

*Farmers' Varietal Identification in a Reference Sample of Local Phaseolus Species in the Sierra Juárez, Oaxaca, Mexico*

**Daniela Soleri, Margaret Worthington, Flavio Aragón-Cuevas, Steven E. Smith & Paul Gepts**

**Economic Botany**

ISSN 0013-0001

Volume 67

Number 4

Econ Bot (2013) 67:283-298

DOI 10.1007/s12231-013-9248-1

**Economic Botany** Volume 67 Number 4  
Issued 1 December 2013  
*Devoted to Past, Present, and Future Uses of Plants by People*



Published for The Society for Economic Botany  
by The New York Botanical Garden

 THE NEW YORK  
BOTANICAL GARDEN



 Springer

12231 • ISSN 0013-0001  
67(4) 283-404 (2013)

 Springer

**Your article is protected by copyright and all rights are held exclusively by The New York Botanical Garden. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Farmers' Varietal Identification in a Reference Sample of Local *Phaseolus* Species in the Sierra Juárez, Oaxaca, Mexico<sup>1</sup>

DANIELA SOLERI<sup>\*,2</sup>, MARGARET WORTHINGTON<sup>3,4</sup>, FLAVIO ARAGÓN-CUEVAS<sup>5</sup>, STEVEN E. SMITH<sup>6</sup>, AND PAUL GEPTS<sup>3</sup>

<sup>2</sup>Geography Department, University of California, Santa Barbara, CA, USA

<sup>3</sup>Department of Plant Sciences, University of California, Davis, CA, USA

<sup>4</sup>Present Address: Department of Crop Science, North Carolina State University, Raleigh, NC, USA

<sup>5</sup>Campo Experimental Valles Centrales, Oaxaca, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), Santo Domingo Barrio Bajo, Etla, Oaxaca, México

<sup>6</sup>School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA

\*Corresponding author; e-mail: soleri@geog.ucsb.edu

---

**Farmers' Varietal Identification in a Reference Sample of Local *Phaseolus* Species in the Sierra Juárez, Oaxaca, Mexico.** Farmer-named varieties are often the basis of *in situ* diversity assessment, collections for *ex situ* conservation, and on-farm improvement programs. Such varieties play an important role in sustainable agriculture because of their adaptation to local environmental conditions and consumer tastes. The importance of these varieties has stimulated interest in understanding farmers' varietal classifications. We investigated the empirical basis of, and agreement among, farmers' bean variety classification in a community in the Sierra Juárez, Oaxaca, Mexico. A reference sample of 300 local seeds of three *Phaseolus* species was sorted by nine farmers into named varieties. Nuclear and chloroplast microsatellite markers and seed morphology data were used to a) establish species identities; and test the hypotheses that b) farmer varieties reflect morphological and genetic structures; and c) there is agreement among farmers in variety classification. Because all farmers sorted the same set of seeds the variation in individual farmers' classifications could be documented and compared. Our results indicate an empirical basis for farmer varieties, but without stringent classification rules. Varietal names underestimated diversity present at the community level because of the intravarietal variation present in farmer classifications. There was low classification agreement among farmers, although broad morphological and genetic patterns were present. The variation in farmers' classifications of this *Phaseolus* diversity resulted in both synonymy and homonymy across classifications. The goal of farmers may not be to maintain the same variety across households, but to form a version of a broad type that best fits their own needs and circumstances at one point in space and time. Thus, in both work with farmers and collections of their *Phaseolus* varieties for *ex situ* conservation it should not be assumed that same-named seed lots are redundant units of diversity. Morphological and/or molecular data should, therefore, supplement farmer varietal names in assessments of *in situ* crop diversity, while *ex situ* collections would benefit from the inclusion of multiple accessions of the same variety from different farmers, repeated over time.

**Identificación campesina de variedades en una muestra de referencia de especies locales de *Phaseolus* en la Sierra Juárez, Oaxaca, México.** Los nombres locales de las variedades ofrecidos por los agricultores son, a menudo, la base para la determinación de diversidad in

---

<sup>1</sup> Received 3 October 2012; accepted 18 July 2013; published online 10 October 2013.

Daniela Soleri and Margaret Worthington contributed equally.

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s12231-013-9248-1) contains supplementary material, which is available to authorized users.

situ, las colecciones para la conservación *ex situ*, y para los programas de mejoramiento participativo. Tales variedades juegan un papel importante en la agricultura sostenible debido a su adaptación a las condiciones ambientales locales y los gustos de los consumidores. La importancia de estas variedades ha estimulado el interés en comprender los principios clasificatorios de los agricultores. Nosotros investigamos los principios empíricos y el grado de acuerdo sobre la clasificación de frijoles entre los agricultores de la Sierra Juárez, Oaxaca, México. Una muestra de referencia de 300 semillas locales de tres especies de *Phaseolus* fue clasificada por nueve agricultores en variedades con nombres locales. Morfología de las semillas y marcadores moleculares del núcleo y de lo cloroplasto fueron utilizados para: a) establecer las identidades de las especies; y para probar las hipótesis: b) que variedades identificadas por los campesinos reflejan estructuras morfológicas y genéticas; y c) que existe concordancia entre los agricultores en cuanto a la identificación de las variedades. Debido a que los agricultores clasificaron el mismo conjunto de semillas, la variación entre las clasificaciones de los agricultores como individuos pudo ser documentada y comparada. Nuestros resultados indican que hay una base empírica de la identificación campesina de las variedades, pero sin reglas clasificatorias estrictas. Los nombres de las variedades subestiman la diversidad existente en la comunidad debido a la inconsistencia presente entre las clasificaciones de los agricultores. Hubo un bajo grado de concordancia entre los agricultores, aunque pautas morfológicas y genéticas generales fueron presentes. La variación entre las clasificaciones campesinas de esta diversidad de *Phaseolus* resultó tanto en la sinonimia como homonimia entre las clasificaciones. Es posible que la meta de los agricultores no sea mantener la misma variedad de fíjol entre granja y granja, sino de un tipo general formar una versión que mejor responda a sus propios intereses y circunstancias en el espacio y el tiempo. Así, tanto en el trabajo con los agricultores y en la colección de sus variedades de *Phaseolus* para conservación *ex situ*, no debe asumirse que los lotes con el mismo nombre sean unidades redundantes de diversidad. Los datos morfológicos y/o moleculares deben, por consiguiente, complementar los nombres que proporcionan los agricultores en la determinación de la diversidad *in situ*, mientras que las colecciones *ex situ* podrían beneficiarse de la inclusión de accesiones múltiples de la misma variedad, pero de distintos agricultores, repetidas a lo largo del tiempo.

**Key Words:** common bean, crop genetic resources, farmer classification, farmer variety, germplasm collection, landrace, Oaxaca, Mexico, *Phaseolus*, varietal name.

## Introduction

Farmer-named varieties are the first order of *in situ* diversity assessment (Jarvis et al. 2008) and the basis of collections for *ex situ* conservation (Hammer and Morimoto 2011). Some see these varieties as the foundation for the development of more environmentally and socially sustainable agriculture (IAASTD 2009), and increasingly they form the basis of collaborative, on- and off-farm improvement for local use (Ceccarelli 2012). Interest in these varieties has prompted investigations of what farmer-named varieties represent, that is what these classifications mean biologically, and how they may differ among households and communities.

Ethnobiological research into emic classifications has established that there can be extensive variation in the use of local names even within a community due to individual variation, including degree and expression of knowledge, personal experience, and history (Berlin 1992). This raises the possibility of both synonymy (more than one name referring to the same material) and homonymy (one name applied

to two or more different materials). Etic interpretations of farmers' variety classifications are based on the implicit assumptions that these classifications structure material into units that are discrete, mutually exclusive and consistent, eliminating both synonymy and homonymy, similar to the goal in formal systems of crop germplasm management. And even when scientists know these assumptions may not be valid, in our (FA-C) extensive experience in southern Mexico, time and resource limitations often mean that only a handful of households are sources of collections from a community, and redundant named-varieties are given lower priority. Finally, many working with farmers and their crops have noticed that farmer-named varieties can present a bewildering diversity of names and materials. It is hard to know if this reflects differences in crop materials, farmer knowledge, both, or neither. A recent review by Sadiki et al. (2006) summarized research investigating some of these etic assumptions and found variable levels of "consistency," that is agreement, for different aspects at different scales:

both high and low agreement within communities for variety names (e.g., wheat in Ethiopia, sorghum in Burkina Faso); higher within and decreasing consistency between communities for sets of phenotypic descriptors used to distinguish named varieties (e.g., fava in Morocco); very low agreement in one growing environment and very high in two others for varietal names and associated phenotypic descriptors (e.g., rice in Nepal). Those studies provide valuable insights into how farmers describe and name samples of existing varieties. In our study, instead of asking farmers to name or describe an existing unit (e.g., a local variety), we investigated two issues fundamental to studies of farmers' varietal names: 'How do different farmers structure the diversity available to them?' and, 'How much agreement is there in this classification among farmers?' Presenting farmers with a diverse sample of local seed allowed us to observe each individual's classification of that diversity. Yet, because all farmers sorted the same seed sample, there was a common empirical reference point that made comparisons across different farmers' classifications of that sample possible.

We tested hypotheses with data from a bean seed sorting exercise presented to a small sample of maize and bean farmers in Santa María Jaltianguis, a Zapotec community in the Sierra Juárez of Oaxaca, Mexico. This is a region of particularly rich *Phaseolus* diversity (Kwak et al. 2009) within the larger Mesoamerican area of domestication of *P. vulgaris* L.—an annual species adapted to lower elevations and warmer temperatures (average 24°C), *P. coccineus* L.—a perennial species associated with cooler temperature (~20°C) highlands, often >2100 masl (Escalante et al. 1994), and their ancient hybrid *P. dumosus* Macfad., an annual with intermediate adaptation (Llaca et al. 1994). Farmers in many of the communities in the Sierra Juárez, including Santa María Jaltianguis, cultivate all three *Phaseolus* species in traditional agricultural fields. Harvests are primarily for household consumption but are occasionally marketed in local communities. In this and other regions of Mexico, seed characteristics are important criteria determining consumer acceptability of bean varieties grown primarily for their dry seed. Farmers consider these characteristics when selecting seeds either for consumption, planting, or sale in markets; 75% of Santa María Jaltianguis farmers select their planting seed

from stored material just before sowing (D. Soleri and F. Aragón-Cuevas, unpublished data 2013).

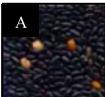
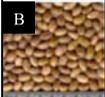
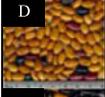
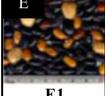
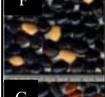
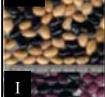
The goal of this study was to better understand similarities and differences in how farmers organize their crop diversity by empirical comparisons among farmer-named bean varieties in a single farmer's classification, and between classifications by different farmers made from the same reference seed sample of local *Phaseolus* diversity. Specifically, we asked 1) Are farmers' varietal classifications empirically supported, i.e. non-random, with respect to neutral genetic markers and seed morphological characteristics? and, 2) Is there agreement among the classifications of different farmers?

## Materials and Methods

### SEED COLLECTION AND CLASSIFICATION

This research builds on a larger study of maize and bean systems in Oaxaca that included farmer interviews and seed collections (D. Soleri and F. Aragón-Cuevas, unpublished data 2013). Santa María Jaltianguis is an agricultural community in the Sierra Juárez with a 2010 population of 575 (INEGI 2011). Households there, like much of the Sierra Juárez (González 2001), work fields distributed across an elevational gradient (~2500–1060 masl), resulting in distinct planting patterns, crop species and varietal combinations associated with the three growing environments they recognize: *tierra fría*, *tierra templada*, and *tierra caliente* (high-, mid-, and low-elevation lands, respectively). For the larger study a spatially stratified random sample of 50 households in Santa María Jaltianguis was drawn from a complete community census, interviews were conducted in 2007 with senior household member(s) responsible for farming, and maize and bean seed collections from many of those same households were made in March 2008. Based solely on seed phenotypes, INIFAP *Phaseolus* expert J.S. Muruaga provided species identification of the bean collections. Based on that identification, seed availability and quality, and with the explicit goal of including intraspecific diversity, 100 seeds were selected from each of the three species present: *P. vulgaris*, *P. dumosus*, and *P. coccineus* (Table 1). Each seed was numbered using permanent ink. From the 50 original interviews, nine farmers (hereafter FMR1–FMR9) were chosen to represent the range in number of bean varieties being grown by a household in the

Table 1. MARCH 2008 SOURCE COLLECTIONS FOR THE 300-SEED SAMPLE USED IN THIS STUDY, FROM 2007 HARVEST.

Source collection	Collection description by farmer		Scientist's visual assessment		N in study sample	Morphological characteristics, by collection							
	Zapotec name (translation)	growing environment in 2007 <sup>a</sup>	<i>Phaseolus</i> spp. in collection	<i>Phaseolus</i> spp. seeds used in sample		weight, gm		perimeter, mm		width:length ratio		B-color value	
						Ave	Std	Ave	Std	Ave	Std	Ave	Std
 <b>A</b>	<i>Daá güin-nii</i> (bean for treating pustules on eyelid)	TC (TT)	<i>vulgaris</i>	<i>vulgaris</i>	10	0.07	0.01	23.26	2.97	0.65	0.03	23.00	0.00
 <b>B</b>	<i>Daá ya-áá</i> (green bean, to eat fresh)	TT	<i>vulgaris</i>	<i>vulgaris</i>	15	0.42	0.07	171.68	33.57	0.70	0.04	107.00	0.00
 <b>C</b>	<i>Daá laá</i> (leafy bean)	TT	<i>vulgaris</i>	<i>vulgaris</i>	30	0.28	0.04	39.60	3.64	0.59	0.03	23.00	0.00
 <b>D</b>	<i>Daá bel-la tuaa</i> (bean the color of maguay worm)	TT	<i>vulgaris</i> and <i>dumosus</i>	<i>vulgaris</i>	15	0.30	0.04	84.29	26.40	0.65	0.04	45.00	0.00
 <b>E</b>	<i>Daá yel-la</i> (milpa beans) + <i>Daá bel-la tuaa</i> (see D) mixed <sup>b</sup>	TT	<i>vulgaris</i> + <i>dumosus</i>	<i>vulgaris</i> + <i>dumosus</i>	90	0.43	0.13	52.90	10.24	0.68	0.07	23.00	0.00
 <b>E1</b>	<i>Daá bel-la tuaa</i>		<i>vulgaris</i>	<i>vulgaris</i>	30	0.30	0.05	44.04	7.93	0.62	0.05	23.00	0.00
 <b>E2</b>	<i>Daá yel-la</i>		<i>dumosus</i>	<i>dumosus</i>	60	0.49	0.11	57.32	8.21	0.71	0.06	23.00	0.00
 <b>F</b>	<i>Daá ix-xiía</i> (field bean, wide bean)	TF	<i>dumosus</i>	<i>dumosus</i>	5	0.58	0.16	193.03	111.04	0.75	0.06	90.00	0.00
 <b>G</b>	<i>Daá laá</i> (see C)	TC	<i>vulgaris</i> and <i>dumosus</i>	<i>dumosus</i>	25	0.67	0.14	119.67	43.72	0.78	0.07	27.00	0.00
 <b>H</b>	<i>Daá yel-la</i> (see E)	TT	<i>vulgaris</i> , <i>dumosus</i> and <i>coccineus</i>	<i>dumosus</i>	10	0.36	0.07	64.45	9.12	0.71	0.05	156.00	0.00
 <b>I</b>	<i>Daá tupii</i> (whistling bean, while dehiscent)	TT (TF)	<i>vulgaris</i> and <i>coccineus</i>	<i>coccineus</i>	60	0.39	0.10	55.04	8.39	0.80	0.07	52.00	29.70
 <b>J</b>	<i>Daá xhidu</i> (light color, ray-like pattern similar to cat eye)	TF (TC)	<i>vulgaris</i> and <i>coccineus</i>	<i>coccineus</i>	40	0.43	0.09	90.40	29.49	0.76	0.05	69.00	18.00

<sup>a</sup> Farmer taxonomy: TF, *tierra fría*; TT, *tierra templada*; TC, *tierra caliente*. High-, mid- and low-elevation fields, respectively

<sup>b</sup> Farmer grows and stores as mixed beans, including these two components that he identified individually at time of collection

2007 season: FMR2, 4, 8, 9 one variety; FMR6 two varieties; FMR1, 3, 5, 7 three varieties. In December 2008, each farmer was individually asked to sort the entire sample of 300 seeds into

varieties so that we would know how to label them in our collection. When farmers completed their sorting we recorded the variety name they assigned to each seed.

Morphological data documented for each seed included visually scored RHS color classes (Royal Horticultural Society 2001) (converted to RGB values [R, G, and B] according to <http://www.azaleas.org/index.pl/rhsmacfan1.html>); and weight (WT, in gm), maximum width (W, in mm), maximum length (L, in mm), perimeter (i.e., circumference of the seed where the two cotyledons meet, PER, in mm), and maximum width:length ratio (WLR) using image analysis software (SigmaScan Pro V.5).

#### MICROSATELLITE MARKER ANALYSIS AND ANALYSIS OF POPULATION STRUCTURE

In April 2009, all 300 seeds used in the sorting exercise were sown in a greenhouse on the University of California, Davis campus. Sixty-five of the seeds failed to germinate and were, therefore, excluded from analyses requiring microsatellite data. Approximately one week after germination, the relative lengths of the hypocotyl and epicotyl shoots of each seedling were recorded as morphological evidence of species membership. DNA extraction and microsatellite marker analysis were then conducted following the methods described in Worthington et al. (2012).

Genotypic diversity present within the sample was evaluated at ten putatively neutral nuclear and three chloroplast DNA (cpDNA) microsatellite (SSR) loci selected based on their wide distribution over the entire *Phaseolus* genome and high polymorphism information content (PIC) values. Selected markers included BMd-20, BMd-41, BMd-42 (Blair et al. 2003), BM143, BM151, BM172, BM210 (Gaitan-Solis et al. 2002), PV-ag003, PV-at007, PV-ctt001 (Yu et al. 2000), ccmp3 (Weising and Gardner 1999), ccSSR9, ccSSR16 (Xu et al. 2002). More information about the SSR marker generation and interpretation can be found in Worthington et al. (2012).

STRUCTURE 2.1 software (Pritchard et al. 2000) was used to infer the number of meaningful populations ( $K_n$ ) within our sample with no *a priori* information other than genotype data following Evanno et al. (2005) as described by Worthington et al. (2012). Chloroplast haplotypes were determined from the data generated with the three cpDNA markers. GenAlEx 6 (Peakall and Smouse 2006) was used to calculate the number of alleles, gene diversity, and observed heterozygosity of each of the populations identified by STRUCTURE.

#### ANALYSIS OF SEED MORPHOLOGICAL VARIATION

Distribution of variation in morphological characteristics of bean seeds in each farmer's and the STRUCTURE classes were analyzed using multinomial logistic regression to provide multivariate models of class membership, linear discriminant functions of those characteristics to estimate 'classification error', and one-way ANOVAs. All analyses were conducted using SAS v. 9.2 (SAS Institute 2002–2011), with a significance level of  $P < 0.05$ . Due to high correlations among some quantitative morphological variables, only variables among which  $r < |0.44|$  were used: seed weight (WT), perimeter (PER), maximum width:length ratio (WLR), and B-color value from RGB conversions of the RHS color classes. Morphological characteristics were individually transformed for each classification system using either the Box-Cox or natural logarithm transformations.

#### FARMER CLASSIFICATION AGREEMENT

All possible pairwise comparisons of the nine farmers' classifications were analyzed using the adjusted Rand index (ARI) (Fisher and Hoffman 1988) to provide a quantitative measure of classification agreement regardless of actual names assigned. ARI is 0 if classification agreement is what would be expected by chance, 1 if agreement is complete, and negative if there is less agreement than would be expected by chance. These statistics were calculated using a SAS macro developed by Fisher and Hoffman (1988), revised by von Borries (2008) and S.R. Bowley (<http://www.plant.uoguelph.ca/research/homepages/sbowley/>). The extent of agreement between farmers' and STRUCTURE classifications was analyzed using the same procedure. It is not possible to define statistical significance of ARI statistics. Therefore, we evaluated the meaning of these statistics using experience-based thresholds that have been derived from a variety of experimental settings (Steinley 2004). The hypothesis that farmer classifications randomly distribute STRUCTURE groups ( $K3-K8$ ) was further evaluated using a Fisher's exact test for each of the total of 46 farmer classes.

To visually assess agreement within the same farmer-named variety, morphology-based principal components analysis (PCA) was performed on all seed lots attributed to a variety and two dimensional scattergrams for mean principal components one and two (PC1 and PC2) for each farmer's seed lot of that variety generated.

## Results

The nine farmers sorted the 300 seeds into anywhere from two to seven varietal classes. There was no relationship between the number of growing environments in which farmers grew beans or number of varieties they reported growing in 2007 or 2008 and the number of classes farmers identified in the sorting exercise (Fisher's exact tests, data not shown).

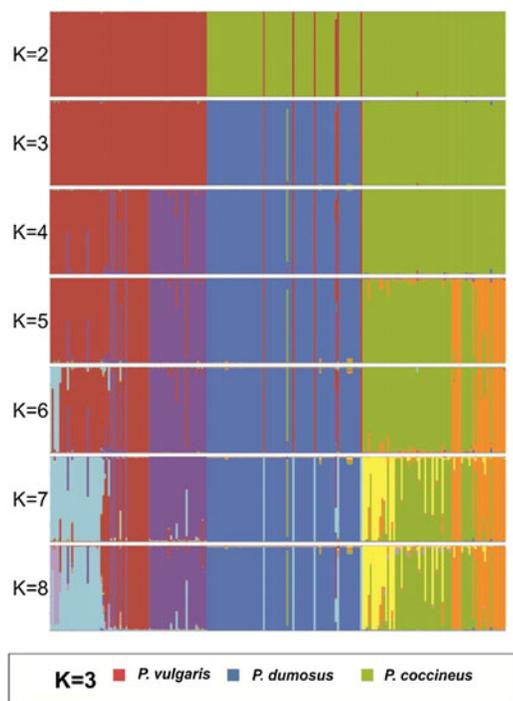
### EMPIRICAL CHARACTERISTICS OF THE SEED SAMPLE

STRUCTURE analysis corroborated almost all of the species designations identified by INIFAP scientist J.S. Muruaga. Assuming two population clusters ( $K=2$ ), STRUCTURE distinguished between samples of *P. vulgaris* and those of *P. dumosus* and *P. coccineus* combined. At three clusters ( $K=3$ ), STRUCTURE differentiated the three species *P. vulgaris*, *P. dumosus*, and *P. coccineus* (Fig. 1). Six seeds identified as *P. dumosus* by Muruaga based on seed morphology were classified as *P.*

*vulgaris* based on nuclear marker data. Another seed, previously identified as *P. dumosus*, was assigned to the *P. coccineus* group by STRUCTURE. Species assignments inferred by STRUCTURE ( $K=2-3$ ) were corroborated by seed morphology, germination habit and cpDNA haplotype data (data not shown). Hypogeal germination was observed in all 75 seeds assigned to the *P. coccineus* group by STRUCTURE. In contrast, cotyledon scars visible above the soil line—characteristic of epigeal germination—were present in all of the seeds assigned to *P. vulgaris* and *P. dumosus*. A total of five cpDNA haplotypes were observed within the 235 sampled seeds (data not shown). All *P. dumosus* samples were marked by the same diagnostic cpDNA haplotype. Two diagnostic cpDNA haplotypes each were identified in the *P. vulgaris* and *P. coccineus* samples.

STRUCTURE runs at higher  $K$  settings ( $K=4-8$ ) revealed evidence of significant substructure within both *P. vulgaris* and *P. coccineus*, whereas the *P. dumosus* group remained a single cohesive block in all  $K$  settings (Fig. 1). *Phaseolus vulgaris* was split into four subpopulations with subdivisions occurring at  $K$  settings of 4, 6, and 8, and *P. coccineus* was split twice, at  $K$  settings of 5 and 7. No biologically meaningful subdivisions of the three species could be inferred at  $K$  settings above eight. Summary statistics calculated in GenAlEx reveal that the *P. dumosus* seeds included in this study had the lowest gene diversity and observed heterozygosity. As expected, seeds of *P. coccineus*, a mostly outcrossing species, had the highest number of alleles and observed heterozygosity of the three species. The *P. vulgaris* seeds had the highest gene diversity, consistent with the significant substructure revealed at high  $K$  settings (Table 2).

Seed morphological characteristics including seed weight (WT), perimeter (PER), maximum width:length ratio (WLR), and B-color value varied across this 300 seed sample. Morphology



**Fig. 1.** STRUCTURE bar plot of the membership coefficients for all 235 seeds. The membership coefficients of each seed were obtained from the least negative log likelihood run of preset  $K$  settings 2–8. The optimal clustering number ( $K = 3$ ) was calculated using the ad hoc parameter  $\Delta K$  (Evanno et al. 2005).

**Table 2.** GENETIC DIVERSITY DETECTED IN THREE *PHASEOLUS* SPECIES USING NUCLEAR MICROSATELLITE MARKERS.

<i>Phaseolus</i> spp. (n seeds)	Number of Alleles	Gene Diversity	Observed Heterozygosity
<i>P. vulgaris</i> (87)	4.7	0.53	0.06
<i>P. dumosus</i> (73)	4	0.23	0.06
<i>P. coccineus</i> (75)	6.3	0.49	0.33
Average (235)	5	0.42	0.15

**Table 3.** DESCRIPTIVE STATISTICS FOR FOUR SEED MORPHOLOGY CHARACTERISTICS IN THE ENTIRE 300-SEED *PHASEOLUS* SAMPLE, AND BY SPECIES.

Material	WT (gm)		PER (mm)		WLR		B-value	
	Ave	CV	Ave	CV	Ave	CV	Ave	CV
Entire sample (300)	0.41	0.37	71.68	0.61	0.71	0.13	46.09	0.75
<i>P. vulgaris</i> (87) <sup>a, b</sup>	0.30c	0.30	65.08a	0.75	0.63c	0.08	37.16a	0.75
<i>P. dumosus</i> (73)	0.52a	0.26	81.96b	0.67	0.73b	0.08	44.64a	1.01
<i>P. coccineus</i> (75)	0.42b	0.21	70.44b	0.38	0.78a	0.08	59.04b	0.45

<sup>a</sup> Species statistics only for 235 seeds for which SSR data available

<sup>b</sup> For species data, Tukey test of means differences on transformed data; untransformed means reported; among species, means in same column followed by different letter are significantly different.

of seeds attributed to the three *Phaseolus* species based on SSR marker data often differed significantly (Table 3).

#### ARE FARMERS' VARIETAL CLASSIFICATIONS EMPIRICALLY SUPPORTED?

##### *Farmer Classification and Seed Morphology*

To determine whether farmers' varietal classifications were empirically supported, we tested the null hypothesis that those classifications were entirely random, and thus the distribution of phenotypic and genotypic variation across these classifications would also be random.

Multivariate models from multinomial logistic regression indicate that in all farmer classifications variation in seed characteristics is partitioned non-randomly, although the contribution of each characteristic varied among farmers (Table 4). Ranking of mean standardized logistic regression coefficients for each set of farmer classifications identifies WT and then PER and B-color together as the first and two second-most important morphological bases for those classifications. Multinomial logistic regression analyses within each of the three *Phaseolus* species were only significant in *P. coccineus* for four farmer classifications (FMR1, 4, 8, 9), with B-color and then WT as the most important traits. STRUCTURE-designated population clusters showed significant partitioning of morphological variation with WLR and PER the most important characteristics.

Although morphologically structured, farmers' classification systems varied in their 'classification error', measured by how closely individual seed membership in each class within a sorting adhered to the morphology-based 'rules' (linear equations of the seed characteristics identified in linear discriminant analysis) that describe that

farmer's classification (Table 4). Although a few classifications had relatively low errors (e.g., FMR7, 0.12; FMR2, 0.15), overall the average level of classification error was 0.25 for farmers and 0.28 for STRUCTURE (*K*3-8).

##### *Farmer Classification and STRUCTURE Population Clusters*

The 235 seeds with SSR data were used to analyze distribution of STRUCTURE population clusters in each farmer's sorting. For example, the STRUCTURE species identities (*K*3; Fig. 1) established the proportion of each species in the sample used to calculate expected species frequencies in each farmer variety under the random distribution hypothesis and a Fisher's exact test conducted for each variety in a farmer's sorting. The same procedure was followed for all *K*<sub>n</sub> population clusters.

Genetic partitioning within farmer classes at all *K* levels was predominantly nonrandom, with between 78-87% of tests significant (data not shown). The four subpopulations of *P. vulgaris* identified at *K*8 were nonrandomly distributed in all farmers' classifications except FMR5 (Table 5). Similarly, the three subspecific groups identified in *P. coccineus* at *K*8 were nonrandomly distributed in six of the nine farmers' sortings. Still, as evident in ARI scores, nonrandom distribution does not necessarily indicate strong agreement.

Figure 2 presents patterns of species combinations and distribution in five varieties commonly included in these farmers' classifications, and the consensus farmer assignments of those varieties to elevational growing environments as reported in 2007 and 2008. Varieties associated with higher elevation growing environments where *P. dumosus*

**Table 4.** SIGNIFICANCE (*P*) VALUES ASSOCIATED WITH  $\chi^2$  TESTS OF WALD STATISTICS FOR MAIN EFFECTS IN MULTINOMIAL LOGISTIC REGRESSION<sup>a</sup> OF FARMER-IDENTIFIED VARIETIES AND FOUR SEED CHARACTERISTICS, RANKING OF MEAN REGRESSION COEFFICIENT AND CALCULATION OF 'CLASSIFICATION ERROR' FOR THOSE CHARACTERISTICS.

Farmer (n varieties identified)	Seed characteristic <i>P</i> value <sup>b</sup>				Rank of mean absolute value of standardized logistic regression coefficients				Classification error <sup>c</sup>
	WT	WLR	PER	B	WT	WLR	PER	B	
FMR1 (6)	<0.01	<0.01	<0.01	<b>0.02</b>	1	3	2	4	0.35
FMR2 (6)	<0.01	<b>0.02</b>	<0.01	0.95	2	4	3	1	0.15
FMR3 (7)	<0.01	<0.01	<0.01	0.82	1	4	3	2	0.28
FMR4 (7)	<0.01	<0.01	<0.01	<0.01	1	4	2	3	0.20
FMR5 (2)	<0.01	<b>0.01</b>	0.80	0.24	1	2	4	3	0.29
FMR6 (4)	<0.01	<0.01	<b>0.01</b>	<0.01	1	2	4	3	0.24
FMR7 (4)	<0.01	<0.01	<0.01	0.65	1	3	2	4	0.12
FMR8 (5)	<0.01	<0.01	<0.01	<0.01	1	4	2	3	0.24
FMR9 (7)	<0.01	<0.01	<0.01	<0.01	1	4	3	2	0.36
				Mean	1.1	3.3	2.8	2.8	0.25
K3	<0.01	<0.01	<0.01	<0.01	4	1	2	3	0.17
K4	<0.01	<0.01	<0.01	<0.01	4	1	2	3	0.27
K5	<0.01	<0.01	<0.01	<0.01	3	1	2	4	0.29
K6	<0.01	<0.01	<0.01	<0.01	1	2	4	3	0.29
K7	<0.01	<0.01	<0.01	<0.01	3	1	2	4	0.33
K8	<0.01	<0.01	<0.01	<0.01	1	2	3	4	0.32
				Mean	2.7	1.3	2.5	3.5	0.28

WT = weight, WLR = maximum width:length ratio, PER = perimeter, B = B-color value.

<sup>a</sup>  $\chi^2$  value for Wald test of the significance of the main effects in the multinomial logistic regression model is statistically significant ( $P < 0.05$ ) in all cases.

<sup>b</sup> *P* values in bold are  $< 0.05$ .

<sup>c</sup> Mean fraction of beans within each farmer's varieties incorrectly classified using a linear discriminant function of seed characteristics. Based on cross-validation of farmer varieties and linear discriminant functions.

and *coccineus* have greater adaptation than *P. vulgaris* displayed patterns of greater species mixing than varieties identified with lower elevations. In varieties dominated by *P. vulgaris* the proportion of material belonging to one of two subpopulations identified in the K4 partitioning is also reported. Together these data reject the null hypothesis of random distribution of *Phaseolus* species and two subpopulations of *P. vulgaris* across farmer-identified varieties, as well as across environments.

#### IS THERE AGREEMENT AMONG THE CLASSIFICATIONS OF DIFFERENT FARMERS?

To ascertain agreement among the classification schemes of different farmers working with the same material we compared agreement between different farmers' sortings of the 300-seed sample into varieties and between farmers' and STRUCTURE classifications, using the ARI. Comparisons among farmers show apparently greater classification agreement than would be

expected by chance in a majority of cases (Table 6), but that agreement is extremely weak: averaged across all farmers, the ARI is  $0.20 \pm 0.08$ . As measured by mean ARI, one farmer's classifications (FMR5) were particularly dissimilar to the others because they were characterized by under-differentiation: FMR5 sorted the 300 seeds into only two varieties (Fig. 3). Eliminating this farmer raises the mean ARI across the remaining eight farmers slightly to  $0.22 \pm 0.09$ . Still, all of these ARI values lie well below the accepted rule of thumb threshold of  $< 0.65$  for poor agreement (Steinley 2004). Conducting the same analyses by species indicates higher levels of agreement for some classifications of *P. vulgaris* seeds, but lower agreement within *P. dumosus* and *coccineus* (Table 6).

The highest between-farmer ARI values occur in comparisons of farmers' sorting of the variety *Daá güin-nii* with mean ARI  $0.43 \pm 0.39$ ; 29% of those comparisons had ARI values  $> 0.86$ , including three with values = 1.0. Those high ARI

**Table 5.** FARMER VARIETAL CLASSIFICATIONS AND THE DISTRIBUTION OF SSR-BASED SUBSPECIFIC POPULATION CLUSTERS. FISHER'S EXACT TEST OF RANDOM DISTRIBUTION; ADJUSTED RAND INDEX (ARI) <sup>a,b</sup> ANALYSIS OF AGREEMENT.

Farmer	N varieties identified containing <i>P. vulgaris</i> material	<i>P. vulgaris</i> (87), K8 four subgroups		N varieties identified containing <i>P. cocineus</i> material	<i>P. cocineus</i> (75), K8 three subgroups	
		Fisher's exact test, <i>P</i>	ARI		Fisher's exact test, <i>P</i>	ARI
FMR1	5	<0.01	0.32	5	<0.01	0.05
FMR2	5	<0.01	0.11	4	0.72	-0.04
FMR3	6	<0.01	0.32	2	0.31	-0.02
FMR4	6	<0.01	0.31	4	<0.01	0.33
FMR5	2	0.06	0.00	2	0.37	0.00
FMR6	4	<0.01	0.23	4	<0.01	0.17
FMR7	4	<0.01	0.33	2	<0.01	0.09
FMR8	5	<0.01	0.27	4	<0.01	0.24
FMR9	6	<0.01	0.33	4	<0.01	0.20
Mean ARI values			0.25			0.11

<sup>a</sup> After von Borries 2008; Bowley (<http://www.plant.uoguelph.ca/research/homepages/sbowley/>)

<sup>b</sup> Experience-based thresholds for evaluating quality of agreement between farmers based on adjusted Rand index (Steinley, 2004): > 0.9 = excellent, 0.9-0.8 = good, 0.8-0.65 = moderate, < 0.65 = poor.

values are for sets of only 10–13 seeds each and all include six small black seeds with an average weight of 0.07 gm compared with 0.41 gm for the 300-seed sample overall.

Figure 3 also shows the influence of seed color on farmer classifications especially of *P. vulgaris*. This contributes to greater agreement between farmers' classifications in that species (Table 6) and the average ARI for agreement between farmers' classifications and B-color value within classifications of that species (0.49) was higher than most other ARI tests, but still poor. Similarly, one-way ANOVAs conducted within *P. vulgaris* seeds only indicate color alone has not defined even these classes (data not shown). While STRUCTURE populations were not distributed randomly in most farmers' classifications as reported above, when considering the entire sample, agreement between the two was very weak, as it was for farmer to farmer comparisons. This lack of agreement among farmers in classification of the same 300 seeds indicates synonymy.

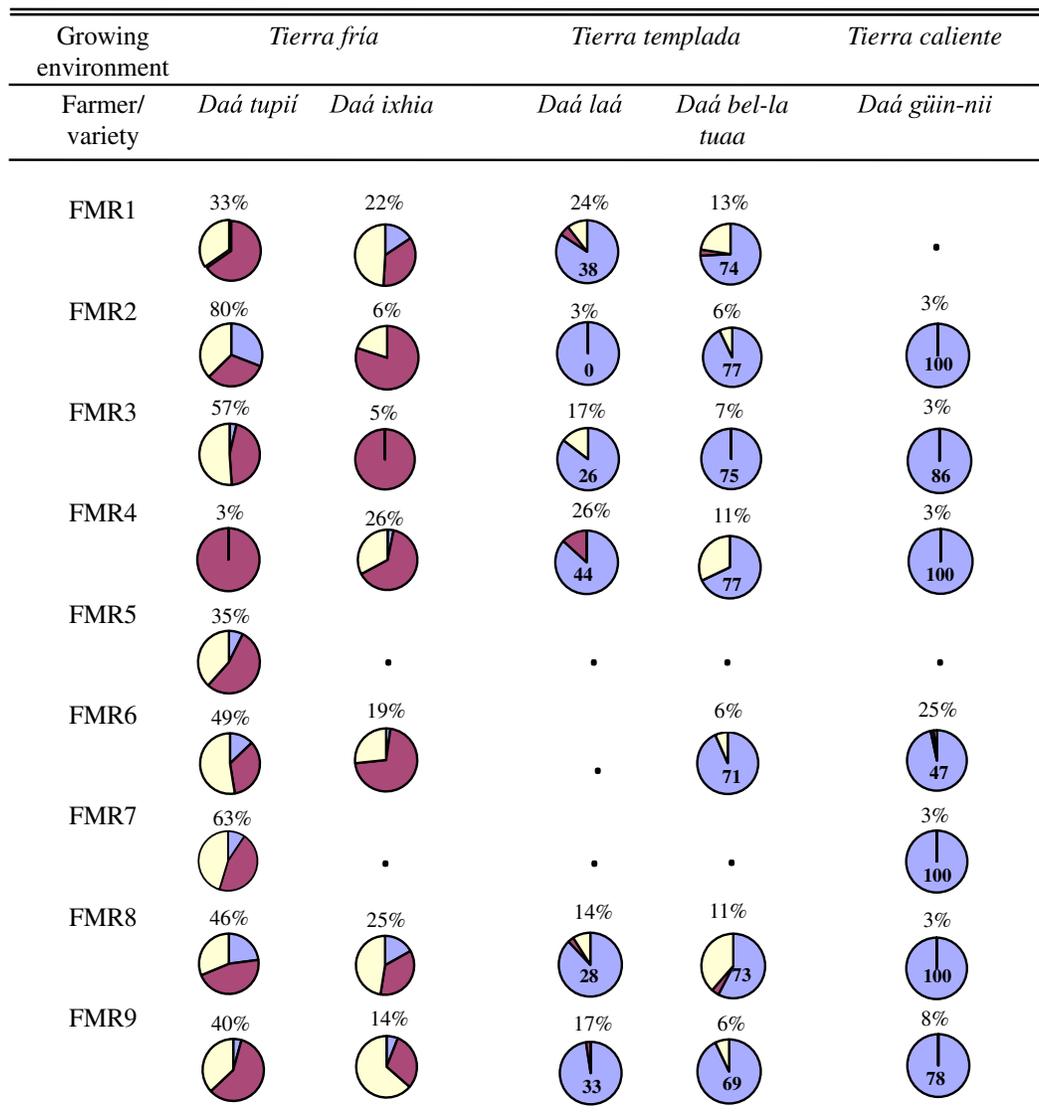
The morphological similarity among seed lots attributed by different farmers to the same named variety is presented in PCA scatter plots of the mean of the first two principal components for each of the five varieties included in Figure 3 (Fig. 4, a-e). Given that all farmers started with the same 300-seed sample, these morphology-based scattergrams for individual varieties display substantial dispersion and homonymy for those

variety names. Still, when integrated into a single PCA (Fig. 4f), varietal clusters are evident.

## Discussion

Farmer-named crop varieties in traditional systems are acknowledged as important first order heuristics for organizing and understanding diversity from both emic and etic perspectives (Hammer and Morimoto 2011:7–8), but these perspectives can result in different meanings (Zent 1996). For formal scientists working with farmers and/or collecting and conserving the diversity those farmers manage, practice tends to be based on simplifying assumptions including that names indicate distinct, mutually exclusive, consistent groupings, at least within a community. From this perspective farmers' varietal names are invaluable, but because of our assumptions, they can also be bewildering, especially in systems where multiple growing environments and interbreeding plant species are present as is true for *Phaseolus* in Santa María Jalteanguis. We asked a sample of farmers there to classify the same, diverse set of local *Phaseolus* seeds into varieties to better understand how classification by different people structures empirically observable variation in the same material, and the level of agreement among those classifications.

We found that farmers' varietal identifications were, for the most part, effectively discriminating



• Farmer did not identify this variety in the seed sorting.

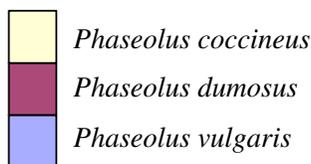


Fig. 2. *Phaseolus* species composition of the same variety as identified by multiple farmers from a 235-seed sample, organized by farmer-identified growing environment. The proportion of the 235 seeds attributed to that variety by the farmer is given above the chart; the proportion of *P. vulgaris* seeds belonging to one of the two subpopulations identified at K4 is given inside charts of varieties with large proportions of *P. vulgaris*.

**Table 6.** COMPARISON OF FARMERS' AND SSR-BASED POPULATION STRUCTURE CLASSIFICATIONS OF A 300 SEED SAMPLE OF THREE *PHASEOLUS* SPECIES AND OF INDIVIDUAL SPECIES USING THE ADJUSTED RAND INDEX (ARI) <sup>a,b,c</sup>

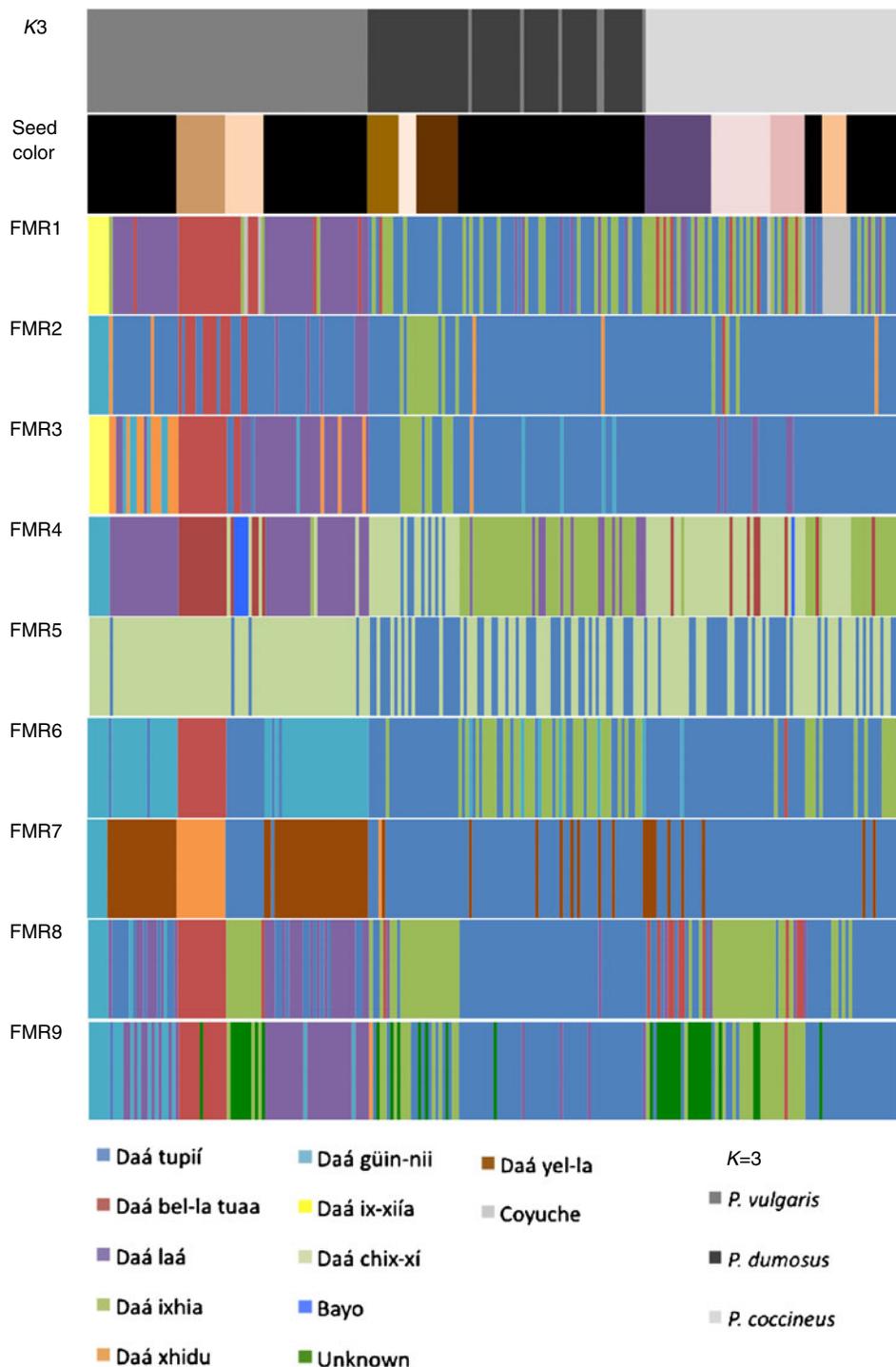
Classification	FMR1	FMR2	FMR3	FMR4	FMR5	FMR6	FMR7	FMR8	FMR9	Mean
FMR1 (6)		0.05	0.22	0.26	<b>0.07</b>	0.23	0.29	0.16	0.23	0.19
FMR2 (6)	0.05		0.25	0.10	0.00	0.06	0.19	0.15	0.07	0.11
FMR3 (7)	0.22	<b>0.25</b>		0.27	-0.01	0.31	<b>0.47</b>	<b>0.25</b>	<b>0.32</b>	0.26
FMR4 (7)	0.26	0.10	0.27		0.07	0.37	0.28	<b>0.31</b>	0.31	0.25
FMR5 (2)	0.07	0.00	-0.01	0.07		0.05	0.01	0.02	0.02	0.03
FMR6 (4)	0.23	0.06	0.31	<b>0.37</b>	0.05		0.41	0.19	0.26	0.23
FMR7 (4)	<b>0.29</b>	0.19	<b>0.47</b>	0.28	0.01	<b>0.41</b>		0.23	0.27	0.27
FMR8 (5)	0.16	0.15	0.25	0.31	0.02	0.19	0.23		0.31	0.20
FMR9 (7)	0.23	0.07	0.32	0.31	0.02	0.26	0.27	0.31		0.22
									Mean	0.20
<i>K3</i> <sup>d</sup>	0.26	0.01	0.28	0.28	<b>0.15</b>	<b>0.29</b>	0.25	0.08	0.24	0.21
<i>K4</i>	0.24	<b>0.05</b>	<b>0.34</b>	0.25	0.06	0.27	<b>0.27</b>	0.12	0.26	<b>0.21</b>
<i>K5</i>	0.25	0.02	0.25	0.29	0.06	0.25	0.21	0.14	0.27	0.19
<i>K6</i>	0.26	0.03	0.27	<b>0.31</b>	0.06	0.25	0.23	0.15	0.27	0.20
<i>K7</i>	0.29	0.02	0.24	0.27	0.04	0.22	0.23	0.18	0.31	0.20
<i>K8</i>	<b>0.30</b>	0.03	0.25	0.27	0.03	0.21	0.23	<b>0.19</b>	<b>0.33</b>	0.20
									Mean	0.20
<i>P. vulgaris</i> only (n=87)										
FMR1 (5)		0.32	0.30	<b>0.64</b>	<b>0.05</b>	0.47	0.63	0.39	0.44	0.41
FMR2 (5)	0.32		0.15	0.32	0.02	0.21	0.36	0.18	0.19	0.22
FMR3 (6)	0.30	0.15		0.27	0.01	0.20	0.31	0.27	0.31	0.23
FMR4 (6)	<b>0.64</b>	0.32	0.27		0.04	0.52	<b>0.78</b>	0.46	0.45	0.44
FMR5 (2)	0.05	0.02	0.01	0.04		0.05	0.00	0.01	0.02	0.02
FMR6 (4)	0.47	0.21	0.20	0.52	<b>0.05</b>		0.69	0.35	0.48	0.38
FMR7 (4)	0.63	<b>0.36</b>	0.31	0.78	0.00	<b>0.69</b>		<b>0.48</b>	<b>0.52</b>	<b>0.47</b>
FMR8 (5)	0.39	0.18	0.27	0.46	0.01	0.35	0.48		0.40	0.32
FMR9 (6)	0.44	0.19	<b>0.31</b>	0.45	0.02	0.48	0.52	0.40		0.35
									Mean	0.31
<i>P. dumosus</i> only (n=73)										
FMR1 (4)		-0.12	-0.06	0.00	-0.01	0.00	<b>0.20</b>	0.00	-0.03	0.00
FMR2 (3)	-0.12		<b>0.48</b>	0.25	-0.02	0.03	-0.09	0.44	0.13	0.14
FMR3 (3)	-0.06	<b>0.48</b>		0.24	-0.02	0.02	0.02	0.39	0.16	0.15
FMR4 (4)	0.00	0.25	0.24		<b>0.02</b>	0.19	0.04	<b>0.53</b>	0.30	0.20
FMR5 (2)	-0.01	-0.02	-0.02	0.02		0.00	0.01	0.01	-0.03	-0.01
FMR6 (3)	0.00	0.03	0.02	0.19	0.00		-0.02	0.20	0.03	0.06
FMR7 (3)	<b>0.20</b>	-0.09	0.02	0.04	0.01	-0.02		0.03	0.02	0.03
FMR8 (4)	0.00	0.44	0.39	<b>0.53</b>	0.01	<b>0.20</b>	0.03		<b>0.44</b>	<b>0.25</b>
FMR9 (5)	-0.03	0.13	0.16	0.30	-0.03	0.03	0.02	0.44		0.13
									Mean	0.10
<i>P. coccineus</i> only (n=75)										
FMR1 (5)		-0.03	-0.01	0.07	0.01	0.01	0.00	0.07	0.07	0.02
FMR2 (4)	-0.03		<b>0.11</b>	-0.04	0.02	-0.09	-0.08	-0.01	-0.03	-0.02
FMR3 (2)	-0.01	<b>0.11</b>		0.01	<b>0.03</b>	0.01	-0.08	0.03	0.06	0.02
FMR4 (4)	0.07	-0.04	0.01		-0.02	<b>0.33</b>	-0.01	<b>0.16</b>	0.14	<b>0.08</b>
FMR5 (2)	0.01	0.02	0.03	-0.02		-0.02	-0.02	0.01	0.02	0.00
FMR6 (4)	0.01	-0.09	0.01	<b>0.33</b>	-0.02		-0.02	0.04	0.00	0.03
FMR7 (2)	0.00	-0.08	-0.08	-0.01	-0.02	-0.02		0.06	-0.01	-0.02
FMR8 (4)	<b>0.07</b>	-0.01	0.03	0.16	0.01	0.04	<b>0.06</b>		<b>0.23</b>	0.07
FMR9 (4)	0.07	-0.03	0.06	0.14	0.02	0.00	-0.01	0.23		0.06
									Mean	0.03

<sup>a</sup> After von Borries 2008; Bowley (<http://www.plant.uoguelph.ca/research/homepages/sbowley/>)

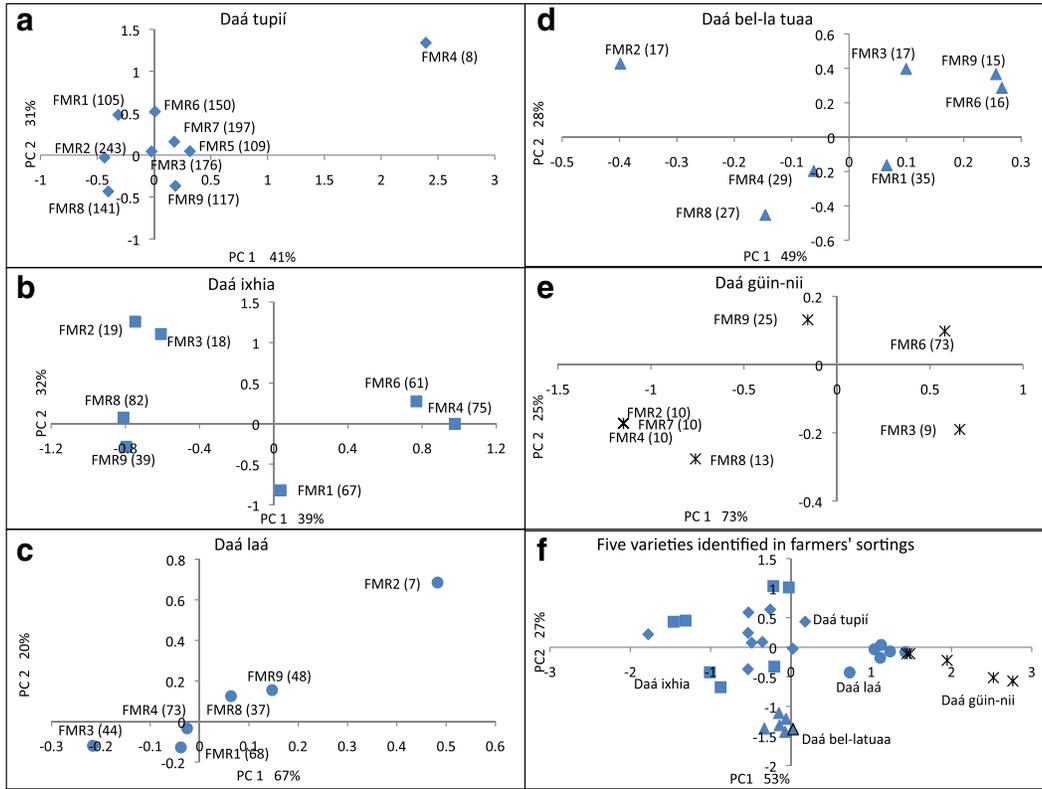
<sup>b</sup> Bold numbers are farmer column maximums of the ARI

<sup>c</sup> Experience-based thresholds for evaluating quality of agreement between farmers based on adjusted Rand index (Steinley, 2004): > 0.9 = excellent, 0.9-0.8 = good, 0.8-0.65 = moderate, < 0.65 = poor.

<sup>d</sup> N=235 for seeds included in comparisons with STRUCTURE-defined classes (*K<sub>i</sub>*)



**Fig. 3.** STRUCTURE bar plot of the membership coefficients for all 235 seeds at  $K = 3$  with the local variety names assigned to each seed by farmers 1–9. Zapotec name translations: *Daá* (bean); *Daá tupií* (whistling on dehiscence); *Daá bel-la tuua* (color of maguey worm); *Daá laá* (leafy); *Daá ixhia* (fibrous net); *Daá güin-nii* (for treating eye pustules); *Daá ix-xiía* (field, or wide); *Daá chix-xí* (mixed); *Daá bayudo* (bay or reddish brown color); *Daá xhidu* (light, ray-like pattern similar to cat eye); *Daá yel-la* (of the *milpa*); *Daá coyuche* (brown).



**Fig. 4.** Scattergrams plotting mean principal components one and two of principal components analysis by variety for five *Phaseolus* varieties commonly identified in the sorting exercise with farmer identity and number of seeds indicated (a-e), and for all five varieties plotted together (f). All PCAs except *Daá güin-nii* and *Daá laá* based on four seed morphology characteristics: weight, perimeter, width:length ratio, B-color value of seed coat. *Daá güin-nii* and *Daá laá* PCAs do not include B-color value due to lack of variation in these varieties. See ESM Table 1 for eigenvector loadings for principal components 1, 2 and 3.

variation in seed morphotypes (Table 4), and that their classifications are based on multiple characteristics. The STRUCTURE species identities based on SSR data also discriminated the morphological variation present in the seed sample (Table 2), a likely reason that farmer-identified varieties organized most of those genetic groupings nonrandomly as well. For example, many farmers' varieties were dominated by one species, especially in the case of *P. vulgaris*, and subspecific partitioning in *P. vulgaris* and *P. coccineus* was nonrandomly distributed in the majority of cases (Table 5, Fig. 2). However, it should not be assumed that farmers are seeking monospecific varieties. Indeed it would be interesting to investigate whether farmer selection is in fact encouraging phenotypic convergence across the three *Phaseolus* species present in many of their higher elevation varieties. That mixing of species is not unusual and was supported by an earlier study of on-farm collections with these same farmers

(Worthington et al. 2012), as well as our finding that one of the seeds in the present study was a *P. vulgaris* x *P. coccineus* hybrid (data not shown).

Still, while the structuring of morphological and genetic variation in farmers' classifications is nonrandom, it is not stringent. Classification errors (Table 4) in both farmer and STRUCTURE groups illustrate the lack of stringency, but also suggest other undocumented characteristics potentially at play. Such classification error may occur even if classification structure is significant when intermediate morphotypes are assigned to classes basically at random. For example, with FMR5 an objective multivariate model can be built for variety groupings based on the relationship between two of the four seed characteristics (WT, WLR; Table 4), but applying this model to predict class membership is not very effective (classification error = 0.29) because it lacks stringency.

There are other possible influences on stringency in this and other classifications. For example, although FMR5 sorted the 300 seeds into only two groups, the name of one of these is *Daá chix-xí* or “mixed beans.” This farmer grew *Daá chix-xí* in both 2007 and 2008, and described it as composed of a number of types. Another factor possibly contributing to low stringency in farmer classifications is that the farmers only had the seed on which to base their classification. However, the majority (75%) of farmers in Santa María Jaltianguis select their planting seed from stored stock. In addition, farmers consciously tolerate variation in a single seed stock. Seven of the 41 seed lot collections made in March 2008 were mixtures of varieties farmers’ identified by individual seed; they reported planting and harvesting these mixtures in their *tierra templada* fields.

Even using the same 300 seeds, agreement among the classifications made by different farmers is weak. There are clearly shared patterns in how farmers organized the *Phaseolus* diversity in this sorting exercise (e.g., Figs. 2–4), but the structure of farmer-identified varieties even as groups regardless of specific names, is varied and idiosyncratic. The variation we observed in each variety identified by different farmers is consistent with the findings of research on farmers’ rice varieties in Maritime Guinea (Barry et al. 2007). Using nuclear genotypes identified with 10 SSR markers to compare accessions of the same variety across farmers, these researchers found that while such accessions shared “the same leading genotype(s) of the variety the frequencies of this (or these) genotype(s) varied between accessions; secondary genotypes were generally specific to each accession” (Barry et al. 2007:1683). Thus, each variety was a collection of different genotypes, with total varietal diversity much greater than that present within any single accession of that variety. In that example and the present *Phaseolus* study, farmer variety names underestimate diversity (c.f. Jarvis et al. 2008: Table 1).

From an etic perspective, variability and low stringency in varietal classifications may be interpreted as a lack of precision. Some studies have shown that when presented with the same predefined seed lot farmers may demonstrate consistency in the name they give it (Sadiki et al. 2006), but our data indicate that even among farmers in the same community they would not

necessarily construct that named variety in the same way. It may be incorrect to assume that farmers are seeking consistent, standardized varieties represented by narrow ideotypes, and so incorrect to expect all farmers to organize available diversity the same way, or that one farmer’s seed lot of a given variety will be the same as another’s, even if they both give the same varietal name when presented with an existing seed lot. Lack of agreement and stringency may reflect individual circumstances and standards. For example, farmers we interviewed (D. Soleri and F. Aragón-Cuevas, unpublished data 2013) reported abandoning distant *tierra caliente* fields, and some *tierra fría* ones, as a result of changing local climate, smaller household size, labor migration and advanced farmer age. One response they described was pooling once distinct bean varieties, much as Hopi farmers did when socioeconomic changes made growing multiple blue maize varieties impractical (Soleri and Cleveland 1993).

Our results indicate farmer-named varieties are empirically meaningful structurings of phenotypic and genetic characteristics, and as such a useful first order tool for quantification of *in situ* crop diversity. However, there is both synonymy, evident in lack of classification agreement, and homonymy with varietal names defining broad domains of phenotypes and genotypes that underestimate diversity present at the community level. Because all farmers in this study started with the same material, we know that the classification variation we observed was due to how those farmers chose to organize the seed sample. Assuming that classifications documented here reflect practice, varietal structure and diversity is individually defined within each farming household’s *Phaseolus* repertoire resulting in substantial intravarietal variation between classifications by different farmers, as reflected in the very low agreement among farmers’ classifications.

The goal of farmers may not be to maintain the same variety across households, but to form a version of a broad type that best fits their own needs and circumstances at one point in space and time. Thus, in both work with farmers and collections of their *Phaseolus* varieties for *ex situ* conservation it should not be assumed that same-named seed lots are redundant units of diversity, and instead include multiple accessions of the same variety from different farmers, repeated over time.

## Acknowledgements

We thank farming households and municipal authorities in Santa María Jaltilanguis; thanks to Professor Eusebio Jiménez Nicolas for translation and correct notation of Zapotec bean variety names; for research assistance J.S. Muruaga-Martínez, Humberto Castro García, Laurentino Juárez Cruz; for advice and editorial comments S.R. Bowley, J. von Braun, D.A. Cleveland; for assistance in the molecular diversity analyses Estevan Quiroga. The suggestions of three anonymous reviewers greatly improved the manuscript. Financial support was provided by NSF (DEB-0409984) to DS, a University of California Mexico-US program (UC MEXUS) to PG, a UC Davis Jastro-Shields award to MW, and a UC Davis Department of Plant Sciences fellowship to MW.

## Literature Cited

- Barry, M. B., J. L. Pham, B. Courtois, C. Billot, and N. Ahmadi. 2007. Rice genetic diversity at farm and village levels and genetic structure of local varieties reveal need for *in situ* conservation. *Genet Resour Crop Evol* 54:1675–1690.
- Berlin, B. 1992. Ethnobiological classification: principles of categorization of plants and animals in traditional societies. Princeton University Press, Princeton, New Jersey.
- Blair, M. W., F. Pedraza, H. F. Buendia, E. Gaitan-Solis, S. E. Beebe, P. Gepts, and J. Tohme. 2003. Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 107:1362–1374.
- Ceccarelli, S. 2012. Landraces: Importance and use in breeding and environmentally friendly agronomic systems. Pages 103–117 in N. Maxted, ed., *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces*. CABI Publishing, Oxfordshire, UK.
- Escalante, A. M., G. Coello, L. E. Eguiarte, and D. Piñero. 1994. Genetic Structure and Mating Systems in Wild and Cultivated Populations of *Phaseolus coccineus* and *P. vulgaris* (Fabaceae). *American Journal of Botany* 81:1096–1103.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611–2620.
- Fisher, D. G. and P. Hoffman. 1988. The adjusted Rand statistic: A SAS macro. *Psychometrika* 53:417–423.
- Gaitan-Solis, E., M. C. Duque, K. J. Edwards, and J. Tohme. 2002. Microsatellite repeats in common bean (*Phaseolus vulgaris*): Isolation, characterization, and cross-species amplification in *Phaseolus* spp. *Crop Science* 42:2128–2136.
- González, R. J. 2001. Zapotec science: farming and food in the northern sierra of Oaxaca. University of Texas Press, Austin.
- Hammer, K. and Y. Morimoto. 2011. Chapter 7: Classifications of infraspecific variation in crop plants. Pages 1–15 in L. Guarino, V. Ramanatha Rao, and E. Goldberg, eds., *Collecting Plant Genetic Diversity: Technical Guidelines - 2011 Update*. Bioversity International, Rome, Italy.
- IAASTD (International assessment of agricultural knowledge, science and technology for development). 2009. Synthesis report with executive summary: a synthesis of the global and sub-global IAASTD reports. Washington, DC ([http://www.agassessment.org/reports/IAASTD/EN/Agriculture%20at%20a%20Crossroads\\_Synthesis%20Report%20\(English\).pdf](http://www.agassessment.org/reports/IAASTD/EN/Agriculture%20at%20a%20Crossroads_Synthesis%20Report%20(English).pdf)).
- INEGI. 2011. Censo de Población y Vivienda 2010. Instituto Nacional de Estadística y Geografía, Mexico, D.F.
- Jarvis, D. I., A. H. D. Brown, P. H. Cuong, L. Collado-Panduro, L. Latournerie-Moreno, S. Gyawali, T. Tanto, M. Sawadogo, I. Mar, M. Sadiki, N. T. N. Hue, L. Arias-Reyes, D. Balma, J. Bajracharya, F. Castillo, D. Rijal, L. Belqadi, R. Ranag, S. Saidi, J. Ouedraogo, R. Zangre, K. Rhrib, J. L. Chavez, D. Schoenu, B. Sthapit, P. De Santis, C. Fadda, and T. Hodgkin. 2008. A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. *Proceedings of the National Academy of Sciences of the United States of America* 105:5326–5331.
- Kwak, M., J. A. Kami, and P. Gepts. 2009. The putative Mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma-Santiago Basin of Mexico. *Crop Science* 49:554–563.
- Llaca, V., A. Delgado Salinas, and P. Gepts. 1994. Chloroplast DNA as an evolutionary marker in the *Phaseolus vulgaris* complex. *Theoretical and Applied Genetics* 88:646–652.
- Peakall, R. and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6:288–295.
- Pritchard, J., M. Stephens, and P. Donnelly. 2000. Inference of population structure using

- multilocus genotype data. *Genetics* 155:945–959.
- Royal Horticultural Society. 2001. Royal Horticultural Society Colour Chart. Botanical color standards. Royal Horticultural Society, London, UK.
- Sadiki, M., D. Jarvis, D. K. Rijal, J. Bajracharya, N. N. Hue, T. C. Camacho-Villa, and L. A. Burgos-May. 2006. Variety names: An entry point to crop genetic diversity and distribution in agroecosystems? in D. Jarvis, C. Padoch, and D. Cooper, eds., *Managing Biodiversity in Agricultural Ecosystems*. Columbia University Press, New York, NY.
- SAS Institute, Inc. 2002–2011. SAS (Statistical Analysis System). Release 9.2. Cary, NC: SAS Institute, Inc.
- Soleri, D. and D. A. Cleveland. 1993. Hopi crop diversity and change. *Journal of Ethnobiology* 13:203–231.
- Steinley, D. 2004. Properties of the Hubert–Arabie Adjusted Rand Index. *Psychological Methods* 9:386–396.
- von Borries, G.F. 2008. Partition clustering of high dimensional low sample size data based on p-values. Doctoral dissertation, Kansas State University, Manhattan, Kansas.
- Weising, K. and R. C. Gardner. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42:9–19.
- Worthington, M., D. Soleri, F. Aragón-Cuevas, and P. Gepts. 2012. Genetic composition and spatial distribution of farmer-managed *Phaseolus* bean plantings: an example from a village in Oaxaca, Mexico. *Crop Science* 52:1721–1735.
- Xu, D. H., J. Abe, J. Y. Gai, and Y. Shimamoto. 2002. Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theoretical and Applied Genetics* 105:645–653.
- Yu, K., S. J. Park, V. Poysa, and P. Gepts. 2000. Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *Journal of Heredity* 91:429–434.
- Zent, S. (10). 1996. Behavioral orientations toward ethnobotanical quantification. Pages 199–239 in Alexiades, MN, ed. *Selected guidelines for ethnobotanical research: a field manual*. Bronx, NY: New York Botanical Garden.