COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE RECURSOS GENÉTICOS Y PRODUCTIVIDAD

GENÉTICA

ÍNDICES DE SELECCIÓN MOLECULAR: UN NUEVO ENFOQUE

J. JESÚS CERÓN ROJAS

T E S I S

PRESENTADA COMO REQUISITO PARCIAL

PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2009

ÍNDICE DE SELECCIÓN MOLECULAR: UN NUEVO ENFOQUE

J. Jesús Cerón Rojas, Doctor

Colegio de Postgraduados, 2009

Con base en la teoría de la descomposición singular de las matrices de varianzas y covarianzas fenotípicas y genotípicas, y la de los marcadores moleculares, se presenta la teoría de dos índices de selección construidos con la información fenotípica y la que proporcionan los marcadores moleculares. Al primer índice de selección se le llama método del índice de selección molecular eigen (*Molecular Eigen Selection Index Method*, o MESIM por sus siglas en inglés) y al segundo, por tratarse de una generalización de MESIM al caso de la selección molecular del genoma completo en varios ambientes, método del índice de selección molecular del genoma completo (*Genome Wide Molecular Eigen Selection Index Method*, o MESIM_{GW}, por sus siglas en inglés). La investigación se dividió en dos partes principales. En la primera se desarrolla la teoría estadística de MESIM y en la segunda parte la de MESIM_{GW} siguiendo las ideas fundamentales de MESIM. Los resultados teóricos y los obtenidos por simulación en computadora indican que MESIM y MESIM_{GW} tienen al menos tres ventajas sobre los índices de selección estándar actualmente disponibles: MESIM y MESIM_{GW} no requieren ponderaciones económicas; las propiedades muestrales del índice y de los estimadores de sus parámetros son conocidas en el contexto asintótico, y el avance genético promedio que se alcanza con ellos es mayor o igual al de los índices de selección estándar.

Palabras clave. Descomposición espectral, matriz inversa generalizada de Moore-Penrose, valores y vectores característicos.

ÍNDICE DE SELECCIÓN MOLECULAR: UN NUEVO ENFOQUE

J. Jesús Cerón Rojas, Doctor

Colegio de Postgraduados, 2009

Using the singular value decomposition of the phenotypic and genotypic variance-covariance matrices and the variance-covariance matrix of molecular markers, we present the theory of two selection indices constructed with phenotype information and that provided by molecular markers. The first selection index is named *Molecular Eeigen Selection Index Method* (MESIM) and the second one is named *Genome Wide Molecular Eeigen Selection Index Method* ($MESIM_{GW}$, GW stand for *genome wide*) because is a generalization of MESIM to the case of genome wide selection in several environments. This work is divided into two principal sections. In the first one the statistical theory of MESIM is presented and in the second part the theory of $MESIM_{GW}$ is presented by following basic ideas of MESIM. Computer simulation and theoretical results indicated that MESIM and $MESIM_{GW}$ have at least three advantages over the standard selection indices currently available: MESIM and $MESIM_{GW}$ do not require economic weights, the sampling properties of MESIM and $MESIM_{GW}$ are known and easy to estimate, and the genetic advance of MESIM and $MESIM_{GW}$ are equal or lager than that of the standard selection indices.

Key words. Eigenvalues and eigenvectors, Moore-Penrose matrix generalized inverse, spectral decomposition.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) *por el financiamiento otorgado para la realización del proyecto de investigación que concluye con el presente trabajo.*

A BIMBO*, por el financiamiento parcial de la primera parte del trabajo de investigación.*

Al ***Dr. José Crossa Hiriart****, jefe de la* Unidad de Biometría y Estadística del Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), *por su invaluable poyo en el presente trabajo de investigación.*

Al ***Dr.*** **Jaime Sahagún Castellanos***, quien fue el primero en creer que el enfoque dado a los índices de selección en la presente investigación era relevante.*

Al ***Dr. Fernando Castillo González****, por haberme permitido trabajar con absoluta libertad en el presente trabajo de investigación.*

Al ***Dr. Amalio Santacruz Varela****, quien en uno de mis peores apuros económicos me tendió la mano.*

Al ***Dr. Ignacio Benítez Riquelme****, quien siempre me orientó en la manera de ordenar los cursos y en el formato que tendría el presente trabajo de tesis.*

Con todos tengo una deuda infinita, y por esta razón considero que el presente trabajo es más de ellos que mío, excepto por los errores y las consecuencias que puedan derivarse de éste, los cuales y las cuales asumo sin ninguna duda.

DEDICATORIA

A

Ignacio y Samuel Cerón Rojas

hermanos entrañables, quienes, durante mis estudios de bachillerato y licenciatura, me tendieron la mano cuando más necesité su apoyo económico.

ÍNDICE GENERAL

INTRODUCCIÓN GENERAL

De acuerdo con Dwivedi *et al*. (2007), de 1995 al 2020 la población mundial pasará de 5,660 a 7,500 millones de personas. Los países en desarrollo contribuirán con el 97.5 % y los países desarrollados con el 2.5 % restante. La demanda de cereales durante tal período se incrementará 39 %; la demanda de carne 58 % y la demanda de raíces y tubérculos en 37 %. Los países en desarrollo serán responsables de aproximadamente el 85 % del incremento en la demanda global de cereales y carne. En respuesta a la alta demanda de carne, la demanda de cereales para la alimentación de ganado será el doble en los países en desarrollo. La demanda de maíz (*Zea mays*) en tales países se incrementará más rápido que para cualquier otro cereal y sobrepasará la demanda de arroz (*Oryza sativa*) y trigo (*Triticum aestivum*) en el 2020. Para satisfacer tal demanda, en el mundo se tendrá que incrementar en 40 % la producción de grano en el 2020. Como la ganancia que se obtiene con el mejoramiento convencional está disminuyendo gradualmente, el aumento adicional del rendimiento deberá generarse combinando el mejoramiento convencional con la genómica y las tecnologías de transgénicos para equipar a los cultivos con resistencia a estrés biótico y abiótico y con capacidad para la adaptación a diversos nichos ecológicos.

El desarrollo de la genómica ha proporcionado nuevas herramientas para el descubrimiento y la identificación de nuevos genes y alelos. Tales herramientas pueden acelerar la eficiencia de los programas de mejoramiento a través de la selección asistida por marcadores moleculares (*Marker Assisted Selection* o MAS por sus siglas en inglés) (Xu y Crouch, 2008). Los marcadores moleculares (MM) son secuencias específicas de ADN en el genoma de plantas y animales utilizados para indicar la presencia cercana de un gen de

interés; pueden ser de tamaño muy diferente, desde pequeñas secuencias hasta grandes fragmentos de ADN que pueden contener algún gen. Las ventajas principales que tienen los MM son: (a) permiten detectar pequeñas variaciones con mínima cantidad de material; (b) no afectan al fenotipo; (c) pueden detectarse en cualquier estado de desarrollo de la planta; (d) se distribuyen a todo lo largo del genoma; (e) la mayoría de ellos son codominantes; (f) no tienen efectos epistáticos ni pleiotrópicos; y (g) el polimorfimo es enorme (Arús y Moreno-González 1993; Cubero, 2003). Actualmente se cuenta con una gran variedad de MM que son de utilidad en el mejoramiento genético de plantas y animales. Algunos de ellos son: RFLP (*Restriction Fragment Length Polymorphism*), RAPD (*Random Amplified Polimorphic* ADN), AFLP (*Amplified Fragment Length Polymorphism*), SCAR (*Sequence Characterised Amplified Region*); STS (*Sequence Tagged Sites*); SNP (*Single Nucleotide Polymorphisms*), etc. Detalles específicos de los MM anteriores pueden verse en Cubero (2003).

Índice de selección molecular

Lande y Thompson (1990) establecieron las bases teóricas de MAS para la selección de caracteres cuantitativos utilizando estudios de simulación por computadora. Consideraciones teóricas adicionales han contribuido de manera más completa al entendimiento de aspectos fundamentales en el desarrollo de MAS respecto al tipo de población, tamaño de muestra, tamaño del genoma y número de MM (Zhang y Smith 1992, 1993; Gimelfarb y Lande 1994, 1995; Whittaker, 2003).

El procedimiento de Lande y Thompson (1990) está basado en un índice en el que se combina la información fenotípica con la información proporcionada por los loci de los

caracteres cuantitativos (QTL o *Quantitative Trait Loci*, por sus siglas en inglés) ligados a los MM. Esto se debe a que no es posible identificar todos los QTL que afectan al carácter de interés (Li, 1998), es decir, a menos que todos los QTL que afectan a los caracteres de interés sean identificados, en MAS debería combinarse la información fenotípica con los efectos de los QTL asociados a los MM , también denominados efectos MQTL, para asegurar la eficiencia de la selección (Dekkers y Settar, 2004). En la construcción del índice de selección de Lande y Thompson (1990) se requiere: (1) identificar el ligamiento entre el MM y el QTL en un mapa de QTLs o MMs; (2) estimar los efectos MQTLs; y (3) combinar los efectos MQTLs con la información fenotípica para clasificar los individuos mediante un índice de selección y, subsecuentemente, desarrollar líneas, variedades o poblaciones de interés. Los efectos MQTLs pueden identificarse y estimarse con base en el desequilibrio de ligamiento creado al cruzar líneas endogámicas o poblaciones divergentes (Jansen, 2003).

Mapas de QTLs

En el problema de identificar y localizar QTLs, el fenotipo del carácter cuantitativo y la constitución alélica de los MMs son observables, mientras que la constitución genética de los QTLs es desconocida. La identificación, localización y estimación de los efectos MQTLs están basadas en el estudio de ligamiento entre los QTLs y los MMs. Las pruebas de ligamiento son potentes y específicas para identificar los QTLs, pero su localización únicamente puede alcanzarse hasta un cierto nivel de precisión que presenta una región que, potencialmente, puede incluir cientos de genes (Borecki y Suárez, 2001).

La identificación de QTLs se hace recurriendo a pruebas de hipótesis estadísticas que permiten identificar el ligamiento entre marcadores y QTLs en la progenie de una retrocruza o F_2. Éstos son los diseños experimentales ideales porque los progenitores en la población F_1 tienen la misma fase de ligamiento, toda la progenie es informativa y el *desequilibrio de ligamiento*[1] es máximo (Beavis, 1998).

La metodología estadística estándar para detectar ligamiento entre un marcador y un QTL recurre al estadístico de prueba t de Student o F de Snedecor (Soller *et al.,* 1976) o al estadístico de razón de máxima verosimilitud cuando se involucran pares de marcadores adyacentes a un QTL; aunque, generalmente, se requerirá la aplicación de este último para estimar la localización del QTL (Jensen, 1989; Knapp *et al.,* 1990; Lander y Botstein, 1989; Van Ooijen, 1992, 1999).

Los métodos convencionales para la identificación de QTLs están basados en la comparación de modelos con un QTL *vs* un modelo donde ningún QTL está presente. Por ejemplo, en el mapeo[2] por intervalos estándar, la verosimilitud para un presunto QTL es evaluada en cada localización del genoma asumiendo que ningún QTL está en el intervalo bajo estudio *vs* un QTL está en dicho intervalo, sin embargo, es muy probable que los QTLs localizados en otra parte del genoma afecten al estadístico de prueba, disminuyendo, así, la potencia de la prueba y proporcionado estimaciones sesgadas del efecto y localización del QTL de interés (Lander y Botstein, 1989; Knapp, 1991). Además, aún si no

[1] Asociación no aleatoria entre MMs y QTLs.
[2] La palabra “mapeo” es una traducción literal de la palabra inglesa “mapping”.

existe un QTL en el punto de prueba pueden aparecer los llamados QTLs "fantasmas[3]" (Haley y Knott, 1992; Martínez y Curnow, 1992).

De acuerdo con Carbonell (1997), la construcción de mapas de QTLs persigue, entre sus principales objetivos: (1) localizar los genes que contribuyen a la variación genética de los caracteres de importancia económica; (2) utilizar la información de los mapas de QTL´s en los programa de selección asistida con marcadores moleculares en plantas y animales; y (3) clonar molecularmente los genes subyacentes en los caracteres cuantitativos (un objetivo a largo plazo).

La identificación y localización de QTLs sigue un algoritmo básico: (a) se construye un mapa de ligamiento con marcadores moleculares; y (b) se hacen pruebas de significancia estadística asumiendo que ningún QTL *vs* al menos un QTL está en la región de interés para identificar QTLs utilizando estadísticos de prueba como t de Student , F de Snedecor , o la razón de máxima verosimilitud y, posteriormente, se estiman los efectos MQTL.

Selección del genoma completo (G*enomewide Selection*)

El índice de selección de Lande y Thompson (1990) involucra dos etapas en su construcción: estimar los efectos MQTL y combinar en un índice la información fenotípica y los efectos MQTLs para maximizar la respuesta a la selección. Sin embargo, el proceso de identificación de QTL está sujeto a errores ya que las estimaciones de los efectos MQTL pueden estar sesgadas y dependen de la distribución de los MQTL y el nivel de significancia de la prueba (Knapp, 1998; Piyasatian *et al*., 2007; Bernardo, 2008). Lange y

[3] Un QTL "fantasma" es un efecto de cosegregación entre QTL´s y marcadores no adyacentes o distantes la cual puede ocurrir debido al azar o a efectos de selección en la población experimental.

Whittaker (2001), en poblaciones endogámicas, y Meuwissen *et al*. (2001), en poblaciones no endogámicas, han considerado una ligera modificación al índice de selección de Lande y Thompson (1990): en lugar de estimar los efectos MQTL y enseguida construir el índice de selección, incorporaron los MM como variables aleatorias adicionales en el índice, lo cual permite predecir los valores genotípicos en una sola etapa. A tal procedimiento lo llamaron selección del genoma completo (*genome-wide selection* o GWS, por sus siglas en inglés). En GWS se utilizan todos los MM sin importar si éstos están ligados a los efectos aditivos de los QTL y sin estimar los efectos MQTL (Bernardo y Yu, 2008). De acuerdo con Wong y Bernardo (2008), al hacer el supuesto de que los MM son variables aleatorias en GWS se evita el problema de sobreparametrización que podría ocurrir en el índice de selección de Lande y Thompson (1990). Además, en GWS no se violan los supuestos de normalidad multivariada y cambios de frecuencia alélicos pequeños, lo que sí ocurre en el caso de Lande y Thompson (1990), ya que la selección se hace con base en los efectos mayores de los QTL. Utilizando simulación por computadora Lange y Whittaker (2001) compararon GWS con el en índice de selección de Lande y Thompson (1990) y concluyeron que el avance genético o la ganancia promedio alcanzada por GWS es superior al último método. Es decir, la selección del genoma completo es una forma de selección apoyada en MM donde se utilizan todos los MM considerando que éstos están ligados a los efectos aditivos de los QTLs, pero, sin estimar los efectos MQTL.

Bernardo y Yu (2007) enfrentan el problema de la GWS de manera diferente a la de Lange y Whittaker (2001) y Meuwissen *et al*. (2001), ya que estiman los efectos aditivos ligados a los marcadores moleculares utilizando el modelo lineal mixto; a los estimadores que utilizan se les llama mejores predictores lineales insesgados (*Best Linear Unbiased*

Prediction o BLUP, por sus siglas en inglés) y utilizan tales estimaciones para predecir el valor reproductivo de las plantas o animales candidatos a selección. Bernardo y Yu (2007) reportan ganancias de rendimiento de grano en maíz del 18 al 43 % superiores a las alcanzadas con el procedimiento de Lande y Thompson (1990). Resultados similares han sido reportados por Wong y Bernardo (2008) en aceite de palma.

Un nuevo enfoque de los índices de selección apoyados en marcadores moleculares

Recientemente, Cerón-Rojas *et al*. (2006) desarrollaron un índice de selección (IS) basado en el eigen-análisis: el método del índice de selección eigen (*Eigen Selection Index Method*, o ESIM por sus siglas en inglés) en el cual los elementos del primer vector característico de la matriz de varianzas y covarianzas (o de correlaciones) fenotípicas de los caracteres de interés se utilizan como ponderaciones en el IS y el primer valor característico es utilizado en la respuesta a la selección. ESIM no requiere estimaciones de la matriz de varianzas y covarianzas genotípicas, ni la asignación de ponderaciones económicas y proporciona una respuesta a la selección similar a la que se obtiene del IS de Smith (1936). De manera similar, Cerón-Rojas *et al*. (2008), siguiendo las ideas fundamentales del índice de selección restringido de Kempthorne y Nordskog (1959), construyeron un ESIM que permite fija un número de caracteres cuando se seleccionan individuos para el siguiente ciclo de selección al que llamaron el método del índice de selección restringido eigen (*Restricted Eigen Selection Index Method* o RESIM por sus siglas en inglés).

El presente trabajo de investigación se dividió en dos partes. En la primera de ellas se desarrolla la teoría de un índice de selección apoyado en MM basado en el método

desarrollado por Cerón-Rojas et al. (2008) y en las ideas básicas de Lande y Thompson (1990). En este índice, denominado índice de selección molecular eigen (*Molecular Eigen Selection Index Method*, o MESIM por sus siglas en ingles), de manera similar que en el método de Lande y Thompson (1990) se utilizan los efectos MQTLs y los valores fenotípicos en el índice para hacer la selección de los individuos de interés. La eficiencia de este método en relación con la del método de Lande y Thompson (1990) se evaluó con simulación por computadora.

En la segunda parte se extiende la teoría de MESIM al caso del IS molecular del genoma completo (*Genome Wide Molecular Eigen Selection Index Method* MESIM_{GW}, por sus siglas en inglés) en varios ambientes y varios caracteres de manera simultánea. Un problema que se presenta cuando se incorpora al índice de selección un gran número de MM ligados es que la inversa estándar de la matriz de varianzas y covarianzas de los MM no existe, por lo que es necesario recurrir a la teoría de las inversas generalizadas para la construcción del índice de selección en el contexto de GWS. Lange and Whittaker (2001) han sugerido una inversa generalizada diferente a la de Moore-Penrose y señalan que aunque las estimaciones de los coeficientes del índice de selección dependen de la elección particular de la inversa generalizada, las estimaciones del índice no dependen de tal inversa. Sin embargo, las estimaciones del índice son una función de las estimaciones de las ponderaciones del índice, por lo que no es claro como es que las estimaciones del índice serán independientes de las estimaciones de las ponderaciones del índice. Creemos que es evidente que los valores estimados del índice dependen del tipo de inversa que se elija, lo que puede conducir a un sin fin de valores estimados, por lo que debería elegirse una

inversa generalizada que sea única, ya que esto permitirá que en todo caso los valores estimados del índice sean comparables. La inversa generalizada de Moore-Penrose es única y existe para matrices cuadradas o rectangulares y cuando la matriz es no singular, la inversa generalizada de Moore-Penrose se transforma en la inversa estándar, lo que no ocurre con otro tipo de inversa generalizada (Kollo y von-Rosen, 2005; Schott, 2005). Para evaluar la eficiencia de los métodos indicados, se utilizó, nuevamente, simulación por computadora.

A MOLECULAR SELECTION INDEX METHOD BASED ON EIGENANALYSIS[§]

J. Jesús Cerón-Rojas, Fernando Castillo-González, Jaime Sahagún-Castellanos,

Amalio Santacruz-Varela, Ignacio Benítez-Riquelme, y José Crossa

RESUMEN

El índice de selección molecular tradicional empleado en la selección asistida por marcadores moleculares maximiza la respuesta a la selección combinando la información de los marcadores moleculares ligados a los *loci* de los caracteres cuantitativos y los valores de los fenotipos de tales caracteres de los individuos de interés. Este estudio propone un índice de selección molecular basado en el método eigenanálisis (*Molecular Eigen Selection Index Method*, o MESIM por sus siglas en inglés), donde el primer eigenvector es utilizado como un criterio del índice de selección, y sus elementos determinan la proporción de la contribución de los caracteres al índice de selección. Este artículo desarrolla el esquema teórico de MESIM. Resultados de simulación muestran que las medias genotípicas y la respuesta esperada a la selección de MESIM para cada carácter es igual o mayor que la alcanzada por el índice de selección molecular tradicional. Cuando varios caracteres se seleccionan de manera simultánea, MESIM es muy eficiente para los caracteres con relativamente baja heredabilidad. Las ventajas principales de MESIM sobre el índice de selección molecular tradicional son que sus propiedades estadísticas muestrales son conocidas y que no requiere ponderaciones económicas y de esta manera puede utilizarse en aplicaciones prácticas cuando todas o algunas de las características necesitan mejorarse de manera simultánea.

Palabras clave: Índices de selección, eigen-análisis, *loci* de caracteres cuantitativos, marcadores moleculares

---

[§] Capítulo publicado como artículo científico en Genetics 180: 547-557 (2008).

ABSTRACT

The traditional molecular selection index (MSI) employed in marker-assisted selection maximizes the selection response by combining information on molecular markers linked to quantitative trait loci (QTLs) and phenotypic values of the traits of the individuals of interest. This study proposes an MSI based on an eigenanalysis method (molecular eigen selection index method, MESIM) where the first eigenvector is used as a selection index criterion, and its elements determine the proportion of the trait's contribution to the selection index. This paper develops the theoretical framework of MESIM. Simulation results show that the genotypic means and the expected selection response from MESIM for each trait are equal to or greater than those from the traditional MSI. When several traits are simultaneously selected, MESIM performs well for traits with relatively low heritability. The main advantages of MESIM over the traditional molecular selection index are that its statistical sampling properties are known, and that it does not require economic weights and thus can be used in practical applications when all or some of the traits need to be improved simultaneously.

Key words: selection index, eigenanalysis, QTL, molecular markers

INTRODUCTION

Marker-assisted selection (MAS) is an important breeding tool in which molecular marker alleles linked to quantitative trait loci (QTLs) that control phenotypic variables of important traits are selected. Marker-assisted selection can be more efficient than selecting individuals based on phenotypic trait values. Progeny of specific progenitors can be selected based on molecular markers as long as these are associated with breeding values of the traits under consideration. This is one form of MAS (DEKKERS and DENTINE 1991; ARUS and MORENO-GONZALEZ 1993). Another form of MAS is based on the molecular selection index (MSI) proposed by LANDE and THOMPSON (1990). In MSI the selection response is maximized by combining information on molecular markers linked to QTLs and the phenotypic values of the traits of interest.

To construct an MSI, it is necessary to identify the linkage between the molecular marker and the QTL, the estimated effect of the QTL linked to the molecular marker (MQTL effect), and the combination of MQTL effects and phenotypic information that allows genotypes to be classified and selected using a selection index. The MQTL effects can be identified and estimated through the linkage disequilibrium that arises when crossing inbred lines or divergent populations (ZHANG and SMITH 1992, 1993; XIE and XU 1998). The MSI depends on various factors, such as number and density of molecular markers associated with QTLs, population size, trait heritability, additive genetic variances that can be explained by molecular markers, and precision of the estimated effect of gene substitution (DEKKERS and DENTINE 1991; MOREAU *et al*. 2000).

The MSI is an application of the selection index methodology proposed by SMITH (1936), in which MQTL effects are incorporated. As proposed by LANDE and THOMPSON (1990), the MSI performs a linear regression of phenotypic values on the coded values of the molecular markers such that selected molecular markers are those statistically linked to QTLs that explain most of the variability in regression models. The coefficient of regression of the molecular marker is the MQTL effect. Statistical models and methods for mapping QTLs and estimating their MQTL effects have been developed (JANSEN 2003). Several authors have pointed out the effectiveness of the MSI in inbred populations with large population sizes and traits with low heritability values (ZHANG and SMITH 1992, 1993; GIMELFARB and LANDE 1994, 1995; WHITTAKER 2003) when only one trait (and its associated molecular score) is considered.

The selection index theory was originally developed by SMITH (1936) and generalized by KEMPTHORNE and NORDSKOG (1959) for a restrictive selection index. The standard selection index is defined as a linear combination of the observed phenotypic values of the traits of interest with the traits' previously defined economic weights. Selection indices are based on improving one trait by incorporating information on related traits (WEI *et al*. 1996; FALCONER and MACKAY 1997) or incorporating information on MQTL effects by means of the MSI; other selection indices are based on improving several traits simultaneously, which requires assigning economic weights to each trait, as proposed by SMITH (1936).

MOREAU *et al*. (2000) and WHITTAKER (2003) found that the MSI is more effective than Smith's selection index only in early generation testing and has the additional disadvantage of increased costs due to molecular marker evaluation. Selection intensity

must also be considered because it affects genetic marker means and the ability to detect QTLs (WU *et al.* 2000). Furthermore, since selection increases the frequency of the QTL's favorable allele, as well as the allele of the molecular marker linked to it, total variability in the selected sample is reduced (MACKINNON and GEORGES 1992).

The MSI has the same advantages and disadvantages as Smith's selection index; it is simple to use but its sampling statistical properties and selection response are unknown, except in the case of two traits (HAYES and HILL 1980). Even for two traits, the statistical properties of Smith's selection index and its selection response, obtained using the delta method, are difficult to use and evaluate (HARRIS 1964); furthermore, it is not easy to consistently assign economic weights to the traits.

Recently, CERÓN-ROJAS *et al*. (2006) developed a selection index based on eigenanalyses of the phenotypic variance-covariance (or correlation) matrix of the traits of interest (called ESIM, for eigen selection index method). The authors showed that ESIM does not require economic weights nor estimates of the genotypic variances-covariances. In ESIM the elements of the first eigenvector determine the proportion each trait contributes to the selection index, and the first eigenvalue is used in the selection response. From a theoretical perspective, CERÓN-ROJAS *et al*. (2006) demonstrated that selection responses from Smith's selection index and from ESIM are the same, except for differences in selection index coefficients due to the different estimation methods. In addition, the ESIM of CERÓN-ROJAS *et al*. (2006) allows constructing a function to estimate gains (or losses) between selection cycles and predicting the selection response for future selection cycles. Following the restrictive selection index of KEMPTHORNE and NORDSKOG (1959),

CERÓN-ROJAS *et al.* (2008) developed a restrictive ESIM (RESIM) that facilitates maximizing the genetic progress of some characters while leaving the others unchanged.

In this paper we develop a molecular selection index (molecular eigen selection index method, MESIM) based on the restrictive eigenvalue selection index method (RESIM) of CERÓN-ROJAS *et al.* (2008) and the molecular selection index developed by LANDE and THOMPSON (1990) using the selection index methodology proposed by SMITH (1936), in which MQTL effects are incorporated. Simulated data were generated for comparing the selection response based on various selection indices: (1) MESIM versus LANDE and THOMPSON (1990); (2) RESIM versus the restrictive selection index of KEMPTHORNE and NORDSKOG (1959), and (3) ESIM versus the Smith selection index (SMITH 1936). Practical and theoretical properties of estimators from MESIM, RESIM, ESIM, LANDE and THOMPSON, the SMITH selection index, and the restrictive selection index of KEMPTHORNE and NORDSKOG are discussed. The efficiency of MESIM, LANDE and THOMPSON, ESIM, the SMITH selection index, and the restrictive selection index of KEMPTHORNE and NORDSKOG is evaluated using the genotypic means of the selected individuals. The theory of RESIM is described in CERON-ROJAS *et al*. (2008).

THEORY OF SELECTION INDICES

Smith's selection index

Details of Smith's selection index are given in CERÓN-ROJAS *et al*. (2006, 2008). A brief description follows. Smith's selection index is based on the linear combinations

$$\text{SI} = Y = \boldsymbol{\beta}'\mathbf{p} \text{ and } Z = \boldsymbol{\theta}'\mathbf{g} \quad (1)$$

where $\mathbf{p}' = [p_1 \quad ... \quad p_q]$ is the vector of the phenotypic values, and $\boldsymbol{\beta}' = [\beta_1 \quad ... \quad \beta_q]$ is the vector of coefficients of $\mathbf{p}$; Z is the breeding value, $\mathbf{g}' = [g_1 \quad ... \quad g_q]$ is the vector of genotypic values, and $\boldsymbol{\theta}' = [\theta_1 \quad ... \quad \theta_q]$ is the vector of economic weights. The phenotypic values p_j (j=1, 2 ,..., q) are modeled as $p_j = g_j + \varepsilon_j$, where g_j is the genotypic value of the jth trait and ε_j is the environmental component. Assuming that g_j and ε_j are independent, and that g_j represents only additive effects, $Z = \boldsymbol{\theta}'\mathbf{g}$ denotes the breeding value (HAZEL 1943; KEMPTHORNE and NORDSKOG 1959). Hence, selection based on $Y = \boldsymbol{\beta}'\mathbf{p}$ leads to a selection response

$$R = k\sigma_Z\rho_{YZ} = k\sigma_Z \frac{\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta}}\sqrt{\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}}} \qquad (2)$$

where $\boldsymbol{\Sigma}$ and $\mathbf{S}$ are the variance-covariance matrices of genotypic and phenotypic values, respectively, k is the standardized selection differential, $\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\beta}$ is the covariance between Y and Z, $\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}$ is the variance of Y, $\sigma_Z^2 = \boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta}$ is the variance of Z, and ρ_{YZ} is the correlation between Y and Z.

In Smith's selection index, the vector $\boldsymbol{\beta}_S = \mathbf{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{\theta}$ (the subscript S denotes Smith's method) where $\mathbf{S}^{-1}$ is the inverse of the phenotypic variance-covariance matrix, $\mathbf{S}$, permite construir el SI ($Y = \boldsymbol{\beta}_S'\mathbf{p}$) that maximizes la correlacion con $Z = \boldsymbol{\theta}'\mathbf{g}$.

Molecular selection index

LANDE and THOMPSON (1990) extended Eq. 1 to include the case where information on QTLs associated with molecular markers is available and denoted the molecular selection index as

$$Y_M = \boldsymbol{\beta}'_p\mathbf{p} + \boldsymbol{\beta}'_m\mathbf{m} \qquad (3)$$

$$= [\boldsymbol{\beta}'_p \quad \boldsymbol{\beta}'_m] \begin{bmatrix} \mathbf{p} \\ \mathbf{m} \end{bmatrix}$$

where $\boldsymbol{\beta}_p$ is a vector of phenotypic weights, $\boldsymbol{\beta}_m$ is the vector of weights of the molecular score, $\mathbf{p}$ is the vector of phenotypic values, and $\mathbf{m}' = [m_1 \quad \ldots \quad m_N]$, where each m_j (j=1, 2, …, N; N= number of molecular scores) is the j[th] molecular score given by the the sum of the products of the estimated additive effect of the QTL linked to the molecular marker (MQTL effects) multiplied by the coded values of their corresponding molecular markers. The response to this molecular selection index may be written as

$$R_M = k\sigma_M\rho_{Y_M Z_M} = k\sigma_M \frac{\boldsymbol{\theta}'_M\boldsymbol{\Sigma}_M\boldsymbol{\beta}_M}{\sqrt{\boldsymbol{\theta}'_M\boldsymbol{\Sigma}_M\boldsymbol{\theta}_M}\sqrt{\boldsymbol{\beta}'_M\mathbf{S}_M\boldsymbol{\beta}_M}} \qquad (4)$$

where $\mathbf{S}_M = \begin{bmatrix} \mathbf{S} & \mathbf{M} \\ \mathbf{M} & \mathbf{M} \end{bmatrix}$, $\boldsymbol{\Sigma}_M = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{M} \\ \mathbf{M} & \mathbf{M} \end{bmatrix}$;

k has been defined as in Eq. 2, $\sigma^2_M = \boldsymbol{\theta}'_M\boldsymbol{\Sigma}_M\boldsymbol{\theta}_M$ is the variance of the breeding value ($Z_M = \boldsymbol{\theta}'_1\mathbf{g} + \boldsymbol{\theta}'_2\mathbf{m}$), $\boldsymbol{\theta}'_M = [\boldsymbol{\theta}'_1 \quad \boldsymbol{\theta}'_2]$ is a vector of economic weights (in the standard molecular selection index, $\boldsymbol{\theta}_2$ is a vector of zeros), $\boldsymbol{\beta}'_M = [\boldsymbol{\beta}'_P \quad \boldsymbol{\beta}'_m]$ is a vector containing phenotypic ($\boldsymbol{\beta}_p$) and molecular ($\boldsymbol{\beta}_m$) weight scores; $\boldsymbol{\Sigma}$ and $\mathbf{S}$ are the variance-covariance matrices defined in Eq. 2, and $\mathbf{M} = \text{Var}(\mathbf{m})$ is the variance-covariance matrix of the

molecular scores when two or more traits are considered (LANDE and THOMPSON 1990). Only statistically significant additive MQTL effects are included in $\mathbf{m}$.

The vector $\boldsymbol{\beta}_{MSI} = \mathbf{S}_M^{-1}\boldsymbol{\Sigma}_M\boldsymbol{\theta}_M$ allows constructing the molecular selection index $Y_{MSI} = \boldsymbol{\beta}'_{MSI}\mathbf{p}_{pm}$ which has maximum correlation ($\rho_{Y_M Z_M}$) with $Z_M = \boldsymbol{\theta}'_1\mathbf{g} + \boldsymbol{\theta}'_2\mathbf{m}$ (the subscript MSI in $\boldsymbol{\beta}_{MSI}$ denotes LANDE and THOMPSON's molecular selection index method). In $Y_{MSI} = \boldsymbol{\beta}'_{MSI}\mathbf{p}_{pm}$, $\mathbf{p}'_{pm} = \begin{bmatrix}\mathbf{p}' & \mathbf{m}'\end{bmatrix}$ (Eq. 3). The variance of Y_{MSI} is $Var(Y_{MSI}) = \boldsymbol{\theta}'_M\boldsymbol{\Sigma}_M\mathbf{S}_M^{-1}\boldsymbol{\Sigma}_M\boldsymbol{\theta}_M$ and the maximized selection response can be written as $R_{MSI} = k\sqrt{\boldsymbol{\beta}'_{MSI}\mathbf{S}_M\boldsymbol{\beta}_{MSI}}$. Estimators of $\boldsymbol{\beta}_p$ and $\boldsymbol{\beta}_m$ ($\hat{\boldsymbol{\beta}}_p$ and $\hat{\boldsymbol{\beta}}_m$) for various traits are obtained directly from the estimators of $\boldsymbol{\Sigma}$, $\mathbf{S}$, and $\mathbf{M}$ ($\hat{\boldsymbol{\Sigma}}$, $\hat{\mathbf{S}}$, and $\hat{\mathbf{M}}$), and from the vector $\boldsymbol{\theta}_M$.

MESIM

Using a concept similar to that of KEMPTHORNE and NORDSKOD (1959), which maximizes the selection response (Eq. 2) by maximizing the square of the correlation between Y and Z (Eq. 1), and utilizing basic concepts from CERÓN-ROJAS et al. (2008), it can be shown that Eq. 4 is maximized by maximizing $\rho^2_{Y_M Z_M}$. The key point when maximizing $\rho^2_{Y_M Z_M}$ is that the variances (or standard deviations) of $Y_M = \boldsymbol{\beta}'_p\mathbf{p} + \boldsymbol{\beta}'_m\mathbf{m}$ and $Z_M = \boldsymbol{\theta}'_1\mathbf{g} + \boldsymbol{\theta}'_2\mathbf{m}$ are constants in each selection cycle. Thus, the selection of genotypes can be done using either Y_M or $Y_M / \sqrt{\boldsymbol{\beta}'_M\mathbf{S}_M\boldsymbol{\beta}_M}$. Because of this fact, when maximizing $\rho^2_{Y_M Z_M}$

it is possible to impose restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$ such that, in MESIM, it is required to maximize

$$\Phi = \left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M\right)^2 - \mu\left(\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M - 1\right) - \omega\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M - 1\right)$$

with respect to $\boldsymbol{\beta}_M$, $\boldsymbol{\theta}_M$, μ, and ω, where $\boldsymbol{\beta}_M$ is the vector of MESIM coefficients, $\boldsymbol{\theta}_M$ is the vector of economic weights, and μ and ω are Lagrange multipliers. In MESIM it is assumed that $\boldsymbol{\theta}_M$ is not a vector of constants.

When Φ is derived with respect to $\boldsymbol{\beta}_M$ and $\boldsymbol{\theta}_M$ (Appendix) and the result is set to the null vector, it follows that

$$\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M\right) \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M - \mu \mathbf{S}_M \boldsymbol{\beta}_M = \mathbf{0} \tag{5}$$

$$\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M\right) \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M - \omega \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = \mathbf{0} \tag{6}$$

Because the two restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$, when Eq. 5 is multiplied by $\boldsymbol{\beta}'_M$ and Eq. 6 is multiplied by $\boldsymbol{\theta}'_M$, the result is $\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M\right)^2 = \omega = \mu$. Hence, μ maximizes $\rho^2_{Y_M Z_M}$ under the restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$.

The following task is to determine the vector $\boldsymbol{\beta}_M$ that allows constructing Y_M that maximizes its correlation with $Z_M = \boldsymbol{\theta}'_1 \mathbf{g} + \boldsymbol{\theta}'_2 \mathbf{m}$. The Appendix shows that the required $\boldsymbol{\beta}_M$ is the solution to the following equality

$$(\mathbf{Q} - \mu \mathbf{I})\boldsymbol{\beta}_M = \mathbf{0}, \tag{7}$$

where $\mathbf{Q} = \mathbf{S}_M^{-1} \boldsymbol{\Sigma}_M$. Thus, for MESIM, the value that maximizes $\rho^2_{Y_M Z_M}$ under restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$ is the first eigenvalue (μ) of matrix $\mathbf{Q}$, and the vector that

allows constructing Y_M (with maximum correlation with $Z_M = \boldsymbol{\theta}_1'\mathbf{g} + \boldsymbol{\theta}_2'\mathbf{m}$) is the first eigenvector of matrix $\mathbf{Q}$ ($\boldsymbol{\beta}_M$).

Let μ and $\boldsymbol{\beta}_M = \boldsymbol{\beta}_{MESIM}$ be the first (largest) eigenvalue and its corresponding $\mathbf{Q}$ eigenvector, respectively; then, the selection index in the context of MESIM is $Y_{MESIM} = \boldsymbol{\beta}_{MESIM}'\mathbf{p}_{pm}$ ($\mathbf{p}_{pm}' = [\mathbf{p}' \quad \mathbf{m}']$) and, because $(\boldsymbol{\theta}_M'\boldsymbol{\Sigma}_M\boldsymbol{\beta}_M)^2 = \mu$, the maximized selection response can be written as $R_{MESIM} = k\sqrt{\mu}$. From $(\mathbf{Q} - \mu\mathbf{I})\boldsymbol{\beta}_M = \mathbf{0}$ it is possible to determine the $\boldsymbol{\beta}_M$ coefficients of $Y_M = \boldsymbol{\beta}_p'\mathbf{p} + \boldsymbol{\beta}_m'\mathbf{m}$ (Eq. 3), $\boldsymbol{\beta}_M' = [\boldsymbol{\beta}_P' \quad \boldsymbol{\beta}_m']$. Although the partial derivatives of Φ are obtained with respect to $\boldsymbol{\beta}_M$ and $\boldsymbol{\theta}_M$, in estimating Y_{MESIM} and $R_{MESIM} = k\sqrt{\mu}$, the vector of economic weights ($\boldsymbol{\theta}_M$) is not required because $\boldsymbol{\beta}_M$ and μ are obtained directly from matrix $\mathbf{Q}$.

Note that when information on the QTLs linked to the molecular markers is not incorporated into the selection index, *i.e.*, when $Y = \boldsymbol{\beta}'\mathbf{p}$, $Z = \boldsymbol{\theta}'\mathbf{g}$, and $R = k\sigma_Z \dfrac{\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta}}\sqrt{\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}}}$, then Eq. 7 can be written as

$$(\mathbf{S}^{-1}\boldsymbol{\Sigma} - \mu\mathbf{I})\boldsymbol{\beta} = \mathbf{0} \qquad (8)$$

from where it is evident that $\mathbf{Q} = \mathbf{S}^{-1}\boldsymbol{\Sigma}$. Equation 8 can be considered a variant of the procedure developed by CERÓN-ROJAS et al. (2006) for cases where the assumption of ESIM ($\boldsymbol{\Sigma}\boldsymbol{\theta} = \boldsymbol{\beta}$) is relaxed.

As indicated by CERON-ROJAS *et. al*. (2008), the maximized selection response, $R_{MSI} = k\sqrt{\boldsymbol{\beta}_{MSI}'\mathbf{S}_M\boldsymbol{\beta}_{MSI}}$ or $R_{MESIM} = k\sqrt{\mu}$, gives a general theoretical assessment of the gain

for all traits considered simultaneously but does not provide genetic gains per trait at each selection cycle. Alternatively, the expected selection response (BAKER 1986; VAN VLECK 1993) determines the expected genetic gain per trait per selection cycle $\mathbf{G} = k\frac{\boldsymbol{\Sigma}_M\boldsymbol{\beta}_M}{\sqrt{\boldsymbol{\beta}'_M\mathbf{S}_M\boldsymbol{\beta}_M}}$. However, $\mathbf{G}$ estimates the expected value of the genetic gains with low precision; thus in our simulated data we used the genotypic means of the selected individuals and the regression of the genotypic means of the selected individuals on the selection cycles for evaluating the efficiency of MESIM RESIM, ESIM, LANDE and THOMPSON, the restrictive selection index of KEMPTHORNE and NORDSKOG, and the SMITH selection index on the response to selection.

Matrix $\mathbf{Q}$ is square and nonsymmetric of order $q \times q$ (where q is the total number of variables: phenotypic and molecular scores)

$$\mathbf{Q} = \mathbf{S}_M^{-1}\boldsymbol{\Sigma}_M = \begin{bmatrix} (\mathbf{S}-\mathbf{M})^{-1}(\boldsymbol{\Sigma}-\mathbf{M}) & \mathbf{0} \\ \mathbf{I}-(\mathbf{S}-\mathbf{M})^{-1}(\boldsymbol{\Sigma}-\mathbf{M}) & \mathbf{I} \end{bmatrix} \quad (9)$$

Therefore, it is not possible to construct a subset of orthogonal vectors from Eq. 7. However, orthogonal vectors from $\mathbf{Q}$ can be calculated by means of singular value decomposition (SVD) (MARDIA *et al*. 1982). Using SVD, $\mathbf{Q}$ can be written as

$$\mathbf{Q} = \mathbf{UDV}' \quad (10)$$

where the columns of matrix $\mathbf{U}$ ($\mathbf{U}'\mathbf{U} = \mathbf{I}$) are the left singular vector of $\mathbf{Q}$, and the columns of matrix $\mathbf{V}$ ($\mathbf{V}'\mathbf{V} = \mathbf{I}$) are the right singular vector of $\mathbf{Q}$; $\mathbf{D}$ is a diagonal matrix with the square root of the eigenvalues (singular values) of $\mathbf{QQ}'$ or $\mathbf{Q}'\mathbf{Q}$ (the eigenvalues of $\mathbf{QQ}'$ and $\mathbf{Q}'\mathbf{Q}$ are the same).

The problem now is to determine: From where should the first singular vector for constructing Y_{MESIM} be taken, from $\mathbf{U}$ or from $\mathbf{V}$? Note that Eq. 10 can be written as $\mathbf{QV} = \mathbf{UD}$, from where it is evident that if μ is the first singular value of $\mathbf{Q}$, and $\mathbf{v}_1$ and $\mathbf{u}_1$ are its associated left and right first singular vectors, respectively, then $\mathbf{Qv}_1 = \mu\mathbf{u}_1$, from where $\mathbf{u}_1 = \mu^{-1}\mathbf{S}_M^{-1}\boldsymbol{\Sigma}\mathbf{v}_1$. Let $\boldsymbol{\beta}_{MESIM} = \mathbf{u}_1$, then $\boldsymbol{\beta}_{MESIM}$ is a linear transformation of $\mathbf{v}_1$. The estimators of $\mu = \mu_{MESIM}$ and $\boldsymbol{\beta}_{MESIM}$ are obtained from $\hat{\mathbf{Q}}\hat{\mathbf{Q}}'$, such that $(\hat{\mathbf{Q}}\hat{\mathbf{Q}}' - \hat{\mu}^2_{MESIM}\mathbf{I})\hat{\boldsymbol{\beta}}_{MESIM} = \mathbf{0}$. According to ANDERSON (2003), $\hat{\mu}^2_{MESIM}$ and $\hat{\boldsymbol{\beta}}_{MESIM}$ are the maximum likelihood estimators of the eigenvector and eigenvalue of $\mathbf{QQ}'$, respectively, and are asymptotically consistent and unbiased. The estimators of $\mathbf{Q}$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{D}$ are $\hat{\mathbf{Q}}$, $\hat{\mathbf{U}}$, $\hat{\mathbf{V}}$, and $\hat{\mathbf{D}}$, respectively, so $\hat{\mathbf{Q}} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{V}}'$. These results allow estimating Y_{MESIM} as $\hat{Y}_{MESIM} = \hat{\boldsymbol{\beta}}_{MESIM}\,\mathbf{p}_{pm}$. Asymptotically, $E(\hat{Y}_{MESIM}) \approx Y_{MESIM}$ (where $\approx$ means approximately).

When only one trait and its molecular scores are considered, $\mathbf{S}_M = \begin{bmatrix} s^2 & \sigma_m^2 \\ \sigma_m^2 & \sigma_m^2 \end{bmatrix}$, $\boldsymbol{\Sigma}_M = \begin{bmatrix} \sigma_g^2 & \sigma_m^2 \\ \sigma_m^2 & \sigma_m^2 \end{bmatrix}$, and $\mathbf{Q} = \begin{bmatrix} \dfrac{\sigma_g^2 - \sigma_m^2}{s^2 - \sigma_m^2} & 0 \\ \dfrac{s^2 - \sigma_g^2}{s^2 - \sigma_m^2} & 1 \end{bmatrix}$. When $\sigma_m^2 = 0$, then $\mathbf{S}_M = \begin{bmatrix} s^2 & 0 \\ 0 & 0 \end{bmatrix}$, $\boldsymbol{\Sigma}_M = \begin{bmatrix} \sigma_g^2 & 0 \\ 0 & 0 \end{bmatrix}$ and $\mathbf{Q} = \begin{bmatrix} h^2 & 0 \\ 0 & 0 \end{bmatrix}$, where s^2 and σ_g^2 are the phenotypic and genotypic variances of the trait, respectively, σ_m^2 is the variance of the molecular score associated with the trait under selection, and $h^2 = \dfrac{\sigma_g^2}{s^2}$.

SIMULATED DATA

We have simulated genotypes from a population with the aim of comparing theoretical and practical results from MESIM, RESIM, ESIM, the restrictive selection index of KEMPTHORNE and NORDSKOG (1959), the SMITH (1936) selection index, and the LANDE-THOMPSON (1990) molecular selection index. The simulator system used in this study, developed by WANG et al. (2004), has two main engines, QU-GENE and QuCim, which require different input data. To simulate a population, the input file for QU-GENE should contain the genetic structure of the genotypes for each specific trait, *i.e.*, number of genes (or QTLs); gene effect for each trait including additive, dominance, and epistasis; linkage among the genes in one chromosome; trait heritability; etc. Component QU-GENE can generate genotypes making up populations of cross-pollinated or self-pollinated species, or create different environmental conditions where the simulated genotypes will be evaluated. On the other hand, the input file for QuCim must have the type of crosses and the selection method to be used in each breeding strategy. Selection methods that can be simulated in QuCim include mass selection, pedigree system, bulk population system, backcross breeding, top-cross breeding, doubled haploid breeding, marker-assisted selection for one trait, and many combinations and modifications of these (WANG *et al.* 2004). The simulator provides, for each genotype in the population, the true genotypic value as well as the phenotypic value of the traits under study.

Generating a doubled haploid population for selection

The original data were taken from an actual doubled haploid maize mapping population of 236 genotypes with five traits; QTLs for all five traits were mapped. The five traits measured were: male flowering time (MFL) (days), female flowering time (FFL) (days), plant height (PHT) (cm), ear height (EHT) (cm), and 100-kernel weight (HKF) (grams). This data file was used to generate 200 doubled haploid genotypes that form the reference population (cycle 0). Using a selection pressure of 10% (k=1.755), 20 genotypes were selected under MESIM, the LANDE-THOMPSON, ESIM, RESIM, SMITH selection index, and the restrictive selection index of KEMPTHORNE and NORDSKOG . These 20 selected doubled haploids were then crossed in diallel fashion, and a new population of 200 doubled haploids was generated. This was repeated during five selection cycles for all five traits. The efficiency of the indices was compared using the mean genotypic value and the regression of the mean genotypic value of the selected genotypes on the selection cycles. We used phenotypic, genotypic, and molecular score variance-covariance matrices for estimating the singular vectors and singular values.

We also generated populations based on selection of individual traits with the objective of comparing MESIM and the LANDE-THOMPSON (1990) molecular selection index method for the simultaneous selection of five traits.

Sign of the coefficients, economic weights, and expected genetic gains

When using MESIM, ESIM, and RESIM, it is often necessary to change the sign of the coefficients of the first singular eigenvector in order to select the genotypes according to the desired genetic advance, that is, for traits such as MFL, FFL, PHT, and EHT, the

signs are always negative (decreasing the mean genotypic value), whereas for HKF the signs are always positive (increasing the mean genotypic value).

Concerning the economic weights for the LANDE-THOMPSON molecular selection index, the restrictive selection index of KEMPTHORNE and NORDSKOG (1959), and the SMITH (1936) selection index, economics weights were assigned following SMITH et al. (1981). Then, one set had coefficients of 1 or -1, and the other had the heritability of each trait multiplied by 1 or -1, depending on the trait. Therefore, for MFL, FFL, PHT, EHT, and HKF, the first set of coefficients was -1, -1, -1, -1, and 1, respectively, whereas the second set of coefficients was $-h^2_{MFL}$ (-0.51), $-h^2_{FFL}$ (-0.46), h^2_{PHT} (-0.38), $-h^2_{EHT}$ (-0.52), and h^2_{HKF} (0.27); all coefficients of the molecular markers associated with the traits were equal to zero. All five traits were simultaneously selected under MESIM, the LANDE-THOMPSON, ESIM, and the SMITH selection index, whereas for the restrictive selection index of KEMPTHORNE and NORDSKOG and RESIM, the traits that were unchanged were MFL and PHT.

Furthermore, MESIM and LANDE-THOMPSON were compared when traits were selected individually. When selection was performed on individual traits, the LANDE-THOMPSON molecular selection index based on heritabilities as economic weights was not applied, and only the index based on coefficients 1 and -1 (depending on the trait of interest), and 0 for the molecular scores was employed.

RESULTS AND DISCUSSION

The genotypic means under MESIM and LANDE-THOMPSON when selection is practiced on traits individually (not simultaneously on various traits) are shown in Table 1. Because genetic variability became exhausted, only two selection cycles were run. The MESIM-selected genotypes had better genotypic means than those selected under the LANDE-THOMPSON index for all five traits. To clarify the interpretation of the MESIM, consider, for example, the first selection cycle on the individual trait MFL. The estimated phenotypic, genotypic, and molecular score variances in the original population were $\hat{s}^2 = 33.489$, $\hat{\sigma}_g^2 = 18.156$, and $\hat{\sigma}_m^2 = 2.248$, respectively, from where

$$\hat{\mathbf{S}}_M = \begin{bmatrix} 33.489 & 2.248 \\ 2.248 & 2.248 \end{bmatrix}, \hat{\mathbf{\Sigma}}_M = \begin{bmatrix} 18.156 & 2.248 \\ 2.248 & 2.248 \end{bmatrix}, \hat{\mathbf{Q}} = \begin{bmatrix} 0.51 & 0 \\ 0.49 & 1 \end{bmatrix}, \text{ and}$$

$$\hat{\mathbf{Q}}\hat{\mathbf{Q}}' = \begin{bmatrix} 0.26 & 0.25 \\ 0.25 & 1.24 \end{bmatrix}.$$

The first singular value and its associated singular vector are $\hat{\mu}_1 = 1.14$ and $\hat{\boldsymbol{\beta}}'_{MESIM} = [0.2333 \quad 0.9724]$, respectively. However, because MFL decreases, it is necessary to multiply the elements of $\hat{\boldsymbol{\beta}}_{MESIM}$ by -1 such that the selection index in the context of MESIM is $\hat{Y}_{MESIM} = -0.233MFL - 0.9724m_{MFL}$, where MFL denotes the trait of interest, and m_{MFL} is the molecular score associated with MFL. In this case, the total expected genetic response can be partitioned into two components, the coefficient related to the phenotypic values per se and those related to the molecular scores. Value -0.233 is the phenotypic coefficient, and -0.972 is the molecular score coefficient.

Table 1. Mean genotypic values under MESIM and LANDE-THOMPSON molecular selection indices when traits are selected individually until genetic variability is exhausted (cycle 2). The traits were male flowering (MFL), female flowering (FFL), plant height (PHT), ear height (EHT), and 100-kernel weight (HKF) for one and two selection cycles for simulated data using phenotypic, genotypic, and molecular score variance-covariance matrices. The signs and economic weights of the selection indices for each trait are shown in parentheses.

	MESIM genotypic means					LANDE-THOMPSON genotypic means				
Selection cycles	MFL (-)	FFL (-)	PHT (-)	EHT (-)	HKF (+)	MFL (-1)	FFL (-1)	PHT (-1)	EHT (-1)	HKF (1)
0	98.54	98.89	139.61	88.37	20.45	98.54	98.89	139.61	88.37	20.45
1	93.89	97.03	124.89	75.83	22.85	93.23	96.91	132.87	72.64	20.80
2	91.66	93.83	120.62	63.33		92.08	94.36	127.18	66.61	

When selection is practiced on all five traits simultaneously, then weights -1, -1, -1, -1, and 1 for each trait are used; the heritability of the traits are also used as weights. The LANDE-THOMPSON molecular selection index is denoted as LANDE-THOMPSON1 when -1,-1,-1,-1, and 1 are used as weights, and when heritabilities are used as economic weights, it is denoted LANDE-THOMPSON2. Similarly, the standard SMITH selection index is denoted as SMITH 1 in the first case and SMITH 2 in the second case; and the KEMPTHORNE-NORDKOG restricted selection indices are denoted as KN1 and KN2, respectively.

For the trait HKF (100-kernel weight), the selection gain per cycle for MESIM (0.50 grams) was greater than that obtained by LANDE-THOMPSON1 (0.21 grams) and LANDE-THOMPSON2 (0.31 grams) (Table 2). However, for MFL (male flowering time), the opposite was true, that is, LANDE-THOMPSON1 (-0.91 days) and LANDE-THOMPSON2 (-0.83 days) under both sets of economic weights were more effective than

MESIM (-0.71 days) for maturity (Table 2). Comparing the genotypic means when individual traits are selected (Table 1) with those obtained when five traits are simultaneously selected (Tables 2), it is evident than the genotypic means are higher when only one trait is under selection. Correlations between traits play an important role in the correlated response of other traits.

Table 2. Mean genotypic values and gain per cycle of the 20 genotypes selected under MESIM, LANDE-THOMPSON1 (economic weights are 1s and -1s) and LANDE-THOMPSON2 (economic weights are heritability of the traits) molecular selection index for five traits selected simultaneously, male flowering (MFL), female flowering (FFL), plant height (PHT), ear height (EHT), and 100-kernel weight (HKF) for five selection cycles for simulated data using phenotypic, genotypic, and molecular score variance-covariance matrices. The signs and economic weights of the selection indices for each trait are shown in parentheses.

	MESIM					LANDE-THOMPSON1					LANDE-THOMPSON2				
Selection cycle	MFL (-)	FFL (-)	PHT (-)	EHT (-)	HKF (+)	MFL (-1)	FFL (-1)	PHT (-1)	EHT (-1)	HKF (1)	MFL (-0.5)	FFL (-0.46)	PHT (-0.38)	EHT (-0.5)	HKF (0.27)
0	98.5	98.9	139.6	88.4	20.4	98.5	98.9	139.6	88.4	20.4	98.5	98.9	139.6	88.4	20.4
1	97.0	98.4	123.5	74.3	21.3	98.7	99.5	130.5	80.7	20.8	98.0	98.7	127.1	76.1	21.3
2	96.5	99.1	118.7	70.9	21.3	97.0	98.4	123.6	71.3	20.6	96.9	98.2	129.1	74.3	19.6
3	96.1	96.5	119.2	64.8	21.6	96.0	97.9	122.5	66.6	21.5	95.7	97.0	121.8	68.5	20.6
4	95.9	95.8	117.4	60.8	22.7	95.3	98.4	119.2	64.1	21.6	95.0	97.2	120.4	70.7	21.9
5	94.4	95.6	114.9	59.4	23.0	94.4	96.6	117.6	59.9	21.2	94.8	96.4	119.9	67.9	22.1
Gain per cycle	-0.71	-0.78	-4.04	-5.48	0.50	-0.91	-0.44	-4.15	-5.62	0.21	-0.83	-0.51	-3.59	-3.55	0.31

Regarding the SMITH SI and ESIM, the genotypic means of the selected genotypes are shown in Table 3. In this case, for four of the five traits, MFL, FFL, EHT, and HKF, the selection gain per cycle of ESIM was greater than that obtained with the SMITH SI. Concerning KN RSI and RESIM (keeping MFL and PHT unchanged), the genotypic means of the selected genotypes are shown in Table 4. For HKF, the selection gain per cycle for

RESIM (0.48 grams) was greater than that obtained using KN1 RSI (0.27 grams) and KN2 RSI (0.19 grams). However, for FFL, the opposite was true, that is, KN1 RSI (-1.05 days) and KN2 RSI (-1.10 days) under both sets of economic weights were more effective than RESIM (-0.92 days) for maturity. The effective selection gain per cycle estimated as the linear regression of the mean genotypic trait value on the selection cycle is also shown in the last row of Tables 3 and 4.

Table 3. Mean genotypic values of the 20 genotypes selected under ESIM and Smith SI 1 and 2, for five traits, male flowering (MFL), female flowering (FFL), plant height (PHT), ear height (EHT), and 100-kernel weight (HKF), during five selection cycles for simulated data using phenotypic and genotypic variance-covariance matrices. The gain per cycle is the regression coefficient of the mean genotypic values regressed on the selection cycles. The signs and economic weights of the SIs for each trait are shown in parentheses.

	ESIM					SMITH SI 1					SMITH SI 2				
Selection cycle	MFL (-)	FFL (-)	PHT (-)	EHT (-)	HKF (+)	MFL (-1)	FFL (-1)	PHT (-1)	EHT (-1)	HKF (+1)	MFL (-0.51)	FFL (-0.46)	PHT (-0.38)	EHT (-0.52)	HKF (0.27)
0	98.5	98.9	139.6	88.4	20.4	98.5	98.9	139.6	88.4	20.4	98.5	98.9	139.6	88.4	20.4
1	95.0	96.4	132.7	75.3	20.2	98.6	99.3	124.1	71.6	20.4	98.5	99.2	123.1	72.1	20.6
2	95.2	95.6	125.1	63.6	21.0	101.0	100.5	114.9	64.0	20.6	97.4	99.1	116.3	66.8	21.0
3	94.0	94.2	123.3	58.0	21.2	98.4	99.1	111.6	59.6	20.6	96.3	98.9	112.5	62.6	21.1
4	93.3	93.4	123.3	57.5	21.6	94.6	97.6	111.6	59.5	21.0	94.4	98.8	111.7	61.4	21.2
5	92.6	93.3	122.2	57.5	22.6	94.5	97.5	111.2	58.4	21.5	94.5	99.3	111.3	60.5	21.4
Gain per selection cycle	0.19	-0.92	-0.18	-3.45	0.48	0.05	-1.05	-1.15	-4.53	0.27	0.02	-1.10	0.58	-3.44	0.19

Table 4. Mean genotypic values of the 20 genotypes selected under RESIM, KN1 SRI, and KN2 SRI, for traits female flowering (FFL), ear height (EHT), and 100-kernel weight (HKF) during five selection cycles for simulated data using phenotypic and genotypic variance-covariance matrices. The gain per cycle is the regression coefficient of the mean genotypic values regressed on the selection cycles. The signs and economic weights of the SIs for each trait are shown in parentheses. The restrictive traits are male flowering (MFL) and plant height (PHT).

	RESIM					KN1 SRI					KN2 SRI				
Selection cycle	MFL (-)	FFL (-)	PHT (-)	EHT (-)	HKF (+)	MFL (-1)	FFL (-1)	PHT (-1)	EHT (-1)	HKF (+1)	MFL (-0.51)	FFL (-0.46)	PHT (-0.38)	EHT (-0.52)	HKF (0.27)
0	98.5	98.9	139.6	88.4	20.4	98.5	98.9	139.6	88.4	20.4	98.5	98.9	139.6	88.4	20.4
1	97.4	96.5	141.3	84.1	20.2	97.4	96.3	141.2	84.2	20.2	97.7	96.33	140.4	82.8	20.0
2	97.7	94.8	143.6	80.2	20.6	99.4	95.6	143.5	80.3	20.5	98.8	95.2	145.1	81.6	20.5
3	97.2	94.0	141.2	76.4	21.7	98.2	94.7	139.8	74.4	21.1	97.6	94.0	143.1	75.9	20.1
4	98.4	94.5	140.7	76.7	22.4	97.4	93.9	137.7	70.5	21.4	97.7	93.7	145.2	75.5	21.1
5	99.2	93.9	139.9	71.1	22.1	98.9	93.3	135.2	67.8	21.4	98.7	93.1	141.9	71.58	21.1
Gain per selection cycle	0.19	-0.92	-0.18	-3.45	0.48	0.05	-1.05	-1.15	-4.53	0.27	0.02	-1.10	0.58	-3.44	0.19

Figs. 1-3 show the genotypic means for HKF, FFL, and MFL for five selection cycles when the genotypes are selected under different selection indices. Increasing trends in the genotypic means of the selected genotypes for the five selection cycles under MESIM, LANDE-THOMPSON 1 and 2, ESIM, SMITH 1 and 2, RESIM and KEMPTHORNE-NORDKOG for 100-kernel weight (HKF) are shown in Fig. 1. Clearly, MESIM selected genotypes with higher HFK in all cycles. For FFL (Fig. 2) ESIM was the best in all cycles, whereas MESIM was better than LANDE-THOMPSON 1 and 2 in the last three cycles. For MFL, Fig. 3 shows that MESIM results are similar to those of LANDE-THOMPSON 1 and 2. However, ESIM is still the selection index that gave the highest response to selection. Furthermore, note that since MFL was unchanged when applying the restricitve selection indices (RESIM, KN1, and KN2), their genotyic means did not change over the selection cycles and stayed around the mean of cycle 0 (Fig. 3).

As previously indicated, the molecular selection indices (MESIM and LANDE-THOMPSON) depend on the heritability of teach trait. According to LANDE and THOMPSON (1990), ZHANG and SMITH (1992, 1993), GIMELFARB and LANDE (1994, 1995), and WHITTAKER (2003), the molecular selection index is expected to be more efficient than the standard selection indices (i.e., ESIM and SMITH's selection index) when the heritability of the trait is low. Fig. 1 shows the genotypic means of 100-kernel weight (HKF) with a heritability of 0.27, whereas Figs. 2 and 3 depict the genotypic means of the selected genotypes for female flowering time (FFL), and male flowering time (MFL), with heritabilities of 0.46 and 0.51, respectively. This would explain why MESIM was more efficient than the other indices for selecting the genotypes with the highest genotypic means. Detail descriptions of ESIM, RESIM, and the SMITH selection index can be found in CERON- ROJAS *et al*. (2008). For the other traits, the gains of MESIM over LANDE-THOMPSON 1 and 2 are not as clear as for HKF and FFL (Tables 2-4). However, when traits are selected individually, the genotypic mean obtained for MESIM is higher than that achieved by LANDE-THOMPSON for most traits (Table 1).

Fig. 1

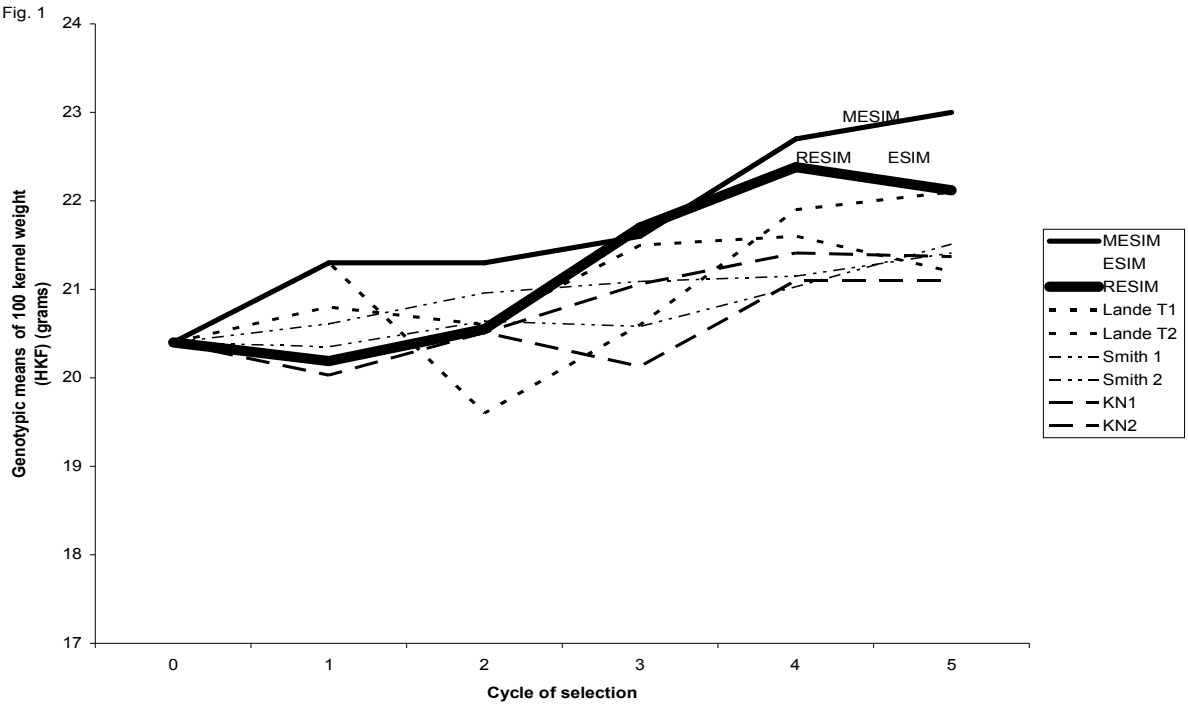

Fig. 1. Mean of the genotypic values of the selected genotypes under under MESIM, LANDE-THOMPSON (Lande T1 and Lande T2) molecular selection indices, ESIM, SMITH selection indices (Smith 1 and 2) RESIM, and KEMPTHORNE-NORDSKOG restricted selection indices (KN1 and KN2) during five selection cycles of traits 100-kernel weight (HKF) (grams) using simulated data. The simultaneously selected traits were male flowering (MFL), female flowering (FFL), plant height (PHT), ear height (EHT), and 100-kernel weight (HKF). The weights used for MFL, FFL, PHT, EHT, and HKF under the LANDE-THOMPSON molecular selection indices, SMITH selection index, and KEMPTHORNE-NORDSKOG restricted selection index were -1, -1, -1, -1, and 1, respectively, and the heritability of the corresponding traits.

Fig. 2

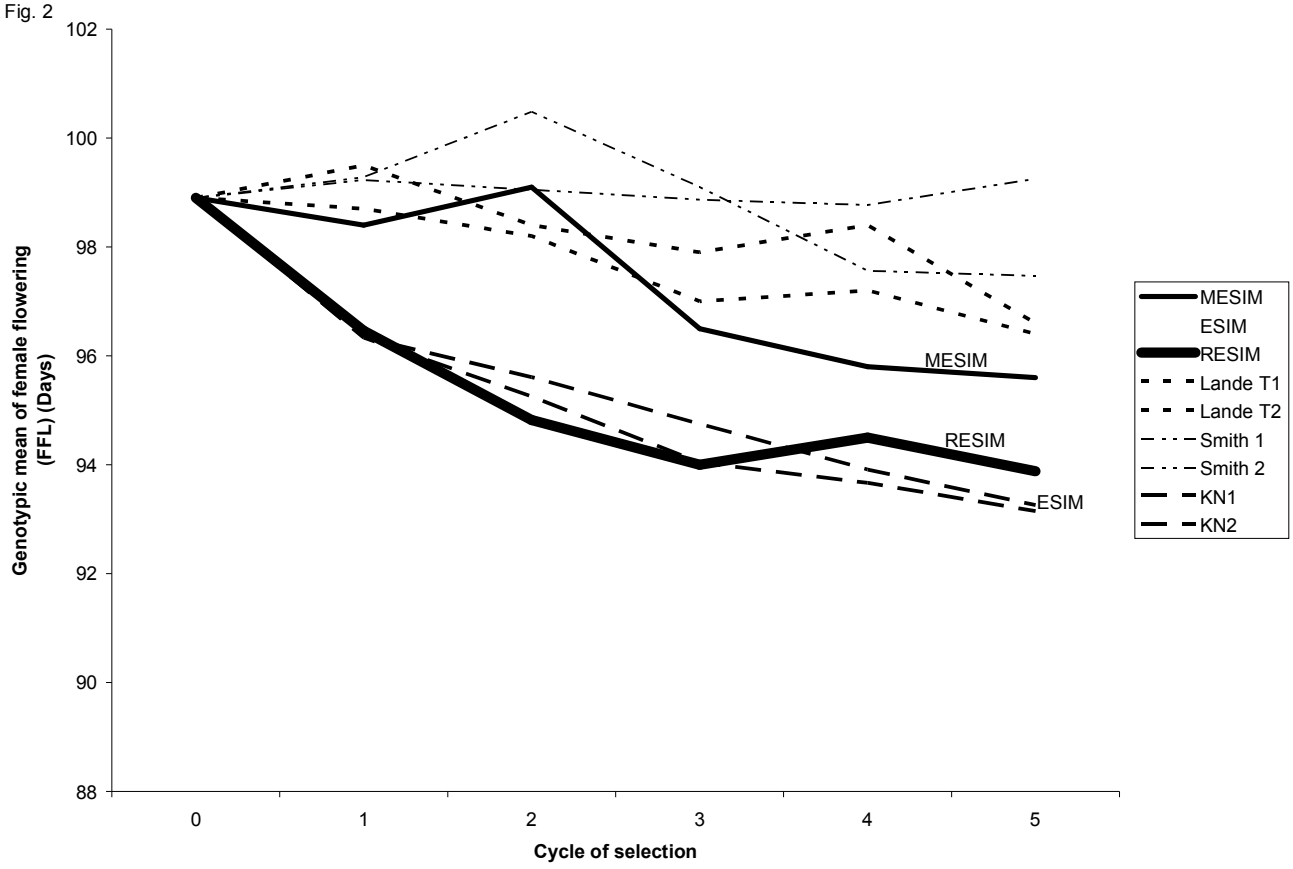

Fig. 2. Mean of the genotypic values of the selected genotypes under MESIM, LANDE-THOMPSON (Lande T1 and lande T2) molecular selection indices ESIM, SMITH SI (Smith 1 and 2) RESIM, and KEMPTHORNE-NORDSKOG restricted selection index (KN1 and KN2) for five selection cycles of trait female flowering (FFL) (days) using simulated data. The simultaneously selected traits were male flowering (MFL), female flowering (FFL), plant height (PHT), ear height (EHT), and 100-kernel weight (HKF). The weights used for MFL, FFL, PHT, EHT, and HKF under the LANDE-THOMPSON molecular selection indices, SMITH selection index, and KEMPTHORNE-NORDSKOG restricted selection index were -1, -1, -1, -1, and 1, respectively, and the heritability of the corresponding traits.

Fig. 3

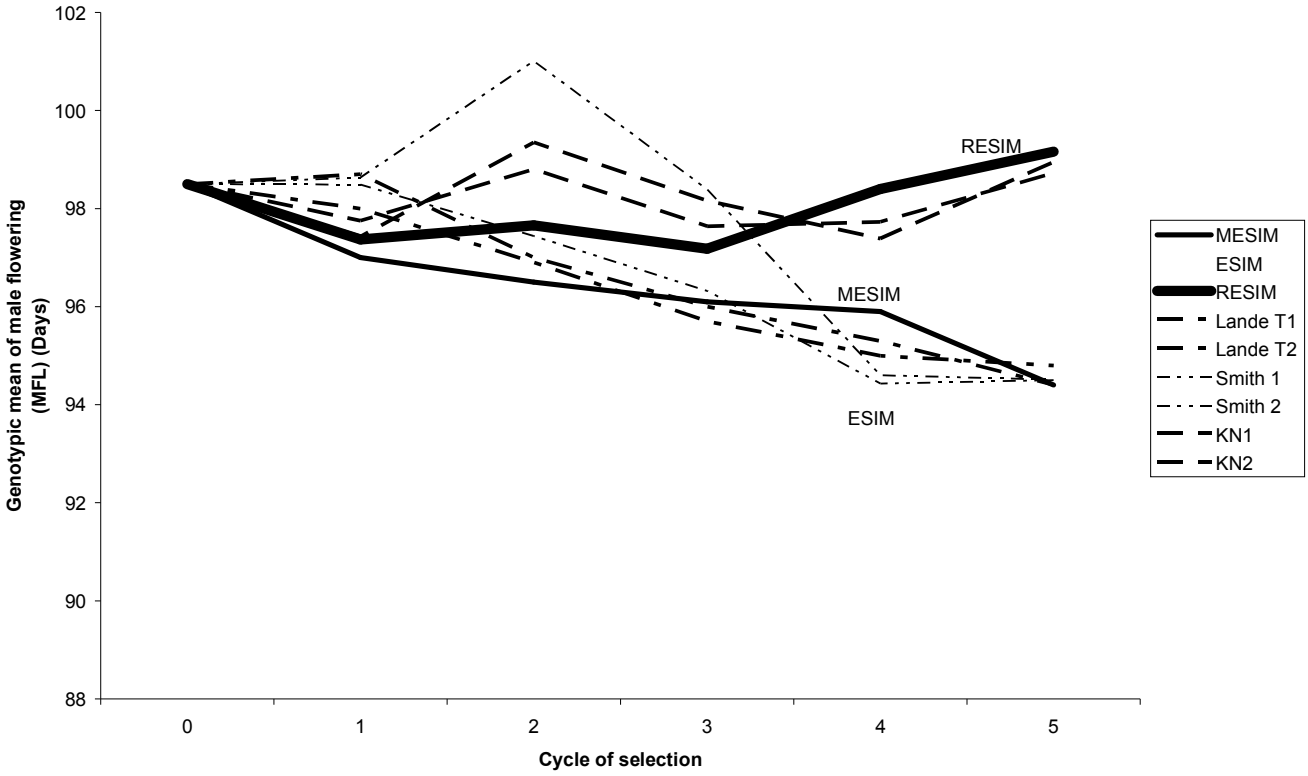

Fig. 3. Mean of the genotypic values of the selected genotypes under MESIM, LANDE-THOMPSON (Lande T1 and Lande T2) molecular selection indices ESIM, SMITH SI (Smith 1 and 2) RESIM, and KEMPTHORNE-NORDSKOG restricted selection index (KN1 and KN2) for five selection cycles of trait female flowering (MFL) (days) using simulated data. The simultaneously selected traits were male flowering (MFL), female flowering (FFL), plant height (PHT), ear height (EHT), and 100-kernel weight (HKF). The weights used for MFL, FFL, PHT, EHT, and HKF under the LANDE-THOMPSON molecular selection indices, SMITH selection index, and KEMPTHORNE-NORDSKOG restricted selection index were -1, -1, -1, -1, and 1, respectively, and the heritability of the corresponding traits.

It is worth noting that when the eigenvectors are obtained from the variance-covariance phenotypic and genotypic matrices, then MESIM, ESIM, and RESIM assign weights proportional to the heritability of the trait, that is, the higher the heritability, the more weight, and vice versa. As mentioned by CERON-ROJAS *et al*. (2006), a solution would be to use the phenotypic and genotypic correlation matrices. Another solution would be to use the inverse of $\mathbf{Q}$ ($\mathbf{Q}^{-1}$) and thus give more weight to traits with low heritability. The latter solution for constructing MESIM comes naturally from Eq. 7, since $(\mathbf{Q} - \mu\mathbf{I})\boldsymbol{\beta}_M = \mathbf{0}$ and be written as $\mathbf{Q}\boldsymbol{\beta}_M = \mu\boldsymbol{\beta}_M$, from where $\mu^{-1}\boldsymbol{\beta}_M = \mathbf{Q}^{-1}\boldsymbol{\beta}_M$. Then the equation to obtain the eigenvectors is $(\mathbf{Q}^{-1} - \mu^{-1}\mathbf{I})\boldsymbol{\beta}_M = \mathbf{0}$, in which case $\mathbf{Q}^{-1} = \boldsymbol{\Sigma}_M^{-1}\mathbf{S}_M = \begin{bmatrix} (\boldsymbol{\Sigma} - \mathbf{M})^{-1}(\mathbf{S} - \mathbf{M}) & \mathbf{0} \\ \mathbf{I} - (\boldsymbol{\Sigma} - \mathbf{M})^{-1}(\mathbf{S} - \mathbf{M}) & \mathbf{I} \end{bmatrix}$; when only one trait and its molecular scores are considered, then $\mathbf{Q}^{-1} = \begin{bmatrix} \dfrac{s^2 - \sigma_m^2}{\sigma_g^2 - \sigma_m^2} & 0 \\ \dfrac{\sigma_g^2 - s^2}{\sigma_g^2 - \sigma_m^2} & 1 \end{bmatrix}$, and when $\sigma_m^2 = 0$, $\mathbf{Q}^{-1} = \begin{bmatrix} \dfrac{1}{h^2} & 0 \\ 0 & 0 \end{bmatrix}$, from where it is evident that traits with low heritability will have higher weights.

Finally, it is worth noting that although MESIM, ESIM, and RESIM may occasionally not to turn out to be the indices with the highest selection gains, they have the statistical properties of the principal components. According to OKAMOTO (1969), these are optimal properties established in terms of maximization and minimization. Thus the first component has the largest variance and the smallest loss of information (RAO 1964). On the other hand, statistical properties of other selection indices are unknown.

This research found that MESIM has three advantages over LANDE-THOMPSON 1 and 2: first, it can be used to solve practical problems faced by breeders attempting to select plants or animals for the next generation when no estimates of economic weights are available. Even if economic weights are available, in practice it is very unlikely that they would maximize the derivative of $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M$ with respect to $\boldsymbol{\beta}_M$ and to $\boldsymbol{\theta}_M$ (under the imposed restrictions). Furthermore, if two breeders are interested in improving, say, *n* traits, it is very unlikely that they would assign the same weights to them. Second, estimates of MESIM have known statistical sampling properties, but estimates for the LANDE-THOMPSON molecular selection index are unknown. Third, results from MESIM using simulated data show that realized genetic gains for various traits simultaneously are similar to, or higher than, those obtained by LANDE-THOMPSON (1990).

CONCLUSIONS

This research presents a molecular selection index based on principles developed by CERÓN-ROJAS et al. (2008). Simulated results show that when genotypes are selected based on individual traits, MESIM increased the response to selection over the LANDE-THOMPSON. When several traits are selected simultaneously, MESIM outperformed LANDE-THOMPSON for traits with low heritability. For traits with high heritability, ESIM performed very well. One of the most important results of MESIM is that $\hat{\boldsymbol{\beta}}_{MESIM}$ is the maximum likelihood estimate of $\boldsymbol{\beta}_{MESIM}$, whereas $\hat{\boldsymbol{\beta}}_{MSI}$ is an estimate of $\boldsymbol{\beta}_{MSI}$, whose sampling properties are unknown. MESIM can be considered a generalization of ESIM

(CERÓN-ROJAS et al. 2006) when information on QTLs is incorporated through molecular markers. The sampling properties of ESIM (and therefore of MESIM) and its selection response are known, and its estimators showed desirable statistical properties such as consistency and asymptotic unbiasedness.

It should be pointed out that MESIM is more general than ESIM (CERÓN-ROJAS et al. 2006) because the basic underlying assumption made in ESIM, $\boldsymbol{\Sigma\theta} = \boldsymbol{\beta}$, is relaxed in MESIM. MESIM maximizes the selection response by maximizing the square of the correlation between Y_M and Z_M, $\rho^2_{Y_M Z_M}$, which is the same as maximizing $\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M\right)^2$. This basic idea, used for developing a restrictive selection index (CERÓN-ROJAS et al. 2008), is valid for MESIM when no restrictions are imposed on any of the traits.

Some advantages of MESIM over MSI should be pointed out: (1) the sampling properties of MESIM, $\hat{R}_{MESIM}$, are known and easy to evaluate; (2) the MESIM eigenvalue and eigenvector are estimated by the maximum likelihood method; and (3) a restrictive SI can be developed from MESIM when only some markers and/or traits are used. In summary, the results of this study indicate that MESIM is a generalization of ESIM when information on QTLs linked to molecular markers is incorporated.

The availability of abundant molecular markers can help to achieve faster breeding progress than with traditional breeding methods or marker-assisted selection by means of genome wide selection (BERNARDO and YU 2007). The MESIM could be a valid option for a genome-wide selection method because the serious problem of parameter identification create by the collinearity of the markers is overcome by the singular value decomposition method of MESIM. Furthermore, MESIM naturally perform crossproduct

between all traits-environments combinations, and markers; thus it implicitely introduces estimates of particular epistatic interactions into the seletion index. Further research on the used of MESIM in genome wide selection is required.

APPENDIX

Theoretical derivation of MESIM

The procedure shown below is a slight modification of that used by CERÓN-ROJAS et al. (2008) within the context of a restricted selection index method based on eigenanalysis (RESIM). In this case, $(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M)^2$ must be maximized under the restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$, i.e., we should maximize

$$\Phi = (\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M)^2 - \mu(\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M - 1) - \omega(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M - 1)$$

with respect to $\boldsymbol{\beta}_M$, $\boldsymbol{\theta}_M$, μ, and ω, where $\boldsymbol{\beta}_M$ is the vector of MESIM coefficients, $\boldsymbol{\theta}_M$ is the vector of economic weights, and μ and ω are Lagrange multipliers. In MESIM it is assumed that $\boldsymbol{\theta}_M$ is not a vector of constants.

When Φ is derived with respect to $\boldsymbol{\beta}_M$, $\boldsymbol{\theta}_M$, μ, and ω, and the result is set to the null vector, it follows that

$$(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M) \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M - \mu \mathbf{S}_M \boldsymbol{\beta}_M = \mathbf{0} \qquad [A.1]$$

$$(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M) \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M - \omega \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = \mathbf{0} \qquad [A.2]$$

$$\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1 \qquad [A.3]$$

$$\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1 \qquad [A.4]$$

where Eqs. A.3 and A.4 denote the restrictions imposed for the maximization of $(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M)^2$ with respect to $\boldsymbol{\beta}_M$ and $\boldsymbol{\theta}_M$. Because the restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$, when Eq. A.1 is multiplied by $\boldsymbol{\beta}'_M$ and Eq. A.2 is multiplied by $\boldsymbol{\theta}'_M$, both equations can be written as

$$(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M)^2 - \mu = 0$$

$$(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M)^2 - \omega = 0$$

Clearly, $(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M)^2 = \omega = \mu$. Therefore, μ maximizes $\rho^2_{Y_M Z_M}$ under the restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$.

The following problem is to determine the vector $\boldsymbol{\beta}_M$, which allows constructing the selection index Y_M that has maximum correlation with Z_M. Because $(\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M)^2 = \omega = \mu$, Eqs. A.1 and A.2 can be written as

$$\sqrt{\mu}\, \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M - \mu \mathbf{S}_M \boldsymbol{\beta}_M = \mathbf{0} \qquad \text{[A.5]}$$

$$\sqrt{\mu}\, \boldsymbol{\Sigma}_M \boldsymbol{\beta}_M - \mu \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = \mathbf{0} \qquad \text{[A.6]}$$

Multiplying Eq. A.5 by $\mu^{-1/2} \boldsymbol{\Sigma}_M^{-1}$, we obtain that $\boldsymbol{\theta}_M - \sqrt{\mu} \boldsymbol{\Sigma}_M^{-1} \mathbf{S}_M \boldsymbol{\beta}_M = \mathbf{0}$, from where $\boldsymbol{\theta}_M = \sqrt{\mu} \boldsymbol{\Sigma}_M^{-1} \mathbf{S}_M \boldsymbol{\beta}_M$ is computed.

Substitute, in Eq. A.6, $\sqrt{\mu} \boldsymbol{\Sigma}_M^{-1} \mathbf{S}_M \boldsymbol{\beta}_M$ for $\boldsymbol{\theta}_M$ and get $\boldsymbol{\Sigma}_M \boldsymbol{\beta}_M - \mu \mathbf{S}_M \boldsymbol{\beta}_M = \mathbf{0}$, from where Eq. 7 (see the text) is obtained

$$(\mathbf{Q} - \mu \mathbf{I}) \boldsymbol{\beta}_M = \mathbf{0}$$

where $\mathbf{Q} = \mathbf{S}_M^{-1} \boldsymbol{\Sigma}_M$, and μ and $\boldsymbol{\beta}_M$ are the eigenvalue and eigenvector of $\mathbf{Q}$, respectively. Thus, for MESIM, the values that maximize $\rho^2_{Y_M Z_M}$ under the restrictions $\boldsymbol{\beta}'_M \mathbf{S}_M \boldsymbol{\beta}_M = 1$ and

$\boldsymbol{\theta}'_M \boldsymbol{\Sigma}_M \boldsymbol{\theta}_M = 1$ are the eigenvalues (μ) of the matrix $\mathbf{Q}$ and its eigenvector vector, $\boldsymbol{\beta}_M$, that allows constructing the index $Y_{MESIM} = \boldsymbol{\beta}'_{MESIM}\mathbf{p}$, that maximizes its correlation with $Z_M = \boldsymbol{\theta}'_1\mathbf{g} + \boldsymbol{\theta}'_2\mathbf{m}$.

The authors are grateful to Jiankang Wang for his valuable assistant and help when running the simulation software QU-GENE. The authors thank the Associate Editor and two anonymous reviewers for their comments and suggestions, which significantly improved the quality of this paper. The authors are thankful to BIMBO-Mexico for partially funding this research.

LITERATURE CITED

ANDERSON, T. W., 2003 *An Introduction to Multivariate Statistical Analysis*. 3rd Ed. John Wiley and Sons, New Jersey.

ARUS, P., and J. MORENO-GONZALEZ, 1993 Marker-assisted selection, pp. 315-331 in Plant Breeding: Principles and Prospects. Hayward, M. D., N.O. Bosemark, and I. Romagosa (eds.). Chapman and Hall, University Press, Cambridge, Great Britain.

BAKER, R. J., 1986 *Selection Indices in Plant Breeding*. CRC Press Inc., Boca Raton, Florida.

BERNARDO, R. and J. YU, 2007 Prospects for genomewide selection for quantitative traits in maize. Crop Science 47:1082-1090.

CERÓN-ROJAS, J. J., J. CROSSA, J. SAHAGÚN-CASTELLANOS, F. CASTILLO-GONZÁLEZ, and A. SANTACRUZ-VARELA, 2006 A selection index method based on eigenanalysis. Crop Science 46:1711-1721.

CERÓN-ROJAS, J. J., J. SAHAGÚN-CASTELLANOS, F. CASTILLO-GONZÁLEZ, A. SANTACRUZ-VARELA, and J. CROSSA, 2008 A restricted selection index method based on eigenanalysis. Journal of Agriculture, Biology and Environmental Statistics 13 (4): 440-457.

DEKKERS, J. C. M., and M. R. DENTINE, 1991 Quantitative genetic variation associated with chromosomal markers in segregating populations. Theoretical and Applied Genetics 81:212-220.

FALCONER, D. S., and T. F. C. MACKAY, 1997 Introduction to Quantitative Genetics. Longman, New York, 464 pp.

GIMELFARB, A., and R. LANDE, 1994 Simulation of marker-assisted selection in hybrid populations. Genetical Research 63:39-47.

GIMELFARB, A., and R. LANDE, 1995 Marker-assisted selection and marker-QTL associations in hybrid populations. Theoretical and Applied Genetics 91:522-528.

HARRIS, D. L., 1964 Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20:46-72.

HAYES, J. F., and W. G. HILL, 1980 A reparameterization of a genetic selection index to locate its sampling properties. Biometrics 36:237-248.

HAZEL, L. N., 1943 The genetic basis for constructing a selection index, pp. 316-330 in Papers on Quantitative Genetics and Related Topics. Department of Genetics, North Carolina State College, Raleigh, North Carolina, USA.

JANSEN, R. C., 2003 Quantitative trait loci in inbred lines, Vol. I, pp. 445-476 in Handbook of Statistical Genetics, 2nd Edition. Balding, D.J., M. Bishop and C. Cannings (eds.). John Wiley and Sons, England.

KEMPTHORNE, O., and A. W. NORDSKOG, 1959 Restricted selection indices. Biometrics 15:10-19.

LANDE, R., and R. THOMPSON, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

MACKINNON, M. J., and M. A. J. GEORGES, 1992 The effects of selection on linkage analysis for quantitative traits. Genetics 132:1177-1185.

MARDIA, K. V., J. T. KENT, and J. M. BIBBY, 1982 *Multivariate Analysis*. Academic Press Inc, New York, USA.

MOREAU, L., S. LEMARIE, A. CHARCOSSET, and A. GALLAIS, 2000 Economic efficiency of one cycle of marker-assisted selection efficiency. Crop Science 40:329-337.

OKAMOTO, M., 1969 Optimality of principal components. In *Multivariate Analysis* II, ed. P. R. Krishnaiah, 673-685. New York, Academic Press.

RAO, C.R., 1964 The use and interpretation of principal component analysis in applied research. In: *Multivariate Statistical Methods: Within-Groups Covariation*. Bryant E. H., and W. R. Atchley (editors). Halsted Press, Pennsylvania.

SMITH, H. F., 1936 A discriminant function for plant selection, pp. 466- 476 in Papers on Quantitative Genetics and Related Topics. Department of Genetics, North Carolina State College, Raleigh, North Carolina, USA.

SMITH, O. S., A. R. HALLAUER, and W. A. RUSSELL (1981). Use of index selection in recurrent selection programs in maize. *Euphytica* 30:611-618.

VAN VLECK, L. D., 1993 *Selection Index and Introduction to Mixed Model Methods*. CRC Press Inc., Boca Raton, Florida.

WANG, J., M. VAN GINKEL, R. TRETHOWAN, G. YE, I. DELACY, D. PODLICH, and M. COOPER, 2004 Simulating the effects of dominance and epistasis on selection response in the CIMMYT wheat breeding program using QuCim. Crop Science 44:2006-1018.

WEI, M., A. CABALLERO, and W. G. HILL, 1996 Selection response in finite populations. Genetics 144:1961-1974.

WHITTAKER, J.C., 2003 Marker-assisted selection and introgression, Vol. I, pp. 554-574 in Handbook of Statistical Genetics, 2nd Edition. Balding, D.J., M. Bishop, and C. Cannings (eds.). John Wiley and Sons, England.

WU, R., Z. B. ZENG, S. E. MCKEAND, and D. M. O'MALLEY, 2000 The case for molecular mapping in forest tree breeding. Plant Breeding Reviews 19:41-68.

XIE, C., and S. XU, 1998 Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. Heredity 8:489-498.

ZHANG, W., and C. SMITH, 1992 Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theoretical and Applied Genetics 83:813-820.

ZHANG, W., and C. SMITH, 1993 Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. Theoretical and Applied Genetics 86:492-496.

Evaluating and comparing genome wide and marker assisted selection indices[ϒ]

J. Jesús Cerón-Rojas, Fernando Castillo-González, Jaime Sahagún-Castellanos, Amalio Santacruz-Varela, Ignacio Benítez-Riquelme, Jiankang Wang, and José Crossa

Resumen

En el contexto del mejoramiento de plantas y animales, la teoría de los índices de selección (IS) está basada en la maximización de la correlación entre el IS y el valor reproductivo para que la respuesta a la selección sea máxima. Ésta es la idea subyacente del IS fenotípico, el IS molecular de Lande y Thompson (LT), el método del índice de selección molecular eigen (*Molecular Eigen Selection Index Method*, o MESIM por sus siglas en inglés) de Cerón-Rojas, y el IS molecular del genoma completo de Lange y Whittaker (LW). Esta investigación tuvo los siguientes objetivos: (1) extender la teoría de MESIM al caso del IS molecular del genoma completo (*Genome Wide Molecular Eigen Selection Index Method* $MESIM_{GW}$, por sus siglas en inglés), utilizando la inversa generalizada de Moore-Penrose; (2) utilizar el IS de LW con la inversa generalizada de Moore-Penrose y comparar sus las ganancias genéticas con las obtenidas con $MESIM_{GW}$; y (3) comparar los resultados de LW y $MESIM_{GW}$ con las obtenidas al utilizar el IS de LT y MESIM. Cinco criterios de evaluación fueron utilizados para comparar las ganancias genéticas obtenidas por los diferentes ISs. Los resultados teóricos se ilustran por medio de (1) dos conjuntos de datos simulados: una población doble haploide (DH) y una población F_2 segregante con 468 marcadores moleculares utilizando dos tamaños de muestra (200 y 500 genotipos); y (2) un conjunto de datos reales de un estudio en una población de maíz F_3 .Los resultados mostraron ventajas de los ISs molecular del genoma completo sobre el IS de LT y MESIM, e indicaron que, en la población DH, el avance genético fue mayor que en la población F_2 .

Palabras clave: Índice de selección, eigen-análisis, selección del genoma completo, inversa generalizada Moore-Penrose.

[ϒ] Capítulo enviado para su revisión y posible publicación a Genetics Research.

Abstract

In the context of animal and plant breeding, the underlying theory of the selection index (SI) is based on maximizing the correlation between the SI and the breeding value such that the genetic response in each selection cycle is maximized. This is the underlying idea of phenotypic SIs, the molecular marker (MM) assisted (MAS) SI of Lande and Thompson (LT), the Molecular Eigen Selection Index Method (MESIM) of Cerón-Rojas, and the genome wide (GW) molecular marker SI of Lange and Whittaker (LW). This research had the following objectives: (1) to extend the MESIM theory to the case of genome wide MM selection index, $MESIM_{GW}$, using the Moore-Penrose generalized inverse; (2) to use the LW SI with the Moore-Penrose generalized inverse and compare its genetic gains with those based on $MESIM_{GW}$; and (3) to compare results from LW and $MESIM_{GW}$ with those obtained using MAS LT SI and MESIM. Five evaluation criteria were used to compare the genetic gains obtained from the different SIs. We illustrated the theoretical results using (1) two simulated data sets: a doubled haploid (DH) population and F_2 segregating population with 468 MM using two sample sizes (200 and 500 genotypes); and (2) a real data set from a study on an F_3 maize population. Results showed the advantages of genome wide SIs over MAS SIs, and indicated that, in the DH population, the genetic advances were greater than in the F_2.

Key words: selection index, eigenanalysis, genome wide molecular marker selection, Moore-Penrose generalized inverse.

Introduction

In the context of animal and plant breeding, the underlying theory of the selection index (SI) is based on the linear combination $Y = \boldsymbol{\beta}'\mathbf{p}$, where Y is the SI, $\mathbf{p}$ is the vector of the phenotypic values, and $\boldsymbol{\beta}$ is the vector of the coefficients of $\mathbf{p}$. The basic idea is to maximize the correlation between Y and the breeding value $Z = \boldsymbol{\theta}'\mathbf{g}$ (where $\mathbf{g}$ is the vector of genotypic values, and $\boldsymbol{\theta}$ is the vector of economic weights), ρ_{YZ}, with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, such that the genetic response in each selection cycle is maximized. This is the underlying idea of the phenotypic SI of Smith (1936), the molecular marker (MM) assisted (MAS) SI of Lande and Thompson (1990) (LT), and the genome wide (GW) molecular marker SI of Lange and Whittaker (2001) (LW). In this case, the principal differences between these three SIs are based on the type of information they incorporate into Y because they use the same mathematical method to estimate $\boldsymbol{\beta}$. While Smith's SI utilizes only phenotypic values, the LT SI uses phenotypic information plus MM scores representing the additive values of the most important QTLs; the LW SI incorporates into Y both phenotypic information and all the MMs as additional random variables.

Let $\mathbf{G}$ and $\mathbf{P}$ be the variance-covariance matrices of genotypic and phenotypic values, respectively, and assume the economic weights, $\boldsymbol{\theta}$, are known, then in the Smith's SI, $\boldsymbol{\beta} = \mathbf{P}^{-1}\mathbf{G}\boldsymbol{\theta}$, maximizes ρ_{YZ} and the selection response. Now let $\mathbf{s}$ be the vector of the MM scores, $\mathbf{p}'_{LT} = [\mathbf{p}' \quad \mathbf{s}']$, and $\boldsymbol{\beta}'_{LT} = [\boldsymbol{\beta}' \quad \boldsymbol{\beta}'_s]$, where $\boldsymbol{\beta}$ is the vector of trait phenotypic ($\mathbf{p}$) weights, and $\boldsymbol{\beta}_s$ is the vector of the MM scores ($\mathbf{s}$) weights. Furthermore, suppose that matrices $\mathbf{T}$ and $\mathbf{W}$ comprise $\mathbf{G}$ and $\mathbf{P}$ (as previously defined) and the variance-covariance matrix of the MM scores, $\mathbf{S} = \mathrm{Var}(\mathbf{s})$; then, in the MAS LT SI, $\boldsymbol{\beta}_{LT} = \mathbf{T}^{-1}\mathbf{W}\boldsymbol{\theta}_{LT}$ (the subscript LT denotes the Lande and Thompson MAS SI; $\boldsymbol{\theta}'_{LT} = [\boldsymbol{\theta}' \quad \mathbf{0}']$, $\boldsymbol{\theta}$ is the vector of economic weights, and $\mathbf{0}$ is a vector of zeros), allows constructing the LT SI $Y_{LT} = \boldsymbol{\beta}'_{LT}\mathbf{p}_{LT}$ which has maximum correlation with $Z = \boldsymbol{\theta}'\mathbf{g}$. For the case of LW SI, $\boldsymbol{\beta}_{LW} = \boldsymbol{\Gamma}^{-}\boldsymbol{\Sigma}\boldsymbol{\theta}_{LW}$ (matrix $\boldsymbol{\Gamma}^{-}$ is a generalized inverse of $\boldsymbol{\Gamma}$, and $\boldsymbol{\theta}'_{LW} = [\boldsymbol{\theta}' \quad \mathbf{0}']$ is the genome wide vector of economic weights) allows constructing the genome wide molecular

marker LW SI $Y_{LW} = \boldsymbol{\beta}'_{LW}\mathbf{p}_{LW}$ (the subscript LW denotes the Lange and Whitaker GW molecular SI), which has maximum correlation with the breeding value $Z = \boldsymbol{\theta}'\mathbf{g}$. Matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are variance-covariance matrices comprising $\mathbf{G}$ and $\mathbf{P}$ (as previously defined), the variance-covariance matrix of the genotyped molecular markers, $\mathbf{M} = Var(\mathbf{m})$ ($\mathbf{m}$ is the random vector of coded values of the molecular makers), and the covariance matrix of the QTL additive effects linked to MMs with genotypic values ($\hat{\mathbf{G}}_M$); $\boldsymbol{\beta}'_{LW} = [\boldsymbol{\beta}' \quad \boldsymbol{\beta}'_m]$ is the genome wide vector containing phenotypic ($\boldsymbol{\beta}$) and molecular marker ($\boldsymbol{\beta}_m$) weights. When there are large numbers of correlated markers and a limited number of records in GW SI, the estimate of the variance matrix of the molecular markers, $\mathbf{M}$, is singular; therefore, the standard inverse of $\boldsymbol{\Gamma}$ does not exist, and a generalized inverse method should be used for constructing the GW SI (Lange and Whittaker 2001).

The previous SIs differ in the type of information they incorporate into Y, which is reflected in the form of $\boldsymbol{\beta}$ that maximizes ρ_{YZ}. An additional difference between the SIs is based on the mathematical method used to maximize ρ_{YZ}; this has been demonstrated by Cerón-Rojas et al. (2008a), who derived the theory of the Eigen Selection Index Method (ESIM) using the underlying concept of canonical correlation and the basic ideas of Kempthorne and Nordskog (1959), who showed that maximizing ρ_{YZ} is equivalent to maximizing ρ^2_{YZ}. Similarly, in the context of MAS, Cerón-Rojas et al. (2008b) derived the theory of the Molecular Eigen Selection Index Method (MESIM) by maximizing ρ^2_{YZ}. Like the MAS LT SI, MESIM requires identifying the linkage between the MM and the QTL, the estimated effect of the QTL linked to the MM, and the combination of the molecular scores effects and phenotypic information that allows genotypes to be classified and selected using the selection index.

The underlying idea of ESIM and MESIM is that when several random variables belong to two groups and the objective is to maximize the correlation between the two groups, the canonical correlation is a natural and efficient method to use (Muirhead 2005). Consider two sets of random variables: phenotypic variables ($\mathbf{p}$) and genotypic variables ($\mathbf{g}$) with joint multivariate normal distribution. The canonical correlation analysis allows

maximizing the correlation between any two groups of random variables or any two linear combinations of random variables from each group, such as $Y = \boldsymbol{\beta}'\mathbf{p}$ and $Z = \boldsymbol{\theta}'\mathbf{g}$. Anderson (2003) pointed out that Y will better predict Z if ρ_{YZ} is maximized. This is important in the context of the selection index theory and its application in plant and animal breeding, because by maximizing ρ_{YZ}, it is likely that the individuals selected based on Y will indeed be those with the largest breeding values Z for those traits. For MESIM, it is important to determine the vector $\boldsymbol{\beta}_{MESIM}$ that allows constructing Y_{MESIM} that maximizes its correlation with $Z = \boldsymbol{\theta}'\mathbf{g}$. The required $\boldsymbol{\beta}_{MESIM}$ is the solution to the equation $(\mathbf{K} - \mu\mathbf{I})\boldsymbol{\beta}_{MESIM} = \mathbf{0}$, where $\mathbf{K} = \mathbf{T}^{-1}\mathbf{W}$ (T and W are the same as those previously defined for MAS LT SI). For MESIM, the value that maximizes ρ_{YZ}^2 is the first eigenvalue (μ) of matrix $\mathbf{K}$, and the vector $\boldsymbol{\beta}_{MESIM}$ is the first eigenvector of matrix $\mathbf{K}$. Similar to MAS LT SI, in MESIM few MMs are used; therefore, matrix $\mathbf{T}$ is likely to be non singular, and a unique standard inverse can be computed.

In the context of GWSI, when there are more markers than records, Lange and Whittaker (2001) have suggested using a generalized inverse of $\boldsymbol{\Gamma}$ ($\boldsymbol{\Gamma}^-$) different from the Moore-Penrose generalized inverse, and have pointed out that although the estimated index weights depend on the particular choice of generalized inverse, the SI value does not. However, since the SI is a function of the estimated index weights, it is not clear how the SI could be independent of the estimated index weights. Because there are several possible generalized inverses that could be used, there should be, at least in theory, several SI that could be constructed. Thus, it is important to construct a SI using a generalized inverse that is unique. The Moore-Penrose generalized inverse is unique and exists for square and rectangular matrices; when the matrix is non singular, the Moore-Penrose generalized inverse is equal to the standard inverse, which does not occur with the other generalized inverses (Kollo and von-Rosen 2005; Schott 2005).

The research described in this paper has the following objectives: (1) to extend the theory of the MESIM to the case of genome wide MM selection index, MESIM_{GW} (GW for genome wide) using the Moore-Penrose generalized inverse; (2) to use the LW SI with

the Moore-Penrose generalized inverse and examine how the genetic gains based on LW SI compare with those based on $MESIM_{GW}$; and (3) to compare results from LW and $MESIM_{GW}$ with those obtained using MAS LT SI and MESIM.

We illustrate the theoretical results using (1) two simulated data sets: one comprising doubled haploids (DH) and another one including an F_2 segregating population for estimating the performance of LT, ESIM, $MESIM_{GW}$ and LW selection indices, and (2) a real data set from a QTL mapping study in an F_3 maize population. In the two simulated cases, the phenotypic data include two traits and three environments, and the genotypic data include 468 MMs in GW SIs. For both populations, DH and F_2, we used two population sizes: 200 and 500 genotypes. The two sample sizes were included with the objective of evaluating the efficiency of GW with respect to MAS because, in general, there will be more variables than genotypes in GW than in MAS; for this reason, it will be necessary to use the Moore-Penrose generalized inverse in the GW SIs. When 200 genotypes are used, the GW SI must be constructed with the Moore-Penrose generalized inverse because more variables than genotypes are present. When 500 genotypes are simulated, the standard inverse can be employed. The standard way of evaluating the performance of various SIs is to compare their selection responses. However, in this research, we present a more complete approach for evaluating the performances of each SI, which includes five different evaluation criteria.

Statistical Methods

Using the Moore-Penrose generalized inverse in the LW and $MESIM_{GW}$ selection indices

The canonical correlation analysis using generalized inverses when the variance and covariance matrices in both groups of variables are of incomplete rank is similar to the case where the matrices are of full rank (Mardia et al. 1982). However, if a matrix $\boldsymbol{\Gamma}$ (where $\boldsymbol{\Gamma}$ is as defined in the introduction) is of incomplete rank, its generalized inverse $\boldsymbol{\Gamma}^{-}$ is not unique; therefore, it is necessary to find a generalized inverse $\boldsymbol{\Gamma}^{+}$ of $\boldsymbol{\Gamma}$ that is unique under

certain circumstances. Only when $\boldsymbol{\Gamma}^{+}$ is unique, will the prediction of breeding values using different selection indices be comparable. Nashed (1973) pointed out that a useful generalized inverse (1) should always exist; (2) should reduce to the standard inverse ($\boldsymbol{\Gamma}^{-1}$) when $\boldsymbol{\Gamma}$ is of full rank; and (3) should have some properties of the standard inverse, such as (3a) $(\boldsymbol{\Gamma}^{-1})^{-1} = \boldsymbol{\Gamma}$ and (3b) $(\boldsymbol{\Gamma}^{-1})' = (\boldsymbol{\Gamma}')^{-1}$.

A sufficient and necessary condition for $\boldsymbol{\Gamma}^{-}$ to be a generalized inverse of $\boldsymbol{\Gamma}$ is that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{-}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}$ (Kollo and von-Rosen 2005); however, this inverse does not achieve condition (3a) because $\boldsymbol{\Gamma}^{-}$ is the generalized inverse of $\boldsymbol{\Gamma}$, but $\boldsymbol{\Gamma}$ is not the generalized inverse of $\boldsymbol{\Gamma}^{-}$. According to Albert (1973), the Moore-Penrose generalized inverse ($\boldsymbol{\Gamma}^{+}$) is the simplest inverse of any matrix of incomplete rank, and (as shown in Appendix A) under certain conditions, it has desirable properties such as uniqueness and transitivity, that is, $\boldsymbol{\Gamma}^{+}$ is the generalized inverse of $\boldsymbol{\Gamma}$, and $\boldsymbol{\Gamma}$ is the generalized inverse of $\boldsymbol{\Gamma}^{+}$ (Kollo and von-Rosen 2005). Appendix A shows that the definition of $\boldsymbol{\Gamma}^{+}$ is explicitly related to the spectral decomposition when $\boldsymbol{\Gamma}$ is a symmetric matrix of incomplete rank. We will use the Moore-Penrose generalized inverse $\boldsymbol{\Gamma}^{+}$ to compare results from LW and MESIM_{GW}.

The general genome wide selection index

In this section, we present a general theoretical framework of a genome wide SI that will result in the LW SI or MESIM_{GW}, depending on the mathematical method used for estimating $\boldsymbol{\beta}$. As already mentioned, the type of information incorporated into the SI allows classifying the three previously defined SIs as phenotypic SI, MAS SI, and GW SI. Another way to distinguish the different SIs is through the mathematical method used for maximizing the correlation between Y and the breeding value Z; this method will determine the efficiency of the SIs. As for the genome wide SIs, it should be possible to hypothesize a general underlying model such that the difference between genome wide SIs (i.e., LW SI and MESIM_{GW}) is due to the mathematical method used to maximize the correlation between the SI and the breeding value.

Assume that the same genetic markers are typed in every individual, then a general genome wide molecular marker selection index can be denoted as

$$Y_M = \boldsymbol{\beta}'\mathbf{p} + \boldsymbol{\beta}'_m\mathbf{m} = [\boldsymbol{\beta}' \quad \boldsymbol{\beta}'_m]\begin{bmatrix}\mathbf{p}\\ \mathbf{m}\end{bmatrix} = \boldsymbol{\beta}'_M\mathbf{p}_M \qquad (1)$$

where $\boldsymbol{\beta}'_M = [\boldsymbol{\beta}' \quad \boldsymbol{\beta}'_m]$ and $\mathbf{p}'_M = [\mathbf{p}' \quad \mathbf{m}']$; $\boldsymbol{\beta}$ is the vector of weights for the phenotypic traits, $\boldsymbol{\beta}_m$ is the vector of MM weights, $\mathbf{p}$ is the vector of the phenotypic traits, and $\mathbf{m}$ is the genotyped molecular marker random vector, where the homozygous genotyped molecular marker takes values of 1 and -1, and the heterozygous genotyped molecular marker takes a value of 0 for cases of segregating populations such as F_2 populations, or values of 1 and -1 for doubled haploids denoting the presence or absence of the MM; a similar designation is given to the genotyped QTLs. The general selection response to Y_M can be written as

$$R_M = k\sigma_{Z_M}\rho_{Y_M Z_M} = k\sigma_{Z_M}\frac{\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\beta}_M}{\sqrt{\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M}\sqrt{\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M}} \qquad (2)$$

where $\boldsymbol{\Gamma} = \begin{bmatrix}\mathbf{P} & \mathbf{G}_M\\ \mathbf{G}'_M & \mathbf{M}\end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix}\mathbf{G} & \mathbf{G}_M\\ \mathbf{G}'_M & \mathbf{M}\end{bmatrix}$ are variance-covariance matrices comprising the phenotypic values variance-covariance matrix, $\mathbf{P}$; the genotypic variance-covariance matrix of the genotypic values, $\mathbf{G}$; the variance-covariance matrix of the genotyped molecular markers, $\mathbf{M} = Var(\mathbf{m})$, and the covariance matrix between the genotypic values and the QTL additive effects linked to molecular markers, $\mathbf{G}_M$; k is the standardized selection differential; $\sigma^2_{Z_M} = \boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M$ is the variance of the genome wide breeding value, $Z_M = \boldsymbol{\theta}'\mathbf{g} + \boldsymbol{\theta}'_m\mathbf{m}$; $\sigma^2_{Y_M} = \boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M$ is the variance of the genome wide selection index Y_M; $\sigma_{Y_M Z_M} = \boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\beta}_M$ and $\rho_{Y_M Z_M}$ are the covariance and the correlation between Y_M and Z_M, respectively; $\boldsymbol{\theta}'_M = [\boldsymbol{\theta}' \quad \boldsymbol{\theta}'_m]$ is the genome wide vector of the coefficients of $\mathbf{g}'_M = [\mathbf{g}' \quad \mathbf{m}']$ (in the GW LW selection index, $\boldsymbol{\theta}_m$ is a vector of zeros) where $\mathbf{g}$ is the vector of genotypic values; and $\mathbf{m}$ is the random vector of MM code values. Since MMs ($\mathbf{m}$) do not have an effect on $\mathbf{p}$ and are not affected by environmental conditions, the covariance matrix

between the genotypic values and the QTL additive effects linked to molecular markers and the covariance matrix between the phenotypic values and the QTL additive effects linked to molecular markers are equal, which explains why matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ only differ for $\mathbf{P}$ and $\mathbf{G}$. When there are more markers than records, matrix $\mathbf{M}$ is singular, as are matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$.

The Lange and Whittaker genome wide selection index

The objective now is to find a mathematical method that will allow maximizing $\rho_{Y_M Z_M}$ such that the expected genetic gain in each selection cycle is the maximum, the mean squared error of prediction is the minimum, the effectiveness of the prediction of the breeding value is maximum, and the relative efficiency of the SI relative to any other SI is maximum. Furthermore, we need a mathematical method that will allow finding the statistical sampling properties of the estimators of Y_M, $\boldsymbol{\beta}_M$, and $\rho_{Y_M Z_M}$, at least within the asymptotic context. In the context of the Lange and Whittaker (2001) SI, the estimator of $\boldsymbol{\beta}_M$, $\hat{\boldsymbol{\beta}}_{LW} = \hat{\boldsymbol{\Gamma}}^{+}\hat{\boldsymbol{\Sigma}}\boldsymbol{\theta}_{LW}$, allows constructing the genome wide molecular marker LW SI estimator $\hat{Y}_{LW} = \hat{\boldsymbol{\beta}}'_{LW}\mathbf{p}_{LW}$, which has maximum correlation with the breeding value $Z_M = \boldsymbol{\theta}'\mathbf{g} + \boldsymbol{\theta}'_m\mathbf{m}$, where $\hat{\boldsymbol{\Gamma}}^{+}$ is the Moore-Penrose generalized inverse estimator. A problem with $\hat{\boldsymbol{\beta}}_{LW}$ is that its statistical sampling properties in asymptotic contexts are difficult to obtain. Sampling properties of the estimator $\hat{\boldsymbol{\beta}}_{LW}$ are important because they make it possible to determine how near or how far $\hat{\boldsymbol{\beta}}_{LW}$ is from the population parameter, $\boldsymbol{\beta}_{LW}$. In addition, this facilitates obtaining information on the estimator of selection index, $\hat{Y}_{LW}$, because $\hat{Y}_{LW}$ is a function of $\hat{\boldsymbol{\beta}}_{LW}$; therefore if, for example, $\hat{\boldsymbol{\beta}}_{LW}$ is unbiased, then $\hat{Y}_{LW}$ will be unbiased, if vector $\mathbf{p}_M$ is considered fixed. The difficulties of determining the statistical sampling properties of $\hat{\boldsymbol{\beta}}_{LW}$ are that: (1) for deriving the expectation of

$\hat{\boldsymbol{\beta}}_{LW}$ [$E(\hat{\boldsymbol{\beta}}_{LW})$], $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Sigma}}$ may not be statistically independent; and (2) deriving the variance of $\hat{\boldsymbol{\beta}}_{LW}$ [$Var(\hat{\boldsymbol{\beta}}_{LW})$] is not simple, even in the unlikely case that $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Sigma}}$ are independent.

The genome wide Molecular Eigen Selection Index Method (MESIM$_{GW}$)

Cerón-Rojas et al. (2008a, b) derived the theory of ESIM and MESIM under two basic restrictions, $\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M = 1$, which simplify the algebraic development of the selection indices. The restrictions are valid because the objective is to maximize $\rho_{Y_M Z_M}$, which is invariant to changes in scale (Kempthorne and Nordskog 1959). Cerón-Rojas et al. (2008a, b) noted that the key point when maximizing $\rho_{Y_M Z_M}$ is that the variances (or standard deviations) of Y_M and Z_M are constants in each selection cycle. Thus, genotypes can be selected using either Y_M or Y_M / σ_{Y_M}. For example, let us define variable $Y^* = \frac{Y_M}{\sigma_{Y_M}} = \frac{1}{\sigma_{Y_M}}\boldsymbol{\beta}'_M\mathbf{p}_M$ and variable $Z^* = \frac{Z_M}{\sigma_{Z_M}} = \frac{1}{\sigma_{Z_M}}\boldsymbol{\theta}'_M\mathbf{g}_M$, where $\mathbf{g}'_M = [\mathbf{g}' \quad \mathbf{m}']$; then the variances of Y^* and Z^* are equal to 1, and the covariance between Y^* and Z^* is equal to the correlation between Y_M and Z_M; from this, the restrictions $\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M = 1$ when maximizing $\rho^2_{Y_M Z_M}$ are valid.

In this study, we derive the theory of MESIM$_{GW}$ by relaxing the two restrictions $\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M = 1$, and assuming that $\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M = \sigma^2_{Y_M}$, $\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M = \sigma^2_{Z_M}$, and $0 < \sigma^2_{Y_M}, \sigma^2_{Z_M} < \infty$. In MESIM$_{GW}$ it is necessary to maximize

$$\Phi = (\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\beta}_M)^2 - \mu(\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M - \sigma^2_{Y_M}) - \omega(\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M - \sigma^2_{Z_M})$$

with respect to $\boldsymbol{\beta}_M$, $\boldsymbol{\theta}_M$ (in MESIM$_{GW}$, $\boldsymbol{\theta}_M$ does not necessarily denote economic weights), μ, and ω, where $\boldsymbol{\beta}_M$ is the vector of MESIM$_{GW}$ coefficients, $\boldsymbol{\theta}_M$ is the vector of Z_M coefficients, and μ and ω are Lagrange multipliers. Appendix B shows that: (1) $(\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\beta}_M)^2 = \sigma^2_{Z_M}\omega = \sigma^2_{Y_M}\mu = \varphi$; therefore, $\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\beta}_M = \sqrt{\varphi}$, and (2) the estimator of

$\boldsymbol{\beta}_M$ ($\hat{\boldsymbol{\beta}}_{MESIM_{GW}}$), which allows constructing the selection index estimator of Y_M ($\hat{Y}_{\text{MESIM}_{\text{GW}}}$) that has maximum correlation with Z_M, is obtained from

$$(\boldsymbol{\Sigma} - \lambda\boldsymbol{\Gamma})\boldsymbol{\beta}_M = \mathbf{0} \qquad (3)$$

or equivalent by $(\mathbf{Q} - \lambda\mathbf{I}^*)\boldsymbol{\beta}_M = \mathbf{0}$, where $\mathbf{Q} = \boldsymbol{\Gamma}^+\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}^+$ denote the Moore-Penrose generalized inverse, $\mathbf{I}^* = \boldsymbol{\Gamma}^+\boldsymbol{\Gamma}$, and $\lambda = \dfrac{\varphi}{\sigma^2_{Z_M}\sigma^2_{Y_M}}$ (see Appendix B), from where $\varphi = \lambda\sigma^2_{Z_M}\sigma^2_{Y_M}$. Note that when $\sigma^2_{Y_M} = 1$ and $\sigma^2_{Z_M} = 1$, then $\varphi = \lambda$, which means that, in MESIM_{GW}, λ is proportional to φ and that the restrictions $\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M = \sigma^2_{Y_M}$ and $\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M = \sigma^2_{Z_M}$ affect only the maximization of $\rho^2_{Y_M Z_M}$ by a proportionality constant $1/\sigma^2_{Z_M}\sigma^2_{Y_M}$.

It should be pointed out that Eq. 3 was obtained under restrictions $\boldsymbol{\beta}'_M\boldsymbol{\Gamma}\boldsymbol{\beta}_M = \sigma^2_{Y_M}$ and $\boldsymbol{\theta}'_M\boldsymbol{\Sigma}\boldsymbol{\theta}_M = \sigma^2_{Z_M}$. It can be shown that Eq. 3 is the same as that obtained under restrictions $\sigma^2_{Y_M} = 1$ and $\sigma^2_{Z_M} = 1$, therefore, under both set of restrictions, the estimator of $\boldsymbol{\beta}_M$ ($\hat{\boldsymbol{\beta}}_{MESIM_{GW}}$), which allows constructing the selection index estimator of Y_M ($\hat{Y}_{\text{MESIM}_{\text{GW}}}$) that has maximum correlation with Z_M, is the same. Thus, the restrictions used by Cerón-Rojas et al. (2008a, b) are also valid for MESIM_{GW}, and the $\boldsymbol{\beta}_M$ that allows constructing $Y_{\text{MESIM}_{\text{GW}}}$ is the first eigenvector of matrix $\mathbf{Q}$. When $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are correlation matrices, $0 \leq \lambda \leq 1$ (Anderson 2003).

Estimating $\boldsymbol{\beta}_{MESIM_{GW}}$ and Y_M

Matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are not full rank and symmetric, but matrix $\mathbf{Q} = \boldsymbol{\Gamma}^+\boldsymbol{\Sigma}$ is square, nonsymmetric, and not full rank of order $q \times q$ (where q is the total number of traits and MMs). The rank of $\mathbf{Q}$ is $\psi = \text{minimum}[\text{rank}(\boldsymbol{\Gamma}^+), \text{rank}(\boldsymbol{\Sigma})]$, that is, $\mathbf{Q}$ has a maximum of ψ eigenvalues that are different from zero. To estimate $\boldsymbol{\beta}_{MESIM_{GW}}$, we use the singular value decomposition theory, which allows us to write $\mathbf{Q}$ as $\mathbf{Q} = \mathbf{LDH}'$, where the columns of

matrix $\mathbf{L}$ ($\mathbf{L}'\mathbf{L} = \mathbf{I}$) are the left singular vector of $\mathbf{Q}$ (or the eigenvector of $\mathbf{QQ}'$), and the columns of matrix $\mathbf{H}$ ($\mathbf{H}'\mathbf{H} = \mathbf{I}$) are the right singular vector of $\mathbf{Q}$ (or the eigenvector of $\mathbf{Q}'\mathbf{Q}$); $\mathbf{D}$ is a diagonal matrix with singular values of $\mathbf{QQ}'$ or $\mathbf{Q}'\mathbf{Q}$ (or the square root of the eigenvalues of $\mathbf{QQ}'$ or $\mathbf{Q}'\mathbf{Q}$, which are the same). The estimators of $\lambda = \lambda_{MESIM_{GW}}$ and $\boldsymbol{\beta}_M = \boldsymbol{\beta}_{MESIM_{GW}}$ are obtained from $\hat{\mathbf{Q}}\hat{\mathbf{Q}}'$, such that $(\hat{\mathbf{Q}}\hat{\mathbf{Q}}' - \hat{\lambda}^2_{MESIM_{GW}}\mathbf{I})\hat{\boldsymbol{\beta}}_{MESIM_{GW}} = \mathbf{0}$; $\hat{\lambda}^2_{MESIM_{GW}}$ and $\hat{\boldsymbol{\beta}}_{MESIM_{GW}}$ are the maximum likelihood estimators of the eigenvector and eigenvalue of $\mathbf{QQ}'$, respectively, and are unbiased. The estimators of $\mathbf{Q}$, $\mathbf{L}$, $\mathbf{H}$, and $\mathbf{D}$ are $\hat{\mathbf{Q}}$, $\hat{\mathbf{L}}$, $\hat{\mathbf{H}}$, and $\hat{\mathbf{D}}$, respectively, so $\hat{\mathbf{Q}} = \hat{\mathbf{L}}\hat{\mathbf{D}}\hat{\mathbf{H}}'$. These results allow estimating $Y_M = Y_{MESIM_{GW}}$ as $\hat{Y}_{MESIM_{GW}} = \hat{\boldsymbol{\beta}}'_{MESIM_{GW}}\mathbf{p}_M$.

Criteria for evaluating the different selection indices

To compare the genetic gains obtained from MESIM_{GW} and LW under the Moore-Penrose generalized inverse with those obtained from the MAS LT SI and MESIM, we used five criteria: (1) criterion 1, the SI mean squared error of the prediction of the breeding value, and the effectiveness of the prediction of the breeding value of each SI; (2) criterion 2, the relative efficiency of the SI when predicting the breeding value; this is given by the ratio of the correlation between the SI and the breeding value of two SIs (e.g., $\hat{\rho}_{Y_{LW}Z} / \hat{\rho}_{Y_{LT}Z}$, where $\hat{\rho}_{Y_{LW}Z}$ and $\hat{\rho}_{Y_{LT}Z}$ denote the correlation of the Lange-Whittaker GW SI and the Lande-Thompson MAS SI values, respectively); (3) criterion 3, the regression of the true genotypic means of the selected individuals (known in the simulated data sets but unknown in real data) on the selection cycles; (4) criterion 4, the effective genetic gain measured as the mean of the individuals selected in the current selection cycle (say, cycle 2) minus the mean of the individuals selected in the previous cycle (say, cycle 1); and (5) criterion 5, the expected genetic gains. The SI mean squared error of prediction has been suggested by Lange and Whittaker (2001) as a criterion to evaluate SIs. The SI Y will better predict the breeding value, Z, when the correlation between Y and Z is large (Anderson 2003). Therefore, it is desired that the SI mean squared error of the breeding

prediction will be minimum and the effectiveness of the prediction of a SI will be greater when the correlation between the SI and the breeding value is larger. The relative efficiency of the SIs and the expected genetic gains have been suggested by Baker (1986) as useful tools to evaluate the SIs. The regression on the selection cycles of the true genotypic mean of the selected individuals has been utilized by Cerón-Rojas et al. (2008a, b) to evaluate the genetic advance for each selection cycle. Criteria 1 and 2 allow estimating the precision of the SI for predicting the breeding value, whereas criteria 3, 4, and 5 allow estimating the average genetic gain per selection cycle.

Criterion 1- Mean squared error of prediction and effectiveness of Y_M for predicting Z_M

In genome wide or MAS SIs, the aim is to predict Z_M using Y_M by maximizing $\rho^2_{Y_M Z_M}$. Suppose that we can approximate Z_M using a multiple of Y_M, say bY_M; if $E(Z_M) = E(Y_M) = 0$, then the mean squared error of this approximation ($Z_M - bY_M$) is $E(Z_M - bY_M)^2 = \sigma^2_{Z_M}(1 - \rho^2_{Y_M Z_M}) + (b\sigma_{Y_M} - \rho_{Y_M Z_M}\sigma_{Z_M})^2$. This equation is minimum when $(b\sigma_{Y_M} - \rho_{Y_M Z_M}\sigma_{Z_M})^2 = 0$, and this occurs when $b = \frac{\sigma_{Z_M}}{\sigma_{Y_M}}\rho_{Y_M Z_M} = \frac{\sigma_{Z_M Y_M}}{\sigma^2_{Y_M}}$, where $\sigma_{Z_M Y_M}$ is the covariance between Y_M and Z_M, and $\sigma^2_{Y_M}$ is the variance of Y_M. Thus, b is essentially the response to selection or the linear regression coefficient of Z_M on Y_M. Anderson (2003) called the term $\sigma^2_{Z_M}(1 - \rho^2_{Y_M Z_M})$ the mean squared error of prediction, and he considers the ratio $\frac{\sigma^2_{Z_M}(1 - \rho^2_{Y_M Z_M})}{\sigma^2_{Z_M}} = 1 - \rho^2_{Y_M Z_M}$ as the effectiveness of Y_M when predicting Z_M. According to Anderson (2003), Y_M will better predict Z_M when $\rho^2_{Y_M Z_M}$ is maximized, and when $\rho^2_{Y_M Z_M}$ is large, Y_M will be more effective for predicting Z_M.

Criterion 2 -- The relative efficiency of the SIs

The relative efficiency between any two SIs is defined as the ratio between the two correlation of the SIs and the breeding value (Baker 1986). For example, suppose that one wants to calculate the relative efficiency of the genome wide LW selection vs LT MAS SI; then the relative efficiency of both SIs can be estimated as $\frac{\hat{\rho}_{Y_{LW}Z_M}}{\hat{\rho}_{Y_{LT}Z}}$, where $\hat{\rho}_{Y_{LW}Z_M}$ denotes the estimator of the correlation defined in Eq. (2), and $\hat{\rho}_{Y_{LT}Z}$ is the estimator of the correlation in the context of the LT MAS SI, $0 < \hat{\rho}_{Y_{LW}Z_M}, \hat{\rho}_{Y_{LT}Z} \leq 1$. A more convenient way of interpreting this ratio is to express it as a percentage, i.e., $\frac{\hat{\rho}_{Y_{LW}Z_M}}{\hat{\rho}_{Y_{LT}Z}} x100$. Therefore, if for example $\frac{\hat{\rho}_{Y_{LW}Z_M}}{\hat{\rho}_{Y_{LT}Z}} x100 = 180$, it can be concluded that the genome wide LW SI is 80% more efficient than the LT MAS SI when predicting Z_M.

Criterion 3 – Regression of the genotypic means of the selected individuals on the selection cycles

Since the simulator engine used in this study (i.e., QU-GENE) provides the true genotypic value for each individual and each trait, the regression of the true genotypic means of the selected individuals on the selection cycle gives the average genetic gains per selection cycle.

Criterion 4 – Effective genetic gain (EFG)

This criterion measures the genetic advance and is defined as the difference between the genotypic mean of a selection cycle minus the genotypic mean of the selected individuals in the previous cycle. For example, the genotypic mean of the selected individuals in cycle 2 minus the genotypic mean of the selected individuals in cycle 1 denotes the genetic gain achieved from cycle 1 to cycle 2. If the trait under selection increases, the effective genetic gain will be positive, whereas if the trait decreases, the

effective genetic gain will usually be negative. To compare criterion 4 with criterion 3, it is necessary to compute the average EFG across all selection cycles.

Criterion 5 – Expected genetic gain (EGG)

In general, in the context of GW SI and MAS SI, the expected genetic gain for an specific vector of traits, can be denoted as $\mathbf{G} = k\frac{\boldsymbol{\Sigma}\boldsymbol{\beta}_M}{\sqrt{\boldsymbol{\beta}_M'\boldsymbol{\Gamma}\boldsymbol{\beta}_M}}$, where $\boldsymbol{\Sigma}$, $\boldsymbol{\beta}_M$, and $\boldsymbol{\Gamma}$ are defined as in Eq. (2), and $\mathbf{G}$ is the vector of average genetic gains of the traits per selection cycle. When using simulated data, the vector $\mathbf{G}$ denotes the average changes of the genotypic values of the traits in each selection cycle as a result of the selection made using a specific selection index. This criterion is related to the genetic gains per selection cycle, and as in EFG (criterion 4), it is necessary to average $\mathbf{G}$ (EGG) across selection cycles.

Materials

Simulated data

We used the genetic and breeding simulation tool QuLine, previously called QuCim (Wang et al. 2003, 2004), to simulate genotypes from a population with the aim of assessing theoretical and practical results from the MESIM_{GW} and LW molecular marker selection indices, and MESIM and LT MAS selection indices applied to the case of three environments. To simulate the genotypic and phenotypic values of individuals in a breeding population, a genetic model (which is called a gene and environment, GE, system in QuLine) needs to be defined first. The information required for defining a GE system includes number of genes (or QTLs), gene effect for each trait, linkage among the genes in

one chromosome, genotypic trait heritability, and trait means. On the other hand, a breeding strategy to generate various breeding populations needs to be defined as well.

By defining the breeding strategy, QuLine translates the complicated breeding process into one the computer can deal with and simulate. QuLine allows several breeding strategies to be defined simultaneously; they are all contained in one input file. The program then makes the same virtual crosses for all defined strategies in the first breeding cycle. A breeding strategy in QuLine is defined as all the crossing, seed propagation, and selection activities in an entire breeding cycle. A breeding cycle begins with crossing and ends with the generation in which the selected advanced lines are returned to the crossing block as new parents. Selection methods that can be simulated in QuLine include mass selection, pedigree system, bulk population system, backcross breeding, top-cross breeding, doubled haploid breeding, marker-assisted selection for one trait, and many combinations and modifications of these. The simulator provides the true genotypic value for each genotype in the population, as well as the phenotypic value of the traits under study.

Generating simulated doubled haploid and F_2 populations for selection

We followed the procedures described by Zhang and Smith (1992), in which the additive effects of the QTLs required for defining a GE system in QuLine were obtained from a generated Normal distribution of gene effects (positive and negative) that contribute to total additive genetic variance. Figs. 1 and 2 show the distribution of the effects of the QTLs on the simulated traits, female flowering time (FFL) (days) and grain yield (GY) (grams per plot) measured in three different environments for a maize population of 10 chromosomes. The heritability of FFL in environments 1, 2, and 3 were 0.385, 0.579, and 0.653, respectively, and those for GY in environments 1, 2, and 3 were 0.260, 0.506, and 0.200, respectively.

Figure 1.

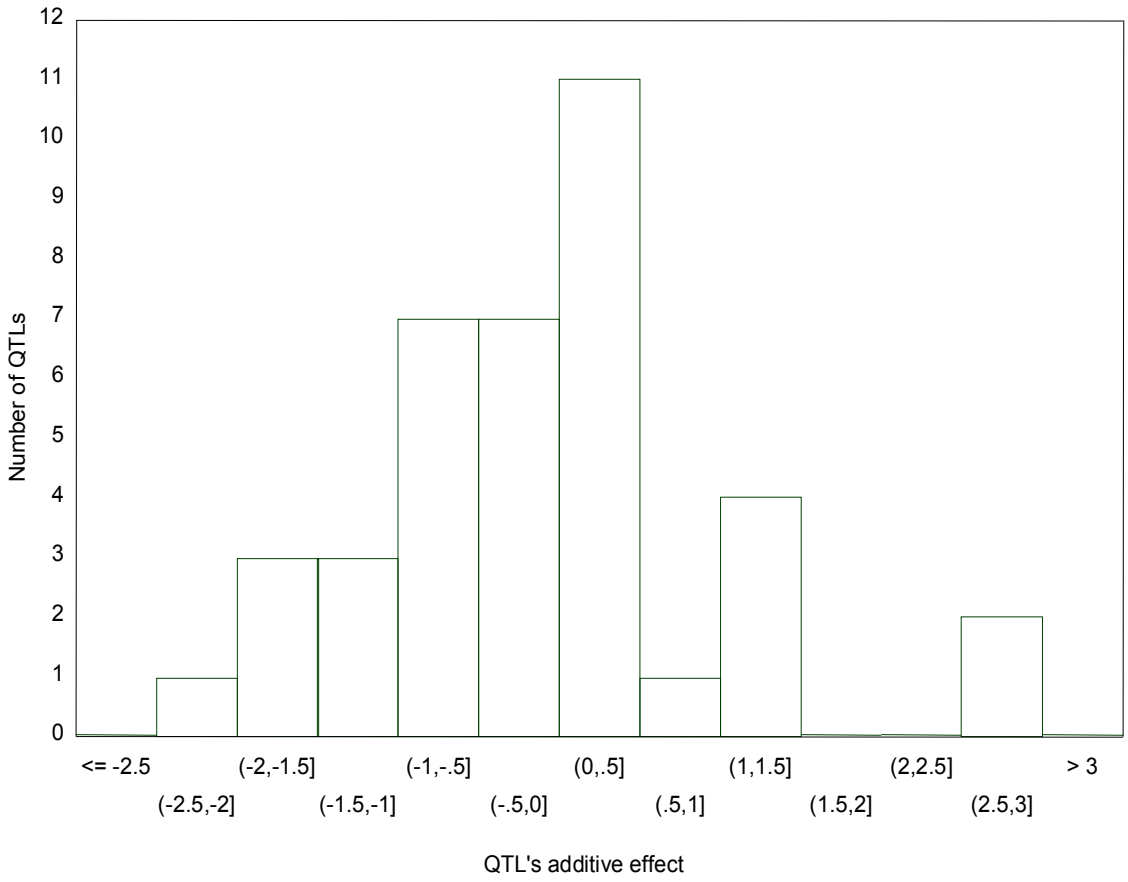

Figure 1. Distribution of the additive effects of 49 simulated QTLs affecting female flowering (days).

Figure 2.

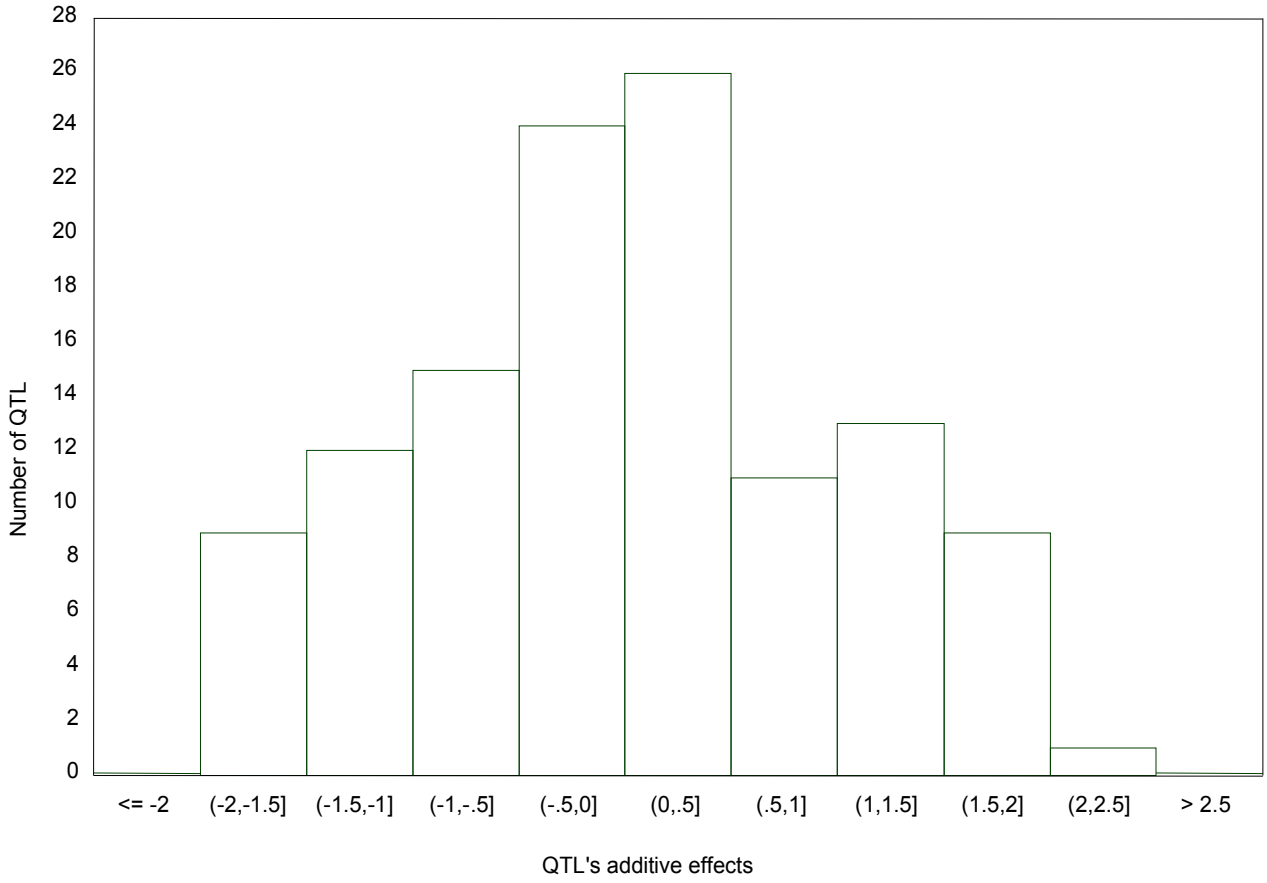

Figure 2. Distribution of the additive effects of 120 simulated QTLs affecting grain yield (grams per plot).

A total of 468 MMs were distributed every 5 cM over the 10 chromosomes. Also, 49 QTLs for FFL were randomly distributed, with a total of 13 QTLs for each environment and 120 QTLs for GY with 40 QTLs in each environment (Table 1). The data were used to generate two doubled haploid (DH) populations of 200 and 500 genotypes, and two F_2 populations of 200 and 500 genotypes that formed the reference population (cycle 0).

Sample sizes of 200 and 500 allowed us to evaluate the efficiency of the GW SI versus MAS, and the efficiency of the Moore-Penrose generalized inverse versus the standard inverse. The two traits were considered simultaneously in the three environments, and 20 and 50 genotypes were selected under $MESIM_{GW}$, LW, MESIM, and LT using 10% (k=1.755) selection pressure. The 20 and 50 selected genotypes from the F_2 and doubled haploid populations were then crossed in diallel fashion, and two new F_2 and doubled haploid populations of 200 and 500 were generated. This was repeated during five selection cycles for the two traits and the two populations in the three environments using all 468 MMs in GW SIs. Each trait-environment combination was considered a variable in the SIs. Therefore, in each selection cycle, the SI and the breeding value comprised six variables (two traits and three environments), plus all the MMs for the GW SIs, and the six variables and MM scores for the MAS SIs.

Table 1. Chromosome, number of molecular markers (MMs) distributed every 5 cM per chromosome, number of QTLs randomly distributed per chromosome, total number of MMs and QTLs (MM +QTL) per chromosome, and chromosome length in cM.

Chromosome	Number of MMs	Number of QTLs	MM+QTL	Chromosome length (cM)
1	48	16	64	235
2	47	17	64	230
3	51	14	65	250
4	48	14	62	235
5	43	19	62	210
6	48	20	68	235
7	47	11	58	230
8	47	18	65	230
9	46	13	59	225
10	43	17	60	210
Total	468	159	627	2290

Sign of the coefficients and economic weights of the selection indices

When using $MESIM_{GW}$ and $MESIM$, it is often necessary to change the sign of the coefficients of the first singular eigenvector in order to select the genotypes according to the desired genetic advance, that is, for trait FFL, the sign is always negative (decreasing the mean genotypic value), whereas for GY, the sign is always positive (increasing the mean genotypic value). However, when $MESIM_{GW}$ is used, the number of QTLs affecting the trait and the molecular markers linked to the QTLs are unknown, so the sign (direction) of the molecular markers on the $MESIM_{GW}$ cannot be modified as in the case of MESIM, where it is possible to modify the direction of the coefficients of molecular scores.

As for the economic weights of the LW and LT selection indices, they were assigned following Smith et al. (1981). Therefore, for FFL and GY, the economic weights were -1 and 1, respectively, in the three environments. The two traits in the three environments, as well as all the molecular markers, were simultaneously considered for the selection indices $MESIM_{GW}$, LW, $MESIM$, and LT.

Real data

A real maize population with 240 F_3 genotypes and 125 MMs were used. This data set gave rise to the simulated data set described above. Selection was based on five traits evaluated in three environments. The traits were male flowering time (MFL) (days), female flowering time (FFL) (days), ear height (EHT) (centimeters), plant height (PH)(centimeters), and grain yield (GY) (grams per plot). The 240 families were planted in the field using an incomplete block design with two replications. The signs of the scores in $MESIM_{GW}$ and $MESIM$, and of the economic weights for LW and LT, were similar to those used for the simulated data, that is, the LW selection index had economic weights of -1, -1, -1, -1, and 1 for MFL, FFL, EHT, PH and GY, respectively, in the three environments. All economic weights of molecular markers for LW and LT were equal to

zero. All trait-environment combinations were simultaneously used under the $MESIM_{GW}$, LW, MESIM, and LT selection indices.

Results

The results of this study are presented in two main sections divided into subsections. The first main section includes results from the simulation data, and the second main section describes results from the real data. The first main section has a subsection describing results from criteria 1, 2, 3, 4 and 5, and the genetic gains of the selected individuals across environments for each selection cycle in graphic form.

Simulated data

Criteria 1 and 2

As previously indicated, two reference populations were initially generated (DH and F_2) each with 200 and 500 genotypes and 468 MMs (Table 1). Using 10 % selection pressure, 20 and 50 genotypes were selected with $MESIM_{GW}$, LW, MESIM, and LT based on female flowering time (FFL) and grain yield (GY) in the three environments. Tables 2 and 3 show the estimated correlations between the SIs and their corresponding breeding values, the mean squared error of prediction and the effectiveness of SIs for predicting the breeding value (criterion 1), and the relative efficiency of the SIs (criterion 2) using 200 DH and F_2 genotypes, respectively. For all selection cycles, the GW SIs had a correlation of 1 with the breeding values and with a mean square prediction error of 0; thus both GW SIs had the same efficiency for predicting the breeding value.

Regarding the MAS SIs, the estimated mean squared error of prediction was smaller for MESIM than for LT, perhaps because for MESIM, the variance of the breeding values was smaller than that for the LT (data not shown). The estimated effectiveness of MESIM and LT for predicting the breeding value was similar in both populations, and the

average relative efficiency for each selection cycle of MESIM with respect to LT was 8 % for DH and -4 % for F_2.

The average relative efficiency of LW and $MESIM_{GW}$ versus the LT SI (criterion 2) was around 80% for the DH and F_2 populations, whereas the average relative efficiency of LW and $MESIM_{GW}$ versus MESIM was 68.18% for the DH population and 87.02% for the F_2 population. These results indicated that including all MMs in the GW SI increases their correlation with the breeding values and, therefore, improves their prediction power as compared to the MAS SIs, where only a few markers closely linked to QTLs are included in the SI. Similar results were obtained when 500 DH and F_2 genotypes were used (see Tables 1c and 2c, Appendix C), where the error mean squared of predictions was always smaller for GW SIs than for MAS SIs. Since the correlation between the selection indices and the breeding values was not perfect (not exactly equal to 1, although very close), the average relative efficiency for each selection cycle of the GW SI versus the MAS SI was smaller than that obtained when 200 individuals were employed (Tables 2 and 3).

Table 2. Correlation between the selection index and the breeding value for each selection index (ρ_{YZ}), mean squared error of prediction (MSE) and effectiveness of SIs for predicting breeding value (criterion 1), and relative efficiency (criterion 2) of the $MESIM_{GW}$ and Lange-Whittaker (LW) selection indices with respect to the Lande-Thompson (LT) SI (%) and MESIM. Traits included in the indices were days to female flowering (FFL) and grain yield (GY) in three environments during five selection cycles using 200 simulated genotypes from a doubled haploid population. (Correlation between the GW SI and their breeding values is 1; therefore, the mean squared error of prediction is 0).

	ρ_{YZ}				Criterion 1				Criterion 2			
Selection cycle	MESIM	LT	$MESIM_{GW}$	LW	MESIM MSE	LT MSE	MESIM effectiveness	LT effectiveness	$\frac{MESIM_{GW}}{MESIM}$	$\frac{LW}{MESIM}$	$\frac{MESIM_{GW}}{LT}$	$\frac{LW}{LT}$
1	0.863	0.853	1.0	1.0	646.5	6796.3	0.255	0.272	115.8	115.8	117.2	117.2
2	0.666	0.605	1.0	1.0	602.1	4621.9	0.556	0.634	150.0	150.0	165.2	165.2
3	0.504	0.543	1.0	1.0	447.9	4265.4	0.746	0.704	198.3	198.3	183.9	183.9
4	0.692	0.448	1.0	1.0	335.3	2952.6	0.520	0.799	144.3	144.3	222.9	222.9
5	0.430	0.519	1.0	1.0	505.4	3587.3	0.815	0.730	232.2	232.2	192.5	192.5
							Average relative efficiency		168.18	168.1	176.3	176.3
							Average relative efficiency in %		68.18	68.1	76.3	76.3

Table 3. Correlation between the selection index and breeding value for each selection index (ρ_{YZ}), mean squared error of prediction (MSE) and effectiveness of SIs for predicting breeding value (criterion 1), and the relative efficiency (criterion 2) of the $MESIM_{GW}$ and Lange-Whittaker (LW) selection indices with respect to the Lande-Thompson (LT) SI (%) and MESIM. Traits included in the indices were days to female flowering (FFL) and grain yield (GY) in three environments during five selection cycles using 200 simulated genotypes from an F_2 population. (Correlation between the GW SI and their breeding values is 1; therefore, the mean squared error of prediction is 0).

	ρ_{YZ}				Criterion 1				Criterion 2			
Selection cycle	MESIM	LT	$MESIM_{GW}$	LW	MESIM MSE	LT MSE	MESIM effectiveness	LT effectiveness	$\frac{MESIM_{GW}}{MESIM}$	$\frac{LW}{MESIM}$	$\frac{MESIM_{GW}}{LT}$	$\frac{LW}{LT}$
1	0.635	0.742	1.0	1.0	902.3	6048.7	0.597	0.450	157.48	157.48	134.84	134.84
2	0.573	0.617	1.0	1.0	647.4	6796.4	0.672	0.619	174.52	174.52	161.98	161.98
3	0.494	0.580	1.0	1.0	355.0	5417.3	0.756	0.663	202.36	202.36	172.37	172.37
4	0.554	0.476	1.0	1.0	520.2	3202.6	0.693	0.774	180.42	180.42	210.17	210.17
5	0.454	0.457	1.0	1.0	252.9	2482.9	0.794	0.791	220.31	220.31	218.77	218.77
							Average relative efficiency		187.02	187.02	179.62	179.62
							Average relative efficiency in %		87.02	87.02	79.62	79.62

Criteria 3, 4, and 5

Tables 4 and 6 show results for GW SIs for 200 genotypes from the DH and F_2 populations, respectively, while Tables 5 and 7 show results for MAS SIs for 200 genotypes from the DH and F_2 populations, respectively. Similarly, Tables 3c and 5c (Appendix C) show results for GW SIs for 500 genotypes from the DH and F_2 populations, respectively, and Tables 4c and 6c (Appendix C) show results for MAS SIs for 500 genotypes from the DH and F_2 populations, respectively.

Table 4. Mean genotypic values of the genotypes selected and the regression coefficient (Reg Coef) (criterion 3) using the $MESIM_{GW}$ and Lange-Whittaker (LW) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 200 simulated genotypes from a doubled haploid population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW , the economic weights are 1 for GY and -1 for FFL.

	$MESIM_{GW}$ Criterion 3						LW Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	113.6	556.4	119.6	604.7	123.6	603.8	113.6	556.4	119.6	604.7	123.6	603.8
1	99.4	682.9	118.6	689.8	121.7	629.0	101.2	665.3	116.7	701.5	116.2	636.6
2	96.9	746.9	118.3	717.3	118.5	648.8	90.5	728.0	115.4	732.0	110.6	651.4
3	96.1	766.8	116.0	776.5	115.2	664.8	86.8	774.9	114.5	768.7	107.4	667.1
4	94.9	795.2	114.5	810.3	109.4	659.1	84.3	798.1	112.2	785.1	102.6	695.1
5	99.7	811.2	113.8	805.0	123.9	698.9	77.0	810.4	113.0	803.8	101.3	703.7
Reg Coef	-2.39	46.59	-1.25	40.63	-1.11	16.62	-6.78	49.01	-1.35	36.66	-4.44	19.73

	$MESIM_{GW}$ Criterion 4 (EFG)						LW Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-14.2	126.5	-1.0	85.0	-1.9	25.2	-12.4	108.9	-2.9	96.8	-7.4	32.9
2	-2.5	64.0	-0.4	27.5	-3.2	19.8	-10.7	62.7	-1.3	30.6	-5.6	14.8
3	-0.8	19.9	-2.3	59.2	-3.4	16.0	-3.7	47.0	-0.9	36.7	-3.2	15.7
4	-1.2	28.4	-1.5	33.9	-5.8	-5.7	-2.5	23.1	-2.3	16.4	-4.8	27.9
5	4.8	16.0	-0.7	-5.4	14.5	39.8	-7.2	12.3	0.8	18.7	-1.4	8.6
Mean EFG	-2.8	50.9	-1.2	40.1	0.1	19.0	-7.3	50.8	-1.3	39.8	-4.5	20.0

	$MESIM_{GW}$ Criterion 5 (EGG)						LW Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-12.3	122.3	-3.0	92.7	-3.9	39.1	-13.2	116.4	-3.3	92.7	-5.7	46.2
2	-3.3	64.7	-1.2	39.4	-3.4	9.6	-9.6	53.9	-2.6	54.5	-5.4	15.6
3	-3.7	30.6	-2.3	60.5	-3.5	20.4	-6.2	39.5	-0.9	30.2	-1.4	21.9
4	-1.9	31.9	-0.6	34.4	-2.2	1.7	-2.1	28.5	-0.7	17.1	-3.8	23.0
5	4.6	9.6	-0.4	0.6	5.9	42.7	-4.6	16.1	-0.4	16.7	-1.4	15.1
Mean EGG	-3.3	51.8	-1.5	45.5	-1.4	22.7	-7.2	50.9	-1.6	42.2	-3.5	24.4

Table 5. Mean genotypic values of the genotypes selected and the regression coefficient (Reg Coef) (criterion 3) using MESIM and Lande-Thompson (LT) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 200 simulated genotypes from a doubled haploid population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW the economic weights are 1 for GY and -1 for FFL.

	MESIM Criterion 3						LT Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	113.6	556.4	119.6	604.7	123.6	603.8	113.6	556.4	119.6	604.7	123.6	603.8
1	99.4	648.0	118.5	677.7	118.6	619.2	100.8	642.0	118.0	677.4	117.6	619.7
2	98.3	674.5	117.6	697.8	111.4	631.4	97.0	682.8	117.4	706.4	113.8	621.8
3	86.0	695.6	115.4	721.7	116.1	641.5	87.0	698.3	116.4	744.0	117.2	622.1
4	84.2	721.8	116.9	764.9	121.5	647.1	87.5	722.4	115.5	773.0	120.3	625.6
5	89.7	725.8	111.5	778.9	121.0	665.0	87.7	739.8	115.3	779.4	113.9	646.3
Reg Coef	-5.07	31.13	-1.36	33.04	0.01	11.42	-5.13	33.53	-0.86	34.23	-1.06	6.59

	MESIM Criterion 4 (EFG)						LT Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-14.2	91.5	-1.2	73.0	-5.0	15.5	-12.8	85.6	-1.6	72.6	-6.0	15.9
2	-1.1	26.5	-0.9	20.0	-7.3	12.2	-3.8	40.8	-0.6	29.0	-3.8	2.1
3	-12.3	21.2	-2.1	23.9	4.8	10.1	-10.0	15.5	-1.0	37.6	3.4	0.4
4	-1.8	26.2	1.5	43.2	5.4	5.5	0.5	24.1	-0.9	28.9	3.1	3.4
5	5.5	4.0	-5.3	14.0	-0.5	18.0	0.3	17.4	-0.1	6.4	-6.4	20.8
Mean EFG	-4.8	33.9	-1.6	34.8	-0.5	12.3	-5.2	36.7	-0.9	34.9	-1.9	8.5

	MESIM Criterion 5 (EGG)						LT Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-12.7	90.2	-4.9	92.1	-4.4	26.0	-7.0	53.7	-2.1	50.9	-3.2	18.0
2	-6.0	53.9	-3.1	54.0	0.6	6.9	-1.9	20.3	-0.9	17.3	-1.7	9.7
3	-6.5	32.6	-2.6	27.9	-0.9	6.4	-2.6	15.8	-0.9	20.0	-1.3	1.7
4	1.6	17.2	-3.7	52.9	-0.2	5.6	-0.9	8.7	-0.1	13.5	0.1	4.2
5	2.0	14.8	-3.3	18.7	-3.3	13.4	-0.8	9.1	0.1	10.7	-1.9	14.0
Mean EGG	-4.3	41.7	-3.5	49.1	-1.6	11.7	-2.7	21.5	-0.8	22.5	-1.6	9.5

Table 6. Mean genotypic values of the genotypes selected and the regression coefficient (Reg Coef) (criterion 3) using $MESIM_{GW}$, and Lange-Whittaker (LW) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 200 simulated genotypes from an F_2 population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW, the economic weights are 1 for GY and -1 for FFL.

	$MESIM_{GW}$ Criterion 3						LW Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	114.0	551.0	120.3	605.8	125.8	602.1	114.0	551.0	120.3	605.8	125.8	602.1
1	105.1	646.0	120.2	659.1	123.4	631.2	105.6	642.1	120.6	662.6	123.7	633.3
2	100.4	715.9	118.5	700.6	122.6	652.1	103.3	690.0	116.8	709.4	123.0	660.8
3	98.1	758.7	117.9	739.3	126.1	664.9	98.1	729.1	116.4	740.6	119.8	689.3
4	91.8	773.2	118.1	797.7	125.1	667.9	94.3	756.1	114.1	772.3	117.0	709.1
5	94.5	799.5	117.0	805.8	117.5	705.2	99.3	780.2	111.9	810.3	112.5	711.8
Reg Coef	-3.99	47.63	-0.67	41.56	-0.94	18.24	-3.22	43.63	-1.77	39.51	-2.57	22.98

	$MESIM_{GW}$ Criterion 4 (EFG)						LW Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-8.9	95.0	-0.1	53.3	-2.4	29.1	-8.4	91.1	0.3	56.8	-2.1	31.2
2	-4.7	69.9	-1.7	41.5	-0.8	20.9	-2.3	47.9	-3.9	46.8	-0.7	27.5
3	-2.3	42.8	-0.6	38.7	3.5	12.8	-5.2	39.1	-0.4	31.2	-3.1	28.5
4	-6.3	14.5	0.2	58.4	-1.0	3.0	-3.8	27.0	-2.4	31.8	-2.8	19.8
5	2.6	26.3	-1.2	8.2	-7.5	37.3	5.0	24.1	-2.1	38.0	-4.6	2.7
Mean EFG	-3.9	49.7	-0.7	40.0	-1.7	20.6	-2.9	45.8	-1.7	40.9	-2.7	21.9

	$MESIM_{GW}$ Criterion 5 (EGG)						LW Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-10.4	97.6	-1.0	61.2	-2.1	27.0	-6.3	52.0	-0.8	36.7	-1.8	18.3
2	-6.0	74.6	0.2	44.6	-2.8	19.3	-5.0	54.2	-1.7	45.5	-1.7	30.1
3	-3.2	43.7	-0.8	45.6	-0.4	14.8	-4.4	38.2	0.9	40.8	-2.0	22.5
4	-4.0	18.9	-1.2	53.6	-0.1	-5.3	-2.0	28.1	-1.8	38.5	-3.2	16.0
5	3.0	25.9	-0.3	11.8	-5.7	37.5	-0.8	24.7	-1.1	32.7	-2.5	3.8
Mean EGG	-4.1	52.1	-0.6	43.4	-2.2	18.7	-3.7	39.5	-0.9	38.8	-2.2	18.1

Table 7. Mean genotypic values of the genotypes selected and the regression coefficient (Reg Coef) (criterion 3) using MESIM, and Lande-Thompson (LT) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 200 simulated genotypes from an F_2 population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW the economic weights are 1 for GY and -1 for FFL.

	MESIM Criterion 3						LT Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	114.0	551.0	120.3	605.8	125.8	602.1	114.0	551.0	120.3	605.8	125.8	602.1
1	108.7	610.3	118.4	646.8	121.9	617.5	108.6	604.6	118.8	654.2	124.6	608.2
2	102.5	643.2	117.8	679.9	128.9	641.0	100.9	641.3	121.8	692.7	131.3	629.6
3	101.1	642.6	119.5	687.9	126.7	655.1	99.1	681.5	119.2	736.2	130.6	642.5
4	98.3	669.9	118.2	708.6	123.9	651.6	94.2	698.6	115.4	746.6	127.5	659.3
5	100.8	699.8	115.6	721.2	119.2	664.0	92.1	718.8	114.1	762.6	127.4	667.2
Reg Coef	-2.82	26.35	-0.64	22.01	-0.83	12.17	-4.41	33.18	-1.25	31.56	0.46	14.05

	MESIM Criterion 4 (EFG)						LT Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-5.4	59.4	-1.9	41.0	-3.9	15.4	-5.4	53.6	-1.5	48.4	-1.2	6.1
2	-6.2	32.9	-0.7	33.1	7.0	23.5	-7.7	36.7	3.0	38.4	6.7	21.3
3	-1.4	-0.5	1.7	8.0	-2.2	14.2	-1.9	40.1	-2.6	43.5	-0.7	12.9
4	-2.8	27.3	-1.3	20.8	-2.8	-3.5	-4.9	17.2	-3.9	10.4	-3.2	16.8
5	2.5	29.9	-2.6	12.6	-4.7	12.4	-2.1	20.2	-1.3	16.0	0.0	7.9
Mean EFG	-2.7	29.8	-0.9	23.1	-1.3	12.4	-4.4	33.6	-1.3	31.4	0.3	13.0

	MESIM Criterion 5 (EGG)						LT Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-8.6	63.4	-1.6	49.6	-2.2	18.8	-5.3	35.8	-0.5	32.1	-1.2	11.0
2	-5.4	37.4	-2.3	28.5	0.2	16.0	-4.2	24.2	0.2	23.7	0.9	13.6
3	-1.1	22.5	1.1	14.8	-1.2	23.6	-2.1	21.7	-1.3	24.1	0.4	3.7
4	0.4	38.0	-1.6	36.1	-3.3	2.4	-2.3	9.4	-1.4	7.4	-2.0	8.1
5	-0.5	7.3	-2.5	10.7	-2.9	18.3	-0.9	8.3	-0.9	9.1	-3.0	3.5
Mean EGG	-3.0	33.7	-1.4	27.9	-1.9	15.8	-3.0	19.9	-0.8	19.3	-1.0	8.0

Each table is divided into three parts; the first part contains the mean of the selected individuals and the regression coefficients (criterion 3), the second part includes the effective genetic gains (EFG) (criterion 4) and the average EFG, and the third part contains the expected genetic gains (EGG or $\mathbf{G}$) (criterion 5) and the average EGG. For each of the six trait-environment combinations (FFL1, FFL2, FFL3, GY1, GY2, GY3), there are three values corresponding to the three evaluation criteria examined, that is, the coefficient of regression, the average EFG, and the average EGG.

Comparing MESIM$_{GW}$ *versus* LW *and* MESIM *versus* LT

In general, the results do not clearly determine which of the two GW SIs or which of the two MAS SIs is superior to the corresponding pair, except in some cases described as follows. There are cases, such as FFL for 200 genotypes from the DH population (Table 4) and FFL for 500 genotypes from the F_2 population (Table 5c, Appendix C), where LW was superior to MESIM$_{GW}$ in the three environments for the three evaluation criteria. Furthermore, for FFL with 500 genotypes from the DH population, LW was superior to MESIM$_{GW}$ seven times of a total of nine possible cases (Table 3c, Appendix C). For GY for 200 genotypes from the F_2 population (Table 6), MESIM$_{GW}$ was better than LW in six out of nine possible cases, and for GY for 500 genotypes from the DH population, LW was better than MESIM$_{GW}$ in six out of nine possible cases (Table 3c, Appendix C).

Similar results were found when comparing MAS SI, MESIM, and LT. For example, for GY for 200 genotypes from the F_2 population (Table 7), LT was better than the MESIM in six out of nine possible cases, and for GY for 500 genotypes from the F_2 population (Table 6c), MESIM was better than LT in seven out of nine possible cases. For the remaining cases, LW does not seem to be clearly superior to MESIM$_{GW}$, and MESIM is not better than LT for the DH and F_2 populations using sample sizes of 200 and 500 genotypes.

Comparing GW SIs *versus* MAS SIs

When comparing GW vs MAS SIs for GY, it is clear that the estimated average genetic gains for each selection cycle of GW SIs are higher than those obtained by MAS SIs in the three environments according to evaluation criteria 3, 4, and 5. As for trait FFL, LW is superior to LT in the three environments for the three evaluation criteria. It is observed for FFL that LW is better

than MESIM in two environments, and when comparing $MESIM_{GW}$ vs MESIM and $MESIM_{GW}$ vs LT, they show similar genetic gains in two environments. In summary, for criteria 3, 4, and 5, GW SIs give higher genetic gains than those obtained by the MAS SI.

Comparing the* DH *population to the* F_2 *population

The average genetic gains measured by criteria 3, 4, and 5 showed that for GY the gains are higher for DH than for F_2 for GW SIs as well as for MAS SIs (Table 8). This comparison was made taking the average coefficient of regression, the average EFG, and the average EGG across the three environments for both traits. For example, if we consider the DH population and the coefficient of regression (Table 4), then the first cell of Table 8 is $\frac{-2.39-1.25-1.1}{3}=-1.58$.

Except for the mean of $MESIM_{GW}$ in the F_2 population for the trait FFL, the remaining averages are higher in the DH population than in the F_2 for both traits, FFL and GY, in the three environments for GW SIs and for MAS SIs. Similar results were obtained for 500 genotypes (data not shown).

Table 8. Average of the mean genetic gains of the regression coefficient (Coef Reg) (criterion 3), effective genetic gains (EFG) (criterion 4), and the effective genetic gains (EGG or **G**) (criterion 5) across three environments for the 200 simulated genotypes for the traits female flowering (FFL) and grain yield (GY) in a doubled haploid population (DH) and an F_2, when employing GW SIs (LW, $MESIM_{GW}$) and MAS SIs (LT, MESIM).

	DH							
	FFL				GY			
Evaluation criteria	$MESIM_{GW}$	LW	MESIM	LT	$MESIM_{GW}$	LW	MESIM	LT
Criterion 3 (Coef Reg)	-1.58	-4.19	-2.14	-2.35	34.62	35.13	25.20	24.78
Criterion 4 (EFG)	-1.30	-4.37	-2.31	-2.66	36.68	36.86	26.99	26.70
Criterion 5 (EGG)	-2.07	-4.10	-3.13	-1.70	40.00	39.17	34.17	17.83
Average	-1.65	-4.22	-2.53	-2.23	37.10	37.05	28.78	23.11
	F_2							
	FFL				GY			
Evaluation criteria	$MESIM_{GW}$	LW	MESIM	LT	$MESIM_{GW}$	LW	MESIM	LT
Criterion 3 (Coef Reg)	-1.87	-2.52	-1.43	-1.74	34.62	35.13	20.18	26.26
Criterion 4 (EFG)	-2.08	-2.43	-1.64	-1.77	36.77	36.22	21.74	25.98
Criterion 5 (EGG)	-2.32	-2.28	-2.09	-1.57	38.06	32.14	25.81	15.72
Average	-2.09	-2.41	-1.72	-1.69	36.48	34.50	22.58	22.65

Genetic gains across environments

Figures 3 and 4 depict the genotypic means for GY and FFL, respectively, evaluated across the three environments for five selection cycles when genotypes were selected using $MESIM_{GW}$, LW, MESIM, and LT selection indices in the DH population using 200 genotypes. The means of the two GW SIs ($MESIM_{GW}$ and LW) increased GY over the genotypic means of the selected individuals using the LT and MESIM SIs (Fig. 3). For FFL (Fig. 4), LW was the best selection index in terms of decreasing the maturity of the selected lines, followed by LT, MESIM and $MESIM_{GW}$. Similarly, Figures 5 and 6 show the genotypic means for GY and FFL evaluated across the three environments simultaneously for five selection cycles when 20 genotypes were selected using $MESIM_{GW}$, LW, MESIM, and LT SIs from an F_2 population using 200 genotypes. The means of $MESIM_{GW}$ and LW surpassed the genotypic mean of the selected genotypes using MESIM and LT, as shown in Fig. 5. For FFL, LW was the best SI in terms of decreasing the maturity of the lines, followed by $MESIM_{GW}$, LT, and MESIM.

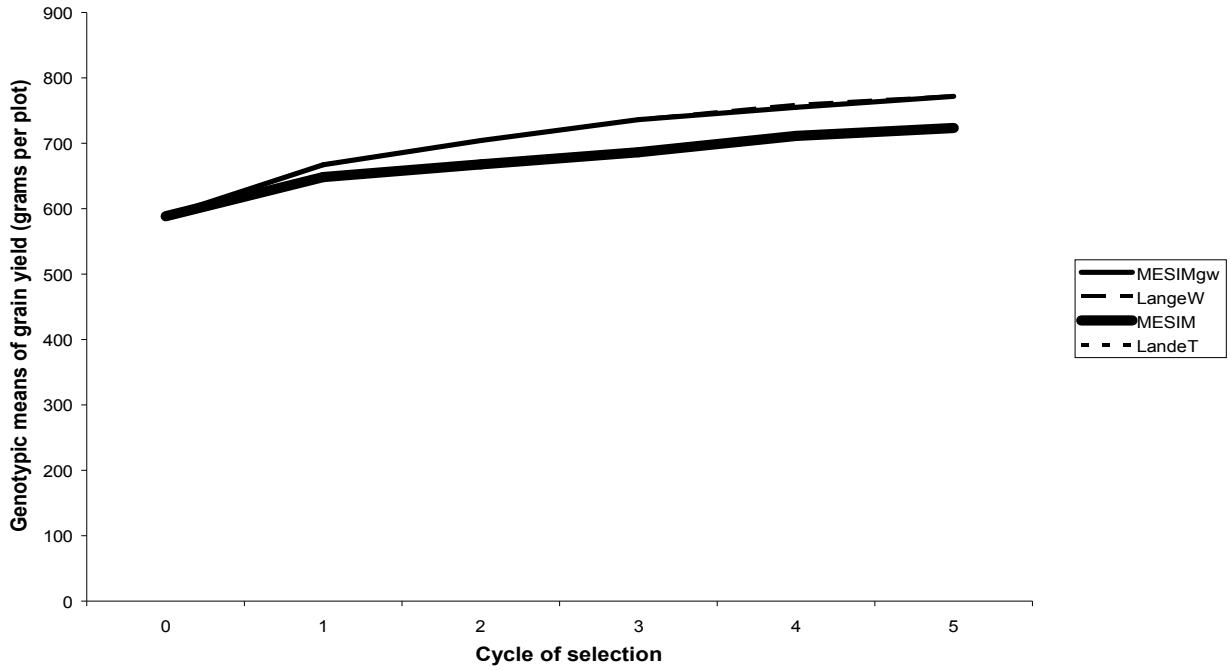

Figure 3. Mean of the genotypic values of grain yield (GY) (grams per plot) across three environments of genotypes selected using $MESIM_{GW}$, Lange-Whittaker (LangeW), MESIM, and Lande-Thompson (LandeT) when female flowering (FFL) and GY are selected in three environments simultaneously during five selection cycles using 200 simulated genotypes from a doubled haploid population. The economic weights used for FFL and GY under the Lange-Whittaker (LangeW) and Lande-Thompson (LandeT) were -1 and 1, respectively.

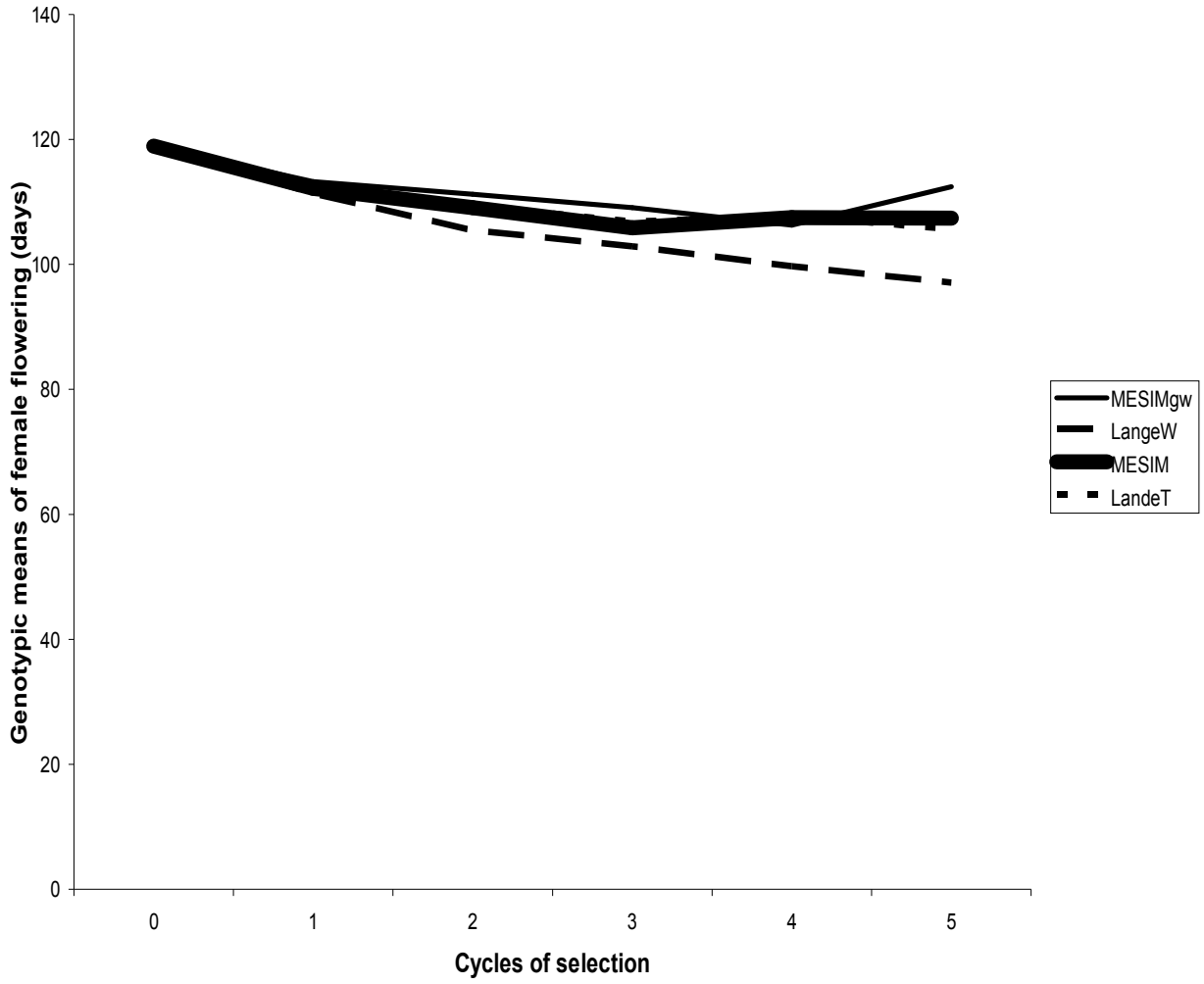

Figure 4. Mean of the genotypic values of female flowering (FFL) (days) across three environments of genotypes selected using $MESIM_{GW}$, Lange-Whittaker (LangeW), MESIM, and Lande-Thompson (LandeT), when FFL and grain yield (GY) are selected in three environments simultaneously during five selection cycles using 200 simulated genotypes from a doubled haploid population. The economic weights used for FFL and GY under the Lange-Whittaker (LangeW) and Lande-Thompson (LandeT) were -1 and 1, respectively.

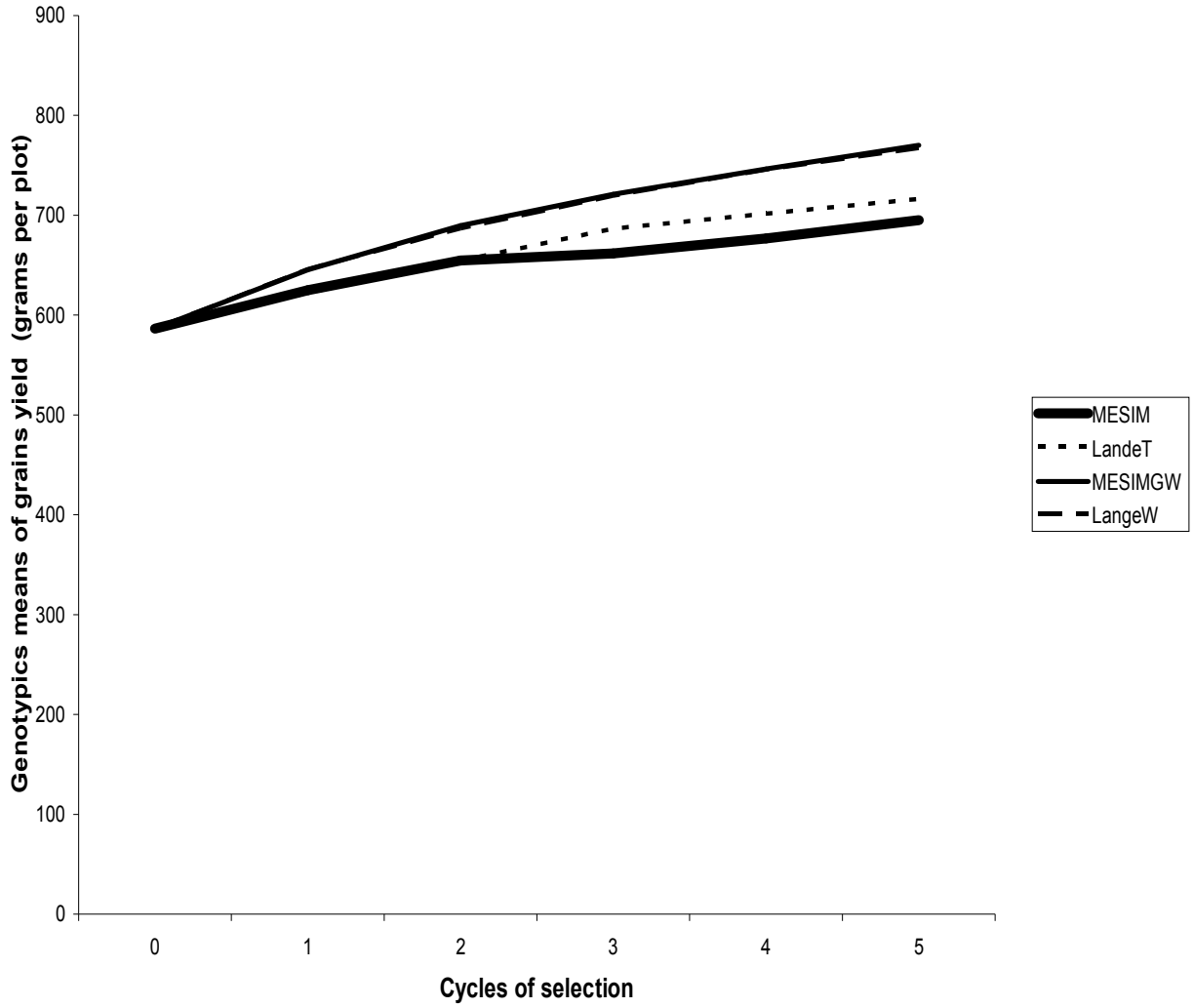

Figure 5. Mean of the genotypic values of grain yield (GY) (grams per plot) across three environments of genotypes selected using $MESIM_{GW}$, Lange-Whittaker (LangeW), MESIM, and Lande-Thompson (LandeT), when female flowering (FFL) and GY are selected in three environments simultaneously during five selection cycles using 200 simulated genotypes from an F_2 population. The economic weights used for FFL and GY under the Lange-Whittaker (LangeW) and Lande-Thompson (LandeT) were -1 and 1, respectively.

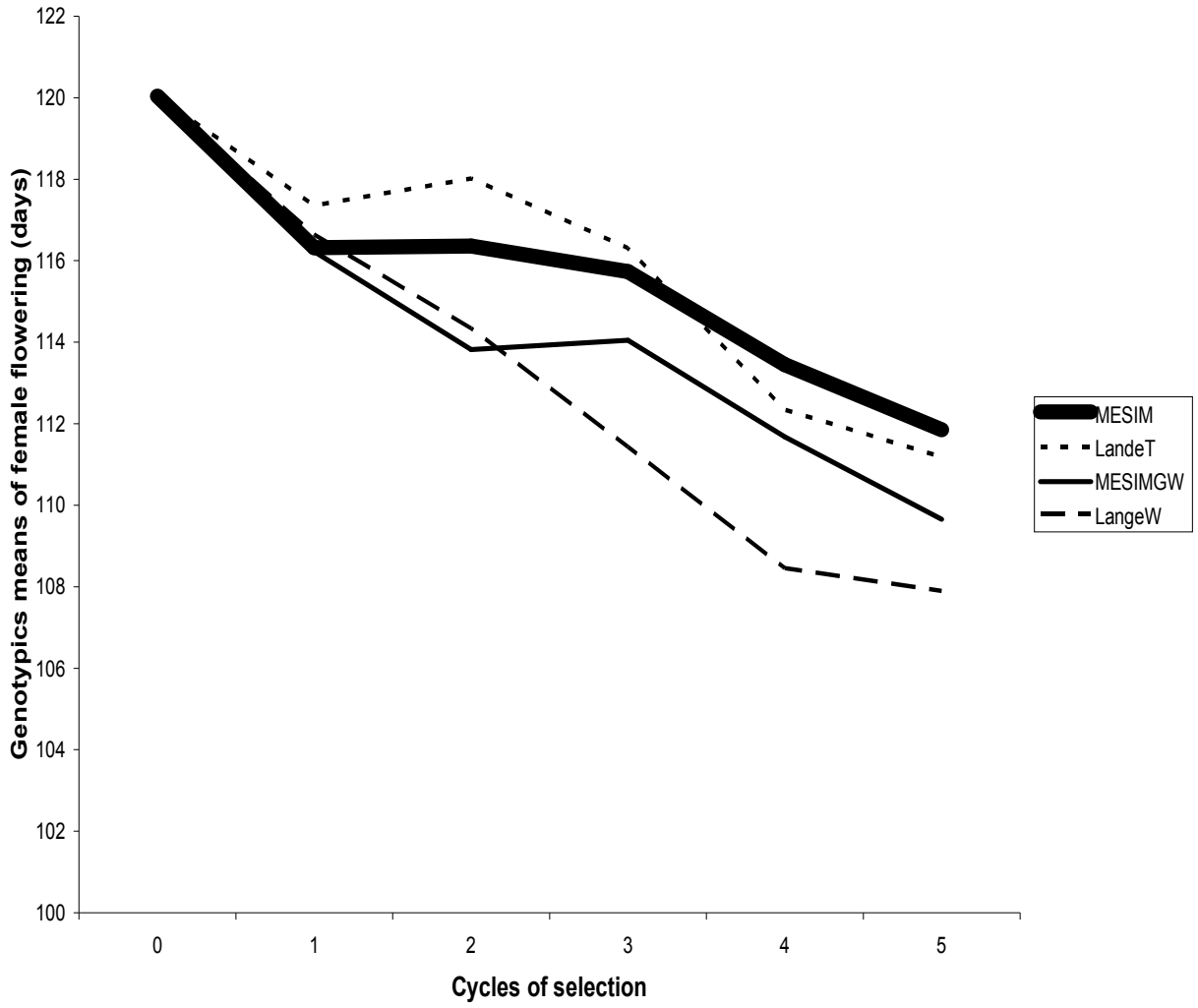

Figure 6. Mean of the genotypic values of female flowering (FFL) (days) across three environments of genotypes selected using $MESIM_{GW}$, Lange-Whittaker (LangeW), MESIM, and Lande-Thompson (LandeT), when FFL and grain yield (GY) are selected in three environments simultaneously during five selection cycles using 200 simulated genotypes from an F_2 population. The economic weights used for FFL and GY under the Lange-Whittaker (LangeW) and Lande-Thompson (LandeT) were -1 and 1, respectively.

Real data

Table 9 shows the mean phenotypic values for one selection cycle of the 24 selected genotypes (10%) obtained using a real maize population with 240 F_3 genotypes and 125 molecular markers for $MESIM_{GW}$, LW, MESIM, and LT when all five traits are selected in three different environments. Note that $MESIM_{GW}$ was more efficient than LW for selecting shorter and earlier maize genotypes with higher grain production in all three environments, except for PHT in environments 1 and 3 (PHT1 and PHT3). These results indicate that when 15 traits were selected simultaneously using all 125 available molecular markers, $MESIM_{GW}$ was better than LW for 13 of the 15 traits.

Results from the real data indicated that, at least for GY, $MESIM_{GW}$ would be more efficient than LW when trait heritability is low (Table 8). For example, heritability of GY in environments 1, 2, and 3 was h^2_{GY1}=0.26, h^2_{GY2}=0.51, and h^2_{GY3}=0.18, respectively, and $MESIM_{GW}$ had a phenotypic mean of selected individuals higher than LW in the three environments. According to these results, the efficiency of the molecular marker selection indices $MESIM_{GW}$ and LW depends on trait heritability. In general, results indicated that MESIM was better than LT, except for GY in environments 1 and 2 (GY1 and GY2).

Table 9. Mean phenotypic values of genotypes selected using $MESIM_{GW}$, Lange-Whittaker (LW), MESIM, and Lande-Thompson (LT) for male flowering (MFL), female flowering (FFL), plant height (PHT), ear height (EHT), and grain yield (GY) in three environments from the real data set of an F_3 maize population using phenotypic, genotypic, and molecular marker variance-covariance matrices for one selection cycle. The economics weights for LT and LW were 1 for GY and -1 for MFL and FFL, EHT and PHT. The heritability of each trait in each environment is shown in parentheses.

	Genotypic means														
	Environment 1					Environment 2					Environment 3				
Selection indices	MFL1 (0.39)	FFL1 (0.38)	EHT1 (0.31)	PHT1 (0.19)	GY1 (0.26)	MFL2 (0.44)	FFL2 (0.58)	EHT2 (0.43)	PHT2 (0.29)	GY2 (0.51)	MFL3 (0.75)	FFL3 (0.65)	EHT3 (0.48)	PHT3 (0.39)	GY3 (0.18)
$MESIM_{GW}$	98.8	98.5	75.0	141.8	166.7	98.1	98.1	72.1	131.6	109.3	102.1	103.4	71.0	116.1	144.0
LW	100.8	100.7	78.1	141.5	164.9	98.8	98.9	76.5	134.4	100.8	102.9	104.7	73.0	113.8	81.8
MESIM	99.1	98.8	73.6	138.7	130.2	97.8	96.2	72.0	134.0	120.4	102.2	102.7	69.3	107.3	112.0
LT	99.5	99.5	76.9	142.1	203.8	98.1	97.9	75.9	136.9	159.7	101.7	102.5	71.3	114.0	98.6
Mean of the original population	101.8	102.0	80.6	140.5	75.5	100.4	100.7	78.9	134.3	58.6	103.3	105.1	74.7	116.2	97.2

Discussion

This research evaluated and compared MESIM (Cerón-Rojas et al. 2008b) and the Lange and Whittaker (2001) genome wide molecular marker selection indices in the context of three environments using the Moore-Penrose generalized inverse vs MAS LT, and MESIM. Results from real data showed that, in general, when several traits were selected in various environments simultaneously, MESIM_{GW} and MESIM increased the phenotypic means over the mean of individuals selected by the LW and LT selection indices.

Results from simulation showed the advantages of MESIM_{GW} and LW over MESIM and LT in terms of average genetic gains for GY in the three environments according to evaluation criteria 3, 4, and 5, and indicated that, in the DH population, the advances were greater than in the F_2 for GW SIs, as well as MAS. Results of evaluation criteria 1 and 2 clearly indicated that incorporating all the MMs into the SI increased the relative efficiency and decreased the mean squared error of prediction of the GW SI as compared to that of the MAS SIs. A surprising result was that GW SIs with 200 and 500 genotypes showed similar average genetic gains, but the correlation of the GW SIs with the breeding value was 1 when using 200 genotypes, with an error mean squared of zero, whereas when using 500 genotypes, this correlation was high but not equal to 1.

An advantage of GW SIs over MAS SIs is that GW SIs are easier to construct. In MAS SIs, it is necessary to combine the additive effects of the QTLs and also the phenotypic information in the SI in order to maximize the selection response. This has to be done in two stages for MAS SI. On the other hand, in GW SIs, the indices are constructed by incorporating the MMs as additional random variables in the SI, which allows predicting genotypic values in only one step. This prevents over-parameterization problems that may occur in MAS SIs (Wong and Bernardo 2008).

One of the most important results of MESIM_{GW} and MESIM is that $\hat{\boldsymbol{\beta}}_{MESIM_{GW}}$ and $\hat{\boldsymbol{\beta}}_{MESIM}$ are the maximum likelihood estimators of $\boldsymbol{\beta}_{MESIM_{GW}}$ and $\boldsymbol{\beta}_{MESIM}$, whereas $\hat{\boldsymbol{\beta}}_{LW}$ and $\hat{\boldsymbol{\beta}}_{LT}$ are estimators of $\boldsymbol{\beta}_{LW}$ and $\boldsymbol{\beta}_{LT}$, whose sampling properties are difficult to evaluate. MESIM_{GW} can be considered a generalization of MESIM (Cerón-Rojas et al. 2008b) when individuals are

selected based on their performance for traits measured in several environments where additional random variables are represented by molecular markers. The sampling properties of $MESIM_{GW}$ and MESIM and their selection responses are known, and their estimators showed desirable statistical properties such as asymptotic unbiasedness.

MESIM maximizes the selection response by maximizing the square of the correlation between the MAS SI and Z. This basic idea, used for developing a molecular selection index based on eigenanalysis (Cerón-Rojas et al. 2008b), is valid for $MESIM_{GW}$ when molecular markers are incorporated as additional traits.

$MESIM_{GW}$ has some advantages over LW. First, it can be used to solve practical problems faced by breeders attempting to select plants or animals for the next generation when no estimates of economic weights are available. Even if economic weights are available, in practice it is very unlikely that they would maximize the derivative of $\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M$ with respect to $\boldsymbol{\beta}_M$ and $\boldsymbol{\theta}_M$. Second, if two breeders are interested in improving, say, n traits, it is very unlikely that they would assign the same weights to them. Third, estimates of $MESIM_{GW}$ have known statistical sampling properties that are easy to evaluate.

On the other hand, both $MESIM_{GW}$ and LW have the main advantage of considering all possible types of cross-products, i.e., marker × marker, trait × trait in one environment, trait × trait in different environments, marker × trait in one environment, and marker × trait in different environments. Therefore, these selection indices implicitly consider the bi-genetic epistatic interaction networks that could potentially affect the expression of complex phenotypic traits of economic importance, such as grain yield in plants or meat and milk production in animals. Furthermore, $MESIM_{GW}$ and LW consider possible epistatic interaction networks in complex *inter loci* interaction under different environmental conditions. Thus, it is expected that variability due to complicated interactions between epistatic networks × environments could be captured under the cross-products considered by $MESIM_{GW}$ and LW.

Since singular value decomposition is a natural mathematical method for dimension reduction, $MESIM_{GW}$ should be useful for direct use in marker-assisted recurrent selection. Although $MESIM_{GW}$ is very straightforward, two difficulties are encountered when this method is used: (1) it is not possible to change the direction of the marker scores of the individuals of the

first eigenvector, and (2) when applying $MESIM_{GW}$ or LW, the user may have difficulty manipulating large variance-covariance matrices with hundreds of thousands of cross-products. However, new computational codes have been developed in SAS for performing all these phenotypic and MM indices which run in very reasonable periods of time (G. Alvarado, personal communication). Finally, it is worth noting that $MESIM_{GW}$ and LW showed a mean squared prediction error equal to zero and a relative efficiency for predicting breeding values that was much higher than that of MAS SI LT and MESIM. In general, the GW SIs showed higher gains per selection cycle, and higher expected and effective genetic gains than the MAS SIs.

These results agree with those obtained by Lange and Whittaker (2001), where GW SI was superior to MAS SI. The actual values of the genotypic means of the selected individuals and the expected genetic gains are generally higher for GY and FFL with the GW SIs than with the MAS SIs. Similar response patterns were obtained when evaluating and comparing LT, MESIM, LW, and $MESIM_{GW}$, when the population size was 500 individuals for the DH and F_2 populations (Tables 3c-6c, Appendix C).

Appendices

Appendix A

The Moore-Penrose's generalized inverse

The Moore-Penrose generalized inverse of a symmetric matrix can be expressed in terms of the components of the spectral decomposition (Schott 2005, Kollo and von-Rosen 2005). In the LW and $MESIM_{GW}$ SIs, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are symmetric matrices of incomplete rank. In this study, when defining the Moore-Penrose generalized inverse, we will use the notation $\boldsymbol{\Gamma}^{+}$ and $\boldsymbol{\Sigma}^{+}$.

Suppose matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are of order q ×q and assume, for simplicity, that both matrices have the same rank, $rank(\boldsymbol{\Gamma}) = r = rank(\boldsymbol{\Sigma})$. Because both $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are symmetric matrices their spectral decomposition gives that $\boldsymbol{\Gamma} = \mathbf{U}\mathbf{D}_{\boldsymbol{\Gamma}}\mathbf{U}'$ and $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{D}_{\boldsymbol{\Sigma}}\mathbf{V}'$, where the columns of matrix $\mathbf{U}$ ($\mathbf{U}\mathbf{U}' = \mathbf{I}_{qxq}$) are the eigenvectors of $\boldsymbol{\Gamma}$, and the columns of matrix $\mathbf{V}$ ($\mathbf{V}\mathbf{V}' = \mathbf{I}_{qxq}$) are the eigenvectors of $\boldsymbol{\Sigma}$; and $\mathbf{D}_{\boldsymbol{\Gamma}}$ and $\mathbf{D}_{\boldsymbol{\Sigma}}$ are diagonal matrices with the eigenvalues of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, respectively (Noble 1973). Suppose that the eigenvalues different from zero of matrix $\mathbf{D}_{\boldsymbol{\Gamma}}$ are $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_r$ and that $\alpha_{r+1} = \alpha_{r+2} = ... = \alpha_q = 0$, then $\mathbf{D}_{\boldsymbol{\Gamma}}$ can be written as

$\mathbf{D}_\Gamma = \begin{bmatrix} \mathbf{D}_\alpha & \mathbf{0}_{rx(q-r)} \\ \mathbf{0}_{(q-r)xr} & \mathbf{0}_{(q-r)x(q-r)} \end{bmatrix}$, where $\mathbf{D}_\alpha = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ 0 & 0 & & 0 \\ 0 & 0 & \dots & \alpha_r \end{bmatrix}$. In matrix $\mathbf{D}_\Gamma$, $\mathbf{D}_\alpha$ is $r \times r$, whereas the matrices $\mathbf{0}'s$ are of different order, as indicated in $\mathbf{D}_\Gamma$. Matrix $\mathbf{D}_\Sigma$ can be written similarly as matrix $\mathbf{D}_\Gamma$.

Note that $\mathbf{D}_\alpha$ is full rank; therefore, its standard inverse ($\mathbf{D}_\alpha^{-1}$) exists. The Moore-Penrose generalized inverse of $\mathbf{D}_\Gamma$ ($\mathbf{D}_\Gamma^+$) is defined as $\mathbf{D}_\Gamma^+ = \begin{bmatrix} \mathbf{D}_\alpha^{-1} & \mathbf{0}_{rx(q-r)} \\ \mathbf{0}_{(q-r)xr} & \mathbf{0}_{(q-r)x(q-r)} \end{bmatrix}$, and the Moore-Penrose generalized inverse of $\mathbf{\Gamma}$ is defined as $\mathbf{\Gamma}^+ = \mathbf{U}\mathbf{D}_\Gamma^+\mathbf{U}'$. Matrix $\mathbf{\Gamma}^+$ has the following properties: (1) $\mathbf{\Gamma}\mathbf{\Gamma}^+\mathbf{\Gamma} = \mathbf{\Gamma}$, (2) $\mathbf{\Gamma}^+\mathbf{\Gamma}\mathbf{\Gamma}^+ = \mathbf{\Gamma}^+$, (3) $(\mathbf{\Gamma}^+\mathbf{\Gamma})' = \mathbf{\Gamma}^+\mathbf{\Gamma}$, (4) $(\mathbf{\Gamma}\mathbf{\Gamma}^+)' = \mathbf{\Gamma}\mathbf{\Gamma}^+$. Properties (1) and (2) give the transitivity character of $\mathbf{\Gamma}^+$, i.e., if $\mathbf{\Gamma}^+$ is the generalized inverse of $\mathbf{\Gamma}$, then $\mathbf{\Gamma}$ is the generalized inverse of $\mathbf{\Gamma}^+$, and the four properties of the Moore-Penrose generalized inverse together give the uniqueness character of $\mathbf{\Gamma}^+$ (Schott, 2005).

Since $\mathbf{D}_\Gamma^+$ also has the above four properties, it is not difficult to show that (1) $\mathbf{\Gamma}\mathbf{\Gamma}^+\mathbf{\Gamma} = \mathbf{\Gamma}$ because $\mathbf{\Gamma}\mathbf{\Gamma}^+\mathbf{\Gamma} = (\mathbf{U}\mathbf{D}_\Gamma\mathbf{U}')(\mathbf{U}\mathbf{D}_\Gamma^+\mathbf{U}')(\mathbf{U}\mathbf{D}_\Gamma\mathbf{U}') = \mathbf{U}\mathbf{D}_\Gamma\mathbf{D}_\Gamma^+\mathbf{D}_\Gamma\mathbf{U}' = \mathbf{U}\mathbf{D}_\Gamma\mathbf{U}' = \mathbf{\Gamma}$. The other three properties of $\mathbf{\Gamma}^+$ can also be shown. To show that $\mathbf{\Gamma}^+$ is unique, note that $\mathbf{\Gamma}$ and $\mathbf{\Gamma}^+$ are symmetrics. Suppose that $\mathbf{C}$ is another symmetric generalized inverse of $\mathbf{\Gamma}$ that has the same four properties as a Moore-Penrose generalized inverse, such that $\mathbf{\Gamma}^+ \neq \mathbf{C}$. Then, according to Kollo and von-Rosen (2005), $\mathbf{\Gamma}^+ = \mathbf{\Gamma}^+\mathbf{\Gamma}\mathbf{\Gamma}^+ = \mathbf{\Gamma}^+(\mathbf{\Gamma}\mathbf{\Gamma}^+)' = \mathbf{\Gamma}^+(\mathbf{\Gamma}^+)'\mathbf{\Gamma}' = \mathbf{\Gamma}^+(\mathbf{\Gamma}^+)'(\mathbf{\Gamma}\mathbf{C}\mathbf{\Gamma})'$ $= \mathbf{\Gamma}^+(\mathbf{\Gamma}^+)'\mathbf{\Gamma}'(\mathbf{\Gamma}\mathbf{C})' = \mathbf{\Gamma}^+\mathbf{\Gamma}\mathbf{\Gamma}^+\mathbf{\Gamma}\mathbf{C} = \mathbf{\Gamma}^+\mathbf{\Gamma}\mathbf{C} = (\mathbf{\Gamma}^+\mathbf{\Gamma})'\mathbf{C}\mathbf{\Gamma}\ \mathbf{C} = \mathbf{\Gamma}'(\mathbf{\Gamma}^+)'(\mathbf{C}\mathbf{\Gamma})'\mathbf{C} = \mathbf{\Gamma}'\mathbf{C}'\mathbf{C} = (\mathbf{C}\mathbf{\Gamma})'\mathbf{C} = \mathbf{C}$, which contradicts $\mathbf{\Gamma}^+ \neq \mathbf{C}$; thus $\mathbf{\Gamma}^+ = \mathbf{C}$ is unique. Finally, it is clear that the Moore-Penrose generalized inverse of $\mathbf{\Sigma}$ is $\mathbf{\Sigma}^+ = \mathbf{V}\mathbf{D}_\Sigma^+\mathbf{V}'$.

Appendix B

Theoretical derivation of MESIM_{GW}

Cerón-Rojas et al. (2008a, b) derived the theory of ESIM and MESIM under the restrictions $\boldsymbol{\beta}'_M \boldsymbol{\Gamma} \boldsymbol{\beta}_M = 1$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\theta}_M = 1$. Here we relax these two restrictions and assume that $\boldsymbol{\beta}'_M \boldsymbol{\Gamma} \boldsymbol{\beta}_M = \sigma^2_{Y_M}$, $\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\theta}_M = \sigma^2_{Z_M}$, and $0 < \sigma^2_{Y_M}, \sigma^2_{Z_M} < \infty$. Then, for MESIM_{GW}, it is necessary to maximize

$$\Phi = \left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M\right)^2 - \mu\left(\boldsymbol{\beta}'_M \boldsymbol{\Gamma} \boldsymbol{\beta}_M - \sigma^2_{Y_M}\right) - \omega\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\theta}_M - \sigma^2_{Z_M}\right)$$

with respect to $\boldsymbol{\beta}_M$, $\boldsymbol{\theta}_M$, μ, and ω, where $\boldsymbol{\beta}_M$ is the vector of MESIM_{GW} coefficients, $\boldsymbol{\theta}_M$ is the vector of Z_M coefficients (in MESIM_{GW}, $\boldsymbol{\theta}_M$ does not necessary denote economic weights), and μ and ω are Lagrange multipliers. When Φ is derived with respect to $\boldsymbol{\beta}_M$, $\boldsymbol{\theta}_M$, μ, and ω, and the result is set to the null vector, it follows that

$$\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M\right) \boldsymbol{\Sigma} \boldsymbol{\theta}_M - \mu \boldsymbol{\Gamma} \boldsymbol{\beta}_M = \mathbf{0} \qquad \text{(B.1)}$$

$$\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M\right) \boldsymbol{\Sigma} \boldsymbol{\beta}_M - \omega \boldsymbol{\Sigma} \boldsymbol{\theta}_M = \mathbf{0} \qquad \text{(B.2)}$$

$$\boldsymbol{\beta}'_M \boldsymbol{\Gamma} \boldsymbol{\beta}_M = \sigma^2_{Y_M} \qquad \text{(B.3)}$$

$$\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\theta}_M = \sigma^2_{Z_M} \qquad \text{(B.4)}$$

where Eqs. B.3 and B.4 denote the restrictions imposed for the maximization of $\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M\right)^2$ with respect to $\boldsymbol{\beta}_M$ and $\boldsymbol{\theta}_M$. Because $\boldsymbol{\beta}'_M \boldsymbol{\Gamma} \boldsymbol{\beta}_M = \sigma^2_{Y_M}$ and $\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\theta}_M = \sigma^2_{Z_M}$, when multiplying Eq. (B.1) by $\boldsymbol{\beta}'_M$ and when multiplying Eq. (B.2) by $\boldsymbol{\theta}'_M$, we get

$$\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M\right)^2 - \mu \sigma^2_{Y_M} = \mathbf{0}$$

$$\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M\right)^2 - \omega \sigma^2_{Z_M} = \mathbf{0}$$

from where $\left(\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M\right)^2 = \sigma^2_{Z_M} \omega = \sigma^2_{Y_M} \mu = \varphi$; therefore, $\boldsymbol{\theta}'_M \boldsymbol{\Sigma} \boldsymbol{\beta}_M = \sqrt{\varphi}$. Because $\sigma^2_{Z_M} \omega = \varphi$ and $\sigma^2_{Y_M} \mu = \varphi$, then $\omega = \dfrac{\varphi}{\sigma^2_{Z_M}}$ and $\mu = \dfrac{\varphi}{\sigma^2_{Y_M}}$; therefore, Eqs. (B.1 and B.2) can be written as

$$\boldsymbol{\Sigma} \boldsymbol{\theta}_M - a \boldsymbol{\Gamma} \boldsymbol{\beta}_M = \mathbf{0} \qquad \text{(B.5)}$$

$$\boldsymbol{\Sigma} \boldsymbol{\beta}_M - c \boldsymbol{\Sigma} \boldsymbol{\theta}_M = \mathbf{0} \qquad \text{(B.6)}$$

where $a = \frac{\varphi}{\sigma^2_{Y_M}}$ and $c = \frac{\phi}{\sigma^2_{Z_M}}$. Eq. (B.5) can be written as

$$\boldsymbol{\Sigma\theta}_M = a\boldsymbol{\Gamma\beta}_M \tag{B.7}$$

Matrix $\boldsymbol{\Sigma}$ does not have a standard inverse, but because of property (1) of the Moore-Penrose generalized inverse (Appendix A), Eq. (B.7) can be pre multiplied by $\boldsymbol{\Sigma\Sigma}^+$ such that $\boldsymbol{\Sigma\theta}_M = a\boldsymbol{\Sigma\Sigma}^+\boldsymbol{\Gamma\beta}_M$, from where $\boldsymbol{\Sigma}(\boldsymbol{\theta}_M - a\boldsymbol{\Sigma}^+\boldsymbol{\Gamma\beta}_M) = \mathbf{0}$. Because $\boldsymbol{\Sigma} \neq \mathbf{0}$, then $(\boldsymbol{\theta}_M - a\boldsymbol{\Sigma}^+\boldsymbol{\Gamma\beta}_M) = \mathbf{0}$ and thus $\boldsymbol{\theta}_M = a\boldsymbol{\Sigma}^+\boldsymbol{\Gamma\beta}_M$. When substituting $\boldsymbol{\theta}_M$ in Eq. B.6, we get

$\boldsymbol{\Sigma\beta}_M - ca\boldsymbol{\Sigma\Sigma}^+\boldsymbol{\Gamma\beta}_M = \mathbf{0}$, or equivalently

$$\boldsymbol{\Sigma\beta}_M = \lambda\boldsymbol{\Sigma\Sigma}^+\boldsymbol{\Gamma\beta}_M \tag{B.8}$$

where $\lambda = ca = \frac{\varphi}{\sigma^2_{Z_M}\sigma^2_{Y_M}} = \rho^2_{Y_M Z_M}$. Because of property (2) of the Moore-Penrose generalized inverse, pre multiplying Eq. B.8 by $\boldsymbol{\Sigma}^+$ gives $\boldsymbol{\Sigma}^+\boldsymbol{\Sigma\beta}_M = \lambda\boldsymbol{\Sigma}^+\boldsymbol{\Gamma\beta}_M$, from where $\boldsymbol{\Sigma}^+(\boldsymbol{\Sigma} - \lambda\boldsymbol{\Gamma})\boldsymbol{\beta}_M = \mathbf{0}$. Because $\boldsymbol{\Sigma}^+ \neq \mathbf{0}$, then $(\boldsymbol{\Sigma} - \lambda\boldsymbol{\Gamma})\boldsymbol{\beta}_M = \mathbf{0}$, or equivalently

$$\boldsymbol{\Sigma\beta}_M = \lambda\boldsymbol{\Gamma\beta}_M \tag{B.9}$$

Pre multiplying Eq. (B.9) by $\boldsymbol{\Gamma}^+$ gives $\boldsymbol{\Gamma}^+\boldsymbol{\Sigma\beta}_M = \lambda\boldsymbol{\Gamma}^+\boldsymbol{\Gamma\beta}_M$, from where

$$(\boldsymbol{\Gamma}^+\boldsymbol{\Sigma} - \lambda\boldsymbol{\Gamma}^+\boldsymbol{\Gamma})\boldsymbol{\beta}_M = \mathbf{0} \tag{B.10}$$

Because $\boldsymbol{\Gamma}^+\boldsymbol{\Gamma} = \mathbf{U}\mathbf{D}^+_{\boldsymbol{\Gamma}}\mathbf{U}'\mathbf{U}\mathbf{D}_{\boldsymbol{\Gamma}}\mathbf{U}' = \mathbf{U}\mathbf{D}^+_{\boldsymbol{\Gamma}}\mathbf{I}_{qxq}\mathbf{D}_{\boldsymbol{\Gamma}}\mathbf{U}' = \mathbf{U}\mathbf{D}^+_{\boldsymbol{\Gamma}}\mathbf{D}_{\boldsymbol{\Gamma}}\mathbf{U}' = \mathbf{U}\mathbf{I}^*\mathbf{U}' = \mathbf{I}^*\mathbf{U}\mathbf{U}' = \mathbf{I}^+\mathbf{I}_{qxq} = \mathbf{I}^*$

where $\mathbf{I}^* = \begin{bmatrix} \mathbf{I}_{rxr} & \mathbf{0}_{rx(q-r)} \\ \mathbf{0}_{(q-r)xr} & \mathbf{0}_{(q-r)x(q-r)} \end{bmatrix}$, and $\mathbf{I}_{rxr} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}$

then Eq. (B.10) is $(\mathbf{Q} - \lambda\mathbf{I}^*)\boldsymbol{\beta}_M = \mathbf{0}$, where $\mathbf{Q} = \boldsymbol{\Gamma}^+\boldsymbol{\Sigma}$, and λ and $\boldsymbol{\beta}_M$ are the eigenvalue and eigenvector of $\mathbf{Q}$, respectively, and $\boldsymbol{\Gamma}^+$ is the Moore-Penrose generalized inverse of $\boldsymbol{\Gamma}$. Thus, for MESIM_{GW}, the eigenvector of matrix $\mathbf{Q}$, $\boldsymbol{\beta}_M$, allows constructing the index $Y_{\text{MESIM}_{\text{GW}}}$ that has maximum correlation with Z_M.

Appendix C

Results using a population size of 500 individuals

Table 1c. Correlation between the selection index and the breeding value for each selection index (ρ_{YZ}), mean squared error of prediction (MSE) and effectiveness of SIs for predicting breeding value (criterion 1), and the relative efficiency (criterion 2) of the $MESIM_{GW}$ and Lange-Whittaker (LW) selection indices with respect to the Lande-Thompson LT SI (%) and MESIM. Traits included in the indices were days to female flowering (FFL) and grain yield (GY) in three environments simultaneously during five selection cycles using 500 simulated genotypes from a doubled haploid population. (Correlations between the GW SI and their breeding values are close to 1; therefore, the mean squared error of prediction is small, and the effectiveness is greater than that of the MAS SI.)

	ρ_{YZ}				Criterion 1				Criterion 2			
Selection cycles	MESIM	LT	$MESIM_{GW}$	LW	MESIM MSE	LT MSE	MESIM effectiveness	LT effectiveness	$\frac{MESIM_{GW}}{MESIM}$	$\frac{LW}{MESIM}$	$\frac{MESIM_{GW}}{LT}$	$\frac{LW}{LT}$
1	0.863	0.833	0.999	0.999	560.9	7269.0	0.256	0.306	115.8	115.8	119.9	119.9
2	0.567	0.677	0.999	0.998	776.3	6678.8	0.678	0.542	176.0	175.9	147.5	147.4
3	0.684	0.638	0.999	0.999	386.3	5832.4	0.532	0.593	146.0	145.9	156.5	156.5
4	0.638	0.516	0.993	0.996	177.5	3732.3	0.594	0.734	155.8	156.3	192.6	193.2
5	0.515	0.432	0.990	0.990	501.6	1781.3	0.735	0.813	115.8	115.8	119.9	119.9
							Average relative efficiency		157.2	157.2	169.1	169.2
							Average relative efficiency in %		57.2	57.2	69.1	69.2

Table 2c. Correlation between the selection index and the breeding value for each selection index (ρ_{YZ}), mean squared error of prediction (MSE) and effectiveness of SIs for predicting breeding value (criterion 1), and the relative efficiency (criterion 2) of the $MESIM_{GW}$ and Lange-Whittaker (LW) selection indices with respect to the Lande-Thompson LT SI (%) and MESIM. Traits included in the indices were days to female flowering (FFL) and grain yield (GY) in three environments simultaneously during five selection cycles using 500 simulated genotypes from an F_2 population. (Correlations between the GW SI and their breeding values are close to 1; therefore, the mean squared error of prediction is small, and the effectiveness is greater than that of the MAS SI.)

	ρ_{YZ}				Criterion 1				Criterion 2			
Selection cycles	MESIM	LT	$MESIM_{GW}$	LW	MESIM MSE	LT MSE	MESIM effectiveness	LT effectiveness	$\frac{MESIM_{GW}}{MESIM}$	$\frac{LW}{MESIM}$	$\frac{MESIM_{GW}}{LT}$	$\frac{LW}{LT}$
1	0.593	0.693	0.998	0.998	746.7	6377.4	0.649	0.520	168.4	168.4	144.0	144.0
2	0.545	0.651	0.997	0.995	699.8	5633.3	0.703	0.577	182.9	182.6	153.1	152.9
3	0.635	0.548	0.991	0.995	251.6	4005.0	0.597	0.699	156.1	156.6	180.7	181.3
4	0.649	0.507	0.994	0.990	334.4	3306.9	0.579	0.743	153.1	152.5	196.0	195.2
5	0.474	0.513	0.987	0.994	534.0	3610.9	0.775	0.736	208.3	209.7	192.3	193.6
							Average relative effeciency		174.0	173.2	173.4	174.0
							Average relative efficiency in %		74.0	73.2	73.4	74.0

Table 3c. Mean genotypic values of the genotypes selected and the regression coefficient (Reg. Coef.) (criterion 3) using $MESIM_{GW}$ and Lange-Whittaker (LW) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 500 simulated genotypes from a doubled haploid population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW , the economic weights are 1 for GY and -1 for FFL.

	$MESIM_{GW}$ Criterion 3						LW Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	114.2	552.4	120.2	601.0	124.2	602.9	114.2	552.4	120.2	601.0	124.2	602.9
1	102.1	678.2	118.8	684.1	121.7	628.9	99.3	660.2	117.9	699.0	119.0	638.7
2	92.2	738.5	117.1	765.2	121.6	653.4	92.2	725.5	119.2	751.7	121.7	670.2
3	87.1	794.7	115.5	756.4	119.4	696.9	85.5	761.2	117.2	789.9	118.8	696.0
4	88.8	808.0	114.2	795.8	116.8	695.4	84.5	796.7	116.0	813.8	115.7	714.2
5	84.8	816.9	113.2	823.1	113.8	695.8	82.2	812.6	115.1	827.4	110.6	723.6
Reg Coef	-5.49	50.52	-1.44	41.05	-1.97	20.21	-6.03	49.89	-0.95	43.27	-2.31	24.45

	$MESIM_{GW}$ Criterion 4 (EFG)						LW Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-12.1	125.8	-1.4	83.1	-2.5	26.0	-14.9	107.8	-2.2	98.0	-5.2	35.7
2	-9.8	60.3	-1.7	81.1	-0.1	24.5	-7.1	65.3	1.3	52.7	2.7	31.5
3	-5.1	56.2	-1.6	-8.8	-2.2	43.5	-6.7	35.7	-1.9	38.2	-2.9	25.8
4	1.7	13.3	-1.3	39.4	-2.6	-1.5	-1.1	35.5	-1.2	23.9	-3.0	18.2
5	-4.0	8.9	-0.9	27.3	-3.0	0.4	-2.3	15.9	-0.9	13.6	-5.1	9.4
Mean EFG	-5.9	52.9	-1.4	44.4	-2.1	18.6	-6.4	52.0	-1.0	45.3	-2.7	24.1

	$MESIM_{GW}$ Criterion 5 (EGG)						LW Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-14.9	125.9	-1.5	80.3	-2.9	36.4	-16.1	114.0	-2.3	89.5	-4.0	44.5
2	-9.5	51.1	-2.0	83.8	-0.5	15.5	-8.5	68.9	1.6	46.5	2.1	30.8
3	-3.1	58.4	-0.4	0.8	0.7	38.6	-6.2	43.7	-1.8	39.0	-2.1	24.2
4	1.0	11.2	-0.7	39.5	-1.3	5.6	-1.2	35.0	-2.4	21.5	-3.7	16.8
5	-4.4	9.9	0.0	27.6	-1.5	6.8	-3.0	17.8	-0.6	14.1	-3.2	12.8
Mean EGG	-6.2	51.3	-0.9	46.4	-1.1	20.6	-7.0	55.9	-1.1	42.1	-2.2	25.8

Table 4c. Mean genotypic values of the genotypes selected and the regression coefficient (Reg. Coef.) (criterion 3) using MESIM, and Lande-Thompson (LT) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 500 simulated genotypes from a doubled haploid population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW , the economic weights are 1 for GY and -1 for FFL.

	MESIM Criterion 3						LT Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	114.17	552.42	120.16	600.99	124.20	602.92	114.17	552.42	120.16	600.99	124.20	602.92
1	101.74	632.63	116.81	690.03	122.56	621.21	100.06	643.40	118.12	687.50	121.46	629.68
2	101.25	677.08	113.69	689.10	115.44	650.93	91.69	699.96	115.99	725.42	117.91	658.80
3	98.09	686.90	112.13	743.63	120.76	654.70	85.74	744.90	112.86	765.92	120.18	677.95
4	93.69	734.68	112.87	758.19	119.61	677.83	84.11	763.21	110.33	774.08	118.26	695.69
5	96.73	767.86	111.78	761.81	111.54	683.91	85.35	767.24	108.35	781.08	115.15	697.44
Reg Coef	-3.27	39.80	-1.58	30.37	-1.91	16.53	-5.65	42.24	-2.44	34.31	-1.50	19.71

	MESIM Criterion 4 (EFG)						LT Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-12.46	80.23	-3.39	89.03	-1.64	18.31	-14.14	91.00	-2.08	86.50	-2.74	26.78
2	-0.48	44.45	-3.12	-0.93	-7.12	29.72	-8.37	56.56	-2.13	37.91	-3.55	29.12
3	-3.17	9.83	-1.56	54.53	5.32	3.77	-5.95	44.94	-3.13	40.50	2.27	19.16
4	-4.40	47.78	0.75	14.55	-1.15	23.13	-1.63	18.31	-2.53	8.16	-1.92	17.73
5	3.04	33.18	-1.10	3.62	-8.08	6.08	1.24	4.03	-1.98	7.00	-3.11	1.75
Mean EFG	-3.49	43.09	-1.68	32.16	-2.53	16.20	-5.77	42.97	-2.37	36.02	-1.81	18.91

	MESIM Criterion 5 (EGG)						LT Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-15.00	80.85	-4.23	92.12	-0.78	21.90	-8.70	50.73	-1.45	49.32	-1.50	16.68
2	-4.60	54.63	-1.91	9.13	-7.52	12.90	-4.78	33.52	-0.01	22.95	-1.52	12.37
3	-4.50	20.05	-2.68	57.17	1.27	4.37	-3.33	26.64	-1.35	21.09	0.85	11.75
4	-1.06	38.69	-2.25	13.41	-5.62	14.44	-0.60	15.00	-1.24	10.18	-0.74	9.00
5	1.65	34.33	-1.77	6.99	-9.70	5.74	1.78	5.92	-1.48	9.55	-1.64	3.41
Mean EGG	-4.7	45.7	-2.6	35.8	-4.5	11.9	-3.1	26.4	-1.1	22.6	-0.9	10.6

Table 5c. Mean genotypic values of the genotypes selected and the regression coefficient (Reg. Coef.) (criterion 3) using $MESIM_{GW}$, and Lange-Whittaker (LW) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 500 simulated genotypes from an F_2 population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW , the economic weights are 1 for GY and -1 for FFL.

	$MESIM_{GW}$ Criterion 3						LW Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	114.6	551.5	120.0	602.1	125.0	602.6	114.6	551.5	120.0	602.1	125.0	602.6
1	105.3	637.2	119.2	664.4	124.1	624.9	104.6	638.6	119.3	661.3	123.8	628.3
2	103.8	712.0	119.5	671.7	122.2	638.5	98.8	691.5	118.5	710.9	120.2	656.3
3	99.6	730.4	117.7	734.2	124.5	649.6	97.8	733.1	116.3	751.7	115.1	670.8
4	97.0	771.8	117.8	751.4	122.4	676.0	95.2	769.5	116.4	775.1	110.7	692.6
5	99.2	756.6	116.0	788.0	120.2	695.5	90.8	797.0	115.4	797.1	107.0	705.9
Reg Coef	-3.03	41.36	-0.74	35.80	-0.77	17.97	-4.23	47.48	-0.97	38.78	-3.84	20.68

	$MESIM_{GW}$ Criterion 4 (EFG)						LW Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-9.3	85.7	-0.8	62.3	-1.0	22.3	-10.0	87.2	-0.7	59.2	-1.3	25.7
2	-1.5	74.8	0.3	7.3	-1.9	13.6	-5.9	52.9	-0.9	49.6	-3.6	28.0
3	-4.1	18.4	-1.8	62.5	2.3	11.1	-1.0	41.6	-2.2	40.8	-5.1	14.5
4	-2.7	41.4	0.1	17.1	-2.0	26.4	-2.6	36.3	0.2	23.4	-4.4	21.8
5	2.3	-15.2	-1.8	36.7	-2.2	19.5	-4.5	27.5	-1.1	22.0	-3.7	13.2
Mean EFG	-3.08	41.03	-0.80	37.18	-0.97	18.58	-4.77	49.10	-0.93	38.99	-3.61	20.65

	$MESIM_{GW}$ Criterion 5 (EGG)						LW Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-10.2	87.1	-1.2	66.6	-0.8	23.9	-10.7	86.0	-1.4	62.1	-2.2	31.7
2	-2.5	81.5	-1.0	15.8	-1.2	13.6	-7.1	58.8	-1.3	45.3	-2.3	25.6
3	-2.8	20.6	-0.9	62.6	0.5	17.9	-4.2	46.0	-1.4	35.9	-2.3	20.7
4	-1.6	44.6	0.1	20.1	-0.5	32.2	-3.4	35.4	-0.3	25.6	-1.6	20.1
5	0.9	-11.8	-1.3	33.1	-0.6	27.4	-3.8	30.2	-1.0	21.0	-1.0	15.5
Mean EGG	-3.3	44.4	-0.9	39.6	-0.5	23.0	-5.8	51.3	-1.1	38.0	-1.9	22.7

Table 6c. Mean genotypic values of the genotypes selected and the regression coefficient (Reg. Coef.) (criterion 3) using MESIM, and Lande-Thompson (LT) selection indices, effective genetic gain (EFG) (criterion 4) and mean EFG, and expected genetic gain (EGG) (criterion 5) and mean EGG for six trait-environment combinations including days to female flowering (FFL) and grain yield (GY) in three environments (FFL1, FFL2, FFL3, GY1, GY2, GY3) selected simultaneously during five selection cycles using 500 simulated genotypes from an F_2 population. The signs and economic weights of the selection indices for each trait are shown in parentheses; for selection index LW , the economic weights are 1 for GY and -1 for FFL.

	MESIM Criterion 3						LT Criterion 3					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	114.62	551.46	120.02	602.14	125.04	602.60	114.62	551.46	120.02	602.14	125.04	602.60
1	110.54	610.86	118.28	639.05	121.51	615.38	109.54	605.50	117.94	652.36	122.81	614.82
2	105.72	656.21	117.25	658.09	120.30	629.99	102.49	647.43	118.22	688.48	121.50	630.18
3	99.99	681.12	114.93	698.58	117.39	649.79	100.57	678.35	116.96	719.31	115.85	635.13
4	96.23	692.71	111.91	735.41	116.91	661.17	97.60	689.45	117.51	735.89	112.05	643.88
5	96.65	719.69	110.52	736.07	118.60	677.06	96.00	715.00	116.68	761.14	110.92	657.89
Reg Coef	-3.96	31.76	-1.97	28.55	-1.40	15.13	-3.74	31.44	-0.55	30.75	-3.10	10.53

	MESIM Criterion 4 (EFG)						LT Criterion 4 (EFG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-4.06	59.36	-1.72	36.95	-3.49	12.78	-5.06	54.00	-2.06	50.26	-2.19	12.22
2	-4.82	45.35	-1.03	19.04	-1.21	14.61	-7.05	41.93	0.28	36.12	-1.31	15.36
3	-5.73	24.91	-2.32	40.49	-2.91	19.80	-1.92	30.92	-1.25	30.83	-5.65	4.95
4	-3.76	11.59	-3.02	36.83	-0.48	11.38	-2.97	11.09	0.54	16.58	-3.79	8.75
5	0.42	26.98	-1.39	0.66	1.69	15.90	-1.60	25.56	-0.83	25.24	-1.13	14.01
Mean EFG	-3.59	33.64	-1.90	26.80	-1.28	14.89	-3.72	32.70	-0.66	31.81	-2.82	11.06

	MESIM Criterion 5 (EGG)						LT Criterion 5 (EGG)					
	Environment 1		Environment 2		Environment 3		Environment 1		Environment 2		Environment 3	
Selection Cycle	FFL1 (-)	GY1 (+)	FFL2 (-)	GY2 (+)	FFL3 (-)	GY3 (+)	FFL1 (-1)	GY1 (+1)	FFL2 (-1)	GY2 (+1)	FFL3 (-1)	GY3 (+1)
0	---	---	---	---	---	---	---	---	---	---	---	---
1	-7.50	57.82	-1.34	40.03	-2.09	15.77	-4.85	31.55	-0.61	29.44	-0.86	9.45
2	-7.21	49.13	-0.47	14.02	-1.86	17.55	-4.05	24.35	-0.10	26.34	-0.57	8.88
3	-6.20	20.78	-2.81	44.27	-3.15	10.35	-3.10	14.23	-0.44	20.68	-1.96	1.09
4	-3.48	11.26	-3.76	39.42	-0.14	1.77	-1.62	12.17	-0.54	13.21	-1.73	4.55
5	-1.89	30.58	-1.57	9.13	0.08	15.19	-1.83	14.87	-0.47	13.66	-1.23	3.89
Mean EGG	-5.3	33.9	-2.0	29.4	-1.4	12.1	-3.1	19.4	-0.4	20.7	-1.3	5.6

References

Albert A (1973) Statistical applications of the pseudo inverses, in: Generalized inversed and applications. Proceedings of an advanced seminar sponsored by the Mathematics Research Center, the University of Wisconsin-Madison. Edited by Nashed MZ, p 525.

Anderson TW (2003) An introduction to multivariate statistical analysis. 3rd ed. John Wiley and Sons, New Jersey, pp 497-500.

Baker RJ (1986) Selection indices in plant breeding. CRC Press Inc., Boca Raton, Florida, pp 87 and 118.

Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Crossa J (2008a) A restricted selection index method based on eigenanalysis. Journal of Agriculture, Biology and Environmental Statistics 13 (4):440-457.

Cerón-Rojas JJ, Sahagún-Castellanos J, Castillo-González F, Santacruz-Varela A, Benítez-Riquelme I, Crossa J (2008b) A molecular selection index method based on eigenanalysis. Genetics 180:547-557.

Kempthorne O, Nordskog AW (1959) Restricted selection indices. Biometrics 15:10-19.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

Kollo T, von-Rosen D (2005) Advanced multivariate statistics with matrices. Springer, The Netherlands, pp 15-18.

Lange C, Whittaker JC (2001) On the prediction of genetic values in marker-assisted selection. Genetics 159:1375-1381.

Mardia KV, Kent JT, Bibby JM (1982) Multivariate analysis. Academic Press Inc, New York, p 287.

Muirhead RJ (2005) Aspects of multivariate statistical theory, John Wiley and Sons, New York, p 549.

Nashed MZ (1973) Generalized inversed and applications. Proceedings of an advanced seminar sponsored by the Mathematics Research Center, the University of Wisconsin-Madison, p xi.

Noble, A (1973) Methods for computing the Moore-Penrose generalized inverse, and related matters, in: Generalized inversed and applications. Proceedings of an advanced seminar

sponsored by the Mathematics Research Center, the University of Wisconsin-Madison. Edited by Nashed MZ, pp 293-294.

Schott JR (2005) Matrix analysis for statistics. 2rd ed. John Wiley and Sons, New Jersey, pp 179-186.

Smith HF (1936) A discriminant function for plant selection in: Papers on quantitative genetics and related topics. Department of Genetics, North Carolina State College, Raleigh, North Carolina, pp. 466-476.

Smith OS, Hallauer AR, Russell WA (1981) Use of index selection in recurrent selection programs in maize. Euphytica 30:611-618.

Wang J, van Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, Delacy IH, Cooper M, Rajaram S (2003) Comparison of two breeding strategies by computer simulation. Crop Sci. 43:1764-1773.

Wang J, van Ginkel M, Trethowan R, Ye G, Delacy I, Podlich D, Cooper M (2004) Simulating the effects of dominance and epistasis on selection response in the CIMMYT wheat breeding program using QuCim. Crop Science 44:2006-2018.

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815-824.

Zhang W, Smith C (1992) Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theor Appl Genet 83:813-820.

DISCUSIÓN GENERAL

Una de las aplicaciones principales de la genética cuantitativa es el mejoramiento genético de los cultivos por medio de la selección de las características de variedades de interés agronómico. Ejemplos típicos son el rendimiento en maíz, la tasa de crecimiento de ganado vacuno, la tasa de producción de leche, el contenido de proteína en soya, etc. En el mejoramiento de plantas y animales el objetivo es crear poblaciones más productivas cuyos productos sean de mayor calidad. De ordinario es posible incrementar la media de una característica agronómica en aproximadamente 1% por año. Los valores medios de las características agronómicas de interés en una población cambian como resultado de la elección de individuos deseables como progenitores de la generación siguiente. Si el carácter tiene una base genética, las frecuencias de los alelos deseables se incrementarán, y se observará un cambio favorable en la media de la población en la característica deseable (Saxton, 2004).

La selección efectiva para el mejoramiento de plantas está basada en la separación eficiente de la variabilidad genética heredable de la que no lo es. En las primeras fases de la domesticación de plantas, ésta estuvo basada en el entendimiento de las leyes de la herencia, basada en la selección de fenotipos del carácter. En tal situación a la selección no se le puede llamar, en sentido estricto, "selección genotípica". La selección asistida por marcadores moleculares (MAS por sus siglas en inglés) ofrece la posibilidad de la selección genotípica indirecta, la cual tiene varias ventajas sobre los métodos de selección tradicionales, principalmente porque no se confunde la variabilidad genética con la no genética (Koebner *et al.*, 2001).

En teoría, se ha demostrado que MAS produce una mayor respuesta a la selección que la alcanzada por la selección fenotípica cuando los caracteres de interés tienen distribución normal

(Knapp, 1998). El volumen de publicaciones sobre la aplicación de los marcadores moleculares (MM) en el mejoramiento de plantas se ha incrementado dramáticamente durante la última década.; sin embargo, llevar a la practica la teoría de las publicaciones promisorias requiere la solución de varias limitantes genéticas y logísticas raramente señaladas en los artículos publicados. Así, es necesario resolver varios problemas para que MAS alcance su potencial completo en los programas de mejoramiento. Entre tales problemas se incluyen el desarrollo de sistemas altamente precisos para el mapeo de QTLs, un mejor entendimiento de la interacción genotipo ambiente y la epistasis, y el desarrollo de herramientas computacionales adecuadas para las necesidades de los programas de mejoramiento molecular (Xu y Crouch, 2008).

Arús y Moreno-González (1993), y Xu y Crouch (2008) enlistan los casos en que MAS puede ser más eficiente que la selección fenotípica convencional:

1. Cuando el carácter no se expresa de manera temprana en la planta, tal como el fruto o las características de las flores o caracteres adultos en especies con un período juvenil;
2. Cuando el gene de interés es recesivo;
3. Caracteres difíciles de manejar a través de la selección fenotípica convencional debido a que son costosos o difíciles de medir o tienen baja heredabilidad;
4. Caracteres cuya selección depende de ambientes específicos o etapas de desarrollo que afectan la expresión de los fenotipos de interés;
5. Cuando existe un requerimiento de operaciones especiales con el fin de que el gene se exprese, como el caso de selección para enfermedades o resistencia a plagas.

En la actualidad se cuenta con varios mapas de QTLs disponibles para los más importantes cultivos de interés agronómico. La identificación y los efectos MQTL han sido establecidos para diversas características de interés económico en plantas y animales y la investigación sobre la

validación del ligamiento existente entre marcadores moleculares (MM) y QTL está incrementándose. Actualmente los investigadores utilizan de manera rutinaria genes de interés agronómico basados en los mapas de QTLs en los programas de MAS (Dwivedi, 2007).

Correlaciones canónicas e índices de selección

Cuando las variables de interés pertenecen a un sólo conjunto de variables aleatorias, el análisis de componentes principales puede ser de utilidad cuando lo que se desea es recuperar la máxima variabilidad con el menor número de combinaciones lineales; sin embargo, cuando las variables pertenecen de manera natural a dos conjuntos de variables aleatorias, y lo que se desea es maximizar la correlación existente entre ellas, el análisis de correlaciones canónica es mucho más eficiente (Muirhead, 2005). Así, considérese el vector de variables fenotípicas ($\mathbf{p}$) y el vector correspondiente de variables genotípicas ($\mathbf{g}$). En este caso, efectivamente, las variables del vector de variables fenotípicas, $\mathbf{p}$, y las variables del vector de variables genotípicas, $\mathbf{g}$, pertenecen de manera natural a dos conjuntos de variables. El análisis de correlaciones canónicas permite expresar la correlación entre $\mathbf{p}$ y $\mathbf{g}$ en su forma más simple por medio de transformaciones lineales de $\mathbf{p}$ y $\mathbf{g}$, es decir, $Y = \boldsymbol{\beta}'\mathbf{p}$ y $Z = \boldsymbol{\theta}'\mathbf{g}$. En la presente investigación se maximizó la correlación de las transformaciones lineales $Y = \boldsymbol{\beta}'\mathbf{p}$ y $Z = \boldsymbol{\theta}'\mathbf{g}$, y esto permitió desarrollar la teoría de MESIM y de MESIM_{GW}.

Los índices de selección se construyen con el fin de seleccionar para la mejor expresión de varios caracteres de manera simultánea. En la práctica, el índice de selección debe elegirse con base en la información con que se cuenta. Los índices de Lange y Whittaker (2001) y MESIM_{GW} requieren básicamente la misma información, lo mismo ocurre con MESIM y el índice de selección de Lande y Thompson (1990), pero el método de selección de Bernardo y Yu (2007)

requiere información diferente: en éste se utilizan los mejores predictores lineales (BLUPs por sus siglas en inglés). Esto significa que es procedente comparar el índice de Lange y Whittaker (2001) *vs* $MESIM_{GW}$ y MESIM *vs* Lande y Thompson (1990), pero podría no ser estrictamente válida la comparación del índice de Lange y Whittaker (2001) y $MESIM_{GW}$ *vs* MESIM y el índice de Lande y Thompson (1990), aunque es común hacer tal tipo de comparaciones, como lo hicieron Bernardo y Yu (2007) al comparar su método con el índice de selección de Lande y Thompson (1990).

En la construcción de los índices de selección se requiere la información completa de todos los individuos que conforman la población, lo que no ocurre con métodos de selección como la selección simultanea de caracteres independientes (Saxton, 2004). Así, los índices de selección implican mayor trabajo y costo de producción, sin embargo, las predicciones que se obtienen con el índice de selección son más precisas y eficientes que con cualquier otro método de selección (Hazel y Lush, 1942), lo cual justifica la inversión y el trabajo adicional.

Los índices de selección de Lange y Whittaker (2001), Lande y Thompson (1990), $MESIM_{GW}$ y MESIM son métodos de selección indirecta; es decir, con ellos se selecciona, utilizando la información fenotípica y la que proporcionan los MM, el valor genotípico neto de los caracteres de interés; por ello, maximizar la correlación entre la combinación lineal conformada por los fenotipos y los marcadores moleculares, y la combinación lineal conformada por los genotipos y los marcadores moleculares, es de suma importancia, y el método matemático que se utilice para la maximización de tal correlación, determinará la eficiencia y la aplicabilidad de los índices de selección.

CONCLUSIÓN GENERAL

En el presente trabajo se presentaron dos nuevos enfoques de la selección asistida por marcadores moleculares que tienen como base las ideas de Lande y Thompson (1990), Lange y Whittaker (2001) y Cerón-Rojas et al. (2008). Por ello la investigación se dividió en dos partes principales. En la primera se construyó un índice de selección basado en marcadores moleculares llamado MESIM (*Molecular Selection Index Method*). Los resultados teóricos y los obtenidos por simulación indican que MESIM tiene al menos tres ventajas sobre el método de Lande y Thopmson (1990): no requiere ponderaciones económicas; las propiedades muestrales del índice y de los estimadores de sus parámetros son conocidas en el contexto asintótico y el avance genético promedio que se alcanza con él es mayor o igual al de Lande y Thompson (1990). En la segunda parte se extendió la teoría básica de MESIM al caso de la selección del genoma completo incorporando los marcadores moleculares como variables aleatorias en el índice de selección, tal y como los propusieron Lange y Whittaker (2001). A tal procedimiento se le llamó índice de selección molecular del genoma completo basado en eigen análisis, o $MESIM_{GW}$ (por sus siglas en inglés). $MESIM_{GW}$ se comparó fundamentalmente con el índice de selección de Lange y Whittaker (2001) y, de acuerdo con los resultados teóricos y con base en los resultados de simulación por computadora es posible concluir que $MESIM_{GW}$ tiene ventajas sobre el índice de selección de Lange y Whittaker (2001) similares a las que tiene MESIM sobre el índice de selección de Lande y Thompson (1990).

LITERATURA CITADA

Arús, P., and J. Moreno González, 1993 Marker-assisted selection. *In* Hayward, M. D., Bosemark, N.O., and Romagosa, I.: Plant Breeding: Principles and Prospects. Chapman and Hall. University Press, Cambrige, Great Britain, pp. 315-331.

Beavis, W. D., 1998 QTL analyses: power, precision, and accuracy , *In*: Paterson A. H. (ed.) Molecular Dissection of Complex Traits. CRC Press, Boca Raton, New York, USA, pp. 145-163 .

Bernardo, R., and J. Yu, 2007 Prospects for genomewide selection for cuantitative traits in maize. Crop Science 47: 1082-1090.

Borecki, I. B., and B. K. Suarez, 2001 Linkage and association: basics concepts. Advances in Genetics 42: 45-66.

Carbonell, A. C., 1997 Methods for QTL analysis. V Meeting of International Biometry Network for Central America, the Caribbean, Mexico, Colombia and Venezuela, Xalapa, Veracruz, México; 53 p.

Cerón-Rojas, J. J., J. Crossa, J. Sahagún-Castellanos, F. Castillo-González, and A. Santacruz-Varela, 2006 A selection index method based on eigenanalysis. Crop Science 46: 1711-1721.

Cerón-Rojas, J. J., J. Sahagún-Castellanos, F. Castillo-González, A. Santacruz-Varela, and J. Crossa, 2008 A restricted selection index method based on eigenanalysis. Journal of Agricultural, Biological, and Environmental Statistics 13: 440-457.

Cubero, J. I., 2003 Introducción a la Mejora Genética Vegetal. Mundi-Prensa, España, 566 p.

Dekkers, J. C. M. , and P. Settar, 2004 Long-term selection with known quantitative trait loci. Plant Breeding Reviews 24: 311-335.

Dwivedi, S. L., J. H. Croch, D. J. Mackhill, Y. Xu, M. B. Blair, M. Ragot, H. D. Upadhyaya, and R. Ortiz, 2007 The molecularization of public sector crop breeding: progress, problems, and prospects. Advances in Agronomy 95: 163-315.

Gimelfarb A. and R. Lande, 1994 Simulation of marker-assisted selection in hybrid populations. Genet. Res. Camb. 63: 39-47.

Gimelfarb A. and R. Lande, 1995 Marker-assisted selection and marker-QTL associations in hybrid populations. Theoretical and Applied Genetics 91: 522-528.

Haley, C. S. and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69: 315-324.

Hazel, L. N., y J. L. Luz, 1942 The efficiency of three methods of selection. Journal of Heredity 33: 393-399.

Jansen, R. C., 2003 Quantitative trait loci in inbred lines, Vol. I, pp. 445-476 in Handbook of Statistical Genetics, 2nd Edition. Balding, D.J., M. Bishop and C. Cannings (eds.). John Wiley and Sons, England.

Jensen, J., 1989 Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. Theoretical and Applied Genetics 78: 613-618.

Kempthorne, O., and A. W. Nordskog, 1959 Restricted selection indices. Biometrics 15: 10-19.

Knapp, S. J., W. C. Bridges, and D. Birkes, 1990 Mapping quantitative trait loci using molecular marker linkage maps. Theoretical and Applied Genetics 79: 583-592.

Knapp, S. J., 1991 Using molecular markers to map multiple quantitative trait loci: models for backcross, recombinant inbred and doubled haploid progeny. Theoretical and Applied Genetics 81: 333-338.

Knapp, S. J., 1998 Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Science 38: 1164-1174.

Koebner R. M. D., W. Powell, P. Donini, 2001 Contributions of DNA molecular marker technologies to the genetics and breeding of wheat and barley. Plant Breeding Reviews 21: 181-220.

Kollo T, and D. von-Rosen, 2005 Advanced Mmultivariate Statistics with Matrices. Springer, The Netherlands.

Lande, R., and R. Thompson, 1990 Efficiency of marker-assisted in the improvement of quantitative traits. Genetics 124: 743-756.

Lander, E. S., and D. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP's linkage maps. Genetics 121: 185-199.

Lange, C., and J. C. Whittaker, 2001 On prediction of genetic values in marker-assited selection. Genetics 159: 1375-1381.

Li, Z., 1998 Molecular analysis of epistasis affecting complex traits, *In*: Paterson A. H. (ed.) Molecular Dissection of Complex Traits. CRC Press, Boca Raton, New York, USA, pp. 119-1130.

Martinez, O., and R. N. Curnow, 1994 Missing markers when estimating quantitative trait loci using regression mapping. Heredity 73: 198-206.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

Muirhead, R. J., 2005 Aspects of Multivariate Statistical Theory, John Wiley and Sons, New York.

Piyasatian, N., R. L. Fernando, and J. C. M. Dekkers, 2007 Genomic selection for marker-assisted improvement in line crosses. Theoretical and Applied Genetics 115: 665-674.

Saxton, A. M., 2004 Genetic selection, *In*: Genetic Analysis of Complex Traits Using SAS. Saxton, A. M. (ed.), SAS Publishing, SAS Institute Inc., SAS Campus Drive, Cary, North Carolina, USA, pp. 55-67.

Schott, J.R., 2005 Matrix Analysis for Statistics. Ed. 2, John Wiley and Sons, New Jersey.

Soller, M., T. Brody, and A. Genizi, 1976 On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theoretical and Applied Genetics 47:35-39.

Smith, H. F., 1936 A discriminant function for plant selection, *In*: Papers on Quantitative Genetics and Related Topics. Deparment of Genetics, North Carolina State College, Raleigh, North Carolina, pp. 466-476.

Van Ooijen, J. W., 1992 Accuracy of mapping quantitative trait loci in autogamous species.

Whittaker, J. C., 2003 Marker-assisted selection and introgression, pp. 554-574 *In*: Handbook of Statistical Genetics. Vol. I, ed. 2, D. J. Balding, M. Bishop y C. Cannings (eds.). John Wiley & Sons, Chichester, U. K.

Wong, C. K., and R. Bernardo, 2008 Genomewide selection in oil palm: increasing selection gain per unit time and cost with small population. Theoretical and Applied Genetics 116: 815-824.

Xu, Y., and J. H. Crouch, 2008 Marker-assisted selection in plant breeding: from publications to practice. Crop Science 48: 391-407.

Zhan W., and C. Smith 1992 Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theoretical and Applied Genetics 83: 813-820.

Zhan W., and C. Smith 1993. Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. Theoretical and Applied Genetics 86: 492-496.