DISTINGUISHING RICE (*ORYZA SATIVA* POACEAE) FROM WILD *ORYZA* SPECIES THROUGH PHYTOLITH ANALYSIS, II: FINALIZED METHOD[1]

ZHIJUN ZHAO, DEBORAH M. PEARSALL, ROBERT A. BENFER, JR., AND DOLORES R. PIPERNO

Zhao, Zhijun (*Smithsonian Tropical Research Institute, Unit 0948, APO. AA 34002-0948, U.S.A.*), **Deborah M. Pearsall, Robert A. Benfer, Jr.** (*Department of Anthropology, 107 Swallow Hall, University of Missouri, Columbia, MO 65211*), and **Dolores R. Piperno** (*Smithsonian Tropical Research Institute, Unit 0948, APO. AA 34002-0948, U.S.A.*). DISTINGUISHING RICE (*ORYZA SATIVA* POACEAE) FROM WILD *ORYZA* SPECIES THROUGH PHYTOLITH ANALYSIS, II: FINALIZED METHOD. *Economic Botany 52(2) 134–145. 1998. Asian rice is an important grain, not only in its homeland but in many areas of the world. Preliminary studies suggested that phytolith analysis, the identification of opaline silica bodies, provided a reliable way of identifying rice, especially in situations where preservation of charred botanical remains was poor. Results of this follow-up study, which incorporates all Asian wild* Oryza *species and a diverse array of traditional* Oryza sativa *cultivars, confirm that rice can be identified with a high level of certainty by the size and qualitative features of a distincitve phytolith, the double-peaked glume cell.*

DISCERNER ENTRE LE RIZ (*ORYZA SATIVA* POACEAE) ET LES ESPÈCES SAUVAGES D'*ORYZA* PAR L'ANALYZE PHYTOLITHIQUE, II: MÉTHODE FINALE. *Le riz d'Asie est une céréale importante, non seulement dans son pays d'origine, mais à travers le monde. Des études préliminaires suggèrent que l'analyze phytolithique—identification de particules de silica opalisé—fournit un moyen sur pour identifier le riz, surtout dans les cas de mauvais état de conservation des restes brulés organiques. Les résultats de cette étude de suite, incorporant toutes les espèces asiatiques d'*Oryza *sauvage et un groupe divers de formes traditionellement cultivées d'* Oryza sativa, *confirment le fait que le riz peut être identifié avec un haut degrès d'assurance par la taille et les traits qualitatifs d'un phytolithe distinctif, la glume bidentée.*

〈利用植硅石分析方法对稻谷(Oryza sativa)的鉴定：II〉 亚洲稻谷，不仅在亚洲而且在世界许多地区，均为一种重要的谷物。初步的研究（见本刊49(2):183-196）揭示了植硅石分析可为考古发现的稻谷遗存的鉴定提供一种有效的方法，特别是在那些植物遗骸保存条件不好的地区。基于对所有亚洲野生稻品种以及各类传统栽培稻的对比分析，此次进一步的研究结果证实，根据其形态不同，产生于稻类植物颖壳上的双峰型植硅石可将栽培稻与野生稻区分开。

Key Words: *Oryza sativa*; phytoliths; Asia, China; discriminant analysis.

Rice is one of the most important grain crops in the world today, a mainstay of diet not only in its area of origin, eastern Asia, but throughout the tropical latitudes of the New and Old Worlds. Archaeologists have been hampered in their search for the area or areas of origin of rice, however, because organic remains of food plants often preserve poorly in lowland tropical environments. For this reason, we explored the potential of phytolith analysis, the identification of opaline silica bodies, as a tool in the search for early rice domestication (Pearsall et al. 1995).

The preliminary study (Pearsall et al. 1995) identified a number of phytolith types that char-

[1] Received 20 December 1996; accepted 25 October 1997.

acterized the members of the tribe Oryzeae: bilobate short cells with scooped ends, thick crosses, epidermal long cells with numerous small rounded projection, thick epidermal cells with one prominent projection, and thick keystone bulliform cells. Refer to Pearsall et al. (1995) for descriptions and illustrations of these types. We also found that the genus *Oryza* could be distinguished from other genera in the tribe by the presence of seed epidermal cells with large conical hairs (double-peaked glume cells), deeply serrated or sinuous-edged seed epidermal cells, and moderately to highly flared keystone bulliform cells. In regions where *Oryza sativa* is the only member of the genus present, genus-level diagnostics can be used to identify rice, i.e., the bulliform method developed by Fujiwara (1993) or by the presence of double-peaked glume cells, as indicated in our work.

To explore whether rice could be identified in regions where it overlapped with wild rice species, we studied the double-peaked glume cell in detail (Pearsall et al. 1995). By applying multiple linear discriminant analysis to a data set of glume cell measurements from six wild rice species and five domesticated rice specimens, we derived formulae that permitted correct classification of domesticated rice glume cells in 81.6% of the test cases.

Here we present the results of a rigorous test of the discriminant analysis approach to identifying rice developed during our preliminary study that incorporates (1) data from all Asian wild rice species, and (2) a large, diverse group of traditional Chinese rice cultivars. We demonstrate that formulae developed from discriminant analysis permit classification of unknown double-peaked glume cells into three groups—domestic, wild, or indeterminate—with a high degree of accuracy. Qualitative characteristics of the glume cells further aid in the identification of domesticated and wild forms.

Methods and Materials

Recognizing the patterns of variability of diagnostic phytolith types is a key to identifying plant taxa from archaeological soil samples. The ideal diagnostic type is one whose morphological characteristics are absolutely unique for its producer. Many such diagnostic phytoliths exist (Piperno 1988; Pearsall 1989; Rapp and Mulholland 1992; Pearsall and Piperno 1993). However, in some plant families, it is not always easy to find this kind of one-to-one diagnostic type at the species-level. One of the reasons for this difficulty is that the characteristics of some phytoliths may be within a quantitative range overlapping two or more species (Pearsall et al. 1994). However, if the range of characteristics is separable by statistical means, the degree to which phytolith types can be diagnosed can be explored. In practice, these kinds of phytolith types can still be used to distinguish plant species.

Statistical analysis has been applied to phytolith analysis for many purposes, such as interpretations of phytolith assemblages (Powers-Jones and Padmore 1993), separations of plant taxa (Mulholland 1993), and phytolith identification (Pearsall et al. 1995; Pearsall and Piperno 1990). In this study, a multivariate technique, multiple linear discriminant function analysis, is used to develop a species-level diagnostic phytolith for domestic rice.

Discriminant Analysis

Linear discriminant analysis, also known as linear discriminant function analysis, is a statistical procedure for identifying boundaries between groups of objects by analyzing relationships between these objects by quantitative variables (Kachigan 1991). Discriminant analysis is also applied to classify objects into respective groups by analyzing the relationships between variables of the objects and the boundaries defined in terms of these variables.

Since linear discriminant analysis is applied only in cases in which heterogeneity is expected, normal distributions of predictor variables are neither anticipated nor required. In place of tests of significance that require a normal distribution of the residuals of predicted scores, we employ a training-set/test-set criterion for evaluation of the effectiveness of the discriminant models. We construct non-linear models by adding variables that are the squares and products of selected measurements. The final model is evaluated for efficiency by reserving random cases for testing that were not employed in building the model. Since discriminant functions maximize differences among groups (Kachigan 1991; Benfer and Benfer 1981), they are more powerful than weaker cluster analysis programs that lack this desirable feature (see Benfer 1975 for a discussion of the use of cluster analysis of untransformed variables versus those produced by prin-

ciple components, linear combinations of original variables).

In linear discriminant function analysis, new variables, linear combinations of the raw variables, are created such that, as noted above, each new variable has the property of minimizing variability within groups while maximizing variability among groups. A critical or cutoff score is that value that divides membership in one group from the other in the simple two group design presented here. Each phytolith receives a score calculated as that linear combination of scores that best distinguishes domestic rice from wild rice. Below we discuss a strategy that varies the criterion for prediction of group membership in order to make for more secure identification of domestic than of wild rice.

Classification by discriminant analysis groups the discriminant scores of objects by the cutoff score. Objects with discriminant scores greater than the cutoff score are assigned to one group, and those less than the cutoff score to the other group in a simple, two-group example. Multiple discrimiant function analysis can separate members of three or more groups.

Discriminant analysis deals with three aspects of data: groups, objects and variables. In this study, species of *Oryza* can be seen as groups, the double-peaked glume cell is the object, and measurable characteristics of this phytolith type are the variables.

Specimens

A total of 106 comparative specimens of *Oryza*, supplied by the International Rice Research Institute (IRRI), were studied. IRRI numbered specimens are curated at the MU Phytolith laboratory; researchers interested in consulting the collection should contact Pearsall (anthdp@showme.missouri.edu). These specimens consist of 27 accessions of domestic rice, originated from China, and 79 specimens of nine wild rice species distributed geographically in South and Southeast Asia: *O. granulata* Nees et Arn. (3 specimens), *O. longiglumis* Jansen (5), *O. meyeriana* Baill. (4), *O. ridleyi* Hook.f. (9), *O. minuta* J. S. Presl (9), *O. officinalis* Wall. (16), *O. nivara* Sharma et Shastry (15), *O. rufipogon* Griff. (11), and *O. sativa* var. *spontanea* (7). *O. nivara* and *O. rufipogon* are the Asian members of the "AA" genome group that includes domesticated rice (Chang 1976); these wild species are considered ancestral to rice by most scholars.

Procedures

The double peaked glume cell is a three-dimensional geometric body; the most common orientation observed in slide-mounted phytolith extracts is a trapezoidal side view with a concavity on the top that gives it the "two-peaked" shape (Fig. 1). Because the glume cells are very difficult to rotate on the microscope slide, the trapezoidal side view was used for all measurements.

Three measurements, top width, bottom width, and height, are not sufficient to determine the shape and size of the glume cells, because some showed a significant difference between the heights of two peaks as well as the depth of curve between two peaks. Thus five measurements (Fig. 2) were taken for each individual glume cell using an eyepiece micrometer: width of the top (TW, the distance between the two peaks), width of the middle (MW, the width at the point where the hair attaches to the base), the height of each peak (H1 and H2, measured from the tip to the base of the hair; H2 is the smaller measurement), and the depth of the curve (CD, measured from the tip of H1 to the lowest point of the curve). It might be possible to obtain other useful measurements by computer image analysis (i.e., Ball et al. 1996); we opted to focus on measurements that could be taken rapidly at the microscope.

Phytoliths were extracted by chemical oxidation and slide-mounted in Canada Balsam following procedures outlined in Pearsall (1989). To minimize bias, all specimens were assigned lab numbers before extraction; only these numbers were used to label extracts and slides. Each slide was scanned by Zhao until 25 individual double peaked glume cells were encountered. Specimens were identified to species after all slides were completed. Double peaked glume cells were found to be abundant in *O. sativa*, *O. minuta*, *O. officinalis*, *O. nivara*, *O. rufipogon* and *O. spontanea*. However, *O. granulata*, *O. longiglumis*, *O. meyeriana* and *O. ridleyi* proved to be poor producers of double peaked glume cells; it was not always possible to obtain measurements on 25 bodies. A total of 2359 double peaked glume cells were measured from 106 specimens, including 663 bodies from *O. sativa* and 1696 bodies from the wild species.

A statistical package for Macintosh computer, SYSTAT, was used for the discriminant analysis.

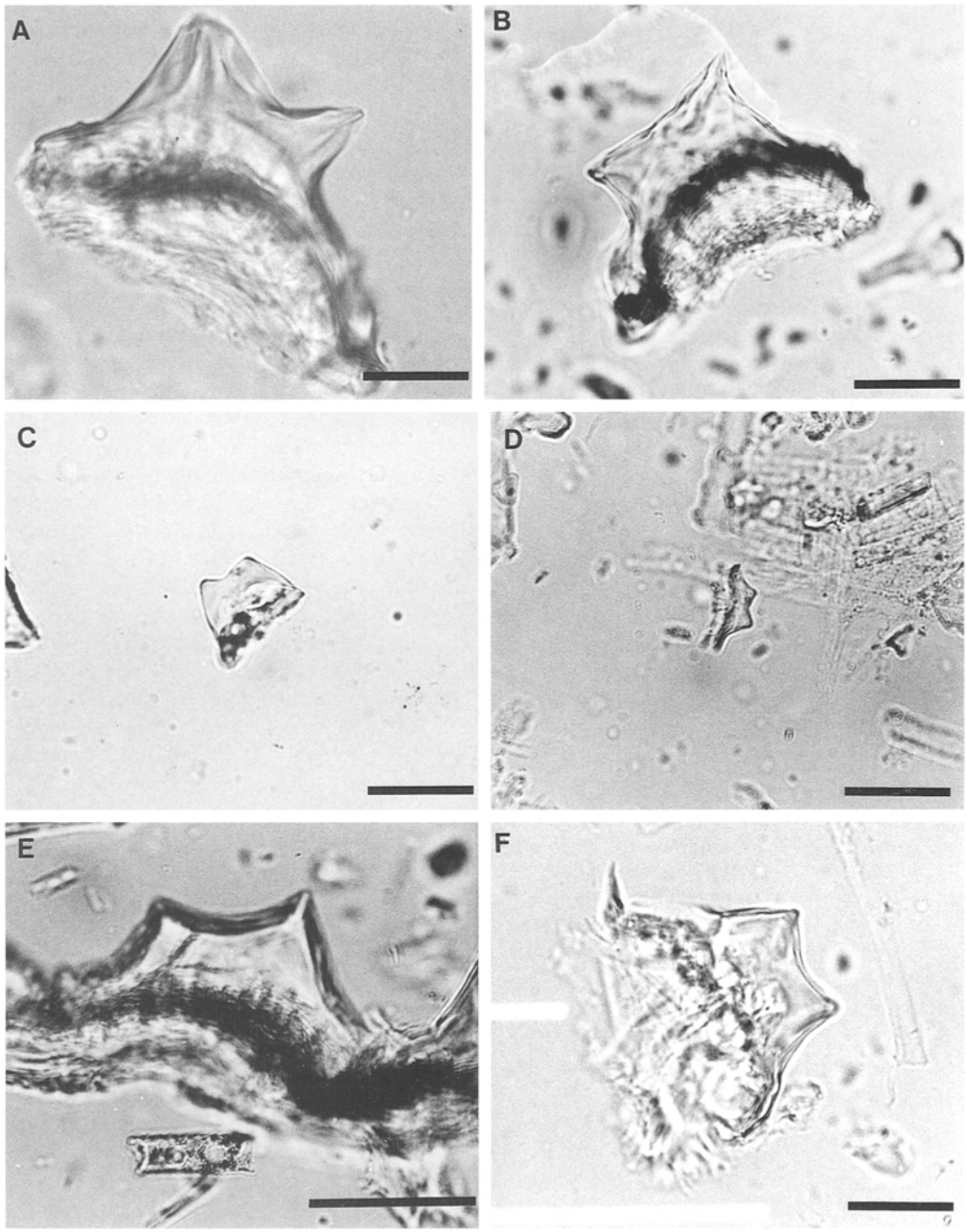

Fig. 1. Double-peaked glume cells: (A) *Oryza minuta*; (B) *O. officinalis*; (C) *O. granulata*; (D) *O. ridleyi*; (E) *O. nivara*; (F) *O. sativa*. Scale bar = 25 micrometers.

In the data matrix, five measurements from each of 2359 glume bodies are used as predictor variables, and 106 specimens are the objects in which all wild species made up one group and the domesticated rice specimens the other.

Results of Discriminant Analysis

Initial Results

The results of the first discriminant analysis of the data are given in a frequency table (Table

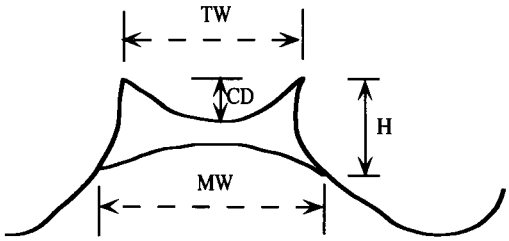

Fig. 2. Double-peaked glume cell measurements.

TABLE 1. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS (FIVE VARIABLES).

	Predicted		
Actual	Domestic	Wild	Total
Domestic	522	141	663
Wild	486	1210	1696
Total	1008	1351	2359

1) and as a percentage of correct classification (Table 2). The results are that 79% of the glume cells from domestic rice are correctly classified in the predicted domestic group, and 71% of the glume cells from wild species are correctly classified in the predicted wild group. These encouraging results indicate that the characters of double peaked glume cells overlap but have a strong tendency of polarization between domestic rice and wild rice species, as was the case for the preliminary study (Pearsall et al. 1995). This kind of overlapping range is separable by an approach described below.

From the above results, nearly 29% of the glume cells from wild species were misclassified as domesticated. A degree of error was expected based on the preliminary study, but this is too high to make a satisfactory assignment in diagnosing domesticated rice. According to the low score of canonical correlations between two groups (0.43), the number of variables in the data set may be not sufficient, as canonical correlations represent the correlation between two sets of derived weighted combination of variables. Canonical correlations must increase with the number of variables, so we next evaluate the magnitude of improvement possible by adding additional ones.

RESULTS BY ADDING MORE VARIABLES

In order to achieve a satisfactory assignment from discriminant analysis, there must be enough variables to provide sufficient information about the separation of two groups (Lachenbruch 1975; Benfer and Benfer 1981). Double peaked glume cells are difficult to rotate on slides and, practically speaking, it is difficult to obtain more measurements. However, new independent variables can be made on the same objects by transforming original measurements, such as using the square of a measurement. Although ratios, non-linear combinations of variables, may achieve successful prediction rates in a linear discriminant function approach, we find their use confusing as they have often been wrongly assumed to remove the effect of one measurement while in fact the correlation usually persists (Benfer and Benfer 1981).

In order to examine whether a non-linear model would improve predictive efficiency, we added squares and products of the original five variables, namely TW^2, MW^2, $H1^2$, $H2^2$, CD^2, MW*TW, and H1*H2. We also included a size variable to possibly improve interpretation, TW+MW+H1. Because there are such a large number of combinations of these 13 variables (the above 8 variables added to the original 5), we did not compute all possible discriminant functions in order to find the best one. The sample size is too small to avoid chance playing a large role in the results. Instead, we employed the intuition of the investigators to consider several selected sets. One, which added $H1^2$, $H2^2$ and CD^2 to the original 5 variables, did increase the canonical correlation substantially, from 0.43 to 0.52. This improvement was reflected by an increase of degree of correct classification of domestic rice (from 79% to 86%), however, the degree of error from wild species worsened (Tables 3 and 4). Therefore, the method of adding more variables provided some help but did not solve the problem central to the study, i.e., minimizing the error from wild species.

TABLE 2. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS (FIVE VARIABLES).

	Predicted		
Actual	Domestic	Wild	Total
Domestic	78.73	21.27	28.11
Wild	28.66	71.34	71.89
Total	42.73	57.27	100

TABLE 3. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS (EIGHT VARIABLES).

	Predicted		
Actual	Domestic	Wild	Total
Domestic	572	91	663
Wild	508	1188	1696
Total	1080	1279	2359

COST OF ERRORS AND PRIOR PROBABILITIES

The results of discriminant analysis minimize the number of classification errors, without regard to the cost of error for each group. But, depending on the purpose of a study, the costs of error from different groups are usually not equal. For example, in the case where the emergence of rice domestication is the issue, the mistake of identifying wild rice as domestic would be more serious than identifying domesticated rice as wild. The discriminant analysis procedure can be modified to take into account the differential costs of alternative errors (Kachigan 1991: 231). The more costly error can be minimized by putting more weight on the less costly error, in order to produce a satisfactory assignment. This adaptation can be incorporated into the analysis using the Bayesian theorem (Mardia et al. 1979). According to the Bayesian theorem, one commonly has certain prior knowledge about one's data. If knowledge, *prior probabilities*, in statistical terms, can be modified in light of the data, more satisfactory posterior probabilities can be obtained from discriminant analysis (Lee 1989). The cutoff score obtained from a discriminant analysis can be altered by assigning different weights of prior probabilities.

In reality, it makes sense to suppose that various populations have different prior probabilities. To search for the emergence of domesticated rice, for example, one may expect that wild rice is intrinsically more likely to be found archaeologically in the early transition period of rice domestication, based on the hypothesis that wild rice collection preceded rice domestication (Zhao 1996). To transform this prior knowledge into discriminant analysis, therefore, more prior probabilities should be given to the group of wild rice and less prior probabilities to the group of domestic rice (total prior probabilities must sum to 1).

TABLE 4. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS (EIGHT VARIABLES).

	Predicted		
Actual	Domestic	Wild	Total
Domestic	86.27	13.73	28.11
Wild	29.95	70.05	71.89
Total	45.78	54.22	100

TABLE 5. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS BY ADJUSTING PRIOR PROBABILITIES OF 0.28 FOR DOMESTIC AND 0.72 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	345	318	663
Wild	161	1535	1696
Total	506	1853	2359

RESULTS BY ADJUSTING PRIOR PROBABILITIES

To minimize the error from wild rice, several tests were carried out by giving more prior probabilities to the wild rice group and correspondingly less prior probabilities to the domestic group, such as 0.6 to wild group vs. 0.4 to domestic group, 0.7 to wild group vs. 0.3 to domestic group, 0.8 to wild group vs. 0.2 to domestic group, and so on. Adjusting prior probabilities of 0.28 for the predicted domestic group and 0.72 for the predicted wild group produced the most satisfactory assignment.

By adjusting prior probabilities by 0.28 and 0.72, the degree of error from wild species sharply dropped from 30% to 9% (Tables 5 and 6). As was expected, more domestic rice bodies were consequently misclassified as wild. However, those predicted as domestic rice were commonly correctly predicted, because discriminant analysis did succeed in identifying 90% of wild

TABLE 6. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS BY ADJUSTING PRIOR PROBABILITIES OF 0.28 FOR DOMESTIC AND 0.72 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	52.04	47.96	28.11
Wild	9.49	90.51	71.89
Total	21.45	78.55	100

TABLE 7. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS BY ADJUSTING PRIOR PROBABILITIES OF 0.56 FOR DOMESTIC AND 0.44 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	597	66	663
Wild	600	1096	1696
Total	1197	1162	2359

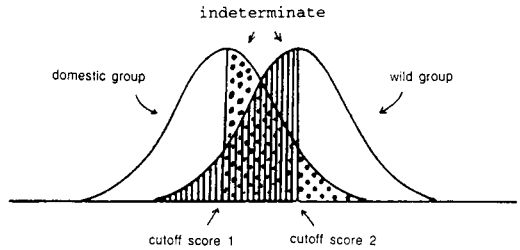

Fig. 3. Conceptual diagram of three groups produced by two alternated cutoff scores.

rice species. The significance of this result is that some domestic rice glume cells may be identified as wild in practice, but those identified as domestic rice (more than 50% of total) are identified with more confidence. Note that a higher precision of identification is the goal in this study.

The same method can also be used to achieve a satisfactory assignment for the case of identifying wild rice from domestic rice. This time, larger prior probabilities are given to the domestic group in order to minimize the error from domestic rice. Discriminant function was calculated by adjusting several pairs of prior probabilities; adjusting prior probabilities by 0.56 for domestic and 0.44 for wild gave the most satisfactory assignment, i.e., the degree of error from domestic rice decreased to 10%, associated with a slight increase of misclassified wild rice species (Tables 7 and 8). This suggests that those predicted as wild-rice glume cells (65%) are commonly correctly predicted.

INDETERMINATE-GROUP DESIGN

The method of adjusting prior probabilities successfully provided two satisfactory assignments, one for identifying domestic rice and the other for identifying wild rice species. Consequently, two alternative cutoff scores were produced, one for minimizing the error from wild rice species and the other for minimizing the error from domestic rice. By combining these two cutoff scores into the same data set, the data are eventually divided into three predicted groups instead of two (Fig. 3). The first group is a predicted domestic group including commonly correctly classified domestic rice with a few misclassified wild species. The second group is a predicted wild group including commonly correctly classified wild species with a few misclassified domestic rice. The third group is an indeterminate group which consists of most misclassified cases from both domestic rice and wild rice species.

By combining the results from Table 6 (discrimination by the first cutoff score) and Table 8 (discrimination by the second cutoff score), the percentages of glume cells in the three predicted groups can be established (Table 9).

For the purpose of prediction, therefore, a group of unknown glume cells will be predicted into three groups, those predicted in the domestic group are more likely to be from domestic rice, those predicted in the wild group are more likely to be from wild rice species, and those predicted in the indeterminate group are unidentifiable. As the result, some glume cells recovered from archaeological samples will fail to be identified because they fall in the indeterminate group, but those that can be identified can be done so with confidence, because the indeterminate group includes most of the ambiguous cases. Therefore, the design of three groups pro-

TABLE 8. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS BY ADJUSTING PRIOR PROBABILITIES OF 0.56 FOR DOMESTIC AND 0.44 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	90.05	9.95	28.11
Wild	35.38	64.62	71.89
Total	50.74	49.26	100

TABLE 9. PERCENTAGES OF GLUME CELLS IN THREE PREDICTED GROUPS.

	Predicted			
Actual	Domestic	Indeterminate	Wild	Total
Domestic	52.04	38.01	9.95	100
Wild	9.49	25.89	64.62	100

TABLE 10. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS OF TRAINING SET BY ADJUSTING PRIOR PROBABILITIES OF 0.28 FOR DOMESTIC AND 0.72 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	271	262	533
Wild	132	1222	1354
Total	403	1484	1887

TABLE 12. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS OF TEST SET BY ADJUSTING PRIOR PROBABILITIES OF 0.28 FOR DOMESTIC AND 0.72 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	67	63	130
Wild	34	308	342
Total	101	371	472

vides a practical solution to the use of double-peaked glume cell as a diagnostic phytolith in areas of overlap between domesticated rice and related wild *Oryza* species.

VALIDATION

For the validation test, the original data set was randomly reassigned into two sets of data, a training set and a test set. The training set contains 80% of the original data and the rest (20%) belongs to the test set. The training set was analyzed by discriminant analysis in the same way as the original data set. Then the predictor formulas derived from the training set were used to calculate the predictions of the data in the test set.

The first test of validation was carried out against the results obtained from the discriminant analysis that was run by adjusting prior probabilities of 0.28 (domestic) and 0.72 (wild). The training set was run again with a discriminant analysis based on eight variables after adjusting prior probabilities to 0.28 for the domestic group and 0.72 for the wild group. The training set results (Tables 10 and 11) show almost no difference from those obtained from the analysis of the original data set.

Two prediction formulas were derived by using group classification function coefficients and group classification constants obtained from the training set. The formulas were then used to calculate the probability of correct classification with the test set. Comparisons of predicted with actual are shown in Table 12 and Table 13. The results of the test set show fewer than 10% errors in correct classification of wild rice, a rate only fractionally higher than observed with the full data set. This indicates that the results of the original discriminant analysis are reliable.

Using the same method, the test of validation was also carried out against the results obtained from the discriminant analysis which was run by adjusting prior probabilities to 0.56 (domestic) vs. 0.44 (wild). The results from the training set are shown in Table 14 and Table 15. Almost no difference exists between these results and the original data. Two prediction formulas were also obtained from this training set. These formulas were then used to calculate the probability of correct classification with the test set. The results are shown in Table 16 and Table 17. Once again, the results of the test set show no significant departure from the original data.

APPLYING THE RESULTS

To apply the results of discriminant analysis, prediction formulas are first derived by using the group classification function coefficients and

TABLE 11. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS OF TRAINING SET BY ADJUSTING PRIOR PROBABILITIES OF 0.28 FOR DOMESTIC AND 0.72 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	50.84	49.16	28.25
Wild	9.75	90.25	71.75
Total	21.36	78.64	100

TABLE 13. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS OF TEST SET BY ADJUSTING PRIOR PROBABILITIES OF 0.28 FOR DOMESTIC AND 0.72 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	51.54	48.46	27.54
Wild	9.94	90.06	72.46
Total	21.40	78.60	100

TABLE 14. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS OF TRAINING SET BY ADJUSTING PRIOR PROBABILITIES OF 0.56 FOR DOMESTIC AND 0.44 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	481	52	533
Wild	494	860	1354
Total	975	912	1887

TABLE 16. FREQUENCIES OF GLUME CELLS IN PREDICTED GROUPS OF TEST SET BY ADJUSTING PRIOR PROBABILITIES OF 0.56 FOR DOMESTIC AND 0.44 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	119	11	130
Wild	139	203	342
Total	258	214	472

group classification constants obtained from the discriminant analysis. These formulas can then be used to predict unknown (i.e., archaeological) glume cells.

The first set of formulas is derived from the discriminant analysis run to obtain a satisfactory assignment for identifying domestic rice, i.e., the one which was adjusted by prior probabilities to 0.28 for the domestic group and 0.72 for the wild group.

(1) *Formulas for Maximizing Accuracy of Domesticated Rice Prediction:*

Prediction of domestic rice

$$= -19.027 - 0.129(TW) + 0.116(MW) + 0.676(H1) + 3.101(H2) + 0.921(CD) - 0.028(H1^2) - 0.079(H2^2) - 0.047(CD^2)$$

Prediction of wild rice

$$= -14.124 - 0.085(TW) + 0.113(MW) + 0.7(H1) + 2.288(H2) + 1.338(CD) - 0.021(H1^2) - 0.066(H2^2) - 0.067(CD^2)$$

The formulas are used as follows: first, measure a double-peaked glume cell recovered from an archaeological sample, insert the measurements into each formula, and compare the two scores. If the score from the prediction of domestic rice is larger in absolute value than that from the prediction of wild rice, this particular glume cell is likely from domestic rice. If the score from the prediction of wild rice is larger, the origin of this particular glume cell remains unclear.

The second set of formulas is derived from the discriminant analysis run to obtain a satisfactory assignment for identifying wild rice species, i.e., the one which was adjusted by prior probabilities of 0.56 for the domestic group and 0.44 for the wild group.

(2) *Formulas for Maximizing Accuracy of Wild Rice Prediction:*

Prediction of wild rice

$$= -14.617 - 0.085(TW) + 0.113(MW) + 0.7(H1) + 2.288(H2) + 1.338(CD) - 0.021(H1^2) - 0.066(H2^2) - 0.067(CD^2)$$

Prediction of domestic rice

$$= -18.334 - 0.129(TW) + 0.116(MW) + 0.676(H1) + 3.101(H2) + 0.921(CD) - 0.028(H1^2) - 0.079(H2^2) - 0.047(CD^2)$$

TABLE 15. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS OF TRAINING SET BY ADJUSTING PRIOR PROBABILITIES OF 0.56 FOR DOMESTIC AND 0.44 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	90.24	9.76	28.25
Wild	36.48	63.52	71.75
Total	51.67	48.33	100

TABLE 17. PERCENTAGES OF GLUME CELLS IN PREDICTED GROUPS OF TEST SET BY ADJUSTING PRIOR PROBABILITIES OF 0.56 FOR DOMESTIC AND 0.44 FOR WILD.

	Predicted		
Actual	Domestic	Wild	Total
Domestic	91.54	8.46	27.54
Wild	40.64	59.36	72.46
Total	54.66	45.34	100

The method of using these formulas is similar to the one described above. If the score from the prediction of wild rice is larger in absolute value than that from the prediction of domestic rice, this particular glume cell is likely from wild rice. If the score from the prediction of domestic rice is larger, the origin of this particular glume cell remains unclear.

In practice, therefore, the measurements of an unknown glume cell should be inserted into both sets of formulas. If it is successfully predicted as domestic rice using the first formulas, and fails to be predicted as wild rice using second formulas, it is likely to be from domestic rice. If results are the reverse, it is likely to be from wild rice species. However, if it is predicted neither as domestic rice by the first formulas nor as wild rice species by the second formulas, this particular glume cell belongs to the indeterminate group, i.e., it is not identifiable.

The indeterminate-group design provides a high probability for correct identification. In order to improve further the ability to identify domestic rice, obviously more than one glume cell should be measured. If the number of individual glume cells from a given archaeological sample predicted in one group is significant larger than in the other groups, the probability of correct classification should be very high. For example, in the case where ten glume cells are found from a soil sample, if nine of them are predicted in the domestic group, and one in the indeterminate group or in the wild group, a joint probability can be calculated that suggests that domestic rice was present in the soil sample (Pearsall et al. 1995).

Quantitative methods for identifying diagnostic phytoliths, such as the discriminant analysis approach for identifying double-peaked glume cells of domestic rice just described, are usefully supplemented with qualitative ones.

Qualitative Features of Double-peaked Glume Cells

As a result of undertaking this intensive study of *Oryza* double-peaked glume cells, Zhao, who scanned all the slides, became very familiar with these phytoliths. This familiarity permits us to discuss how double-peaked glume cells vary among groups of rice species. These qualitative features, expressed as central tendencies (Table 18), may aid the analyst in identifying the origin of individual glume bodies.

TABLE 18. CENTRAL TENDENCIES OF DOUBLE PEAKED GLUME CELLS IN GROUPS OF *ORYZA* SPECIES. REFER TO FIG. 1 FOR EXAMPLES.

(1) O. sativa, O. nivara, and O. spontanea:
surface of hair: smooth
drop of curve between peaks: shallow, but present
symmetry: very symmetrical (little to no difference in heights of the two hairs; little to no difference in slopes of the outer sides of the hairs)
variation among glume cells: low within each species

(1-1) O. nivara:
size of glume cell: tends to be smaller than *O. sativa*
surface of hair: tends to be smoother than *O. sativa*, almost "shiney"
drop of curve between peaks: tends to be deeper than *O. sativa*

(2) All other species:
surface of hair: rough in appearance
drop of curve between peaks: deep or absent
symmetry: tend to be asymmetrical (hairs tend to be different heights; slopes of the outer sides of the hairs often differ; there may be little slope [straight or even incurving sides])
variation among glume cells: high within each species

(2-1) O. granulata, O. meyeriana, O. ridleyi, O. longiglumis:
size of glume cells: very small in comparison to group 1 and the rest of group 2 (*O. rufipogon, O. officinalis, O. minuta*).

What these central tendencies suggest is that most of error in distinguishing double-peaked glume cells of domestic rice from wild rice comes from *O. nivara*, the annual wild species most closely related to rice, and "weedy" rice, *O. spontanea* (Fig. 4). Interestingly, the glume cells of *O. rufipogon*, the perennial wild species most closely related to rice, are much easier to discriminate. The other wild species contribute much less error. While we did not study *O. australiensis* (Australia, "EE" genome) or *O. schlechteri* (New Guinea), since the focus of our work was China, it is unlikely either of these species would complicate identification of rice

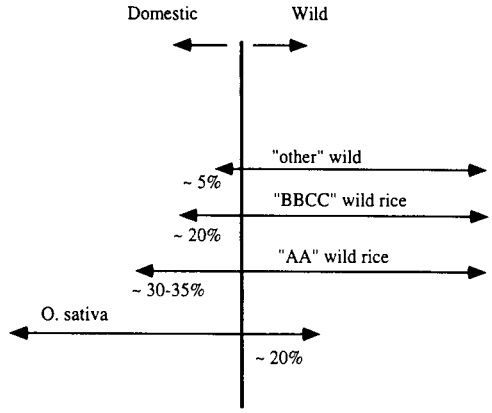

Fig. 4. Schematic representation of sources of error in classifying double-peaked glume cells. Percentage error is estimated. "AA" group: *Oryza nivara*, *O. rufipogon*, *O. spontanea*; "BBCC"/"CC" group: *O. officinalis*, *O. minuta*; "other" group: *O. granulata*, *O. meyeriana*, *O. ridleyi*, *O. longiglumis*.

in their areas of distribution, given the qualitative patterns discussed here.

Conclusions

Study of a well-silicified, easily recognizable phytolith type, the double-peaked glume cell, from wild rice species native to eastern Asia and a variety of traditional cultivated rice varieties from China has resulted in an accurate method of discriminating between domesticated and wild rice in archaeological settings. Formulas developed from discriminant analysis permit classification of unknown glume bodies into three groups—domestic, wild, or indeterminate—with a high degree of confidence. Qualitative characteristics of glume bodies may be used to provide additional confirmation of the identification. Application of this method should aid archaeologists in tracing the early history of rice domestication.

Acknowledgments

This research was supported by National Science Foundation grants to Deborah M. Pearsall and Dolores R. Piperno. Robert A. Benfer, Jr. guided the statistical analyses. Laboratory facilities for this research were provided by the Department of Anthropology, American Archaeology Division. MU graduate assistant Sara Stewart extracted the specimens. We gratefully acknowledge the help of the International Rice Research Institute, Philippines; this study would not have been possible without the specimens provided to us by IRRI. The discussion of statistical methods was made clearer as a result of comments from an anonymous reviewer. This research is part of Zhao's doctoral research, directed by Pearsall.

Literature Cited

Ball, T., J. S. Gardner, and J. D. Brotherson. 1996. Identifying phytoliths produced by the inflorescence bracts of three species of wheat (*Triticum monococcum* L., *T. dicoccon* Schrank., and *T. aestivum* L.) using computer-assisted image and statistical analyses. Journal of Archaeological Science 23:619–632.

Benfer, R. A., Jr. 1975. Classification and Sampling. Pages 227–250 *in* J. W. Mueller, ed., Sampling in Archaeology. University of Arizona Press, Tuscon.

Benfer, R. A., Jr., and A. N. Benfer. 1981. Automatic classification of inspectional categories: Multivariate theories of archaeological data. American Antiquity 46:381–396.

Chang, T-T. 1976. Rice. Pages 98–104 *in* N. W. Simmonds, ed., The evolution of crop plants. Longman, London.

Fujiwara, H. 1993. Research into the history of rice cultivation using plant opal analysis. Pages 147–158 *in* D. M. Pearsall and D. R. Piperno, eds., Current research in phytolith analysis: applications in archaeology and paleoecology. MASCA, University of Pennsylvania Museum, Philadelphia, PA.

Kachigan, S. K. 1991. Multivariate statistical analysis: a conceptual introduction. Radium Press, New York.

Lachenbruch, P. A. 1975. Discriminant analysis. Hafner Press, New York.

Lee, P. M. 1989. Bayesian statistics: an introduction. Oxford University Press, New York.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. Multivariate analysis. Academic Press, New York.

Mulholland, S. C. 1993. A test of phytolith analysis at Big Hidatsa, North Dakota. Pages 131–148 *in* D. M. Pearsall and D. R. Piperno, eds., Current research in phytolith analysis: applications in archaeology and paleoecology. MASCA, University of Pennsylvania Museum, Philadelphia, PA.

Pearsall, D. M. 1989. Paleoethnobotany. A Handbook of Procedures. Academic Press, San Diego.

———, **and D. R. Piperno.** 1990. Antiquity of maize cultivation in Ecuador: summary and reevaluation of the evidence. American Antiquity 55:324–337.

———, **and** ———, **eds.** 1993. Current research in phytolith analysis: applications in archaeology and paleoecology. MASCA, University of Pennsylvania Museum, Philadelphia, PA.

———, **and R. A. Benfer, Jr.** 1994. Identifying crops through phytolith analysis. Paper presented at the 59th Annual Meeting of the Society for American Archaeology, Anaheim, CA.

———, **D. R. Piperno, E. H. Dinan, M. Umlauf, Z. Zhao, and R. A. Benfer, Jr.** 1995. Distinguishing rice (*Oryza sativa* Poaceae) from wild Oryza species through phytolith analysis: results of preliminary research. Economic Botany 49:183–196.

Piperno, Dolores R. 1988. Phytolith analysis. An archaeological and geological perspective. Academic Press, San Diego.

Powers-Jones, A. H., and J. Padmore. 1993. The use

of quantitative methods and statistical analysis in the study of opal phytoliths. Pages 47–56 *in* D. M. Pearsall and D. R. Piperno, eds., Current research in phytolith analysis: applications in archaeology and paleoecology. MASCA, University of Pennsylvania Museum, Philadelphia, PA.

Rapp, George, Jr., and Susan C. Mulholland, eds. 1992 Phytolith systematics. Emerging issues. Plenum Press, New York.

Zhao, Z. 1996. Rice domestication in the Middle Yangtze region, China: an application of phytolith analysis. Unpublished doctoral dissertation, Department of Anthropology, University of Missouri, Columbia.

Book Review

The Evolutionary Biology of Plants. Karl J. Niklas. 1997. The University of Chicago Press, 5801 South Ellis Avenue, Chicago, IL 60637. xx + 449 pp. (paperback). US$19.95, £15.95 (paperback), $65.00, £51.95 (hardcover). ISBN 0-26-58083-0 (paperback); 0-26-58082-2 (hardcover).

This is a wonderful plant evolutionary biology book that I highly recommend. All botany instructors will want a copy within arm's reach, and the level of presentation lends itself to assigned undergraduate and graduate readings. One of the book's strengths is coverage of a multitude of topics. The numerous figures are clear and the writing style is lucid. The book is a convenient size and has an attractive cover.

In part one, "Evolutionary Basics," examples are drawn from both wild and cultivated plants. I enjoyed the discussion of species and speciation, because the author sifted through the prominent ideas of workers past and present to provide a straight-forward synthesis. Part two, "Life's Chronicles: The Fossil Record," made fascinating reading: it provided a summary of current thinking about the origin of life on earth and the invasion of land. In part three, "Adaptive Walks: A Hypothesis," computer simulations of the evolution of phenotypes are presented. In part four, "Long-Term Trends," I was pleased to see a discussion of analogy, homology, divergence, and convergence in a phylogenetic context, and an introduction to genome structure and rates of genomic evolution for plastids, mitochondria, and nuclei.

This book is loaded with interesting facts and insights about so many topics. For example, if plants and animals are viewed together, sympatric speciation may be more common than allopatric speciation. Photosynthesis evolved twice. The evolution of heterospory (convergently derived many times) may have conferred other advantages besides those afforded by outcrossing. The discussion of the many ways that arborescence has been attained by plants (e.g., dicots, palms, tree ferns, extinct horsetails and lycopods), along with the clear, accompanying diagrams, provides a wonderful study of convergent adaptation.

Throughout the text words are defined when they first appear. Definitions are clear and concise, except for one. Anisospory is defined as a condition involving "spores that look alike but develop into sexually dimorphic gametophytes." Here I was left wondering about the intended meaning of "look alike" because anisospory, as originally defined, refers to *two sizes* of spores (produced in the same sporangium) that develop into sexually dimorphic gametophytes (Vitt, D. H. 1968. Sex determination in mosses. The Michigan Botanist 7:195–203).

The text is not without typographical errors, some of which could have been avoided with electronic spell checking by the publisher ("againt" p. 84, "confrounted" p. 306) and others that must have been missed by proofreaders ("among natural*ly* populations," p. 34; "decease" instead of "decrease" p. 39; "microgametophyte" instead of "gametophyte," line 33, p. 194; "each phenotype [in] this domain," p. 215). But the few errors detract little from the quality of this book. This masterful work was a pleasure to read because it integrated morphological, ecological, genetic and population genetic thinking leading to fertile summaries, and for so many topics, interesting synthesis of ideas.

Thomas Mione
Biological Sciences
Central Connecticut State University
New Britain, CT 06050-4010, U.S.A.