Molecular Systematics of Plants

Pamela S. Soltis, Douglas E. Soltis, and Jeff J. Doyle, Editors

Chapman and Hall

New York

London
Molecular Systematics and Crop Evolution

John Doebley

During the past decade, the tools of molecular biology have been applied to plant systematics with remarkable success. New insights have been gained into such topics as phylogenetic reconstruction, introgression, genomic evolution, and levels of genetic variation in natural populations. Molecular methods have provided greater resolution than was previously possible with other approaches. Moreover, variation in DNA sequences is more readily subjected to statistical analysis than many previous types of data, and it can be less ambiguous, making interpretation of data more straightforward.

In this chapter, I review the application of molecular systematics to questions surrounding the origins and evolution of crop plants, specifically (1) identifying the progenitors of crops, (2) assaying levels of genetic variation in crops and their ancestors, and (3) determining the extent of introgression between crops and their relatives. Each of these issues has been addressed for several crops using molecular methods, and molecular evidence has provided definitive results in many cases. I will also outline issues in crop evolution that, although amenable to study with molecular methods, have received little or no study to date. These issues include (1) the genetic basis of morphologic changes induced by human selection and (2) the nature of molecular evolution under domestication.

Methodologic Considerations

Angiosperms possess three separate genomes: chloroplast DNA (cpDNA), mitochondrial DNA (mtDNA), and nuclear DNA (nDNA). Molecular analyses of each of these are applicable to the study of crop evolution, although their utility may vary depending on the question of interest. For studies of crop

Preparation of this chapter was supported in part by NSF grant number BSR-8806889. I thank Drs. K. Hosaka, K. Rasmussen, D. Spooner, and J. Wendel, and the editors for their many helpful suggestions which have greatly improved this chapter.
evolution, the most practical method of analysis of these genomes is restriction fragment length polymorphism (RFLP) analysis because it enables greater sampling of species and populations than the more labor-intensive method of nucleotide sequencing.

Chloroplast DNA is maternally inherited in most angiosperms, making it particularly suitable for identifying maternal parental species of polyploid crops. It also is typified by conservative rates of both structural rearrangements and sequence evolution (Palmer, 1987). This enables the identification of specific mutational events with great certainty, increasing confidence that restriction fragments, which appear to be the same, are in fact the same. Such unambiguous data are readily treated by statistical or phylogenetic analyses. The conservatism of cpDNA also results in generally low levels of intraspecific variation, often reducing its utility at this taxonomic level (but see Chapter 6, this volume).

The mitochondrial genome, like cpDNA, is usually maternally inherited in angiosperms. However, this genome is typified by a much higher rate of structural rearrangement (Palmer and Herbon, 1988). This feature generates higher levels of intraspecific RFLPs, but lowers confidence that two restriction fragments with the same electrophoretic mobility are indeed the same. Also, single structural changes may alter several restriction fragments, making accurate quantification of mtDNA data more difficult. Indeed, in the absence of detailed restriction mapping, it is practically impossible to sort out restriction site mutations and structural rearrangements in mtDNA.

The complexity of nDNA offers opportunity to study repetitive sequences, such as ribosomal genes (rDNA) (Doyle and Beachy, 1985; Zimmer et al., 1988), and single-copy or low-copy-number sequences or genes (Song et al., 1988a, b; Keim et al., 1989; Havey and Muehlbauer, 1989). Nuclear DNA sequences provide evidence for both the maternal and paternal lineages, and levels of polymorphism for nDNA are suitable for intraspecific analyses. Evolutionary rates for the nuclear genome vary depending upon the specific portion of the genome analyzed; however, in general, nDNA has a high rate of both sequence and structural evolution as compared to cpDNA. For nuclear low-copy-number sequences, detailed restriction maps provide the most precise data (Gepts and Clegg, 1989), but these can be assembled only through tedious laboratory analyses that, if performed, limit the number of accessions and genes that can be analyzed. Scoring presence/absence of specific low-copy-number restriction fragments is more practical; however, as with mtDNA, one can have only limited confidence that two restriction fragments with similar mobilities are indeed the same. Similarly, a single structural change in nDNA can produce several restriction fragment changes, creating uncertainties as how best to quantify the data.

Many questions surrounding crop evolution require quantification of molecular data, especially to obtain an accurate measure of genetic similarity between taxa or an estimate of the levels of variation. For cpDNA, crops and their relatives can be readily analyzed for the presence/absence of particular restriction sites and
insertion/deletion events. These events can then be used to construct phylogenetic
trees using parsimony (Palmer et al., 1985; Doebley et al., 1987; Ogihara and
Tsunewaki, 1988). For nuclear and mitochondrial DNA, the higher levels of
polymorphism create a more complex situation, especially if the molecular basis
of the observed differences in restriction fragments is not understood. The best
approach to mtDNA restriction fragment data may be to calculate distances based
on the shared fragment method (McClean and Hanson, 1986; Graur et al., 1989).
Because mtDNA RFLPs are apt to mix both restriction site changes and structural
rearrangements, distances based upon them will not provide an accurate estimate
of the proportion of substitutions per site, but should still provide a general
measure of the degree of phenetic similarity. For rDNA, detailed restriction
mapping has proven a practical approach (Zimmer et al., 1988). In the absence
of detailed restriction mapping, low-copy-number nuclear sequence data are
probably most accurately quantified when RFLPs are identified as alleles at a
particular locus (Havey and Muehlbauer, 1989; Keim et al., 1989). This proce-
dure increases confidence that structural rearrangements that alter several restric-
tion fragments will not be weighted too heavily, and the data can be subjected to
multivariate, phylogenetic, and statistical analyses, including measures of genetic
distance and heterozygosity (Havey and Muehlbauer, 1989).

Origins of Crop Species

A principal concern of crop evolutionists has been the identification of the wild
progenitors of domesticated species. This has often been difficult with classical
taxonomic methods, because crops usually exhibit wide morphologic departures
from their relatives. Genetic, cytogenetic, and biochemical analyses have contrib-
uted greatly to this subject (Simmonds, 1976; Doebley, 1989); however, much
remains to be learned. The question of crop origins can be pursued at several
different levels. First, some crops may have a clearly identified wild progenitor
species, although the progenitor may be polymorphic and geographically wide-
spread. Here, the question of interest may be to pinpoint the geographic region
in which domestication occurred. Second, for some crops, there may be uncer-
tainty as to which of several distinct wild taxa gave rise to the cultigen. Third,
for polyploid crops, there is interest in identifying their diploid parental species.
Finally, the question of whether a crop has been domesticated once or several
times is often an issue.

Because domestication is a relatively recent event (the last 10,000 years),
one expects crops to show relatively little genetic differentiation from their
progenitors. This is true despite the often large differences in gross morphology
between crop and progenitor because artificial (human) selection is likely to have
acted primarily on the small suite of genes controlling morphologic traits of
interest to humankind, leaving the vast majority of the genome to evolve at a
much slower pace. Thus, for the conservatively evolving chloroplast genome,
few (one or two) or no restriction site differences should be observed between
crop and progenitor for studies using 20 to 25 restriction enzymes. Moreover,
crop and progenitor should appear on the same (or very close) branchlet(s) of the
cpDNA phylogenetic tree. For mtDNA and nDNA, which evolve more rapidly,
greater differentiation is expected, although the progenitor should show greater
similarity to the crop than to other congeneric wild species.

**Barley**

Clegg et al. (1984a) and Neale et al. (1988) examined cpDNA variation in
barley (*Hordeum vulgare* ssp. *vulgare*) and its presumed progenitor (*H. vulgare*
ssp. *spontaneum*). Each of these studies revealed a small number (3 to 5) of
variable restriction sites that defined several cpDNA genotypes. These studies
revealed that the most common cpDNA genotypes in cultivated barley were also
present in wild barley, reflecting the close genetic relationship between these two
subspecies. Further, this work showed greater cpDNA diversity in wild barley as
compared to cultivated barley, which one would expect if the former were the
progenitor species and the latter were the derivative.

**Cotton**

One of the most definitive molecular analyses of the origin of a crop has been
that of cotton (Wendel, 1989). There are four separate species of cultivated cotton
including two New World tetraploids, *Gossypium hirsutum* and *G. barbadense*.
Cytogenetic analysis has shown that these tetraploids combine the A genome (of
Asia-Africa) and the D genome (of the Americas). An unresolved question
surrounding the origin of the New World cottons is when the A and D genome
diploids hybridized to form the ancestor of the cultivated tetraploids. Various
authors have proposed that hybridization took place (1) before Africa and South
America were fully separated by continental drift (100 million or more years ago),
(2) very recently when humans carried A genome species from Africa to the New
World (during the past 15,000 years), or (3) within the past several million years
after natural dispersal of an A genome species to the New World. The weight of the
biosystematic evidence favors the third explanation; however, this topic is still
controversial. Wendel (1989) showed that the degree of cpDNA divergence be-
tween the New World cottons and the A genome diploids indicates that the former
arose 1–2 million years ago, clearly eliminating the first alternative and providing
strong support for the third. Wendel (1989) also demonstrated that an Asian-African
(A genome) species was the maternal parent of both New World tetraploids.

**The “Irish” Potato**

*Solanum* is a large, pantropical genus with between 1,400 and 2,000 species.
Within it, the tuber-bearing species, including the “Irish” potato, belong to the
subgenus *Potatoe* section *Petota*. Within this group, approximately six wild diploid species (x = 12) are taxonomically similar enough to domesticated potatoes to have potentially contributed to their origin (Hawkes et al., 1979; Hawkes, 1989). The cultivated potatoes include diploid, triploid, tetraploid, and pentaploid forms, which have been variously treated taxonomically. *Solanum goniocalyx*, *S. phureja*, and *S. stenotomum* are the primary cultivated diploids, and *S. tuberosum*, the potato of commerce, is a tetraploid. *Solanum tuberosum* includes two principal cultivated forms, ssp. *andigena* and ssp. *tuberosum*. The former subspecies arose in South America, and the latter is thought to have arisen twice by parallel evolution, once in southern Chile and once in Europe. Two questions surround the origin of cultivated potatoes: (1) which diploid(s) gave rise to the cultivated potatoes (both diploid and tetraploid) and (2) what is the relationship between ssp. *tuberosum* and ssp. *andigena*?

Hosaka (1986) and Hosaka and Hanneman (1988b) examined cpDNA variation in the cultivated potatoes and their relatives. These authors defined eight cpDNA types based on restriction site and insertion/deletion mutations. They also surveyed a large number of wild and cultivated accessions for these cpDNA variants (Hosaka and Hanneman, 1988b). Their work provides many valuable insights into the origin of the cultivated potatoes. First, ssp. *andigena* possesses several cpDNA types that are also found in the cultivated diploids, strongly supporting the hypothesis that ssp. *andigena* was derived from the cultivated diploids. Second, the diversity of cpDNA types in ssp. *andigena*, along with other biosystematic data, suggests that ssp. *andigena* had multiple origins via repeated cycles of hybridization and introgression. Third, the occasional presence of wild-type cpDNAs (*S. sparsipilum* or *S. chacoense*) in ssp. *andigena* suggests occasional introgression from the wild forms into the domesticates. Fourth, the cpDNA data suggest a very close relationship among the three species of diploid cultivated potatoes (*S. goniocalyx*, *S. phureja*, and *S. stenotomum*).

Hosaka and Hanneman (1988a) also used cpDNA variation to assess the origin of ssp. *tuberosum*. They reported that ssp. *tuberosum* has a specific cpDNA type (T) not found in any diploid taxa, although it is found at low frequency in ssp. *andigena*. From the Chilean-Argentine border. When combined with other data, cpDNA data provide strong support for the hypothesis that Chilean ssp. *tuberosum* was derived from Andean ssp. *andigena*. Chloroplast DNA data also indicate that the first potatoes introduced into Europe belonged to ssp. *andigena*. However, these apparently did not, for the most part, survive the potato blights of the 1840s and were replaced by the American cultivar "Rough Purple Chili," which belongs to ssp. *tuberosum*. Thus, molecular data argue against the hypothesis that ssp. *tuberosum* arose independently under domestication in Europe, but rather suggest that the modern European potato was derived from Chilean ssp. *tuberosum*.

**Lentil**

the RFLPs as alleles at loci (probes), these authors calculated genetic distances among the species and subspecies. A UPGMA dendrogram based on these distances placed *Lens culinaris* ssp. *orientalis* closest to the cultivated lentil, in agreement with previous isozymic and morphologic data that indicated that this wild subspecies is the ancestor of the cultivated lentil. The RFLP data also demonstrated that the cultivated lentil possesses less genetic diversity than ssp. *orientalis*, consistent with the hypothesis that the former was a recent (during the past 10,000 years) derivative of the latter.

Maize

Timothy et al. (1979), Doebley et al. (1987), and Doebley (1990) examined cpDNA variation among *Zea mays* ssp. *mays* (maize) and its wild relatives (the teosintes). The teosinte species, *Z. diploperennis*, *Z. perennis*, and *Z. luxurians*, possessed cpDNAs that differed substantially (19 mutations) from that of maize. However, the four cpDNA genotypes found in maize were also present in the teosintes, *Z. mays* ssp. *mexicana* and ssp. *parviglumis*. These four cpDNA types are very similar to one another, each distinguished by a single restriction site variant or length mutation. These cpDNA results are consistent with previous cytologic and allozymic data that suggest one of these two wild subspecies was the progenitor of maize. The fact that four cpDNA types were found in both teosinte and maize implies either multiple domestications or introgression of wild cytoplasms into cultivated maize. A preponderance of other data suggests that the latter is the case (Doebley, 1990). The distribution of these cpDNA types in maize and teosinte was not sufficiently correlated with geography to enable the authors to discern a likely center for the origin of maize based on these data.

Further support that teosinte is ancestral to maize comes from restriction analysis of rDNA (Zimmer et al., 1988) and sequence analysis of heterochromatic DNA (Dennis and Peacock, 1984). Both of these studies showed that maize is essentially identical to Mexican annual teosinte (*Z. mays* ssp. *mexicana* or *parviglumis*), but distinct from the other teosinte species (*Z. diploperennis* and *Z. luxurians*). This result indicates that maize and Mexican annual teosinte share a more recent common ancestor than do Mexican annual teosinte and the other teosinte species.

Mustards and Cole Crops

The genus *Brassica* contains several important crop species, including the diploids, black mustard (*B. nigra*), turnip (*B. rapa*), and cabbage (*B. oleracea*), and all three possible allopolyploid hybrids between them, Abyssinian mustard (*B. carinata*), leaf mustard (*B. juncea*), and rape (*B. napus*). The relationships among these species have been established based on cytogenetic data (Fig. 9.1). Analysis of cpDNA showed that *B. nigra* was the maternal parent of *B. carinata*, whereas *B. rapa* was the maternal parent of both *B. napus* and *B. juncea* (Erickson
et al., 1983; Palmer et al., 1983). Phylogenetic analysis of low-copy-number nDNA supported the relationships outlined in Fig. 9.1, and indicated that the nuclear genomes of the amphidiploids are more similar to their maternal than their paternal parents (Song et al., 1988a). Detailed analysis of low-copy-number nDNA sequences of *B. rapa* indicated that this species consists of two well-defined groups, European and east Asian, which may represent two separate domestications from separate wild populations (Song et al., 1988b).

**Pea**

Palmer et al. (1985) analyzed cpDNA variation in *Pisum sativum* (pea) and its wild relatives, *P. humile, P. elatius,* and *P. fulvum.* Among 13 accessions of cultivated pea, they found two cpDNA types. The more common (12 of the 13 accessions) of these cpDNA types was similar to that of an accession of *P. humile* from northern Israel, where cytologic and morphologic data place the origin of domestication for pea. The second cpDNA type, which was found in a single accession of cultivated pea, differed from the first type by five mutations. As noted by the authors, it represents either a separate domestication event or introgression of a foreign cytoplasm into cultivated pea.

**Soybean**

Relatively few comparative studies of nDNA in crops and their progenitors have been conducted; however, soybean (*Glycine max*) and its presumed progenitor (*G. soja*) have been the subject of several such studies. Doyle and Beachy (1985) examined the 18S–25S ribosomal genes of over 40 diverse accessions of wild and cultivated soybean and failed to find a single variant within or between these species. Similarly, Doyle (1988) found only a single repeat length size (345 bp) for the 5S ribosomal genes of soybean. Both this repeat size plus a smaller variant (334 bp) were found in the 5S ribosomal genes of *G. soja.* Data from both ribosomal genes suggest a very close relationship between soybean and its presumed progenitor.

Keim et al. (1989) examined RFLPs of 17 low-copy-number nuclear sequences
in cultivated and wild soybean (*G. soja*). This study uncovered several facts that appear inconsistent with the simple scenario that *G. max* is a recent, cultivated derivative of *G. soja*. First, there is greater genetic diversity within soybean than within *G. soja*. Second, a principal component analysis of the RFLP data revealed little overlap between these two species, suggesting that most soybean accessions have become differentiated from their presumed progenitor for these 17 nuclear sequences. Third, one RFLP, present in 90% of all soybean accessions, was absent from *G. soja*, a pattern that would not be anticipated if the former were recently derived from the latter. These results contrast with typical observations for allozymic variation between a crop and its progenitor for which one normally finds (1) the same common alleles in the domesticate and its ancestor and (2) greater genetic variation in the progenitor (Doebley, 1989). Too little is known about RFLPs of low-copy-number nDNA sequences among species to draw hard conclusions from the results of Keim et al. (1989). It will be of interest to see data of this nature accumulate for other crops and their wild relatives.

**Squash, Pumpkin, and Ornamental Gourds**

Chloroplast DNA variation was analyzed in *Cucurbita*, including five cultivated and ten wild species (Fig. 9.2) (Doebley et al., unpublished data). These analyses revealed a close association (two restriction site differences) between the cultivated hubbard squash (*C. maxima*) and its presumed progenitor, *C.
andreana. They also revealed a close association (one restriction site difference) between the cultigens cushaw (*C. mixta*) and butternut squash (*C. mochata* from Mexico), suggesting that these two cultigens may share the same maternal progenitor species. *Cucurbita sororia* (four accessions), thought to be the progenitor of cushaw, differed from it by four restriction site mutations. This is a somewhat greater degree of divergence than one would expect between a crop and its progenitor, although a form of *C. sororia* more similar to cushaw may be found among additional accessions of this wild species. Another anomalous result was that *C. mochata* from South America differed from Mexican varieties of this species by four restriction sites, somewhat more than anticipated among varieties of a cultigen. This suggests separate origins for this domesticate in Mexico and South America or perhaps introgression. Finally, cpDNA analyses revealed two distinct cpDNA forms in the cultivated species, *C. pepo* (squash, ornamental gourd, and pumpkin). The first of these, which was found in ornamental gourd and crookneck squash, was identical to that of the wild species *C. texana*, agreeing with recent allozymic data that suggest this wild species had a role in the origin of *C. pepo* (Decker and Wilson, 1987; Decker, 1988). The second cpDNA form, which was found in zucchini and Mexican pumpkin, differed from the first by six restriction site mutations. This cpDNA type is unknown in any wild form, including a suspected progenitor, *C. fraterna*. This too agrees with previous allozymic data (Wilson, 1989) and suggests that the wild progenitor of pumpkin has yet to be discovered.

**Sunflower**

Rieseberg and Seiler (1990) examined cpDNA variation in cultivated and wild accessions of *Helianthus annuus* (sunflower) and found four closely related cpDNA genotypes among 11 accessions of wild/weedy sunflower. The cpDNAs of 23 diverse accessions of cultivated sunflower were all the same, and they were identical to one of the four wild-type cpDNAs. The cpDNA type associated with the cultigen is widespread in the wild sunflower (Missouri to California), and thus these data do not help pinpoint the geographic origin of the sunflower. However, when combined with other data, the lack of polymorphism in the cpDNA of the sunflower suggests that this crop was domesticated only once.

**Wheat**

The cultivated wheats are part of a polyploid complex in the genus *Triticum*. There are four primary taxa of cultivated wheat, each of which has its own presumed ancestor: the diploid *T. monococcum* var. *monococcum* (AA genome) derived from the wild-type *T. monococcum* var. *boeoticum* (AA); the tetraploid *T. turgidum* var. *dicoccum* (AABB) derived from the wild-type *T. turgidum* var. *dicoccoides* (AABB); the tetraploid *T. timopheevii* var. *timopheevii* (AAGG)
derived from the wild-type *T. timopheevii* var. *araraticum* (AAGG); and the hexaploid *T. aestivum* (AABBDD) derived from a cross of the cultivated *T. turgidum* (AABB) and the wild *T. tauschii* (DD) (Feldman, 1976). As indicated, each of the cultivated polyploids has a wild ancestral form of the same ploidal level with the exception of *T. aestivum*, which presumably arose by polyploidy under domestication.

The origin of the A, B, and D genomes of hexaploid wheat (*T. aestivum*) has not been fully resolved. Based on cyto genetic analyses, two of the genomes can be clearly associated with specific, wild, diploid taxa: (1) *T. monococcum* var. *boeoticum* for the A genome, and (2) *T. tauschii* for the D genome. The donor of the B genome is uncertain; however, several species of *Triticum* have been favored based on cyto genetic, taxonomic, and biochemical data, including *T. speltoides*, *T. longissimum*, and *T. bicerne*. The donor of the G genome of *T. timopheevii* is also uncertain, although *T. speltoides* has been suggested by some authors (see Feldman, 1976).

Identification of the wild species that contributed the B genome to the wheats has been attempted with cpDNA and mtDNA analyses. This is possible because it was the ancestral B genome species that was the maternal (cytoplasmic) parent of *T. turgidum* and *T. aestivum*. Thus, these species may be said to contain B cytoplasm. Tsunewaki and Ogihara (1983) examined cpDNA of the wheats and concluded that the maternal or B genome parent was *T. longissimum*; however, Bowman et al. (1983) were unable to confirm this result. Subsequent analyses suggested that the *T. longissimum* cytoplasm examined by Tsunewaki and Ogihara (1983) had been mislabeled (Ogihara and Tsunewaki, 1988). Recent comprehensive analyses revealed the cpDNA of *T. turgidum* and *T. aestivum* to be distinct (four restriction site mutations and two insertion differences from *T. speltoides*) within the genus (Ogihara and Tsunewaki, 1988). Analysis of mtDNA of the cultivated wheats and their relatives also demonstrated the mtDNA of the B cytoplasm to be unique within the genus (Graur et al., 1989). This is a curious result because it shows that the cytoplasmic genome of the tetraploid, *T. turgidum* (AABB), and the hexaploid, *T. aestivum* (AABBDD), is unknown in any diploid species. Either there are additional species or populations that have not been discovered and analyzed, or the donor of the B genome and B cytoplasm is extinct. The fact that considerable heterogeneity has been found among the mtDNA of *T. speltoides* suggests that more comprehensive sampling would be informative (Breiman, 1987).

Determination of other progenitor species of the wheats has been much more straightforward. Ogihara and Tsunewaki (1988) analyzed cpDNA of *T. timopheevii* and found it to be identical to that of *Aegilops australi* (a form of *T. speltoides*), its presumed maternal progenitor. Analysis of mtDNA (Graur et al., 1989) and cpDNA (Ogihara and Tsunewaki, 1988) support the hypothesis that *T. turgidum* var. *dicoccoides* was the maternal progenitor of *T. turgidum* var. *turgidum*, which in turn was the maternal progenitor of *T. aestivum*. 
Other Crops

Several other studies have examined organellar DNA in crops and their relatives. Clegg et al. (1984b) and Gepts and Clegg (1989) found no restriction site differences between cpDNAs of pearl millet (*Pennisetum glaucum* ssp. *glaucum*) and its presumed wild progenitor (ssp. *monodii*). Perl-Treves and Galun (1985) reported no restriction site or other differences between cpDNAs of *Cucumis sativus* (cucumber) and its presumed wild progenitor, *C. hardwickii*. These same authors also examined cpDNA of melon (*Cucumis melo* var. *melo*) and its presumed progenitor (C. *melo* var. *agrestis*) and found only a single, small deletion to differentiate them. Duvall and Doebley (1990) found little or no differentiation (0 to 2 mutations) between the cpDNAs of accessions of sorghum (*Sorghum bicolor* ssp. *bicolor*) and its presumed progenitor (*S. bicolor* ssp. *arundinaceum*). McClean and Hanson (1986) studied mtDNA of tomato (*Lycopersicon esculentum* var. *esculentum*) and its wild relatives. Based on the shared restriction fragment method, they calculated divergence values among the taxa. These revealed *L. esculentum* var. *cerasiforme* to be most similar to tomato, supporting the hypothesis that it is the progenitor of tomato (Rick and Fobes, 1975).

The origins of several polyploid crops have been subjected to molecular analyses. First, tobacco (*Nicotiana tabacum*) is a tetraploid known only in cultivation. Cytotaxonomic evidence suggests that *N. sylvestris* and *N. tomentosiformis* are its parental diploids (Gerstel, 1976). Analysis of both cpDNA (Kung et al., 1982) and mtDNA (Bland et al., 1985) demonstrated that tobacco inherited these cytoplasmic genomes from *N. sylvestris*. Second, finger millet (*Eleusine coracana*) is a tetraploid cereal crop of Africa. Analysis of cpDNA restriction patterns indicated that finger millet inherited its chloroplast genome from *E. indica* (Hilu, 1988). Third, the evolution of citrus fruits is known to have involved frequent hybridizations. Analysis of cpDNA revealed that sour orange (*Citrus aurantium*), grapefruit (*C. x paradisi*), lemon (*C. limon*), and sweet orange (*C. sinensis*) all share the same cpDNA genotype (Green et al., 1986). This cpDNA type was the same as that of *C. grandis*, the presumed maternal progenitor of these cultivated fruit species. Finally, coffee (*Coffea arabica*) is a tetraploid tree species of Africa. Molecular analyses revealed no differences between the mtDNA and cpDNA of coffee and *C. eugenioides*, a diploid previously thought to be one of the ancestors of coffee (Berthou et al., 1983).

In summary, it is clear that molecular analyses have been successfully applied to many questions surrounding the origins of crop plants. Our knowledge of the ancestry of cotton, the cole crops, sunflower, and pea have all been enhanced substantially. In other cases, such as the origin of the B genome of hexaploid wheat and the origin of pumpkin, molecular analyses have not identified any likely progenitor species. In these cases, it appears that all relevant wild species have yet to be discovered, or perhaps that they
exist in germplasm collections but have yet to be analyzed. This situation calls attention to the need for more field work, taxonomic studies, and evaluation of germplasm collections.

**Genetic Variation in Crop Plants and their Wild Ancestors**

Although there is little hard evidence concerning the beginnings of plant domestication, it seems a reasonable assumption that the first farmers experimented with only a small fraction of the variation present within the progenitor species of today's crops. Further, as the domestication process proceeded, these farmers probably selected only the best phenotypes when setting seed aside for sowing the following year. For these reasons, one expects a significant loss of genetic variation over the six to ten millennia since most crops were first domesticated. Competing forces such as introgression from wild relatives and selection against loss of fitness due to inbreeding have probably counterbalanced the expected loss of genetic variation.

Many authors investigating molecular variation in crops and their progenitors have addressed the above issue. In general, these authors report lower levels of genetic variation in crops than in their progenitors. Studies of cpDNA in barley (Clegg et al., 1984a; Neale et al., 1988) and sunflower (Rieseberg and Seiler, 1990) have found greater variation in the progenitors of these crops than in the crops themselves. Analyses of ribosomal genes in soybean (Doyle, 1988), barley (Allard, 1988), and pearl millet (Gepts and Clegg, 1989) also suggest that greater variation is found in the progenitor than in its associated crop. Examination of low-copy-number nDNA sequences in lentil shows considerably less total genetic diversity ($H_s$) in lentil (0.18) than in its progenitor (0.27) (Hovey and Muehlbauer, 1989).

A few studies reached different conclusions concerning this issue. First, equivalent levels of polymorphism for mtDNA and cpDNA were reported in wild and cultivated barley (Holwerda et al., 1986), in contrast with the two analyses of barley cpDNA cited above. Second, Gepts and Clegg (1989) reported equivalent levels of restriction fragment variation at the alcohol dehydrogenase loci in wild and cultivated pearl millet in contrast with the results of these same authors for rDNA in pearl millet. Third, Keim et al. (1989) showed greater diversity for low-copy-number nDNA sequences in soybean (diversity = 0.37) than in its presumed progenitor, *G. soja* (0.22). They obtained this result despite equivalent sample sizes (ten accessions of *G. max* plant introductions; eight accessions of *G. soja*). This result agrees with the results of Shoemaker et al. (1986) who found three cpDNA types in soybean but only one in *G. soja*.

Table 9.1 summarizes the molecular evidence for loss of genetic variation in the cpDNAs of crops. This table includes only those studies for which the mutational basis (restriction site change versus structural change) of the observed
Table 9.1. A comparison of cpDNA variation in crops and their progenitors.*

<table>
<thead>
<tr>
<th>Crop Progenitor</th>
<th>D</th>
<th>TS</th>
<th>P</th>
<th>N</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Hordeum vulgare</em> ssp. vulgare</td>
<td>0.08</td>
<td>5</td>
<td>0.20</td>
<td>9</td>
<td>Clegg et al. (1984a)</td>
</tr>
<tr>
<td><em>H. vulgare</em> ssp. spontaneum</td>
<td>0.43</td>
<td>5</td>
<td>1.00</td>
<td>11</td>
<td>Clegg et al. (1984a)</td>
</tr>
<tr>
<td><em>Hordeum vulgare</em> ssp. vulgare</td>
<td>0.03</td>
<td>3</td>
<td>0.33</td>
<td>51</td>
<td>Neale et al. (1988)</td>
</tr>
<tr>
<td><em>H. vulgare</em> ssp. spontaneum</td>
<td>0.45</td>
<td>3</td>
<td>1.00</td>
<td>30</td>
<td>Neale et al. (1988)</td>
</tr>
<tr>
<td><em>Pisum sativum</em></td>
<td>0.10</td>
<td>6</td>
<td>0.50</td>
<td>13</td>
<td>Palmer et al. (1985)</td>
</tr>
<tr>
<td><em>P. humilis</em></td>
<td>0.43</td>
<td>6</td>
<td>0.67</td>
<td>4</td>
<td>Palmer et al. (1985)</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em> ssp. bicolor</td>
<td>0.06</td>
<td>8</td>
<td>0.13</td>
<td>3</td>
<td>Duvall and Doebly (1990)</td>
</tr>
<tr>
<td><em>S. bicolor</em> ssp. arundinaceum</td>
<td>0.35</td>
<td>8</td>
<td>1.00</td>
<td>6</td>
<td>Duvall and Doebly (1990)</td>
</tr>
<tr>
<td><em>Helianthus annuus</em> cultivated</td>
<td>0.00</td>
<td>4</td>
<td>0.00</td>
<td>23</td>
<td>Rieseberg and Seiler (1990)</td>
</tr>
<tr>
<td><em>H. annuus</em> wild/weedy</td>
<td>0.25</td>
<td>4</td>
<td>1.00</td>
<td>11</td>
<td>Rieseberg and Seiler (1990)</td>
</tr>
<tr>
<td><em>Zea mays</em> ssp. mays</td>
<td>0.22</td>
<td>3</td>
<td>0.67</td>
<td>80</td>
<td>Doebly (1990)</td>
</tr>
<tr>
<td><em>Z. mays</em> ssp. <em>mays</em></td>
<td>0.38</td>
<td>3</td>
<td>1.00</td>
<td>31</td>
<td>Doebly (1990)</td>
</tr>
<tr>
<td><em>Glycine max</em></td>
<td>0.20</td>
<td>2</td>
<td>1.00</td>
<td>46</td>
<td>Close et al. (1989)</td>
</tr>
<tr>
<td><em>G. soja</em></td>
<td>0.47</td>
<td>2</td>
<td>1.00</td>
<td>8</td>
<td>Close et al. (1989)</td>
</tr>
</tbody>
</table>

* Diversity (D), total number of polymorphic sites in crop and its progenitor (TS), proportion of the total sites that are polymorphic (P), and number of accessions analyzed (N) for pairs of crops and their presumed progenitors. Diversity is defined as the average probability that two cpDNAs will differ at a polymorphic site.

Differences had been determined. The cpDNA data were used to calculate diversity (D) as the average probability per variable site that two accessions will differ (Clegg et al., 1984a). Also presented is the proportion of all polymorphic sites (i.e., sites polymorphic in either crop or progenitor) that were polymorphic in the crop and in the progenitor (P). The data for seven studies (six crops) show consistently lower levels of cpDNA variation in crops as compared to their progenitors (Table 9.1). Furthermore, these data agree with a previous review of allozymic variation, which suggested that crops possess less genetic variation than their progenitors (Doebly, 1989). Nevertheless, the cpDNA data should be viewed cautiously as sample sizes are generally small, and sampling bias is a possibility. Finally, although these data on cpDNA are of interest, more detailed studies of the levels of genetic variation in the nuclear genome are needed. Studies of the nuclear genome such as that of Havey and Muehlbauer (1989) have the potential of giving a very precise assessment of the effects of domestication on genetic variation within crops.

**Introgression between Crops and their Relatives**

Introggression between crops and their relatives is a critical concern of crop evolutionists, because it is a potential source of new variation during crop evolu-
tion and it may give rise to weedy intermediates between crops and their relatives. Ample opportunity exists for such introgression, because crops and their wild relatives usually grow sympatrically and often lack barriers to hybridization. Thus, one would expect introgression to be a common phenomenon. However, disruptive selection may act to restrict introgression. Crops are adapted to human needs and their wild relatives to survival in nature. Hybrids between the two satisfy the demands of neither nature nor humankind, and thus, will be strongly selected against. Because crops and their progenitors differ for many traits controlled by at least dozens of genes, a large portion of their genomes will be linked to one or more of these genes. Thus, large portions of their genomes may be hindered from introgressing freely because of such linkage.

Molecular and biochemical techniques offer a powerful means of detecting introgression and distinguishing it from other phenomena, such as joint retention of the ancestral condition, clinal variation, and convergence. Nevertheless, I am unaware of any studies explicitly designed to determine the extent of introgression between crops and their relatives using molecular data. In fact, with the exception of several studies that employed isozyme analysis (see Doebley, 1989, for a review), reports of introgression between crops and their relatives have been based almost exclusively on field observations of morphology. A few cases of introgression between crops and their relatives have been discovered during studies whose primary intention was to examine molecular variation in a crop or to elucidate its phylogeny.

Palmer et al. (1983) found two accessions of rape (B. napus) that possessed a foreign cpDNA type as compared with other accessions of this species. Analysis of rDNA of these two accessions indicated that their nuclear genomes were not different from other B. napus. This strongly indicates introgression of a foreign cytoplasm into the nuclear background of B. napus. Doebley and Sisco (1989) reported a similar situation in maize where a form of cytoplasmic male sterile maize (CMS-S) possesses the cytoplasm of the teosinte Zea mays ssp. mexicana as the result of introgression. Hosaka and Hanneman (1988b) found foreign (wild-type) cpDNAs in nine accessions of the potato out of more than 140 accessions analyzed. This suggests introgression of the cytoplasms of wild species into the nuclear background of potato, although, in the absence of analysis of the nuclear genome, multiple domestications cannot be conclusively rejected. A situation similar to that with potato was described by Palmer et al. (1985) for pea. Doebley et al. (1987) and Doebley (1990) demonstrated that maize and the Mexican annual teosintes share four distinct cpDNA genotypes. Because evidence from the nuclear genome indicates that maize was domesticated only once (Doebley, 1990), the shared presence of these four cpDNA types probably represents introgression of the cytoplasms of teosinte into maize or the reverse.

All of the above evidence for introgression between crops and their relatives involves the introgression of a foreign cytoplasm into a distinct nuclear background. Introgression of this nature is relatively easy to document because cyto-
plasmic genomes can provide several linked markers (restriction site mutations) that, unlike markers in the nuclear genome, cannot recombine. A more difficult task is the documentation of introgression for the nuclear genome, although this is possible through the analysis of RFLPs of low-copy-number nuclear sequences or rDNA. In such studies, it will be important to eliminate explanations other than introgression, such as convergence, joint retention of the ancestral condition (allele), and multiple domestications from distinct wild types.

**Future Prospects**

To understand fully the domestication process, one must explain the genetic basis of the morphologic and physiologic changes that differentiate crops from their progenitors. These changes are many, including retention of the seed/fruit on the plant at maturity, loss of germination inhibitors, increase in seed/fruit size, changes in starch, sugar, and protein content of the seed/fruit, loss of bitter substances in the seed/fruit, increase in vegetative vigor, increase in apical dominance, restoration of fertility to sterile floral parts, synchrony of flowering, more determinate growth, and a smaller number of larger fruits or inflorescences (Harlan, 1975; Simmonds, 1979). For some of these traits, such as retention of seed, genetic analyses indicate that they are under simple genetic control, often involving one or two genes (Harlan et al., 1973). However, there is little evidence concerning genetic control of most traits distinguishing crops from their progenitors. Harlan (1975) stated that the number of genes controlling wild versus domesticated morphologies is often rather small. Evidence for this point of view comes from studies of F₂ populations derived from wild–domesticate crosses. In such populations, plants resembling the parental types can be recovered at relatively high frequencies (Beadle, 1972).

The joining of molecular biology and quantitative genetics offers crop evolutionists a powerful opportunity to investigate genetic control of the morphologic evolution of crop species. Through the use of RFLPs as markers to locate quantitative trait loci, the minimum number of genes controlling a trait and the chromosomal locations of these genes can be discerned (Edwards et al., 1987; Paterson et al., 1988). Moreover, the relative contributions of these genes to the total variance for a trait can be estimated, and thus, major loci distinguished from modifier loci. Thus, it will be possible to construct a detailed picture of the genetic basis of the morphologic differences between crops and their progenitors.

This type of analysis has been applied to the origin of maize from its progenitor teosinte (Doebley et al., 1990). The results indicated that each morphologic trait distinguishing maize from teosinte is under the control of four or more chromosomal regions. However, the chromosomal regions affecting any one morphological trait tend to differ markedly in the magnitude of their effects. Regions with major effects on the traits that distinguish maize and teosinte are restricted to five of the ten chromosomes (Fig. 9.3), whereas regions with small
Figure 9.3. Schematic diagram of the 10 chromosomes of maize and teosinte showing the locations of regions with large effects on the morphologic traits that distinguish maize from its progenitor, teosinte (stippled blocks). Cross-bars indicate the locations of molecular marker loci (RFLPs or isozymes) used to detect the presence of morphologic trait loci. Thickness of the cross-bars represents the probability that segregation at a particular marker locus fit Mendelian expectations. The scale indicates the recombination fraction (r) between marker loci (Doebly et al., 1990).
effects are found throughout the entire genome. This explains why maize-like and teosinte-like plants can be recovered in relatively high frequencies in F$_2$ populations derived from maize–teosinte hybrids (Beadle, 1972), despite the fact that a large number (20 or more) of loci were involved in the domestication of maize.

Harlan (1975, p. 140) comments that one should expect considerable linkage of genes controlling the differences between wild and cultivated morphologies. This opinion is based largely on theoretical arguments that groups of genes controlling a trait will respond more readily to selection when they are linked. Proposed cases of this phenomenon include wheat (see Harlan, 1975) and maize (see Beadle, 1972). In the case of maize, the genes distinguishing maize from teosinte (its progenitor) are said to be assembled in five or six blocks (Beadle, 1972). Galinat (1988) speculates that cryptic structural changes within these blocks suppress crossing over, and thereby, act to preserve the integrity of the blocks. Thus, when a maize–teosinte hybrid is formed, parental types will be recovered among its progeny at higher than expected frequencies because of suppression of crossing over within the blocks. These ideas on linkage relationships can be readily tested through the analysis of linkage for low-copy-number nDNA sequences and quantitative trait loci controlling the differences between crops and their progenitors. In a maize–teosinte F$_2$ population, restriction to recombination was observed throughout the entire genome, as opposed to being confined to regions containing the morphologic trait loci (Doebly et al., 1990).

With the start of domestication, incipient crop species were placed under dramatically different selective regimes from those which they experienced in nature. Therefore, it is possible that the rate of evolution for some genes may differ under domestication as compared to their rate under natural selection. Furthermore, the domestication process may have subjected incipient cultigens to periods of intense stress, either through inbreeding, hybridization, or other means. During these episodes, transposable elements may have been mobilized, causing both restructuring of the genome and elevated mutation rates (McClintock, 1984). One wonders if the Northern Flint Corns, which differ so dramatically from all other types of corn, are not the result of some such process (Doebly et al., 1986). Molecular analyses of the genomes of crops and their progenitors are certain to provide new insights into these questions.

Conclusions

Molecular systematics has contributed substantially to our understanding of crop evolution in several capacities. First, it has clarified the phylogenies and origins of many crop species, including cotton and potato, whose origins have been surrounded by uncertainty. Second, molecular analyses have provided support for previous allozymic evidence that crops possess lower levels of genetic diversity than their progenitors, although more studies addressing this issue are
clearly needed. Third, molecular evidence has revealed several cases in which the cytoplasmic genomes of cultigens have been replaced by those of their relatives through introgressive hybridization. Molecular analysis of introgression of nDNA between crops and their relatives promises to provide new insights into this critical issue in crop evolution. Finally, molecular approaches will play an important role in understanding the morphologic evolution of crops and the evolution of the genomes of crop species under domestication.

References


Harlan, J.R. (1975) *Crops and Man*, Crop Science Society of America, Madison, WI.


Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E., and Tanksley,


