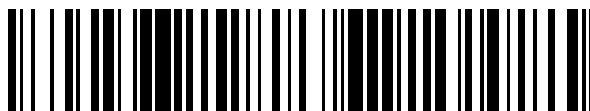


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 708 758**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.10.2014 PCT/EP2014/072953**

87 Fecha y número de publicación internacional: **07.05.2015 WO15063008**

96 Fecha de presentación y número de la solicitud europea: **27.10.2014 E 14793824 (5)**

97 Fecha y número de publicación de la concesión europea: **05.12.2018 EP 3063294**

54 Título: **Predicción de características híbridas**

30 Prioridad:

30.10.2013 DE 102013111980

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

11.04.2019

73 Titular/es:

**HI FIDELITY GENETICS LLC (100.0%)
Corporation Trust Center 1209 Orange Street
Wilmington, DE 19801, US**

72 Inventor/es:

**SCHOLTEN, STEFAN;
THIEMANN, ALEXANDER;
SEIFERT, FELIX;
FRISCH, MATTHIAS y
MELCHINGER, ALBRECHT**

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 708 758 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Predicción de características híbridas

5 La presente invención está relacionada con un método para predecir el nivel de heterosis en plantas, tal y como se determina en las reivindicaciones.

10 Un híbrido se define como la descendencia o el resultado de un cruce entre dos padres o progenitores, normalmente líneas endogámicas, con un acervo genético diferente. Las líneas endogámicas que se usan, por ejemplo, en el mejoramiento comercial del maíz híbrido (Duvick & Cassman; 1999; 'Post-green revolution trends in yield potential of temperate maize in the north-central United States'; *Crop Sci.*; 39: 1622-1630) son individuos sin heterocigosidad que se producen mediante retrocruzamiento constante o como haploides duplicados por medio de técnicas biotecnológicas (Geiger & Gordillo; 2009; 'Doubled haploids in hybrid maize breeding'; *Maydica*; 54: 485-499).

15 Cuando se cultivan o desarrollan como híbridos y se producen cruzando dos progenitores genéticamente diferentes, muchas especies animales y vegetales presentan unas mayores tasas de crecimiento y producen una mayor biomasa y, en el caso de las cosechas y los animales de granja, proporcionan un mayor rendimiento y una mayor productividad. Este fenómeno se conoce como vigor híbrido o heterosis (Shull; 1908; 'The composition of a field of maize'; *American Breeders Association*; 4: 296-301) y puede aplicarse a casi todos los aspectos de la biología y a todas las características de las plantas y los animales cuando un híbrido supera a sus progenitores.

20 El alcance de la heterosis en plantas y animales puede variar enormemente y se calcula como el incremento en comparación con el valor promedio de ambos progenitores (heterosis de ambos padres o MPH, por las siglas en inglés de 'mid-parent heterosis') o como el incremento en comparación con el progenitor con el mayor rendimiento (heterosis del mejor padre o BPH, por las siglas en inglés de 'best-parent heterosis').

25 La heterosis tiene una importancia enorme en muchos cultivos, así como en el mejoramiento animal y vegetal, y la producción de híbridos que tengan un nivel elevado de heterosis resulta conveniente para el mejoramiento (Duvick; 1986; 'Plant breeding: past achievements and expectations for the future'; *Econ. Bot.*; 40: 289-297). A pesar de los análisis genéticos intensivos, la base molecular del fenómeno de la heterosis aún no se entiende completamente. Para explicar los beneficios que se derivan de la complementación, la combinación o la interacción de dos alelos diferentes en los híbridos, se usan modelos genéticos no exclusivos de dominancia (Bruce; 1910; 'The Mendelian theory of heredity and the augmentation of vigor'; *Science* 32: 627-628; Davenport; 1908; 'Degeneration, albinism and inbreeding'; *Science* 28: 454-455), sobredominancia (Shull; 1908; 'The composition of a field of maize'; *American Breeders Association* 4: 296-301; East; 1936; 'Heterosis'; *Genetics* 21: 375-397) y epistasia o epistasia (Stuber; 1994; 'Heterosis in plant breeding'; *Plant Breed Rev* 12: 227-251; Goodnight; 1999; 'Epistasis and heterosis', en Coors JG, Pandey S, editores; 'The Genetics and Exploitation of Heterosis in Crops'; *Madison: American Society of Agronomy*; págs. 59-67). También se han creado modelos hipotéticos que explican estas interacciones mediante redes reguladoras de genes (Omholt et al.; 2000; 'Gene regulatory networks generating the phenomena of additivity, dominance and epistasis'; *Genetics* 155: 969-980). Sin embargo, particularmente los mecanismos de heterosis con rasgos cuantitativos complejos como el rendimiento o el crecimiento vegetativo no se entienden bien (Birchler et al.; 2010; 'Heterosis'; *Plant Cell* 22: 2105-2112).

30 Las explicaciones genéticas de los modelos propuestos para al menos una parte de la heterosis observada en híbridos no proporcionan información cuantitativa sobre el alcance de la heterosis.

35 Por consiguiente, aparte de las explicaciones genéticas, se intentó describir la heterosis a nivel molecular. Se partió del supuesto de que las diferencias cuantitativas de las reservas de ARNm de algunos genes que existen entre los progenitores y sus híbridos contribuyen a la base molecular de la heterosis. Esta expresión génica diferencial de los híbridos en comparación con los progenitores endogámicos pudo observarse en el pasado en diversos estudios. Estos estudios mostraron grandes cambios en el transcriptoma de los híbridos en comparación con sus progenitores (Stupar et al.; 2008; 'Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis'; *BMC Plant Biol.* 8: 33; Guo et al.; 2006; 'Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis'; *Theor. Appl. Genet.* 113: 831-845; Swanson-Wagner et al.; 2006; 'All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents'; *Proc. Natl. Acad. Sci.*; EE. UU.; 103: 6805-6810; Meyer et al.; 2007; 'Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern'; *Plant Mol. Biol.*; 63: 381-391; Jahnke et al.; 2010; 'Heterosis in early seed development: a comparative study of F1 embryo and endosperm tissues 6 days after fertilization'; *Theor. Appl. Genet.* 120: 389-400; Uzarowska et al.; 2007; 'Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height'; *Plant Mol. Biol.* 63: 21-34; Stupar & Springer; 2006; 'Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid'; *Genetics* 173: 2199-2210), que eran o bien aditivos o no aditivos. Una expresión génica aditiva determina una expresión híbrida de acuerdo con el promedio de la expresión génica parental. Una expresión híbrida aditiva puede explicarse mediante una expresión génica diferencial cis-regulada. En una expresión no aditiva, la expresión híbrida difiere de la expresión promedio de ambos progenitores, lo que sugiere una contribución diferencial de los transfactores (Wittkopp et al.; 2004; 'Evolutionary changes in cis

and trans gene regulation'; Nature 430: 85-88).

Por lo tanto, los estudios suponen que están involucrados ambos genes aditivamente expresados y diferencialmente cis-regulados (Guo et al.; 2004; 'Allelic variation of gene expression in maize hybrids'; Plant Cell 16: 1707-1716; Springer & Stupar; 2007; 'Allelic variation and heterosis in maize: How do two halves make more than a whole?'; Genome Res. 17: 264-275; Thiemann et al.; 2010; 'Correlation between parental transcriptome and field data for the characterization of heterosis in *Zea mays*'; L. Theor. Appl. Genet. 120: 401-413), así como genes no aditivos diferencialmente trans-regulados. Los posibles efectos trans que pueden desempeñar un papel son ARNs pequeños, que pueden afectar, entre otros, al epigenoma (Ha et al.; 2009; 'Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids'; Proc. Natl. Acad. Sci.; EE. UU. 106: 17835-17840). Se observaron cambios en el epigenoma, a nivel de la metilación del ADN, entre híbridos y progenitores endogámicos de *Arabidopsis* y arroz (Groszmann et al.; 2011; 'Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor'; Proc. Natl. Acad. Sci.; EE. UU.; 108: 2617-2622; He et al.; 2010; 'Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids'; Plant Cell 22: 17-33). En el caso del arroz se observó un vínculo entre los patrones de metilación del ADN y la expresión génica en híbridos (Chodavarapu et al.; 2012; 'Transcriptome and methylome interactions in rice hybrids'; Proc. Natl. Acad. Sci.; EE. UU.; 109: 12040-12045). En el caso del maíz y la remolacha azucarera, pudo demostrarse que los cambios en la metilación del ADN no se limitan a los transposones, sino que también se dan en las regiones de codificación de genes (Zhao et al.; 2007; 'Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids'; Plant Science 172: 930-938; Zhang et al.; 2007; 'Endosperm-specific hypomethylation, and meiotic inheritance and variation of DNA methylation level and pattern in sorghum (*Sorghum bicolor* L.) inter-strain hybrids'; Theor. Appl. Genet. 115: 195-207).

Los ARNs pequeños están estrechamente vinculados con la metilación del ADN (Lister et al.; 2008; 'Highly integrated single-base resolution maps of the epigenome in Arabidopsis'; Cell 133: 523-36). Contribuyen a la estabilidad del genoma y son unos reguladores importantes de la expresión génica (Van Wolfswinkel & Ketting; 2010; 'The role of small non-coding RNAs in genome stability and chromatin organization'; J. Cell Sci.; 2010; 123: 1825-39). Los ARNs pequeños tienen una longitud de 15 nt a 40 nt y provocan o bien un silenciamiento génico transcripcional (TGS, por sus siglas en inglés) o bien un silenciamiento génico post-transcripcional (PTGS, por sus siglas en inglés). Dependiendo de las diferentes vías o rutas biosintéticas, se dividen en siRNAs (ARNs pequeños de interferencia; también llamados ARNips) y mi-ARNs o micro-ARNs (microRNA o miRNA, en inglés) (Bartel; 2004; 'MicroRNAs: Genomics, biogenesis, mechanism, and function'; Cell 116: 281-297). Los mi-ARNs se derivan de los, así llamados, genes MIR que codifican transcriptos monocatenarios y que luego forman una estructura de horquilla y, posteriormente, se convierten en mi-ARNs maduros gracias a las proteínas DICER (Bartel; 2004; 'MicroRNAs: Genomics, biogenesis, mechanism, and function'; Cell 116: 281-297). Sin embargo, los siRNAs provienen de ARNs bicatenarios y tienen una gran variedad de lugares de origen y/o biogénesis. Los ARNs bicatenarios que pueden llevar a la formación de siRNAs se derivan, por ejemplo, de regiones de secuencia complementarias inversas (repetición invertida) de ARN transcrito, pares de transcripción naturales cis-antisentido, polimerasas de ARN dependiente del ARN (RDRs), la replicación de virus de ARN o regiones genómicas ricas en retroelementos (Khraiwesh et al.; 2012; 'Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants'; Biochim. Biophys. Acta 1819: 137-48). Los ARNs bicatenarios se convierten en siRNAs bicatenarios cortos mediante DICER. Son, entre otras cosas, importantes represores de transposones y secuencias virales (Mallory & Vaucheret; 2006; 'Functions of microRNAs and related small RNAs in plants'; Nat. Genet. 38: S31-S36).

La primera prueba de la participación de ARNs pequeños en el efecto de heterosis vino del maíz y *Arabidopsis* y mostró una expresión diferencial de ARNs pequeños entre híbridos y líneas endogámicas (Barber et al.; 2012; 'Repeat associated small RNAs vary among parents and following hybridization in maize'; Proc. Natl. Acad. Sci.; EE. UU.; 106: 10444-10449; Groszmann et al.; 2011; 'Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor'; Proc. Natl. Acad. Sci.; EE. UU.; 108: 2617-2622). En estos estudios, los siRNAs de 24 nt se redujeron en los híbridos y los mi-ARNs reguladores expresados no aditivamente. Los siRNAs reducidos estaban relacionados principalmente con genes reguladores y sus regiones laterales y también mostraron una asociación con la expresión génica en los híbridos. Se ha sugerido que estos -así llamados- epialelos regulados epigenéticamente podrían contribuir a la heterosis mediante la variación híbrida (Groszmann et al.; 2011; 'Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor'; Proc. Natl. Acad. Sci.; EE. UU.; 108: 2617-2622). En el caso del maíz, se ha identificado un tipo determinado de siRNAs largos de 22 nt expresados diferencialmente (Nobuta et al.; 2008; 'Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant'; Proc. Natl. Acad. Sci.; EE. UU.; 105: 14958-14963). Un estudio realizado en dos líneas endogámicas y sus híbridos recíprocos reveló que los siRNAs de 22 nt expresados diferencialmente entre las líneas endogámicas y los híbridos provienen de ciertos retrotransposones y podrían contribuir a la heterosis debido a la variabilidad (Barber et al.; 2012; 'Repeat associated small RNAs vary among parents and following hybridization in maize'; Proc. Natl. Acad. Sci.; EE. UU.; 106: 10444-10449). Un estudio posterior sobre el maíz con híbridos recíprocos y sus progenitores reveló unas alteraciones epigenéticas complejas en los híbridos en comparación con sus progenitores basándose en la metilación del ADN, la modificación de la histona y los ARNs pequeños (Shen et al.; 2012; 'Genome-Wide Analysis of DNA Methylation and Gene Expression Changes in Two Arabidopsis Ecotypes and Their Reciprocal Hybrids'; The Plant Cell 24: 875-892). Scholten et al. plantean la hipótesis de que los ARNs pequeños no codificantes tienen un potencial de predicción ('Diagnostics in

plant breeding'; [20130101]; págs. 265ff).

Pero las observaciones realizadas hasta la fecha no dan muestra alguna de una influencia directa y cuantitativa de los ARNs pequeños sobre la heterosis.

5 Actualmente, la producción de híbridos con una heterosis fuerte se lleva a cabo principalmente basándose en pruebas de campo de prueba y error. A tal efecto, se cruzan progenitores genéticamente diferentes y se cultiva su descendencia para analizar sus características (Windhausen et al.; 2012; 'Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments'; G3 2: 1427-1436). Puesto que
10 características importantes como la producción solo pueden medirse en fases tardías del ciclo vital, estas pruebas requieren mucho tiempo. Además, en algunos casos la distancia genética entre dos progenitores tiene una correlación inconsistente con la heterosis y, por lo tanto, resulta insuficiente para predecir el rendimiento del híbrido (Melchinger; 1999; 'Genetic diversity and heterosis'; en: Coors JG, Pandey S (editores); 'The genetics and exploitation of heterosis in crops'; ASA-CSSA; Madison; págs. 99-118). Debido a estas limitaciones, durante el
15 proceso de mejoramiento se analizan muchos híbridos no aptos, lo que conlleva unos costes elevados. Más particularmente, a tenor del gran número de líneas endogámicas que se producen ahora en los programas de mejoramiento comercial de híbridos en cada ciclo de mejoramiento, los posibles híbridos no pueden analizarse por completo debido al elevado coste de las pruebas de campo, lo cual provoca pérdidas significativas de compañeros de cruce potencialmente extraordinarios (Schrage et al.; 2006; 'Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL'; Theor. Appl. Genet. 113: 1037-1047).

Se han desarrollado métodos que permiten realizar una predicción de características híbridas basándose en marcadores genéticos que indican polimorfismos de la secuencia de ADN entre las líneas parentales (Schrage et al.;
25 2006; 'Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL'; Theor. Appl. Genet. 113: 1037-1047; Schrage et al.; 2009; 'Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses'; Theor. Appl. Genet. 118: 741-751). La predicción precisa y fiable en base a los marcadores genéticos sigue siendo un desafío (Windhausen et al.; 2012; 'Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different
30 Breeding Populations and Environments'; G3 2: 1427-1436; Reif et al.; 2012; 'Genomic prediction of sunflower hybrid performance'; Plant Breed; 132: 107-114). También se usaron marcadores genéticos en combinación con datos de metabolitos de líneas parentales para predecir rasgos o características híbridas (Gärtner et al.; 2009; 'Improved Heterosis Prediction by Combining Information on DNA- and Metabolic Markers'; PLOS ONE 4: e5220; Riedelsheimer et al.; 2012; 'Genomic and metabolic prediction of complex heterotic traits in hybrid maize'; Nature
35 Publishing Group 44: 217-220). Estos modelos experimentales usados para integrar múltiples niveles de datos muestran la gran demanda y el enorme interés que existen para mejorar la precisión y la fiabilidad de los métodos para la predicción de rasgos híbridos.

A partir de estudios en los que se observó que la expresión génica se ve alterada en los híbridos en comparación
40 con las líneas parentales, se concluyó que los efectos híbridos están relacionados con los patrones de expresión génica modificada (Hochholdinger & Hoecker; 2007; 'Towards the molecular basis of heterosis'; Trends in Plant Science 12: 427-432). Las predicciones sobre el rendimiento y la heterosis de los híbridos podrían realizarse en base a los perfiles de transcripción parentales en el maíz (Fu et al.; 2012; 'Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene
45 expression data'; Theor. Appl. Genet. 124: 825-833; Frisch et al.; 'Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize'; Theor. Appl. Genet. 120: 441-450) y en Arabidopsis (Stokes et al.; 2010; 'An association transcriptomics approach to the prediction of hybrid performance'; Mol. Breeding 26: 91-106).

50 Asimismo, en el caso del tomate (Shivaprasad et al.; 2011; 'Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs'; The EMBO Journal 31: 257-266), Arabidopsis (Groszmann et al.; 2011; 'Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor'; Proc. Natl. Acad. Sci.; EE. UU.; 108: 2617-2622), el maíz ('Repeat associated small RNAs vary among parents and following hybridization in maize'; Proc. Natl. Acad. Sci.; EE. UU.; 109: 10444-10449) y el arroz (Chodavarapu et al.;
55 2012; 'Transcriptome and methylome interactions in rice hybrids'; Proc. Natl. Acad. Sci.; EE. UU.; 109: 12040-12045), se descubrió recientemente que la expresión de ARNs pequeños difiere entre progenitores y sus híbridos. Más particularmente, las diferencias en la expresión de ARNs de 24 nt originaron la hipótesis de un componente epigenético de la heterosis.

60 Sin embargo, todos los métodos conocidos anteriormente no pueden aplicarse para predecir cuantitativamente la heterosis o el rendimiento de un híbrido específico y, por lo tanto, para determinar las líneas parentales más prometedoras. Aún no se conoce un método que permita, en base a las líneas parentales, una predicción exacta de la cantidad o nivel de heterosis o vigor híbrido de los híbridos resultantes. Por consiguiente, el objetivo de la presente invención es proporcionar un método así que permita, en base a las líneas parentales, una predicción
65 exacta de la cantidad o nivel de heterosis o vigor híbrido de los híbridos resultantes.

De acuerdo con la invención, este objetivo se alcanza mediante un método en el que se identifican las moléculas de ARN pequeño o sRNA (por sus siglas en inglés) que están relacionadas con un rasgo o característica híbrida y se analizan las líneas parentales que son adecuadas para la producción de híbridos a fin de hallar los niveles de expresión de las moléculas de ARN pequeño identificadas; de manera que la identificación de las moléculas de ARN pequeño comprende los pasos de:

- a) determinación del alcance de la expresión de rasgos o características en diferentes híbridos producidos a partir de líneas parentales genéticamente diferentes; y
- b) análisis de las líneas parentales de los híbridos testados en cuanto a la expresión de su ARN pequeño.

De este modo, la invención permite la selección de organismos adecuados para la producción de híbridos vegetales que tengan un mayor vigor híbrido o heterosis en uno o más rasgos híbridos, como la producción, la fertilidad, la resistencia al estrés, etc.

El método de acuerdo con la invención, tal y como se especifica en las reivindicaciones, puede iniciarse con moléculas de ARN pequeño identificadas previamente, de manera que se seleccionan las líneas parentales más prometedoras en base a la expresión de estas moléculas de ARN pequeño. Por otra parte, es posible identificar nuevas moléculas de ARN pequeño. Esto resulta necesario, por ejemplo, cuando para ciertos organismos no se ha designado ninguna molécula correspondiente de ARN pequeño o los organismos no se han analizado hasta el momento en busca de un rasgo particular.

Por consiguiente, una realización preferida de la invención proporciona la identificación de moléculas de ARN pequeño, que comprende los pasos de:

- a) cultivar plantas o animales de líneas parentales genéticamente diferentes;
- b) cruzar dichas plantas o animales para producir híbridos;
- c) determinar el alcance de la expresión de rasgos o características en híbridos diferentes;
- d) analizar las líneas parentales de los híbridos testados en cuanto a la expresión de su ARN pequeño.

El análisis de las líneas parentales en cuanto al nivel de expresión de las moléculas de ARN pequeño identificadas se realiza, por ejemplo, determinando el número de moléculas de ARN pequeño. De este modo, se determina la expresión diferencial de ARN pequeño entre las líneas parentales genéticamente diferentes.

De manera sorprendente, se descubrió que los perfiles de expresión de los -así llamados- ARNs pequeños o 'sRNAs' (por sus siglas en inglés) que tienen una longitud de entre 15 y 40 nucleótidos, que no codifican proteínas, permiten sacar conclusiones sobre los rasgos híbridos. Más particularmente, estos sRNAs pueden usarse para hacer predicciones sobre el alcance de la heterosis u otras características de los híbridos resultantes del cruce de plantas o animales. De este modo, la presente invención proporciona unos beneficios significativos para el proceso de mejoramiento, ya que las predicciones sobre los rasgos híbridos pueden realizarse con una generación de adelanto sin necesidad de recoger datos de campo para los genotipos exactos. En comparación con un método que utiliza perfiles de expresión del ARNm, la presente invención se distingue por una exactitud de predicción particularmente alta.

Las mediciones de la expresión de sRNAs en organismos individuales o en un grupo de organismos, como en líneas endogámicas o híbridos con un esquema de cruce factorial de una población reproductora, preferiblemente se realizan utilizando una secuenciación de alto rendimiento (por ejemplo, pirosecuenciación, secuenciación mediante hibridación, secuenciación de ADN semiconductor de iones, secuenciación por síntesis de puentes, secuenciación de dos bases, secuenciación de extremo pareado) o una secuenciación de alto rendimiento mediante la medición de la reacción de moléculas individuales (por ejemplo, protones, fluoróforos) o métodos convencionales, por ejemplo, el método de Maxam y Gilbert (Maxam & Gilbert; 1977; 'A new method for sequencing DNA'; Proc. Natl. Acad. Sci.; EE. UU.; 74: 560-564) o el método dideoxi de Sanger (Sanger & Coulson; 1975; 'A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase'; Journal of Molecular Biology; 94: 441-448), a fin de registrar cuantitativamente de la manera más exhaustiva posible los sRNAs individuales existentes. Si un conjunto predeterminado de sRNAs se usara para la predicción, los datos de expresión del sRNA pueden medirse utilizando métodos de PCR; por ejemplo, RT-PCR cuantitativo (Varkonyi-Gasic et al.; 2007; 'Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs'; Plant Methods 3: 12) o experimentos con chips o micromatrices (Bowtell & Sambrook; 2003; 'DNA microarrays: a molecular cloning manual'; Cold Spring Harbor Laboratory Press; Cold Spring Harbor; Ney York; EE. UU.).

Preferiblemente, los datos de expresión del sRNA deben normalizarse para obtener una comparabilidad óptima entre las diferentes mediciones. Para este fin pueden usarse diversos métodos. Para los datos de expresión a partir de la secuenciación, los valores de expresión pueden ajustarse, por ejemplo, mediante un factor de escala como la profundidad de secuenciación. De manera alternativa, puede realizarse una normalización mediante controles endógenos o artificiales.

Otro método posible es la normalización por cuantiles, promedio o de varianza (McCormick et al.; 2011;

'Experimental design, preprocessing, normalization, and differential expression analysis of small RNA sequencing experiments'; Silence 2: 2).

5 Los datos de características utilizados para la asociación con los datos de expresión del sRNA pueden obtenerse de forma cualitativa y/o cuantitativa a partir de experimentos de campo/en invernadero y experimentos de laboratorio o durante el proceso de producción o de reproducción. Los datos cuantitativos o cualitativos de las características pueden obtenerse a partir de líneas endogámicas, híbridos, transgénicos u otros genotipos.

10 El análisis correlativo de los datos de sRNA con los datos de características puede llevarse a cabo usando un conjunto fijo y predeterminado de sRNAs o un conjunto de sRNAs específico para híbridos de un estudio de asociación.

15 Preferiblemente, la asociación de los datos de expresión del sRNA con los datos cualitativos de las características se lleva a cabo mediante una prueba de probabilidad binomial. Los individuos se dividen en dos o más grupos cualitativos dependiendo de sus características, y se analiza la probabilidad de una asociación no aleatoria de diversos patrones de expresión del sRNA (ver Ejemplo 1). Por ejemplo, los patrones de expresión del sRNA pueden ser expresiones diferenciales entre diferentes individuos (por ejemplo, progenitores híbridos), así como diferencias en los valores de expresión absolutos. La identificación de sRNAs expresados diferencialmente puede obtenerse estableciendo una expresión mínima y un pariente mínimo y/o una diferencia absoluta en la expresión. De manera alternativa, una expresión diferencial puede establecerse mediante pruebas o test estadísticos (por ejemplo, una prueba F, una prueba T, ANOVA o análisis de la varianza). Preferiblemente, se usan réplicas biológicas y/o técnicas.

25 La asociación de los datos de expresión del sRNA con los datos cualitativos de las características puede realizarse tal y como se ha explicado en relación a los datos cualitativos de las características, de manera que los valores cuantitativos de las características se dividen preferiblemente en dos tipos o clases cualitativas (por ejemplo, valores característicos altos y bajos). De manera alternativa a una prueba de probabilidad binomial, los sRNAs relacionados pueden identificarse mediante métodos de regresión paramétricos/no paramétricos (por ejemplo, regresión lineal) u otros métodos matemáticos de reconocimiento de patrones (por ejemplo, SVM, Random Forest, redes neurales).

30 Además del estudio cuantitativo de los datos de expresión de sRNAs individuales para la asociación con datos de las características, estos datos de expresión también pueden integrarse basándose en ciertos criterios (por ejemplo, anotaciones/regiones de secuencias genómicas). Una posibilidad adicional es el estudio meramente cualitativo de sRNAs (por ejemplo, la expresión a ciertos niveles) para la asociación con datos característicos mediante, por ejemplo, una prueba de probabilidad binomial.

35 La predicción de datos de las características puede realizarse usando los datos de expresión del sRNA de un individuo o en base a los datos de expresión de los progenitores de una prole. Para realizar una predicción, la expresión del sRNA y los datos de las características de otros individuos (para la predicción de las características de los individuos) o los datos de expresión del sRNA de los progenitores y los datos de las características de su prole deben existir hasta cierto punto. La predicción se realiza determinando los parámetros de predicción (por ejemplo, los parámetros de regresión) basados en los individuos conocidos y aplicando estos parámetros a los datos de expresión del sRNA del individuo que tenga el rasgo que se va a predecir.

45 Los parámetros de predicción pueden determinarse basándose en los datos absolutos de expresión del sRNA o, alternativamente, mediante mediciones de distancia (por ejemplo, entre progenitores de una prole o individuos y un individuo de referencia).

50 En resumen, la invención comprende el uso de análisis de expresión del sRNA de plantas, o de sus híbridos y/o líneas endogámicas o cualesquiera otros genotipos, tal y como se determina en las reivindicaciones, (1) para realizar predicciones sobre el alcance de la heterosis y otros rasgos o características en plantas o animales y (2) para identificar compañeros de cruce que proporcionen combinaciones ventajosas respecto a sus perfiles de sRNA y cuyas proles proporcionen mejores prestaciones respecto a uno o más rasgos. De este modo, por ejemplo, puede identificarse un conjunto de 11 272 sRNAs, lo cual permite realizar la predicción de la heterosis para la producción de granos en el maíz y también puede usarse como un conjunto completo o parcial para la predicción de características análogas en otras especies vegetales.

60 La invención permite que un cultivador realice una predicción de la heterosis de diferentes rasgos y la predicción de características no heteróticas examinando los perfiles de expresión del sRNA. La invención también permite la predicción de características mediante el análisis de las etapas muy tempranas (por ejemplo, plántulas) de la misma generación o de la siguiente. El material de muestras de sRNA que se usa para generar los datos de expresión del sRNA puede diferir respecto a la etapa de desarrollo o el tejido de las características expresadas. Esto significa que los tejidos utilizados para la investigación de sRNAs no deben ser idénticos a los de las mediciones de las características. Esto conlleva un menor tiempo y menos dinero gastado en el cultivo, y también a lo largo de todo el proceso de selección para la reproducción. Además, la utilización de etapas tempranas de desarrollo permite unas condiciones de cultivo controladas (por ejemplo, un invernadero, una cámara de crecimiento) y, de este modo, se puede realizar una predicción replicable o reproducible. En el caso del mejoramiento vegetal, pueden evitarse las

pruebas de campo testando las plántulas de las líneas parentales o los genotipos parentales para predecir las proles.

5 A efectos de la presente solicitud, las plantas incluyen, por ejemplo, cereales como el trigo, la cebada, el arroz o el maíz, vegetales como el pimentón, las cebollas, las zanahorias o los tomates, frutas como la manzana, la pera, la cereza o la uva, y otros vegetales económicamente importantes como las legumbres, las hierbas (por ejemplo, Miscanthus) o algas para la producción de biomasa o árboles para obtener madera (por ejemplo, el chopo).

10 Un requisito para la heterosis en plantas es el mejoramiento de líneas endogámicas o líneas poco heterocigotas para la producción de híbridos. Esto resulta más fácil en plantas con una alogamia que se da de manera natural (polinización cruzada). También existe la posibilidad de usar la esterilidad masculina, que se divide en la -así llamada- 'Esterilidad masculina génica (nuclear)' (o GMS, por sus siglas en inglés) y la 'Esterilidad masculina citoplásmica' (o CMS, por sus siglas en inglés) (Bruce; 1910; 'The Mendelian theory of heredity and the augmentation of vigor'; Science 32: 627-628).

15 Otras aplicaciones de la presente divulgación incluyen los animales, excepto los humanos, como los mamíferos, las aves y los peces. Los ejemplos en los que la heterosis desempeña un papel importante incluyen los animales de granja, como el ganado, los cerdos, las ovejas, las cabras, los pollos o los pavos. Los peces de cultivo piscícola en los que la heterosis es importante incluyen, por ejemplo, el salmón o la carpa. También es posible aplicar la divulgación a animales reproductores no relevantes agrariamente como caballos de carrera o perros. Para la presente invención, la asociación directa y cuantitativa de sRNAs con la heterosis u otras características se utiliza para la predicción.

20 Los estudios también han revelado grandes diferencias en los perfiles de expresión del sRNA entre líneas endogámicas diferentes y entre líneas endogámicas y sus proles híbridas. Las investigaciones sobre las poblaciones de sRNA de diversas líneas endogámicas y de los correspondientes datos de campo exhaustivos permiten determinar el impacto directo de algunos sRNAs sobre la heterosis.

25 La investigación de un total de 21 líneas endogámicas de dos grupos heteróticos de maíz (maíz mellado y maíz duro) condujo a la anteriormente mencionada identificación de 11 272 sRNAs cuya expresión parental diferencial está relacionada con la heterosis de la producción de grano. De estos, 6915 fueron negativos y 4357 fueron positivos y estaban relacionados con la heterosis de la producción de grano.

30 Por consiguiente, la heterosis en híbridos de maíz puede predecirse con éxito en base a la expresión parental diferencial de sRNA. Puesto que la heterosis es un fenómeno generalizado que no se limita al maíz u otros granos, sino que también se da en animales, la utilización de sRNAs descrita en el presente documento, tanto en un conjunto completo como parcialmente, también es bastante concebible en otras especies vegetales y animales, siempre y cuando estos se conserven y se prediga un rasgo heterótico similar al de la producción de grano. De lo contrario, el método de acuerdo con la invención requiere la identificación de nuevos conjuntos predictivos de sRNAs mediante la asociación de datos de expresión del sRNA y los respectivos datos de características de los progenitores y los híbridos. Con estos sRNAs predictivos es posible realizar una predicción para otros compañeros de cruce.

35 La invención que se describe en el presente documento, y que se especifica en las reivindicaciones, es complementaria con otros métodos usados para la predicción de características híbridas y puede combinarse en modelos integrados con diferentes niveles de datos, como marcadores genéticos (por ejemplo, SNPs y otros marcadores de ADN), mRNA, proteínas o datos de metabolitos (Riedelsheimer et al.; 2012; 'Genomic and metabolic prediction of complex heterotic traits in hybrid maize'; Nature Publishing Group 44: 217-220).

40 La presente invención también está relacionada con el uso de moléculas de sRNA para la predicción de rasgos híbridos en plantas, tal y como se especifica en las reivindicaciones. Más particularmente, las moléculas de sRNA son moléculas de sRNA parental expresadas diferencialmente que están relacionadas con el alcance o magnitud de la expresión del rasgo híbrido. El número de moléculas de sRNA permite realizar una buena predicción sobre el alcance de la heterosis.

55 Ejemplo

Figura 1: regresión lineal de las distancias binarias combinadas D_b de 98 híbridos

60 Figura 2: precisión o exactitud de la predicción -basada en el sRNA- de la heterosis para la producción de grano

Se midió la expresión del sRNA de líneas endogámicas de maíz de los dos grupos heteróticos (maíz duro y mellado). El maíz es adecuado para estos estudios debido al rango comparativamente grande de niveles de heterosis.

65 Se cultivaron un total de 21 líneas endogámicas con un esquema de cruce de 7x14 en condiciones controladas.

Cuatro de las siete líneas de maíz duro tenían un acervo genético de 'European Flint' y las tres restantes, 'Flint/Lancaster'. Ocho de las líneas de maíz mellado tenían un acervo genético de 'Iowa Stiff Stalk Synthetic' y seis, de 'Iodent'. Los datos de campo de los 98 híbridos del esquema de cruce y los de las líneas endogámicas se obtuvieron en diferentes lugares de Alemania. Las pruebas de campo se realizaron con una disposición experimental de 2 filas con 2-3 réplicas o duplicados. La producción de grano se midió en Mg/ha con una humedad de grano de 155g/kg. La heterosis de cada par de línea endogámica y su híbrido se determinó como MPH (por las siglas en inglés de 'mean parental heterosis', es decir, 'heterosis parental promedio') en Mg/ha con una humedad de grano de 155 g/kg.

Se agruparon cinco réplicas biológicas de plántulas de cada línea endogámica 7 días después de la siembra y se aisló el ARN del total del material biológico. El 'sRNAoma' se analizó mediante secuenciación profunda de Illumina.

Los adaptadores de secuenciación se eliminaron de los datos de secuenciación en bruto y se eliminaron las regiones de secuencia que tuvieran una calidad de secuenciación baja inferior a un 99,9%. Todas las secuencias redundantes con una longitud de entre 15 nt y 40 nt se combinaron para el siguiente análisis y se determinó su número de secuencia (expresión del sRNA).

Los datos de expresión del sRNA se normalizaron por cuantiles con una modificación que evitaba la asignación de datos de expresión normalizados a sRNAs sin expresar (Bolstad et al.; 2003; 'A comparison of normalization methods for high density oligonucleotide array data based on variance and bias'; Bioinformatics 19: 185-93).

Los valores de expresión normalizados por cuantiles se normalizaron a un número de secuencia por millón de secuencias por biblioteca de secuenciación (recuento de lecturas por millón de lecturas normalizadas por cuantiles, o 'rpmqn', por sus siglas en inglés). Esto permite realizar una comparación de diferentes bibliotecas de secuencias y diferentes profundidades de secuenciación.

Se asumió que los sRNAs que tenían una expresión mínima de 0,5 rpmqn se expresaban por definición. Los expresados diferencialmente fueron los sRNAs con una diferencia mínima de la expresión con un factor ≥ 2 , o, si un progenitor estaba por debajo de la expresión mínima, la expresión mínima del otro progenitor debía multiplicarse por la diferencia de expresión mínima ($= 1$ rpmqn).

La asociación de sRNAs parentales expresados diferencialmente con la heterosis se basó en el método de Frisch et al. (Frisch et al.; 2010; 'Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize'; Theor. Appl. Genet. 120: 441-450) con la ampliación a sRNAs negativamente asociados. Según sus niveles de heterosis, los 98 híbridos se dividieron en dos tipos o clases (heterosis baja/alta) y, para cada sRNA en ambos tipos, se determinó el número de híbridos con expresión parental diferencial. Para cada sRNA se determinó o_l y o_h , en correspondencia con el número de híbridos en los tipos de heterosis baja y alta, cuyos progenitores tienen una expresión diferencial del sRNA examinado. La probabilidad de distribución binomial calculada posteriormente representa la probabilidad de expresión diferencial, que se distribuye de manera desigual en los dos tipos, y, por lo tanto, se relaciona con el MPH de producción de grano. La probabilidad se calcula de acuerdo con la Fórmula 3 como una función del número de expresión diferencial del sRNA en los tipos o clases definidas (Fórmula 1 o 2).

$$k_{min} = o_h, k_{max} = (o_h + o_l) \quad \forall \quad o_l \leq o_h \quad (1)$$

$$k_{min} = 0, k_{max} = o_l \quad \forall \quad o_l > o_h \quad (2)$$

$$P_f = \sum_{k=k_{min}}^{k_{max}} Bin_{n,p}(k) \quad \text{con} \quad n = (o_h + o_l), p = \frac{1}{2} \quad (3)$$

Los valores-p $< 0,05$, de acuerdo con la corrección FDR de Benjamini-Hochberg (Benjamini & Hochberg; 1995; 'Controlling the false discovery rate: a practical and powerful approach to multiple testing'; J. R. Statist. Soc. B 57: 289-300), indican una asociación significativa del sRNA testado con el MPH de producción de granos. Se determinó la distancia binaria combinada D_b de cada híbrido basándose en un conjunto de análisis de sRNAs que comprendía el número $n_{f, pos}$, correspondiente a los sRNAs asociados positivamente, y el número $n_{f, neg}$, correspondiente a los sRNAs asociados negativamente, así como los números de sRNAs expresados diferencialmente, $n_{d, pos}$, de los sRNAs asociados positivamente y $n_{d, neg}$, de los sRNAs asociados negativamente, del conjunto de análisis (Fórmula 4):

$$D_b = \frac{\sqrt{\frac{n_{d,pos}}{n_{f,pos}}} \cdot n_{f,pos} + \left(1 - \sqrt{\frac{n_{d,neg}}{n_{f,neg}}}\right) \cdot n_{f,neg}}{n_{f,pos} + n_{f,neg}} \quad (4)$$

5

De los datos de expresión del sRNA de las 21 líneas endogámicas, se identificaron un total de 11 272 sRNAs que están asociados o relacionados significativamente con la heterosis de la producción de granos. 4357 de estos sRNAs están positivamente asociados y 6915 están negativamente asociados con la heterosis de la producción de granos.

10

Una regresión lineal de las distancias binarias combinadas D_b , basada en los 11 272 sRNAs -de los 98 híbridos testados- significativamente asociados con la heterosis de la producción de granos, mostró una alta correlación de $r = 0,933$ con el rasgo asociado de heterosis de la producción de granos (Figura 1).

15

Esta alta correlación sugiere una elevada exactitud de predicción de los sRNAs. Para confirmar este supuesto, se llevaron a cabo validaciones cruzadas de tipo-0 y tipo-2 como se describen en Frisch et al. (Frisch et al.; 2010; 'Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize'; Theor. Appl. Genet. 120: 441-450) con un total de 100 realizaciones. A tal fin, se determinaron dos conjuntos no superpuestos de progenitores endogámicos en cada validación, el conjunto de determinación a través del cual se determinan los parámetros de predicción, y el conjunto de predicciones para cuyos híbridos se predice la heterosis de la producción de granos. Para cada validación cruzada, se determinan todos los sRNAs relacionados o asociados con la heterosis de la producción de granos ($p > 0,05$, sin corrección de errores) basándose en los híbridos posibles del conjunto de determinación. Basándose en los sRNAs asociados determinados del conjunto de destino, puede determinarse la distancia binaria combinada D_b de los progenitores del conjunto de determinación de cada uno de los híbridos. Los parámetros a (intercepto o intersección del eje y) y b (pendiente) de la ecuación lineal (Fórmula 5) pueden calcularse para el conjunto de determinación mediante el método de mínimos cuadrados (minimización de menos cuadrados) utilizando los valores H y D_b conocidos para los híbridos del conjunto de destino. Los parámetros de predicción calculados (a y b) y la distancia binaria D_b , que se calcula individualmente para cada híbrido del conjunto de predicción en base a los sRNAs asociados del conjunto de destino, se usan entonces para predecir la heterosis de la producción de granos H para cada uno de estos híbridos (Fórmula 5).

20

25

30

$$H = a + b \cdot D_b \quad (5)$$

35

En cada prueba de validación cruzada, la distancia binaria combinada D_b se calcula para cada híbrido de parejas de progenitores endogámicos del conjunto de predicción de acuerdo con la Fórmula 4, y el valor característico de la heterosis de la producción de granos H se calcula por medio de los parámetros de regresión determinados a y b determinados mediante el conjunto de destino, tal y como se ha explicado anteriormente, de acuerdo con la Fórmula 5. La precisión o exactitud de la predicción se determina por medio de los coeficientes de correlación de los valores característicos conocidos y predichos para la heterosis de la producción de grano de los híbridos del conjunto de predicción. Los resultados de las validaciones de tipo-0 y tipo-2 se muestran en la Figura 2.

40

45

El método de acuerdo con la presente invención permite realizar unas predicciones de gran calidad sobre los datos de expresión del sRNA para la heterosis y, más particularmente, para la producción de grano.

50

55

60

65

REIVINDICACIONES

- 5 1. Un método para predecir el nivel o cantidad de heterosis en plantas, de manera que se identifican las moléculas de sRNA (o ARN pequeño) que están relacionadas con un rasgo híbrido y se analizan las líneas parentales que son adecuadas para la producción de híbridos a fin de hallar el nivel de expresión de las moléculas de sRNA identificadas; y de manera que la identificación de las moléculas de sRNA comprende los siguientes pasos:
- 10 a) determinar el alcance de la expresión de rasgos o características en diferentes híbridos producidos a partir de líneas parentales genéticamente diferentes; y
b) analizar las líneas parentales de los híbridos testados en cuanto a la expresión de su sRNA.
- 15 2. Un método de acuerdo con la reivindicación 1, que se caracteriza por el hecho de que las moléculas de sRNA tienen una longitud de entre 15 y 40 nucleótidos.
3. Un método de acuerdo con cualquiera de las reivindicaciones 1 o 2, que se caracteriza por el hecho de que las moléculas de sRNA no codifican proteínas.
- 20 4. Un método de acuerdo con cualquiera de las reivindicaciones 1 a 3, que se caracteriza por el hecho de que se lleva a cabo una determinación de la expresión del sRNA diferencial entre líneas parentales genéticamente diferentes.
5. Un método de acuerdo con cualquiera de las reivindicaciones 1 a 4, que se caracteriza por el hecho de que el rasgo o característica es el rendimiento o producción.
- 25 6. Un método de acuerdo con cualquiera de las reivindicaciones 1 a 5, que se caracteriza por el hecho de que se realiza un análisis en plántulas sobre la expresión del sRNA de las líneas parentales.
7. El uso de moléculas de sRNA para la predicción del nivel o cantidad de heterosis en plantas.
- 30 8. El uso de acuerdo con la reivindicación 7, que se caracteriza por el hecho de que las moléculas de sRNA son moléculas de sRNA parental expresado diferencialmente que están relacionadas con el alcance o magnitud de la expresión de un rasgo.
- 35
- 40
- 45
- 50
- 55
- 60
- 65

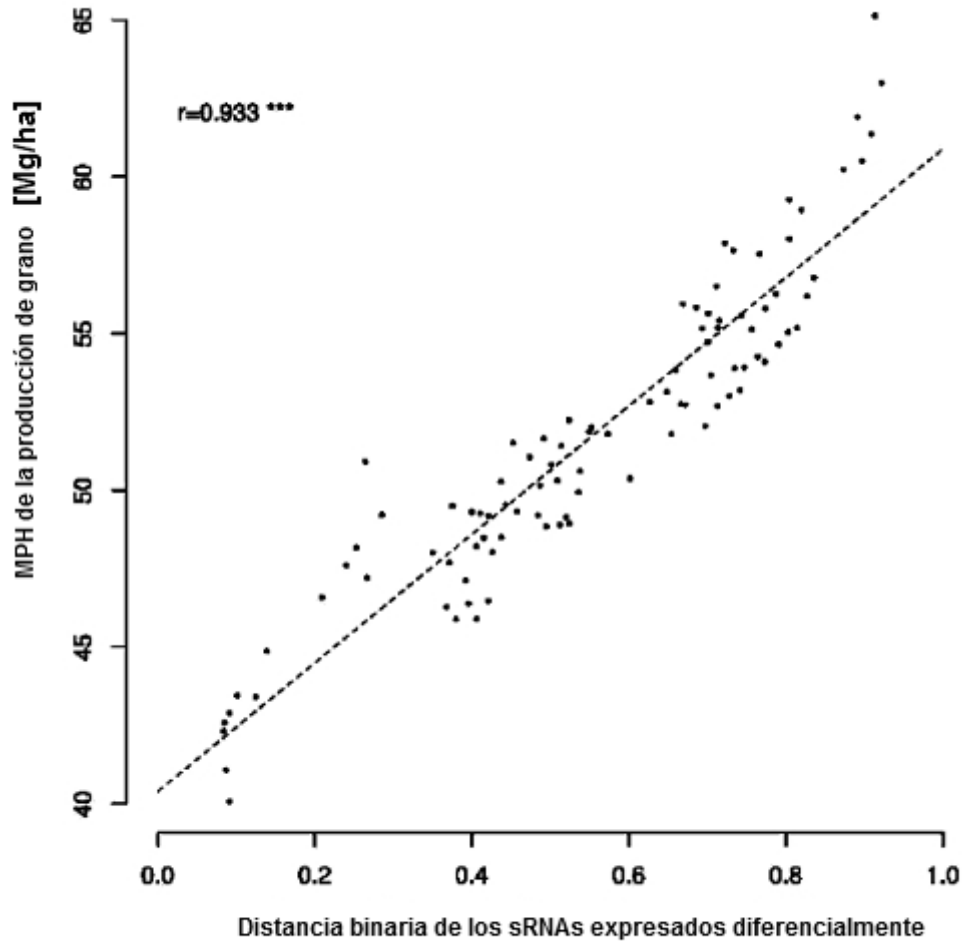


Figura 1

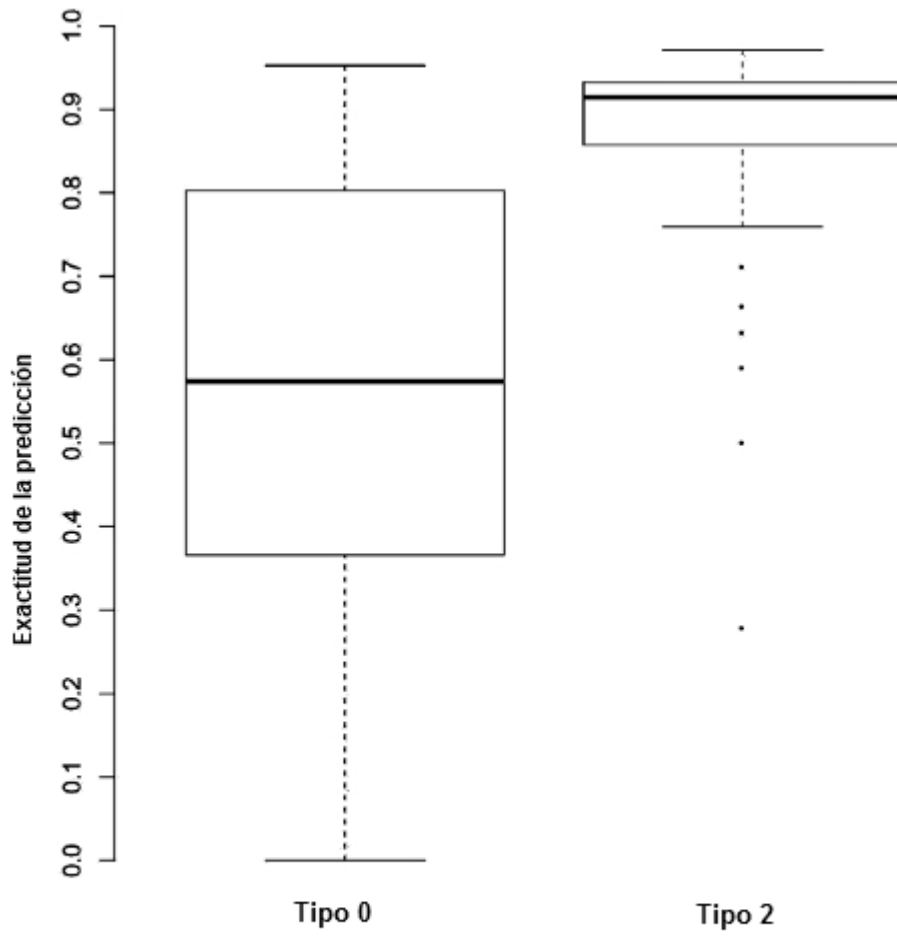


Figura 2