DOI: 10.1534/genetics.104.033175

The Population Structure of African Cultivated Rice *Oryza glaberrima* (Steud.): Evidence for Elevated Levels of Linkage Disequilibrium Caused by Admixture with *O. sativa* and Ecological Adaptation

Mande Semon,* Rasmus Nielsen,† Monty P. Jones‡ and Susan R. McCouch*,[1]

**Department of Plant Breeding and* †*Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853 and* ‡*FARA, Accra, Ghana*

Manuscript received July 7, 2004
Accepted for publication November 30, 2004

ABSTRACT

Genome-wide linkage disequilibrium (LD) was investigated for 198 accessions of *Oryza glaberrima* using 93 nuclear microsatellite markers. Significantly elevated levels of LD were detected, even among distantly located markers. Free recombination among loci at the population genetic level was shown (1) by a lack of decay in LD among markers on the same chromosome and (2) by a strictly increasing composite likelihood function for the recombination parameter. This suggested that the elevation in LD was due not to physical linkage but to other factors, such as population structure. A Bayesian clustering analysis confirmed this hypothesis, indicating that the sample of *O. glaberrima* in this study was subdivided into at least five cryptic subpopulations. Two of these subpopulations clustered with control samples of *O. sativa*, subspecies *indica* and *japonica*, indicating that some *O. glaberrima* accessions represent admixtures. The remaining three *O. glaberrima* subpopulations were significantly associated with specific combinations of phenotypic traits—possibly reflecting ecological adaptation to different growing environments.

ORYZA glaberrima (Steud.) is a form of cultivated rice that was domesticated in the Niger River delta ~3500 years ago (Viguier 1939) and is widely grown in West Africa today. Germplasm collections of *O. glaberrima* at the West Africa Rice Development Association (WARDA), the International Institute for Tropical Agriculture and the International Rice Research Institute (IRRI) include ~2800 accessions. Estimates of genetic diversity in *O. glaberrima* based on RFLP and isozyme markers are significantly lower than those in cultivated Asian rice, *O. sativa* (Second 1982, 1986; Wang *et al.* 1992). Despite this fact and the smaller number of accessions available for study, *O. glaberrima* harbors a rich reservoir of genes that have allowed the species to survive and prosper in West Africa with minimal human intervention (Jones *et al.* 1997). Recently, new, high-yielding varieties suitable for cultivation in the West African region (NERICA varieties) have been developed from interspecific crosses between *O. glaberrima* and *O. sativa* (http://www.warda.org).

There are believed to be three centers of domestication for *O. glaberrima*, in Mali, the Sene-Gambia, and Guinea (Portéres 1970), and this may have contributed to the broad ecological adaptation of African rice cultivars today. While genetic evidence points to a common ancestral gene pool (*O. barthii*) for all *O. glaberrima* domesticates, the recent nature of rice domestication in West Africa offers a view into the early stages of genetic differentiation among subpopulations. As a self-pollinating species that has undergone a founder effect, the degree of variation between populations is expected to exceed the variability observed within a field or population, and thus *O. glaberrima* offers an unusual opportunity to detect and characterize the nature of emerging population structure.

In domesticated species, levels of gene flow among subpopulations are determined by both human-mediated movement of germplasm along major trade routes and natural gene flow resulting from cross-pollination. Between the 15th and 17th centuries, *O. sativa* was introduced to West Africa by Arab traders and Portuguese navigators and rapidly spread throughout West Africa. Today, both *O. glaberrima* and *O. sativa* are commonly grown in mixtures by farmers in upland and rainfed lowland environments. Natural intermediates between the two species have been reported, but the outcrossing rate is estimated to be low (between 2 and 5%). While *O. glaberrima* can be crossed with *O. sativa*, the F_1 offspring are male sterile and can survive only if pollinated by the mother species. When this occurs, interspecific hybrid progeny can be very productive, as demonstrated by the new, high-yielding varieties for the West Africa region (Jones *et al.* 1997; http://www.warda.org).

It is of interest to understand the structure and evolution of *O. glaberrima* in West Africa. Information on diversity and population structure is expected to assist plant breeders in the selection of parents for crossing,

[1]*Corresponding author:* Plant Breeding Department, Cornell University, 162 Emerson Hall, Ithaca, NY 14853. E-mail: srm4@cornell.edu

providing a more rational basis for expanding the gene pool and for identifying materials that harbor alleles of value for plant improvement. It is also of interest to understand the early stages of speciation and to gain an appreciation of how populations diverge and differentiate over time.

In this study, we used 93 SSR markers to evaluate the extent of linkage disequilibrium and population structure in a collection of 198 accessions of *O. glaberrima*. We were interested in assessing the relative importance of introgressions, admixture, geography, and ecological specialization in shaping the pattern of genetic variability in the species and, in particular, evaluating the influence of *O. sativa* on the genetic structure of *O. glaberrima*, given that the two species have been grown together in Africa for 300–500 years. Finally, we aimed to provide some insight into the long history of failures as well as the more recent successes in interspecific hybridization between *O. sativa* and *O. glaberrima*.

MATERIALS AND METHODS

Plant material: One hundred ninety-eight accessions of *O. glaberrima* were obtained from the germplasm collection at WARDA and purified by harvesting single panicles from individual plants grown in the field at M'bé, Ivory Coast, during June–October 1998. Nine accessions of *O. sativa*, representing both the *indica* and *japonica* subspecies, were used as an outgroup. Variety names and accession numbers are given in supplemental Table 1 at http://www.genetics.org/supplemental. The 198 accessions are part of a larger collection of 1130 *O. glaberrima* accessions maintained at WARDA and were selected to maximize geographical and morphological diversity.

DNA extraction, PCR conditions, and allele detection: DNA was extracted from single plants grown in the Guterman greenhouse at Cornell. DNA was extracted using the micropreparation method as described by Dellaporta *et al.* 1983. Ninety-three nuclear microsatellite markers were used for genotyping. PCR conditions for each primer pair were as described by Chen *et al.* (1997) and Temnykh *et al.* (2000, 2001). Amplification products were size separated using polyacrylamine gel electrophoresis and detected using silver staining as described by Panaud *et al.* (1996) or using an ABI fluorescent detection system as described by Coburn *et al.* (2002). The semiautomated method of marker detection was made using an ABI 373 sequencer and Genotyper software (Applied Biosystems, Foster City, CA).

Multiplex design: The molecular weight ranges of microsatellite alleles have been reported for *O. sativa* (Cho *et al.* 2000; Coburn *et al.* 2002) but are not identical to those of *O. glaberrima*; thus the size range of alleles in *O. glaberrima* was initially estimated using a subset of 20 diverse accessions, including two *O. sativa* varieties as checks. These size estimates were then used as the basis for the multiplexing scheme that was developed for the remaining accessions (as described in Coburn *et al.* 2002). Map positions for all SSR markers were based on the SSR map reported by Chen *et al.* (1997) and Temnykh *et al.* (2000, 2001) and available at http://www.gramene.org.

Linkage disequilibrium analysis: Classical measures of linkage disequilibrium consider only diallelic loci (Lewontin 1964; Hill and Robertson 1968). To generate diallelic data from the microsatellite data, for each locus we identified the most frequent allele and combined all the remaining alleles into a second "allele" class. Other methods for generating diallelic data were also explored and led to the same qualitative conclusions. The two measures of linkage disequilibrium (LD), standardized disequilibrium coefficients (D'; Lewontin 1964) and squared allele-frequency correlation (r^2; Hill and Robertson 1968), were then calculated for all pairs of loci. A linear regression was performed to investigate the decay of linkage disequilibrium with genetic distance. Also, the composite likelihood estimator of Hudson (2001) was applied to the data, to estimate the likelihood function for the population recombination parameter ρ. ρ is the population recombination rate ($2N_eR$, where N_e is the effective population size and R is the recombination fraction between loci). The composite likelihood function is constructed by multiplying together the pairwise likelihood for all sites. This composite likelihood function is then maximized with respect to ρ to provide an estimate of this parameter. Finally, to test for significant linkage disequilibrium, a test of association was performed for each pair of loci by randomly permuting the genotypes in one of the loci, using both D' and r^2 as test statistics.

Population structure analysis: Three different methods were used to assess population structure. First, a Bayesian clustering analysis was undertaken using the program STRUCTURE (Pritchard *et al.* 2000). This method uses multilocus genotypes to infer the fraction of an accession's genetic ancestry that belongs to a population, for a given number of populations (K). Because the species is highly selfing, we relaxed the assumption of Hardy Weinberg equilibrium and treated the data as haploid, so that groupings are based only on LD. The posterior probabilities were estimated using a Markov chain Monte Carlo method (MCMC). The results were based on 3,200,000 iterations of this chain, following a burn-in period of 800,000 iterations. The MCMC chain was run multiple times, using a correlated allele frequency model (prior mean is 0.01, prior SD = 0.05 and Lambda set at 1.0 in the advance option of the STRUCTURE program). In this study, an accession was assigned to a cluster if at least 75% of its genome value was estimated to belong to that cluster.

Second, a principal coordinate analysis was conducted, on the basis of similarity measures (S_{ij}) using the DCENTER and EIGEN procedures in NTSYS (Rohlf 1997) to cluster *O. glaberrima* accessions into major groups. A neighbor-joining (NJ) tree (Saitou 1987) was created with the TREE option of NTSYS and a goodness of fit of the clustering to the data was calculated using the COPH and MXCOMP procedures.

Third, a Mantel test was used to assess the hypothesis of genetic isolation by geographic distance. To measure genetic distances we used $\delta\mu^2$ (Goldstein and Pollock 1997), which is the square of the mean difference in repeat scores between subpopulations (a subpopulation here is a group of individuals from a country of origin).

Phenotypic evaluation: Each of the 198 accessions was grown in a three-row plot (one plant per hill) under sprinkler-irrigated conditions at WARDA M'bé in Ivory Coast in June 1998. Phenotypic data, corresponding to 18 yield components, were recorded on 10 plants from the inner rows of each plot and averaged across the 10 plants. Evaluation was similar to that described in Xiao *et al.* (1998) and the Standard Evaluation System for Rice (http://www.knowledgebank.irri.org/ses/SES.htm) as follows: (1) *Days to heading* was evaluated as the average number of days from seeding until 10% of the panicles had headed, (2) *days to maturity* was evaluated as the average number of days from seeding until 75% of grains were mature on the panicles, (3) *tiller number per plant* was measured as the average number of tillers per plant calculated for the 10 plants at maximum tillering, (4) *plant height* was measured as the average height of the 10 plants in centimeters from the soil

surface to the tip of the tallest panicle (awns excluded), (5) *panicle length* was measured as the average number of centimeters from the panicle neck to the panicle tip (excluding the awn) based on an evaluation of all panicles from the 10 plants, (6) *number of branches per panicle* was measured as the average number of branches per primary panicle calculated for the 10 plants at maturity, (7) *panicles per square meter* was the number of panicles on the 10 plants times 2.5 (panicles having less than five seeds were not counted for any of the panicle traits), (8) *panicles per plant* was the average number of panicles on the 10 plants, (9) *spikelets per panicle* was the average number of spikelets on the 10 plants, (10) *grains per panicle* was measured as the average number of filled spikelets per plant calculated for the 10 plants, (11) *spikelets per plant* was the average number of spikelets on the 10 plants, (12) *grains per plant* was measured as the average number of filled spikelets per plant calculated for the 10 plants, (13) *seed set rate* was the number of filled spikelets divided by the total number of spikelets per panicle evaluated for each panicle on all 10 plants, (14) *grain length* was measured as the actual measurement of length in millimeters of filled grains averaged over 10 randomly selected filled grains, (15) *grain width* was measured as the actual measurement of width in millimeters as the distance across the fertile lemma and the palea at the widest point, averaged over 10 randomly selected filled grains, (16) *grain length to width ratio* was measured as the ratio of grain length/grain width, (17) *yield per plant* was the average weight (in grams) of bulk-harvested grain per plant, and (18) *1000-grain weight* was the average weight of 1000 filled spikelets, measured in grams, averaged over three samples taken from bulk-harvested grain from the 10 plants.

Testing for association between genetic groups and phenotypic traits: To compare the phenotypes of the genetic groups identified by STRUCTURE, analysis of variance (ANOVA) was employed, followed by multiple means comparisons among the 18 phenotypic traits. To determine the level of significance for all pairwise comparisons, a Bonferroni correction for multiple testing was applied.

RESULTS

Diversity: Abundant SSR diversity was observed within the 198 accessions of *O. glaberrima*. An average of 9.4 alleles per locus was detected among the 93 SSRs analyzed, with a range of 2–27 alleles/locus, a mean polymorphism information content (PIC) value of 0.34, a gene-diversity (He) value of 0.27, and an allele size range of 67–388 bp. As summarized in Figure 1, the accessions were collected from 12 different countries throughout West Africa and collection sites represented a variety of ecological zones. A majority of the accessions were identified with Nigeria and Liberia, reflecting the location of the international centers that were responsible for the initial collections.

Population structure analysis: When the 198 *O. glaberrima* accessions and nine *O. sativa* control varieties were analyzed for population structure using a model-based approach (Pritchard *et al.* 2000), we identified five genetically distinct groups or admixtures thereof within *O. glaberrima*. Estimated likelihood values for a particular value of K were highly variable among runs, but the likelihood function appeared to be an increasing function of K for all examined values of K. With large amounts of data and small local deviations from Hardy Weinberg equilibrium, STRUCTURE may tend to identify more populations than are biologically relevant (Falush *et al.* 2003). While we chose a value of $K = 5$ for the final analysis, other values of K are possible and would not qualitatively affect our conclusions.

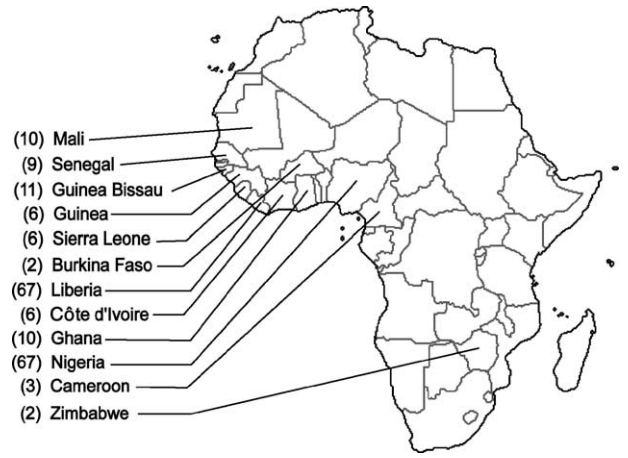

Figure 1.—Countries of origin of the 198 *O. glaberrima* accessions characterized in this study. Numbers in parentheses are the number of accessions surveyed for the corresponding country.

Table 1 summarizes the proportion of shared ancestry among the five genetically defined groups detected in the total collection of 207 accessions. Three of the groups (groups 1–3) were characteristic of *O. glaberrima* while two groups (groups 4 and 5) share ancestry with the two subspecies of *O. sativa*. Among the 198 accessions of *O. glaberrima*, 110 shared ≥75% ancestry with one of the five groups and were classified as members of that group. Group 1 was represented by 43 accessions (colored red in Figure 2A) that were distributed across all countries, but were less prevalent in collections from Liberia and Nigeria. Group 2, with 27 individuals (colored green in Figure 2A), was populated by accessions collected predominantly in Nigeria. Group 3 consisted of 34 individuals (colored yellow in Figure 2A), predominantly from Liberia (see supplemental Table 1 at http://www.genetics.org/supplemental/). No *O. glaberrima* accessions shared ≥75% ancestry with group 4, the *O. sativa*, ssp. *indica*-like group. Group 5, the *O. sativa*, ssp. *japonica*-like group (colored pink in Figure 2A), was represented by six accessions from Guinea and Sierra Leone. Eighty-eight (44%) of the *O. glaberrima* accessions were classified as admixtures, with varying levels of ancestry shared among the five groups. Three of these admixed accessions shared between 30 and 49% ancestry with the *indica*-like subpopulation (see supplemental Table 1 at http://www.genetics.org/supplemental/).

Genetic distance-based analysis: When principal coordinate analysis (PCA) or NJ was used to evaluate genetic relationships, the same eight *O. glaberrima* accessions clustered closely with the nine *O. sativa* accessions. This is illustrated for PCA in Figure 2B and for NJ in the unrooted

TABLE 1

Bayesian partitioning of ancestry within a collection of 198 accessions of *O. glaberrima* and nine cultivars of *O. sativa* based on SSR markers

No.	Origin	Groups					No. of accessions	Species
		1	2	3	4[a]	5[b]		
1	Burkina Faso	0.37	0.50	0.02	0.10		2	*O. glaberrima*
2	Cameroon	0.41	0.50	0.03	0.07		2	*O. glaberrima*
3	Ivory Coast	0.73	0.16	0.02	0.05	0.04	6	*O. glaberrima*
4	Ghana	0.47	0.36	0.09	0.07	0.01	10	*O. glaberrima*
5	Guinea Bissau	0.70	0.15	0.10	0.05		11	*O. glaberrima*
6	Guinea Bissau	0.31	0.02	0.01	0.07	0.59	6	*O. glaberrima*
7	Liberia	0.28	0.08	0.60	0.04		67	*O. glaberrima*
8	Mali	0.42	0.22	0.25	0.10		10	*O. glaberrima*
9	Nigeria	0.30	0.56	0.09	0.04	0.01	67	*O. glaberrima*
10	Senegal	0.63	0.25	0.08	0.03		9	*O. glaberrima*
11	Sierra Leone	0.40	0.12	0.06	0.03	0.39	6	*O. glaberrima*
12	Zimbabwe	0.88	0.09	0.03			2	*O. glaberrima*
13	Africa-Asia				0.08	0.91	9	*O. sativa*

[a] Group 4 represents shared ancestry with *O. sativa* (ssp. *indica*).
[b] Group 5 represents shared ancestry with *O. sativa* (ssp. *japonica*).

phylogram in Figure 2C. Neither of these distance-based methods had the resolution to discern significant fine population structure within *O. glaberrima*, as evidenced by the dendrogram in Figure 2D.

LD decay: To rigorously assess the interpretation of fine population structure within *O. glaberrima* as detected by STRUCTURE, we tested several of the assumptions underlying this approach. First, model-based population structure analysis requires a set of markers that are unlinked at the population genetic level. To confirm that no physical linkage was detected among our set of 93 markers, we tested for evidence of LD among markers on the same chromosome. Little or no decay of LD was observed as a function of genetic distance in the set of 198 accessions (Figure 3, A and B) at the marker density used, indicating effectively free recombination among the 93 loci. Furthermore, significant associations were observed for many of the comparisons among loci separated by >100 cM (25,000 kb; Figure 3, C and D). This clearly demonstrated that the loci are not in linkage equilibrium, even at long genomic distances. In addition, the composite likelihood surface is a strictly increasing function of ρ resulting in an estimate of $\hat{\rho} = \infty$ (Figure 3E). The fact that the data show strong evidence of LD, even at long distances, while there is no decay in LD, and the composite likelihood estimate of $\hat{\rho} = \infty$, argues that a factor such as population structure must be acting to increase levels of LD.

Isolation by distance: "Isolation by distance" refers to the fact that distance-dependent gene flow generally limits the genetic differences among natural populations (Slatkin 1993). In the presence of isolation by distance, populations in geographic proximity to each other will be more similar at the genetic level than those that are far away from each other. As illustrated in Figure 4, there was no evidence for isolation by distance (*P*-value = 0.40) in the *O. glaberrima* accessions studied here. The same can be seen in Figure 2C where samples collected in the same country are not always the most closely related genetically. These results underscore the inherent differences between natural and cultivated populations. In natural populations, geographical barriers often limit pollen dispersal, while in populations of cultivated species, both seed and pollen dispersal are often a consequence of human activity and artificial (human) selection tends to reinforce existing groups.

Subpopulation origin and identity: As presented in supplemental Table 1 (at http://www.genetics.org/supplemental/), 67% of the accessions of *O. glaberrima* from 11 countries shared at least 1% of ancestry with *O. sativa*. Eleven of these accessions shared significant (>52%) ancestry with *O. sativa* (hereafter referred to as "interspecific admixed accessions") and were collected from Guinea Conakry, Sierra Leone, and Nigeria.

Of the six accessions collected in Guinea Conakry in this study, four (YS168, YS179, YS230, and YS351) shared at least 87% ancestry with *O. sativa* (Figure 2, B and C). Similarly, 50% of accessions from Sierra Leone (Pa DC Kono, DC Kono, and Saliforeh) shared at least 54% of their ancestry with *O. sativa* cultivars. In contrast, only 11% (7 of 67) of varieties from Nigeria shared significant (12–22%) ancestry with *O. sativa* (save for an exceptional individual, 5486TOG, that shared 96% of ancestry with *O. sativa*—49% with the *indica*-like and 47% with the *japonica*-like group). This is consistent with the fact that Sierra Leone and Guinea Conakry, but not Nigeria, are believed to be the primary ports of entry for *O. sativa* into West Africa.

All the countries of collection have some samples that

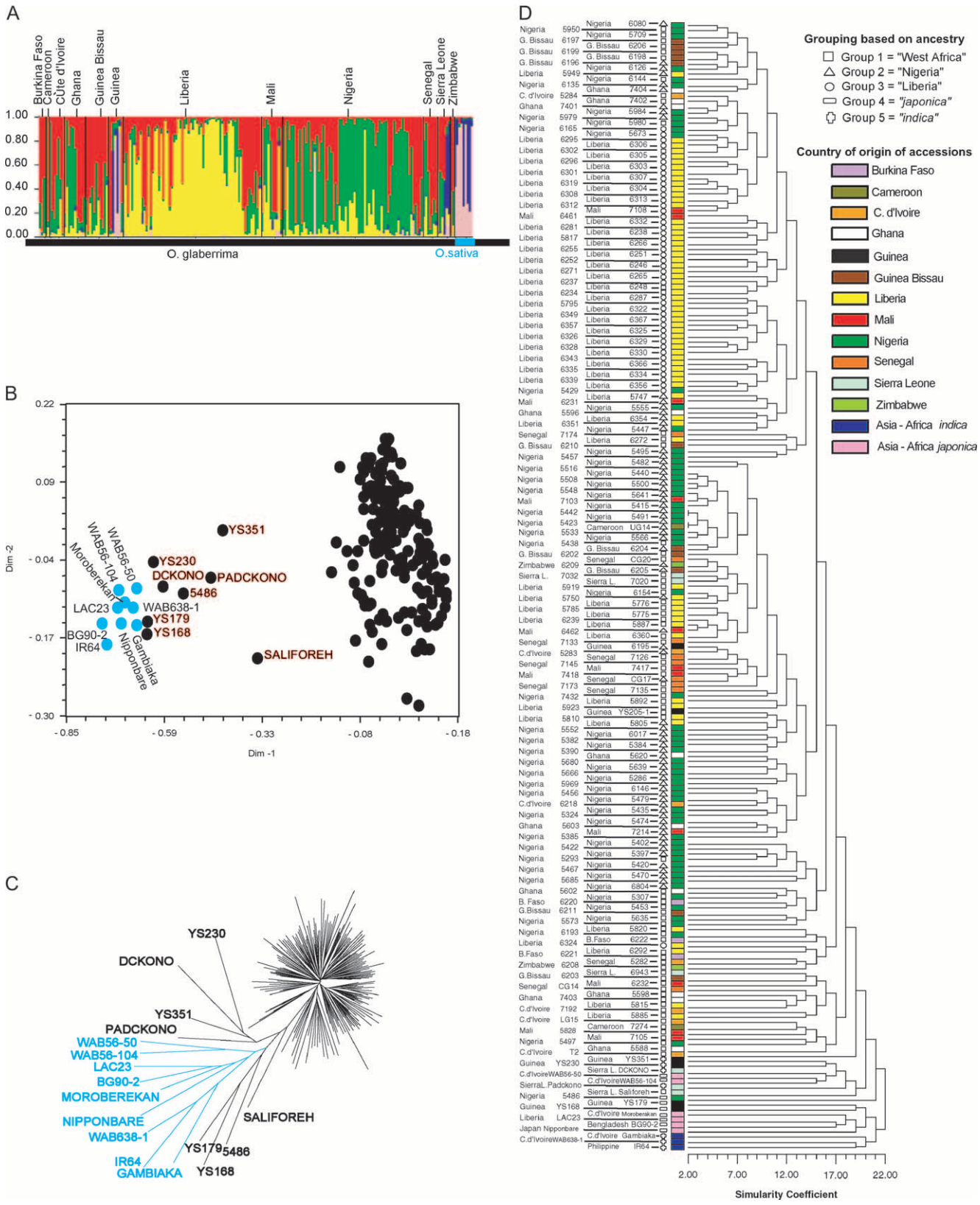

FIGURE 2.—(A) Ancestries of 207 individuals (198 *O. glaberrima* and 9 *O. sativa* accessions) estimated from 93 nuclear SSR loci using STRUCTURE (PRITCHARD *et al.* 2000). Yellow, green, and red refer to the three subpopulations detected in *O. glaberrima*; pink and blue represent ancestries corresponding to the two subspecies of *O. sativa* (*japonica* and *indica*, respectively). (B) Principal coordinate analysis showing partitioning of the 198 *O. glaberrima* accessions into two subpopulations. Light blue circles represent *O. sativa* varieties; black circles represent *O. glaberrima* accessions; *O. glaberrima* varieties clustering with *O. sativa* are named. (C) Unrooted phylogram (based on neighbor joining) representing shared allele frequencies among the 207 accessions; *O. sativa* varieties are shown in light blue; *O. glaberrima* varieties clustering with *O. sativa* are named. (D) Dendrogram (based on neighbor joining) representing shared allele frequencies among 198 *O. glaberrima* and 9 *O. sativa* accessions.

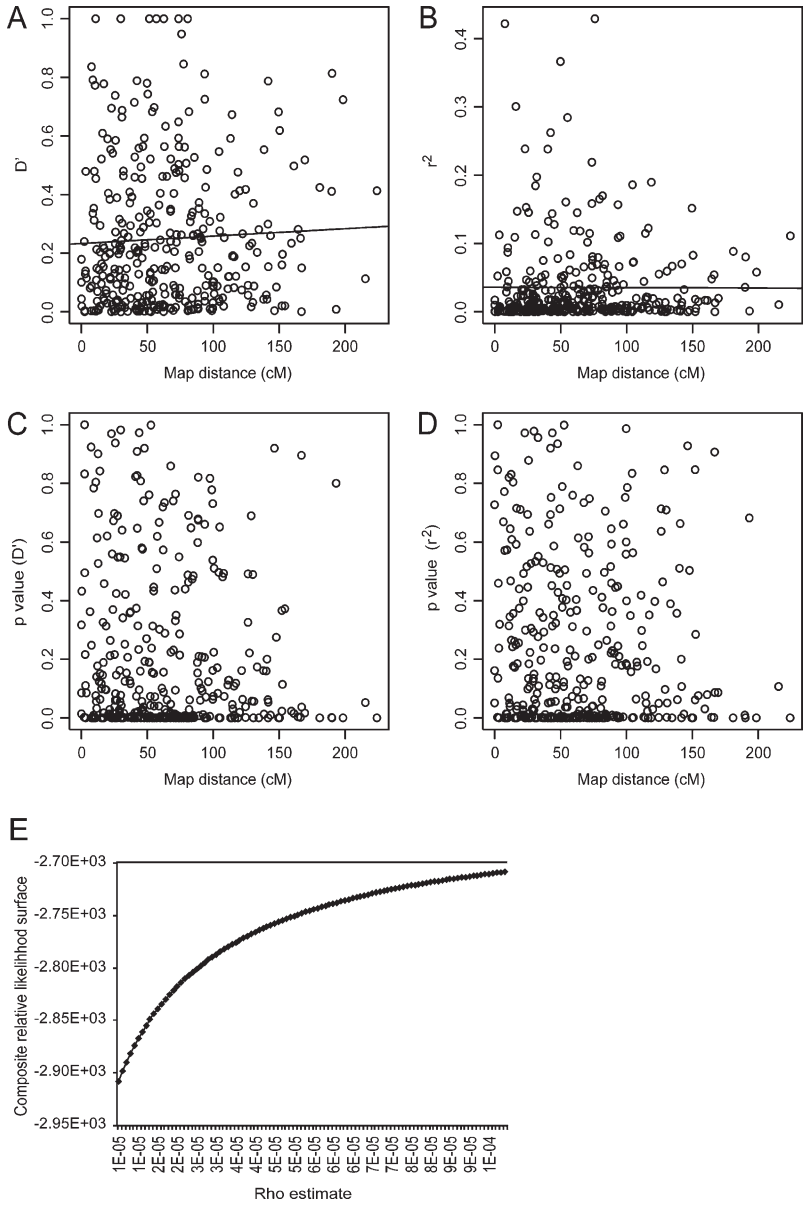

FIGURE 3.—A and B show measures of LD decay (D' and r^2) in a sample of 198 *O. glaberrima*, plotted as a function of the genetic distance between markers. C and D show measures of significance for D' and r^2, plotted as a function of the genetic distance between markers. (E) Composite relative likelihood surface as a function of rho (ρ) based on 198 *O. glaberrima* accessions genotyped with 93 nuclear SSR markers.

share ancestry with *O. sativa*, except those from Zimbabwe (Table 1). Samples from Cameroon and Guinea Bissau share ancestry only with *O. sativa*, ssp. *indica*, while the six *O. glaberrima* accessions from Guinea Conakry and the six from Sierra Leone share an average of 59 and 39%, respectively, of their ancestry with *O. sativa*, ssp. *japonica*. These two groups share only 7 or 3%, respectively, with ssp. *indica*. The same six samples from Guinea Conakry share 31% of their ancestry with *O. glaberrima* group 1 and those from Sierra Leone share 40% ancestry with group 1, compared to 2 or 12%, respectively, with group 2 and 1% or 6%, respectively, with group 3 accessions. This supports the conclusion that many African rice varieties represent admixtures between *O. glaberrima*, group 1, and *O. sativa*.

Genetic differentiation between subpopulations of *O. glaberrima*: The three *O. glaberrima* subpopulations identified by STRUCTURE that do not cluster with *O. sativa* show significant pairwise and overall F_{ST} values ($P < 0.0001$), indicating that *O. glaberrima* subpopulations are significantly differentiated from each other.

Statistics such as the number of alleles, gene diversity, and PIC values offer views into how diversity is partitioned within each group of germplasm. Group 1, the largest and most geographically distributed group of *O. glaberrima* germplasm, embodied more genetic diversity than either group 2 or group 3 (Table 2). In contrast, group 3 had significantly lower PIC and slightly lower gene-diversity values than either group 1 or group 2, suggesting that it may be more homogeneous than the other groups.

The three groups also differ for several population genetic statistics that offer insight into their respective demographic histories. When pairwise F_{ST} values are compared to estimate the degree of differentiation between pairs of subpopulations, it can be concluded that

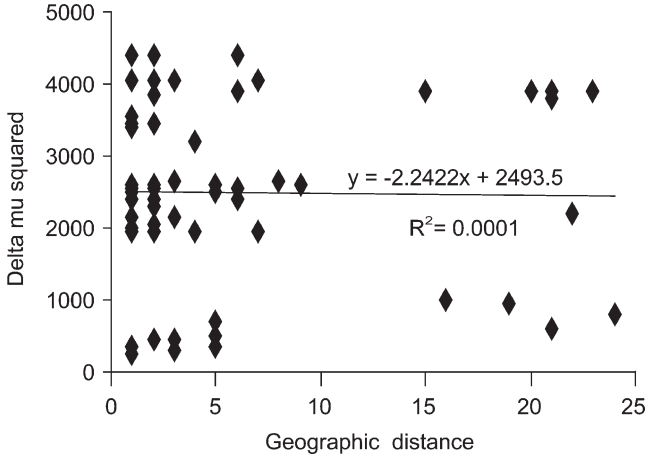

FIGURE 4.—Isolation by distance correlation for the 198 *O. glaberrima*, collected from 12 different African countries.

TABLE 2

Genetic diversity and pairwise differentiation (F_{ST}) among the three groups identified within *O. glaberrima* using a Bayesian analysis (composed of individuals with shared ancestry ≥75%)

	Diversity				Differentiation		
Group	n^a	A^b	He	PIC	1	2	Overall
1	43	4	0.29	0.27	—	—	
2	27	3	0.27	0.25	0.16*	0.18*	
3	34	3	0.22	0.20	0.09*	—	
	—	—			—	—	0.14*

He, heterozygosity; PIC, polymorphism information content. *F_{ST} values are all significant at $P < 0.0001$.

[a] Number of accessions.

[b] Allele number.

group 1 and group 3 are less differentiated from each other than either is from group 2. As summarized in Table 2, only 9% of the variation embodied by groups 1 and 3 is detected as between-group variation while ~16 and 18% of the variation differentiates group 2 from group 1 or group 3, respectively. The remaining portion of the variation represents within-group variation.

Comparison of genetic groups for phenotypic traits: Once the identity of each group or subpopulation had been determined on the basis of genetic markers, it was of interest to determine whether there were any significant differences in phenotypic means for the 18 traits among the three subpopulations of *O. glaberrima*. To address this question, we used analysis of variance followed by a multiple means comparison.

ANOVA showed significant differences among means of the three groups for 14 of 18 phenotypic traits surveyed. Confidence intervals for the differences in the means, computed using a Bonferroni correction for multiple testing of the three-sampled means, indicated differences for the 14 phenotypic traits shown in Table 3. Groups 2 and 3 showed the most extreme phenotypes, with significant differences for 13 of 14 traits. Four traits were particularly informative in distinguishing between pairs of groups, namely flowering time, days to maturity, tiller number and plant height. Group 1 was taller, on average, than either of the other groups; group 2 was particularly late in flowering, heading, and maturity; and group 3 had a larger number of tillers/plant and longer grains (higher grain length/width ratio) than the other two groups. Yield and yield components (spikelets per panicle, panicles per plant, spikelets per plant, and seed set rate) were significantly lower for group 2 individuals compared to group 1 or group 3, as evaluated at M'bé, WARDA headquarters station in Ivory Coast (an upland environment) during August 1998. These results demonstrate that population groups, identified solely on the basis of molecular marker analysis, may also have distinct phenotypic characteristics.

DISCUSSION

The genetic diversity of the WARDA *O. glaberrima* germplasm collection cannot be characterized on the basis of the geographic location where the samples were obtained. We found no evidence of isolation by distance using the measures of geography in this study even though population structure was detected in the species. This feature underscores the important role of humans in the dissemination of *O. glaberrima* throughout West Africa. Human-mediated gene flow occurs rapidly (on an evolutionary scale) and is not linearly related to geographical distance. Furthermore, artificial selection tends to enhance population structure and has undoubtedly contributed to the maintenance of the three subpopulations identified in this study.

Both distance-based and model-based analysis indicated a major subdivision within *O. glaberrima*, with a small group of accessions clustering closely with *O. sativa* control varieties. The introgression of *O. sativa* DNA into *O. glaberrima* germplasm appears to have created intermediate types that cannot be easily distinguished at the phenotypic level from native cultivars of *O. glaberrima*. Nevertheless, it has given rise to significant population structure within *O. glaberrima* and has contributed in important ways to the differentiation of *O. glaberrima* subgroups. Further, many (67%) accessions of *O. glaberrima* carry genetic evidence of some level of admixture with *O. sativa*. This could be accounted for by the introduction of *O. sativa* into West Africa between the 15th and 17th centuries, possibly by Arab traders and surely by Portuguese navigators off the Atlantic Coast. This genetic profile of *O. glaberrima* is also entirely consistent with the cultural history of rice cultivation patterns in West Africa where *O. glaberrima* is often grown in mixtures with *O. sativa*. In addition to the diagnosis of acces-

TABLE 3

Mean trait differences among pairs of subpopulations detected in *O. glaberrima*

				Pairwise group comparisons using Bonferroni correction (95% confidence interval)					
	Means for group			1 and 2		1 and 3		2 and 3	
Traits	1	2	3	L	R	L	R	L	R
Days to heading	102.64	116.34	102.36	*6.31*	*21.09*[a]	−7.56	7.00	*−21.61*	*−6.35*[a]
Days to maturity	135.49	143.76	130.17	−1.46	18.00	−14.90	4.26	*−23.57*	*−3.61*[a]
Tiller number	21.15	23.06	26.34	−1.03	4.85	*2.32*	*8.06*[a]	*0.25*	*6.31*[a]
Plant height	125.35	113.91	107.39	−23.87	0.99	*−30.29*	*−5.63*[a]	−19.44	6.40
Panicle no./sq. meter	247.39	211.5	348.36	−74.01	2.23	*63.15*	*138.79*[a]	*98.74*	*174.98*[a]
Spikelet no./panicle	120.78	95.25	118.20	*−40.55*	*−10.51*[a]	−17.36	12.20	*7.45*	*38.45*[a]
Panicle no./plant	17.82	13.08	21.76	−10.69	1.21	−1.91	9.79	*2.55*	*14.81*[a]
Spikelet no./plant	1157.90	771.40	1514.50	*−596.64*	*−176.36*[a]	*149.74*	*563.46*[a]	*527.53*	*958.67*[a]
Grain width	2.74	2.73	2.46	−0.15	0.15	*−0.42*	*−0.13*[a]	*−0.43*	*−0.12*[a]
Grain length/width ratio	3.04	3.03	3.36	−0.20	0.19	*0.13*	*0.51*[a]	*0.13*	*0.53*[a]
Grain no./panicle	85.79	37.09	90.25	*−64.78*	*−32.62*[a]	−11.23	20.15	*36.58*	*69.74*[a]
Grain no./plant	789.00	292.60	1130.60	*−699.81*	*−292.99*[a]	*141.37*	*541.83*[a]	*629.33*	*1046.67*[a]
Yield/plant	16.26	7.33	21.64	*−13.30*	*−4.55*[a]	*1.08*	*9.69*[a]	*9.79*	*18.83*[a]
Seed set rate	0.68	0.38	0.77	*−0.42*	*−0.20*[a]	−0.03	0.19	*0.27*	*0.51*[a]
1000 grain wt.	21.54	24.25	21.04	ANOVA test not significant					
No. of branches	13.30	12.27	12.65	ANOVA test not significant					
Grain length	8.17	8.25	8.24	ANOVA test not significant					
Panicle length	23.49	25.03	22.82	ANOVA test not significant					

Italic indicates significant mean difference ($P < 0.05$) between pairs of subpopulations. L, left side of confidence interval; R, right side of confidence interval.

[a] Intervals not including 0 indicate that pairwise mean differences are significant for the morphological traits.

sions reflecting *O. sativa*-*O. glaberrima* admixture, model-based analysis also identified three additional, genetically distinct subpopulations that were specific to *O. glaberrima*. One of the three subpopulations was found widely distributed throughout West Africa, while the other two were each predominantly found in Nigeria or Liberia, respectively. Nigeria and Liberia are geographically isolated enough from each other to be considered as divergent demes.

Portères (1970) proposed the hypothesis that the earliest domestication of *O. glaberrima* was achieved under floodwater cultivation, as is still practiced in the inland delta of the Niger River. This primary form of "floating" rice cultivation was stable and localized in riparian and lucustrine situations that characterize the upper Gambia and Senegal. Knowledge of this form of rice culture is believed to have spread from farmer to farmer by way of the valleys. In the secondary centers, "nonfloating" races were selected and techniques of cultivating rice in brackish water evolved. At the same time the dispersal of cultivated forms to other areas, including the Guinea highlands, led to the selection of "upland" races adapted to dry cultivation, mainly by swidden methods. Along with the development of these cultivation techniques, nonfloating and upland races were selected, making possible the large-scale spread of the crop into what is now the rice zone of West Africa (Harris 1976). The three recognizable genetic groups within *O. glaberrima* that cannot be explained by *O. sativa* introgression may correspond to the floating, nonfloating, and upland types reported by Portères (1970).

The *O. glaberrima* populations found predominantly in Nigeria (group 2) or Liberia (group 3) demonstrated slightly lower levels of genetic diversity and a more constrained geographic distribution than the population found throughout West Africa (group 1). Together, these observations are consistent with the hypothesis that groups 2 and 3 may be derivatives of the more widely dispersed group 1 and may represent ecologically specialized subgroups. If this is the case, the subpopulation found throughout West Africa (He = 0.29) may correspond to the ancestral floating type described by Portères (1970). Group 2, the subpopulation that is highly represented in Nigeria (He = 0.27) is likely to represent the nonfloating type, arriving in Nigeria via the Niger River, where it is widely cultivated along the shore. Finally, group 3, the subpopulation found in high frequency in Liberia (He = 0.22) may correspond to the upland rice type. Phenotypically, the floating types might be expected to exhibit taller height as compared to the nonfloating and upland types, which is consistent with our data.

The possibility that the subpopulation structure identified with SSRs corresponds to the distinct ecotypes that have evolved over the history of rice cultivation in West Africa is an intriguing hypothesis. In this study, all

the germplasm was evaluated only under upland conditions (with overhead irrigation) in a single location. In future work, we aim to develop a more comprehensive view of the phenotypic plasticity and adaptive potential of this germplasm to further test this hypothesis.

The two groups of *O. glaberrima* that cluster with *O. sativa*, subspecies *indica* and *japonica* accessions, originated from Guinea Conakry, Sierra Leone, and Nigeria. This result is consistent with reports suggesting that in West Africa, Asian rice was introduced into the rice zone from two principal areas of entry, one being between the Casamance and Cacheao Rivers along the present boundary between Senegal and Guinea Bissau and the other being in Sierra Leone and adjacent parts of Guinea Conakry and Liberia (Portères 1970). In this study, 67% of accessions from Guinea Conakry and 50% of accessions from Sierra Leone shared ancestry with *O. sativa* cultivars. In contrast, only 1 of 67 accessions (1.5%) from Nigeria (5486TOG) grouped with *O. sativa*, suggesting that it might have been introduced by traders or exchanged by farmers, but that Nigeria was not a major port of entry of *O. sativa* into West Africa.

We conclude that the high levels of LD observed in *O. glaberrima* germplasm are primarily caused by population structure driven by introgression with *O. sativa* and the differentiation of ecotypes adapted to different growing environments in West Africa.

We thank the Rockefeller Foundation for a Ph.D. and postdoctoral fellowship for Mande Semon and the Japan International Research Center for Agricultural Science for financial support for the phenotypic evaluation done in M'be, Ivory Coast; Rasmus Nielsen was supported by National Science Foundation grant DEB-0089487. We gratefully acknowledge Ed Buckler and Steve Kresovich for critical reading of this article prior to submission and Lois Swales for help with formatting of this manuscript.

LITERATURE CITED

Chen, X., S. Temnykh, Y. Xu, X. G. Cho and S. R. McCouch, 1997 Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor. Appl. Genet. **95:** 553–567.

Cho, Y. G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich *et al.*, 2000 Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. **100:** 713–722.

Coburn, J. R., S. V. Temnykh, E. M. Paul and S. R. McCouch, 2002 Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.). Crop Sci. **42:** 2092–2099.

Dellaporta, S. L., H. Wood and J. B. Hicks, 1983 A plant DNA minipreparation. Plant Mol. Biol. Rep. **1:** 19–21.

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

Goldstein, D. B., and D. D. Pollock, 1997 Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. J. Hered. **88:** 335–342.

Harris, D., 1976 *Origins of African Plant Domestication*, pp. 331–341. Mouton, Paris.

Hill, W., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159:** 1805–1817.

Jones, M. P., M. Dingkuhn, G. K. Aluko and M. Semon, 1997 Interspecific *O. sativa* L. × *O. glaberrima* Steud: progenies in upland rice improvement. Euphytica **92:** 237–246.

Lewontin, R. C., 1964 The interaction of selection and linkage. II. Optimum models. Genetics **50:** 757–782.

Panaud, O., X. Chen and S. R. McCouch, 1996 Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (Oryza sativa L.). Mol. Gen. Genet. **252:** 597–607.

Portères, R., 1970 Primary cradles of agriculture in the African continent, pp. 43–58 in *Papers in African Prehistory*, edited by J. Fage and R. Olivier. Cambridge University Press, Cambridge, UK.

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

Rohlf, F., 1997 *NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System.* Applied Biostatistics, Setauket, NY.

Saitou, N., 1987 Neighbor-joining method. Mol. Biol. Evol. **4:** 406–425.

Second, G., 1982 Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Jpn. J. Genet. **57:** 25–57.

Second, G., 1986 Isozymes and phylogenetic relationship in *Oryza*, pp. 27–39 in *Rice Genetics: Proceedings of the International Rice Genetics Symposium*, May 27–31, 1985. International Rice Research Institute, Manila.

Slatkin, M., 1993 Isolation by distance in equilibrium and nonequilibrium populations. Evolution **47:** 264–279.

Temnykh, S., W. D. Park, N. Ayres, S. Cartinhour, N. Hauck *et al.*, 2000 Mapping and genome organization of microsatellite sequences in rice (Oryza sativa L.). Theor. Appl. Genet. **100:** 697–712.

Temnykh, S., G. DeClerck, A. Lukashova, K. Lipovich, S. Cartinhour *et al.*, 2001 Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11:** 1441–1452.

Viguier, P., 1939 *La Riziculture Indigène au Soudan Français.* Larose, Paris.

Wang, Z., G. Second and S. Tanksley, 1992 Polymorphism and phylogenetic relationship among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. **83:** 565–581.

Xiao, J., J. Li, S. Grandillo, S. N. Ahn, L. Yuan *et al.*, 1998 Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics **150:** 899–909.

Communicating editor: M. Nordborg