

Cytoplasmic DNA markers, phylogeny, and systematics in Oryzeae

G. Second

Studies of cytoplasmic DNA restriction fragment length polymorphism (RFLP) were undertaken to investigate phylogenetic relationships among the Oryzeae (mostly genus *Oryza*, section *Oryza*, including not only cultivated rice but also the whole tribe Oryzeae). Comparisons with nuclear encoded markers such as isozymes and nuclear DNA RFLP markers were made. It appears that the diversity observed at the level of total chloroplast DNA generally supports the scenario of evolution proposed earlier on the basis of isozymes. Some discrepancies are easily resolved as cases of polyphyletic origin (introgression, allotetraploidization) of newly evolved species, assuming maternal inheritance of cytoplasm. The case of the *O. latifolia* allotetraploid complex from America, showing a unique chloroplast DNA although it was assumed to have evolved recently, could not be resolved by conventional interpretation, however. A hypothesis of recombinational origin, assuming occasional biparental inheritance of cytoplasm, was supported by additional RFLP observations. A new hypothesis for the origin of the D genome, unknown in nature at the diploid level, is thus proposed. Diversity at the mitochondrial DNA level, as seen through hybridization with 11 mitochondrial gene probes, paralleled that of the chloroplast DNA between genomes. It was low within species or even genomes. There was, however, more polymorphism in newly evolved species, including cultivated rice. Also, species with the CD genome showed particular relatedness with the E genome at the mitochondrial DNA level. The importance of the Oryzeae in evolutionary studies that bear on plant breeding is stressed. Phylogenetic analysis provides a basis for a revised systematics of genus *Oryza*, section *Oryza*, in which ancestral species are distinguished from "species" or forms that have evolved since the disturbance of the paleodistribution of *Oryza* by man, including cultivated and weedy forms as well as wild "species."

A sound taxonomy and the evaluation of germplasm diversity must reflect phylogenetic relationships, particularly when the treatment of genetic resources is concerned.

Isozymes, which are primary products of genes, provided a new insight into genetic diversification of the section *Oryza* (=Eu-*Oryza*) of genus *Oryza*, including cultivated

475



Fonds Documentaire ORSTOM
Cote : B* 15191 Ex: 1

rice (summarized in Second 1985b). The observed variation clearly corroborated the earlier definition of genomes by cytogeneticists, and it was possible to adopt their nomenclature, from the A genome to the E genome. Much variation within genomes was also observed.

Reasoning from the geographic structure of the genetic variation seen at the isozyme level, in the framework of paleogeography and general knowledge of the botany and genetics of rice, I propose this scenario for the evolution of *Oryza* that summarizes the inferred phylogenesis.

Section *Oryza* is composed of two natural groups of species with an origin in the early differentiation of *Oryza* in Eurasia during the Tertiary Era: 1) the "Sativa" group, including all cultivated, weedy, and wild forms or species sharing the A genome; and 2) the "Latifolia" group of species with the diploid genomes B, C, and E and the tetraploid genomes BC and CD, all wild or weedy. Because rice seeds do not migrate naturally over long distances (a postulate validated by the consistency of the evolutionary model summarized in this paper), pan-tropical distribution of both groups resulted from the combined processes of 1) natural migration by land from Eurasia to Africa and Australasia along favorable routes (which occurred temporarily until the Pleistocene) and 2) the intervention of man, who carried seeds of wild species mixed with cultivated varieties in his migrations. (Most wild species considered here are sometimes observed in cultivated fields. Besides, wild species have received attention from man for medicinal or mythical purposes.) By transporting seeds across natural barriers to migration and by perturbing the natural habitats, man unconsciously promoted the evolution of new wild species and weedy forms of rice.

If this is correct, the surprisingly close genetic relationship between all *Oryza* species found in America (A and CD genomes) and some of the Old World species allows the conclusion that *Oryza* has been introduced to America by man. The origin of the D genome of the CD allotetraploid American species, however, remains unclear.

Additional genetic molecular markers are now available in an infinite number since the development of restriction fragment length polymorphism (RFLP) techniques for both nuclear and cytoplasmic encoded DNA. A nuclear DNA RFLP survey of the *Oryza* section of *Oryza* was recently made (unpubl. data). Not only does it provide additional information on some accessions, but new accessions were made available and studied for the first time. Additional information mostly confirmed the picture of genetic variation obtained at the isozyme level. This means that the genetic structure revealed is very strong and can be evidenced by using only a few loci. The only conflicting relationships were found in forms whose polyphyletic origin is consistent with the above-mentioned scenario: depending on the genetic loci considered, the affinity is then generally with one or the other parental species. In such cases, a limited number of loci are unable to provide an accurate picture. Polyphyletic origin was shown in particular for *O. rufipogon* from America (= *O. glumaepatula*) and from western India as well as for some weedy forms of *O. barthii* (= *O. breviligulata*). Information on newly collected accessions concerned mostly *O. officinalis* from China and the perennial form of *O. rufipogon* in Australia. *O. officinalis* from China was

shown to be different at numerous loci from *O. officinalis* from India. This variation, however, was unable to account for the origin of the D genome, which appeared to be a highly polymorphic genome not closely related to any known diploid form (the possibility of its future discovery at the diploid level appears to be very unlikely in America; Second 1990). Note, however, that neither *O. eichingeri* from Sri Lanka (Vaughan 1990) nor *O. officinalis* from Papua New Guinea has been studied. Comparing the CD genomes with diploid genomes, the E genome appeared the closest, after the C genome, confirming a trend also observed at the isozyme level. The perennial form of *O. rufipogon* from Australia appeared to be close to Asian *O. rufipogon*, contrary to the annual form of *O. rufipogon* from Australia (= *O. meridionalis*), which was confirmed to be highly divergent.

The potential of RFLP studies in phylogenetic investigations of the Oryzeae has not yet been fully utilized. Probes better characterized in terms of the evolutionary conservatism of the DNA fragments they hybridize to, mapped with confidence and used in large numbers, should give a complete picture of the genetic relationships among the Oryzeae tribe, both across wide genetic distances as well as between closely related forms. Introgressed forms in particular should be characterized in terms of which chromosome fragment(s) was introgressed.

We will restrict ourselves here to the information obtained from RFLP studies of the DNA encoded in the chloroplast and the mitochondria. Because the cytoplasm is generally maternally inherited in grasses, as in most plants, these markers should trace the phylogeny of the cytoplasm. In the case of genetic divergence in complete isolation during a long evolutionary time, both the nuclear and the cytoplasmic encoded markers are expected to diverge simultaneously and thus to show the same relatedness with other genomes. However, when independent evolutionary lines merge in allotetraploidization, hybridization, or introgression, nuclear and cytoplasmic encoded markers may be expected at times to show contrasting relationships. Moreover, we should bear in mind that nuclear and cytoplasmic DNA are coadapted (not all possible combinations are equally fit); that some DNA sequences occur ubiquitously in nuclear, chloroplast, and mitochondrial DNA; and that cytoplasm can occasionally be inherited biparentally, as is suggested in particular for rice (Second et al 1990).

Chloroplast DNA RFLP and phylogeny in the Oryzeae

Chloroplast DNA (ctDNA) is the most conserved during evolution, compared with mitochondrial and nuclear DNA. For this reason, and also because it has never been shown to recombine in nature in higher plants, it is thought to be an ideal marker of phylogenetic relationships, particularly above the species level.

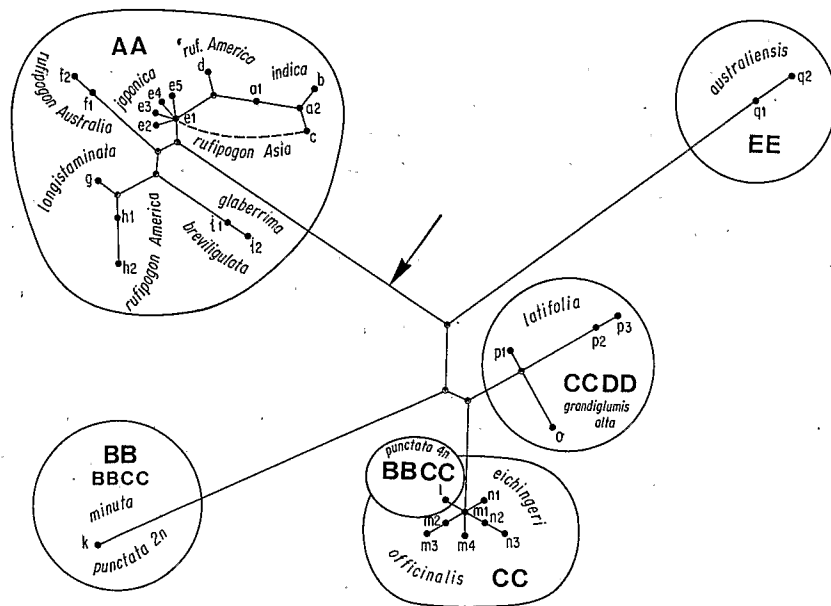
Section Oryza

An innovative nonaqueous technique (Dally and Second 1989) was used to purify ctDNA from 320 single plants representing the breadth of the available variation within section Oryza. Complex restriction patterns obtained with the enzymes *EcoRI* and

*Ava*I were studied in all plants, and 32 plastotypes (chloroplast genotypes) were distinguished. These 32 plastotypes were further characterized for their restriction patterns with *Bam*HI, *Pst*I, *Hind*III, *Bst*EII, and *Sma*I enzymes. The restriction patterns were observed directly under fluorescence in ethidium bromide, but at least one specimen of each DNA pattern was saved through Southern blot transfer for later hybridization with selected probes. The molecular size of restriction fragments was carefully determined by comparison with a DNA "ladder" of known molecular size. The data permitted determination of the type of mutations, either additions/deletions or site changes, that explained the observed variation between patterns. Because reproduction of the ctDNA molecule is clonal (at least generally), a cladistic analysis (by successive branching, according to shared mutations) was preferred to the multivariate analysis adopted for nuclear encoded markers for which the importance of recombination in reproduction is paramount (Dally 1988, Dally and Second 1990).

As the complete sequence of ctDNA from a japonica cultivar is now available (Hiratsuka et al 1989), it is possible to compare the observed patterns for the most common japonica-related plastotype with the patterns deduced from the sequence. The comparison shows that the estimation of molecular sizes of restricted fragments by Dally (1988) was accurate (generally less than 5% error). Major discrepancies occurred only in some fragments of *Ava*I and *Sma*I restriction patterns. Hybridization with probes showed they were due to errors of the sequence in the corresponding CG-base-rich restriction sites.

Major additions/deletions, which are detected with several enzymes, and many of the presumed restriction sites can be readily located on the physical map of rice ctDNA deduced from its sequence. Systematic hybridization of selected probes (covering the whole ctDNA molecule) with the DNA patterns saved on filters allowed resolution of most remaining ambiguities. Revision of the type of mutation presumed was necessary only in some cases where there were a large number of mutations to be explained



1. Cladogram showing relatedness among 32 plastotypes distinguished in *Oryza* section (Dally 1988). Capital letters stand for nuclear genomes and small letters for plastotypes. Length of a branch is approximately proportional to number of mutations specific to that branch. Arrow indicates "root." Dashed line indicates a suggested possible origin in recombination of the *c* plastotype. As discussed in text, CCDD

longistaminata occurred (Second 1985a). In both cases, the chloroplast genome of the most recent immigrant from Asia was retained.

The fact that all forms of *O. rufipogon* from Australia (and also the "Oceanian" form of *O. rufipogon* from New Guinea [Ishii 1991]) share the same two unique plastotypes supports their recognition as a distinct species, *O. meridionalis* Ng, with three subspecies: annual Australian, perennial Australian, and New Guinean. The perennial subspecies is, however, similar to Asian *O. rufipogon* at the morphological level as well as at most nuclear isozyme and RFLP loci. It may thus be interpreted as a result of hybridization between Asian *O. rufipogon* and *O. meridionalis*.

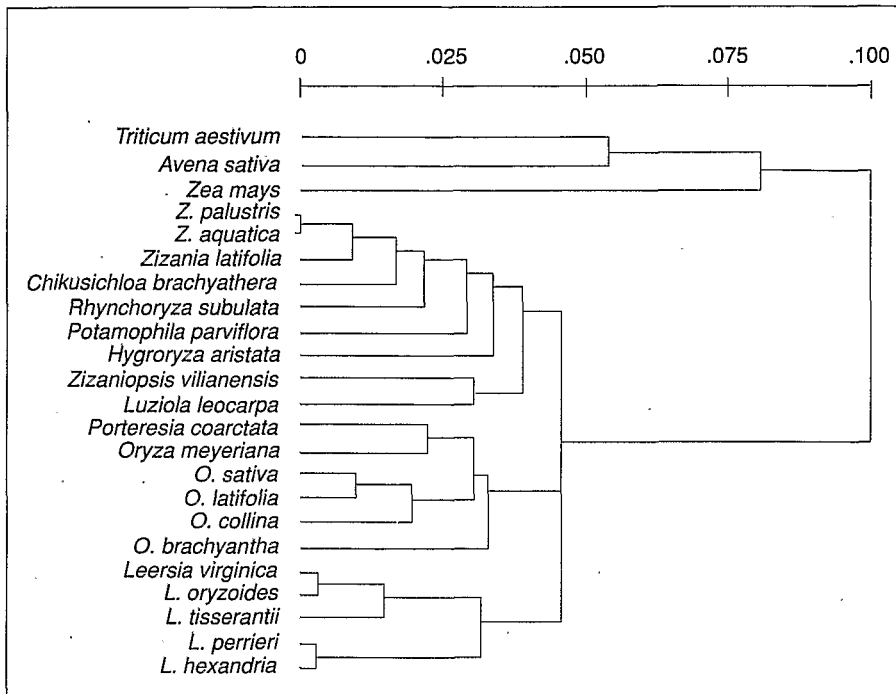
The high heterogeneity of the plastotypes of *O. rufipogon* from America and their close relation to Old World plastotypes again point to its polyphyletic and recent origin.

The Latifolia group of species appears composed of four clusters of plastotypes. Three of them correspond to the diploid genomes B, C, and E. The allotetraploids BC share a plastotype identical or close to either the B or C genome plastotypes. At least two events of allotetraploidization are then suggested for the origin of the BC genome. The fourth cluster corresponds to the CD genome, which appears particularly polymorphic.

The high degree of divergence of the B and E genome plastotypes is in line with their presumed long period of evolution in isolation in Africa and Australasia, respectively. Application of the molecular clock concept, based on published estimations of mutation rates for ctDNA, point, for both of them, to a time of divergence, around 15 million yr ago, in agreement with the proposed evolutionary scenario (Dally 1988). From the same perspective, however, the divergence of CD genome plastotypes is in clear contradiction with the presumed recent emergence of that genome in America. Conventional interpretation would point to an ancient differentiation of the CD genome or of one of its (missing) progenitors. Because this violates the postulate that seeds of rice do not spontaneously cross large stretches of ocean, of which we are confident because it appears to be validated by other observations, an alternative explanation was sought. Evidence for an exceptional case of rapid differentiation of that plastotype was found in its similarity, not only to C genome plastotypes, but also to the B genome plastotype (Dally and Second 1990). Mapping of the respective mutations and further characterization with additional restriction enzymes did point to an origin of the CD plastotypes in a combination (recombination or an unknown mechanism?) of derived mutations characteristics of both the C and B plastotypes (unpubl. data), but more data are necessary to support a positive conclusion.

Tribe Oryzeae

A preliminary analysis of phylogenetic relationships among tribe Oryzeae (as understood in Clayton and Renvoize 1986) at the ctDNA RFLP level was conducted by Zhang and Second (1990). Figure 2 illustrates the main results, which agree fairly well with the modern treatment of the genera *Leersia* and *Oryza*. A striking feature is, however, that two morphologically similar *Leersia* species, *L. tisserantii* from Africa



2. Average linkage dendrogram based on genetic distances for ctDNA RFLP among individual representatives of various species and genera of tribe Oryzeae and 3 other genera of grasses (Zhang and Second 1990).

and *L. perrieri* from Madagascar, appear to be widely divergent at the ctDNA level, as much so as the most divergent species in the genus *Oryza*. Both of them used to be placed in the genus *Oryza*, indicating their morphological relatedness to that genus. A tempting hypothesis is thus that these two *Leersia* species are relics of a common ancestor of both *Leersia* and *Oryza*. The common ancestor was born on the African plate and diverged early because of the separation of Madagascar from it. A similar hypothesis was proposed by Second (1985c) on the basis of the biogeography of the Oryzeae. A prediction made from that model was that species such as *L. tisserantii* and *L. perrieri* could be diploids, while all the *Leersia* species so far observed for chromosome number were reported to be polyploids (Pyrah 1969). A comparison of the number of bands in nuclear DNA RFLP patterns (per plant, among species) fulfills this prediction. Another prediction of the model was that the introduction of *Oryza* in

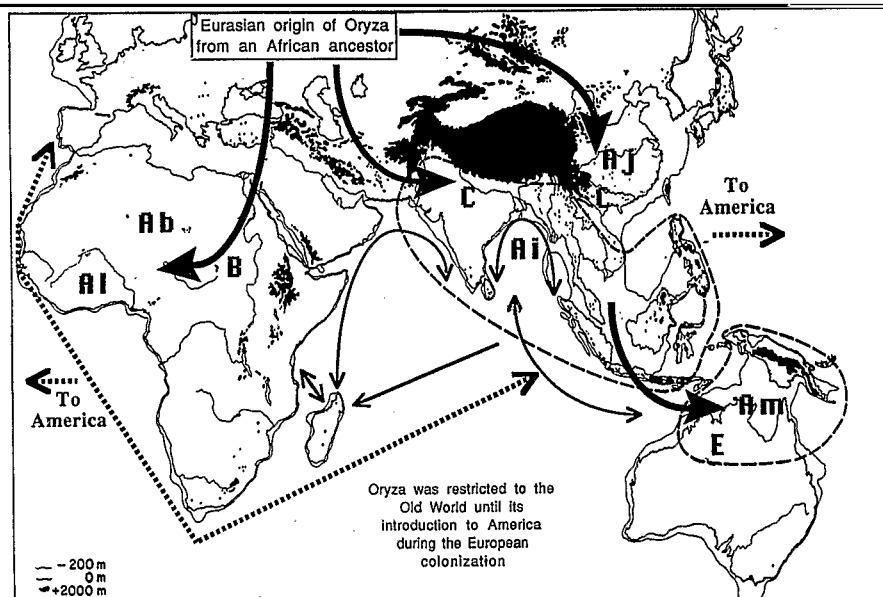
Mitochondrial DNA RFLP in section *Oryza*

Mitochondrial genes used as probes were hybridized on filters prepared from *Eco*RI-restricted total DNA and exhausted for their potential to hybridize with single-copy probes to a detectable level (unpubl. data). Checks for cross hybridization with ctDNA were made on filters prepared from purified ctDNA. Preliminary results are presented for 7 probes showing polymorphism out of 11 used.

A low degree of polymorphism of the mitochondrial genome (chondriome) was detected within species (although several cases of polymorphism were found in cultivated varieties), but a high degree of polymorphism was observed between the main nuclear genomes. Table 1 omits within-species variation. Most of the probes distinguish the basic genomes A, B, C, and E. BC allotetraploids have an affinity with either the B or C mitochondrial type according to their relationships at the ctDNA level. However, there is some variation, compared with the parental genomes, in the allotetraploid species, and also in those C genome species considered to have recently introgressed some DNA from the B or E genome (*O. eichingeri*, *O. rhizomatis*). A similar trend of species considered to be recently derived to show more polymorphism was observed at the ctDNA level (Dally and Second 1990) but is more developed at the mitochondrial level. The CD species chondriome is similar to the E species chondriome in three cases out of seven probes showing polymorphism, reinforcing the tendency that its relationship to other species cannot be explained in a conventional

Conclusion

The pattern of variation observed at the cytoplasmic DNA level fits nicely within the framework of the phylogenetic relationships worked out from the nuclear encoded markers. It reveals additional details, particularly where polyphyletic origins are concerned. The conservation of ctDNA allows one to obtain information across the whole tribe Oryzeae and provides evidence of surprising relationships. A summary of the evolutionary scenario for section *Oryza* is given in Figure 3. Differentiation of genomes in the *Latifolia* group occurred in allopatry on different continents: genome B in Africa, C in Asia, and E in Australasia. The absence of genome differentiation (as seen at the chloroplast DNA level) among the *Sativa* group is attributed to its



3. Deduced paleodistribution of genus *Oryza* before man promoted its migration across natural barriers (modified from Second 1985c). Map shows main physical constraints that, along with favorable or unfavorable climate, regulate possibilities of natural migration: thin line along coast is -200 m isochore and shaded areas represent high elevations; during glacial ages, blackened areas show zones above

Oryza, section Oryza

SATIVA GROUP (Genome A)

Ancestral species

O. rufipogon Griff.
(a "complex species")

O. barthii A. Chev.
= *O. breviligulata*
A. Chev. & Roehr.

O. longistaminata
A. Chev. & Roehr.

O. meridionalis Ng
annual ssp
perennial ssp
New Guinean ssp

Cultivated forms

*O. sativa***:
indica ssp
japonica ssp:
tropical
temperate
(+intermediates
and others)

O. glaberrima

Derived forms or "species"

Weedy forms

Lists by no means exhaustive*

O. spontanea:
"red rice" (USA)
"akai-mai" (Japan)
"crodo" (France)
"purure" (Zanzibar)
etc...

O. stapfii

O. madagascariensis
"Obake" forms =
O. longistaminata/O. sativa

Newly evolved wild forms

O. nivara (in West Indies
= *O. rufipogon/O. breviligulata*)

American *O. rufipogon*:
(= *O. rufipogon/O. longistaminata/O. breviligulata*, and others...)

Australian red rice

?

LATIFOLIA GROUP (Genomes B to E)

Ancestral species

O. officinalis Wall ex Watt
(a "complex species",
genome C)
Chinese ssp
South Asian ssp

O. punctata Kotschy
(genome B)

O. australiensis Domin
(genome E)

Newly evolved "species"

Some names of imprecise application*

**Genome C introgressed
from B or E(?)**

Weedy *O. officinalis*
O. eichingeri complex
O. thizomatis

Allotetraploid complex:

genome BC
O. punctata 4n
(= *O. schweinfurthiana*?)
O. malampuzhaensis
O. minuta

genome CD
genome D= modified B+E?
O. latifolia
O. alta
O. grandiglumis

*See Vaughan (1989 and 1990) for synonyms.

**Of hybrid origin between geographic forms of *O. rufipogon*.

4. The systematics of *Oryza*, section *Oryza*. Two largely independent groups of species are recognized: Sativa and Latifolia. Taxa are arranged by lines and columns. Columns are arranged according to an "evolutionary concept" of species: ancestral forms (in boxes), cultivated, weedy, and newly evolved wild forms (following the disturbance of distribution and habitats by man). Lines are arranged according to a "biological species concept" considering mostly the main genetic reproductive barriers. In Sativa group, for example, a single species would represent *O. rufipogon* and *O. sativa*, including all weedy forms. It should be called *O. sativa*. However, such possible revision of the nomenclature would not bear on a single universally recognizable species concept. Moreover, it does not appear to be practical. The present treatment is thus preferred.

On herbarium specimens, many taxa cannot be determined with accuracy morphologically, although knowing the geographical origin and habitat helps to discriminate them (see Vaughan 1989). On live accessions, there is clear congruence of morphology, cytogenetics, molecular markers for nucleus, molecular markers for cytoplasm, geographical origin, ecology, and the knowledge of reproductive barriers to characterize ancestral species. However, both Asian species *O. rufipogon* and *O. officinalis* appear to be "complex species." As ancestral species, they include subspecies differentiated on both sides of the Himalaya. Moreover, they include a complex intermingling of ancestral and derived forms.

ecological adaptation, including temporary pools in arid savannah; this allowed it to migrate between continents of the Old World later than the forms of the Latifolia group, which are adapted primarily to more humid climatic areas (at least for its representatives from Asia). The A genome is, however, the most diverse of all genomes distinguished by cytogeneticists. Subgenome differentiation of the A genome clearly parallels in its geographic distribution the genome differentiation of the Latifolia group.

Confidence in the proposed evolutionary scenario led to preliminary evidence that a combination of two widely different plastotypes might explain the rapid differentiation of the CD genome plastotype from America. From an historical point of view, it makes sense that, after the emergence of BC allotetraploids, with both B and C plastotypes, in the Old World, and possible intervention of the E genome, the two of them were introduced to America (possibly as weeds), and their hybridization allowed the evolution of the CD plastotypes and nuclear genomes. The origin of the D genome, so far mysterious, could be in its rapid differentiation at the tetraploid level. A combination of the B and E genomes appears likely to be involved. This hypothesis for the origin of the D genome, presented here for the first time, is suggestive of the profound genome modifications that could be possible in breeding work. Adaptation of forms with new characteristics to new environments, such as those of the American continent, seems to have systematically involved recombination between African and Asian lineages. Transposed to plant breeding, this is the concept of transgressive variation through recombination.

The exact processes of recombination in the ctDNA molecule need to be worked out more precisely by sequencing. If proven, the fact that we were led to the likely existence of that phenomenon would exemplify that the Oryzae provide an outstanding model for the study of plant evolution under domestication or human disturbance. The Oryzae model, with a whole gradation of divergence time (from millions of years to a few centuries, as calibrated by events in the paleoenvironment and in history) and a wide variety of life history types and ecological situations, is a powerful tool for addressing evolutionary questions that have bearing in the breeding of rice at the molecular as well as organismic levels.

Phylogenetic information provides a basis for a revised systematics of *Oryza*, section *Oryza*, shown in Figure 4, in which ancestral species are distinguished from

- Dally A M, Second G (1989) Chloroplast DNA isolation from higher plants: an improved non-aqueous method. *Plant Mol. Biol. Rep.* 7(2):135-143.
- Dally A M, Second G (1990) Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section *Oryza*). Cladistic-mutation and genetic-distance analysis. *Theor. Appl. Genet.* 80:209-222.
- Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C R, Meng B Y, Li Y Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. *Mol. Gen. Genet.* 217:185-194.
- Ichikawa H, Hirai A, Katayama T (1986) Genetic analyses of *Oryza* species by molecular markers for chloroplast genomes. *Theor. Appl. Genet.* 72:353-358.
- Ishii T (1991) Cytoplasmic and nuclear differentiation in A genome diploid species of rice as revealed by the restriction length polymorphism analysis of DNAs. Ph D dissertation, Kyoto University. 115 p.
- Ishii T, Terachi T, Tsunewaki K (1986) Restriction endonuclease analysis of chloroplast DNA from cultivated rice species, *Oryza sativa* and *Oryza glaberrima*. *Jpn. J. Genet.* 61:537-541.
- Ishii T, Terachi T, Tsunewaki K (1988) Restriction endonuclease analysis of chloroplast DNA from A-genome diploid species of rice. *Jpn. J. Genet.* 63:523-536.
- Pyrh G L (1969) Taxonomic and distributional studies in *Leersia* (Gramineae). *Iowa State J. Sci.* 44:215-270.
- Second G (1985a) Evolutionary relationships in the Sativa group of *Oryza* based on isozyme data. *Genet. Sel. Evol.* 17(1):89-114.
- Second G (1985b) Isozymes and phylogenetic relationship in *Oryza*. Pages 27-39 in *Rice genetics*. International Rice Research Institute, P.O. Box 933, Manila, Philippines.
- Second G (1985c) Relations évolutives chez le genre *Oryza* et processus de domestication des riz. *Coll. Etudes et Theses, ORSTOM. Paris.* 189 p.
- Second G (1987) Field notes of, and observations on, a collection of wild rice species in Australia, 3-31 May 1987. Trip report. 12 p. (mimeo.).
- Second G (1990) Additional observations and collection of the *Oryza latifolia* Desv. species complex. *Rice Genet. Newsl.* 6:73-76.
- Second G, Dally A M, Zhang S H (1990) Biparental inheritance of chloroplast DNA in rice. *Rice Genet. Newsl.* 6:150-153.
- Vaughan D A (1989) The genus *Oryza* L.: current status of taxonomy. *IRRI Res. Pap. Ser.* 138.
- Vaughan D A (1990) A new rhizomatous *Oryza* species (Poaceae) from Sri Lanka. *Bot. J. Linnean Soc.* 103:159-163.
- Zhang S H, Second G (1990) Phylogenetic analysis of the tribe Oryzaeae. Total chloroplast DNA restriction fragment analysis. A preliminary report. *Rice Genet. Newsl.* 6:76-80.

Notes

Author's address: G. Second, Institut Français de Recherche Scientifique pour le Développement en Coopération, Montpellier, France, and Plant Breeding, Genetics, and Biochemistry Division, International Rice Research Institute, P.O. Box 933, Manila, Philippines.

Acknowledgments: I thank J. Davis for stimulating discussions. I express my deep appreciation to S. Tanksley and his collaborators for their hospitality in his laboratory.

Citation information: International Rice Research Institute (1991) *Rice genetics II*. P.O. Box 933, Manila, Philippines.