Universidad Autónoma de Baja California
Instituto de Ingeniería

Maestría y Doctorado en Ciencias e Ingeniería

Título:

"Implementación de un Modelo Estadístico para la Estimación de Valores Genómicos de Crianza de Ganado Bovino, Utilizando Marcadores SNP"

TESIS PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS

PRESENTA:
NATANIEL ANGULO CUEVAS

DIRECTORES DE TESIS
DR. RAFAEL VILLA ANGULO
DR. VICTOR MANUEL GONZALES VIZCARRA

Mexicali, B.C. **Mayo, 2015**

RESUMEN

La Estimación de los Valores de Crianza tradicional combina solo datos fenotípicos y de pedigrí, hoy en día, con la inclusión de marcadores moleculares, tales como los SNPs (Single Nucleotid Polimorfims, por sus siglas en ingles), en la integración del análisis genético en los esquemas de selección, ya es posible integrar datos genotípicos para realizar estas estimaciones. Avances en las tecnologías de genotipificación de salida masiva para bovinos llevaron al desarrollo de los SNPchips, de los cuales hay de diferentes capacidades. Para la Estimación de los Valores Genómicos de Crianza (GEBV, por sus siglas en ingles) existen diferentes métodos basados en el contexto del Modelo Lineal de Mezclas (Linear Mixed Models), en los que encontramos el GBLUP (Genomic Best Linear Unbiased Prediction) el cual es un método que utiliza relaciones genómicas para realizar dichas estimaciones. Para esto, es usada una matriz de relaciones genómicas, estimada de marcadores moleculares (como SNPs). Esta matriz define la covarianza entre individuos en un grupo de estudio basada en similitud observada a nivel genómico, en vez de la similitud basada en pedigrí. En este trabajo de tesis se reporta la implementación y pruebas de una herramienta bioinformática que implementa el GBLUP para estimar los valores genómicos de ganado bovino. Esta herramienta es parte integral del primer sistema tecnológico para la implementación de un programa de mejora genética basado en Selección Genómica, en el Estado de Baja Californias.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

CAPÍTULO 1. INTRODUCCIÓN

1.1. Antecedentes.

Las especie bovina constituyen un grupo muy importante de animales, no solo por su posibilidad de explotación y aporte en la economía de muchos países, sino por ser de las primeras especies domesticadas por el hombre en la era Neolítica (~10,000 AC), hecho que le ha permitido acompañarlo en su evolución, y en gran parte de sus rutas migratorias, a través de la historia. El ganado bovino representa así junto con los cerdos, perros y gatos, una clase de mamíferos placentados que ha coevolucionado con los humanos, y su estudio permite no sólo ampliar el conocimiento de estas especies, sino brindar huellas importantes sobre la evolución e historia natural de las misma [1].

La selección de animales para reproducción se ha venido practicando desde los inicios del proceso de domesticación; A principios la selección se basaba exclusivamente en la apariencia física y la capacidad productivas (rasgos fenotípicos) de los animales. Sin embargo, en los últimos dos siglos se ha venido implantando la sistematización dirigida por parte de los productores, y se han diseñado programas de crianza de animales

beneficiando funciones especificas, tales como producción de carne, de leche, facilidad de parto y longevidad, entre otras. En los últimos 50 años, con los avances en la técnicas de análisis genético, y más recientemente (últimos 5 años) análisis genómico, ha sido posible evaluar el mérito genético de los animales, lo que ha permitido tomar decisiones de selección más eficaces. La reciente inclusión de la evaluación genómica se proyecta como la tecnología más prometedora para implementar estrategias de selección y garantizar la capacidad productiva de las nuevas generaciones de ganado bovino [2].

1.2. Selección Genómica.

1.2.1 Genómica

La Genómica es el conjunto de ciencias y técnicas dedicadas al estudio integral del funcionamiento, el contenido, la evolución y el origen de los genomas (la totalidad de la información genética que posee un organismo o una especie en particular) [3]. Existen muchas áreas relacionadas con la genómica que se han estado desarrollando a lo largo de los años, algunas de las más importantes por su potencial tanto económico como social y ambiental son la medicina genómica, la genómica forense, la genómica ambiental, la genómica industrial, y en especial la genómica agropecuaria, que es el área de aplicación de este trabajo de tesis.

El desarrollo de la Genómica Agropecuaria podría detonar en México la sustentabilidad del sector agropecuario. En especial la mejora genética de las especies, basada en el análisis de su genoma, cambiará los paradigmas de producción y conservación de recursos

naturales. En México ya se están haciendo algunos esfuerzos por adoptar las tecnologías genómicas para la mejora de la producción de leche y carne proveniente de la especie bovina, tal es el caso del proyecto de mejora genética basado en técnicas de Selección Genómica, que se desarrolla actualmente por investigadores los Institutos de Investigaciones en Ciencias Veterinarias e Ingeniería de la Universidad Autónoma de Baja California; con el cual se pretende generar nuevos linajes de ganado bovino que serán mejor adaptados a las condiciones climáticas de las zonas áridas de Baja California, así como más resistentes a los patógenos propios de la regios, y más productivos [4].

1.2.2 Genotipo y Fenotipo.

El genotipo de un animal representa el gen o grupo de genes responsable de un rasgo en particular. En un sentido más general, el genotipo describe todo el grupo de genes que un individuo ha heredado. Como contraste, el fenotipo es el valor que toma un rasgo; en otras palabras, es lo que puede ser observado o medido. Por ejemplo, el fenotipo puede ser rasgos físicos y/o fisiológicos tales como las enfermedades, el color de piel, la producción individual de leche de una vaca, la facilidad de parto, etc. [5].

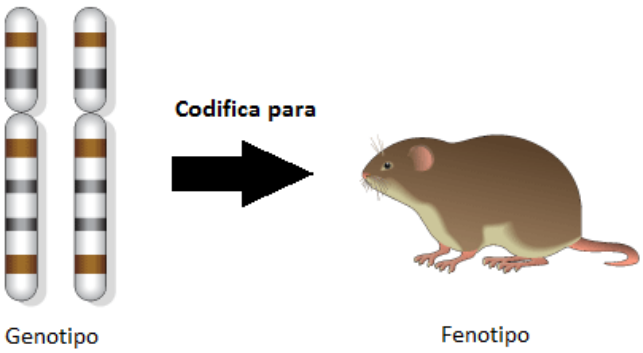


Figura 1. Genotipo y Fenotipo. Los genotipos corresponden a la información genética de los distintos rasgos, mientras que los fenotipos corresponden al producto físico o fisiológico que puede ser medido y caracterizan los distintos rasgos.

Existe una diferencia importante entre genotipo y fenotipo. El genotipo es esencialmente una característica fija del organismo; permanece constante a lo largo de la vida del animal y no es modificado por el medio ambiente. Cuando solamente uno o un par de genes son responsables por un rasgo, el genotipo permanece generalmente sin cambios a lo largo de la vida del animal (p. ej., color de pelo). En este caso, el fenotipo otorga una buena indicación de la composición genética del individuo. Aún así, para algunos rasgos, el fenotipo cambia constantemente a lo largo de la vida del individuo como respuesta a factores ambientales. En este caso, el fenotipo no es un indicador confiable del genotipo. Esto generalmente se presenta cuando muchos genes se involucran en la expresión de un rasgo, tal como producción de leche [5].

1.2.3 Selección asistida por marcadores.

Con el método de selección tradicional, que utiliza información fenotípica y de pedigrí, se logra tener una efectividad del 30 – 40% en las predicciones. Esto se debe en gran parte a que, aun cuando logramos la mejora genética al ver un aumento en nuestra producción, el conjunto de genes que han sido favorecidos (o desfavorecidos) en nuestro programa de selección sigue siendo una caja negra. Se sabe que un buen balance entre el manejo adecuado de la crianza y la selección minuciosa de animales con características fenotípicas mayores en los rasgos que deseamos, ha conducido a lograr progreso, pero de igual forma se sabe que muchos de los rasgos observables (fenotipos) de fisiología del animal, de sus

propiedades bioquímicas, de su morfología, de desarrollo, comportamiento y producción, son epigenéticos. Esto es, son regulados por la interacción de los genes con el medio ambiente. Lo que hace que según los factores de manejo, la raza, la región geográfica, la época del año y el estado de lactancia, entre otros, sumado a la naturaleza dominante de algunos genes que posee el animal, al efecto aditivo de genes que actúan de forma independiente y al efecto epispástico de genes que actúan de forma interactiva para controlar los rasgos del animal, influya de forma determinante en el rendimiento productivo.

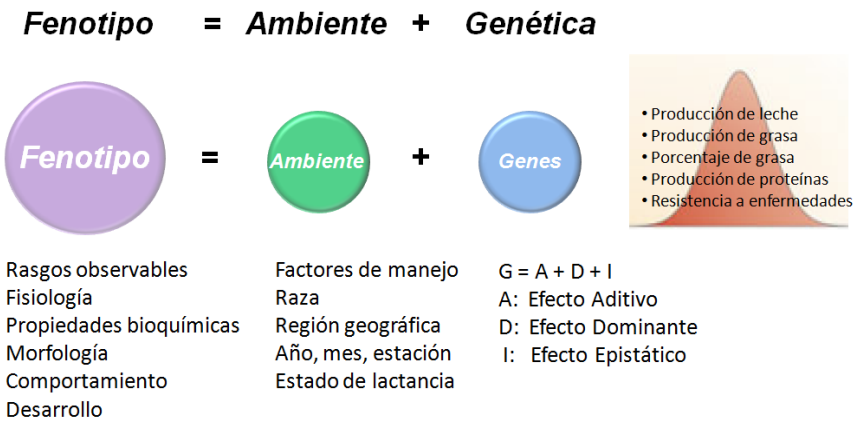


Figura 2. La mayor parte de los rasgos de importancia económica en el ganado bovino son epigenéticos. La mayor parte de los fenotipos son regulados por la interacción de los genes con el medio ambiente.

Un adelanto crucial en la integración del análisis genético en los esquemas de selección fue la inclusión de los marcadores moleculares, con los cuales se integra de forma específica el efecto de los genes en el rendimiento productivo del animal. Un marcador de ADN es una posición, o región, dentro de un cromosoma, físicamente identificable y cuya heredabilidad puede ser monitoreada [6]. La función inicial y básica de los marcadores es identificar la expresión de genes relacionados con las características de importancia

económica, y utilizar esta información para hacer una estimación más acertada de los Valores de Crianza (EBV, “Estimated Breeding Values”, por sus siglas en Ingles).

Sin embargo, aún cuando la idea de usar marcadores de ADN para incrementar la tasa de ganancia genética en el ganado lechero, ha existido por décadas(ej. [7]), la adopción de la selección asistida por marcadores (MAS, “Marker Asisted Selection”, por sus siglas en inglés) en la industria lechera había estado limitada, hasta hace muy poco tiempo. Existían diferentes razones para esto; para muchos rasgos cuantitativos, de importancia económica, tales como los rasgos de producción y salud, en el ganado lechero, una gran cantidad de locus afectan a cada rasgo, donde cada locus captura una proporción limitada de la varianza genética. Consecuentemente, solo una ganancia relativamente pequeña era posible con el número limitado de marcadores que se tenían disponibles, y el costo de genotipado de estos era alto. En adición, la complejidad para el cálculo de los valores genéticos de crianza incluyendo la información de los marcadores era una barrera más para la aplicación de la selección asistida por marcadores [7].

En el año 2001, investigadores de Holanda y Australia demostraron a través de simulaciones por computadora, que usando marcadores de ADN distribuidos a lo largo de todo el genoma permitía capturar de forma simultánea los efectos de todos los genes que afectan y regulan las características de importancia económica, aumentando con esto de forma substancial la precisión de la selección asistida por marcadores [7]. Este trabajo inicial sobre los beneficios de utilizar información de todo el genoma sirvió para establecer las necesidades tanto de información como de tecnología para lograr los avances demostrados en las simulaciones. Estos requerimientos incluían: 1) Contar con

poblaciones de animales con mediciones fenotípicas precisas y sus muestras de ADN genómico, 2) Marcadores de polimorfismo de ADN posicionados en todo el genoma, 3) Tecnología de laboratorio para realizar análisis de una alta densidad de marcadores y 4) Algoritmos y métodos para el análisis de datos genotípicos y fenotípicos. Estos requerimientos, aunados a otros beneficios de gran importancia, impulsaron a varios países, a través de sus centros de investigación y universidades a generar dos consorcios con el fin de secuenciar el genoma y crear un mapa de haplotipos de la evolución *Bos Tarus*.

1.2.4 Avances de la tecnología de genotipificación de salida masiva para bovinos.

Los proyectos de secuenciación de genomas completos, en especial los genomas humano y bovino, han sido detonadores del desarrollo de nuevas tecnologías para la captura de información genética proveniente de muestras de ADN. Un ejemplo de ellas son las tecnologías de salida masiva para extracción de genotipos ("HTGT High-Throughput Genotyping Technologies", por sus siglas en inglés) la cuales han jugado un rol invaluable en la captura de diversidad genética y variaciones hereditarias entre individuos. La disponibilidad de las HTGTs ha permitido la extracción de genotipos de especies adicionales, que han servido como organismos modelos para resolver la complejidad de la evolución humana y para extrapolar de una forma efectiva, información genética por medicina comparativa (veterinaria) a medicina humana.

En especial en la especie bovina, con el proyecto de secuenciación del genoma fue posible descubrir miles de marcadores de ADN, en forma de Polimorfismo de Nucleótido Simple

("SNP Single Nucleotide Polimorphism" por sus siglas en Ingles) [8]. Para hacer esto se realizó un estudio de descubrimiento de SNPs obteniendo como resultado más de 6 millones de SNPs putativos.

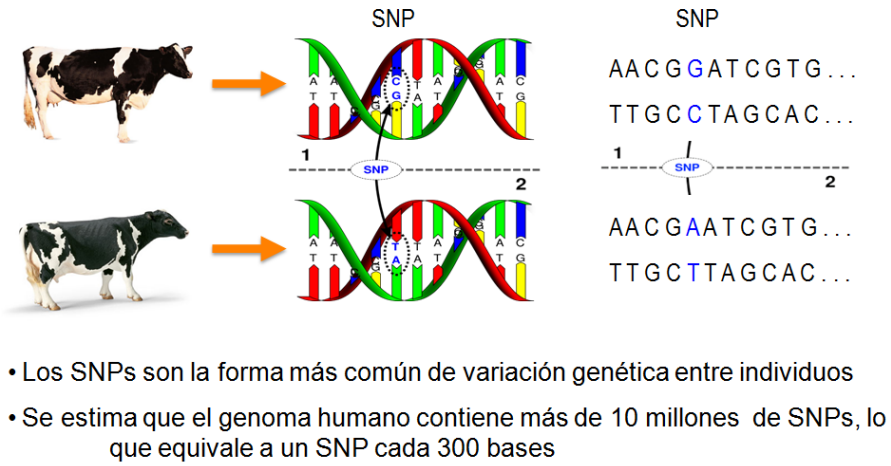


Figura 3. La forma más común de polimorfismos en los genomas son los SNPs.

Inicialmente de validaron 54,609 SNPs (50K) muestreando 565 animales de 19 razas distintas ubicadas alrededor del mundo, y se realizaron los primeros estudios de caracterización de la estructura genética de la especie [8].

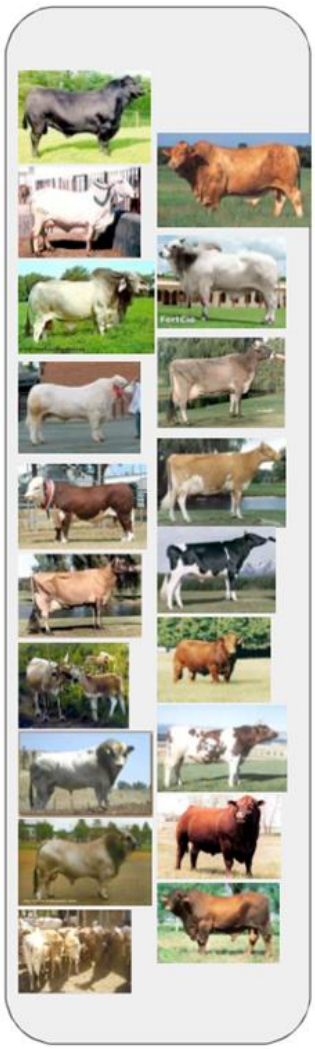


TABLE 3: BOVINESNP50 BEADCHIP CONTENT VALIDATION

BREED	SAMPLES	POLYMORPHIC LOCI*	MEAN MAF†	MEDIAN MAF†
Angus	60	41,491	0.21	0.21
Beefmaster	24	42,925	0.22	0.21
Bos indicus Gir	24	23,971	0.11	0.02
Bos indicus Nelore	21	25,814	0.11	0.02
Brahman	25	30,284	0.13	0.08
Brown Swiss	24	36,347	0.19	0.17
Charolais	26	42,589	0.22	0.21
Guernsey	21	38,632	0.19	0.17
Hereford	32	42,992	0.20	0.23
Holstein	64	42,730	0.22	0.22
Jersey	28	35,976	0.18	0.14
Limousin	45	42,821	0.22	0.22
N'Dama	25	29,049	0.14	0.08
Norwegian Red	21	42,782	0.22	0.21
Piedmontese	24	42,185	0.22	0.21
Red Angus	15	40,188	0.21	0.20
Romagnola	24	38,830	0.20	0.19
Santa Gertrudis	24	42,064	0.22	0.21
Sheko	20	35,726	0.17	0.12
Outgroup‡	18	11,206	0.05	0.00
Overall	565	47,545	0.25	0.24

Figura 4. Razas utilizadas para la validación de los 54,608 SNPs. Se tomaron muestras de 565 animales pertenecientes a 21 razas de ganado bovino y se realizaron los primeros estudios de estructura genética de genoma completo (imagen tomada de Illumina.com).

La validación de los SNPs facilitó el diseño e implementación del primer Chip de Genotipificación de Salida Masiva para ganado bovino (Bovine SNPchip50 comercializado por Ilumina Inc. desde 2009). El SNPchip50 [9], en solo unos meses se convirtió en el estándar para realizar estudios genéticos de las razas de ganado pertenecientes a la evolución *Bos Taurus*.

Figure 5. Características del SNPchip50, y la tecnología de genotipificación de salida masiva. Estándares para realizar Genotipificación masiva en programas de mejora genéticas asistida por marcadores.

Actualmente existen en el mercado seis SNPchips distintos para ganado bovino, cuya capacidad va desde baja hasta alta densidad. Cinco de estos son de la compañía Illumina: Bovine3K [10], BovineLD [11], BovineSNP50v.1 [9], BovineSNP50v.2 [9], y el BovineHD [12], los cuales tienen capacidad de 2900, 6909, 54001 y 777962 SNPs respectivamente. Uno es de la empresa Affimetrix: Axiom Bos 1 [13] el cual tiene una capacidad de 648875 SNPs. Los propósitos específicos de estos chips van desde cálculo del mérito neto para características productivas, hasta calculo de valores genómicos y estudios de asociación de SNPs con enfermedades y distintos fenotipos. La tabla 1 presenta un resumen de los chips incluyendo el propósito específico de cada uno de ellos.

Tabla 1. Resumen de SNPchips comerciales y el tipo de análisis que puede ser realizado con ellos.

SNPchip	Total de SNPs	Compañía	Propósito principal
Bovine3K	2,900	Illumina	Merito genético, verificación de porcentaje, trazabilidad de ganado
BovineLD	6,909	Illumina	Valor genómico de crianza (GEBV), variación genética
BovineSNP50v.1	54,001	Illumina	Selección genómica, Identificación de QTLs, Evaluación de mérito genético, comparación genética
BovineSNP50v.2	54,609	Illumina	Selección genómica, Identificación de QTLs, Evaluación de mérito genético, comparación genética
BovineHD	777,962	Illumina	Variación genética dentro de poblaciones, selecicon genómica, identificación de QTLs, evaluación de mérito genético, mapeo entre razas, desequilibrio de ligamiento, comparación genética, caracterización de razas para estudios de biodiversidad
Axiom BOS 1	648,875	Affimetrix	Merito genético, estudios de asociación, evaluación genómica, estudios de respuesta a fármacos, desequilibrio de ligamiento

1.2.5 Selección Genómica.

La revolución de la selección genómica empezó con dos desarrollos claves; el primero fue la reciente secuenciación del genoma bovino [14], y el segundo el desarrollo del Bovine SNPchip50 [9]. La Selección genómica esta revolucionando a la industria del ganado y se refiere a la toma de decisiones de selección basada en el valor genómico de cría de los animales (GEBV, “Genomic Estimated Breeding Value” por sus siglas en inglés). Los GEBVs

son calculados como la suma de los efectos de una cantidad grande de marcadores moleculares (SNPs), distribuidos a lo largo de todo el genoma. Los genotipos y/o haplotipos de los marcadores involucrados capturan los efectos de todos los Locus de Características Cuantitativas (QTL, "Quantitative Trait Loci" por sus siglas en inglés) que contribuyen a la variación de un rasgo de valor económico. El efecto de los QTLs, inferidos ya sea de los haplotipos o genotipos de marcadores de tipo polimorfismo de nucleótido simple (SNPs), es primeramente estimado en una población grande de referencia con información fenotípica. En las generaciones subsecuentes solo es necesaria la información de los marcadores para calcular los GEBVs.

La tecnología de selección genómica involucra tres pasos principales; 1) captura de las muestras y obtención del ADN genómico, 2) aplicación de los SNPchips para obtención de los genotipos y 3) análisis bioinformático de la información para obtener los GEBVs. La captura de la muestras pueden ser hechas de esperma, sangre, saliva, segregación nasal o una pinchada en la oreja, entre otras. Estas muestras deben ser transportadas al laboratorio para su procesado químico de extracción de ADN. El ADN debe ser ADN genómico completo y purificado. El ADN es amplificado y depositado en los SNPchip. Se realiza una reacción de hibridización del ADN al SNPchip para posteriormente correr una reacción de extensión en la cual el valor alélico de cada posición polimórfica queda visible por elementos fluorescentes. Por medio de un escáner laser son medidos los valores de los alelos en lo polimorfismos y se genera una base de datos con la información genética de cada uno de los animales de la muestra.

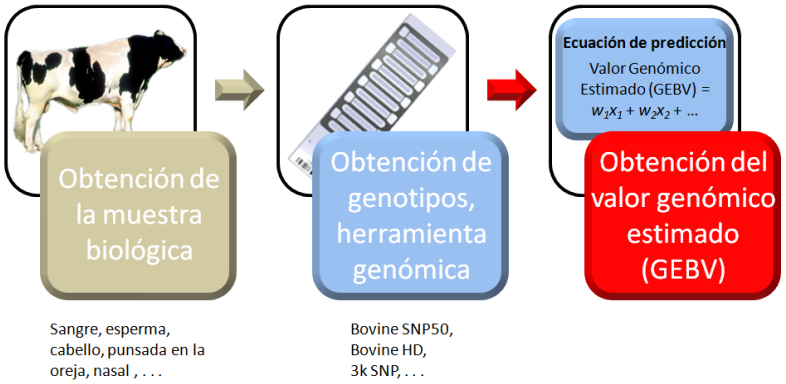


Figura 6. Pasos de la Selección Genómica. La selección genómica involucra tres pasos principales: 1) Captura de la muestras y obtención de ADN genómico, 2) Extracción de genotipos y 3) Análisis de la información.

Métodos bioinformáticos de análisis son aplicados a los datos para hacer estudios de asociación de genes con los rasgos de importancia, con enfermedades y para obtener los GEBVs. Toda esta información resultante es integrada a los criterios para seleccionar los sementales y las vacas que serán utilizadas para generar las nuevas crías en el hato.

La selección genómica ha sido adoptada por la mayor parte de los países productores de leche de Europa y por Estados Unidos y Canadá en América. La figura tabla 2 muestra los resultados en la efectividad de hacer selección basado en datos genómicos.

Tabla 2. Resultados de la efectividad de la selección genómica en cuatro países de Europa y América.

País	Número de toros	Número de SNPs	Características de producción	Método estadístico de cálculo	Certeza de los GEBV
Nueva Zelanda (Harris et al, 2008)	4,500	44,146	Leche, peso al nacimiento, fertilidad, longevidad, Células Somáticas	BLUP, Bayes A, Bayes B	50-60% (34% PABV)
USA y Canadá (VanRaden, 2009)	3,576	38,416	Merito neto, producción de leche, producción de grasa, longevidad	Similar al BLUP	50% (27% PABV%)

Holanda (De Roos, 2009)	1,583	57,660	Producción de grasa, producción de proteínas, patas y piernas, profundidad de la ubre, células somáticas, fertilidad	-	Aumento de (20% PABV)
Australia (Hayes et al, 2009)	798	38,259	Producción de proteínas, fertilidad, Índice de selección Australiano, Ranquin de ganancia Australiano	BLUP, Bayes A	14-55% (41% PABV)

La selección genómica es la tecnología que ha generado el incremento más grande en la tasa de mejora genética para la industria lechera en los últimos 20 años. Y, aun cuando se había trabajado de forma simulada para proyectar sus beneficios, es hasta ahora que se está probando su verdadero potencial. El incremento en la efectividad de la GEBVs (Valor genético basado en información genómica) sobre los EBV (valor genético basado en información de fenotipos y pedigrí) para toros jóvenes sin record de sus hijas, es impresionante, va del 2% al 20%. El incremento en la efectividad de los GEBVs está siendo usado de dos formas por las compañías de crianza de sementales. En algunos casos grupos grandes de toros jóvenes están siendo muestreados para seleccionar un número reducido de toros y llevarlos a programas de progenie. Esto reduce el costo de los programas de crianza y resulta en una ganancia genética extra. Otras compañías están ofertando grupos de toros jóvenes basados únicamente en su GEBVs, tan rápido como estos están listos para reproducción. Esto resulta en una ganancia genética mayor, como resultado de reducir la longitud del intervalo generacional.

1.3. Métodos Estadísticos utilizados para estimar valor genómico de crianza.

Un aspecto fundamental del cual depende, en gran parte, el éxito de un programa de mejora genética, es la integración y procesamiento adecuados de una gran cantidad de datos de genotipos, de pedigrí, de salud, y del medio ambiente en el cual ha crecido cada animal. Para este efecto se han desarrollado distintos métodos estadísticos y matemáticos [15,16,17]. Los métodos para estimar valores genómicos de crianza podemos dividirlos en dos grupo; los que utilizan Modelos Bayesianos [16, 17], y los que utilizan modelos de Regresión Aleatoria [15]. Los métodos que utilizan modelos Bayesianos están basados en hacer asociación multi-locus, y son resueltos con métodos matemáticos como Modelos de Markov entrenados con el método Monte Carlo, o Máxima Verosimilitud; mientras que los métodos que utilizan regresión aleatoria están basados en modelar sistemas de variables aleatorias utilizando Modelos Lineales de Mezclas, para calcular efectos aleatorios (genéticos y medio ambiente), partiendo de efectos fijos (pedigrí). Estos métodos son poco entendidos, razón por la cual es limitada la comunidad de científicos que están desarrollando sistemas de software bioinformático para programas de mejora genética utilizando selección genómica.

1.4. Planteamiento del Problema.

Un problema histórico con el ganado lechero de las regiones áridas de Baja California es la baja eficiencia productiva en los meses de verano, debido al estrés calórico provocado por las drásticas condiciones climáticas. Este problema provoca una reducción en la producción de leche hasta en 25% y permite tan sólo 10% de concepciones, y ha tenido un impacto tanto económico como social incalculable. En adición, en los últimos 5 años, la

situación de la producción lechera en el estado se ha agudizado provocando la mayor crisis productiva en los últimos 50 años. El factor detonador es la escases de alimentos para el ganado, porque aun cuando el cultivo de alfalfa y otros granos en el estado de Baja California son buenos, el problema de desabasto debido a la preferencia de los productores de exportar a Estados Unidos las cosechas, no ha cesado, y se proyecta que siga en aumento. En los últimos tres años, el 80% de las cosechas de alfalfa no han sido utilizadas para alimentación del ganado local. Esto ha provocado la desaparición de más del 50% de los hatos lecheros del estado, con el sacrificio de más de 15,000 animales, mismos que salieron del inventario habitual de producción.

Ante esta realidad, el productor lechero debe procurar en hacer mas eficientes sus recursos, buscando tener cada vez mejores animales, capaces de responder a las nuevas exigencias que implica la producción de leche en las condiciones actuales, dotados de una genética superior, aparejados con nuevas estrategias de alimentación, administración, manejo zootécnico, medicina preventiva, crianza de reemplazos y biotecnología de la reproducción. Es de crucial importancia el establecimiento de programas integrales a ser implementados en los hatos lecheros del estado, para mejorar la producción ganadera; en especial la mejora genética del ganado, beneficiando con programas de reproducción los rasgos de interés económico, así como la resistencia a efectos de estrés calórico y/o una baja susceptibilidad a enfermedades de predisposición genética, aunada a un programa de salud animal, de bioseguridad y de registro productivo adecuado.

Como respuesta a esta importante necesidad, empresarios productores ganaderos del estado, en coordinación con grupos de investigación de la Universidad Autónoma de Baja

California, emprendieron una iniciativa para generar programas de mejora genética basados en el genoma para contrarrestar los efectos de estrés calórico, aumentar la resistencia a enfermedades típicas de la región y generar linajes de ganado más adaptado y productivo en Baja California. La iniciativa se generó con el propósito de caracterizar a nivel de diferencias genómicas el ganado susceptible a sufrir estrés calórico, localizar los genes involucrados con el estrés calórico y otras patologías que afectan la producción de leche, diseñar e imprentar estrategias de reproducción de ganado óptimo genéticamente para la producción de leche bajo las condiciones ambientales de las regiones áridas del estado de Baja California y desarrollar la tecnología de instrumentación necesaria para experimentos de exploración y análisis de ganado [18].

A la fecha (Mayo 2015), se realizó un muestreo de 150 animales en distintas regiones del estado de Baja California, se extrajo ADN de la muestras y está en proceso la medición de genotipos con el SNPchip LD (6,909 SNPs), mismos que serán utilizados para hacer estudios de asociación y encontrar genes relacionados con la susceptibilidad a enfermedades típicas (Tuberculosis, Cetosis, Mastitis, entre otras), para caracterizar la estructura genética del ganado de Baja California, para estimar los valores genómicos de crianza e implementar el primer programa de mejora genética basado en Selección Genómica. Dentro de los avances que se han logrado están: la caracterización de variaciones genómicas en alta densidad, hecha en 12 vacas lecheras la ciudad de Mexicali con el SNPchip Axiom Bos 1 [19], y la caracterización de variables de manejo, de salud y de producción del Modelo Experimental de Producción Lechera del Instituto de

Investigaciones en Ciencias Veterinarias [20], donde se implementará el programa de Selección Genómica.

Para continuar con la implementación del programa de Selección Genómica es necesario desarrollar las herramientas bioinformática para administrar toda la información de salud, de manejo y producción de los animales, para la estimación de los Valores Genómicos de Crianza, para realizar las proyecciones de cruzamiento vaca-semental, para medir el progreso genético y para evaluar la eficacia del programa de mejora. En este trabajo de tesis se propone estudiar e implementar métodos estadísticos para estimar Valores Genómicos y desarrollar una herramienta bioinformática que será utilizada como base para la implementación de un software administrador que apoyará el desarrollo del primer programa de Selección Genómica en ganado bovino en el Estado de Baja California, y en México. Para el desarrollo de esta tesis de platean los siguientes objetivos y metas:

1.5 Objetivo y metas.

1.5.1 Objetivo.

El objetivo planteado para este trabajo de tesis es el siguiente:

Desarrollar una herramienta estadística para estimar Los valores Genómicos de Crianza, utilizando marcadores SNP, en ganado bovino del estado de baja california, con el fin de implementar un programa de mejora genética usando Selección Genómica.

1.5.2 Metas.

Paras lograr el objetivo se plantean las siguientes metas:

2 Realizar un estudio sobre los distintos métodos estadísticos para estimar el Valor Genómico de Crianza.

3 Adaptar el método estadístico elegido para estimar los Valores Genómicos que para las distintas características productivas del ganado de Baja California.

4 A partir del método estadístico adaptado desarrollar una herramienta para la estimación de los valores genómicos.

5 Estimación de los valores genómicos de crianza utilizando la herramienta desarrollada y genotipos de ganado bovino.

6 Validación de los resultados, contrastándolos con otras herramientas.

CAPÍTULO 2. MODELO ESTADISTICO PARA ESTIMACION DE VALORES GENOMICOS DE CRIANZA.

El nombre Modelos Lineales de Mezclas viene del hecho de que estos modelos son lineales en los parámetros, y que las covariables, o variables independientes, pueden implicar una mezcla de efectos fijos y aleatorios [21]. Estos modelos son útiles en ajustes donde las mismas unidades estadísticas son medidas repetidamente (estudios longitudinales). Debido a estas particularidades, basaremos nuestro estudio en un método que utilice los modelos lineales de mezclas.

Los modelos lineales de mezclas son asumidos en muchas de las aplicaciones genéticas y pueden ser representados matricialmente de la siguiente forma:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \tag{1}$$

Ó

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} X_{11}X_{12} \\ X_{21}X_{22} \\ . \\ . \\ . \\ X_{n1}X_{n2} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ b_n \end{bmatrix} + \begin{bmatrix} Z_{11}Z_{12} \\ Z_{21}Z_{22} \\ . \\ . \\ . \\ Z_{n1}Z_{n2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix}$$

Donde $\boldsymbol{y}$ es un vector de $\boldsymbol{n}$ observaciones (fenotipos, mediciones repetidas a través del tiempo), $\boldsymbol{X}$ y $\boldsymbol{Z}$ son matrices conocidas, $\boldsymbol{b}$ es un vector de efectos fijos, $\boldsymbol{u}$ es un vector de efectos aleatorios y $\boldsymbol{e}$ vectores de valores residuales aleatorios [22] con:

$$\boldsymbol{Var}\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix} \boldsymbol{\sigma^2}$$

Donde $\boldsymbol{\sigma^2}$ es un escalar, posiblemente desconocido, $\boldsymbol{G} = \boldsymbol{Var}(\boldsymbol{u})$ y $\boldsymbol{R} = \boldsymbol{Var}(\boldsymbol{e})$.

2.1. Método BLUP (Best Linear Unbiased Prediction).

En la crianza animal, BLUP ó Mejor Predicción Lineal Insesgada, es una técnica para la estimación del merito genético. En general, este es un método de estimación de efectos aleatorios. El contexto del BLUP es el modelo lineal de mezclas.

Por insesgado se refiere a que ***E(predictor) = E(k'b+m'u) = k'b***. Henderson [15] mostró que la Mejor Predicción Lineal Insesgada (BLUP, por sus siglas en ingles) de ***k'b+m'u*** es

$$\boldsymbol{k'\hat{b}} + \boldsymbol{m'GZ'V^{-1}}(\boldsymbol{y} - \boldsymbol{X\hat{b}}) \qquad (2)$$

donde ***V = R + ZGZ'***, $\boldsymbol{\hat{b}}$ es cualquier solución a (3), la ecuación generalizada de mínimos cuadrados

$$\boldsymbol{XV^{-1}\hat{b}} = \boldsymbol{X'V^{-1}y} \qquad (3)$$

La dificultad con esto es que ***V*** frecuentemente es una matriz tan grande que la inversión de esta puede no ser viable. Un método alternativo fue sugerido por Henderson [15]. La predicción de ***k'b+m'u*** es $\boldsymbol{k'\hat{b} + m'\hat{u}}$, donde $\hat{\boldsymbol{b}}$ y $\hat{\boldsymbol{u}}$ son cualquier solución para (4);

$$\begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix} \quad (4)$$

Esta solución es conocida como las Ecuaciones del Modelo Mixto ó MME (Mixed Model Equations, por sus siglas en ingles).

Henderson comprobó que $\hat{\boldsymbol{b}}$ de (4) es una solución a (3), y que $\hat{\boldsymbol{u}}$ de (4) es igual a $\boldsymbol{GZ'V^{-1}(y - X\hat{b})}$ de (2) [15].

La ventaja computacional de (4) sobre (2) es que no se requiere ***V*** ni su inversa. ***R*** tiene las mismas dimensiones que ***V***, pero ***R*** es usualmente una matriz identidad [15].

El método BLUP tiene algunas variantes, en las que podemos encontrar el ABLUP (Traditional BLUP) y GBLUP (Genomic BLUP). La diferencia entre estas dos variantes del método BLUP recae en que para el ABLUP o BLUP tradicional, la Matriz de Relaciones Genéticas, llamada matriz A, utilizada para la solución, está basada solo en los datos de pedigrí de los animales en el grupo de estudio, mientras que en el GBLUP o BLUP Genómico, dicha matriz es definida como Matriz de Relaciones Genómicas, llamada matriz G, y se calcula a partir de la matriz de incidencia de marcadores genéticos (SNPs), los cuales son obtenidos a través de la genotipificación del ADN de cada animal., utilizando SNPchips.

2.2. Método GBLUP (Genomic Best Linear Unbiased Prediction).

GBLUP es un método que utiliza relaciones genómicas para estimar el mérito genético de un individuo. Para este propósito, se necesita una matriz de relaciones genómicas, estimada de la información de los marcadores de ADN. La matriz define la covarianza entre individuos basado en similitud observada a nivel genómico, en lugar de la similitud esperada basada en el pedigrí, de modo que las predicciones de mérito genético estimadas pueden ser más exactas.

2.2.1. Matriz de Relaciones Genómicas.

Para obtener la matriz de relaciones genómicas, ó matriz ***G***, partimos de la Matriz ***MAF*** (Minor Allele Frequency matrix). Para genotipos, los elementos en esta matriz ***MAF*** son valores de 0, 1 ó 2, 0 para homocigotos alelo mayor, 1 para heterocigotos y 2 para homocigotos alelo menor, las dimensiones de la matriz ***MAF*** son número de individuos (***n***) por número de marcadores (***m***)[23] .

El siguiente paso es obtener la matriz ***M*** del resultado de ***MAF -1***, con lo que los valores de los elementos en la matriz ***M*** serian de -1 para homocigotos alelo mayor, 0 para heterocigotos y 1 para homocigotos alelo menor. También se necesita calcular una matriz ***P***, las columnas de la matriz $\boldsymbol{P}$ son las frecuencias de alelos expresadas como $\boldsymbol{P_i = 2(p_i - 0.5)}$, donde $\boldsymbol{p_i}$ es la frecuencia de alelo menor del marcador ***i***. Obteniendo las matrices ***M*** y ***P*** ya podemos calcular la matriz ***Z***, para esto sustraemos la matriz ***P*** de ***M*** (***Z = M - P***). Sustraer ***P*** de ***M*** da mayor crédito a los alelos raros que a los alelos comunes cuando calculamos relaciones genómicas[23].

Para calcular la matriz ***G*** ó matriz de relaciones genómicas utilizando marcadores SNPs existen diferentes métodos de los cuales vamos a exponer tres de estos métodos [23]. El ***primero de los métodos*** para obtener ***G*** usa la siguiente fórmula,

$$\boldsymbol{G} = \frac{\boldsymbol{ZZ'}}{\boldsymbol{2\sum p_i(1-p_i)}} \tag{5}$$

La división entre $\boldsymbol{2\sum p_i(1-p_i)}$ escala a ***G*** para ser análoga a la matriz de relaciones genéticas (matriz ***A*** del método ABLUP). Este método requiere el cálculo de la frecuencia de los alelos en la matriz de incidencia de los marcadores [23].

El ***segundo método*** para obtener ***G*** esta dado por la siguiente fórmula,

$$\boldsymbol{G} = \boldsymbol{ZDZ'} \tag{6}$$

Donde ***D*** es una diagonal con $\boldsymbol{D_{ii}} = \frac{\boldsymbol{1}}{\boldsymbol{m[2\,p_i(1-p_i)]}}$. Al igual que el primer método requiere el cálculo de la frecuencia de los alelos en la matriz de incidencia de los marcadores [23]. Esta fórmula fue propuesta para estudios genéticos en humanos .

El ***tercer método*** para obtener ***G*** no requiere la frecuencia de los alelos y en cambio se ajusta a la homocigocidad media por regresión ***MM'*** en ***A***(matriz de relaciones Genéticas) para obtener ***G*** [23] usando el modelo

$$\boldsymbol{MM'} = \boldsymbol{g_0 11'} + \boldsymbol{g_1 A} + \boldsymbol{E} \tag{7}$$

donde $\boldsymbol{g_0}$ es la intersección y $\boldsymbol{g_1}$ es la pendiente. Y para obtener ***G*** se usa la formula

$$\boldsymbol{G} = \frac{\boldsymbol{MM' - g_0 11'}}{\boldsymbol{g_1}} \tag{8}$$

La matriz ***G*** es positiva semidefinida para los 2 primeros métodos pero puede ser singular si el número de loci es limitado o si 2 individuos tienen genotipos idénticos; ***G*** debe ser singular si el numero de marcadores genéticos (SNPs) es menor al número de individuos. Gemelos idénticos o clones pueden causar singularidad aun en ***A***. Una matriz ***Gw*** mejorada y no singular se puede obtener como el ponderado (***w***), $\boldsymbol{wG + (1 - w)A}$ si el numero de marcadores es limitado y ***A*** no es singular. Donde $\boldsymbol{w = \frac{0.05^2}{\left(0.05^2 + \frac{0.125}{m}\right)}}$, ***m*** es el numero de marcadores. ***G*** debería obtener más peso que ***A*** si ***m*** > 50 y casi todo el peso (>0.99) si ***m*** > 5000 [23].

2.2.2. Estimado del Valor Genómico de Crianza (GEBV).

La solución propuesta por Henderson para calcular el valor genómico de crianza implica resolver (4) el sistema de Ecuaciones del Modelo Mixto (MME), el cual, adecuado para datos genómicos quedaría como

$$\begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}\lambda} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix} \qquad (9)$$

donde $\boldsymbol{\lambda = \sigma_e^2/\sigma_u^2}$, el factor de índice de reducción en GBLUP, $\boldsymbol{\sigma_e^2}$ es la varianza del vector de residuales y $\boldsymbol{\sigma_u^2}$ es la varianza aditiva genética [24].

También, la ecuación de índice de selección es una solución al GBLUP para predecir el valor genómico $\boldsymbol{u}$ directamente y evitar todos los pasos que implicaría resolver las Ecuaciones del Modelo Mixto. Esta ecuación nos permite obtener los Estimados del Valor Genómico de Crianza (***GEBV***, Genomic Estimated Breeding Value) en un solo paso.

Cuando derivamos la matriz $\boldsymbol{G}$ utilizando los primeros 2 métodos expuestos en el capítulo 2, utilizando las frecuencias de los alelos, la ecuación de índice de selección para obtener los $\boldsymbol{GEBV}(\widehat{\boldsymbol{u}})$ que se sugiere es,

$$\boldsymbol{GEBV}(\widehat{\boldsymbol{u}}) = \boldsymbol{G}\,[\boldsymbol{G} + \boldsymbol{R}\boldsymbol{\lambda}]^{-\mathbf{1}}\,(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}}) \tag{10}$$

donde , $\boldsymbol{X}\widehat{\boldsymbol{b}}$ (efectos fijos) es la media de las observaciones ($\boldsymbol{y}$) [24].

Otra solución, si utilizamos la matriz G derivada el tercer método descrito (de regresión), para obtener los Estimados del Valor Genómico de Crianza se sugiere utilizar una solución diferente, esta es representada con la siguiente ecuación,

$$\boldsymbol{GEBV}(\widehat{\boldsymbol{u}}) = \,[\boldsymbol{R}^{-\mathbf{1}} + \,\boldsymbol{G}^{-\mathbf{1}}\boldsymbol{\lambda}]^{-\mathbf{1}}\,\boldsymbol{R}^{-\mathbf{1}}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{b}}) \tag{11}$$

Esta solución resulta más eficiente, ya que $\boldsymbol{G}$ se invierte solo una vez y entonces rasgos adicionales con heredabilidad diferente o R pueden procesarse con iteraciones [24].

2.3. Interpretación práctica del Valor Genómico de Crianza (GEBV).

El BLUP con su naturaleza de modelo mezclas integra datos de distinta naturaleza para estimar el efecto que estos tienen, de forma simultánea y/o correlacionada, sobre un evento o medición. En el contexto de la mejora genética integra por un lado datos del pedigrí, en los cuales viene implícita la aptitud heredada de su linaje, incluyendo el grado adaptabilidad que se ha logrado a las condiciones ambientales, o el grado afectación de la misma; la genética propia del individuo, la cual es mérito individual y actúa respondiendo a las condiciones de crianza, de manejo, de salud, de medio ambiente y de presión de producción, sumado a factores aleatorios impredecibles, para lograr los valores fenotípicos que lo caracterizan.

De lo anterior que, el Valor Genómico de Crianza, medido como efectos aleatorios en el modelo de mezclas, indica como la genética propia del individuo influye en las observaciones de su rendimiento.

CAPITULO 3. IMPLEMENTACION Y PRUEBAS DEL MODELO.

Para la implementación y pruebas del método se utilizó lenguaje R. R es un lenguaje y entorno para computación estadística y generación de gráficas, es sumamente extensible [25]. Una de las fortalezas de R es la facilidad con la que se pueden diseñar gráficas con la calidad necesaria para incluirlas en una publicación científica. R está disponible como "Software Libre" bajo los términos de la "Free Software Foundation´s GNU General Public License" en forma de código fuente. Este compila y corre sobre una amplia variedad de plataformas UNIX y sistemas similares, Windows y Mac OS.

Durante este capítulo se explicará paso a paso los datos que se utilizaran, el manejo de los archivos que contienen estos datos, y como a partir de estos datos se van derivando las matrices necesarias para el cálculo de los valores genómicos, bajo ambiente R.

3.1. Datos Necesarios.

Para el cálculo de los Valores Genómicos (GEBV´s), utilizando el método GBLUP es preciso contar con los genotipos (matriz de incidencia de los marcadores de ADN, SNPs), fenotipos

(estudio longitudinal del rasgo que queremos mejorar) y la información de pedigrí de los individuos en el grupo de estudio. Los genotipos son obtenidos a partir de la genotipificación del ADN de cada individuo utilizando SNPchips tales como los de la marca Illumina [9][11][12] expuestos anteriormente. A las lecturas resultantes (marcadores SNPs) se debe de aplicar ciertos filtros para elevar la calidad de estos datos, como eliminar SNPs cuyos datos faltantes sean mayores a 10% (estos datos faltantes pueden deberse a errores de lectura en el SNPchip), eliminar SNPs que tengan una frecuencia de alelo menor <0.05, eliminar SNPs monomórficos (SNPs que solo tienen el mismo alelo para todos los individuos), entre otros. Los fenotipos y pedigrí es información la cual de manera estándar los administradores de hatos llevan registros y actualizan constantemente.

3.1.1. Matriz de incidencia de Genotipos de los marcadores (SNPs).

La matriz de genotipos, es la matriz que contiene el valor en pares de los SNPs para cada individuo en el grupo de estudio. En pares se refiere a que el genotipo de un SNP específico consta de dos alelos, de los cuales, uno fue heredado por el padre y el otro por la madre del individuo. Por lo general, esta información se maneja en archivos donde los datos en un renglón son separados por comas, y con extensión .CSV.

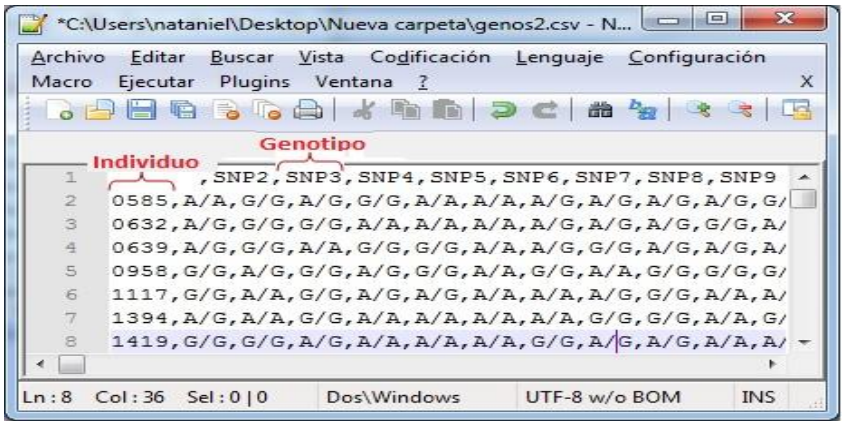


Figura 7. Tabla de genotipos vista en un editor de texto.

El primer renglón corresponde al nombre o identificación de cada uno de los SNPs, a partir del segundo renglón, como podemos observar el primer dato de cada renglón corresponde al nombre del individuo, seguidos de los genotipos, separados por "," (coma).

3.1.1.1. Matriz de Frecuencia de Alelo Menor.

La matriz ***MAF*** (Minor Allele Frequency), es una matriz donde los genotipos son representados con un numero, ya sea 0 para homocigoto alelo mayor, 1 para heterocigotos y 2 para homocigoto alelo menor, a esta codificación se le denomina contenido genético, esta información es usualmente manejada en un archivo con extensión .CSV.

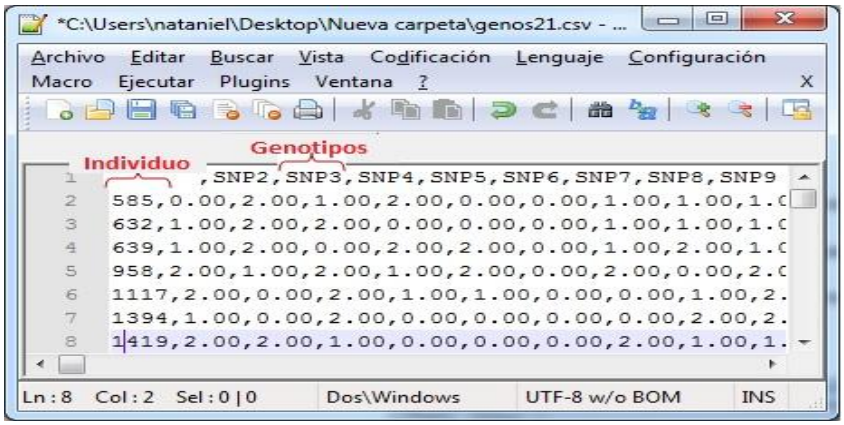
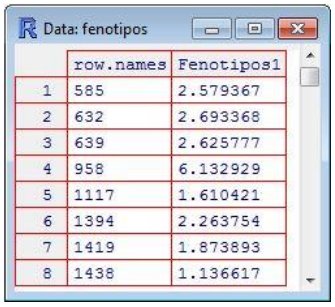


Figura 8. Tabla de genotipos vida en un editor de texto, codificada en 0, 1 y 2.

3.1.2. Fenotipos.

Como se mencionó, los fenotipos corresponden al estudio longitudinal de un rasgo en específico, o bien, mediciones continuas del rasgo en el que estamos interesados durante un periodo de tiempo, ya sea producción de leche, carne, etc.



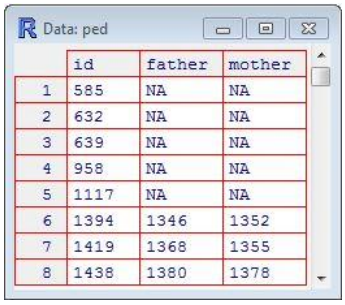
	row.names	Fenotipos1
1	585	2.579367
2	632	2.693368
3	639	2.625777
4	958	6.132929
5	1117	1.610421
6	1394	2.263754
7	1419	1.873893
8	1438	1.136617

Figura 9. Tabla de fenotipos vista en ambiente R.

Como se observa en la tabla, la primera columna corresponde al nombre del individuo y la segunda columna al valor del fenotipo. En esta tabla también se pueden incluir datos de más de un fenotipo, solo incrementando las columnas a la derecha y nombrando la columna con el número de fenotipo y/o nombre del rasgo.

3.1.3. Pedigrí.

La información del pedigrí nos ayudara a establecer una relación de parentesco entre los individuos en el grupo de estudio, a demás, a partir de esta información podemos calcular la matriz de relaciones genéticas (matriz ***A***).



	id	father	mother
1	585	NA	NA
2	632	NA	NA
3	639	NA	NA
4	958	NA	NA
5	1117	NA	NA
6	1394	1346	1352
7	1419	1368	1355
8	1438	1380	1378

Figura 10. Tabla de pedigrí vista en ambiente R.

Como se observa en la ilustración, esta tabla cuenta con 3 columnas, la primera corresponde al nombre del individuo, la segunda al nombre del su padre y la tercera al nombre de su madre. Los ***NA*** que se observan en la tabla de pedigrí de la Ilustración 5, se refiere a que el nombre del padre o de la madre, según corresponda a la columna, se desconoce.

La razón de por qué usar archivos donde los datos son separados por "," (coma), es que con este tipo de archivos se nos facilita la organización y manejo de los datos en ambiente R, a demás que este tipo de archivos son también manipulables en la hoja de cálculo de Excel.

3.2. Calculo de la Matriz de Relaciones Genómicas utilizando frecuencias de los alelos en ambiente R.

Para explicar la implementación en R del cálculo de la matriz G vamos a asumir que tenemos 4 individuos con 6 SNPs cada uno.

Matriz de Genotipos.

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Ind1	**GG**	**AA**	**AT**	**CC**	**AC**	**AG**
Ind2	**GA**	**GG**	**TT**	**CC**	**AA**	**GG**
Ind3	**AA**	**AG**	**AT**	**TT**	**CC**	**AA**
Ind4	**GG**	**AA**	**TT**	**CT**	**AA**	**GG**

Esta matriz de genotipos es convertida a contenido genético contando los alelos de menor frecuencia, y otorgando el valor correspondiente a cada genotipo (0, 1 o 2). A esta matriz se le denomina Matriz ***MAF*** (Minor Allese Frequency) de tamaño n (número de individuos) x m (número de marcadores).

Matriz *MAF*.

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Ind1	**0**	**0**	**1**	**0**	**1**	**1**
Ind2	**1**	**2**	**0**	**0**	**0**	**0**
Ind3	**2**	**1**	**1**	**2**	**2**	**2**
Ind4	**0**	**0**	**0**	**1**	**0**	**0**

Para iniciar nuestro calculo en R debemos tener el archivo de datos con los genotipos codificados como se observa en la matriz ***MAF***, abrir este archivo utilizando la funcion read.csv() (con los atributos necesarios, como el nombre y extensión del archivo), direccionando a una objeto llamado "***MAF*"** el cual contendrá el contenido del archivo,

>MAF = read.table("MAF.csv",sep=",")

En R, se denomina objeto a las entidades que este crea y manipula, estos pueden ser variables, variables indexadas (arreglos), cadenas de caracteres, funciones, etc.

El siguiente paso es obtener la matriz ***M***, para esto se resta 1 a todos los elementos de la matriz MAF generando valores de -1, 0 y 1. Para obtener la matriz ***M*** en entorno R utilizamos la siguiente instrucción,

>M = MAF - 1

Matriz *M*

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Ind1	**-1**	**-1**	**0**	**-1**	**0**	**0**
Ind2	**0**	**1**	**-1**	**-1**	**-1**	**-1**
Ind3	**1**	**0**	**0**	**1**	**1**	**1**
Ind4	**-1**	**-1**	**-1**	**0**	**-1**	**-1**

A partir de la matriz ***M*** ya estamos listos para calcular la matriz de relaciones genómicas entre cada par de individuos (matriz ***G***).

A demás podemos obtener cierta información relevante a través del desarrollo de la matriz ***M*** en ***MM'*** y ***M'M***, como a continuación describiremos. La Matriz ***MM'***, que es el resultado de la multiplicación de la matriz ***M*** por su transpuesta ***M'***.

>MMt = M %*% t(M)

Como se observa el operador ***** se refiere a multiplicación, y cuando está entre el símbolo **%** se indica que son matrices de dimensiones diferentes. La funcion t() obtiene la transpuesta de una matriz.

Matriz *MM'*

	SNP1	SNP2	SNP3	SNP4
Ind1	3	0	-2	2
Ind2	0	5	-3	2
Ind3	-2	-3	4	-3
Ind4	2	2	-3	5

Los elementos de la diagonal en esta matriz cuentan el número de SNPs homocigotos de cada individuo, y los elementos fuera de la diagonal miden el número de alelos compartidos por pariente. La matriz M'M, que es el resultado de la multiplicación de la transpuesta de M (M') por M.

>MtM= t(M)%*% M

Matriz *M'M*.

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Ind1	3	2	1	2	2	2
Ind2	2	3	0	0	0	0
Ind3	1	0	2	1	2	2
Ind4	2	0	1	3	2	2
Ind4	2	0	2	2	3	3
Ind4	2	0	2	2	3	3

En esta matriz los elementos de la diagonal cuentan el número de individuos homocigotos para cada SNP, los elementos fuera de la diagonal miden el número de veces que alelos en diferente SNP fueron heredados por el mismo individuo.

El siguiente paso es calcular la matriz P, la matriz ***P*** contiene la frecuencia de los alelos expresada como $\boldsymbol{P_i = 2(p_i - 0.5)}$, donde $\boldsymbol{p_i}$ es la frecuencia de alelo menor del **SNP** $\boldsymbol{i}$. Partimos de la Matriz ***MAF*** para calcular las frecuencias de alelo menor de cada **SNP**. Las frecuencias de alelo menor según nuestra matriz ***MAF***, serian para el **SNP1=0.375,**

SNP2=0.375, SNP3=0.25, SNP4=0.375, SNP5=0.375 y SNP6=0.375. Entonces nuestra matriz ***P*** queda de la siguiente forma.

Matriz *P*.

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Ind1	-0.25	-0.25	-0.5	-0.25	-0.25	-0.25
Ind2	-0.25	-0.25	-0.5	-0.25	-0.25	-0.25
Ind3	-0.25	-0.25	-0.5	-0.25	-0.25	-0.25
Ind4	-0.25	-0.25	-0.5	-0.25	-0.25	-0.25

El siguiente paso es calcular la matriz **Z**, la cual es el resultado de sustraer ***P*** de ***M***

>Z = M - P

Matriz *Z*.

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Ind1	-0.75	-0.75	0.5	-0.75	0.25	0.25
Ind2	0.25	1.25	-0.5	-0.75	-0.75	-0.75
Ind3	1.25	0.25	0.5	1.25	1.25	1.25
Ind4	-0.75	-0.75	-0.5	0.25	-0.75	-0.75

Para calcular la matriz G utilizando las frecuencias observadas de los alelos empleamos la formula $\boldsymbol{G} = \frac{ZZ'}{2\sum p_i(1-p_i)}$ descrita en el Capítulo 2. Donde $2\sum p_i(1-p_i)$ = 2.71875. El producto de ***Z*** por su transpuesta ***Z'*** es igual a ***ZZ'***.

>ZZt= Z %*% t(Z)

Matriz *ZZ'*.

	Ind1	Ind2	Ind3	Ind4
Ind1	2.0625	-1.1875	-1.1875	0.3125
Ind2	-1.1875	3.5625	-2.4375	0.0625
Ind3	-1.1875	-2.4375	6.5625	-2.9375
Ind4	0.3125	0.0625	-2.9375	2.5625

Por lo tango la matriz ***G***, de tamaño ***n*** x ***n***, sería

>G = ZZt / 2 $\sum \mathbf{p_i(1 - p_i)}$

	Ind1	Ind2	Ind3	Ind4
Ind1	**0.758621**	**-0.43678**	**-0.43678**	**0.114943**
Ind2	**-0.43678**	**1.310345**	**-0.89655**	**0.022989**
Ind3	**-0.43678**	**-0.89655**	**2.413793**	**-1.08046**
Ind4	**0.114943**	**0.022989**	**-1.08046**	**0.942529**

Obteniendo la matriz ***G*** utilizando el método de frecuencias observadas para calcular los Valores Genómicos se sugiere utilizar la ecuación de índice de selección (10) vista en el capítulo 2 de este trabajo, la cual es, $\boldsymbol{GEBV}(\hat{\boldsymbol{u}}) = \boldsymbol{G}\,[\boldsymbol{G} + \boldsymbol{R}\lambda]^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{b}}\right)$, donde ***R*** es una matriz identidad de tamaño ***n*** x ***n***, y $\boldsymbol{\lambda} = \boldsymbol{\sigma_e^2}/\boldsymbol{\sigma_u^2}$. $\boldsymbol{\sigma_e^2}$ y $\boldsymbol{\sigma_u^2}$ son obtenidas utilizando el paquete regress() ya implementado en R.

Como ya vimos durante el Capitulo 2, también se puede usar el método de regresión de ***MM'*** en ***A*** como se muestra en (7) y utilizando (8). Si calculamos la matriz G por el método de regresión se recomienda usar (11) [24], también vista en el Capitulo 2.

Como observamos en el ejemplo de cómo calcular la matriz G, R nos facilita el manejo de datos indexados (matrices) al poder utilizar operadores aritméticos básicos como "*", "-", "/" para la manipulación de estos datos, entonces para realizar todos los cálculos se creó un Script en lenguaje R, el cual al ejecutarlo de manera secuencial desde la consola de comandos en R, calcula los valores genómicos utilizando dos diferentes métodos para calcular la matriz G (Frecuencias de los alelos y Regresión) y como solución final usando (9), (10) y (11) para calcular los GEBV's. También se desarrolló un ambiente gráfico que sea amigable al usuario, del que se creó un archivo ejecutable con extensión .EXE, el cual corre el Script de R en modo consola de comandos de MS-DOS evitando al usuario

interactuar con la consola de comandos de R. Este ambiente grafico se le llamo GEBV.exe y cuenta con una serie de ventanas las cuales se explicara a continuación su funcionamiento.

3.3. Interfaz gráfica implementada.

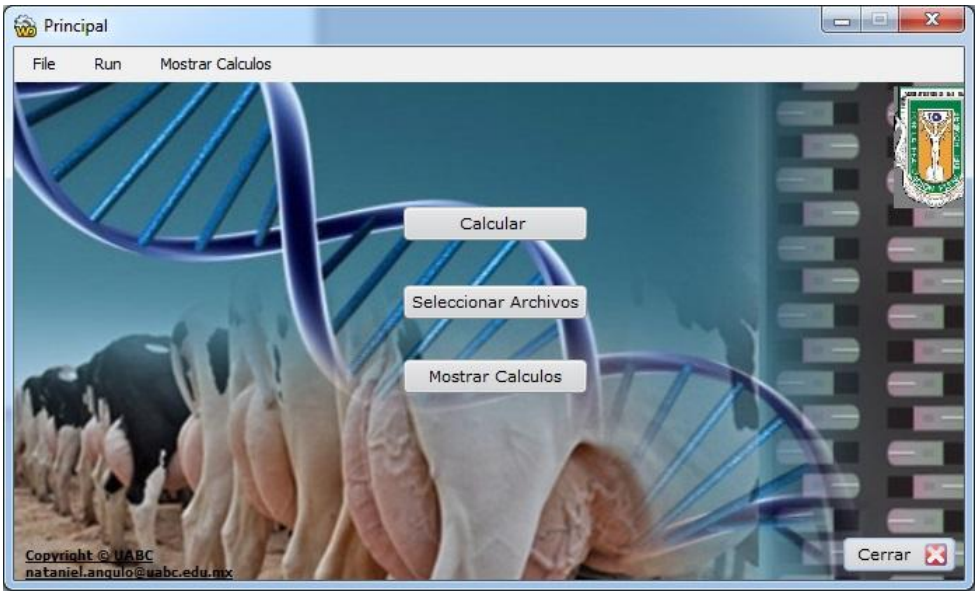


Figura 11. Ventana Principal de GEBV.exe.

La ventana Principal es el cuerpo de este sistema de ventanas, desde esta se pueden operar todas las funcionalidades de la aplicación. En el centro de esta ventana se encuentran tres botones.

Figura 12. Botones en el centro de la Ventana Principal de GEBV.exe.

3.3.1. La opción Calcular.

El botón "***Calcular***" ejecuta el Scritp en R que calcula los valores genómicos, los resultados son guardados en varios archivos con extensión .CSV. Al dar clic en el botón "***Calcular***" se despliega la siguiente ventana como se aprecia en la imagen.

Figura 13. Ventana Calculando de GEBV.exe.

La barra de progreso nos indica el avance de los cálculos.

3.3.2. Seleccionar Archivos.

Este botón nos despliega la ventana que nos permite seleccionar los archivos necesarios para calcular los GEBV's. Estos archivos son, el de genotipos, fenotipos y pedigrí. Anterior mente especifico el formato en el que deben de estar cada uno de ellos para poder utilizarlos en el Script en R. Para seleccionar los archivos se debe de dar clic en el icono al final de la barra de selección (como se muestra en el círculo rojo), al seleccionar la ubicación de los 3 archivos se debe dar clic en el botón "***Guardar Archivos***", que se encuentra en la parte inferior de la ventana "***Seleccionar Archivos***".

Figura 14. Ventana Seleccionar Archivos de GEBV.exe.

Al dar clic en el botón "***Guardar Archivos***" se desplegara un aviso con la ubicación y nombre de los archivos seleccionados.

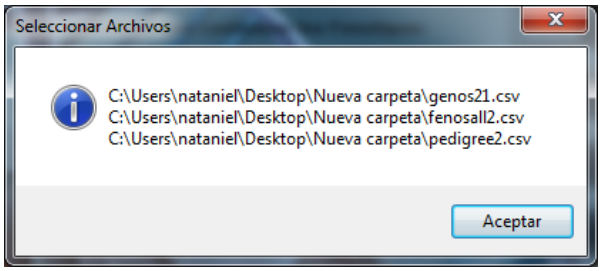


Figura 15. Sub-ventana Seleccionar Archivos de GEBV.exe.

Al dar clic en aceptar se termina el proceso de seleccionar los archivos y automáticamente nos envía a la ventana "***Principal***".

3.3.3. Mostrar Cálculos.

Esta ventana muestra los cálculos obtenidos. Los cuales son desplegados en tablas como se muestra en la imagen.

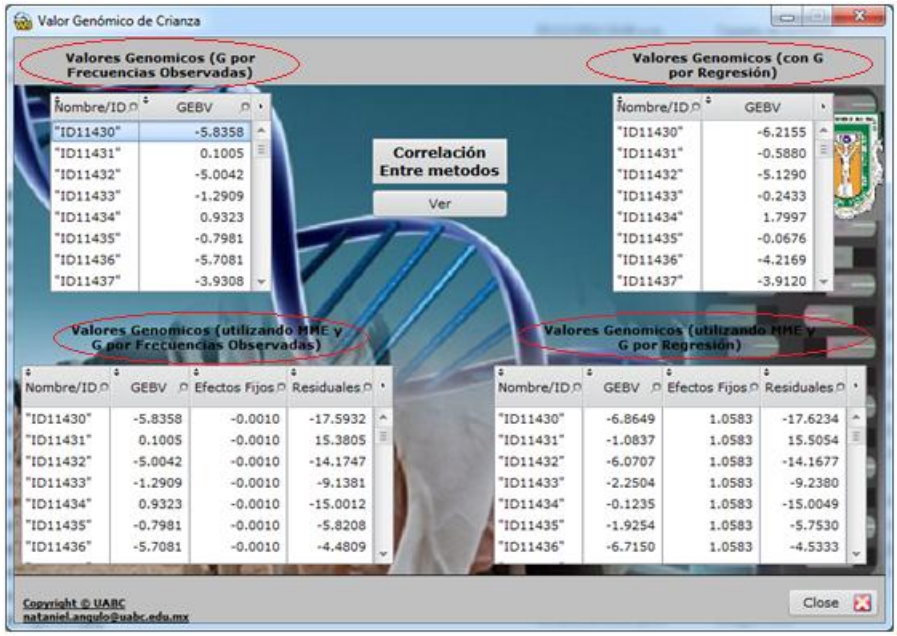


Figura 16. Ventana Valores Genómicos de Crianza de GEBV.exe.

En el encabezado de cada tabla se indica el método por el cual fueron calculados los valores genómicos. Cada una de las tablas cuenta con una columna de "***Nombre/ID***" seguida de otra llamada "***GEBV***", las cuales corresponden a la identificación y valor genómico respectivamente de cada animal. En las dos tablas que se encuentran en la parte inferior de esta ventana para calcular el valor genómico se utilizaron las ecuaciones del modelo mixto (MME), a través de este método podemos obtener conjuntamente el vector de efectos fijos ($\widehat{\boldsymbol{b}}$) y el vector de residuales (***e***).

Dentro de esta ventana podemos exportar los resultados de los cálculos, para ellos basta hacer un clic en la esquina superior derecha de la tabla que queremos exportar los datos

como se muestra en la imagen, se despliega un menú Pop-up, el cual nos muestra diferentes opciones para este propósito.

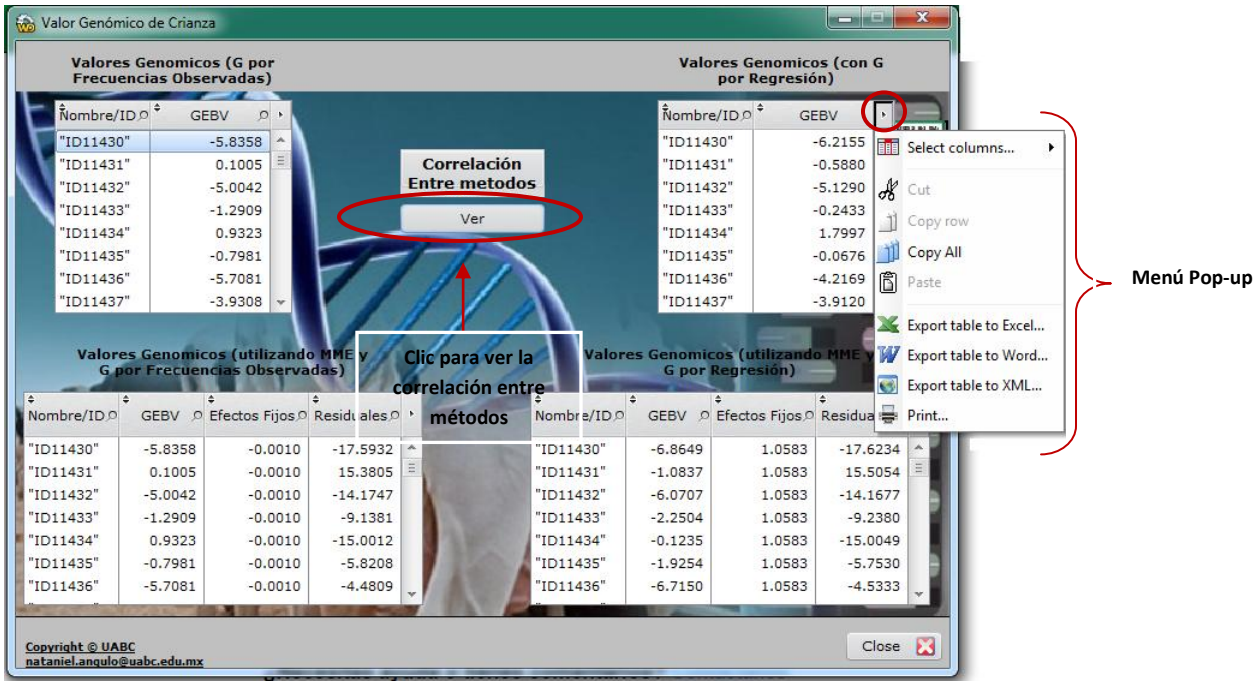


Figura 17. Descripción de ventana Valores Genómicos de Crianza de GEBV.exe.

En esta ventana se encuentra un botón llamado "***Correlación Entre Métodos***", la cual nos abre una sub-ventana con las gráficas de correlación entre todos los métodos como se muestra a en la siguiente imagen.

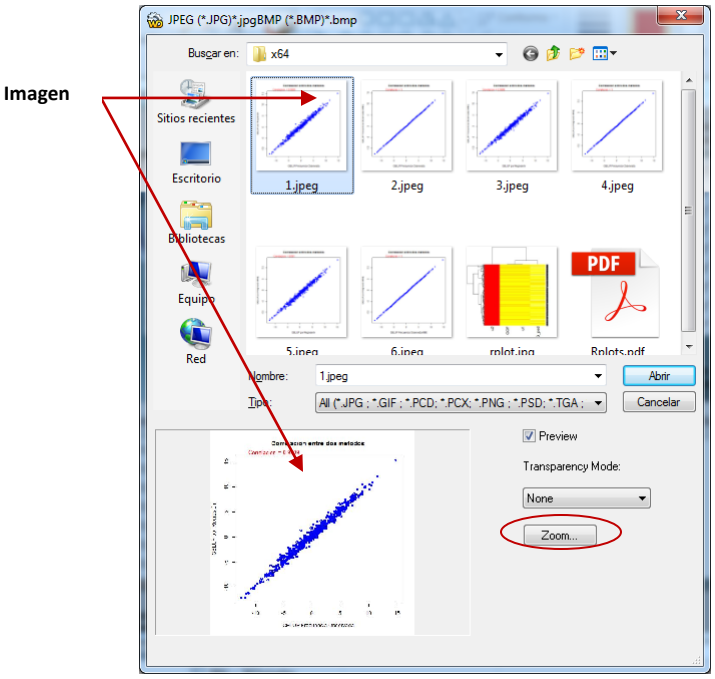


Figura 18. Ventana para mostrar Graficas de GEBV.exe.

Al dar un clic sobre la imagen seleccionada la podremos ver en modo "***Preview***", para extender la imagen debemos dar clic en el botón "***Zoom***".

La ventana "***Principal***" cuenta con una barra de menú en su parte superior.

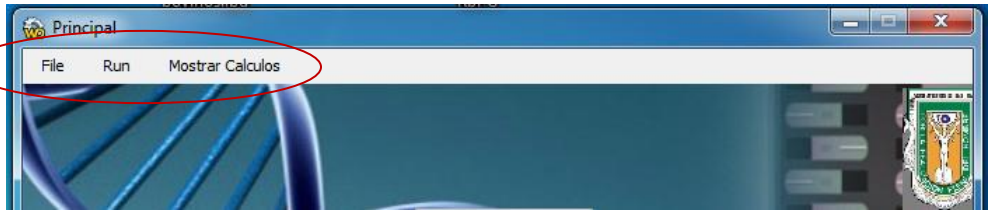


Figura 19. Barra de Menú en la Ventana Principal de Crianza de GEBV.exe.

En el submenú File podemos encontrar, dos opciones, "***Seleccionar Archivo***" y "***Cerrar***". "***Seleccionar Archivo***" nos lleva a la ventana de donde seleccionamos los tres archivos

necesarios para realizar los cálculos de los valores genómicos, y la opción "***Cerrar***" sirve para cerrar por completo la aplicación.

En la barra de menú superior encontramos las opciones "***Run***" y "***Mostrar Cálculos***". Ambas opciones no cuentan con submenú. "***Run***" abre la ventana del cálculo de los valores genómicos y "***Mostrar Cálculos***" nos despliega el resultado de los cálculos una vez realizados.

Esta aplicación se compilo para sistema operativo Windows, y se creó un archivo de instalación llamado "***Instalador_GEBV.EXE***". Al dar doble clic sobre el archivo "***Instalador_GEBV.EXE***" nos abrirá el asistente de instalación el cual nos guiara durante el proceso de instalación. Como se muestra en la imagen.

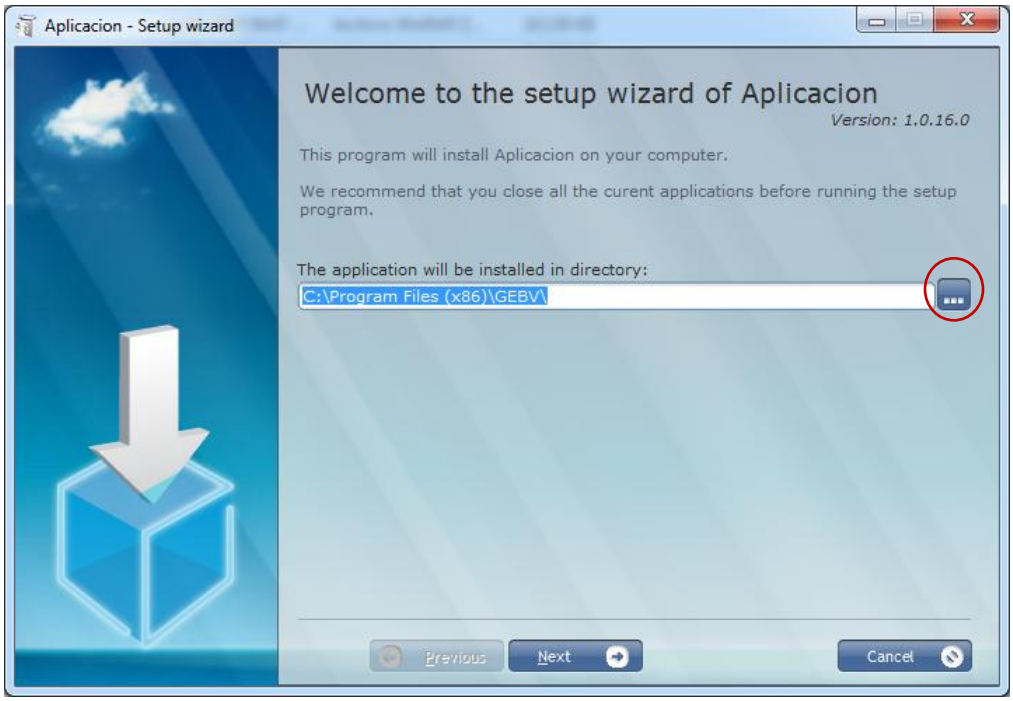


Figura 20. Ventana del Asistente para la instalación de GEBV.exe.

El asistente instalara el programa en "***C:\Program Files (x86)\GEBV***", como se muestra en la siguiente imagen, también es posible dar una ubicación de instalación manualmente solo dando clic en el icono señalado en el círculo. A su vez, el asistente de instalación creara en el Escritorio de Windows los iconos  de nombre "***GEBV***" que dan acceso a la aplicación.

CAPITULO 4. PRUEBAS DEL MODELO Y ANÁLISIS DE RESULTADOS.

Para las pruebas del método implementado se utilizó un conjunto de datos que contiene genotipos, fenotipos y pedigrí de 500 toros. En este conjunto de datos cada individuo está etiquetado con una identificación única. Los Genotipos y el pedigrí fueron tomados de un conjunto de datos de ganado real, mientras que los fenotipos son hipotéticos. Los datos genotípicos consisten en 7250 marcadores SNPs para cada individuo en el grupo [26]. El conjunto de datos se manejo en tres archivos donde los datos fueron separados con "**,**" y con extensión .***CSV***, llamados "***genotipos.csv***", "***fenotipo.csv***" y "***pedigrí.csv***". Se utilizó el Scritp en R para hacer las diferentes corridas y calcular los valores genómicos de los 500 individuos para un fenotipo, utilizando matrices G derivadas de frecuencias alélicas de los marcadores SNPs y por método de regresión, y aplicando las diferentes soluciones expuestas en el Capítulo 2. Los resultados de los cálculos fueron llamados GEBVa (Matriz G por Frecuencias Alélicas y solución con 10) GEBVb (Matriz G por regresión y solución con 11), GEBVc (Matriz G por Frecuencias Alélicas y solución con 9), adicionalmente, se calcularon valores genómicos utilizando la matriz A (relaciones genéticas derivadas solo de

pedigrí)a estos resultados se le llamo A. Se hicieron comparaciones entre los resultados de diferentes métodos.

Tabla 3. Lista parcial de Valores Genómicos calculados.

Nombre/ID	GEBVa	GEBVb	GEBVc	A
ID11430	-5.8358	-6.2155	-5.8358	-11.3695
ID11431	0.1005	-0.5880	0.1005	0.6604
ID11432	-5.0042	-5.1290	-5.0042	-6.5246
ID11433	-1.2909	-0.2433	-1.2909	-4.5475
ID11434	0.9323	1.7997	0.9323	-2.9489
ID11435	-0.7981	-0.0676	-0.7981	-0.5871
ID11436	-5.7081	-4.2169	-5.7081	-3.6692
ID11437	-3.9308	-3.9120	-3.9308	-7.0743
ID11438	-7.8363	-8.5829	-7.8363	-8.1255
ID11439	5.4315	5.9581	5.4315	5.8943
ID11440	-3.2646	-3.1041	-3.2646	-7.8390
ID11441	-8.2524	-7.5188	-8.2524	-10.7945

Con los vectores de los cuatro resultados de valores genómicos se creó una tabla de correlación obteniendo los siguientes resultados.

Tabla 4. Correlación entre los resultados de los diferentes métodos.

	GEBVa	GEBVb	GEBVc	A
GEBVa	1.0000	0.9909	1.0000	0.9110
GEBVb	0.9909	1.0000	0.9909	0.8978
GEBVc	1.0000	0.9909	1.0000	0.9110
A	0.9110	0.8978	0.9110	1.0000

Los resultados de la correlación entre los métodos utilizados nos indica que los cálculos de GEBVa, GEBVb y GEBVc son casi idénticos, ya que se obtuvieron correlaciones de entre 0.99 y 1.00, en estos tres métodos se utilizaron datos genómicos. Los resultados obtenidos

en A se acercan a los obtenidos con los otros 3 métodos, solo que la matriz de relaciones utilizada en A se derivó solamente de los datos de parentesco (pedigrí), lo cual sería la razón de que A no logra capturar efectivamente la relación a nivel genómico entre todos los individuos del grupo.

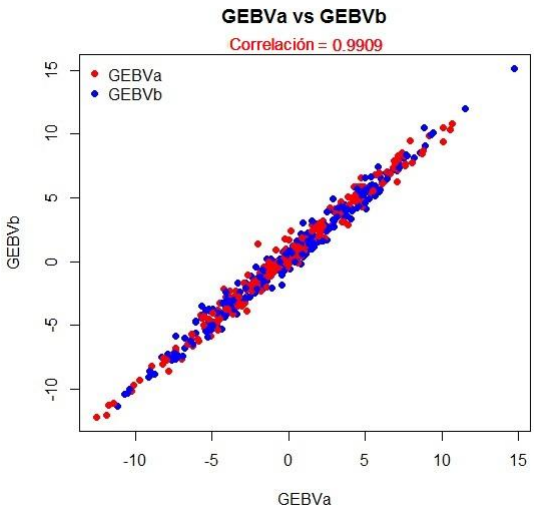


Figura 21. GEBVa vs GEBVb.

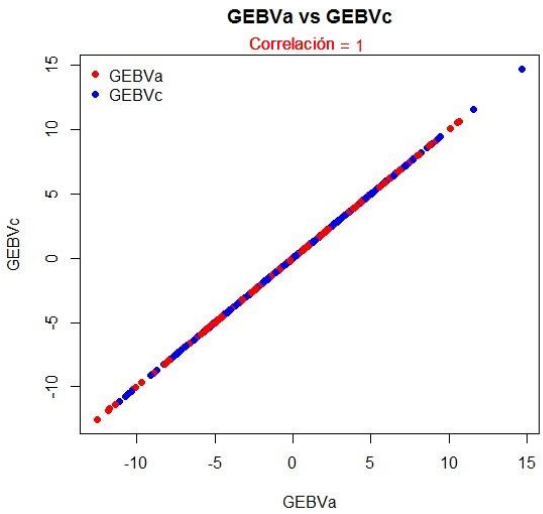


Figura 22. GEBVa vs GEBVc.

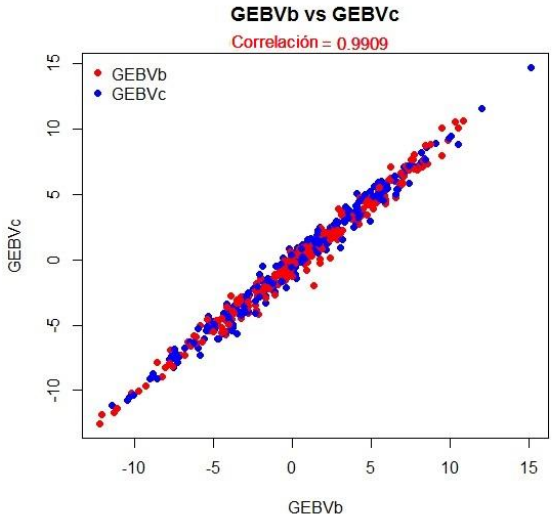


Figura 23. GEBVb vs GEBVc.

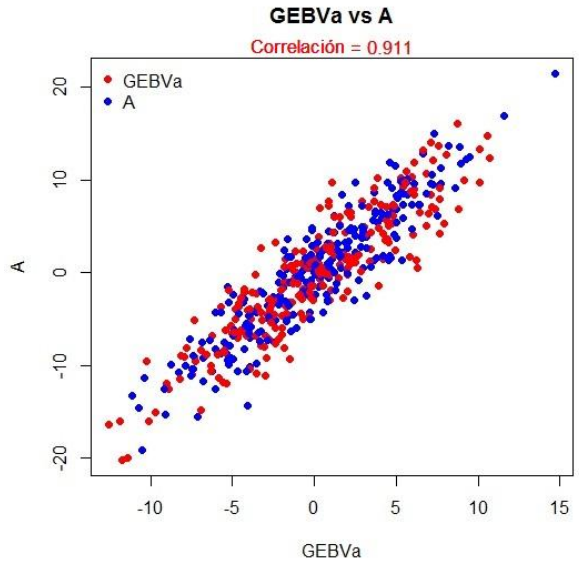


Figura 24. GEBVa vs A.

Las relaciones genómicas derivadas de datos de marcadores proporcionan una estimación más exacta de covarianzas genéticas entre parientes, que a su vez nos lleva a la obtención de predicciones más exactas. A continuación encontramos las matrices de Relaciones Genómicas (G, por frecuencias de los alelos y regresión) y matriz de relaciones genéticas (A, derivada del pedigrí), se puede observar que la matriz G es más precisa al encontrar relación entre los 500 individuos que la matriz A, esto debido a la utilización de datos genómicos, por tal motivo, con G podemos encontrar relación entre individuos aun sin tener datos de su parentesco.

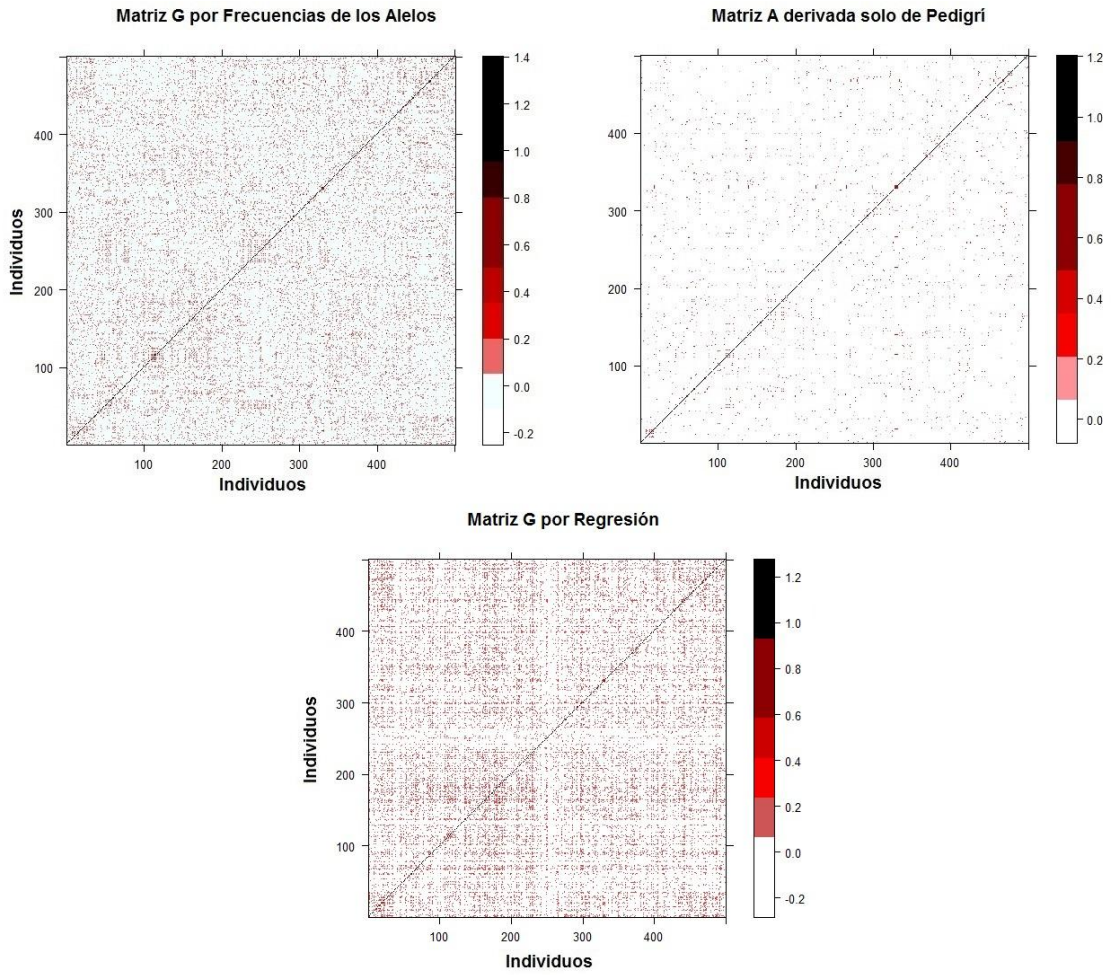


Figura 25. Se observa la mayor precisión de las matrices G para calcular la relación entre individuos.

En el modelo lineal de mezclas el GEBV($\widehat{\boldsymbol{u}}$) se refiere a los efectos aleatorios ocasionados por la incidencia de los marcadores genéticos en cada individuo. A continuación presentamos la correlación entre GEBVa y GEBVb contra las observaciones del fenotipo utilizado en los cálculos.

Tabla 5. Correlación GEBVa vs Fenotipo y GEBVb vs Fenotipo.

	Fenotipo
GEBVa	0.8354
GEBVb	0.8226

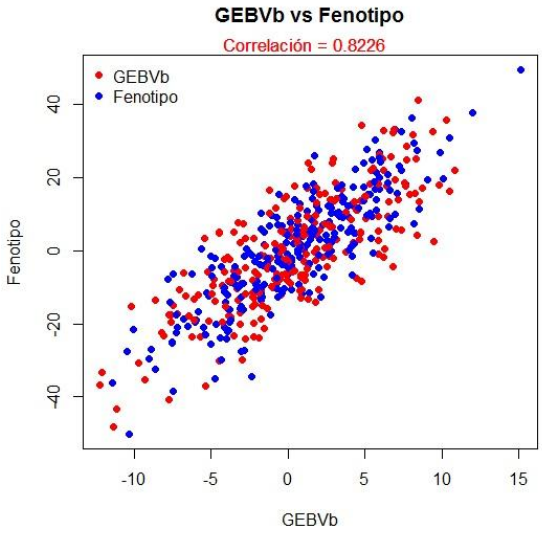


Figura 26. GEBVa vs Fenotipo.

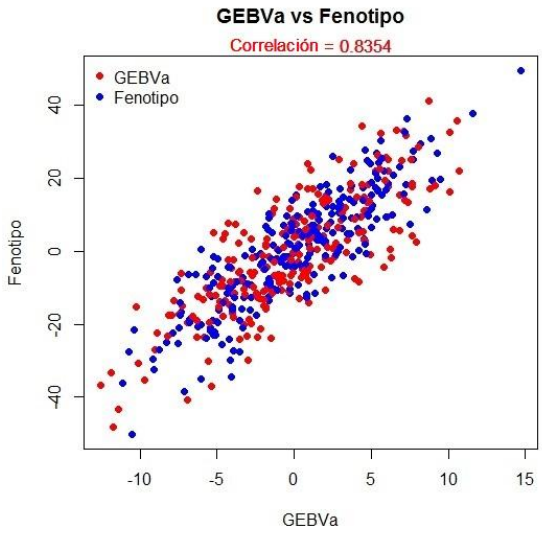


Figura 27. GEBVb vs Fenotipo**.**

Se puede observar como los Valores genómicos (Efectos aleatorios) y los fenotipos observados (mediciones) presentan una correlación positiva de 0.8354 (GEBVa) y 0.8226 (GEBVb). Esto es un indicador de como los efectos aleatorios (genotipos) afectan a la expresión del fenotipo de interés en cada uno de los individuos.

Los cálculos obtenidos con el Scritp presentado en este trabajo de investigación también fueron comparados con resultados obtenidos con paquetes implementados en R, synbreed [26] y rrBLUP[27] obteniendo los mismos valores genómicos. Estas comparaciones se hicieron utilizando los mismos datos que en el Script desarrollado. Para comparar con synbreed se utilizó una función dentro de este paquete llamada gpMod() y el cual a su vez calcula una matriz de relaciones genómicas basada en frecuencia de los alelos y el modelo GBLUP. En rrBLUP se encuentra la funcion kin.blup(), la cual calcula

valores genómicos por el método GBLUP utilizando una matriz de relaciones genómicas basa en frecuencias de los alelos.

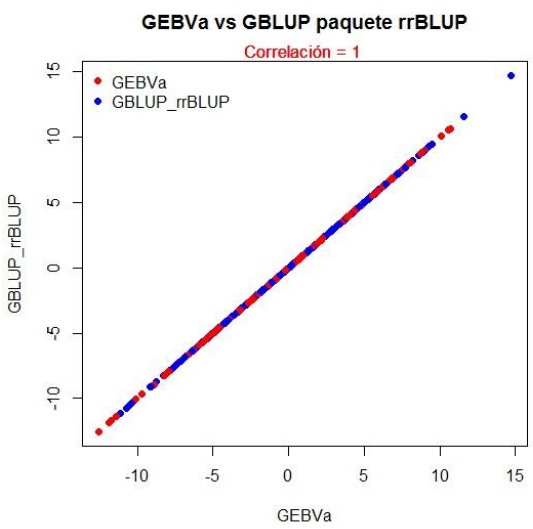


Figura 28.GEBVa vs Paquete rrBLUP.

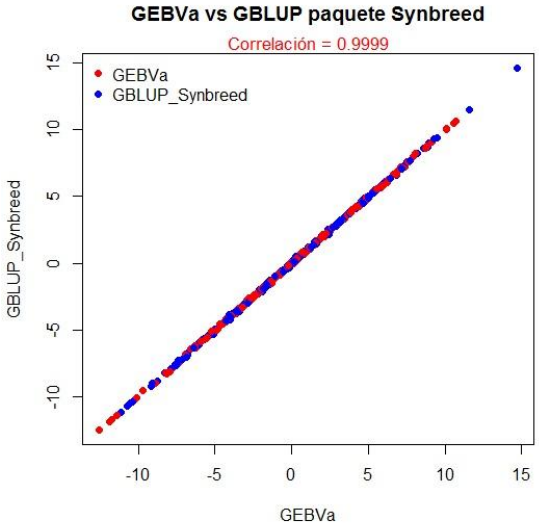


Figura 29. GEBVa vs Paquete Synbreed**.**

La ejecución del Script en R para el cálculo de los valores genómicos utilizando el set de datos presentado al inicio de este capítulo se hizo sobre una PC tipo laptop con sistema operativo Windows 7, procesador Intel® Core™ i3 2.5 Ghz, 8 Gb de memoria RAM, con un tiempo de ejecución de aproximadamente 180 segundos.

CAPITULO 5. CONCLUSIONES Y TRABAJO FUTURO

5.1. Conclusiones.

Calcular valores genómicos utilizando el método GBLUP es prácticamente eficaz y computacionalmente viable para un conjunto de datos como el utilizado (500 individuos x 7250 marcadores SNP), a demás de la facilidad que nos permite el integrar los datos necesarios al modelo. Otro punto a destacar es que los métodos para construir la matriz G a utilizar en GBLUP no son complicados, y podemos integrar datos genómicos y de pedigrí en ésta (Matriz G por Método de Regresión en A). De este modo, se desarrolló una herramienta estadística basada en el método GBLUP para estimar valores genómicos de ganado bovino utilizando marcadores SNP. El objetivo de estimar los valores genómicos en bovino es el de implementar un programa de mejora genética de ganado bovino en hatos del municipio de Mexicali, Baja California.

5.2. Trabajo futuro.

Como trabajo futuro, se desarrollará un software de administración ganadera, en el cual se manejaran todos los datos referentes a un animal en un hato, tales como registros de producción, pedigrí, salud, etc., y adicionalmente integrar datos genotípicos, a demás como parte innovadora en este tipo de software, se integrara la herramienta estadística presentada en este trabajo de investigación, para calcular valores genómicos utilizando el método GBLUP.

GLOSARIO:

Haplotipos: Es un conjunto de polimorfismo de nucleótido simple (SNPs) en un cromosoma particular que están estadísticamente asociados.

Locus: Un locus (en latín, lugar; el plural es loci) es una posición fija en un cromosoma, como la posición de un gen o de un marcador (marcador genético). Una variante de la secuencia del ADN en un determinado locus se llama alelo.

Loci: Plural de Locus.

Modelos de Marcov: Es un tipo especial de proceso estocástico discreto en el que la probabilidad de que ocurra un evento depende del evento inmediatamente anterior.

Método Monte Carlo: Es un método no determinístico o estadístico numérico, usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud. El método se llamó así en referencia al Casino de Monte Carlo (Principado de Mónaco) por ser "la capital del juego de azar", al ser la ruleta un generador simple de números aleatorios.

Máxima Verosimilitud: Es un método habitual para ajustar un modelo y encontrar sus parámetros.

BIBLIOGRAFÍA

[1] Janeth Ortega T. y Luís García P (2011). The bovine genome, methods and results of its analysis. *MVZ Córdoba*, *16*(1), 2410-2424.

[2] Hayes, B.J., Lewin, H.A., Goddard, M.E. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Gentics, 29*(4), 206-214.

[3] Genómica. (s.f.). En Wikipedia. Recuperado el 18 de Mayo de 2015 de http://es.wikipedia.org/wiki/Gen%C3%B3mica.

[4] Martínez, S.E. (12 de Junio de 2012). Tendrá Baja California ganado mejorado genéticamente. LA CRONICA.COM. Recuperado de http://www.lacronica.com/EdicionEnlinea/Notas/Noticias/12062014/852281-Tendra-BC-ganado-mejorado-geneticamente.html.

[5] Michel A. Wattiaux. Dairy Essentials , Babcock Institute: for international dairy research and development (www.babcock.wisc.edu/node/121).

[6] Genetics Home Reference. A service of the U.S. National Library of Medicine. (18 de Mayo 2015). Marker. Recuperado de http://ghr.nlm.nih.gov/glossary=marker.

[7] T. H. E. Meuwissen, B. J. Hayes† and M. E. Goddard., (2001), "Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps". Genetics 157: 1819–1829.

[8] Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS., (2008), "SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries". *Nat Methods*, 5(3):247-252.

[9] Illumina (2012): BovineSNP50 Genotyping BeadChip. Illumina, Inc.

[10] Illumina (2011). GoldenGate Bovine 3K Genotyping BeadChip. Illumina, Inc.

[11] Illumina (2013): BovineLD v1.1 Genotyping BeadChip. Illumina, Inc.

[12] Illumina (2012): BovineHD Genotyping BeadChip. Illumina, Inc.

[13] Affymetrix (2011): Axiom Genome-Wide Bos 1 Array Plate.

[14] The Bovine HapMap Consortium., (2009), "Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds". *Science*, 324(5926):528-532.

[15] Henderson, I. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics,31*(2), 423-447.

[16] Kärkkäinen, H.P., Sillanpää, M.J. (2012). Back to basics for Bayesian model building in genomic selection. *Genetics, 191*(3), 969-87.

[17] Takeshi. H, and Hiroyoshi. I. (2010). EM algorithm for Bayesian estimation of genomic breeding values. *BMC Genetics*, 11:3.

[18] Protocolo. (s.f.). Caracterización de variables de producción, de salud y genéticas en un hato lechero para la implementación de un programa de mejora genética basada en marcadores. Universidad Autónoma de Baja California, Mexicali, Baja California. México.

[19] Salomón-Torres, R. (2014). Algoritmos Bioinformáticos para la Detección de Variaciones Estructurales en el Genoma Completo del Ganado Bovino, Utilizando SNPs de Alta Densidad. Universidad Autónoma de Baja California, Mexicali, Baja California, México.

[20] Arango-Pérez, M.L. (2013). Caracterización del Modelo Experimental de Producción Lechera del Instituto de Investigaciones Veterinarias de la UABC para la implementación de un programa de mejora genética. Universidad Autónoma de Baja California, Mexicali, Baja California, México.

[21] West, B., Welch, K., Galecki, A. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software.* Boca Raton, FL: Taylor & Francis Group, LLC.

[22] Robinson, G.K. (1991).That BLUP is a Good Thing: The Estimation of Random Effects. Statiscal Science, 6(1), 15-51.

[23] VanRaden, P.M. (2008). Efficient Methods to Compute Genomic Predictions. J. Dairy Sci., 91(11), 4414-4423.

[24] Zapata-Valenzuela, J., Whetten R. W., Neal D., Mckeand S. y Isik F. (2013).Genomic Estimated Breeding Values Using Genomic Relationship Matrices in a Cloned Population of Loblolly Pine. *G3 Journal, 3*(5), 909-916.

[25] R Project. (s.f.). What is R?. Recuperado el 4 de Mayo de 2015 de: http://www.r-project.org/about.html.

[26] Wimmer, V., Albrecht, T., Auinger, H. J., Schon, C. C. (2012). synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics, 28(15)*, 2086–2087.

[27] Endelman, J.B. (2011).Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome, 4*(3), 250–255.