

## USDA-ARS refined pea core collection for 26 quantitative traits

Coyne, C.J.<sup>1</sup>, Brown, A.F.<sup>1</sup>,  
Timmerman-Vaughan, G.M.<sup>2</sup>,  
McPhee, K.E.<sup>3</sup>, and  
Grusak, M.A.<sup>4</sup>

<sup>1</sup>USDA-ARS, WRPIS, Pullman, WA, USA  
<sup>2</sup>Inst. for Crop and Food Res., Lincoln, New Zealand  
<sup>3</sup>USDA-ARS, Grain Legume Genet. and Physiol., Pullman, WA, USA  
<sup>4</sup>USDA-ARS, Children's Nutrition Res. Center, Houston, TX, USA

### Introduction

Creation of core subsets of crop germplasm collections was first suggested by Frankel (3) as a way to efficiently utilize the genetic diversity present within the larger collection. Ideally, core collections represent the genetic diversity of a crop species and its wild relatives (1). Core collections have proven to be a successful way for plant scientists from many disciplines (plant genetics, plant physiology, plant pathology) to first access a subset of germplasm to help refine further exploration of the larger germplasm collection held in trust in public institutions worldwide. Among food legume crops distributed from the National Plant Germplasm System (NPGS) repository located in Pullman, WA, USA, the pea core collection (11) has been used frequently to screen for biotic stress resistance (4, 7), and more recently for mineral nutrient analyses (5). All data generated is publicly available through the Germplasm Resources Information Network (GRIN) (<http://www.ars-grin.gov/npgs/>).

The USDA-ARS *Pisum* germplasm collection currently contains 3918 accessions. The first *Pisum* core collection contained 504 accessions and inclusion was based on geographical origin and flower color and was created using a proportional logarithm model to determine number of accessions per country (geographic origin) (11). Since establishment of the core collection, phenotypic data generated by the repository and cooperators has been entered into the NPGS GRIN database by cooperators. This refined core was created using biomass and related character data (8), seed mineral nutrient composition (5), and seed protein concentration (2). These data allow for the application of multivariate statistical procedures such as cluster analysis to understand the USDA pea core collection diversity for 26 quantitative traits.

The purpose of this study was to investigate the possibility of reduce the size of the USDA pea core to approximately 10% of the *Pisum* accessions using 26 quantitative traits without reducing the trait diversity. Published cores range in size from 5 to 20% of various crop germplasm collections using passport, morphological and/or quantitative traits, typically a mixture of data types (6). A recent example using these three data types is for a chickpea core collection of accessions held in India (14). Given the core concept (3), access to other and/or rare alleles is not diminished by whatever percent is finally chosen as the USDA-ARS pea working collection remains the same at 3918 accessions. The 26 traits selected were based on data availability for a significant portion of the USDA pea core collection of some economically important characters of pea. The refined core will be used for replicated field experiments and laboratory molecular studies of allelic diversity of published pea genes and markers in the collection and to begin to identify significant gaps in the core. This baseline information on the refined pea core may contribute to association mapping (linkage disequilibrium) studies in *Pisum*.

### Materials and methods

#### Plant material and trait data

The set of germplasm accessions used in this analysis was the USDA pea core collection and they can be found <http://www.ars-grin.gov/npgs/> under “Observations” and the descriptor “CORE” (11). Quantitative data on 26 traits measured on the first core are listed under their GRIN descriptor names, and published references are listed in Table 1. The pea core accessions and their quantitative trait data used in this analysis are available at <http://www.ars-grin.gov/npgs/> under “Observations”. Geographic origin and flower color were not included in the analysis.

**Table 1. Quantitative trait data used to reduce the size and redundancy in the USDA-ARS pea core collection entered into the GRIN database (2,5,7).**

Field trait measurements	Number of accessions	Seed trait measurements	Number of accessions
Biomass (kg/ha) <sup>a</sup>	390	Ca <sup>b</sup>	481
Seed yield (kg/ha) <sup>a</sup>	389	Mg <sup>b</sup>	481
Straw yield (kg/ha) <sup>a</sup>	389	K <sup>b</sup>	481
Harvest index (yield/biomass) <sup>a</sup>	389	P <sup>b</sup>	481
Days to first flower (50% with open flowers) <sup>a</sup>	390	Fe <sup>b</sup>	481
Days maturity <sup>a</sup>	390	Zn <sup>b</sup>	481
Reproductive days <sup>a</sup>	390	Mn <sup>b</sup>	481
Node to first flower <sup>a</sup>	390	Cu <sup>b</sup>	481
Height to first flower node <sup>a</sup>	390	Ni <sup>b</sup>	481
Height at maturity <sup>a</sup>	390	B <sup>b</sup>	458
Seed weight (g/100 seed) <sup>a</sup>	388	Mo <sup>b</sup>	481
Seed & pod dry weight partitioning (greenhouse) <sup>b</sup>	482	Seed positions (greenhouse) <sup>b</sup>	479
Seed dry weight (greenhouse) <sup>b</sup>	482	Seed protein concentration (greenhouse) <sup>c</sup>	482

<sup>a</sup> McPhee and Muehlbauer, 2001.<sup>b</sup> Grusak et al. 2004.<sup>c</sup> Coyne et al. 2005.

#### Statistical methods

The variables were standardized using the STAND module of NTSYSpc (9). The linear transformation used is of the form:

$$y' = [(y - \hat{y})/\sigma_y^2] - c$$

where  $\hat{y}$  = the mean of all  $y$  values,  $\sigma_y^2$  = the standard deviation of all  $y$  values, and  $c$  = a constant added after the above operations have been performed (9). Dissimilarity coefficients for interval measure (quantitative) data were generated using the SIMINT module of NTSYSpc. The parameter of average taxonomic distance (DIST) module of NTSYSpc was used to generate the matrix.

A dendrogram was generated from the sequential, agglomerative, hierarchical, and nested (SAHN) clustering method using the unweighted pair-group method arithmetic average (UPGMA) (12) using the NTSYSpc SAHN module. The Euclid coefficient (EUCLID module) was used to generate the dissimilarity matrix in Euclidean distances for accessions in the new core. The cut off point for the distance value of closely scored accessions based on the 26 trait measurements was set to result in a core of approximately 310 accessions. Random numbers were assigned to accessions in the same cluster and used to select the accession from each group for the refined core.

#### Comparison of core collections

The means of the original core and the refined core were compared for all 26 traits using ANOVA (Proc GLM) and Tukey's Studentized Range (HSD) modules of SAS (10). The comparison of variances between each of the 26 trait data of the original pea core with the refined pea core was determined using ANOVA (Proc GLM) and Levene's Test for homogeneity of variances modules of SAS (10).

#### Results and Discussion

The purpose of this study was to investigate redundancy in the USDA pea core for 26 quantitative traits and to use the relationships discovered to create a refined core for future allelic diversity studies on economic traits of pea. An underlying assumption was that a core of 504 (~14%) selected in 1995 from the approximately 3,500 pea accessions in the collection at that time may over-represent the collection for these 26 quantitative traits. Additionally, 453 of the ~3500 accessions are Marx Genetic Stocks created by backcrossing to the same parent, so the original core is closer to ~17% of the 1995 collection. Further

**Table 2. Comparison of means and variances between the original geographic core and the refined pea core using 26 quantitative traits indicates that genetic diversity for these traits was maintained (i.e., no significant loss of genetic variance in each trait).**

Traits	Means <sup>a</sup>			Variances <sup>b</sup>				CV (%) <sup>c</sup>	Range (%) <sup>d</sup>
	Original core	Refined core	$\alpha = 0.05$	Original core	Refined core	F value	p		
Biomass (kg/ha)	3330.1	3354.9	NS <sup>e</sup>	1132.5	1182.9	0.62	0.433	34.5	100
Seed yield (kg/ha)	1309.3	1308.7	NS	520.2	544.3	0.59	0.441	40.5	100
Straw yield (kg/ha)	2020.5	2045.8	NS	684.2	718.7	0.80	0.371	34.4	100
Harvest index	38.5	38.2	NS	7.0	7.5	0.72	0.395	18.8	100
Days to first flower	54.6	54.6	NS	5.8	5.9	0.14	0.704	10.7	100
Days to maturity	86.1	86.5	NS	9.8	9.9	0.21	0.643	11.4	100
Reproductive days	31.6	31.9	NS	7.5	7.7	0.23	0.634	23.9	100
Node first flower	15.3	15.2	NS	2.8	2.9	0.40	0.525	18.8	100
Height to first flower node	50.3	50.2	NS	15.1	15.7	0.57	0.452	30.5	100
Height at maturity	65.6	64.8	NS	18.4	18.8	0.14	0.705	28.4	96.5
Seed weight (g/100 seed)	16.2	16.3	NS	5.5	5.5	0.04	0.852	33.7	100
Seed & pod dw partitioning	88.0	88.1	NS	4.4	4.6	0.11	0.735	5.1	100
Seed dry weight	18.6	18.9	NS	7.4	7.4	0.02	0.886	39.5	100
Ca (ppm)	773.8	810.7	NS	321.0	359.9	2.06	0.152	42.7	100
Mg (ppm)	1693.5	1682.6	NS	168.9	183.2	1.19	0.276	10.3	100
K (ppm)	12622.5	12412.4	NS	1657.4	1673.2	0.02	0.887	13.2	100
P (ppm)	5163.6	5035.7	NS	953.8	999.2	0.87	0.350	19.0	100
Fe (ppm)	50.0	51.0	NS	11.7	12.1	0.35	0.552	23.5	91.7
Zn (ppm)	41.9	42.2	NS	11.5	11.7	0.07	0.798	27.5	100
Mn (ppm)	15.9	16.4	NS	4.5	4.8	0.28	0.594	28.8	100
Cu (ppm)	4.4	4.4	NS	1.7	1.8	0.21	0.646	39.4	100
Ni (ppm)	2.4	2.6	NS	1.6	1.8	0.64	0.426	69.0	100
B (ppm)	7.7	7.8	NS	1.6	1.6	0.14	0.709	20.3	100
Mo (ppm)	23.7	23.0	NS	8.4	8.0	1.00	0.319	35.2	82.9
Seed positions	5.7	5.7	NS	1.0	1.0	0.14	0.709	17.4	100
Seed protein concentration(%)	24.1	24.0	NS	3.5	3.5	0.03	0.874	14.7	100

<sup>a</sup>Differences between means were tested by Tukey's Studentized range test (9).

<sup>b</sup>Variances tested using Levene's test for homogeneity (9).

<sup>c</sup>CV = coefficient of variation calculated from ANOVA of the 26 traits between the original core and the refined core.

<sup>d</sup>% range was calculated from the minimum and maximum trait values of the original core and the refined core.

<sup>e</sup>NS = non-significant at the  $\alpha = 0.05$  level (9).

phenotypic and genotypic studies would need to be conducted to actually determine if this is the case. The 310 accessions included in the refined core collection are a subset of the 504 accessions in the core collection.

Comparison of means and variances indicates no significant loss of genetic variation for 26 traits between the original core and the refined core (Table 2). The dendrogram of the refined USDA core collection can be found at <http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492806>, under "Dendrogram of the refined core (Power Point)".

Interestingly, *Pisum sativum* L. subsp. *abyssinicum*, known to be very similar at the molecular level (13), was found grouped closely together using these 26 quantitative traits. The original USDA ARS core lacked representatives from *Pisum fulvum*. We plan to add accessions to fill this obvious gap in the refined core with *Pisum fulvum* and to capture additional diversity of traits. Since 1995, we have added over 400 new accessions, including subspecies not in the 1995 collection from other germplasm collections and new plant explorations to Turkey and central Asia. Examples of additional traits identified would be accessions with improved resistance to

Aphanomyces root rot resistance (7) and Fusarium root rot (4). Additionally, we plan to include accessions representing the extremes found for Mo (ppm) and Fe (ppm) (Table 2). We are exploring the core and refined core genetic diversity at the molecular level and will use this information to further refine the USDA pea core collection.

As Brown (1) predicted, “the composition of a core will change with time, as new data, new material, or requirements come along”. A core collection, especially a heavily used collection such as the USDA pea core collection, will remain useful if it also remains dynamic. Both the original USDA pea core and the refined pea core are found on the GRIN web site under the Observations and Descriptors CORE and REFINED CORE (<http://www.ars-grin.gov/npgs/>).

*Acknowledgments:* USDA-ARS Project 5348-21000-020-00 (Coyne) and USDA Foreign Agriculture Service Project 5348-21000-020-03 (Coyne and Timmerman-Vaughan).

1. Brown, A.H.D. 1989. In: Brown, A.H.D., Frankel, O.H., Marshall, D.R. and Williams, J.T. (eds.) *The Use of Plant Genetic Resources*. Cambridge University Press, Cambridge, UK, pp 136-155.
2. Coyne, C.J., Grusak, M.A., Razai, L., and Baik, B.-K. 2005. *Pisum Genetics* 37: XX-XX
3. Frankel, O.H. 1984. In: Arber, W., Llimensee, K., Peacock, W.J. and Starlinger, P. (eds.) *Genetic Manipulations: Impact of Man and Society*. Cambridge University Press, Cambridge, England, pp 161–170.
4. Grünwald, N. J., Coffman, V.A. and Kraft, J.M. 2003. *Plant Disease* 87: 1197-1200.
5. Grusak, M.A., Burgett, C.L., Knewton, S.J.B., Lopéz-Millán, A.-F., Ellis, D.R., Li, C.-M., Musetti, V.M., and Blair, M.W. 2004. *Proceedings of the 5<sup>th</sup> AEP-2<sup>nd</sup> ICLGG Conference*, pp 37-38.
6. Johnson, R.C. and T. Hodgkin. 1999. *Core Collections for Today and Tomorrow*. International Plant Genet. Resources Inst., Rome, Italy.
7. Malwick, D.K., and Percich, JA. 1999. *Plant Disease* 83: 51-54.
8. McPhee, K.E. and Muehlbauer, F.J. 2001. *Genetic Res. Crop Evol.* 48: 195-203.
9. Rohlf, F.J. 2000. *NTSYSpc: Numerical Taxonomy and Multivariate Analysis System*, version 2.11. Exeter Software, NY.
10. SAS 9.1. 2002-2003. SAS Institute, Cary, NC, USA.
11. Simon, C.J. and Hannan, R.M. 1995. *HortScience* 30: 907.
12. Sneath, P.H.A. and Sokal, R.R. 1973. *Numerical Taxonomy*. W.H. Freeman and Co., San Francisco, USA.
13. Weeden, N.F. and Wolko, B. 2001. *Pisum Genetics* 33: 21-25.
14. Upadhyaya, H.D., Bramel, P.J., and Singh S. 2001. *Crop Sci* 41: 206-210.