

Concordancia entre analistas de laboratorios de latinoamérica para la determinación de la apariencia de grano de arroz pulido mediante el uso de imágenes digitales

Manuel Avila, Eduardo Graterol, Jesús Alezones, Beisy Criollo, Dámaso Castillo, Victoria Kuri, Norman Oviedo, Cesar Moquete, Marbella Romero, Zaida Hanley, Margie Taylor

Fundación para la Investigación Agrícola Danac. San Felipe, Venezuela. Semillas Híbridas de Venezuela, C.A. (SEHIVECA). Araure, Venezuela. Instituto de Investigaciones de Granos (IIG). Artemisa, Cuba. Fondo Latinoamericano para Arroz de Riego (FLAR). Cali-Palmira, Colombia. Semillas del Nuevo Milenio S.A. (SENUMISA). Alajuela, Costa Rica.

RESUMEN. La determinación de la apariencia del grano de arroz es un aspecto clave para evaluar su calidad. Generalmente, este análisis es realizado de manera visual empleando analistas expertos, sin embargo debido a la naturaleza subjetiva de su determinación, los resultados pueden ser divergentes. Con el objetivo de evaluar la concordancia entre analistas de laboratorios latinoamericanos de calidad de arroz en la determinación de la apariencia del grano de arroz pulido con ayuda de imágenes digitalizadas, se realizó un ensayo interlaboratorio con diez analistas e imágenes de 90 granos, capturadas mediante scanner de alta resolución. Los granos fueron clasificados en cuatro categorías incluyendo grano traslúcido, grano yesoso, grano panza blanca y granos dañados. La categorización fue analizada mediante la moda, frecuencia, concordancia relativa y coeficiente de concordancia *Kappa*. Adicionalmente, se elaboró una galería referencial de imágenes de granos típicos de cada categoría, basada en la frecuencia de modas. Los resultados revelaron un valor de *Kappa* de 0,49 que corresponde a una reproducibilidad moderada, atribuida a la subjetividad del análisis visual de las imágenes. Los resultados evidencian la necesidad de uniformizar criterios de evaluación entre analistas para mejorar la confiabilidad en la determinación de la apariencia de granos de arroz.

Palabras clave: Calidad de arroz, interlaboratorio, reproducibilidad

SUMMARY. **Concordance among analysts from Latin-american laboratories for rice grain appearance determination using a gallery of digital images.** The appearance of rice grain is a key aspect in quality determination. Mainly, this analysis is performed by expert analysts through visual observation; however, due to the subjective nature of the analysis, the results may vary among analysts. In order to evaluate the concordance between analysts from Latin-American rice quality laboratories for rice grain appearance through digital images, an inter-laboratory test was performed with ten analysts and images of 90 grains captured with a high resolution scanner. Rice grains were classified in four categories including translucent, chalky, white belly, and damaged grain. Data was categorized using statistic parameters like mode and its frequency, the relative concordance, and the reproducibility parameter *kappa*. Additionally, a referential image gallery of typical grain for each category was constructed based on mode frequency. Results showed a *Kappa* value of 0.49, corresponding to a moderate reproducibility, attributable to subjectivity in the visual analysis of grain images. These results reveal the need for standardize the evaluation criteria among analysts to improve the confidence of the determination of rice grain appearance.

Key words: Rice quality, ring test, reproducibility

INTRODUCCIÓN

Generalmente la apariencia visual del grano de arroz es el primer atributo que percibe el consumidor al momento de su compra. Los granos de aspecto traslúcido son los más apreciados en la mayor parte del mundo (1). Por el contrario, los granos con regiones opacas en el endospermo, de aspecto semejante a la tiza o al yeso, se consideran defectuosos y son indeseables, salvo para las variedades de tipo arbóreo y ce-

roso donde este aspecto es considerado característico. Además de la apariencia, se ha demostrado que los granos defectuosos son más susceptibles a fracturarse durante el procesamiento industrial de arroz, y pueden presentar un grado de cocción diferente al de los granos traslúcidos, lo que afecta su calidad de consumo (2). El origen de la opacidad del grano aún no está del todo claro, pero se ha demostrado que puede estar determinada por factores genéticos, ambientales y de manejo del cultivo (3-5). De acuerdo a normas de calidad

de diversos países (6-8), existen dos tipos de granos defectuosos por opacidad. Estos son los “granos yesosos” o “granos yesados”; definidos como aquellos que presentan un área opaca ventral igual o superior a 50% del área total del grano, y los “granos centro blanco” o “granos panza blanca”, con un área opaca interna menor del 50% del área total del grano. Estos tipos son determinados en la industria para valorar la calidad del grano, y contribuyen a definir el precio del arroz a nivel de los agricultores. Por ello, los programas de mejoramiento genético de arroz incluyen la apariencia de los granos como criterio de selección de los cultivos y así satisfacer las demandas de la industria y los consumidores (9).

Existe controversia en el método de cuantificación de los granos defectuosos en la industria del arroz de mesa, debido a que esta se realiza mayormente de forma visual, y la percepción de la extensión del área yesada puede variar de un analista a otro (10). De hecho, especialistas en clasificación de arroz la han declarado como una evaluación con alto contenido de subjetividad (11). Por ello, se han realizado programas de comparación interlaboratorio que permiten a los participantes evaluar su desempeño, así como detectar y corregir desviaciones de los resultados de ensayo (12). Uno de los elementos críticos de los estudios interlaboratorio es la asignación del valor de referencia contra el cual los resultados serán comparados. Este valor puede ser establecido mediante la designación del laboratorio de referencia o por una medida de consenso entre laboratorios. Ambos casos demandan tiempo y son costosos, ya que los laboratorios deben realizar un número significativo de determinaciones (13).

En los últimos años se han desarrollado instrumentos computarizados de captura y análisis de imágenes digitales para la determinación de granos defectuosos, con el objeto de eliminar las discrepancias en el análisis visual (14-16). Adicionalmente, y a diferencia de los granos, las imágenes pueden ser almacenadas indefinidamente, e incluso distribuidas fácilmente mediante tecnologías de la información como el correo electrónico y localizarse en sistemas en línea, permitiendo tantas verificaciones como sean necesarias. Sin embargo, debido al alto costo de inversión de los equipos requeridos, dichos métodos aún no han sido adoptados por la mayoría de los laboratorios, continuándose así la determinación visual. Por lo anteriormente expuesto, el objetivo del presente trabajo

fue determinar la concordancia entre analistas de laboratorios latinoamericanos de calidad de arroz en la determinación de la apariencia del grano de arroz pulido con ayuda de imágenes digitalizadas

MATERIALES Y MÉTODOS

La preparación de las muestras consistió en tomar al azar 90 granos de arroz pulido de tipo largo, procedentes de diversas variedades de arroz, facilitadas por Fundación Danac, Venezuela. Las imágenes de los granos se tomaron mediante un equipo “scanner” Scanjet 5590 FCLSD-0407 (Hewlett-Packard, USA). Se utilizó la máxima resolución del equipo (2400dpi), sin colocar la cubierta superior del scanner, con ayuda del programa de computación HP Photosmart para Windows® (Hewlett-Packard, USA). De cada grano se realizaron dos imágenes (una de cada cara longitudinal del grano) que fueron almacenadas en formato Joint Photographic Experts Group (JPEG). Finalmente, las imágenes fueron compiladas en una presentación del software Power Point (Microsoft, USA) a razón de seis granos por diapositiva, donde cada imagen ocupaba un área aproximada de 3,77 cm x 1,98 cm, equivalente a 50 veces el área promedio del grano. Para la evaluación, se envió el archivo con las imágenes codificadas y una planilla vía correo electrónico a analistas de institutos de investigación y agroindustrias del arroz de cinco países de Latinoamérica. Se seleccionaron sólo analistas con experiencia mayor de cinco años en el análisis de granos defectuosos, resultando en diez analistas de: Cuba [2], Costa Rica [2], Colombia [1], República Dominicana [1] y Venezuela [4].

La evaluación consistió en observar ambas imágenes de cada grano y, conforme su percepción, asignar una de cuatro posibles categorías: i) grano traslúcido (T): grano arroz entero y pulido de apariencia cristalina o traslúcida; ii) grano yesoso (Y): grano entero y pulido que presenta un área yesada en un 50% o más en su superficie; iii) grano panza blanca (P): grano entero y pulido de apariencia cristalina que presenta en su parte ventral interna, un área yesada menor de 50% del área total del grano; y iv) granos con daños (D): grano entero y pulido que está marcadamente dañado por el proceso de pulido, calor, factores climáticos, hongos o insectos. Dichas categorías se corresponden con las normas de calidad de grano aplicadas en los países de procedencia de los analistas considerados en

el estudio (17-18). Adicionalmente, se le consultó a los analistas acerca de la calidad de las imágenes y de las dificultades en la evaluación en virtud de considerarla como posible fuente de variación.

Análisis de resultados

Con los datos de las evaluaciones se calculó la moda y su frecuencia. La moda se valoró como la medida de consenso entre analistas para asignar la categoría de las imágenes. También se estimó la concordancia relativa entre analistas, el coeficiente de correlación *Kappa* y la reproducibilidad (R), empleados para cuantificar el grado de concordancia entre los observadores en sistemas de mediciones, en escala nominal (8-9). El cálculo de *Kappa* fue realizado por la siguiente ecuación:

$$kappa = \frac{P(A)-P(E)}{1-P(E)}$$

Donde:

P(A) = Probabilidad de concordancia observada.

P(E) = Probabilidad de concordancia si los evaluadores hicieran sus mediciones al azar.

Para el cálculo de R

se empleó la ecuación: $R = 1 - kappa$

Por otra parte, se calculó la concordancia relativa, y se estableció el orden de mérito de los analistas, basado en la proporción porcentual de aciertos con respecto a la moda del número total de evaluaciones realizadas. Finalmente, se estableció una tabla de referencia con una muestra de imágenes seleccionadas para cada categoría de grano.

RESULTADOS

Los datos de las evaluaciones de imágenes de los granos presentaron moda, lo cual permitió asignar una categoría predominante, a excepción de dos granos en los cuales las respuestas estuvieron divididas en 50% entre las categorías P y Y, así como D y T (datos no mostrados). Las frecuencias de las modas variaron ampliamente, sin embargo, la mayoría de las imágenes evaluadas (65,5%) alcanzaron entre 60 y 90% de las frecuencias de modas (Figura 1). Un 18,9% de los granos evaluados presentaron una frecuencia de moda de 100%, es decir, una coincidencia absoluta de las categorías arrojadas por los analistas, mientras que una baja proporción de imágenes presentaron frecuencias de modas inferiores a 50%.

Los valores de concordancia relativa oscilaron entre 22,2% y 100,0%. El 53,4% de las imágenes evaluadas presentaron concordancias relativas entre analistas su-

periores a 62,2%, mientras el resto de las imágenes presentaron niveles de concordancias relativas inferiores a 53,3% (Figura 2). Un grano con elevada concordancia relativa es indicativo que su definición converge a pocas categorías y denota un mayor consenso entre analistas. La probabilidad de concordancia observada entre analistas fue de 0,64; mientras que el coeficiente de concordancia *Kappa* y la reproducibilidad alcanzaron valores de 0,49 y 0,51, respectivamente.

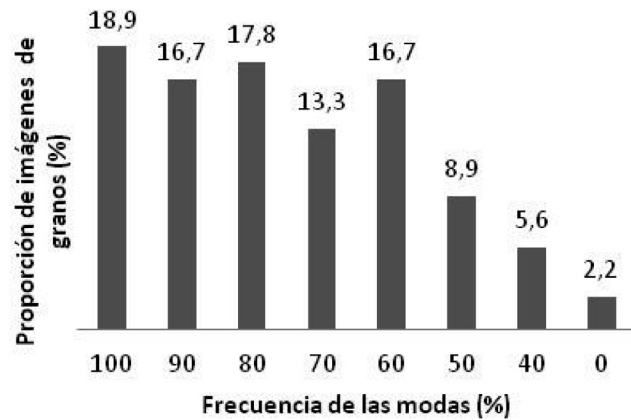


FIGURA 1. Distribución de imágenes de granos con base en la frecuencia de la moda de las respuestas

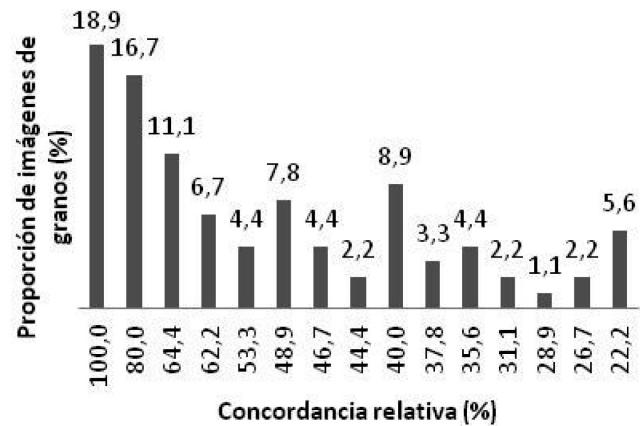
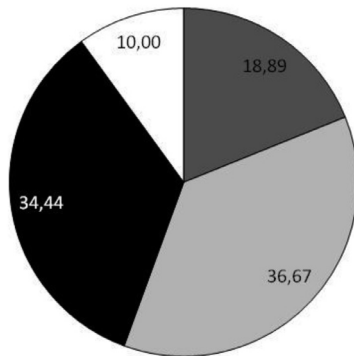


FIGURA 2. Distribución de imágenes de granos con base en la concordancia relativa de las respuestas

Al evaluar la distribución de imágenes de grano en función del número de categorías asignadas, se encontró que una considerable cantidad de imágenes alcanzaron una sola categoría de grano (Figura 3). Sin embargo, la mayor proporción de imágenes presentaron entre dos y tres categorías, incluso se evidenciaron imágenes que presentaron todas las categorías, pero en baja frecuencia. En cuanto al tipo de categoría, se

observó que las imágenes de algunos granos alcanzaron categorías absolutas de T y Y pero no D y P, estas dos últimas siempre estuvieron de forma combinada con el resto de las categorías (Figura 4). Las mayor proporción de imágenes evaluadas presentaron respuestas derivadas de la combinación D, Y y P; mientras que imágenes de granos con las combinaciones T - Y presentaron baja frecuencia.



■ Una categoría ■ Dos categorías ■ Tres categorías □ Cuatro categorías

FIGURA 3. Distribución de imágenes de grano con base en el número de categorías

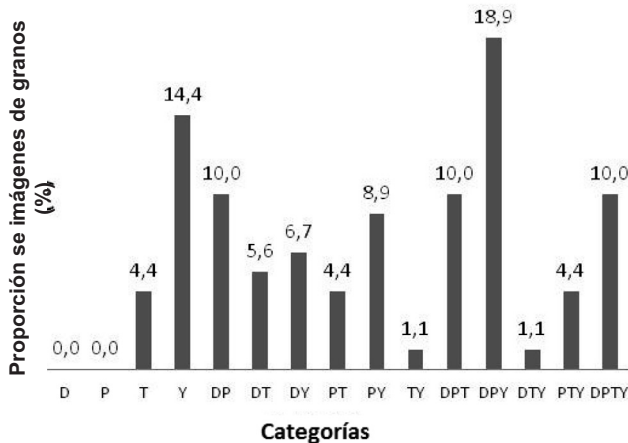


FIGURA 4. Distribución de imágenes de grano con base en las categorías asignadas

Los analistas difirieron en los niveles de aciertos con respecto a las modas de las imágenes de grano, lo cual permitió establecer un orden de mérito (Tabla 1). Este indicó que la mayor parte de los analistas presentaron niveles de aciertos entre 60 y 90%. La galería de imágenes generada representa una muestra de aquellas categorías asignadas para frecuencias de moda superiores a 70%, con lo que se establece un marco de referencia para la evaluación visual de imágenes de granos de arroz (Figura 5).

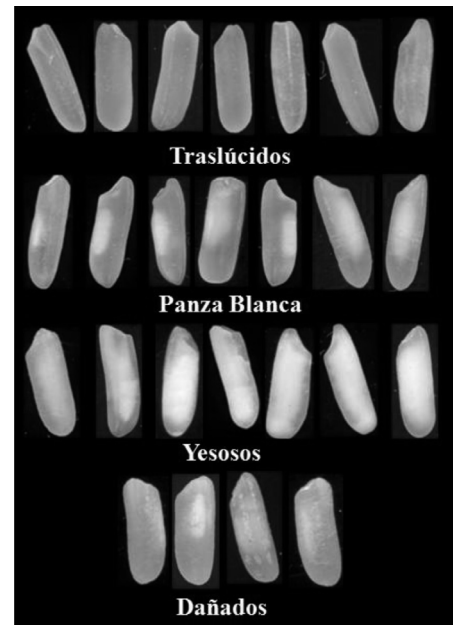


FIGURA 5. Galería de imágenes de granos por categoría

TABLA 1. Orden de merito de los analistas para la evaluación de imágenes de grano de arroz pulido

Analista	Aciertos N°	Aciertos (%)
9	52	57,78
10	60	66,67
2	61	67,78
4	64	71,11
6	68	75,56
8	70	77,78
5	73	81,11
1	76	84,44
7	79	87,78
3	83	92,22
n=90		

DISCUSIÓN

Las frecuencias de moda y concordancias relativas de las respuestas revelaron una percepción diferencial entre los analistas, al momento de definir algunos granos. Estas diferencias pueden ser atribuidas a la sensibilidad visual del analista para identificar algunos tipos de grano, pese a que las categorías empleadas son comúnmente aplicadas en los laboratorios de calidad considerados en el presente estudio (6-8). Los analistas no indicaron problemas para la evaluación relacionados con la calidad de la imagen, por lo que los resultados corroboran la subjetividad del método de evaluación visual tal como fue señalado por Oryza (11).

La asignación consistente para las categorías de T y Y en algunas imágenes, probablemente se debe a que estos presentan menor dificultad para establecer diferencias visualmente. En contraste, la existencia de imágenes con categorías de D o P únicamente bajo la forma combinada, puede ser atribuida a la dificultad de los analistas para cuantificar las áreas yesadas cuando están cercanas al 50% del área total del grano, con lo cual se asignan categorías de Y, P o incluso D. Existen daños mecánicos del grano durante el pulido que resultan en la aparición de áreas opacas externas (no ventral), que de acuerdo a su extensión podrían ser percibidas también como un Y o P.

A pesar de los años de experiencia de los analistas, estos presentaron diferentes niveles de precisión conforme a lo observado en el porcentaje de aciertos con la moda, lo que probablemente puede implicar errores en la tipificación de calidad del grano. El valor de concordancia estimada en función de *Kappa* indicó que los analistas presentaron un nivel de concordancia “moderada” conforme a las categorías propuesta por Landis y Koch (19). En la presente investigación se encontró un mayor nivel de acuerdo entre analistas que en lo señalado por Castillo y Gaviria (10), quienes al evaluar con cinco analistas expertos 50 granos considerados entre Y y P, encontraron entre 40% y 80% de respuestas divergentes.

La tabla referencial de imágenes derivada de este estudio se presenta como una alternativa para uniformizar los criterios de evaluación entre analistas, y contribuir a la mejora de la confiabilidad de los resultados del análisis visual de granos defectuosos de arroz, sin las limitaciones asociadas a la replicación de muestras físicas entre laboratorios.

CONCLUSIÓN

La evaluación interlaboratorio para la determinación de la apariencia del grano mediante imágenes digitales evidenció una concordancia moderada entre analistas, que se atribuye a la subjetividad de la evaluación visual. Los analistas presentaron mayor concordancia para las categorías de granos traslúcidos y yesosos con respecto a granos panza blanca y dañados. Estos resultados evidencian la necesidad de uniformizar criterios para la evaluación visual de la apariencia de granos en arroz. La realización de ensayos interlaboratorio como el realizado en esta investigación, así

como el uso de una tabla referencial de imágenes de granos para cada categoría de apariencia, pueden ser alternativas que contribuyan a mejorar la confiabilidad de los resultados emitidos por los analistas.

AGRADECIMIENTOS

Los autores agradecen a la Ing. Ana Acosta por la captura y organización de las imágenes de granos, y al Dr. Omar Verde por sus recomendaciones para el análisis estadístico de los datos.

REFERENCIAS

1. Tashiro T, Ebata M. The effect of ripening conditions on occurrence of white belly kernel. *Proc. Crop.* 1975; 44: 86-92.
2. Martínez C, Cuevas F. Evaluación de la calidad culinaria y molinera del arroz. Centro Internacional de Agricultura Tropical (CIAT). Colombia: Colombia S.A.; 1989.
3. Lisle A, Fitzgerald M. Finding the link between chalkiness and starch synthesis: Temperature rice conf, 2nd. *Rice Res. Philippines.* 2002; p 485–487.
4. Juliano B. Compiler. An international survey of methods used for evaluation of cooking and eating qualities of milled rice. *IRPS.* 1985; 77: 3–23.
5. Resurrección A, Yoshida S. Effect of temperature during ripening on grain quality of rice. *Soil Sci. Plant Nutr.* 1977; 23:109–112.
6. Grain Inspection, Packers and Stockyards Administration (GIPSA). United States Standards for Rice. [Internet]. 2005 [Citado 30 de julio de 2010]; Disponible en: <http://archive.gipsa.usda.gov/reference-library/standards/ricestandards.pdf>
7. Instituto Colombiano de Normas Técnicas y Certificación (ICONTEC). Norma técnica Colombiana. NTC 519; 2001.
8. Comisión Venezolana de Normas Industriales. (COVENIN) Norma COVENIN 44-90. Arroz paddy. Venezuela;1990
9. Torres E, Jeninngs P, Duque M, Kuri V, Corredor E, Sierra J. Análisis de estabilidad para centro blanco en arroz (*Oryza sativa*). *Foro Arroceros Latinoamericano.* 2002; 8:9-13
10. Castillo A, Gaviria J. Molinería de arroz en los trópicos. Colombia: Ediagro LDTA; 2000.
11. Oryza. What makes rice chalky? Are U.S. hybrid varieties connected to chalkiness? We ask experts. Tell us what by dailynews. [Internet]. 2011 [Citado 14 de junio de 2010]; Disponible en: <http://oryza.com/Rice->

- News/13040.html
12. Dotta G, Calidad de grano de arroz: En camino hacia la uniformidad en el criterio analítico. *Revista Arroz*. 2008; 55:21-30
 13. Cabañas E, Sánchez R, Alva S, Programa de evaluación de la calidad entre laboratorios XXII. La confiabilidad de la media de consenso. *LAB-acta*. 1999; 11:113-117
 14. Castillo A. Molinería (Pilado) de Arroz. Molinería en los trópicos. Tomo II. Colombia: Ediagro LDTA; 2007.
 15. CONARROZ. Sistema estándar de producción agroindustrial de arroz. Analizador de arroz Suzuki S21. [Internet]. 2007 [Citado Mayo 2011]; Disponible en: <http://www.conarroz.com/pdf/11Equiposdelaboratorio.pdf>
 16. Aryan A, Martin M, Blackburn H, Philpot K, Snell P, Reinke R. New tolls for precise & efficient grain evaluation. *IREC Farmers Newsletter*. 2006; 171: 35-37
 17. Futrell, D. When quality is a matter of test, use reliability indexes. *Quality progress*. 1995; 81-86.
 18. Martínez B. Evaluando la repetibilidad y reproducibilidad de sistemas de medición para características de calidad replicables, definidas por atributos múltiples en escala nominal. Instituto Autónomo de Querétaro. [Internet]. 2001 [Citado 20 de mayo de 2009]; Disponible en: http://www.itvictoria.edu.mx/principal/tecnointelecto/vol4_no1/Tecnointelecto%20Vol%204%20No%201%20pp%203441%20Mart%C3%ADnez.pdf
 19. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159-123.

Recibido: 17-02-2012

Aceptado: 26-06-2012