Productive, qualitative and seed image analysis traits of guar (*Cyamopsis tetragonoloba* (L.) Taub)

F. Gresta[1], C. Santonoceto[1], G. Ceravolo[1], C. Formantici[2], O. Grillo[3]*, C. Ravalli[3], G. Venora[3]

[1] *Dipartimento di Agraria, Università Mediterranea di Reggio Calabria, Loc. Feo di Vito, 89122 – Reggio Calabria, Italy*
[2] *Laboratory of Biotechnology Lamberti spa, Via Piave 18, 21041 Albizzate (VA), Italy*
[3] *Stazione Sperimentale di Granicoltura per la Sicilia, Via Sirio 1, 95041 Borgo Santo Pietro, Fraz. Caltagirone*

Supplemental Table S1 - Haralick's descriptors measured as reported in Haralick et al. (1973).

	Feature	Equation
Har 1	Angular second moment	$\sum_i \sum_j p(i,j)^2$
Har 2	Contrast	$\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}, |i,j| = n$
Har 3	Correlation	$\dfrac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ where μ_x, μ_y, σ_x and σ_y are the means and the standard deviations of p_x and p_y.
Har 4	Sum of square: variance	$\sum_i \sum_j (i - \mu)^2 p(i,j)$
Har 5	Inverse difference moment	$\sum_i \sum_j \dfrac{1}{1 + (i-j)^2} p(i,j)$
Har 6	Sum average	$\sum_{n=2}^{2N_g} i p_{x+y}(i)$ where *x* and *y* are the coordinates (row and column) of an entry in the co-occurrence matrix, and $p_{x+y}(i)$ is the probability of co-occurrence matrix coordinates summing to *x+y*.
Har 7	Sum variance	$\sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$
Har 8	Sum entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} = f_8$
Har 9	Entropy	$-\sum_i \sum_j p(i,j) \log[p(i,j)]$

Har 10	Difference variance	$$\sum_{n=0}^{N_g-1} i^2 p_{x-y}(i)$$
Har 11	Difference entropy	$$-\sum_{n=0}^{N_g-1} p_{x-y}(i) \, log\{p_{x-y}(i)\}$$

The basis for these features is the gray-level co-occurrence matrix (G in equation 1). This matrix is square with dimension Ng, where Ng is the number of gray levels in the image. Element [i,j] of the matrix is generated by counting the number of times a pixel (p) with value i is adjacent to a pixel with value j and then dividing the entire matrix by the total number of such comparisons made. Each entry is therefore considered to be the probability that a pixel with value i will be found adjacent to a pixel of value j.

$$G = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1,N_g) \\ p(2,1) & p(2,2) & \cdots & p(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g,1) & p(N_g,2) & \cdots & p(N_g,N_g) \end{bmatrix} \tag{1}$$

Supplementary Information 2 - List of morphometric features measured on seeds, excluding the mean seed weigh, the Elliptic Fourier Descriptors (EFDs) calculated according to Hâruta (2011) and the Haralick's descriptors reported in Supplementary Information 1.

Feature		Description
A	Area	Seed area (mm^2)
P	Perimeter	Seed perimeter (mm)
P_{conv}	Convex Perimeter	Convex perimeter of the seed (mm)
P_{Crof}	Crofton Perimeter	Crofton perimeter of the seed (mm)
P_{conv}/P_{Crof}	Perimeter ratio	Ratio between convex and Crofton's perimeters
D_{max}	Max diameter	Maximum diameter of the seed (mm)
D_{min}	Min diameter	Minimum diameter of the seed (mm)
D_{min}/D_{max}	Feret ratio	Ratio between minimum and maximum diameters
Sf	Shape Factor	Seed shape descriptor = (4 x π x area)/perimeter^2 (normalized value)
Rf	Roundness Factor	Seed roundness descriptor = (4 x area)/(π x max diameter^2) (normalized value)
Ecd	Eq. circular diameter	Diameter of a circle with equivalent area (mm)
F	Fiberlength	Seed length along the fiber axis
C	Curl degree	Ratio between D_{max} and F
$Conv$	Convessity degree	Ratio between P_{Crof} and P
Sol	Solidity degree	Ratio between A and convex area
Com	Compactness degree	Seed compactness descriptor = [$\sqrt{(4/\pi)}\,A$]/D_{max}
EA_{max}	Maximum ellipse axis	Maximum axis of an ellipse with equivalent area (mm)
EA_{min}	Minimum ellipse axis	Minimum axis of an ellipse with equivalent area (mm)
R_{mean}	Mean red channel	Red channel mean value of seed pixels (grey levels)
R_{sd}	Red std. deviation	Red channel standard deviation of seed pixels
G_{mean}	Mean green channel	Green channel mean value of seed pixels (grey levels)
G_{sd}	Green std. deviation	Green channel standard deviation of seed pixels
B_{mean}	Mean blue channel	Blue channel mean value of seed pixels (grey levels)
B_{sd}	Blue std. deviation	Blue channel standard deviation of seed pixels

H_{mean}	Mean hue channel	Hue channel mean value of seed pixels (grey levels)
H_{sd}	Hue std. deviation	Hue channel standard deviation of seed pixels
L_{mean}	Mean lightness ch.	Lightness channel mean value of seed pixels (grey levels)
L_{sd}	Lightness std. dev.	Lightness channel standard deviation of seed pixels
S_{mean}	Mean saturation ch.	Saturation channel mean value of seed pixels (grey levels)
S_{sd}	Saturation std. dev.	Saturation channel standard deviation of seed pixels
D_{mean}	Mean density	Density channel mean value of seed pixels (grey levels)
D_{sd}	Density std. deviation	Density channel standard deviation of seed pixels
S	Skewness	Asymmetry degree of intensity values distribution (grey levels)
K	Kurtosis	Peakness degree of intensity values distribution (densit. units)
H	Energy	Measure of the increasing intensity power (densitometric units)