. Fundamental to this understanding is the realization that both *G. hirsutum* (Upland cotton) and *G. barbadense* (Pima cotton, Egyptian cotton) have polyploid genomes, resulting from a truly remarkable chance biological reunion among ancestral diploid genomes that presently are geographically restricted to different hemispheres.

One generalization that has emerged from the recent massive effort in genome sequencing and mapping in a diversity of organisms is that genome doubling through polyploidy is a prominent process in plant evolution and has played a major role in the evolution of eukaryotic nuclear genomes (Hughes *et al.*, 2000; Makalowski, 2001; Sidow, 1996; Smith *et al.*, 1999; Spring, 1997; Taylor and Brinkmann, 2001; Wolfe, 2001; Wolfe and Shields, 1997). Polyploidization has been especially active and ongoing in higher plants, with up to 70% of all angiosperms having experienced a relatively recent episode of genome doubling (Grant, 1981; Leitch and Bennett, 1997; Masterson, 1994; Soltis and Soltis, 1993, 1999; Soltis *et al.*, 1992; Stebbins, 1950, 1971). Although there are various types of polyploidy (Grant, 1981; Stebbins, 1971), the most common is allopolyploidy, whereby two differentiated genomes, usually from different species, become reunited in a common nucleus as a consequence of a hybridization event. In the simplest case, allopolyploids have one complete diploid set of chromosomes derived from each parental species, and thus contain a doubled complement of genes (homoeologues). Examples of such polyploids abound and include many of

the world's most important agricultural commodities (Hilu, 1993) including cotton.

Gossypium hirsutum and *G. barbadense* are thus classic allopolyploids, resulting from the merger of two formerly isolated diploid genomes. This history may have promoted morphological, ecological, and physiological adaptation, mediated by natural selection on a greatly enhanced level of variation resulting from an instantaneously doubled complement of genes (Fryxell, 1965, 1979; Grant, 1981; Harland, 1936; Ohno, 1970; Otto and Whitton, 2000; Soltis and Soltis, 2000; Stebbins, 1950, 1971; Stephens, 1951a,b). For the same reasons, genome doubling may have offered novel opportunities for agronomic improvement through human selection (Hutchinson *et al.*, 1947; Jiang *et al.*, 1998, 2000b; Wright *et al.*, 1998).

The evolutionary history of polyploid cottons is reviewed here, with a focus on the recent insights gleaned from the happy marriage of phylogenetic analysis with genomic investigations. Our intent is to provide a convenient entry point into a burgeoning and dispersed literature, by summarizing evidence bearing on the origin and diversification of the cotton genus and the trans-oceanic voyage that led to the origin and subsequent evolution of the allopolyploids. In addition to conveying an understanding of organismal context and history, we also discuss briefly the evolution of the modern cotton genome. One surprising and recently revealed aspect of allopolyploid plants is that their genomes need not be strictly additive with respect to the genomes of progenitor diploids. Instead, in some cases the merger of two different genomes in a common nucleus is accompanied by considerable genomic reorganization and non-Mendelian genetic behavior (reviewed by Wendel, 2000; Wendel and Liu, 2002). Although the extent and significance of the various phenomena involved are not yet clear, the myriad underlying molecular mechanisms are clearly relevant to cotton. Accordingly, we draw attention to some of the mysteries associated with allopolyploidy, with a focus on novel gene and genome interactions, as well as implications for crop improvement.

II. TAXONOMIC, CYTOGENETIC, AND PHYLOGENETIC FRAMEWORK

A. Origin and Diversification of the *Gossypieae*, the *Cotton Tribe*

The origin of polyploid cotton can be fully appreciated within the context of the evolutionary history of the genus and its tribe. For more than a century traditional taxonomic methods have been used to explore the natural affinities of *Gossypium*, and more recently this knowledge base has been supplemented by modern approaches involving comparative analysis of DNA sequences. A synthesis of these data has led to a reasonably coherent taxonomic concept

of a group of genera that are aligned into a single small tribe, the *Gossypieae*. This tribe, which includes only eight genera (Fryxell, 1968, 1979), has traditionally been distinguished from other members of the Malvaceae on the basis of morphological features of the embryo, wood and seed-coat anatomy, and by the presence of the punctae or lysigenous cavities ("gossypol glands") that are widely distributed throughout the plant body. More recently, the monophyly of the tribe has been confirmed using comparative analyses of chloroplast DNA restriction site variation (LaDuke and Doebley, 1995) and DNA sequence data (Seelanan *et al.*, 1997; Wendel *et al.*, 2002).

Four of the eight genera in the *Gossypieae* are small, with restricted geographic distributions (Fryxell, 1968, 1979). *Lebronnecia* is a monotypic endemic from the Marquesas Islands. *Cephalohibiscus*, from New Guinea and Solomon Islands, also contains only a single species and is not found in cultivation. Two species from East Africa and Madagascar are described in *Gossypioides*. The Hawaiian endemic *Kokia* includes four species, of which one is extinct. In addition to these four small genera, the tribe includes four moderately sized genera with broader geographic ranges: *Hampea* comprises of 21 neotropical species, *Cienfuegosia* includes 25 species with an aggregate range that includes the neotropics and parts of Africa, and 17 species are recognized in the pantropically distributed *Thespesia*. The largest and most widely distributed genus in the tribe is *Gossypium*, which contains more than 50 species (Fryxell, 1992), including the four domesticated species described before. These cultivated species embody considerable genetic diversity, but this diversity is dwarfed by that included in the genus as a whole (detailed later), whose species have an aggregate geographic range that encompasses most tropical and subtropical regions of the world. *Gossypium* is distinguished from related genera by a combination of characters, including: an undivided style, coriacious capsule containing several seeds per locule, a somatic chromosome number of 26, and the presence of three foliaceous (usually) involucellar bracts subtending each flower. As each of these traits is found in related genera, no unique morphological characters define *Gossypium*.

Recent molecular phylogenetic analyses have clarified several aspects of the evolutionary history of the tribe that are particularly germane to an exploration of polyploid cotton (Cronn *et al.*, 2002b; Seelanan *et al.*, 1997; Wendel *et al.*, 2002). Most important has been the formal demonstration that the diverse group of species recognized as belonging to *Gossypium* do in fact constitute a single natural lineage (a monophyletic group), despite their worldwide distribution and extraordinary morphological and cytogenetic diversity (the latter discussed later). A second significant discovery has been the identity of the closest relatives of *Gossypium*. All phylogenetic data sets concur in revealing that two genera collectively constitute the phylogenetic sister-group to *Gossypium*, i.e., the African–Madagascan genus *Gossypioides* and the Hawaiian endemic genus

Kokia. These latter genera may thus be used as phylogenetic outgroups for purposes of studying gene and genome evolution within *Gossypium*.

An additional insight that emerges from phylogenetic analysis concerns the temporal component to the genealogy, which is evident in sequence divergence data that serve as a proxy for time. Using sequence divergence data from the chloroplast gene *ndhF* and published sequence divergence rates to calibrate a molecular clock, Seelanan *et al.* (1997) suggested that *Gossypium* branched off from *Kokia* and *Gossypioides* approximately 12.5 mya, with the latter two genera becoming separated more recently, perhaps 3 mya. The early estimates are in broad agreement with those recently obtained from a larger data set based on synonymous site divergences calculated from nearly 8000 aligned sites representing 10 different nuclear genes (Cronn *et al.*, 2002b). These latter data show that the mean divergence at synonymous sites between members of *Gossypium* and its closest relatives (*Gossypioides* + *Kokia*) is approximately 7.0%, with the latter two genera being about 2.8% divergent at synonymous sites. Although rates of mutation may vary widely among genes and across lineages (Gaut, 1998; Sanderson, 1998), the magnitude of divergence is consistent with a Miocene separation between *Gossypium* and its closest relatives and a rather recent, perhaps Pliocene divergence between *Gossypioides* and *Kokia*.

We note that the relatively recent split of genera now geographically isolated from one another by thousands of kilometers of open ocean (*Kokia* from Hawaii and *Gossypioides* from Madagascar and East Africa) implies that trans-oceanic dispersal was involved in the evolution of one or both genera. In this respect the *Gossypioides–Kokia* example represents only the latest in a series of examples of long-distance, oceanic dispersal as a factor in the evolution of the cotton tribe and the genus *Gossypium* (DeJoode and Wendel, 1992; Fryxell, 1979; Stephens, 1958, 1966; Wendel, 1989; Wendel and Albert, 1992; Wendel and Percival, 1990; Wendel and Percy, 1990).

B. Emergence and Diversification of the Genus *Gossypium*

A Global Radiation

As noted earlier, a remarkable diversification accompanied the global radiation of *Gossypium*. These morphologies evolved in response to the demands of particular ecological settings and selective environments. Plant habit, for example, ranges from fire-adapted, herbaceous perennials in NW Australia to small trees in SW Mexico that escape the dry season by dropping their leaves. Corolla colors span a rainbow of blue to purple (*G. triphyllum*), mauves and pinks ("Sturt's Desert Rose," *G. sturtianum*, is the official floral emblem of the

Northern Territory, Australia), whites and pale yellows (Mexico, Africa–Arabia) and even a deep sulphur-yellow (*G. tomentosum* from Hawaii). Seed coverings range from nearly glabrous to the naked eye (e.g., *G. klotzschianum* and *G. davidsonii*), to short stiff, dense, brown hairs that aid in wind-dispersal (*G. australe*, *G. nelsonii*), to long, fine white fibers that characterize highly improved forms of the four cultivated species. There are even seeds that produce fat bodies to facilitate ant-dispersal (section *Grandicalyx* cottons from NW Australia, Seelanan *et al.*, 1999). Much of this morphological diversity is described in detail by Fryxell (1979).

The foregoing discussion suggests that the cotton genus has a history that extends back millions of years, so perhaps it is not surprising that the genus achieved worldwide distribution, with several primary centers of diversity in the arid or seasonally arid tropics and subtropics (Table I). Particularly species-rich regions include Australia, especially the Kimberley region in NW Australia, the Horn of Africa and southern Arabian Peninsula, and the western part of central and southern Mexico. Recognition of these groups of related species and their individual constituents reflects accumulated scientific understanding that has emerged from a long history of basic plant exploration and taxonomic and evolutionary study. The taxonomy of the genus has been summarized in several useful volumes (Fryxell, 1979, 1992; Hutchinson *et al.*, 1947; Saunders, 1961; Watt, 1907). The most recent and widely followed taxonomic treatments are those of Fryxell (1979, 1992), in which species are grouped into four subgenera and eight sections (Table I). This classification system is primarily based on morphological and geographical evidence, although most infrageneric alignments are congruent with cytogenetic and molecular data sets as well, as will be discussed later.

At present, *Gossypium* includes approximately 50 species (Fryxell, 1992), but remarkably, new species continue to be discovered (Fryxell *et al.*, 1992, Stewart,

Table I
Diversity and Geographic Distribution of the Major Lineages of *Gossypium*

Genome group	Number of species	Geographic distribution
A	2	Africa, possibly Asia
B	3	Africa, Cape Verde Islands
C	2	Australia
D	13	Primarily Mexico; also Peru, Galapagos Islands, Arizona
E	7+	Arabian Peninsula, Northeast Africa, Southwest Asia
F	1	East Africa
G	3	Australia
K	12	NW Australia
AD	5	New World tropics and subtropics including Hawaii

Craven, Brubaker and Wendel, unpublished). It is a truism that the morphological and ecological breadth encompassed by the wild species of *Gossypium* must have parallels in physiological and chemical diversity. The wild species of cotton, consequently, represent an ample genetic repository for potential exploitation by the cotton breeders. Although these wild species remain a largely untapped genetic resource, examples abound of their productive inclusion in breeding programs (reviewed by McCarty and Percy, 2001). A brief introduction to the major groups of diploid cotton species follows.

Australian Species

Australian cottons (subgenus *Sturtia*) comprise 16 named species as well as a new species that is yet to be named (Stewart, Craven, Brubaker, and Wendel, unpublished). Collectively, these taxa comprise the C-, G-, and K-genome groups, with two, three, and 12 species, respectively. These three groups of species are implicated by DNA sequence data (Liu *et al.*, 2001b; Seelanan *et al.*, 1997, 1999) to be natural lineages, consistent with their formal alignments into the taxonomic sections *Sturtia* (C-genome), *Hibiscoidea* (G-genome), and *Grandicalyx* (K-genome). Relationships among the three groups, however, remain unclear. Some data place *G. robinsonii* as basal within the entire assemblage of Australian species (Wendel and Albert, 1992), suggesting that radiation of *Gossypium* in Australia proceeded eastward from the westernmost portion of the continent. Whether this basal position will withstand the scrutiny of other data sets is an open question, as the most recent analyses (Liu *et al.*, 2001b; Seelanan *et al.*, 1997, 1999) are equivocal in this regard.

With respect to the taxonomy within each of the three Australian genome groups, there is little uncertainty for the C- and G-genome groups, as these are well represented in collections and have been thoroughly studied (Fryxell, 1979, 1992; Liu *et al.*, 2001b; Seelanan *et al.*, 1997, 1999; Wendel and Albert, 1992; Wendel *et al.*, 1991, and references therein). Much less certain is the taxonomy of the K-genome species, which are all placed in section *Grandicalyx*. Recent expeditions to the Kimberley area have enhanced our understanding of diversity within the group and have resulted in the discovery of at least seven new species, six of which have been formally described (Fryxell *et al.*, 1992). These unusual species have a distinctive geography, morphology and ecology, and exhibit a syndrome of features that are characteristic of fire-adaptation. In particular, they are herbaceous perennials with a bi-seasonal growth pattern whereby vegetative growth dies back during the dry season, or as a result of fire, to underground rootstocks that initiate a new cycle of growth with the onset of the next wet season. Species in section *Grandicalyx* have pedicels that recurve following pollination so that the capsules are pendent and open inverted at maturity,

releasing sparsely vestitured, ant-dispersed seeds that bear elaiosomes to aid in attracting ants. Many of these species are poorly represented in collections and are incompletely understood from a taxonomic standpoint. Molecular phylogenetic analyses have yielded conflicting results regarding interspecific relationships in this group (Liu *et al.*, 2001b; Seelanan *et al.*, 1999).

African–Asian Species

Fourteen species from Africa and Arabia are recognized in the most recent taxonomic treatment of the genus (Fryxell, 1992), collectively comprising the subgenus *Gossypium*. The taxonomic section *Gossypium* contains four subsections whereas section *Serrata* contains only *G. trifurcatum*, from deserts in eastern Somalia; this species is poorly understood taxonomically and cytogenetically. The unusual feature of dentate leaves raises the possibility that it may not belong to *Gossypium*, and may instead be better referred to *Cienfuegosia* (Fryxell, 1992), a possibility that requires future evaluation. This latter example underscores the provisional nature of much of the taxonomy of the African–Arabian species of *Gossypium*, which are sorely in need of basic plant exploration and systematic study. Within section *Pseudopambak*, species recognition and definition are in some cases based on limited material (e.g., *G. benadirense*, *G. bricchettii*, *G. vollesenii*) and no analyses have been conducted on cytogenetic characteristics or molecular phylogenetic affinities.

From a cytogenetic standpoint, the African–Arabian species exhibit considerable diversity, collectively accounting for four of the eight genome groups (A-, B-, E-, and F-). The two cultivated cottons of subsection *Gossypium*, *G. arboreum* and *G. herbaceum*, have been extensively studied (reviewed by Wendel *et al.*, 1989) and comprise the A-genome. The three African species in subsection *Anomala* comprise the B-genome, as discussed before. The sole F-genome species, *G. longicalyx*, is cytogenetically distinct (Phillips and Strickland, 1966), morphologically isolated (Fryxell, 1971, 1992; Vollesen, 1987), and is perhaps adapted to more mesic conditions than any other diploid *Gossypium* species. The remaining species, those of subsection *Pseudopambak*, are considered to possess E-genomes, although few of these taxa have been sufficiently studied to verify this supposition.

American Diploid Species

Subgenus *Houzingenia* contains two sections and six subsections, whose species collectively represent the New World D-genome diploids. Given their proximity to American taxonomists, these species have been thoroughly

collected and studied than most, and consequently their taxonomy is reasonably well understood. The subgenus has also received considerable phylogenetic attention (Cronn *et al.*, 1996; Seelanan *et al.*, 1997; Small and Wendel, 2000b; Wendel and Albert, 1992; Wendel *et al.*, 1995b), which provides strong support for the naturalness of most of the recognized subsections. Evolutionary relationships among the apparently natural subsections are less evident, however, although increasing evidence (Cronn *et al.*, 1996; Liu *et al.*, 2001b; Small and Wendel, 2000b) suggests that *G. gossypioides* is basal-most within the subgenus (also Cronn and Wendel, unpublished data).

As the center of diversity for the 13 species of D-genome diploids is western Mexico, it is likely that the lineage became established and initially diversified in this region. Later range extensions are inferred to have arisen from relatively recent (probably Pleistocene) long-distance dispersals, leading to the evolution of endemics in Peru (*G. raimondii*) and the Galapagos Islands (*G. klotzschianum*).

C. Chromosomal Evolution and the Origin of the Polyploids

Genome Size Variation And The Concept of Genome Group

As the genus diversified and spread, it underwent extensive chromosomal evolution, which has been studied by many researchers (reviewed by Endrizzi *et al.*, 1985). Chromosome morphology is similar among closely related species, and this is reflected in the ability of related species to form hybrids that display normal meiotic pairing and high F_1 fertility. In contrast, crosses among more distant relatives are often difficult or impossible to effect, and those that are successful are characterized by meiotic abnormalities. The collective observations of pairing behavior, chromosome sizes, and relative fertility in interspecific hybrids led to the designation of single-letter genome symbols (Beasley, 1941) for related clusters of species. At present, eight diploid genome groups (A–G and K) are recognized (Endrizzi *et al.*, 1985; Stewart, 1995; Wendel *et al.*, 1999). This cytogenetic partition of the genus is largely congruent with taxonomic and phylogenetic divisions, as discussed later.

Although all diploid *Gossypium* species share the same chromosome number ($n = 13$), there is more than a threefold variation in DNA content per genome (Bennett *et al.*, 1997; Bennett *et al.*, 1982; Edwards *et al.*, 1974; Edwards and Mirza, 1979; Kadir, 1976; Michaelson *et al.*, 1991), with 2C contents ranging from approximately 2 pg per 2C nucleus in the New World, D-genome diploids to approximately 7 pg per cell in Australian K-genome species (J. McD. Stewart, personal communication). A-genome species have intermediate values of approximately 3.8 pg per 2C nucleus. The range in genome sizes is even greater when other diploid members of the tribe are considered, nearly sevenfold

variation in DNA content is reported (Wendel *et al.*, 2002) between the largest (*T. populnea*; 2C = 8.2 pg) and smallest (*Gossypioides kirkii* and *K. drynarioides*, each with 2C = 1.2 pg) genomes measured till date in the *Gossypieae*. This extraordinary variation in DNA content is widely believed to be caused by modification of the repetitive DNA fraction, with relatively little change in the absolute amounts of single-copy DNA (Geever *et al.*, 1989). Increasing evidence indicates that both genome size expansion and contraction are common in evolution, not only in *Gossypium* and *Gossypieae* but also in angiosperms as a whole (Wendel *et al.*, 2002).

Genomic Composition of the Polyploids

Cytogenetic investigations, starting as early as the 1920s (Denham, 1924), revealed that in addition to species having a haploid complement of 13 chromosomes, *Gossypium* included taxa with a haploid number of 26. Longley (1933) noted that this doubled chromosome number suggested "a duplication of the chromosomes of an ancestral type." Webber (1935) commented that the formation of 13 bivalents in a hybrid between wild and cultivated American species "support the hypothesis that the species having 26 pairs are allotetraploids," and further suggested that the ancestral diploid donors involved "wild American species,... and Asiatic species." This conclusion was also attained by Skovsted (1934, 1937) based on the analyses of chromosome sizes and pairing behavior in interspecific hybrids. Historically important confirmation of the allopolyploid nature of the American tetraploid cotton species emerged from the work of Beasley (1940) and Harland (1940), who synthesized experimental allotetraploids from A-genome (Asiatic) and D-genome (American) diploids and showed that these could form fertile hybrids with natural American tetraploids.

These classic cytogenetic studies demonstrated that the American tetraploid species are true allopolyploids and that they contain two resident genomes, an A-genome from Africa or Asia, and a D-genome similar to those found in the American diploids. Additional and conclusive support (reviewed by Endrizzi *et al.*, 1985) for the hypothesis of an allopolyploid origin of the American tetraploids emerged in subsequent decades from diverse sources of evidence, including genetic studies of duplicate factors controlling morphology (Stephens, 1944b, 1951a), meiotic pairing behavior in synthetic polyploids (Gerstel and Phillips, 1958; Phillips, 1963, 1964), phytochemical analysis (Parks *et al.*, 1975), isozyme markers (Saha and Zipf, 1998), comparative genetic mapping (Brubaker *et al.*, 1999b; Reinisch *et al.*, 1994) and comparative analysis of DNA sequences (Cronn *et al.*, 1999; Small *et al.*, 1998; Small and Wendel, 2000a). These latter studies provide particularly compelling proof that the allotetraploid (AD-

genome) species formed from hybridization between A- and D-genome ancestors: most nuclear genes are duplicated in the AD-genome cottons, and when both copies are isolated and sequenced, they correspond phylogenetically and phenetically to those of the antecedent A- and D-genome diploids.

Taxonomic Diversity of the Polyploids

Following their initial origin, allopolyploid *Gossypium* spread and diversified, and at present, five distinct species are recognized. It is ecologically noteworthy that two of these are island endemics that must have originated following long-distance dispersal events. *Gossypium darwinii* is native to the Galapagos Islands, where it may form large and continuous populations in some areas (Wendel and Percy, 1990). *Gossypium tomentosum*, from the Hawaiian Islands, has a more diffuse population structure, occurring mostly as scattered individuals and small populations on several islands (DeJoode and Wendel, 1992). A third allopolyploid, *G. mustelinum*, has an island-like distribution in the sense that it is an uncommon species restricted to a relatively small region of northeast Brazil (Wendel *et al.*, 1994). In addition to these three true wild species, there are two cultivated species (*G. barbadense* and *G. hirsutum*), each of which has a large indigenous range, collectively encompassing a wealth of morphological forms that span the wild-to-domesticated continuum (Brubaker *et al.*, 1999a; Brubaker and Wendel, 1994; Fryxell, 1979; Wendel *et al.*, 1992). *Gossypium hirsutum* is widely distributed in Central and South America, the Caribbean, and even reaches distant islands in the Pacific (Solomon Islands, Marquesas). *Gossypium barbadense* has a more southerly indigenous range, centered in the northern third of South America but with a large region of range overlap with *G. hirsutum* in the Caribbean. Some have recognized a sixth allopolyploid species (Fryxell, 1979), *G. lanceolatum* (= *G. hirsutum* "race palmeri"), which is known only as a cultigen. Brubaker and Wendel (1993) reviewed the evidence that bears on the specific status of this taxon and concluded that *G. lanceolatum* is more appropriately considered a variant of *G. hirsutum*.

D. Phylogenetic Relationships and the Temporal Scale of Divergence

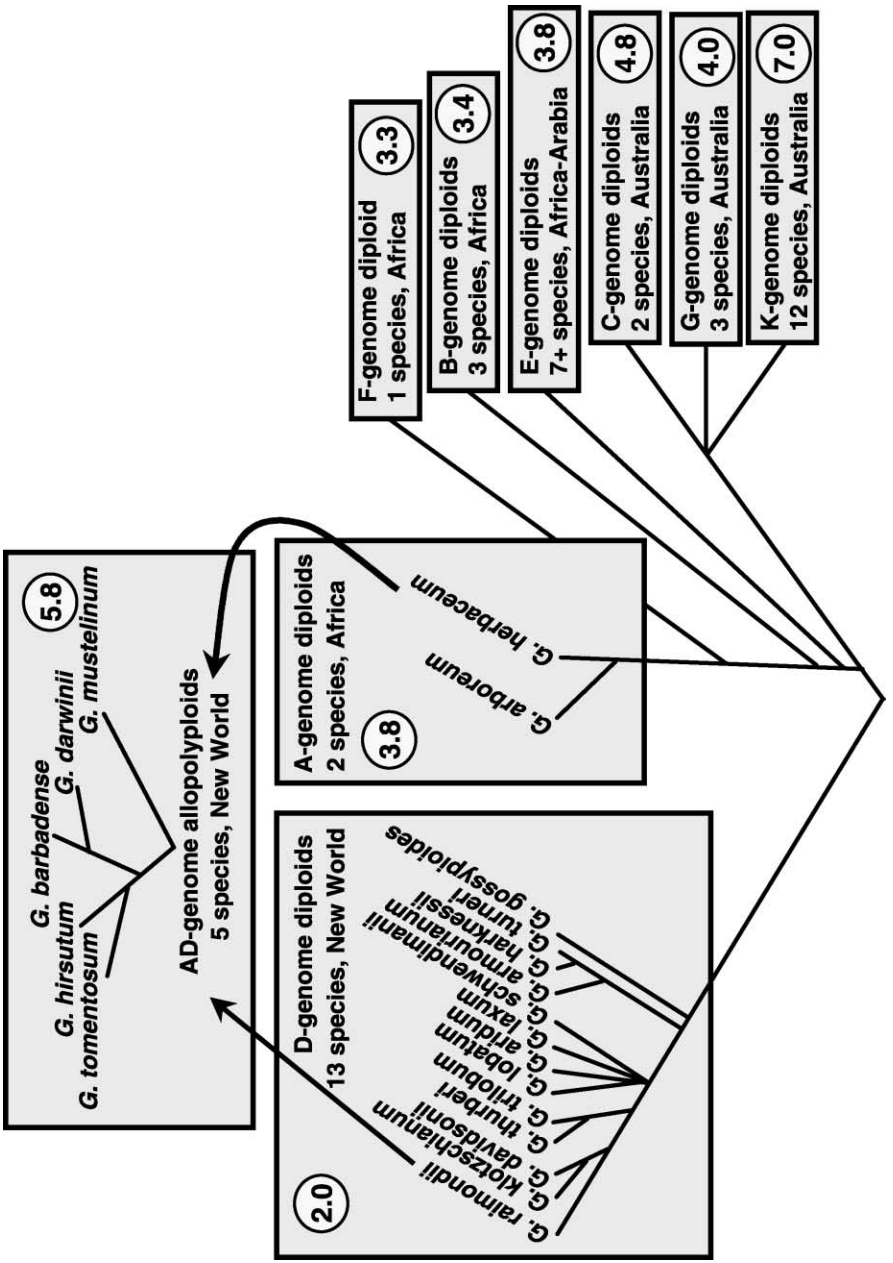
A Genealogical Framework for the Genus

Relationships within and among the various genome groups have been addressed in a number of recent molecular phylogenetic investigations, using variation in cpDNA restriction sites (Wendel and Albert, 1992), DNA sequences for the 5S ribosomal genes and spacers (Cronn *et al.*, 1996), DNA sequence data

for the cpDNA gene *ndhF* and the nuclear 5.8S gene and its flanking internal transcribed spacers (Seelanan *et al.*, 1997), sequence analysis of *Adh* genes (Seelanan *et al.*, 1999; Small and Wendel, 2000b) nuclear introns (Liu *et al.*, 2001b) and most recently, by extensive analyses of DNA sequence data for 16 different loci from the nuclear and chloroplast genomes, collectively representing over 17,000 nucleotides per taxon (Cronn *et al.*, 2002b). Each of these studies shows that the genealogical lineages of species are largely congruent with genome designations and geographical distributions. Accordingly, each genome group corresponds to a single natural lineage, and in most cases, these lineages are also geographically cohesive. This information has been embodied in a synthesis of relationships and genome evolution, as summarized in Fig. 1, which is largely based on the thorough molecular phylogenetic analysis of Cronn *et al.* (2002b).

Several aspects of the phylogenetic history of *Gossypium* bear highlighting. First, there exist four major lineages of diploid species corresponding to three continents: Australia (C-, G-, K-genomes), the Americas (D-genome), and Africa/Arabia (two lineages: one comprising the A-, B-, and F-genomes, and a second containing the E-genome species). Second, the earliest divergence event in the genus separated the New World D-genome lineage from the ancestor of all Old World taxa. Thus, New World and Old World diploids are phylogenetic sister groups. Following this basal-most split in the genus, cottons comprising the Old World lineage was divided into three groups, namely, the Australian cottons (C-, G-, and K-genome species), the African–Arabian E-genome species, and the African A-, B-, and F-genome cottons. Third, the African F-genome clade, which consists of the sole species *G. longicalyx*, is definitively diagnosed as sister to the A-genome species. This identifies the wild forms most closely related to those first domesticated in the A-genome species *G. arboreum* and *G. herbaceum*. Because this relationship is revealed, prospects are raised for ultimately understanding the genetic basis of the origin of modern lint. Fourth, the serendipitous merger of the A- and D-genome at the time of allopolyploid formation represents a chance biological reunion of two genomes descended from the earliest split in the genus. Thus, the two constituent genomes of allopolyploid cotton evolved first in different hemispheres and diverged for millions of years in isolation from one another. Fifth, and finally, the accumulated data indicate that the major lineages of *Gossypium* were established in relatively rapid succession shortly after the genus originated and diverged from the *Kokia–Gossypioides* clade. The evolutionary picture thus envisioned is that there was a rapid and global radiation early in the history of the genus, with temporally closely spaced divergence events.

There remains some uncertainty regarding several of the earliest branch points. These exist because the phylogenies inferred from different molecular data sets differ with respect to the resolution of the B- and E-genome species groups. As discussed by Cronn *et al.* (2002b), the failure to robustly resolve the E-genome probably reflects its early divergence from its nearest relatives; rapid


AD-genome allopolyploids
5 species, New World
5.8
G. barbadense
G. darwinii
G. mustelinum
G. hirsutum
G. tomentosum
D-genome diploids
13 species, New World
2.0
G. raimondii
G. klotzschianum
G. davidsonii
G. thurberi
G. trilobum
G. lobatum
G. arildum
G. laxum
G. schwendimanii
G. armourianum
G. harknessii
G. turneri
G. gossypioides
A-genome diploids
2 species, Africa
3.8
G. arboreum
G. herbaceum
F-genome diploid
1 species, Africa
3.3
B-genome diploids
3 species, Africa
3.4
E-genome diploids
7+ species, Africa-Arabia
3.8
C-genome diploids
2 species, Australia
4.8
G-genome diploids
3 species, Australia
4.0
K-genome diploids
12 species, Australia
7.0

radiation translates into phylogenetically short interior branches, which are by definition more difficult to resolve with statistical confidence than longer branches. Perhaps more problematic is the absence of a consistent phylogenetic signal for the B-genome species. Chloroplast DNA data robustly place the B-genome lineage sister to the combined Australian (C + G)-genome, whereas 12 nuclear loci place the B-genome lineage solidly into an African clade that includes A- and F-genome cottons. On the basis of evolutionary independence of the estimates provided by 12 different nuclear loci, the resolution of the B-genome inferred from analysis of the nuclear data probably more accurately reflects its true history. Taken together, the phylogenetic results suggest that all African–Arabian cottons comprise a single clade, with early and rapid radiations of the E-, B-, and A + F-genome lineages (Fig. 1).

Dating the Divergence Events

A temporal component to the evolutionary history shown in Fig. 1 may be developed from a different use of the DNA sequence data, namely sequence divergence amounts. Previous authors have recognized the extensive morphological variation and global distribution of the genus and have suggested that: (1) *Gossypium* was relatively ancient, with genome groups originating in the Cretaceous, 60–100 mya (Endrizzi *et al.*, 1985, 1989; Geever *et al.*, 1989) and (2) the present distribution of the New World and Old World lineages reflects divergence arising from the breakup of the Gondwanan supercontinent (Endrizzi *et al.*, 1985, 1989; Saunders, 1961). This hypothesis has been frequently cited even though it lacked support from alternative evidence such as macrofossils or palynological surveys. Indeed, the latter contradicts a Cretaceous diversification of the genus, as the oldest pollen referable to the Malvaceae is Eocene (38–46 mya) in age (Muller, 1981, 1984). At present, *Gossypium* fossils are limited to leaf prints

Figure 1 Evolutionary history of *Gossypium*, as inferred from multiple molecular phylogenetic data sets. The closest relative of *Gossypium* is a lineage comprising the African–Madagascan genus *Gossypioides* and the Hawaiian endemic genus *Kokia*. Following its likely Miocene origin, *Gossypium* split into three major lineages: the New World diploids (D-genome); the African–Asian clade (A-, B-, E- and F-genomes); and the Australian group (C-, G-, and K-genomes). This global radiation was mediated by several trans-oceanic dispersal events and was accompanied by considerable morphological, ecological, and cytogenetic differentiation (2C genome sizes shown in circles). Interspecific hybridization is implicated in the evolution of approximately one-fourth of the genus. Allopolyploid cottons formed following trans-oceanic dispersal of an A-genome diploid to the Americas, where the new immigrant underwent hybridization, as female, with a native D-genome diploid. Polyploid cotton probably originated during the Pleistocene, with the five modern species representing the descendants of an early and rapid colonization of the New World tropics and subtropics.

in Hawaiian volcanic sediments dating to *ca.* 0.4 mya (Woodcock and Manchester, 1998), and fossil pollen has yet to be ascribed to the genus.

Molecular data, initially derived from nuclear DNA thermal reannealing experiments (Endrizzi *et al.*, 1989; Geever *et al.*, 1989), and more recently from DNA sequence variation (Cronn *et al.*, 2002b; Seelanan *et al.*, 1997; Wendel and Albert, 1992), have been used to evaluate divergence times among diploid cottons and to estimate the time of formation of the allopolyploid members of the genus. Using thermal stability measures of nucleotide divergence, Geever *et al.* (1989) provided the first estimate of divergence for members of *Gossypium*, arguing for a Cretaceous origin of the genus. This suggestion has been refuted using later data and estimates derived from DNA sequence data, where sequence divergence amounts were used in conjunction with divergence rates (estimated from other taxa) to calibrate a *Gossypium* molecular clock. Direct sequencing of the chloroplast gene *ndhF* (Seelanan *et al.*, 1997) resulted in a mean difference between C-genome cottons and other diploid genome groups of 0.86%. Using a molecular rate calibration based on *rbcL* sequence divergence for other plants, for which the fossil record provides indications of divergence times, the earliest split in *Gossypium* was calculated to have taken place approximately 12 mya. A more extensive analysis was conducted by Cronn *et al.* (2002b), who first estimated rates of divergence at approximately 8000 synonymous sites from across multiple nuclear genes and then pegged these divergence amounts to divergence times using divergence rate calibrations from angiosperm *Adh* genes (see discussion in Cronn *et al.*, 2002b). These calculations yielded results that were in broad agreement with those based on the earlier cpDNA sequence divergence data, suggesting that the genus originated in the last 5–15 million years, probably in the Miocene.

The molecular clock analyses further indicate that the initial split in the genus, namely, separation of the D-genome lineage from the Old World lineage, occurred within the last seven million years. This estimate provides an approximation of the amount of time that the two progenitor genomes (A and D) of allopolyploid cotton evolved in isolation prior to their merger during hybridization and polyploidization. To quantify the concept of "rapid divergence" among the major extant lineages of diploid cotton, Cronn *et al.* (2002b) used a maximum likelihood approach to estimate that the modern diploid lineages of cottons diverged within a time span of at most two million years. Although there are many sources of error in the calculations such as these, they offer useful approximations of absolute and relative dates of diversification events. Irrespective of the inherent imprecision in these estimates, two aspects of evolutionary history appear incontrovertible based on the phylogenetic results and the molecular dating analyses: first, that the extant *Gossypium* lineages diversified sufficiently recently that they achieved their global distribution via an evolutionary history involving at least several long-distance, trans-oceanic dispersals (as noted earlier); and second, that the major divergence events occurred on a temporally compressed scale relative to the age of the genus.

III. SPECIATION MECHANISMS

A. A Fondness for Trans-oceanic Voyages

The aforementioned dispersal history of the genus is so remarkable that it bears further comment. As noted in Section II-A earlier, the apparent propensity for long-distance dispersal appears to be a characteristic of the entire cotton tribe, and as noted in the previous paragraphs, it certainly has played an important role in diversification of *Gossypium* itself. Consideration of the phylogeny in a temporal context and in light of plate tectonic history leads to several diagnoses of intercontinental, presumably trans-oceanic voyages. These would include at least one dispersal between Australia and Africa, another to the Americas (probably Mexico) leading to the evolution of the D-genome diploids, and a second, much later colonization of the New World by the A-genome ancestor of the AD-genome allopolyploids.

Long-distance dispersal clearly has played an important role not only in diversification of major evolutionary lines but also in speciation within *Gossypium* genome groups. Examples include dispersals from southern Mexico to Peru (*G. raimondii*), from northern Mexico to the Galapagos Islands (*G. klotzschianum*), from western South America to the Galapagos Islands (*G. darwinii*), from Africa to the Cape Verde Islands (*G. capitis-viridis*), and from the neotropics to the Hawaiian Islands (*G. tomentosum*) (DeJoode and Wendel, 1992; Wendel and Percival, 1990; Wendel and Percy, 1990).

These latter examples, as well as those noted earlier for the origin of *Kokia* and *Gossypioides* from a common ancestor in the Pliocene, suggest a common dispersal mechanism of oceanic drift. In this respect, it is satisfying that seeds of many species of *Gossypium* are tolerant to prolonged periods of immersion in salt water (Stephens, 1958, 1966). It is astonishing that seeds of the Hawaiian endemic cotton, *G. tomentosum*, are capable of germination after three years immersion in artificial salt water (Fryxell, personal communication). Apparently seeds of some species may retain buoyancy for at least a couple of months, which may be insufficient for trans-oceanic dispersal, perhaps in some cases long-distance dispersal was mediated through natural rafting on floating debris (Stephens, 1966).

B. A Propensity for Interspecific Gene Exchange

In addition to the conventional mechanisms of geographic speciation (Fryxell, 1965, 1979) and speciation promoted by long-distance dispersal, hybridization has played an important role in diversification of *Gossypium* (Table II). This was first discovered during routine phylogenetic studies of Australian species, where

Table II
Species and Lineages of *Gossypium* that may have Hybrid or Introgressant Ancestries

Species or species group	References
G. aridum	DeJoode (1992) and Wendel and Albert (1992)
G. gossypioides	Wendel *et al.* (1995b) and Zhao *et al.* (1998)
G. bickii	Wendel *et al.* (1991) and Liu *et al.* (2001b)
G. cunninghamii	Seelanan *et al.* (1999) and Wendel and Albert (1992)
B-genome: *G. anomalum*, *G. capitis-viridis*, *G. triphyllum*	Cronn *et al.* (2002b)
AD-genome allopolyploids: *G. barbadense*, *G. darwinii*, *G. hirsutum*, *G. mustelinum*, *G. tomentosum*	Reviewed here

molecular markers from the plastid and nuclear genomes were used to document an unusual evolutionary history for *G. bickii* (Wendel *et al.*, 1991). This species is one of three morphologically similar G-genome cottons (along with *G. australe* and *G. nelsonii*) in section *Hibiscoidea*. In contrast to expectations based on this taxonomy, the maternally inherited chloroplast genome of *G. bickii* was shown to be nearly identical to the plastid genome of *G. sturtianum*, a morphologically distant C-genome species from a different taxonomic section (*Sturtia*). In contrast, nuclear markers show the expected relationship, i.e., *G. bickii* shares a more recent common ancestor with its close morphological allies, *G. australe* and *G. nelsonii*, than it does with *G. sturtianum*. This discrepancy was explained by invoking a bi-phyletic ancestry for *G. bickii*, whereby *G. sturtianum*, or a similar species, served as the maternal parent in an ancient hybridization with a paternal donor from the lineage leading to *G. australe* and *G. nelsonii*. Interestingly, no *G. sturtianum* nuclear genes were detected in *G. bickii*, suggesting that the nuclear genomic contribution of the maternal parent was eliminated from the hybrid or its descendent maternal lineage (see also Liu *et al.*, 2001b).

This phenomenon of "cytoplasmic capture" has subsequently been implicated elsewhere in the genus (Table II). It may be that the entire B-genome species-group has an introgressant ancestry, as suggested by Cronn *et al.* (2002b) in noting the conflicting phylogenetic signal for this clade offered by sequence data from the nuclear and cytoplasmic genomes. Another likely example concerns the K-genome species *G. cunninghamii*, which perhaps not coincidentally has an unusual morphology and is geographically widely disjunct from its close relatives. This species is restricted to the Cobourg Peninsula, approximately 500 km distant from the Kimberley region where all other K-genome taxa are found. Analogous to *G. bickii*, the chloroplast genome of *G. cunninghamii* appears to have been donated by the *G. sturtianum* lineage, although in this case the hybridization event appears to have been more ancient (Seelanan *et al.*, 1999; Wendel and Albert, 1992).

A final example of cytoplasmic introgression resulting from hybridization involves *G. aridum*, one of four species of Mexican cottons that comprise subsection *Erioxylon*. These four species form a morphologically coherent and distinctive group of small trees whose shared common ancestry is supported, with one exception, by all molecular data sets (Cronn *et al.*, 1996; DeJoode, 1992; Small and Wendel, 2000b; Wendel and Albert, 1992; Wendel *et al.*, 1995b). The exception is that populations of *G. aridum* from the Mexican state of Colima have a chloroplast genome that is strikingly divergent from that found in the remainder of the species. This alien cytoplasm is inferred to have originated through an ancient hybridization with a member of the *Integrifolia* subsection, whose two extant species (*G. davidsonii* and *G. klotzschianum*) have geographic ranges (Baja California and the Galapagos Islands, respectively) that are distant from the range of *G. aridum*. As was the case with *G. bickii*, the nuclear genome of *G. aridum*, including the Colima populations, exhibits no evidence of this introgression event.

The most recently discovered and mysterious example of interspecific sexual contact in *Gossypium* is one involving not only cytoplasmic introgression between species but also recombination between diverged nuclear genomes. The species in question is *G. gossypioides*, the exclusive member of subsection *Selera* and a taxon with an apparently relictual range, occurring only in several small isolated populations in a single river drainage in Oaxaca, Mexico.

Until recently, there was no indication that *G. gossypioides* had a noteworthy evolutionary history, in that the inferred relationships between *G. gossypioides* and the other D-genome species, based on comparative gross morphology, cytogenetic data, interfertility relationships, and allozyme analysis were congruent (reviewed by Wendel *et al.*, 1995b). Wendel *et al.* (1995b), however, showed that the nuclear ribosomal DNA sequences from *G. gossypioides* are unlike those of any other D-genome taxon in fact, the sequence data implicated extensive recombination with rDNA sequences from A-genome cottons. Subsequent to this finding, *G. gossypioides* was discovered to contain a variety of repetitive DNAs that are shared with A-genome species but are otherwise unknown among D-genome species (Zhao *et al.*, 1998).

Complicating the story even further, recent phylogenetic analysis of DNA sequences reveals that *G. gossypioides* occupies a basal position within the D-genome clade (Cronn *et al.*, 1996; Liu *et al.*, 2001b; Small and Wendel, 2000b), yet it possesses a chloroplast genome very much like that found in *G. raimondii* (Wendel and Albert, 1992), which occupies a more terminal phylogenetic position with the D-genome assemblage. Wendel *et al.* (1995b) attributed these data to an ancient hybridization event, whereby *G. gossypioides* experienced contact with an A-genome, either at the diploid level, or at the triploid level as a consequence of hybridization with a New World allopolyploid, followed by repeated backcrossing of the hybrid into the *G. gossypioides* lineage, thereby restoring the single-copy component of the D-nuclear genome. It may be

that the *G. gossypioides* lineage was spawned by this process, and then later underwent secondary hybridization with a member of the *G. raimondii* lineage to acquire its chloroplast genome. Regardless of the details of this mysterious ancestry, this example highlights an utterly unexpected and as yet unexplained history of interspecific recombination.

The foregoing examples underscore a remarkable biogeographic feature of the genus, namely, relatively frequent hybridization between geographically widely separated lineages that apparently have no opportunity for contact. It may well be that one or more of the foregoing species (or their ancestors) actually originated through an evolutionary process that was "seeded" by a hybridization event. Although this remains an open question with the diploid species just discussed, there remains no doubt about the allopolyploid species, which exemplify a different outcome of the hybridization process while reflecting this biogeographically wondrous capacity for interspecific travel and genomic reunion.

IV. ORIGIN OF THE ALLOPOLYPLOIDS

A. Time of Formation

As discussed in Section IIC, ample evidence establishes that the New World tetraploid cottons are allopolyploids containing one genome similar to those found in the Old World, A-genome diploids and a second genome like those of the New World, D-genome diploids (reviewed by Endrizzi *et al.*, 1985; Wendel, 1989; Wendel *et al.*, 1995b). This implies that the two genomes must have established physical proximity, at least ephemerally, at some time in the past. Because the two parental genomic groups exist in diploid species that presently are half a world apart, a classic botanical mystery emerged: how and when did allopolyploid cotton form? For over 50 years now, this riddle has stimulated interest and speculation.

We previously noted that a number of authors have suggested that the *Gossypium* had an ancient, perhaps Cretaceous or early Tertiary origin (Endrizzi *et al.*, 1985, 1989; Geever *et al.*, 1989; Harland, 1939; Saunders, 1961; Skovsted, 1934; Stebbins, 1947). This idea appears to have been generated largely by the biogeographic realization that the genus has attained a global pattern of distribution during which it acquired impressive morphological variation (Saunders, 1961; Stebbins, 1947). Given these observations, and the present isolation of progenitor diploid genomes in different hemispheres, it seemed rational to suggest that allopolyploid formation was also ancient. According to this hypothesis, hybridization and polyploidization took place prior to the separation of the parental A- and D-genome lineages, which subsequently drifted apart as a consequence of plate tectonic movements. Under this scenario, then,

allopolyploids originated prior to the rifting of the South American and African continents, in the Cretaceous or perhaps the early Tertiary.

At the other end of the spectrum, several authors have been strongly influenced by the occurrence of cultivated species in both the Old and New World, and by the supposition that agronomically advanced fiber probably was developed only once from a truly wild ancestor (Hutchinson, 1959; Hutchinson *et al.*, 1947; Johnson, 1975). These observations were used to invoke a very recent origin of allopolyploids, forwarding a scenario that involved human transfer of an African or Asiatic A-genome cultigen to the New World, followed by deliberate or accidental hybridization with a wild D-genome species. Hutchinson *et al.* (1947), for example, concluded that allopolyploid cotton originated following human-mediated intercontinental transfer of a cultivated A-genome diploid, which was "...carried across the Pacific by man among the seeds of his crop plants and with the tools of his civilization." According to this hypothesis, allopolyploid cotton first formed in agricultural times, perhaps within the last six millennia.

In between these two extremes of a Cretaceous (perhaps 60–100 million) and recent (perhaps 6000 years) origin, which, remarkably, vary by four orders of magnitude, are other proposals. Endrizzi *et al.* (1989) argued for a probable Miocene origin (5–18 mya), based on thermal stability measurements in interspecific DNA-DNA hybridization experiments. This calculation was based, however, on the assumption of an early Cretaceous (100 mya) divergence between the parental diploid groups. Phillips (1963), in a review of cytogenetic evidence, including patterns of intra- and intergenomic chromosome differentiation, presents a persuasive case that polyploid cotton originated "in geologically recent times, probably since the start of the Pleistocene." Fryxell (1965) reached a similar conclusion based on taxonomic and ecological evidence considered in light of our understanding of historical climate change, and stated: "...the amphidiploids, which combine at the tetraploid level the germinal lines of two anciently divergent parts of the genus, are relative newcomers in the evolution of the genus" (he further defines this as mid-Pleistocene).

More recent experiments have uniformly supported this latter view, namely, that allopolyploid *Gossypium* originated prior to the evolution of modern humans but relatively recently in geological terms. Specifically, all molecular data sets support a Pleistocene origin, probably within the last two million years (Cronn *et al.*, 1999; Seelanan *et al.*, 1997; Small *et al.*, 1998; Wendel *et al.*, 1989; Zhao *et al.*, 1995). The first estimate explicitly based on DNA sequence data was that of Wendel (1989), who assayed variation in 560 restriction sites in the chloroplast genomes of diploid and allopolyploid species and translated these data into sequence divergence percentages. These estimates permitted divergence times to be calculated by calibrating with rates from other plants; the conclusion reached was that allopolyploids formed in the mid-Pleistocene. A similar conclusion emerged from a later study (Seelanan *et al.*, 1997) based on direct sequencing

of the cpDNA gene *ndhF*. Additional support was provided by Zhao *et al.* (1995), who reported minimal variation in repetitive DNAs among the various polyploid species.

DNA sequence data for a variety of nuclear genes has recently become available, and notwithstanding the gene-to-gene variation in rates, a consistent picture has emerged regarding divergence amounts between the diploid genome groups and between the diploids and their orthologous counterparts in the allopolyploids. Cronn *et al.* (1999), in a study of 16 low-copy nuclear sequences, reported that mean sequence divergence between the A- and D-genome diploids averaged 2.2%, whereas that between the diploids and their counterparts in the allopolyploid averaged 0.68 and 1.05%, respectively, for the A- and D-genome comparisons. The minimum of these latter two values (0.68%) is the relevant one for the present purposes, in that it provides an estimate of the maximum age of *Gossypium* allopolyploids. Using the formula $T = K/2r$, where K is the divergence amount and r is the rate of synonymous site divergence for nuclear genes from plants (perhaps 2.6×10^{-9}–1.5×10^{-8} substitutions/synonymous site/yr—(Gaut, 1998; Koch *et al.*, 2000; Morton *et al.*, 1996)), one may estimate that allopolyploids formed 0.3–1.3 mya. As the clock calibration utilizes synonymous sites, and because the sequence data include some, presumably more slowly diverging non-synonymous sites, this time estimate is moderately biased downward, perhaps by approximately one-third (Wendel and Cronn, unpublished). Also, given that the generation time is positively correlated with molecular evolutionary rates (Gaut, 1998) and that wild *Gossypium* species are long-lived perennials, it is likely that the more appropriate end of the rate spectrum to use is the slower rates. Hence, it seems probable that *Gossypium* allopolyploids formed in the Mid-Pleistocene, *ca.* 1–2 mya.

Given a Pleistocene origin for allopolyploid cotton species, one may infer that their morphological diversification and spread must have been relatively rapid following polyploidization. Recent phylogenetic analyses have demonstrated that since formation, allopolyploid cottons have radiated into three lineages collectively comprising five species (Cronn *et al.*, 1996; DeJoode and Wendel, 1992; Small *et al.*, 1999; Wendel, 1989; Wendel and Percy, 1990; Wendel *et al.*, 1994, 1995a). The only living descendant of one branch of the first cladogenetic event in the allopolyploids is *Gossypium mustelinum* (Small *et al.*, 1998; Wendel *et al.*, 1994). The other branch is represented by two species-pairs, each containing one of the two cultivated species (*G. barbadense* and *G. hirsutum*) and an island endemic that originated from long-distance dispersals; thus, *G. barbadense* is the sister-species of *G. darwinii*, from the Galapagos Islands (Wendel and Percy, 1990), while *G. hirsutum* is sister to *G. tomentosum*, from the Hawaiian Islands (DeJoode and Wendel, 1992). Collectively, these allopolyploid species exhibit great morphological diversity and have a geographic range that encompasses much of Central America and the Caribbean, the Hawaiian Islands, northern South America, and many distant Pacific Islands (Fryxell, 1979). Yet

molecular data sets indicate low levels of interspecific divergence. For example, in a survey of *ca.* 7000 nucleotides from introns and spacer sequences from the chloroplast genomes of each allopolyploid, relatively few differences were detected between species (Small *et al.*, 1998 #168; Wendel, 1989). Similarly, low levels of interspecific sequence divergence are reported for nuclear genes (Cronn *et al.*, 1996; DeJoode and Wendel, 1992; Small *et al.*, 1998; Wendel *et al.*, 1994; Wendel *et al.*, 1995a). Collectively, the evidence indicates that there must have been a relatively rapid colonization of the New World tropics by the allopolyploids following their formation.

B. Parentage of the Allopolyploids

Since the discovery that allopolyploid *Gossypium* species contain two genomes whose progenitors presently occur in different hemispheres, investigators have attempted to provide pieces to the puzzle of polyploid origin. One particular focus has been the question of parentage; that is, which of the modern species of A- and D-genome diploids best serve as models for the progenitor genome donors? Over the decades a diverse array of tools have been used in an effort to solve this question, from early studies of comparative morphology and segregation analysis, through cytogenetic, comparative phytochemistry, and protein electrophoretic studies, to modern phylogenetic investigations using DNA sequencing of homologous genes. The history of these efforts thus closely parallels the conceptual and methodological development of the fields of biosystematics and taxonomy. Much of this history was reviewed by Endrizzi *et al.* (1985), to which the reader is referred for a lucid discussion of the evidence generated up until that time; only a brief synopsis of this older work is needed here.

Early efforts to identify the A- and D-genome donors focused on genetic and morphological evidence. Stephens (1944a,b), for example, compared allometric patterns of leaf development in intergenomic hybrids to those of the naturally occurring allopolyploids *G. hirsutum* and *G. darwinii* (as *G. barbadense* var. *darwinii*). From these comparisons he offered perhaps the first explicit hypothesis of parentage, and stated (Stephens, 1944b) that "...either (*G. klotzschianum*, its close relative *G. davidsonii*, or *G. raimondii*) in combination with *G. arboreum* would produce a hybrid showing considerable similarity to present-day New World cottons." Additional support for the hypothesis of *G. raimondii* as the D-genome donor emerged from comparative analyses of plant habit and shape, floral features and extrafloral bract morphology in synthetic A × D amphiploids, and from observations of lint characteristics and vigor of intergenomic hybrids (Hutchinson *et al.*, 1945). A similar conclusion was later reached by Fryxell (1965), based on observations of the lint characteristics of diploid and wild polyploid species.

Cytogenetic studies also bolstered the view that *G. raimondii* was a reasonable model of the D-genome donor. Hutchinson *et al.* (1947) reported on multivalent frequencies in synthetic allopolyploid × D-genome hexaploids, using five different D-genome species as parents in crosses with either *G. hirsutum* or *G. barbadense*. They argued that the exceptionally low fertility in the hexaploids involving *G. raimondii* arose from a higher frequency of multivalent formation than observed in the other synthetic hexaploids, implicating *G. raimondii* as closer to the D-genome than the other species tested. This approach, involving comparative analysis of diverse synthetic allohexaploids, became an important and widely used methodology to address the question of polyploid parentage in the subsequent 15 years (reviewed by Endrizzi *et al.*, 1985). Gerstel (1958), for example, studied multivalent frequencies in hexaploids involving both of the two extant A-genome species (*G. arboreum* and *G. herbaceum*) to argue that *G. herbaceum* was more closely related of the two to the A-genome of the natural allopolyploids.

Particularly, clever extensions of the use of synthetic hexaploids involved the study of genetic segregation in testcrosses, under the assumption that the degree of fit to predicted autotetraploid ratios serves as an appropriate proxy for level of relatedness. Phillips (1963, 1964) summarized segregation data for between three and 10 loci in seven synthetic AD × D hexaploids involving seven different D-genome species. Segregation in crosses involving *G. raimondii* exhibited segregation ratios closest to the theoretical expectations of autotetraploids, implicating this species as the closest living relative of the original D-genome donor.

An additional perspective that became evident from the accumulating cytogenetic and segregation data is that the A-genome of allopolyploid cotton is more similar to that of the A-genome diploids than the D-genome of the allopolyploid is to that of the D-genome diploids. Thus, in synthetic allohexaploids multivalent frequencies are higher and genetic segregation more closely approximates autotetraploid ratios for A-genome chromosomes than D-genome chromosomes (Gerstel and Phillips, 1958; Phillips, 1964). Subsequent data from many sources has confirmed this observation. For example, in a recent survey of amplified fragment length polymorphisms (AFLPs) in diploid and polyploid cottons, 368 polymorphic fragments were observed in a collection of *G. barbadense* and *G. hirsutum* cultivars (Abdalla *et al.*, 2001). Of these 368 bands, 143 were shared between at least some of the tetraploids and the A-genome diploids, whereas only 84 were similarly shared with *G. raimondii*. Cronn *et al.* (1999) quantified these relationships using 14,705 base pairs of sequence information for 16 nuclear loci isolated from the D-genome diploid *G. raimondii*, the A-genome diploid *G. arboreum* (or *G. herbaceum*), and the AD-genome tetraploid *G. hirsutum*. Sequence divergence between the diploids and their corresponding genomes in the allopolyploid were 0.68 and 1.05% for the A- and D-genomes, respectively. Thus, *G. arboreum* and *G. herbaceum* may

be thought of as approximately 50% better model of the progenitor A-genome diploid than *G. raimondii* of the D-genome diploid.

The classical inferences of the parentage of allopolyploid cotton have withstood the scrutiny of time and of other techniques. Thus, seed protein electrophoretic (Cherry *et al.*, 1970, 1972), flavonoid (Parks *et al.*, 1975), and isozyme (Saha and Zipf, 1998) surveys were consistent with or directly corroborated the earlier inferences based on genetic and cytogenetic studies. In the last several years, these studies of phenetics, or similarity, have been supplemented by formal phylogenetic analyses, using explicit methods of genealogical reconstruction. Liu *et al.* (2001b), for example, showed that *G. raimondii* is the sister group to a clade composed of all five allopolyploid species, with respect to the D-genome homoeologues of a fatty acid desaturase gene. Four other D-genome species are sequentially basal to this clade, thereby eliminating them as candidates for D-genome donor status. Cladistic analysis of 5S ribosomal DNAs (Cronn *et al.*, 1996) and the internal transcribed spacer region of the 18S–26S arrays (Seelanan *et al.*, 1997) yield similar results, even when sampling all extant D-genome species (see, however Wendel *et al.*, 1995b). Ongoing studies of other nuclear genes, including two different alcohol dehydrogenase genes (Small *et al.*, 1998; Small and Wendel, 2000a,b), provide unambiguous evidence that *G. raimondii* is the closest living relative of the ancestral D-genome donor (Wendel, Cronn and Perkey, unpublished).

With respect to the A-genome parent, most have considered *G. herbaceum* to be closer than *G. arboreum* to the allopolyploid A-genome (Endrizzi *et al.*, 1985). It is important to recognize, however, that *G. herbaceum* is not the actual progenitor of the polyploid A-genome. This is evidenced by its chromosomal and molecular differentiation from the A-genome of the allopolyploids (Brubaker *et al.*, 1999b; Cronn *et al.*, 1996; Gerstel, 1953; Liu *et al.*, 2001b; Small *et al.*, 1998; Wendel, 1989; Wendel and Albert, 1992). Cytogenetic and comparative mapping studies have revealed that these genomes differ by at least two large translocations. Moreover, and more critically with respect to the issue of parentage, all cladistic analyses of molecular sequence data show that the two extant A-genome species are phylogenetically sister to each other and hence are genealogically equidistant from the A-genome of the allopolyploids (Cronn *et al.*, 1996; Liu *et al.*, 2001b; Wendel, 1989; Wendel and Albert, 1992). This leads to an important and perhaps self-evident truth namely, that the actual parents of the allopolyploids are extinct, and hence that reference to their parentage is more appropriately framed in terms of closest living descendants of the donor species. As noted above, the degree of divergence between the ancestral and modern genomes has been quantified by sequence data.

The notion of polyploid parentage is entwined with the biogeography of their formation and is informed by cognizance of their Pleistocene origin. Cytogenetic data, combined with the observation that the only known wild A-genome cotton is African (*G. herbaceum* subsp. *africanum*), has been used to support the suggestion

that polyploidization occurred following a trans-Atlantic introduction to the New World of a species similar to *G. herbaceum*. While trans-Atlantic transfer of an A-genome propagule is plausible, Wendel and Albert (1992) raised the possibility of a pre-Pleistocene A-genome radiation into Asia followed by a trans-Pacific, rather than trans-Atlantic dispersal to the Americas. This possibility is supported by the biogeography of the D-genome species, whose center of diversity is western Mexico, and by the occurrence of wild tetraploids in the Pacific region, including Hawaii and the Galapagos Islands. The relatively recent arrival of *G. raimondii* in Peru also suggests that the initial hybridization event may have taken place in Mesoamerica rather than South America, although other scenarios have been proposed (see discussion in Liu *et al.*, 2001b). We note that present species ranges may be rather different from those that existed at the time that an A-genome seed or seeds managed to find their way to the New World, so by necessity these speculations must be considered tentative.

One aspect of the history of the New World polyploid cottons that has become clear is that they all contain Old World cytoplasms. This was first suggested based on the observation that interspecific hybrids between species with these two genome types are more readily effected with the A-genome parent as female (Phillips, 1963). Subsequent analyses of chloroplast (Galau and Wilkins, 1989; Wendel, 1989) and mitochondrial (Small and Wendel, 1999) DNAs confirmed this early suggestion, thereby demonstrating that the seed parent in the initial hybridization event was an African or Asian A-genome taxon.

Several authors have proposed that allopolyploids formed more than once, that is, they are polyphyletic (Johnson, 1975; Parks *et al.*, 1975). These suggestions were made not on the basis of definitive evidence but were instead speculations for which the alternative of monophyly cannot be excluded (see discussion in Endrizzi *et al.*, 1984, 1985). Indeed biogeographic considerations, with the progenitor genomes existing in different hemispheres, are such that polyphyly is extraordinarily unlikely, especially in light of the foregoing evidence for a Pleistocene origin and the implied requirement for trans-oceanic dispersal. The question of monophyly *versus* polyphyly has only recently been addressed using explicit cladistic methods and molecular sequence data. These data demonstrate that all five allopolyploid species possess a cytoplasm descended from a single source (Small and Wendel, 1999; Wendel, 1989), indicating that there was only a single seed parent in the initial hybridization event. Ongoing studies using nuclear (bi-parentally inherited) genes lead to the same conclusion (Wendel, Cronn and Perkey, unpublished). Hence, evidence indicates that allopolyploid cottons formed only once.

A final and puzzling aspect of the parentage of the polyploids concerns the relationship and biogeography of *G. raimondii* and *G. gossypioides*. As a member of the D-genome diploids, *G. raimondii*, from Peru, belongs to an evolutionary lineage that is otherwise Mexican, and the species occupies a cladistically derived position within the subgenus. Hence, *G. raimondii* represents

the descendant taxon of a relatively recent dispersal to South America. Yet is appear to share a cytoplasm with *G. gossypioides* (Wendel and Albert, 1992), which is narrowly distributed in the SW Mexican state of Oaxaca. As detailed in Section III-B, the genome of *G. gossypioides* mysteriously contains a number of repetitive DNAs that are shared with A-genome species but are otherwise unknown among D-genome species (Wendel *et al.*, 1995b; Zhao *et al.*, 1998). As *G. gossypioides* is the only D-genome diploid that exhibits evidence of genetic "contact" with an A-genome plant, it must have acquired these introgressant genomic components after phylogenetic separation from the lineage leading to *G. raimondii*. Hence, the long-distance dispersal event that led to an ephemeral presence of an A-genome entity in the New World may have occurred after *G. gossypioides* diverged from *G. raimondii*, consistent with other indications of a Pleistocene allopolyploid origin. This evolutionary history raises the possibility that the *G. gossypioides* lineage was involved in the origin of allopolyploid cotton, as suggested earlier based on morphological considerations (Valicek, 1983). Indeed, A-genome introgression into *G. gossypioides* and initial allopolyploid formation may have been spatially and temporally associated events, as recently proposed (Wendel *et al.*, 1995b). This scenario, however, is challenged by recent phylogenetic analyses of nuclear genes, which routinely place *G. gossypioides* as basal within the subgenus, distant from a lineage comprising *G. raimondii* and the D-genome of the allopolyploids. Thus, there are conflicting signals from the chloroplast and nuclear genomes with respect to the relationship between the latter two species, a full understanding of the parentage of the polyploids requires reconciliation of this incongruence, as well as a more complete understanding of the events that led to intergenomic contamination of *G. gossypioides*.

V. POLYPLOID EVOLUTION

A. Repeated Cycles of Genome Duplication

In Section I, it was noted that polyploidy is a prominent process in plant speciation, having played a role in generating a relatively high percentage of existing angiosperm species diversity. Because genome doubling has been continuing since flowering plants first definitively appeared in the Cretaceous, many if not most angiosperm genomes have experienced several cycles of polyploidization at various times in the past. The most ancient of these historical genome doubling events may be difficult to discern, due to evolutionary restoration of diploid-like chromosomal behavior and/or other genomic changes following polyploidization. Nonetheless, most angiosperms appear to have "paleopolyploid" genomes, which are revealed as such through comparative

mapping or other approaches (Brubaker *et al.*, 1999b; Devos and Gale, 2000; Gaut and Doebley, 1997; Gómez *et al.*, 1998; Grant *et al.*, 2000; Lagercrantz, 1998; Muravenko *et al.*, 1998; Paterson *et al.*, 2000; Reinisch *et al.*, 1994; Sossey-Alaouni *et al.*, 1998; Wilson *et al.*, 1999). Throughout the angiosperms, more recent polyploidization events have been superimposed on these more ancient genome doubling events, followed often by additional rounds of diploidization and evolutionary divergence among previously doubled genomic sequences. This cyclic process of duplication and divergence leads to a concept of the modern angiosperm genome as one characterized by a series of nested duplications of varying antiquity, only some of which descend to the present relatively unscathed by subsequent evolutionary disruption. Only the most recent genome duplications, such as that of modern allopolyploid cotton, are likely to be classically recognized as constituting polyploid speciation events.

This episodic process of genome doubling is apparent in the genomes of many of our most important crop plants, including *Brassica* (Lagercrantz, 1998; Lagercrantz and Lydiate, 1996), soybean (Grant *et al.*, 2000; Shoemaker *et al.*, 1996), and many important cereals (Bennetzen and Freeling, 1997; Devos and Gale, 2000; Gaut and Doebley, 1997; Gómez *et al.*, 1998; Kellogg, 1998; Moore *et al.*, 1995; Wilson *et al.*, 1999). Comparative genetic mapping studies reveal that the genomes of *diploid* cotton species harbor multiple instances of nested, duplicated chromosome segments, indicating that *allotetraploid* cotton experienced rounds of polyploidization that are more ancient than the most recent one in the Pleistocene (Brubaker *et al.*, 1999b; Paterson *et al.*, 2000; Reinisch *et al.*, 1994). Hence, present AD-genome allotetraploids appear evolutionarily to be at least paleooctaploid.

The suggestion that diploid cotton itself is a paleopolyploid was first made over 70 years ago on the basis of secondary associations that are visible during meiotic metaphase in diploid cotton (Davie, 1933; Lawrence, 1931; Skovsted, 1933, 1937). Lawrence (1931) may have been the first to draw attention to the possibility that this reflects ancient polyploidy, citing observations of Denham (1924). Davie (1933) discussed at length the likelihood of ancient polyploidy, and additionally highlighted the presence of a single pair of chromosomes that was larger than the rest, as would be expected if the ancestral condition was $n = 7$ and with modern diploid cotton having a haploid complement of 13 chromosomes. Later, Abraham (1940) noted that only "about seven pairs of chromosomes seem to be homologous" in interspecific hybrids between the A-genome species *G. arboreum* and the E-genome taxon *G. stocksii*, suggesting that the remaining unpaired chromosomes have a different origin, and that therefore diploid cottons "represent a secondary condition derived from a lower ancestral number." Saunders (1961) cited genetic evidence, noting that diploid cotton contained duplicate genetic factors (for chlorophyll deficiency, among others), suggesting that this may reflect ancient polyploidy. More recently, BrdU-Hoechst–Giemsa chromosome banding techniques have been applied to diploid *Gossypium*,

leading to the recognition of pairs of chromosomes whose banding patterns are sufficiently similar that they were hypothesized to represent homoeologues from an ancient polyploidization event (Muravenko *et al.*, 1998). Muravenko *et al.* (1998) suggested that "the ancestral cotton genome contained seven homologous chromosome pairs" (presently, $n = 13$ in diploid cotton). In this respect we note that the entire tribe to which *Gossypium* belongs (the *Gossypieae*) is based on a chromosome number of $n = 13$, implying that the paleopolyploidization event evidenced in the chromosome banding and genetic mapping data occurred prior to the origin of the tribe, which may be 20–40 million years old (Seelanan *et al.*, 1997). Paleopolyploidization may conceivably be even more ancient, perhaps even antedating the origin of the Malvaceae.

These studies reveal a history of both ancient and recent polyploidization in *Gossypium*, events that must have profoundly impacted its morphological, ecological and physiological diversification. Given this central significance, it is of interest to attempt to evaluate the molecular evolutionary consequences of allopolyploidy, and to distinguish the phenomena and processes that might characterize the earliest stages of polyploid formation from those that are responsible for longer-term genomic changes. The most experimentally tractable examples of duplication trace to the most recent allopolyploidization event; this was the Pleistocene merger of two differentiated genomes (A and D) into a single nucleus in only one of the two parental cytoplasms (that of the A-genome parent, Small and Wendel, 1999; Wendel, 1989). Though we know relatively little about many if not most details of polyploidization, the initial stages of biological reunion evidently are molded by an array of molecular genetic mechanisms and processes that collectively lead to polyploid stabilization (Table III). This suite of mechanisms and processes may be rather different, at least in part, from those responsible for longer-term gene and genome evolution in polyploids, such as functional diversification of duplicated genes. In the following we review recent efforts to understand the allopolyploid genome of *Gossypium*, with a focus on the evolutionary consequences of gene and genome doubling.

Table III
Molecular Evolutionary Phenomena that Characterize Allopolyploid Evolution in *Gossypium*

Molecular evolutionary phenomenon	References
Chromosome stabilization	Endrizzi *et al.* (1985)
Enhanced genetic recombination	Brubaker *et al.* (1999b)
Interlocus concerted evolution	Wendel *et al.* (1995a)
Independent evolution of duplicated genes	Cronn *et al.* (1996); Liu *et al.* (2001a); Small and Wendel (2000a) and Small *et al.* (1998)
Intergenomic epistasis	Jiang *et al.* (2000b)
Unequal expression of homoeologues	Wendel lab (unpublished)
Intergenomic "horizontal" transfer	Hanson *et al.* (1998) and Zhao *et al.* (1998)

B. Chromosomal Stabilization

Classical cytogenetic evidence indicates that the A- and D-genomes of allopolyploid *Gossypium* are more distinct from one another than are the descendents of their diploid progenitors (reviewed by Endrizzi *et al.*, 1985). For example, allopolyploid-derived haploids form an average of less than one bivalent per cell at meiotic metaphase, whereas chromosomes in hybrids between extant A- and D-genome diploids average 5.8 and 7.8 bivalents. These and similar observations indicate that natural selection has favored the evolution of mechanisms that promote exclusive bivalent formation in the allopolyploid. Neither the pace at which such mechanisms operate nor their nature are understood, but it seems rational to postulate that selection would be most intense in the first generations following allopolyploid formation, where the fitness cost of unbalanced gametes would be the greatest.

Perhaps one component of genome stabilization after polyploidization is reorganization of the two resident genomes so that they are no longer capable of homoeologous pairing. To evaluate this possibility Brubaker *et al.* (1999b) employed a common set of RFLP probes and created genetic maps, using interspecific F_2 progenies for both diploid and allopolyploid cottons (see also Reinisch *et al.*, 1994). Direct comparisons of gene order and synteny among the A- and D-genome maps, as well as those for both genomes of the allopolyploid, permitted direct assessment of the types and magnitudes of structural changes that preceded and followed allopolyploid formation. As expected from the chromosome numbers, 13 sets of homoeologous linkage groups were identified. Map comparisons showed that the two reciprocal translocations that had previously been identified as distinguishing the A-genome diploids from their counterpart chromosomes in the allopolyploid (Endrizzi *et al.*, 1985) arose in the diploid lineage after allopolyploid formation. More importantly, only 19 locus order differences (inversions) were detected among the two diploid and two allotetraploid genomes, and conservation of colinear linkage groups was the rule rather than the exception. Thus, allopolyploidy in *Gossypium* was not accompanied by extensive chromosomal rearrangement. One implication of this and similar studies (Paterson *et al.*, 2000) is that in many cases gross structural rearrangement may not be a particularly significant aspect of the process of polyploid genome stabilization.

C. Increased Recombination in Polyploid *Gossypium*

Comparative genetic mapping experiments are useful not only for detecting evolutionary changes in synteny and gene order, but also provide valuable information on recombination. As noted earlier, A-genome diploids have twice the DNA content per cell as D-genome diploids, with a corresponding difference

in chromosome size. To a large extent, these differences are maintained in allopolyploid *Gossypium* (Endrizzi *et al.*, 1985; Skovsted, 1934), although DNA content is not quite additive and the A-genome of allopolyploid cotton has chromosomes that are slightly smaller than those of their diploid antecedent (Davie, 1933; Endrizzi *et al.*, 1985).

This nearly twofold difference in genome size provides a natural experiment on the relationship between genome size and recombination. These two genomic attributes often are uncoupled (e.g., Ahn *et al.*, 1993; Whitkus *et al.*, 1992), and they appear to be in *Gossypium* as well. Brubaker *et al.* (1999b) showed that recombination rates are essentially conserved between the A- and D-diploid genomes; for a common set of markers, the genetic length of these two genomes by only about 6%. Similarly, at the tetraploid level, recombination in the two resident genomes differed by only 5%. This latter result verifies previous reports of a lack of correlation between genome size and total recombination in a particularly satisfying way, in that the two allopolyploid genomes are in the same nucleus, thereby controlling for the myriad life history, population genetic, and ecological covariables that might be suspected of affecting recombination rates.

Although there is no significant difference in recombination between genomes that vary in size by a factor of two, at either the diploid or allopolyploid level, Brubaker *et al.* (1999b) reported increased recombination in allopolyploid cotton. Specifically, the A- and D-genomes of allopolyploid cotton had, for a common set of markers, map lengths that were 52 and 59% higher, respectively, than those of their diploid counterparts. This suggests that polyploidy itself has promoted higher rates of recombination in *Gossypium*. At present, neither the generality of this conclusion nor the responsible mechanism are understood, but the hypothesis that polyploidy is recombinogenic bears further investigation both in *Gossypium* and in other systems.

D. A Diverse Array of Genic and Genomic Interactions

The most immediate and important genomic consequence of allopolyploid formation in *Gossypium* was simultaneous duplication of all nuclear genes. Many have addressed the potential evolutionary significance of gene duplication (Levin, 1983; Lewis, 1980; Ohno, 1970; Stebbins, 1950; Stephens, 1951a,b). As polyploidy is being so significant in cotton, insight into mechanisms of genomic change in polyploids may ultimately lead to applied benefits (Section VI). Theory suggests that one possible outcome of gene duplication will be relaxation of selection, allowing divergence between the duplicated genes (homoeologues) and the acquisition of new function (Ferris and Whitt, 1979; Hughes, 1994; Hughes *et al.*, 2000; Li, 1985; Lynch and Conery, 2000; Lynch and Force, 2000; Ohno, 1970). Indeed, polyploidization is widely perceived to provide the raw material for the origin of physiological, ecological (Section V)

and morphological novelty (Barrier *et al.*, 1999; Grant, 1981; Levin, 1983; Lewis, 1980; Schranz and Osborn, 2000; Stebbins, 1950). In this regard Stephens' comment half a century ago about the possible significance of divergence following the duplication in cotton is noteworthy: "One might expect (still on *a priori* grounds) that a mechanism in which new functions could be *added* and the old ones *retained* would have considerable selective advantage" (Stephens, 1951b). An alternative outcome of gene doubling is that one member of the duplicated gene pair will become silenced and ultimately degenerate as a pseudogene (Lynch and Conery, 2000; Wendel, 2000). Duplicated genes may also maintain their original function, or aggregate function may be partitioned between the two duplicates (Force *et al.*, 1999; Lynch and Force, 2000).

From a phylogenetic perspective, these various fates of gene duplication may partially be modeled as shown in Fig. 2. The null hypothesis for sequence evolution in allopolyploids derives from the organismal history, if both duplicated genes evolve independently and at equal rates following allopolyploid formation, then each homoeologue should be phylogenetically sister to its orthologue from a diploid cotton, rather than to the other homoeologue. Similarly, if rates of sequence evolution are similar at the diploid and allopolyploid level, branch lengths for the two A-genome sequences (one from the diploid, the other from the allopolyploid) should be similar, as they should for the two D-genome sequences (Fig. 2, center). The utility of the null hypothesis lies in its falsification; if homoeologous sequences interact, for example (Fig. 2, top), a different tree may be recovered, or if there is strong directional selection or pseudogenization (Fig. 2, bottom), rate inequalities may become evident. Additional possibilities include silencing or loss of one of the duplicated copies (Fig. 2, top right) and "horizontal transfer" of sequences from one genome to the other (Fig. 2, bottom right).

Recently there have been several tests of the null expectations of independence and rate equality in allopolyploid *Gossypium* as well as in other plants (reviewed by Wendel, 2000). Wendel *et al.* (1995a) demonstrated interaction among the 18S–26S ribosomal genes that exist at multiple loci in the A- and D-genomes (Ji *et al.*, 1999). Specifically, instead of evolving independently, as expected if they were sequestered in separate genomes of diploid plants, repeats at the different loci in allopolyploid cotton become homogenized to the same sequence (either "A-like" or "D-like") by one or more processes of concerted evolution (reviewed by Elder and Turner, 1995). In four of the five allopolyploid species, interlocus homogenization has created exclusively D-genome like rDNAs, whereas in *G. mustelinum* nearly all rDNA repeats have been homogenized to an A-like form. This example showed that since polyploid formation in the Pleistocene, some 3800 repeats, each approximately 10 kb in length, were "overwritten" with the alternative form originating from the other parental genome, probably through unequal crossing over or gene conversion. Moreover, interlocus concerted evolution was bi-directional, operating in different directions in different allopolyploid lineages.

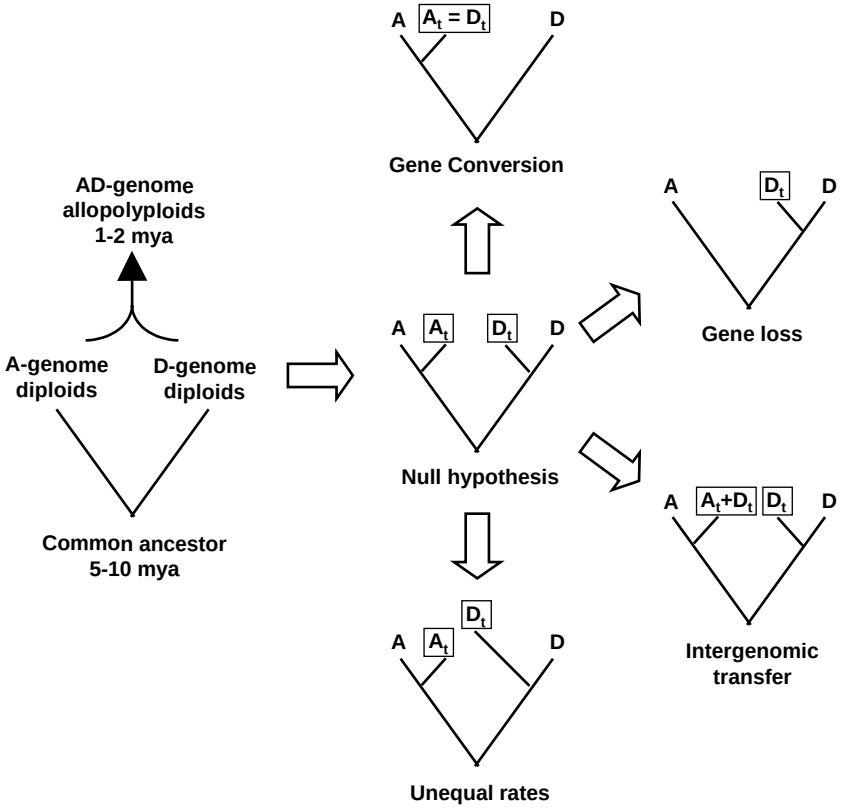


Figure 2 Phylogenetic history of diploid and allopolyploid *Gossypium* species (left) and the various possibilities for the evolution of duplicated genes. Allopolyploids are expected to have duplicated copies (A_t and D_t) of most single-copy and low-copy genes, and duplicated suites of similar repetitive DNAs. In the absence of mutation or selection, homoeologous copies are expected to evolve at equivalent rates and independently of one another, such that they are phylogenetically sister to their counterparts from the progenitor diploids rather than to each other (center). This expectation provides a convenient null hypothesis for diagnosing molecular evolutionary phenomena that accompany genome doubling, such as gene conversion (top), accelerated evolutionary rates (bottom), transfer of sequences between genomes (bottom right), and gene silencing or gene loss (top right). See text for details.

This demonstration that some repeated sequences could interact across genomes in the allopolyploid nucleus led to additional investigations of the scope of the phenomenon. In an analogous study, Cronn *et al.* (1996) showed that the duplicated arrays of tandemly repeated 5S rDNA genes are not homogenized by concerted evolutionary forces in the allopolyploid, in contrast to the 18S–26S arrays. Similarly, low-copy nuclear genes duplicated by allopolyploidy appear to largely evolve independently of one another in the polyploid nucleus (Cronn *et al.*, 1999; Small and Wendel, 2000a). In fact, to date there has been no

convincing demonstration of interlocus gene conversion for single-copy or low-copy nuclear genes in cotton, although PCR-mediated recombinants are often recovered (Cronn *et al.*, 2002a). More generally, the physiological significance of these observations is not clear. Perhaps interlocus interactions of the kind detectable by changes in duplicated gene sequences have not played a particularly prominent role in generating novel physiology, ecology or morphology in the allopolyploids. Perhaps instead other sorts of genomic interactions have been more important.

Clues into the nature of these interactions may possibly be evidenced in the results of other recent investigations. In a study of experimental backcross populations between *G. hirsutum* and *G. barbadense*, Jiang *et al.* (2000b) noted large deficiencies of donor parent (*G. barbadense*) transmission for some chromosomal regions. They attributed this to epistatic interactions affecting chromatin transmission, a high proportion of which were caused by interactions between alleles contributed by the two genomes. Other data show that dispersed repetitive elements have become mobilized as a consequence of polyploidization in cotton, possibly leading to novel regulatory changes or gene functions. The studies of Zhao *et al.* (1998) and Hanson *et al.* (1998) are noteworthy in this respect, using florescent *in situ* hybridization, they showed that dispersed repetitive sequences that are A-genome-specific at the diploid level have colonized the D-genome at the polyploid level. Similarly, Hanson *et al.* (1999) showed that a family of *copia*-like retrotransposable elements "horizontally" transferred across genomes following allopolyploid formation. These and other studies highlight the evolutionary possibility of transposable element spread across genomes following polyploid formation, and raise the possibility that this process has played a role in diversification and adaptation.

In *Gossypium* these intergenomic interactions appear to arise on an evolutionary timescale as opposed to being an immediate consequence of hybridization and polyploidization (Feldman *et al.*, 1997; Liu *et al.*, 1998a,b; Ozkan *et al.*, 2001; Shaked *et al.*, 2001; Song *et al.*, 1995). Liu *et al.* (2001a) used AFLP analysis to evaluate the extent of fragment additivity in nine sets of newly synthesized allotetraploid and allohexaploid *Gossypium*. Approximately 22,000 genomic loci were examined, yet fragment additivity was observed in nearly all cases, even when methylation sensitive and insensitive isoschizomers were used. These indications of genomic additivity and epigenetic stasis during allopolyploid formation provide a contrast to recent evidence from several model plant allopolyploids, most notably wheat (Feldman *et al.*, 1997; Liu *et al.*, 1998a,b; Ozkan *et al.*, 2001; Shaked *et al.*, 2001) and *Brassica* (Song *et al.*, 1995), where rapid and unexplained genomic changes have been reported. In addition, the data contrast with the foregoing account of repetitive DNAs in *Gossypium*, some of which are subject to non-Mendelian molecular evolutionary phenomena such as interlocus concerted evolution and intergenomic colonization. Collectively, these and other recent studies have drawn attention to the "dynamic" (Soltis and Soltis,

1995) nature of polyploids, and underscored the relatively poorly understood and sometimes non-Mendelian mechanisms that may characterize gene and genome evolution in polyploids.

In addition to evolutionary changes in *gene and genome structure*, a key component of polyploid evolution concerns the consequences of genome doubling on *gene expression*. This is a largely unexplored area in *Gossypium*, although ongoing studies indicate a range of responses from full co-expression to gene silencing (Wendel Lab, unpublished). In older polyploids, it is well documented that there is a slow decay of duplicate gene expression due to the deletional/substitutional processes of pseudogenization (Gottlieb, 1982; Haufler, 1987; Soltis and Soltis, 1989). In addition, hybridization and/or polyploidization may stimulate more rapid epigenetic changes that lead to bursts of transposable element activity (Liu and Wendel, 2000; O'Neill *et al.*, 1998), although as discussed above there is little evidence for this in nascent *Gossypium* allopolyploids. Nonetheless, genomic incompatibilities of various kinds and epigenetic responses may accompany polyploid formation or stabilization, as has been shown in *Arabidopsis* (Comai, 2000; Comai *et al.*, 2000; Lee and Chen, 2001) and *Brassica* (Chen and Pikaard, 1997). It is likely that different mechanisms affect gene expression evolution and gene silencing in nascent *versus* stabilized allopolyploids, and equally probable that these processes have been evolutionarily significant (Sections V and VI).

E. Differential Evolution of Cohabiting Genomes

Following the union of two genomes into a single nucleus as a consequence of allopolyploidization it is expected that over time some genes will become mutagenized into pseudogenes whereas others may diverge and acquire new function, as discussed above. On an average, however, one would expect that these and other phenomena that impact the molecular evolution of genes would be equally distributed in the two allopolyploid genomes. This leads to a useful null hypothesis, i.e., evolutionary rates will be equivalent for duplicated homoeologues. A corollary expectation is that both gene copies will accumulate infraspecific diversity at an equivalent rate. In any single case this need not be true, of course, as when there is strong directional selection on one gene copy or pseudogene formation. Nonetheless, the model may be helpful in informing a search for the underlying explanation for differential evolutionary rates or different levels of diversity when these are observed. For example, if one homoeologue becomes pseudogenized while the other remains under purifying selection, then nucleotide diversity is expected to increase in the former locus at a faster rate than in the latter. The fact that duplicated genes reside in the same nucleus greatly simplifies the challenge of isolating potentially important genomic forces from population-level factors that might effect patterns of

diversity, such as breeding system or effective population size. As population-level factors are neutral with respect to the two homoeologues, observed differences in diversity are more easily attributed to genetic or genomic processes.

Gossypium allopolyploids offer a powerful model for these explorations, particularly in as much as the two genomes are known to be largely co-linear yet differ in genome size by a factor of two. An early suggestion of unequal evolutionary rate for the A- and D-genomes was stimulated by the observation that synthetic A-genome × D-genome hybrids can be synthesized exclusively with the A-genome parent as female. Phillips (1963), on noting this, speculated: "If, in the ancestral amphidiploid the D-genome was contained in A-cytoplasm; during "shakedown" of the raw amphidiploid the D-genome might have been genetically and chromosomally more unstable than the A-genome leading to a more rapid genetical and cytological diploidization of the D-genome of the allopolyploid."

This remarkable comment, made nearly four decades ago, has surprisingly found recent support from divergent quarters (Table IV). For example, in a survey of RFLP polymorphism levels detected in allopolyploid cotton, 10% more probes revealed polymorphisms in the D-genome than the A-genome (Reinisch *et al.*, 1994). Similarly, in two independent phylogenetic analyses (Liu *et al.*, 2001b; Small *et al.*, 1998), D-genome sequences in the allopolyploids were found to have longer branches (i.e., faster evolutionary rates) than their homoeologous A-genome sequences. Moreover, inferences of the location of loci controlling quantitative characters repeatedly suggest a higher evolutionary rate in the D-genome than the A-genome. This was shown for fiber-related traits, where 10 of 14 QTLs were located in the D-genome (Jiang *et al.*, 1998), disease resistance, where five of six resistance genes were localized to the D-genome (Wright *et al.*, 1998), and leaf morphology, where 14 of 21 QTL mapped to the D-genome (Jiang *et al.*, 2000a).

A more direct test of the null hypothesis of rate equivalence for homoeologous genes is provided by the measures of nucleotide diversity levels. If evolutionary forces are equivalent for duplicated genes, mutations should accumulate

Table IV
Evidence for Differential Rates of Evolution for the A- and D-Genomes in Allopolyploid *Gossypium*

Type of data or evidence	References
Artificial hybridization data	Phillips (1963)
RFLP polymorphism levels	Reinisch *et al.* (1994)
Branch lengths in phylogenetic analyses	Liu *et al.* (2001b) and Small *et al.* (1998)
QTL analysis	Jiang *et al.* (1998, 2000) and Wright *et al.* (1998)
Surveys of nucleotide polymorphism levels	Small *et al.* (1999) and Small and Wendel (2002)

randomly with respect to homoeologue, and hence in a survey of allelic polymorphism in a sample of individuals, the number of alleles detected should be approximately equal for the two gene copies. This was the approach used by Small *et al.* (1999) in a study of approximately 1 kb of *AdhA* sequence for 22 accessions (44 alleles per genome) of *G. hirsutum* and for five accessions (10 alleles per genome) of *G. barbadense*. In both allopolyploid species, estimates of nucleotide diversity were higher for *AdhA* from the D-genome than from the A-genome, by a factor of two or more. In a follow-up study wherein a 1.3 kb section was sequenced of a second alcohol dehydrogenase gene (*AdhC*) with a faster overall evolutionary rate, the same conclusion was even more emphatically reached (Small and Wendel, 2002). In a survey of 44 alleles from each genome of *G. hirsutum*, 24 different alleles were detected for the D-genome homoeologue whereas only seven allelic variants of the A-genome sequence were observed. To evaluate whether this was a species-specific effect, 12 alleles were sequenced from each genome of a second allopolyploid species, *G. barbadense*. Although diversity levels were lower, the same phenomenon of differential diversity was observed, with three and one alleles detected for the D- and A-genome homoeologues, respectively.

These observations collectively suggest that there has been an overall acceleration in evolutionary rate in the D-genome relative to the A-genome of allopolyploid *Gossypium*. Although this rate enhancement is not always observed (Cronn *et al.*, 1999), the emerging picture is that evolutionary forces operating on the two genomes may be fundamentally different. At present, the responsible forces and underlying molecular mechanisms are obscure, but a logical suggestion is that they are causally connected to the nearly twofold difference in genome size.

VI. ECOLOGICAL CONSEQUENCES OF POLYPLOIDIZATION

The foregoing discussion of the genomic and genetic attributes of allopolyploid cotton demonstrates that polyploid formation has led to a diverse array of genetic and genomic responses, including non-Mendelian transmission. The question naturally arises as to the selective consequences of genome duplication: has allopolyploidy stimulated novel adaptation or physiological capacity? A voluminous literature in plants documents the frequency of polyploids in various habitats, their morphological and physiological attributes, and their ecological success relative to diploids (reviewed by Grant, 1981; Soltis and Soltis, 2000; Stebbins, 1950, 1971). One generalization that has emerged from this accumulated body of work is that polyploidy often is associated with broader ecological amplitude and novel evolutionary opportunity, perhaps

mediated by the increased "buffering" capacity afforded by duplicated genes and the enhanced vigor resulting from the "fixed heterozygosity" of their duplicated genomes. Perhaps, some of the more recently discovered genetic phenomena discussed above also play a role in the success of polyploids, but this direct connection remains to be demonstrated.

With respect to *Gossypium*, allopolyploidy led to the establishment of a new and successful clade as well as the apparent invasion of a new ecological niche. In considering the Pleistocene origin of allopolyploid cotton, Fryxell (1965, 1979) noted that in contrast to the majority of diploid species, allopolyploid species typically occur in coastal habitats, at least those forms that arguably are truly wild (see also Brubaker and Wendel, 1994). Thus, among the five allopolyploid species, two are completely restricted to near coastlines, in that they are island endemics (*G. darwinii* and *G. tomentosum*), and for two others (*G. barbadense* and *G. hirsutum*), wild forms occur disparately in littoral habitats ringing the Gulf of Mexico, northwest South America, and even on distant Pacific Islands. We have previously mentioned this capacity for oceanic dispersal (Fryxell, 1965, 1979; Stephens, 1958, 1966), but here draw attention to the observation that in the case of allopolyploid *Gossypium*, this dispersal capacity was associated with specialization for establishment in coastal communities. Fryxell (1965, 1979) forwarded the tantalizing suggestion that following initial formation, adaptation of the newly evolved allopolyploid to littoral habitats enabled it to exploit the fluctuating sea levels that characterized the Pleistocene. This ecological innovation is envisioned to have not only permitted the initial establishment of the nascent polyploid lineage, but is also suggested to have provided a means for the rapid dispersal of the salt–water tolerant seeds. By this means, perhaps, the mobile shorelines of the Pleistocene facilitated the exploitation of a new ecological niche, and hence colonization of the New World tropics.

VII. POLYPLOIDY AND FIBER

A final consequence of polyploidy is one of paramount agronomic importance, as it concerns fiber. As noted in Section I, four separate species of *Gossypium* were independently domesticated for their seed hairs. The characteristic that attracted the attention of the earliest domesticators, the seed lint itself, however, evolved only once in the progenitor of all four cotton species. This becomes evident from the account of organismal taxonomy and phylogeny provided in Section II, which highlights the fact that the ancestral condition of *Gossypium* species is to have seeds with epidermal seed trichomes that typically are short and tightly adherent to the seed. While mature seeds from wild species exhibit great diversity in fiber length, color, and other properties; it has recently been shown that the earliest developmental stages are similar among all species (Applequist

et al., 2001). Indeed, there is little difference in fiber density or developmental profile for the first several days following anthesis.

To identify developmental differences that might account for variation in fiber length and to place these differences in a phylogenetic context, Applequist *et al.* (2001) conducted scanning electron microscopy of ovules at and near the time of flowering, and generated growth curves for cultivated and wild diploid and allopolyploid species. Trichome initiation was found to be similar in all species, with few notable differences in fiber density or early growth. Developmental profiles of the fibers of most wild species are similar, with fiber elongation terminating at about two weeks post-anthesis. In contrast, growth is extended to 3 weeks in the A- and F-genome diploids. When this observation is considered in light of the phylogeny of the genus (Fig. 1), it becomes clear that this prolonged elongation period represents a key evolutionary event in the origin of long fiber, and that it happened in the common ancestor of these two groups of diploid cottons prior to domestication in Africa. This observation has a fascinating implication; namely, that the domestication of the New World cottons that presently dominate cotton agriculture worldwide was first precipitated by a developmental switch that occurred millions of years ago in a different hemisphere.

Analysis of fiber growth curves reveals that domestication itself has been associated with further prolongation of elongation at both the diploid and allopolyploid levels. This provokes the speculation that the effects of parallel artificial selection for long fiber in the four cultivated species resulted in a genetically convergent or parallel transformation in the developmental program that is responsible for this aspect of fiber development. Applequist *et al.* (2001) further showed that a second evolutionary innovation in fiber morphology was that absolute growth rate is higher in species with long fibers. Thus, wild forms of A-genome diploids, for example, have fiber elongation rates that exceed that of their closest relatives with shorter fiber.

A final intriguing observation from Applequist *et al.* (2001) emerges from noting that fiber growth curves for wild AD-genome allopolyploids are similar to those of the wild A-genome species, but that the fiber of the cultivated allopolyploids is superior to that of the cultivated Old Word diploids. Also, domestication at the allopolyploid level was shown not only to have prolonged the elongation period beyond three weeks, but was also demonstrated to have increased the growth rate in the early stages of trichome expansion. Perhaps, these shifts in developmental profiles were mediated by recruitment of novel expression patterns for D-genome genes, or perhaps novel expression of A-genome genes. It may be well that the genome-wide gene duplication caused by allopolyploidization provided the raw material necessary for the evolution of novel gene expression patterns, which subsequently were exploited by the aboriginal domesticators (and perhaps modern plant breeders) of *G. hirsutum* and *G. barbadense*.

In light of the foregoing, several recent studies (Jiang *et al.*, 1998, 2000b; Wright *et al.*, 1998) have suggested that allopolyploidization provided novel opportunities for agronomic improvement. Jiang *et al.* (1998), for example, used quantitative genetic approaches (QTL analysis) to study fiber-related traits in an interspecific F_2 progeny derived from a cross between *G. hirsutum* and *G. barbadense*. They reported that a majority of loci affecting fiber yield and quality are found in the D-genome rather than the A-genome, possibly explaining the superiority of the lint of the allopolyploids. These studies may constitute actual genetic evidence for a speculation forwarded 65 years ago by Harland (1936), who stated "If as a consequence of polyploidy a large number of genes become duplicated, and the characters governed by such genes are of importance to the species, one of the members may mutate, leaving the character unimpaired, with the further possibility that the mutation may be of benefit to the species." An exciting prospect is that in the near future we will obtain information on the nature of these genes and the mutational or regulatory changes that underlie altered morphology and agronomic improvement.

VIII. CONCLUDING REMARKS

In this review we have attempted to provide a synthesis of the evolutionary history of cotton and its genome. Quite remarkably, this history has included multiple episodes of trans-oceanic dispersal and a surprisingly high frequency of natural interspecific hybridization among lineages that presently are both geographically isolated and intersterile. Moreover, the resulting genomic reunions have led to an array of genetic mechanisms and adaptive responses that are not yet fully understood. We note with wonder the many implausibilities and improbabilities involved in this account, which were revealed over the decades by innumerable investigators in diverse disciplines. The exploration of *Gossypium* and its genome has truly been an interdisciplinary enterprise, enriched by investigations at all levels of biological organization, from molecular to ecological. Indeed, insights into the evolutionary history of *Gossypium* and its genome have not only been deepened by but also have been dependent upon what we termed in Section I "the happy marriage of phylogenetic analysis with genomic investigations."

It seems likely that additional insights will continue to emerge from this interplay between molecular biology and evolutionary systematics. Our understanding of allopolyploid formation is still in its infancy, as relatively little is known about the myriad consequences of genomic merger and the attendant short-term and long-term effects on patterns of gene expression and development. This by necessity limits our understanding of the effects of gene and genome doubling on morphological or physiological attributes and agronomic potential.

Powerful new technologies are beginning to be brought to bear on the problem, such as global expression screens using comparative microarray or cDNA–AFLP analysis. As documented in the present account of the history of our understanding of *Gossypium*, the most profound insights may emerge from a phylogenetically informed implementation of these and related technologies.

REFERENCES

Abdalla, A. M., Reddy, O. U. K., El-Zik, K. M., and Pepper, A. E. (2001). Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. *Theor. Appl. Genet.* **102**, 222–229.

Abraham, P. (1940). Cytological studies in *Gossypium*. I. Chromosome behavior in the interspecific hybrid *G. arboreum* × *G. stocksii*. *Ind. J. Agric. Sci.* **10**, 285–298.

Ahn, S., Anderson, J. A., Sorrells, M. E., and Tanksley, S. D. (1993). Homoeologous relationships of rice, wheat and maize chromosomes. *Mol. Gen. Genet.* **241**, 483–490.

Anonymous, (1997). *In* "Agricultural Statistics Board, Annual Crop Summary". NASS, USDA.

Applequist, W. L., Cronn, R. C., and Wendel, J. F. (2001). Comparative development of fiber in wild and cultivated cotton. *Evol. Dev.* **3**, 3–17.

Barrier, M., Baldwin, B. G., Robichaux, R. H., and Purugganan, M. D. (1999). Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Mol. Biol. Evol.* **16**, 1105–1113.

Beasley, J. O. (1940). The origin of American tetraploid *Gossypium* species. *Am. Nat.* **74**, 285–286.

Beasley, J. O. (1941). Hybridization, cytology, and polyploidy of *Gossypium*. *Chronica Botanica* **6**, 394–395.

Bennett, M. D., Smith, J. B., and Heslop-Harrison, J. S. (1982). Nuclear DNA amounts in angiosperms. *Proc. R. Soc. Lond. B* **216**, 179–199.

Bennett, M. D., Cox, A. V., Leitch, I. J (1997). Angiosperm DNA *c*-Values Database, http://www.rbgkew.org.uk/cval/database1.html.

Bennetzen, J. L., and Freeling, M. (1997). The unified grass genome: synergy in synteny. *Genome Res.* **7**, 301–306.

Brubaker, C. L., and Wendel, J. F. (1993). On the specific status of *Gossypium lanceolatum* Todaro. *Genet. Res. Crop Evol.* **40**, 165–170.

Brubaker, C. L., and Wendel, J. F. (1994). Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Am. J. Bot.* **81**, 1309–1326.

Brubaker, C. L., Bourland, F. M., and Wendel, J. F. (1999a). The origin and domestication of cotton. *In* "Cotton: Origin, History, Technology and Production" (C. W. Smith and J. T. Cothren, Eds.), pp. 3–31. Wiley, New York.

Brubaker, C. L., Paterson, A. H., and Wendel, J. F. (1999b). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome* **42**, 184–203.

Chen, Z. J., and Pikaard, C. S. (1997). Transcriptional analysis of nucleolar dominance in polyploid plants: Biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. *Proc. Natl. Acad. Sci. USA* **94**, 3442–3447.

Cherry, J. P., Katterman, F. R. H., and Endrizzi, J. E. (1970). Comparative studies of seed proteins of species of *Gossypium* by gel electrophoresis. *Evolution* **24**, 431–447.

Cherry, J. P., Katterman, F. R. H., and Endrizzi, J. E. (1972). Seed esterases, leucine aminopeptidases and catalases of species of the genus *Gossypium*. *Theor. Appl. Genet.* **42**, 218–226.

Comai, L. (2000). Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol. Biol.* **43**, 387–399.

Comai, L., Tyagi, A. P., Winter, K., Holmes-Davis, R., Reynolds, S. H., Stevens, Y., and Byers, B. (2000). Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* **12**, 1551–1567.

Cronn, R. C., Zhao, X., Paterson, A. H., and Wendel, J. F. (1996). Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *J. Mol. Evol.* **42**, 685–705.

Cronn, R., Small, R. L., and Wendel, J. F. (1999). Duplicated genes evolve independently following polyploid formation in cotton. *Proc. Natl. Acad. Sci. USA* **96**, 14406–14411.

Cronn, R. C., Cedroni, M., Haselkorn, T., Grover, C., and Wendel, J. F. (2002a). PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor. Appl. Genet.* (in press).

Cronn, R. C., Small, R. L., Haselkorn, T., and Wendel, J. F. (2002b). Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *Am. J. Bot.* **89**, 707–725.

Davie, J. H. (1933). Cytological studies in the Malvaceae and certain related families. *J. Genet.* **28**, 33–67.

DeJoode, D. R (1992). Molecular insights into speciation in the genus *Gossypium* L (Malvaceae). MS Thesis, Iowa State University, Ames.

DeJoode, D. R., and Wendel, J. F. (1992). Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. *Am. J. Bot.* **79**, 1311–1319.

Denham, H. J. (1924). The cytology of the cotton plant. II. Chromosome numbers of old and new world cottons. *Ann. Bot.* **38**, 433–438.

Devos, K. M., and Gale, M. D. (2000). Genome relationships: the grass model in current research. *The Plant Cell* **12**, 637–646.

Edwards, G. A., and Mirza, M. A. (1979). Genomes of the Australian wild species of cotton. II. The designation of a new G genome for *Gossypium bickii*. *Can. J. Genet. Cytol.* **21**, 367–372.

Edwards, G. A., Endrizzi, J. E., and Stein, R. (1974). Genomic DNA content and chromosome organization in *Gossypium*. *Chromosoma* **47**, 309–326.

Elder, J. F., and Turner, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.* **70**, 297–320.

Endrizzi, J. E., Turcotte, E. L., and Kohel, R. J. (1984). *Cotton, Agronomy Monograph No. 24. In* "Qualitative genetics, cytology, and cytogenetics", pp. 81–129. ASA–CSSA–SSSA, Madison.

Endrizzi, J. E., Turcotte, E. L., and Kohel, R. J. (1985). Genetics, cytogenetics, and evolution of *Gossypium*. *Adv. Genet.* **23**, 271–375.

Endrizzi, J. E., Katterman, F. R. H., and Geever, R. F. (1989). DNA hybridization and the time of origin of three species of *Gossypium*. *Evol. Trends Plants* **3**, 115–119.

Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A. A., and Vega, J. M. (1997). Rapid elimination of low-copy DNA sequences in polyploid wheat: A possible mechanism for differentiation of homoeologous chromosomes. *Genetics* **147**, 1381–1387.

Ferris, S. D., and Whitt, G. S. (1979). Evolution of the differential regulation of duplicate genes after polyploidization. *J. Mol. Evol.* **12**, 267–317.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. -L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545.

Fryxell, P. A. (1965). Stages in the evolution of *Gossypium*. *Adv. Front. Plant Sci.* **10**, 31–56.

Fryxell, P. A. (1968). A redefinition of tribe *Gossypieae*. *Bot. Gazette* **129**, 296–308.

Fryxell, P. A. (1971). Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L (Malvaceae). *Evolution* **25**, 554–562.

Fryxell, P. A. (1979). *In* "The Natural History of The Cotton Tribe". Texas A&M University Press, College Station, TX.

Fryxell, P. A. (1992). A revised taxonomic interpretation of *Gossypium* L (Malvaceae). *Rheedea* **2**, 108–165.

Fryxell, P. A., Craven, L. A., and Stewart, J. M. (1992). A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. *Syst. Bot.* **17**, 91–114.

Galau, G. A., and Wilkins, T. A. (1989). Alloplasmic male sterility in AD allotetraploid *Gossypium hirsutum* upon replacement of its resident A cytoplasm with that of D species *G. harknessii*. *Theor. Appl. Genet.* **78**, 23–30.

Gaut, B. S. (1998). Molecular clocks and nucleotide substitution rates in higher plants. *Evolutionary Biology* (M. K. Hecht Ed.), **vol. 30**, pp. 93–120. Plenum Press, New York.

Gaut, B. S., and Doebley, J. F. (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc. Natl. Acad. Sci. USA* **94**, 6808–6814.

Geever, R. F., Katterman, F. R. H., and Endrizzi, J. E. (1989). DNA hybridization analyses of a *Gossypium* allotetraploid and two closely related diploid species. *Theor. Appl. Genet.* **77**, 553–559.

Gerstel, D. U. (1953). Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* **7**, 234–244.

Gerstel, D. U. (1958). Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* **7**, 234–244.

Gerstel, D. U., and Phillips, L. L. (1958). Segregation of synthetic amphiploids in *Gossypium* and *Nicotiana*. *Cold Spring Harbor Sympp. Quant. Biol.* **23**, 225–237.

Gómez, M. I., Islam-Faridi, M. N., Zwick, M. S., Czeschin, D. G., Hart, G. E., Wing, R. A., Stelly, D. M., and Price, H. J. (1998). Tetraploid nature of *Sorghum bicolor* (L.) Moench. *J. Hered.* **89**, 188–190.

Gottlieb, L. D. (1982). Conservation and duplication of isozymes in plants. *Science* **216**, 373–380.

Grant, V. (1981). *In* "Plant Speciation". Columbia University Press, New York.

Grant, D., Cregan, P., and Shoemaker, R. C. (2000). Genome organization in dicots: Genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **97**, 4168–4173.

Hanson, R. E., Zhao, X. -P., Islam-Faridi, M. N., Paterson, A. H., Zwick, M. S., Crane, C. F., McKnight, T. D., Stelly, D. M., and Price, H. J. (1998). Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). *Am. J. Bot.* **85**, 1364–1368.

Hanson, R. E., Islam-Faridi, M. N., Crane, C. F., Zwick, M. S., Czeschin, D. G., Wendel, J. F., Mcknight, T. D., Price, H. J., and Stelly, D. M. (1999). Ty1-*copia*-retrotransposon behavior in a polyploid cotton. *Chromosome Res.* **8**, 73–76.

Harland, S. C. (1936). The genetical conception of the species. *Camb. Philos. Soc. Biol. Rev.* **11**, 83–112.

Harland, S. C. (1939). *In* "The Genetics of Cotton". Jonathan Cape, London.

Harland, S. C. (1940). New allopolyploids in cotton by the use of colchicine. *Tropp. Agric.* **17**, 53–55.

Haufler, C. H. (1987). Electrophoresis is modifying our concepts of evolution in homosporous pteridophytes. *Am. J. Bot.* **74**, 953–966.

Hilu, K. W. (1993). Polyploidy and the evolution of domesticated plants. *Am. J. Bot.* **80**, 1494–1499.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. B* **256**, 119–124.

Hughes, A. L., A, G. J., Garbayo, J. M., and Roberts, R. M. (2000). Adaptive diversification within a large family of recently duplicated, placentally expressed genes. *Proc. Natl. Acad. Sci. USA* **97**, 3319–3323.

Hutchinson, J. B. (1951). Intra-specific differentiation in *Gossypium hirsutum*. *Heredity* **5**, 161–193.

Hutchinson, J. B. (1954). New evidence on the origin of the Old World cottons. *Heredity* **8**, 225–241.

Hutchinson, J. B. (1959). *In* "The Application of Genetics to Cotton Improvement". Cambridge University Press, Cambridge.

Hutchinson, J. B., Stephens, S. G., and Dodds, K. S. (1945). The seed hairs of *Gossypium*. *Ann. Bot.* **9**, 361–368.

Hutchinson, J. B., Silow, R. A., and Stephens, S. G. (1947). *In* "The Evolution of *Gossypium* and the Differentiation of the Cultivated Cottons". Oxford University Press, London.

Ji, Y., De Donato, M., Crane, C. F., Raska, W. A., Nurul Islam-Faridi, M. N., McKnight, T. D., Price, H. J., and Stelly, D. M. (1999). New ribosomal RNA gene locations in *Gossypium hirsutum* mapped by meiotic FISH. *Chromosoma* **108**, 200–207.

Jiang, C., Wright, R., El-Zik, K., and Paterson, A. H. (1998). Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc. Natl. Acad. Sci. USA* **95**, 4419–4424.

Jiang, C., Wright, R. J., Woo, S. S., DelMonte, T. A., and Paterson, A. H. (2000). QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). *Theor. Appl. Genet.* **100**, 409–418.

Jiang, C. -X., Chee, P. W., Draye, X., Morrell, P. L., Smith, C. W., and Paterson, A. H. (2000). Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution* **54**, 798–814.

Johnson, B. L. (1975). *Gossypium* palmeri and a polyphyletic origin of the New World cottons. *Bull. Torrey Bot. Club* **102**, 340–349.

Kadir, Z. B. A. (1976). DNA evolution in the genus *Gossypium*. *Chromosoma* **56**, 85–94.

Kellogg, E. A. (1998). Relationships of cereal crops and other grasses. *Proc. Natl. Acad. Sci. USA* **95**, 2005–2010.

Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* **17**, 1483–1498.

LaDuke, J. C., and Doebley, J. F. (1995). A chloroplast DNA based phylogeny of the Malvaceae. *Syst. Bot.* **20**, 259–271.

Lagercrantz, U. (1998). Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent recombinations. *Genetics* **150**, 1217–1228.

Lagercrantz, U., and Lydiate, D. J. (1996). Comparative genome mapping in *Brassica*. *Genetics* **144**, 1903–1910.

Lawrence, W. J. C. (1931). The secondary association of chromosomes. *Cytologia* **2**, 352–384.

Lee, H. -S., and Chen, Z. J. (2001). Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proc. Natl. Acad. Sci. USA* **98**, 6753–6758.

Leitch, I. J., and Bennett, M. D. (1997). Polyploidy in angiosperms. *Trends Plant Sci.* **2**, 470–476.

Levin, D. A. (1983). Polyploidy and novelty in flowering plants. *Am. Nat.* **122**, 1–25.

W. H. Lewis (ed.) (1980). *In* "Polyploidy: Biological Relevance". Plenum Press, New York.

Li, W. -H. (1985). Accelerated evolution following gene duplication and its implications for the neutralist–selectionist controversy. *In* "Population Genetics and Molecular Evolution" (T. Ohta and K. Aoki, Eds.), pp. 333–353. Springer, Berlin.

Liu, B., and Wendel, J. F. (2000). Retroelement activation followed by rapid repression in interspecific hybrid plants. *Genome* **43**, 874–880.

Liu, B., Vega, J. M., and Feldman, M. (1998). Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. II. Changes in low-copy coding DNA sequences. *Genome* **41**, 535–542.

Liu, B., Vega, J. M., Segal, G., Abbo, S., Rodova, M., and Feldman, M. (1998b). Rapid genomic changes in newly synthesized amphiploids of *Triticum* and *Aegilops*. I. Changes in low-copy non-coding DNA sequences. *Genome* **41**, 272–277.

Liu, B., Brubaker, C. L., Mergeai, G., Cronn, R. C., and Wendel, J. F. (2001a). Polyploid formation in cotton is not accompanied by rapid genomic changes. *Genome* **43**, 874–880.

Liu, Q., Brubaker, C. L., Green, A. G., Marshall, D. R., Sharp, P. J., and Singh, S. P. (2001b). Evolution of the *FAD2-1* fatty acid desaturase 5′UTR intron and the molecular systematics of *Gossypium* (Malvaceae). *Am. J. Bot.* **88**, 92–102.

Longley, A. E. (1933). Chromosomes in *Gossypium* and related genera. *J. Agric. Res.* **46**, 217–227.

Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.

Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473.

Makalowski, W. (2001). Are we polyploids? A brief history of one hypothesis. *Genome Res.* **11**, 667–670.

Masterson, J. (1994). Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* **264**, 421–424.

McCarty, J. C., and Percy, R. G. (2001). Genes from exotic germplasm and their use in cultivar improvement in *Gossypium hirsutum* L. and *G. barbadense* L. *In* "Genetic Improvement of Cotton: Emerging Technologies" (J. N. Jenkins and S. Saha, Eds.), pp. 65–80. Science Publishers, Enfield, NH.

Michaelson, M. J., Price, H. J., Ellison, J. R., and Johnston, J. S. (1991). Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. *Am. J. Bot.* **78**, 183–188.

Moore, G., Devos, K. M., and Wang, Z. (1995). Grasses, line up and form a circle. *Curr. Biol.* **5**, 737–739.

Morton, B. R., Gaut, B. S., and Clegg, M. T. (1996). Evolution of alcohol dehydrogenase genes in the palm and grass families. *Proc. Natl. Acad. Sci. USA* **93**, 11735–11739.

Muller, J. (1981). Fossil pollen records of extant angiosperms. *Bot. Rev.* **47**, 1–142.

Muller, J. (1984). Significance of fossil pollen for angiosperm history. *Ann. Missouri Bot. Gard.* **71**, 419–443.

Muravenko, O., Fedotov, A. R., Punina, E. O., Federova, L. I., Grif, V. G., and Zelenin, A. V. (1998). Comparison of chromosome BrdU–Hoechst–Giemsa banding patterns of the A_1 and $(AD)_2$ genomes of cotton. *Genome* **41**, 616–625.

Ohno, S. (1970). *In* "Evolution by Gene Duplication". Springer, New York.

O'Neill, R. J. W., O'Neill, M. J., and Graves, J. A. M. (1998). Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* **393**, 68–72.

Otto, S. P., and Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437.

Ozkan, H., Levy, A. A., and Feldman, M. (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops–Triticum*) group. *Plant Cell* **13**, 1735–1747.

Parks, C. R., Ezell, W. L., Williams, D. E., and Dreyer, D. L. (1975). The application of flavonoid distribution to taxonomic problems in the genus *Gossypium*. *Bull. Torrey Bot. Club* **102**, 350–361.

Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elsik, C. G., Jiang, C. -X., Katsar, C. S., Lan, Y. -R., Ming, R., and Wright, R. J. (2000). Comparative genomics of plant chromosomes. *Plant Cell* **12**, 1523–1539.

Percy, R. G., and Wendel, J. F. (1990). Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theor. Appl. Genet.* **79**, 529–542.

Phillips, L. L. (1963). The cytogenetics of *Gossypium* and the origin of New World cottons. *Evolution* **17**, 460–469.

Phillips, L. L. (1964). Segregation in new allopolyploids of *Gossypium*. V. Multivalent formation in New World × Asiatic and New World × wild American hexaploids. *Am. J. Bot.* **51**, 324–329.

Phillips, L. L., and Strickland, M. A. (1966). The cytology of a hybrid between *Gossypium hirsutum* and *G. longicalyx*. *Can. J. Genet. Cytol.* **8**, 91–95.

Reinisch, A. J., Dong, J., Brubaker, C. L., Stelly, D. M., Wendel, J. F., and Paterson, A. H. (1994). A detailed RFLP map of cotton, *Gossypium hirsutum* × *G. barbadense*: Chromosome organization and evolution in a disomic polyploid genome. *Genetics* **138**, 829–847.

Saha, S., and Zipf, A. (1998). Genetic diversity and phylogenetic relationships in cotton based on isozyme markers. *J. Crop Prod.* **1**, 79–93.

Sanderson, M. J. (1998). Estimating rate and time in molecular phylogenies: Beyond the molecular clock? *In* "Molecular Systematics of Plants. II. DNA Sequencing" (P. S. Soltis, D. E. Soltis and J. J. Doyle, Eds.), pp. 242–264. Kluwer Academic, Dordrecht.

Saunders, J. H. (1961). *In* "The wild species of *Gossypium* and their evolutionary history". Oxford University Press, London.

Schranz, M. E., and Osborn, T. C. (2000). Novel flowering time variation in the resynthesized polyploid *Brassica napus*. *J. Hered.* **91**, 242–246.

Seelanan, T., Schnabel, A., and Wendel, J. F. (1997). Congruence and consensus in the cotton tribe. *Syst. Bot.* **22**, 259–290.

Seelanan, T., Brubaker, C. L., Stewart, J. M., Craven, L. A., and Wendel, J. F. (1999). Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Syst. Bot.* **24**, 183–208.

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **13**, 1749–1759.

Shoemaker, R. C., Polzin, K., Labate, J., Specht, J., Brummer, E. C., Olson, T., Young, N., Concibido, V., Wilcox, J., Tamulonis, J. P., Kochert, G., and Boerma, H. R. (1996). Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* **144**, 329–338.

Sidow, A. (1996). Gen(ome) duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* **6**, 715–722.

Skovsted, A. (1933). Cytological studies in cotton. I. The mitosis and meiosis in diploid and triploid Asiatic cotton. *Ann. Bot.* **47**, 227–251.

Skovsted, A. (1934). Cytological studies in cotton. II. Two interspecific hybrids between Asiatic and New World cottons. *J. Genet.* **28**, 407–424.

Skovsted, A. (1937). Cytological studies in cotton. IV. Chromosome conjugation in interspecific hybrids. *J. Genet.* **34**, 5–42.

Small, R. L., and Wendel, J. F. (1999). The mitochondrial genome of allotetraploid cotton (*Gossypium* L.). *J. Hered.* **90**, 251–253.

Small, R. L., and Wendel, J. F. (2000). Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). *Genetics* **155**, 1913–1926.

Small, R. L., and Wendel, J. F. (2000). Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium* L. Malvaceae). *Mol. Phyl. Evol.* **16**, 73–84.

Small, R. L., and Wendel, J. F. (2002). Differential evolutionary dynamics of duplicated paralogous *Adh* loci in allotetraploid cotton (*Gossypium*). *Mol. Biol. Evol.* **19**, 597–607.

Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., and Wendel, J. F. (1998). The tortoise and the hare: Choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* **85**, 1301–1315.

Small, R. L., Ryburn, J. A., and Wendel, J. F. (1999). Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Mol. Biol. Evol.* **16**, 491–501.

Smith, N. G. C., Knight, R., and Hurst, L. D. (1999). Vertebrate genome evolution: A slow shuffle or a big bang? *BioEssays* **21**, 697–703.

Soltis, D. E., and Soltis, P. S. (1989). Polyploidy, breeding systems and genetic differentiation in homosporous pteridophytes. *In* "Isozymes in Plant Biology" (D. E. Soltis and P. S. Soltis, Eds.), pp. 241–258. Dioscorides, Portland, OR.

Soltis, D. E., and Soltis, P. S. (1993). Molecular data and the dynamic nature of polyploidy. *Crit. Rev. Plant Sci.* **12**, 243–273.

Soltis, D. E., and Soltis, P. S. (1995). The dynamic nature of polyploid genomes. *Proc. Natl. Acad. Sci. USA* **92**, 8089–8091.

Soltis, D. E., and Soltis, P. S. (1999). Polyploidy: Origins of species and genome evolution. *Trends Ecol. Evol.* **9**, 348–352.

Soltis, P. S., and Soltis, D. E. (2000). The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl Acad. Sci. USA* **97**, 7051–7057.

Soltis, P. S., Doyle, J. J., and Soltis, D. E. (1992). Molecular data and polyploid evolution in plants. *In* "Molecular Systematics of Plants" (P. S. Soltis, D. E. Soltis and J. J. Doyle, Eds.), pp. 177–201. Chapman & Hall, New York.

Song, K., Lu, P., Tang, K., and Osborn, T. C. (1995). Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. *Proc. Natl Acad. Sci. USA* **92**, 7719–7723.

Sossey-Alaouni, K., Serieys, H., Tersac, M., Lambert, P., Schilling, E., Griveau, Y., Kaan, F., and Bervillé, A. (1998). Evidence for several genomes in *Helianthus. Theor. Appl. Genet.* **97**, 422–430.

Spring, J. (1997). Vertebrate evolution by interspecific hybridization—are we polyploid? *FEBS Lett.* **400**, 2–8.

Stebbins, G. L. (1947). Evidence on rates of evolution from the distribution of existing and fossil plant species. *Ecol. Monogr.* **17**, 149–158.

Stebbins, G. L. (1950). *In* "Variation and Evolution in Plants". Columbia University Press, New York.

Stebbins, G. L. (1971). *In* "Chromosomal Evolution in Higher Plants". Edward Arnold, London.

Stephens, S. G. (1944). The genetic organization of leaf-shape development in the genus *Gossypium. J. Genet.* **46**, 28–51.

Stephens, S. G. (1944). Phenogenetic evidence for the amphidiploid origin of New World cottons. *Nature* **153**, 53–54.

Stephens, S. G. (1951). Evolution of the gene—"Homologous" genetic loci in *Gossypium. Cold Spring Harbor Sympp. Quant. Biol.* **16**, 131–140.

Stephens, S. G. (1951). Possible significance of duplication in evolution. *Adv. Genet.* **4**, 247–265.

Stephens, S. G. (1958). Salt water tolerance of seeds of *Gossypium* species as a possible factor in seed dispersal. *Am. Nat.* **92**, 83–92.

Stephens, S. G. (1966). The potential for long range oceanic dispersal of cotton seeds. *Am. Nat.* **100**, 199–210.

Stewart, J. M. (1995). Potential for crop improvement with exotic germplasm and genetic engineering. *In* "Challenging the Future: Proceedings of the World Cotton Research" (G. A. Constable and N. W. Forrester, Eds.), pp. 313–327. CSIRO, Melbourne.

Taylor, J. S., and Brinkmann, J. S. (2001). 2R or not 2R? *Trends Genet.* **17**, 488–489.

Valicek, P. (1983). A contribution to the botanical characteristics of the species *Gossypium gossypioides* (Ulbr.) Standl. *Agric. Tropp. Subtropp. Univ. Agric. Praga* **16**, 185–196.

Vollesen, K. (1987). The native species of *Gossypium* (Malvaceae) in Africa, Arabia and Pakistan. *Kew Bull.* **42**, 337–349.

Watt, G. (1907). *In* "The wild and cultivated cotton plants of the world". Longmans, Green and Co, London.

Webber, J. M. (1935). Interspecific hybridization in *Gossypium* and the meiotic behavior of F_1 plants. *J. Agric. Res.* **51**, 1047–1070.

Wendel, J. F. (1989). New World tetraploid cottons contain Old World cytoplasm. *Proc. Natl Acad. Sci. USA* **86**, 4132–4136.

Wendel, J. F. (1995). Cotton. *In* "Evolution of Crop Plants" (N. Simmonds and J. Smartt, Eds.), pp. 358–366. Longman, London.

Wendel, J. F. (2000). Genome evolution in polyploids. *Plant Mol. Biol.* **42**, 225–249.

Wendel, J. F., and Albert, V. A. (1992). Phylogenetics of the cotton genus (*Gossypium* L.): Character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. *Syst. Bot.* **17**, 115–143.

Wendel, J. F., and Liu, B. (2002). Non-Mendelian phenomena in allopolyploid genome evolution. *Curr. Genom.* (in press).

Wendel, J. F., and Percival, A. E. (1990). Molecular divergence in the Galapagos Island–Baja California species pair *Gossypium klotzschianum* and *G. davidsonii* (Malvaceae). *Plant Syst. Evol.* **171**, 99–115.

Wendel, J. F., and Percy, R. G. (1990). Allozyme diversity and introgression in the Galapagos endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochem. Syst. Ecol.* **18**, 517–528.

Wendel, J. F., Olson, P. D., and Stewart, J. M. (1989). Genetic diversity, introgression and independent domestication of Old World cultivated cottons. *Am. J. Bot.* **76**, 1795–1806.

Wendel, J. F., McD, S. J., and Rettig, J. H. (1991). Molecular evidence for homoploid reticulate evolution in Australian species of *Gossypium*. *Evolution* **45**, 694–711.

Wendel, J. F., Brubaker, C. L., and Percival, A. E. (1992). Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. *Am. J. Bot.* **79**, 1291–1310.

Wendel, J. F., Rowley, R., and Stewart, J. M. (1994). Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst. Evol.* **192**, 49–59.

Wendel, J. F., Schnabel, A., and Seelanan, T. (1995a). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc. Natl Acad. Sci. USA* **92**, 280–284.

Wendel, J. F., Schnabel, A., and Seelanan, T. (1995b). An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Mol. Phyl. Evol.* **4**, 298–313.

Wendel, J. F., Small, R. L., Cronn, R. C., and Brubaker, C. L. (1999). Genes, jeans, and genomes: Reconstructing the history of cotton. *Proceedings of the VII International Symposium of the International Organization of Plant Biosystematists*. *In* "Plant Evolution in Man-made Habitats" (L. W. D. van Raamsdonk and J. C. M. den Nijs, Eds.), pp. 133–161. Hugo de Vries Laboratory, Amsterdam.

Wendel, J. F., Cronn, R. C., Johnston, J. S., and Price, H. J. (2002). Feast and famine in plant genomes. *Genetica* **115**, 37–47.

Whitkus, R., Doebley, J., and Lee, M. (1992). Comparative genome mapping of sorghum and maize. *Genetics* **132**, 1119–1130.

Wilson, W. A., Harrington, S. E., Woodman, W. L., Lee, M., Sorrells, M. E., and McCough, S. R. (1999). Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated Panicoids. *Genetics* **153**, 453–473.

Wolfe, K. H. (2001). Yesterday's polyploids and the mystery of polyploidization. *Nat. Rev. Genet.* **2**, 333–341.

Wolfe, K. H., and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713.

Woodcock, D. W., and Manchester, C. A. (1998). Fossil cotton from the Salt Lake Crater area, O'ahu, Hawaii. Records of the Hawaii Biological Survey for 1997—Part 2. Notes. *In* "Bishop Museum Occasional Papers" (N. L. Evenhuis and S. E. Miller, Eds.), pp. 17–19. Bishop Museum Press, Honolulu.

Wright, R. J., Thaxton, P. M., El-Zik, K. M., and Paterson, A. H. (1998). D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. *Genetics* **149**, 1987–1996.

Zhao, X. -P., Wing, R. A., and Paterson, A. H. (1995). Cloning and characterization of the majority of repetitive DNA in cotton (*Gossypium* L.). *Genome* **38**, 1177–1188.

Zhao, X. -P., Si, Y., Hanson, R. E., Crane, C. F., Price, H. J., Stelly, D. M., Wendel, J. F., and Paterson, A. H. (1998). Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res.* **8**, 479–492.