EWERU Library # 450

Automated Sorting of Glabrous Versus Pubescent Annual Canarygrass Seeds

T. Pearson, D. Knievel, P. Hucl

Abstract. *An automated imaging system was developed to identify and sort annual canarygrass seeds with the glabrous trait in order to screen bulk amounts of seed to segregate populations for this desirable trait. Glabrous seed is preferred since it does not have the irritating silica hairs that pubescent seeds do. Glabrous seeds may also have potential as a food product. The automated system inspected seeds at a rate of approximately four seeds per second (~100 g/h) and had an average accuracy of 90%. The system performed consistently over time across five different cultivars from two growing seasons at four different locations. Cost for all of the parts in the system was approximately $2500, and therefore, was affordable for breeding programs developing glabrous canaryseed and perhaps other glabrous seeds.*

Keywords. *Imaging, Separating, Silica, Value-added.*

Silica hairs from pubescent hulled annual canarygrass seeds (*Phalaris canariensis*) have been an irritant during harvesting and handling of the seed and have been linked to cancer in humans and laboratory animals (O' Neill et al., 1986). While canaryseed (the trade name for annual canarygrass) has primarily been used as birdfeed, it has high nutritional value and could be considered for human food if the pubescent hulls were not present (Abdel-Aal et al., 1997). Glabrous lines were developed by mutating seed of a pubescent cultivar (Hucl et al., 2001). The glabrous trait was simply inherited and was recessive to the pubescent phenotype (Matus-Cadiz et al., 2003). The availability of glabrous cultivars create the opportunity to increase canaryseed use in food products and reduce the many risks from inhaling broken hairs of pubescent seed during harvest and handling. For example, glabrous seed may be used as a substitute for sesame seed, since sesame seed can cause severe allergies in some people (Gangur et al., 2005; Beyer et al., 2007). During variety development, glabrous canaryseed is crossed with pubescent canaryseed to improve agronomics. This requires segregation of pubescent/glabrous seeds in the subsequent generations. A sorter is then needed to separate the two, in order to develop sufficient glabrous seed stocks. Panicles, or individual seeds, from segregating populations currently need to be examined under a microscope to select the glabrous phenotype. This is a laborious, time-consuming task. Glabrous and pubescent seeds have similar color, size, and morphological features. Only upon examination with a low power microscope, can the hairs on the pubescent seeds be detected. It was the objective of this research to develop a low-cost sorting machine that could separate glabrous and pubescent seeds automatically using machine vision. The objective of this research was to develop a low-cost sorting system for glabrous and pubescent canarygrass seeds. Humans are about 95% accurate in separating these seeds but cannot perform this task for more than a few hours a day and can only inspect seeds at a rate of about 20 per min. Successful completion of this objective will allow breeders to test many more early populations which can reduce the time it takes to develop a commercially viable glabrous variety.

Submitted for review in March 2010 as manuscript number BE 8472; approved for publication as a Technical Note by the Biological Engineering Division of ASABE in April 2011.

The authors are **Tom Pearson, ASABE Member Engineer,** Agricultural Engineer USDA ARS, Manhattan, Kansas; **Donna Knievel,** Research Officer and **Pierre Hucl,** Professor University of Saskatchewan, Crop Development Centre, Saskatoon, Saskatchewan, Canada. **Corresponding author:** Tom Pearson, USDA ARS, 1515 College Ave., Manhattan, KS 66502; phone: 785-776-2729; e-mail: thomas.pearson@ars.usda.gov.

Materials and Methods

Imaging and Sorting System Hardware

Figure 1 depicts the machine-vision system linked to a personal computer for inspection of seed. The seeds were dropped from a vibratory feeder (FTOC, FMC Corp., Homer City, Pa.) near the center of a rotating turntable. The turntable was made of glass, 29 cm in diameter, and rotated at four rpm. The seeds were gradually diverted to the outside perimeter of the turntable by a Teflon ring (fig . 1), causing the seeds to separate and form a single line of seeds. The seeds were diverted out to a radius of 27 cm where the linear speed was approximately 5.6 cm/s. The vibratory feeder was adjusted in order to drop approximately 4 seeds/s onto the turntable. When the seeds reached the outside of the turntable, they were usually separated by about 1.5 cm with a range of 0.5 to 2.5 cm. This seed gap was adequate to image only one seed at a time and allowed reliable diversion of the seed with a short burst of air from an electronic air valve (35A–AAA–DDBA–1BA, Mac Valves, Inc., Wixom, Mich.).

A monochrome camera (#DMK21BF04, The Imaging Source Sommerstrabe, Germany) with a total of 640 × 480 pixels was used. The camera transferred images to the computer via an IEEE 1394 cable in direct memory access (DMA) mode using the DirectX application programming interface (DirectX version 9.0, Microsoft Corp. Redmond, Wash.). The camera used a 35-mm C-mount lens (HF35HA-1B, Fujinon, Japan). A 12.7-mm extension tube

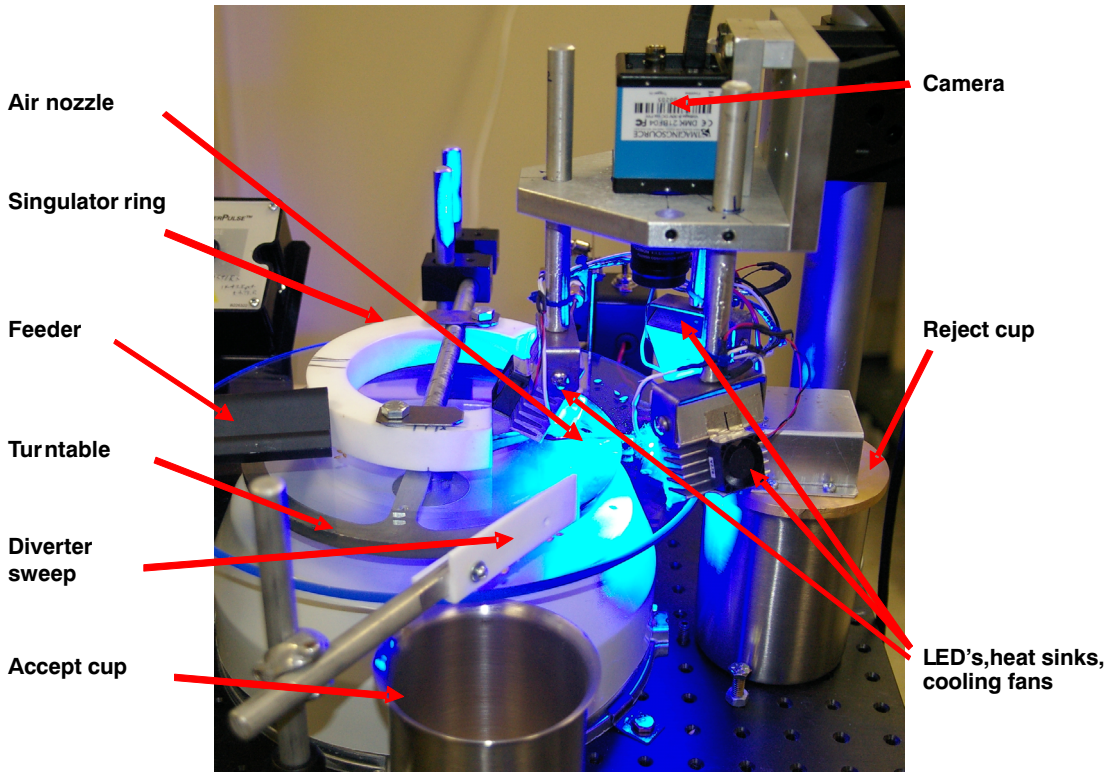

Figure 1. Picture of prototype canaryseed sorting system.

was mounted between the camera body and lens to facilitate close focusing and high magnification. The camera body was mounted 14 cm above the turntable and the resulting image resolution was approximately 82 pixels/mm. At this resolution, annual canarygrass seeds would be approximately 200 pixels wide and 450 pixels long.

The seeds were illuminated with three high power blue (peak emission at 465 nm) light emitting diodes (LED) (B42182, Seoul Semiconductor, Korea). These LEDs had a maximum forward current of 1000 mA, which was provided by a constant current source LED driver (TLM4036DC-1000, Triad Magnetics, Corona, Calif.). Each LED was coupled to a spot lens with a 5° viewing angle (OP005, Dialight, Farmingdale, N.J.) that focused the light onto an area slightly larger than the annual canarygrass seed. Blue LEDs were chosen since the shorter wavelength blue light was more prone to scattering on the hairs than red, green, or infrared light. The LED's had a rated luminus flux of 80 lm, and rated life expectancy of 35,000 h, thus after initially "burning in" for 24 h, per the manufacturers specifications, the light intensity of the LED's will remain virtually constant over long periods (~1000 h). The LEDs were placed close to the turntable and were positioned approximately 120° apart and 3 cm from the spot where kernels were imaged.. This allowed the light emitted from the LEDs to be almost perpendicular to an imaginary line connecting the seed and camera and maximized the imaging of scattered light from the hairs of pubescent seeds. The shutter speed of the camera was set at a constant 1/2500s with a gain of 500 (the range for this camera was 1 to 1000). The lens aperture was stopped down to f12 so the depth of field would be large enough to have the whole seed in focus.

Image Processing

Before performing any sorting, the camera captured a blank image without any seeds in the field of view. The first image processing step for distinguishing glabrous and pubescent seeds was to subtract the blank image, acquired prior to sorting, from an image containing a seed. This helps to segment the kernel from the background by removing dark current noise from each pixel and removes any static objects in the background that might be brighter than the background. When the intensity between pixels from the seed image and blank image was greater than 30, these pixels were considered to be part of the seed. If the difference was less than or equal to 30, the pixels were set to zero and considered background. This enabled nearly flawless segmentation of the seed from the background. The threshold level was determined by trial and error before experiments began, but was kept the same for every subsequent experiment.

The camera was set to capture 30 frames/s. The PC was programmed to inspect each image for the presence of a seed in the center of the image frame. This was accomplished by counting pixels above the threshold along the center line and column of the image. If the counts for the center line and column exceeded 50 pixels, then a seed was considered to be centered in the image. Then, further processing was performed. After processing, the computer ignored the next two images so that the seed would pass out of the center of the imaging area and would not be re-imaged.

To highlight pubescent hairs, every pixel of the monochromatic image was high pass filtered by convolution with two 5×5 compass kernels, $h(u,v)$ and $k(u,v)$, as shown in equations 1 and 2:

$$h(u,v) = \begin{bmatrix} -2 & -2 & -2 & -2 & -2 \\ -1 & -1 & -1 & -1 & -1 \\ 6 & 6 & 6 & 6 & 6 \\ -1 & -1 & -1 & -1 & -1 \\ -2 & -2 & -2 & -2 & -2 \end{bmatrix} \quad (1)$$

and

$$k(u,v) = \begin{bmatrix} -2 & -1 & 6 & -1 & -2 \\ -2 & -1 & 6 & -1 & -2 \\ -2 & -1 & 6 & -1 & -2 \\ -2 & -1 & 6 & -1 & -2 \\ -2 & -1 & 6 & -1 & -2 \end{bmatrix} \quad (2)$$

The values of these kernels were chosen by trial and error. Less high pass filtering did not highlight enough hairs while higher frequency filtering introduced other artifacts in the image that made detection of pubescent hairs problematic. For each pixel, the maximum value from the two convolutions computed by equations 1 and 2 above were saved to generate a new filtered image for further processing.

Figure 2 displays an unprocessed image from the camera, and figure 3 displays a filtered image. The filters highlighted hairs that were either horizontal or vertical in the original image. The hairs usually were aligned parallel to the long axis of the seed, and the seeds were usually aligned vertically in the image.

A two-dimensional histogram was then constructed using the filtered image and their corresponding intensities from the original image. The original intensity image was scaled from 8-bit intensity to 4 bit (16 levels), and the high pass filtered image was scaled to 5 bit (32 levels) for the histogram. Thus, the histogram resulted in 16 × 32 = 512 bins, or features, that could possibly be used for discrimination of glabrous and pubescent seeds.

In addition to the two-dimensional histogram features, the gaps along image rows and columns between brighter spots in the filtered image were computed. To compute these gaps, the filtered image was thresholded at an intensity of 50 and the distance along an image row or column between spots above the threshold value were then computed. The average gap and the number of gaps were recorded. Pubescent seeds tended to have more gaps at shorter intervals than glabrous seeds did (fig. 4). Thus, resulting in a total of 514 features computed for each image that had a seed centered in it.

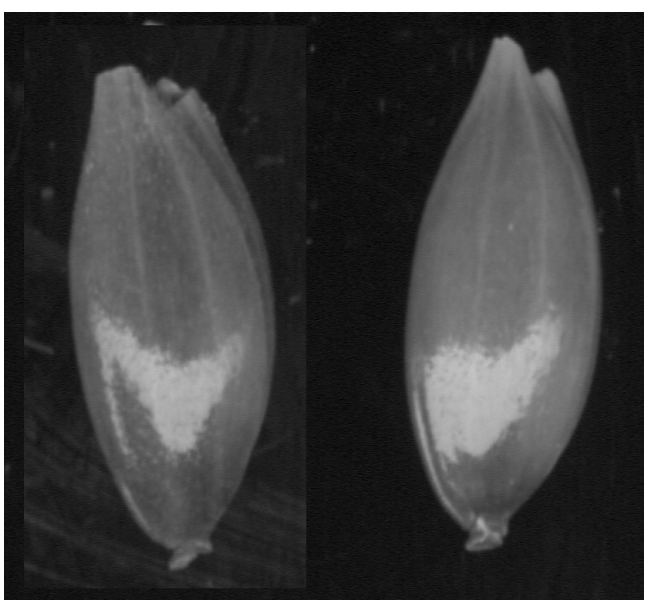

Figure 2. Close-up image of pubescent (left) and glabrous (right). Note the small white specs on the pubescent seed.

Figure 3. Band pass filtering to highlight specks on pubescent seeds.

The computer used a dual core AMD X2 Athlon 6000+ processor with 2GB of 800-MHz DDR2 SDRAM and Windows XP Professional operating system for real-time image processing. All of the image processing steps outlined here required 3 to 4 milliseconds on this computer. Integer math was used to speed up processing; but otherwise, no other special instructions such as discrete math or multiple executions were used.

Classification and Training

Prior to sorting canaryseed samples, the image sorter algorithm was trained using data from 200 typical glabrous seeds and 200 pubescent seeds. For training, the glabrous seeds were CDC Togo variety and the pubescent seeds were of the Keet variety. The training set comprised a mixture from all growing locations. A linear discriminant analysis method (Haff and Pearson, 2006) was used as the classification method to select a small subset of features from the pool of 514 image features. The data was randomly divided into equal training and validation sets. The training set was used to compute the group mean and covariance of the feature subsets to determine the discriminant function parameters. The validation set was used to test the classifications. An exhaustive search was used to test all possible one-, two-, and three-feature combinations. The best discriminating combination was selected for use in sorting. The whole training process took less than 5 min to perform.

Sorter Tests

The sorter was tested on three glabrous cultivars (CDC Maria, CDC Togo, and CDC Bastia) and two pubescent cultivars (Cantate and Keet). Four replicates were grown

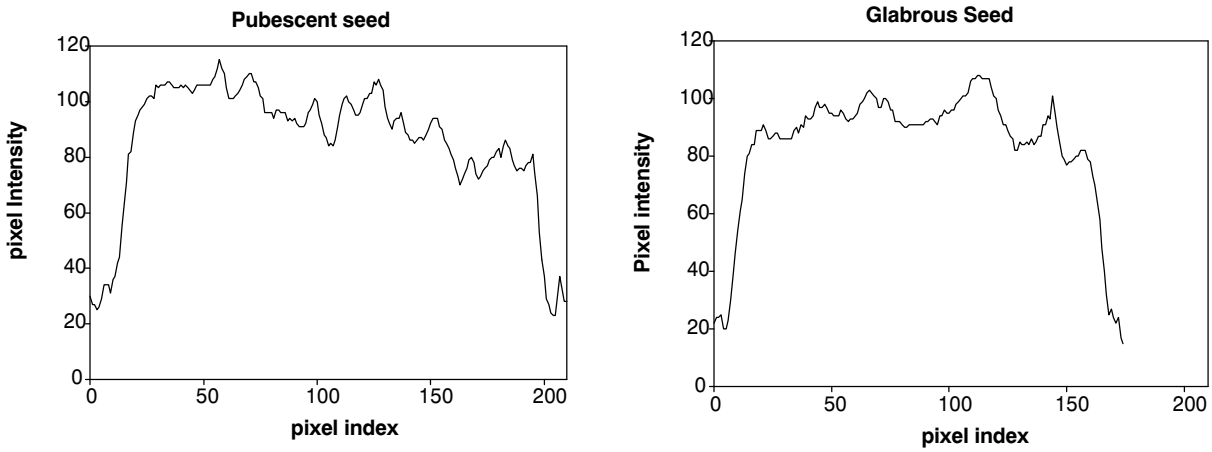

Figure 4. Pixel intensity plots across one image line of a pubescent (left) and glabrous seed (right) image. Note that there are many more small peaks and valleys in the pubescent seed compared with glabrous seed. This can be easily quantified with some more mathematical analysis.

over two years at four different locations: University of Saskatchewan Kernen Crop Research Farm (2007 harvest), Agri and Agrifood Food Canada (AAFC) Swift Current Research Station (2007 harvest), Viterra Watrous Research Farm (2008 harvest), and AFFC Melfort Research Station (2008 harvest). Thus, a total of 20 samples of seed comprising five cultivars and four test sites were studied.

Testing was first performed on pure samples (10 g each, about 1400 seeds) of all cultivars from all locations. This enabled a comparison of the accuracy for each cultivar from each location. Each location/cultivar test was repeated 10 times. Additionally, a mixture of 48 g each of Keet and Cantate pubescent seeds were prepared with 2 g each of CDC Maria, CDC Togo, and CDC Bastia glabrous seeds to form a 100-g mixture of seed from all cultivars containing 96% pubescent and 4% glabrous seed. Only seeds from the Kernen location were used for this test. The sorter was tested over a 5-day period, 10 times each day, with this mixture of pubescent and glabrous seeds used to determine actual sorting accuracy with a mixed sample and sorter stability over time. The sorting test was performed with the glabrous seed being rejected by the air valve since they were the smaller portion. The pubescent seed was allowed to drop off the turntable after moving out of the field of view of the camera. After sorting the entire 100-g sample, 10-g portions of the accept, and all of the reject streams were inspected for accuracy under a 20X magnifying glass.

It is important to note that the sorter was trained only once before all testing began. The LEDs on the sorter remained powered on for the entire week to prevent errors from the LEDs warming up. Other studies using similar LED's for illumination show that the light emission intensities do not change noticeably over several days or weeks. They do require about 30 min to warm up but otherwise can be turned off and on and will emit light at consistent intensities so that the sorter does not need to be re-trained more than once a week but probably less frequent than that. (Pearson, 2010).

Results

The set of three features that gave the best classification accuracy between the Keet (pubescent) seeds and CDC Togo (glabrous) seeds were the average gap size between bright blobs in the filtered image, the number of gaps between white blobs in the filtered image, and a count of pixels having a medium intensity (128 to 143) as well as a medium high pass filter value (112 to 127). These three features had an accuracy of 96% for pubescent seed identification and 88% for glabrous seed identification. Pubescent seeds had lower average gaps between white blobs since the spacing of the hairs was usually only a few pixels. Glabrous seeds had lower numbers of such gaps. Areas on high pass filtered images of glabrous seeds with brighter intensities were usually larger continuous blobs due to very bright spots on the image. The bright spots were most likely caused by direct reflection or portions of hull still attached to the seed. Finally, pubescent seeds had higher counts of pixels with moderate intensities from the original and filtered images. The pubescent seeds had many more bright spots in the filtered images due to light scattered from the hairs.

The best single feature was the number of gaps between white blobs in the filtered image. The best pair of two features was the number of gaps and count of pixels having a medium intensity as well as a medium high pass filter value. However, the average classification accuracy from only one feature and two features was 77% and 85%, respectively. Both are substantially lower than the average of 90% accuracy when three features were selected.

Examination of glabrous seed images from the training set that were misclassified showed that approximately half of the glabrous seeds that were incorrectly classified had cracked pericarp or pieces of hull still attached to them, typical of lab cleaned samples. The seeds with cracked pericarp probably will not germinate and would not be a loss if grouped with pubescent seeds. Glabrous seeds that were not completely hulled and classified as pubescent were a loss; but methods to more completely remove hulls from the seeds will likely improve sorting results for seed selection.

Average classification accuracies of pure seed samples inspected by the sorting system are shown in table 1. Analysis of variance did not detect any significant difference ($p = 0.05$) between average accuracies for locations (i.e. growing environment). Furthermore, there was no significant difference among the three glabrous cultivars or between the two pubescent varieties.

Table 1. Average sorting accuracies for the five canaryseed varieties tested.

Variety	Average Accuracy (%)	Standard Deviation (%)
CDC Maria (glabrous)	87	2.1
CDC Togo (glabrous)	90	1.6
CDC Bastia (glabrous)	88	1.9
Keet (pubescent)	94	2.1
Cantate (pubescent)	93	2.0

Sorter testing results closely matched those of the training accuracies. From the 5-day sorter test, where a mixture of glabrous and pubescent seed was used, the resulting sorting accuracy was 94% for the pubescent seeds and 88% for the glabrous seeds. Again, approximately half of the glabrous seeds that were grouped with the pubescent seeds either had a cracked pericarp or were partially hulled. Some of the glabrous cultivar samples were brighter or more yellowish in color than other seeds. As discussed earlier, this did not affect sorting accuracy. The monochrome image and high-pass filtering apparently were not affected greatly by seed color, as expected. Thus, the sorting method should be robust across different cultivars and production environments.

Testing of sorter accuracy over a 5-day period showed no appreciable change in accuracy. All accuracies for pubescent and glabrous seed were within 3.5% of each other, and no trends, up or down, were observed in the sorting accuracies. The average accuracy for glabrous and pubescent seeds was 87% ± 3.8% and 93% ± 4.4% at the 95% confidence level, respectively. Thus the range of sorting accuracies for all samples fell within, the 95% confidence interval for glabrous and pubescent seeds. In comparison, humans are about 95% accurate in separating these seeds but cannot perform this task for more than a few hours a day and can only inspect seeds at a rate of about 20/min; less than 10% of the rate of this automated system. This allows breeders to test many more early populations in their breeding program which can reduce the time it takes to develop a commercially viable glabrous variety. For this test, the sorter was trained once at the beginning of the test and then tested each day for the next five days. LED lights are known for consistency over time, which is probably one reason for the consistent sorting results. These accuracies were slightly less than what was achieved with the training set and inspection of pure samples because of some material handling errors during diversion of kernels with the air valve. About 1.5% of the pubescent seeds were diverted because they were positioned too close to a glabrous seed that was being diverted. Also, about 1% of the glabrous seeds were not correctly diverted even though they were correctly classified as glabrous. Training the system only took a few minutes, if training samples were readily available. Thus, even if the sorter needed to be trained every week, it should not diminish its usefulness for this application.

Conclusion

An automated system that can separate seeds with silica-based pubescent hairs can have a significant impact on breeding programs attempting to develop glabrous varieties and impact farmer safety as well as food safety. The goal of developing an accurate and inexpensive automated sorting system was achieved. The described system has an average accuracy of 90% for separating glabrous from pubescent canaryseed and has a throughput of approximately 100 g/h (approximately 14,000 seeds/h). The system was not significantly affected by variety, growing location, or growing year. Training of the system was simple and took only a few minutes and should not need to be performed more than once per week. Parts for the system totaled less than $2500. The system should be a useful tool for separating canaryseed by its surface texture and may be useful for sorting other seeds by glabrous/pubescent traits, such as barley seeds or oat groats.

References

Abdel-Aal, E. S. M., P. Hucl, and F. W. Sosulski. 1997. Structural and compositional characteristics of canaryseed (*Phalaris canariensis* L.). *J. Agric. Food Chem*. 45(8): 3049-3055.

Beyer, K., G. Grishina, L. Bardina, and H. A. Sampson. 2007. Identification of 2 new sesame seed allergens: Ses i 6 and Ses i 7. *The J. of Allergy and Clinical Immunology* 119(6): 1554-1556.

Gangur, V., C. Kelly, and L. Navuluri.2005. Sesame allergy: a growing food allergy of global proportions? *Ann Allergy Asthma Immunol.* 2005: 95(1): 4-11.

Haff, R. P., and T. C. Pearson. 2006. Spectral band selection for optical sorting of pistachio nut defects. *Trans. ASABE* 49(4): 1105-1113.

Hucl, P., M. Matus-Cádiz, A. Vandenberg, F. W. Sosulski, E. S. M. Abdel-Aal, G. R. Hughes, and A. E. Slinkard. 2001. CDC Maria annual canarygrass. *Can. J. Plant Sci*. 81(1): 115-116.

Matus-Cádiz, M. A., P. Hucl, and A. Vandenberg. 2003. Inheritance of hull pubescence and seed color in annual canarygrass. *Can. J. Plant Sci.* 83(3): 471-474.

O'Neill, C. H., P. Jordon, T. S. Bhatt, and R. H. Newman. 1986. Silica and Oesophageal Cancer, in Ciba Foundation Symposium 121 - Silicon Biochemistry, eds. D. Evered and M. O'Connor. Chichester, UK: John Wiley & Sons, Ltd.

Pearson, T. C. 2010. High-speed sorting of grains by color and surface texture. *Applied Eng. in Agric*. 26(3): 499-505.