

**Estimación paramétrica de segmentos de identidad por descendencia
mediante una distribución mezcla de Gammas
considerando interferencia meiótica**

*Tesis presentada para optar al título de Magister de la Universidad de Buenos Aires,
Área Biometría y Mejoramiento*

Roberto Ramses Saavedra Flores
Ingeniero Agrónomo Zootecnista - Universidad de Panamá - 2016

Lugar de trabajo: Universidad de Buenos Aires
Facultad de Agronomía – Cátedra de Mejoramiento Genético Animal



Escuela para Graduados Ing. Agr. Alberto Soriano
Facultad de Agronomía – Facultad de Ciencias Veterinarias
Universidad de Buenos Aires

COMITÉ CONSEJERO

Director de Tesis

Rodolfo Juan Carlos Cantet

Ing. Agr. (Universidad de Buenos Aires, Argentina)

M.Sc (Montana State University, EE. UU.)

M.Sc (University of Illinois, EE. UU.)

Ph.D. (University of Illinois, EE. UU.)

Co-director de Tesis

Natalia Soledad Forneris

Ing. Agr. (Universidad de Buenos Aires, Argentina)

Doctora (Universidad de Buenos Aires, Argentina)

JURADOS DE TESIS

JURADO

Luis Gómez-Raya

Lic. Biología (Universidad Complutense de Madrid, España)

M.Sc (Universidad Complutense de Madrid, España)

Ph.D (University of Guelph, Canadá)

JURADO

Juan Pedro Lirón

Lic. Genética (Universidad Nacional de Misiones, Argentina)

Doctor (Universidad Nacional de La Plata, Argentina)

JURADO

Carlos Alberto Martínez Niño

Zoot. (Universidad Nacional de Colombia, Colombia)

M.Sc (Universidad Nacional de Colombia, Colombia)

M.Sc (University of Florida, EE. UU.)

Ph.D (University of Florida, EE. UU.)

Fecha de defensa de tesis: 21 de diciembre de 2021

DEDICATORIA

A toda mi familia

AGRADECIMIENTOS

Primero que todo dar gracias al Estado Panameño, a través de las instituciones IFARHU y SENACYT de financiar mis estudios a través de una beca.

Luego dar gracias al Dr. Rodolfo Cantet por haber aceptado dirigir mi tesis, le estaré siempre agradecido de hacer que pudiese viajar a Buenos Aires a cursar estudios de maestría, además de sus muchos consejos y paciencia. De igual modo a la Dra, Natalia Forneris de haber sido mi codirectora de tesis y de sus muchas sugerencias y consejos.

A los jurados, por su tiempo dedicado a la lectura de esta tesis y sus comentarios oportunos y constructivos.

A la Catedra de Mejoramiento Animal de la FAUBA, específicamente Sebas, Caro, Andres, Dani, Belcy, Esteban, Matías, María José y Martín, le agradezco su generosa y valiosa ayuda.

A todos los profesores que tuve a lo largo de la carrera, por sus enseñanzas, así como al personal de la Escuela para Graduados por su disposición y atención.

A todos los compañeros que tuve a lo largo del cursado, por los gratos momentos compartidos.

A Argentina, por acogerme y ser generosa conmigo, así como lo ha sido con millones de personas a lo largo de la historia, así como a todas las personas que conocí allí.

DECLARACIÓN

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en esta u otra institución.

Roberto Ramses Saavedra Flores

INDICE GENERAL

| | |
|---|------|
| DEDICATORIA | iii |
| AGRADECIMIENTOS | iv |
| DECLARACIÓN | v |
| ÍNDICE GENERAL | vi |
| ÍNDICE DE CUADROS | ix |
| ÍNDICE DE FIGURAS | x |
| ABREVIATURAS | xii |
| RESUMEN | xiii |
| ABSTRACT | xiv |
| | |
| CAPITULO 1. Introducción. | 1 |
| | |
| CAPÍTULO 2. Teoría sobre la fracción de genoma compartido IBD entre Individuos. | 9 |
| 2.1. Medidas de identidad por descendencia para IBD discreto. | 11 |
| 2.1.1 Varianza de la IBD con información genealógica. | 13 |
| 2.2. Medidas de identidad por descendencia para IBD continuo. | 17 |
| 2.3. Estimación de la longitud de los segmentos compartidos entre abuelo y nieto. | 21 |
| 2.3.1. Identificación de los eventos de recombinación. | 23 |
| 2.4. El proceso de interferencia durante la meiosis | 24 |
| | |
| CAPÍTULO 3. Materiales y Métodos. | 30 |
| 3.1. Base de datos analizada. | 31 |
| 3.2. La fracción del genoma del abuelo de referencia presente en sus nietos | 32 |
| 3.2.1. Esperanza y Varianza de la fracción de genoma compartido IBDP y de β^* | 33 |
| 3.3. Estimación de los parámetros de la distribución de IBD compartido. | 36 |
| 3.4. Estimación del parámetro de interferencia (v). | 37 |

| | |
|---|----|
| 3.5. Prueba del cociente de verosimilitud. | 40 |
| CAPÍTULO 4. Resultados | 42 |
| 4.1. Distribución del número de recombinaciones por cromosoma. | 43 |
| 4.2. Parámetros de interferencia estimados bajo el proceso de renovación empleando una distribución Gamma. | 46 |
| 4.3. Presencia de Interferencia en el proceso de recombinación | 48 |
| 4.4. Probabilidad de ausencia de recombinación | 51 |
| 4.5. Proporción de Genoma IBD compartido entre abuelos y nietos | 54 |
| 4.5.1. Valores Esperados y Varianzas de la fracción de genoma compartido IBD y β^* | 55 |
| CAPÍTULO 5. Discusión. | 60 |
| 5.1. Estimación de los parámetros de interferencia bajo una distribución Gamma. | 61 |
| 5.2. Probabilidad de ausencia de recombinación. | 62 |
| 5.3. Proporción de Genoma IBD compartido (IBDP) para abuelo – nieto | 63 |
| 5.4. Implicancias de la varianza del IBDP | 64 |
| CAPÍTULO 6. Conclusiones | 69 |
| BIBLIOGRAFÍA | 70 |
| APÉNDICE A | 74 |

ÍNDICE DE CUADROS

| | |
|---|-----------|
| Cuadro 1: Frecuencia de cromosomas parentales y recombinantes. | 44 |
| Cuadro 2. Estimación del parámetro de forma (v) y la tasa de cambio (τ) de distribuciones Gamma, junto con su desvío estándar respectivo en fase paterna. | 47 |
| Cuadro 3. Estimación del parámetro de forma (v) y la tasa de cambio (τ) de distribuciones Gamma, junto con su desvío estándar respectivo en fase materna. | 47 |
| Cuadro 4. Determinación del nivel de interferencia en los cromosomas en fases paternas. | 50 |
| Cuadro 5. Determinación del nivel de interferencia en los cromosomas en fases maternas.. | 51 |
| Cuadro 6. Probabilidad de ausencia de recombinación por cromosoma..... | 52 |
| Cuadro 7. Valores de las medias (\bar{X}), medianas (\tilde{X}), desvíos estándares (DE), coeficientes de variación (CV), asimetría, curtosis y valores máximos y mínimos esperadas en las IBDP abuelo(a) – nieto(a)..... | 54 |
| Cuadro 8. Valores de los parámetros de forma (v) y la tasa de cambio (τ) de una distribución gamma estimados por ML, tanto en fase materna y paterna..... | 56 |
| Cuadro 9. Valores de las medias (\bar{X}), medianas (\tilde{X}), desvíos estándares (DE), coeficientes de variación (CV), asimetría, curtosis y valores máximos y mínimos esperadas en valores medios de las IBDP abuelo(a) – nieto(a)..... | 57 |
| Cuadro 10. Valores de las medias (\bar{X}), medianas (\tilde{X}), desvíos estándares (DE), coeficientes de variación (CV), asimetría, curtosis y valores máximos y mínimos esperadas en varianzas de las IBDP abuelo(a) – nieto(a)..... | 57 |

ÍNDICE DE FIGURAS

| | |
|---|-----------|
| Figura 1. Estados de Identidad IBD en un locus (Jacquard, 1972). | 13 |
| Figura 2. Identificación de genotipos en fase de individuo con un abuelo homocigota y otro abuelo heterocigoto. | 22 |
| Figura 3: Identificación de genotipos en fase de individuos con abuelos homocigotos con genotipo opuesto | 23 |
| Figura 4. Distribución de las recombinaciones y los segmentos IBD dentro de una gameta. C_1, C_2, C_3 denotan los eventos de recombinación, mientras que l_1, l_2, l_3, l_4 son los segmentos IBD dentro de esa gameta. | 26 |
| Figura 5. Número de recombinaciones promedio por cromosoma, tanto en fase paterna y materna. | 44 |
| Figura 6. Comparación de las distancias genéticas obtenidas por cromosoma y sexo. Se comparan los datos analizados en este trabajo (Genus – Pic) con respecto a valores de dos base de datos distantes presentados en Tortereau <i>et al.</i> 2012 | |
| Figura 7. Distribuciones de los números de recombinaciones encontradas por cromosoma. | 45 |
| Figura 8. Estimaciones de los parámetros de forma por cromosoma. La línea horizontal representa al parámetro de forma cuando hay ausencia de interferencia ($v = 1$). | 48 |
| Figura 9. Probabilidades de ausencia de recombinación para cada cromosoma en cada asociada con su distancia física en Mbp..... | 53 |
| Figura 10. Comparación de las proporciones de gametas no recombinantes con respecto a las probabilidades de no recombinación estimadas. | 53 |
| Figura 11. Distribución de las IBDP observadas (0.5 PG) bajo el método de Cheung (2007) para abuelo – nieto en fase paternas (a) y maternas (b)..... | 55 |
| Figura 12. Distribución de las proporciones de genoma IBD compartido estimadas para fases paternas (a) y maternas (b). | 57 |

| | |
|--|-----------|
| Figura 13. Distribución de las varianzas de las proporciones de genoma IBD compartido estimadas para fases paternas (a) y maternas (b). | 58 |
| Figura 14. Comparación entre las proporciones observadas de IBDP, respecto de las IBDP esperadas abuelo – nieto (a) y abuela – nieto (b)..... | 59 |

ABREVIATURAS

| | |
|-----------|--|
| AR | Regresión Ancestral (del inglés <i>Ancestral Regression</i>). |
| ADN | Ácido Desoxirribonucleico |
| BLUP | Predicción lineal insesgada de mínima varianza (del inglés <i>Best Linear Unbiased Prediction</i>). |
| CV | Coefficiente de Variación. |
| DE | Desvío Estándar |
| IBD | Identidad por Descendencia (del inglés <i>Identity by Descent</i>) |
| IBDP | Proporción de Genoma Compartido IBD (del inglés <i>IBD Genome Shared Porportion</i>). |
| E(IBDP) | Valor Esperado de Genoma Compartido IBD |
| Var(IBDP) | Varianza del Genoma Compartido IBD |
| IBS | Identidad en Estado (del inglés <i>Identity by State</i>). |
| ML | Máxima Verosimilitud |
| SNP | Polimorfismo genético de un solo nucleótido (del inglés <i>Single Nucleotide Polymorphysm</i>). |

RESUMEN

Estimación paramétrica de segmentos de identidad por descendencia mediante una distribución mezcla de Gammas considerando interferencia meiótica.

La interferencia consiste en la ausencia de independencia de los eventos de entrecruzamiento (crossover) de las diferentes cromátidas de un cromosoma. La longitud de los segmentos del genoma idénticos por descendencia compartidos por un par de individuos emparentados (IBDP) es una variable aleatoria que sigue una distribución mezcla de Gammas, bajo un proceso estocástico de renovación que incorpora la interferencia. Bajo este proceso, existe interferencia cuando el parámetro de forma de la distribución Gamma es distinto de 1. El objetivo general de esta tesis fue cuantificar la presencia de interferencia en datos de recombinaciones porcinas y desarrollar la expresión de la varianza del IBDP ($\text{Var}(\text{IBDP})$) bajo herencia por segmentos, para finalmente asociar dicha varianza a la del *proceso conjunto de recombinación* presentado recientemente por Carl Veller. A tal efecto se analizaron los datos de cada uno de los 18 autosomas para las fases paternas de 6441 cerdos y las fases maternas de 3065 animales nacidos entre 2017 y 2020. Se evaluó la presencia (o no) de interferencia estimando los parámetros de forma y escala con el modelo Gamma propuesto por Broman y Weber en 2000. Asimismo, se desarrollaron fórmulas para calcular la esperanza y la varianza del IBDP como funciones de los parámetros de la mezcla de Gammas. Con la información de los segmentos IBD, se calculó $E(\text{IBDP})$ y $\text{Var}(\text{IBDP})$ entre abuelo - nieto de todos los animales genotipificados, con los parámetros de la mezcla de Gammas estimados por máxima verosimilitud. Se observó interferencia en todos los cromosomas, excepto en 7, 9 y 14 (fase paterna), y 2, 15 y 16 (fase materna). La probabilidad estimada de ausencia de recombinación fue inversamente proporcional a la longitud del cromosoma en la fase paterna, mientras que para la fase materna no se observó relación alguna entre la interferencia y la longitud del cromosoma. El uso de información de segmentos IBD reduce la $\text{Var}(\text{IBDP})$ con respecto a la estimada utilizando sólo pedigree. Los resultados sugieren no ignorar la interferencia en el cálculo del IBDP y en su varianza en la especie porcina.

Palabras clave: recombinación, interferencia, identidad por descendencia

ABSTRACT

Parametric estimation of identity by descent segments using a Gamma mixture that considers interference at meiosis.

Crossover interference is the non-independence of crossovers events in a chromosome. The length of genome segments shared identical by descent (IBDP) by a pair of related individuals is a random variable following a distribution that is a mixture of Gammas, under a renewal process that allows for interference in recombination. The estimates of the “shape” parameter of the Gamma mixture display interference when their values are different to one. The main goal of the thesis was to detect the presence of interference in meiotic data from pigs, and to obtain the variance of IBDP ($\text{Var}(\text{IBDP})$) under segmental inheritance, and to relate that ensuing variance to the *aggregate recombination process* discussed recently by Carl Veller. Marker data from paternal phase on 6441 pigs and maternal phase on 3065 pigs born between 2017 to 2020 were employed for the analyses. The presence or absence of interference was checked through the shape and scale parameters of the Gamma mixture for the model proposed by Broman and Weber in 2000. In addition, formulae were derived to calculate the expected value and the variance of IBDP as a function of the parameters of the Gamma mixture. The $E(\text{IBDP})$ and the $\text{Var}(\text{IBDP})$ were calculated for the IBD segments from the relationship between grandparent and grand offspring of all genotyped animals with the previously derived expressions by Maximum Likelihood after fitting the parameters of the Gamma mixture. Except for chromosomes 7, 9 and 14 for the paternal phase and 2, 15, and 16 for maternal phase, interference was detected in all other 15 chromosomes. For the paternal phase, the estimated probability of no recombination was inversely related to chromosome length, but no association was found for the maternal phase. The use of information on IBD segments reduces the variance of IBDP with respect to the variance estimated using only pedigree. The results suggest that interference should not be overlooked when estimating the proportion of genome shared IBD and its variance in pigs.

Key words: recombination, interference, identity by descent.

Capítulo 1: Introducción

Capítulo 1

Introducción

Actualmente metodologías como la estimación de los valores de cría y los componentes de varianza de los efectos genéticos utilizados en la evaluación genética animal requieren utilizar información genómica, además del pedigrí, para cuantificar el grado de similitud genética existente entre dos individuos, condicional al grado de parentesco entre ellos (Legarra et al., 2014; Cantet et al., 2017). A tal efecto, es de suma importancia el desarrollo de modelos teóricos que permitan describir los momentos de las distribuciones del parentesco para cada par de individuos. En su trabajo fundacional, Fisher (1918) desarrolló el modelo de análisis de los efectos genéticos para loci que segregan de manera independiente. Posteriormente, Wright (1921) propuso modelar la similitud genética entre un par de individuos en función de un modelo correlacionar aleélico definido como la relación aditiva; Malecot (1948;1969) posteriormente es definida igual a dos veces la probabilidad que – posteriormente – sería definida como el coeficiente de parentesco. Este último es igual a la probabilidad que dos alelos muestreados al azar en dos individuos distintos provengan de un ancestro común, es decir: sean *idénticos por descendencia* (IBD). El muestreo presupone que cualquier gen muestreado en el genoma de un individuo tiene la misma similitud genética con el gen muestreado en el otro individuo, con lo cual el coeficiente de parentesco será el mismo en cualquier lugar del genoma. Esto permite calcular el valor esperado de la proporción de genes compartidos IBD entre uno y otro animal y, por lo tanto, la covarianza entre cada par de individuos emparentados empleando el método tabular (Henderson 1984).

Sin embargo, estos modelos no consideran que el modo de transmisión del genoma es por segmentos y no mediante la segregación individual e independiente de cada gen (Thompson, 2013). Aquellos genes causales que sean vecinos en un mismo segmento y cromosoma suelen segregarse conjuntamente, en mayor frecuencia que la esperable bajo independencia estadística, asociación que es conocida como *ligamiento*. Accionando con efecto opuesto al ligamiento se halla la *recombinación* del genoma de ambos abuelos dentro de las gametas parentales que al unirse forman el nuevo individuo. Un porcentaje variable según especie y sexo (Broman y Weber, 2000) de los cromosomas no recombina durante la meiosis y son llamados “parentales”, mientras que cuando ocurre al menos un evento de recombinación, el cromosoma es referido como

“recombinante”. Weir y Cockerham (1969) consideraron la similitud genética entre dos sitios de recombinación dentro de un cromosoma a través de la probabilidad de IBD entre *dos loci* generalizando la noción inicial de IBD en loci independientes. La probabilidad de IBD a dos loci es función del efecto de ligamiento, que se expresa por medio de la tasa de recombinación (r) entre ambos sitios en el genoma. En tal sentido, la tasa r es función de la probabilidad de ausencia de recombinación (P_0) del segmento o cromosoma, según propusiera Mather (1938), e igual a $r = 0.5(1 - P_0)$. Esta P_0 permite diferenciar entre relaciones de parentesco de un locus con el mismo valor esperado. Thompson (1988) demostró que mediante la noción IBD para dos loci para las relaciones abuelo – nieto, medio hermanos y tío – sobrino, salvo con recombinación libre ($r = 0.5$) o ausencia de esta ($r = 0$), se obtienen valores diferentes para cada una de estas relaciones. En ausencia de consanguinidad las tres coancestrías en un locus son iguales a 0.125, pero para dos loci muestran valores de coeficientes de coancestría diferentes en función del grado de recombinación. Este análisis sugiere que las probabilidades de IBD pueden ser más informativas de la relación de parentesco cuando pasen de considerar eventos *discretos* crecientes (un locus, dos loci, etc) a un proceso *continuo* de IBD (Donnelly, 1983) asociado con la recombinación de todo el genoma o “proceso conjunto de recombinación” (“aggregate recombination process”, Veller et al, 2020).

El uso de información genómica para calcular predicciones del valor de cría mediante ecuaciones de modelo mixto y predictores BLUP (Legarra et al, 2014) impacta fuertemente en la varianza de las predicciones entre años. En tal sentido, es imprescindible emplear una matriz de covarianzas de los valores de cría que lleve a resultados consistentes entre años. Para ello debe comprenderse cómo es afectada la variabilidad de las estimaciones del parentesco empleando información de pedigree y genómica durante la recombinación que ocurre en las meiosis específica e individuales que dan lugar a un nuevo individuo. Desde este punto de vista, esta tesis trata sobre la estimación de la recombinación meiótica en un “proceso conjunto de recombinación” en una o ambas gametas que forman un individuo (Veller et. (2020), y no de la noción más usual de recombinación entre dos sitios en el genoma, mirada como parámetro poblacional, es decir, la tasa de recombinación. Existen dos fuentes de variación respecto del parecido genético entre individuos emparentados. La primera es la variabilidad *dentro* de las relaciones genómicas para individuos con el mismo grado de parentesco. La segunda es la variabilidad debida al error de muestreo asociado con la

estimación (Visscher *et al.* 2006; Hill y Weir 2011). Para comprender cómo impacta la $\text{Var}(\text{IBD})$ en relaciones de parentesco que tienen un valor esperado similar es importante diferenciar entre la variable aleatoria IBD a un locus (Malecot, 1948, 1969), a dos loci (Weir y Cockerham, 1969) con la variable aleatoria IBDP (Guo, 1995), que involucra la noción de IBD *continuo* (Donnelly, 1983). A diferencia del modelo de segregación independiente descrito previamente (conocido además como IBD puntual o discreto), en la teoría de IBD continuo se considera el proceso biológico en el cual el genoma se hereda en *segmentos* de longitud variable. Esto se debe a que durante la meiosis que genera el material genómico contenido en espermatozoides y óvulos, los cromosomas homólogos intercambian material genético en distintos sitios de entrecruzamiento (crossover) en los diferentes cromosomas (sitios de recombinación), originándose en cada meiosis una variante gamética única. Considere por ejemplo, la relación padre – hijo y la relación entre hermanos enteros. Ambas tienen un valor esperado igual a 0.5, bajo un modelo de segregación independiente y ausencia de consanguinidad. Sin embargo, la $\text{Var}(\text{IBD})$ empleando solo información genealógica (Karigl, 1981), es igual a 0.125 para hermanos enteros mientras que la relación padre – hijo en ausencia de consanguinidad no muestra varianza alguna, i.e. $\text{Var}(\text{IBD}) = 0$. Más difícil aún es diferenciar entre las relaciones abuelo – nieto, medios hermanos y tío – sobrino utilizando solo información del pedigrí, cuyo valor esperado de IBD es 0.25 y la $\text{Var}(\text{IBD}) = 0.0625$ es igual para los tres parentescos. Guo (1995) mostró como calcular la esperanza ($E(\text{IBDP})$) y la varianza ($\text{Var}(\text{IBDP})$) del IBDP. Esta diferencia de variables aleatorias entre el IBD e IBDP de un mismo par de individuos emparentados fue anticipada por Bickeboller y Thompson (1986), quienes encontraron que no existe un número *discreto* de genes IBD que sea equivalente al IBDP, para una relación de parentesco en particular entre dos individuos emparentados.

Con el advenimiento de la tecnología genómica, es posible conocer qué par de bases nitrogenadas está presente en cada sitio marcado de una gameta, así como el origen ancestral (o fase) de dichos marcadores moleculares. Más aún, con información de tres generaciones es posible detectar los crossovers dentro de un cromosoma (Cheung *et al.*, 2007). Esta información permite utilizar modelos que estiman el número y las distancias entre recombinaciones en un cromosoma, o en un segmento comprendido entre dos sitios dentro de un cromosoma, en función de cierta tasa de recombinación r . Asimismo una *función de mapeo* (*map distance*) es una relación entre r y la distancia, (d) es decir $r = M(d)$ para conectar las frecuencia de recombinación y la

distancia física o de mapa (medida en centimorgans cM o en mega-pares de bases, Mpb) entre dos puntos a lo largo del cromosoma.

La función de mapa más empleada es la propuesta por Haldane (1919), bajo el supuesto de que las posiciones entre sitios de recombinación y de los espacios entre dichos sitios, sigan un proceso estocástico Poisson (Ross, 2007). En este proceso, los sitios de recombinación son variables aleatorias discretas con distribución Poisson, mientras que la longitud del ADN entre recombinaciones sigue una distribución exponencial. Existen otras funciones de mapeo como las propuestas por Kosambi (1944) o Felsenstein (1979). Sin embargo, estas funciones no permiten ajustar el *proceso conjunto de recombinación* de Veller et al (2020) cuando hay tres o más recombinaciones (Felsenstein, 1979) debido a su constitución algebraica que impide modelar más de tres sitios dentro. En la búsqueda de funciones de mapeo que puedan ajustar más de tres sitios de recombinación, Liberman y Karlin (1984) definieron la noción *recombinación factible en varios loci (multilocus feasible)*, clase limitada a distribuciones estadísticas de la familia exponencial. En tal sentido, la función de mapeo de Haldane es de factibilidad multilocus, y posee la propiedad conocida como de “falta de memoria” (*memoryless*) donde la probabilidad de ocurrencia de un evento de recombinación es *independiente* de la de la recombinación ocurrida previamente. Como consecuencia, la longitud de los segmentos entre recombinaciones dentro de un mismo cromosoma es totalmente independiente a lo largo del genoma cuando no existe *interferencia* física de las recombinaciones precedentes sobre las posteriores dentro de un mismo cromosoma. Alternativamente, otros modelos como el propuesto por Felsenstein (1979) o Kosambi (1944) no cumplen con la propiedad de factibilidad multipunto. Por otra parte, el modelo de recombinación de la distribución Gamma ajustó mejor que varios otros las distancias entre dos sitios en el genoma (Zhao y Speed, 1996). Asumiendo que el proceso meiótico se ajusta al modelo de Haldane (1919), Hill y Weir (2011) presentaron ecuaciones para calcular la $\text{Var}(\text{IBD})$ de distintas relaciones de parentesco en ausencia de consanguinidad, y también para ciertas relaciones de parentesco en presencia de consanguinidad (Hill y Weir, 2012).

Sin embargo, existen evidencias de asociación o interferencia entre eventos de recombinación en un mismo cromosoma, para distintas especies (Broman y Weber, 2000; Housworth y Stahl, 2003). Este hecho explica por qué en muchas ocasiones los modelos de recombinación que ignoran la interferencia ajustan pobremente los datos (Speed 2005). En presencia de interferencia, Cobbs (1978) sugirió modelar las

recombinaciones empleando la distribución Gamma que permite condicionar un evento aleatorio en el valor anterior de dicha variable, dentro de la misma gameta. McPeck y Speed (1995) encontraron que la distribución Gamma ajustó razonablemente datos meióticos de *Drosophila*, mientras que Broman y Weber (2000) y Housworth y Stahl (2003, 2009) obtuvieron resultados similares con cromosomas humanos. Asimismo, Caballero *et al.* (2019) simularon datos bajo un modelo de recombinación bajo interferencia, y mostraron que para distintas relaciones de parentesco entre colaterales la $\text{Var}(\text{IBDP})$ se redujo en comparación con la situación de ausencia de interferencia. Al considerar el proceso de interferencia se puede estimar P_0 , o probabilidad de ausencia de recombinación, para cada gameta, información de utilidad al estimar la varianza aditiva bajo ligamiento (Cockerham, 1956; Cockerham y Weir, 1969), o la varianza genética aditiva \times aditiva (Cockerham 1984; Totir y Fernando, 1998). Sin embargo, para cada recombinación existe un solo grado de libertad para estimar P_0 entre el antecesor hembra y el macho de los cuales se recombina su material genético. Veller *et al.* (2020) parametrizaron el *proceso conjunto de recombinación* de ese único grado de libertad en una probabilidad que, cuando vale 0, expresa que los segmentos genómicos en esa gameta provienen exclusivamente del antecesor hembra o, más precisamente, de la abuela. Alternativamente cuando el parámetro vale 1, la gameta es enteramente parental del abuelo. Ambos valores son biológicamente improbables. Finalmente, el estado de igualdad de aportes genómicos entre ambos abuelos es reflejado por un valor del parámetro igual a 0.5.

El objetivo general de esta tesis es desarrollar la expresión de la $\text{Var}(\text{IBDP})$ en un proceso de herencia en segmentos, bajo recombinación afectada por interferencia, cuantificando la presencia de éste último proceso en datos reales de recombinaciones en una población porcina altamente selecta, además de la estimación de parámetros concernientes al proceso de recombinación asumiendo interferencia mediante un . Para ello, en el Capítulo 2 describiremos las distintas medidas probabilísticas de IBD, para luego explicar en detalle la modelación del proceso estocástico de interferencia, y finalmente desarrollar fórmulas de $E(\text{IBDP})$ y $\text{Var}(\text{IBDP})$ bajo interferencia, asumiendo que las relaciones de parentesco son función de un proceso IBD continuo. En el Capítulo 3 se describe la base de datos de meiosis porcinas empleada en los análisis, para luego presentar las distribuciones ajustadas para evaluar interferencia. El Capítulo 4 muestra los resultados obtenidos en cuanto al análisis de interferencia, como así también respecto a las fracciones IBDP descritas en el Capítulo 2. Posteriormente, en

el Capítulo 5 se discuten los resultados y en el 6 se presentan las conclusiones generales de la tesis.

**Capítulo 2: Teoría sobre la fracción de genoma compartido
IBD entre Individuos**

Capítulo 2

Teoría sobre la fracción de genoma compartido IBD entre individuos

La fracción de genoma compartido idéntico por descendencia entre un par de individuos (IBDP) es la variable aleatoria resultante de la herencia de material genético ancestral en común para ambos integrantes del par. Su estimación es de importancia fundamental para determinar la similitud genética entre individuos con el objeto de poder así escribir la matriz de varianzas y covarianzas de los valores de cría en modelos de evaluación genética que puedan utilizar el IBDP (Cantet et al., 2017). Los métodos actualmente empleados para calcular la similitud genética bajo herencia mendeliana en animales y plantas consideran un mecanismo de herencia con *efectos aditivos independientes* bajo el supuesto de ausencia de ligamiento (Kempthorne, 1969; Henderson, 1984; Lynch y Walsh, 1998). Sin embargo, los genes causales son segregados en segmentos de genoma, con lo cual los efectos génicos que inducen dichas variantes causales *no son estadísticamente independientes* (Thompson, 2013). Estos segmentos son formados durante la meiosis, donde ocurren eventos de entrecruzamiento entre dos cromátidas no hermanas que intercambian ADN de distinto origen parental. En este capítulo se obtendrán expresiones del valor esperado y la varianza de la proporción de genoma compartido idéntico por descendencia (IBDP) entre abuelos y nietos al ajustar la distribución de probabilidad de los segmentos compartidos dentro de la fase parental de un individuo que son IBD respecto del abuelo y/o la abuela. El interés en esta relación de parentesco en particular es que involucra individuos de tres generaciones y permite estimar los parámetros del modelo de regresión ancestral de Cantet et al. (2017), permitiendo así detectar los sitios de recombinación del material genético proveniente de los abuelos y recibido por sus nietos debido a que se calculan las fases paterna y materna con probabilidad uno. Detectados los sitios de recombinación es posible calcular la longitud de los segmentos heredados en el nieto originados en cada abuelo.

Tanto para la noción de IBD discreta como para la continua, el proceso IBD en distintos pedigrís muestra una gran variabilidad (Donnelly, 1983; Weir y Hill, 2011; Thompson, 2013). Modelar y evaluar esta varianza de las relaciones de parentesco, bajo un modelo de herencia segmental o multiloci, es el principal objetivo de esta tesis, dado la importancia que tiene en la evaluación genética para la selección genómica. En este

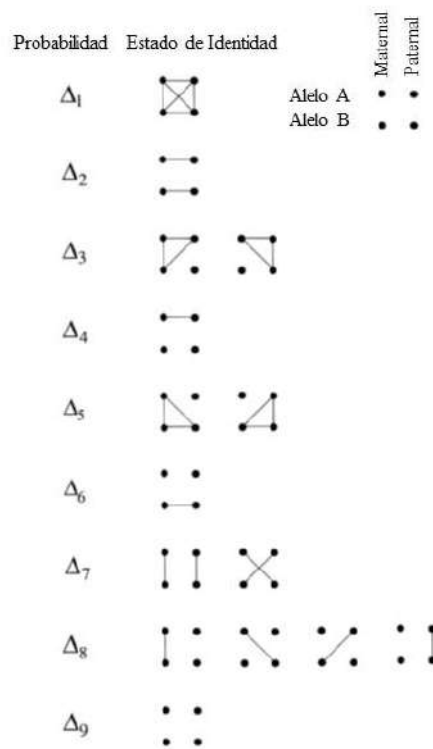
capítulo discutiremos primero las dificultades de la herencia “puntual” o IBD discreto a un locus, para luego considerar la variabilidad de IBD. Posteriormente introduciremos la herencia en segmentos y su θ_{XY} condicional a la información genómica de individuos en *tres generaciones consecutivas* (abuelos, padres, hijos e hijas) siguiendo un algoritmo sugerido por el método de identificación del origen de sitios del genoma compartidos entre abuelo y nieto por Cheung et al. (2007), para luego caracterizar la distribución de dichos segmentos de IBD continuo bajo un proceso de renovación (“renewal process”; Ross, 2007) cuando la recombinación sufre o no interferencia (Cobbs, 1978).

2.1. Medidas de identidad por descendencia para IBD discreto.

Distintos autores analizaron el parecido genético entre parientes (Fisher, 1918; Wright, 1921; Cockerham, 1954; Kempthorne, 1954) y parametrizaron el resultado como una combinación lineal de componentes de varianza de efectos genéticos ponderados por medidas de identidad entre parientes. Malecot (1948, 1969) unificó estas ideas definiendo formalmente el parecido entre individuos emparentados en términos de la *probabilidad de IBD*, $P(\text{IBD})$. Esta probabilidad se asocia con el evento en que variables aleatorias que identifican dos genes homólogos, portados por dos individuos distintos, son iguales debido a que poseen uno o más ancestros en común. La notación clásica de la probabilidad de IBD entre X e Y es $P(X \equiv Y)$. Cuando referimos a un origen en particular (por ejemplo, animales X e Y), el símbolo ‘ \equiv ’ resalta que la igualdad se debe a la ancestría en común del gen recibido por X y su homólogo recibido por Y, siendo ambos genes *copias idénticas* de un gen ancestral. Conceptualmente, la noción de IBD no se limita en que las variables X e Y sean un gen causal, sino que también pueden ser porciones de cromosoma homólogas, cromosomas o gametas de dos individuos distintos o paterna y materna del mismo individuo (Donnelly, 1983; Guo, 1995; Thompson, 2013). En ciertas aplicaciones es usual emplear funciones de $P(X \equiv Y)$ conocidas como relaciones de parentesco (A_{XY}), o dos veces el *coeficiente de coancestría* (θ_{XY}) de Malecot, específicamente: $A_{XY} = 2 P(X \equiv Y) = 2 \theta_{XY}$. Para un individuo X, hijo de S_X (el padre) y D_X (la madre), el evento IBD ocurre con respecto a los genes recibidos de S_X y D_X y la $P(\text{IBD}) = P(S_X \equiv D_X)$ es referida como *coeficiente*

de consanguinidad (F_X , Wright, 1922) del individuo X; siendo entonces F_X igual a la coancestría entre sus padres: $F_X = P(S_x \equiv D_x) = \theta_{S_x D_x}$. Esta igualdad se cumple en presencia de información genealógica solamente y en completa ausencia de información genómica. Dado que $2\theta_{S_x D_x} = A_{S_x D_x}$, la relación entre F_X y A_{SD} es igual a $F_X = 0.5 A_{S_x D_x}$. Si bien originalmente la IBDP en un locus se calculaba mediante el número de pares de genes IBD (n) entre ambos individuos en la relación, i.e. $n = 0, 1, 2$ (Cotterman, 1941), Jacquard (1972) estableció un espacio muestral con nueve (9) estados de identidad entre dos pares de alelos homólogos, los que se observan en la Figura 1.

Figura 1. Estados de Identidad IBD en un locus (Jacquard, 1972).



Los genes del individuo X se encuentran en la parte superior y los del individuo Y en la parte inferior de cada diagrama. Las líneas establecen que ese par de alelos es IBD, indicando las líneas horizontales la presencia de consanguinidad. Nótese que en total existen 15 estados de identidad, pero Jacquard (1972) los redujo a nueve (9) en la situación que es irrelevante o imposible determinar si el origen del IBD es por vía paterna o materna.

2.1.1. Varianza del IBD con información genealógica.

Karigl (1981) generalizó el cálculo de los coeficientes de coancestría para la $P(\text{IBD})$ de un loci en presencia de consanguinidad. Para ello describió cada una de las fórmulas para el cálculo de $P(\text{IBD})$ entre cuatro alelos en dos individuos (los puntos negros en la Figura 1), en número de quince (15) dentro del espacio muestral. Jacquard (1972) redujo estos 15 a nueve estados de identidad $\Delta_1, \dots, \Delta_9$ indicados a la izquierda en la Figura 1. Los 15 estados de IBD de Karigl (1981) se describen en la Tabla 1 de Liu y Weir (2005). Dichos estados son: 1) cuatro genes IBD; 2) tres genes IBD y el restante no IBD; 3) dos pares de genes IBD; 4) un par de genes IBD y los otros dos genes no; 5) ningún par de genes son IBD. Denotando los genes homólogos a y b en X y c y d en Y , las 15 medidas posibles son las siguientes:

| | | |
|-------------------------|---|---|
| Cuatro genes IBD: | $a \equiv b \equiv c \equiv d$ | 1 |
| Tres genes IBD: | $a \equiv b \equiv c$ $a \equiv b \equiv d$ $a \equiv c \equiv d$ $b \equiv c \equiv d$ | 4 |
| Dos pares de genes IBD: | $a \equiv b, c \equiv d$ $a \equiv c, b \equiv d$ $a \equiv d, b \equiv c$ | 3 |
| Un par de genes IBD: | $a \equiv b$ $a \equiv c$ $a \equiv d$ $b \equiv c$ $b \equiv d$ $c \equiv d$ | 6 |
| Ningún par de gen IBD: | | 1 |

Los números a la derecha indican el *número de casos* a considerar para cada medida.

Las IBD para dos loci tienen un espacio muestral similar al de un locus teniendo en cuenta el origen de los alelos, hecho que ignoran los estados de identidad condensados de Jacquard (1972). Es así que el cálculo de $P(X_1 \equiv Y_1, X_2 \equiv Y_2)$ es similar al de la medida a un loci $\Phi_{xy,xy}$ definida por Karigl (1981) o $P(\text{IBD})$ para dos pares de genes ($a \equiv b, c \equiv d$; $a \equiv c, b \equiv d$; $a \equiv d, b \equiv c$, definidos en la página anterior). Cockerham (1971, página 241) desarrolló previamente teoría de $P(\text{IBD})$ a un locus que permite expresar $\Phi_{xy,xy}$ del modo siguiente

$$\Phi_{xy,xy} = 0.25 \left(\begin{array}{l} P(X \equiv Y) + P(X_1 \equiv X_2 \equiv Y) + \\ P(Y_1 \equiv Y_2 \equiv X) + P(X_1 \equiv Y_1, X_2 \equiv Y_2) \end{array} \right) \quad [2.1]$$

Las $P(\text{IBD})$ en [2.1] son respectivamente para un par de alelos (coancestría entre X e Y), dos genes en X y uno en Y , un gen en X y ambos alelos en Y y finalmente dos genes en X y dos en Y . Como los efectos génicos son independientes en el modelo de un loci, los alelos individuales pueden asimilarse como originados en dos, tres o cuatro sitios en [2.1] (Weir y Cockerham, 1969). Consecuentemente, se expresa esta

coancestría como el promedio de la probabilidades correspondientes que ese par de individuos compartan dos alelos (dialélica), tres alelos (trialélica) y cuatro alelos (tetralélicas). La definición de dicha coancestría como medida conjunta de dos pares de genes es igual a $P(X_1 \equiv Y_1, X_2 \equiv Y_2) = 4 \Phi_{xy,xy}$. Con lo cual, la varianza de IBD a dos loci es igual a:

$$\text{Var}(\text{IBD}) = 4 \Phi_{xy,xy} - (2 \theta_{xy})^2 \quad [2.2]$$

Jacquard (1972) observó que

$$\theta_{xy} = P(X \equiv Y) = \Delta_1 + 0.5 (\Delta_3 + \Delta_5 + \Delta_7) + 0.25 \Delta_8 \quad [2.3]$$

Para la probabilidad “trialélica” entre alelos IBD en número de dos en X y uno en Y, provenientes de tres sitios distintos en la generación anterior, se verifica que $\Delta_5 = \Delta_7 = \Delta_8 = 0$, con lo cual

$$P(X_1 \equiv X_2 \equiv Y) = \Delta_1 + 0.5 \Delta_3 \quad [2.4]$$

La restante probabilidad trialélica de un alelo en X que es IBD a ambos alelos en Y, se verifica que $\Delta_3 = \Delta_7 = \Delta_8 = 0$. Como resultado obtenemos

$$P(X \equiv Y_1 \equiv Y_2) = \Delta_1 + 0.5 \Delta_5 \quad [2.5]$$

Para la probabilidad tetralélica entre dos alelos de sendos sitios provenientes de los padres de X e Y, que sean IBD en pares de alelos no homólogos, tenemos que $\Delta_3 = \Delta_5 = \Delta_8 = 0$, y por lo tanto

$$P(X_1 \equiv Y_1, X_2 \equiv Y_2) = \Delta_1 + 0.5 \Delta_7 \quad [2.6]$$

Sustituyendo con [2.3], [2.4], [2.5] y [2.6] en [2.2], se obtiene la siguiente expresión

$$4 \Phi_{xy,xy} = \Delta_1 + 0.5 (\Delta_3 + \Delta_5 + \Delta_7) + 0.25 \Delta_8 + \Delta_1 + 0.5 \Delta_3 + \Delta_1 + 0.5 \Delta_5 + \Delta_1 + 0.5 \Delta_7$$

o, equivalentemente

$$4 \Phi_{xy,xy} = 4 \Delta_1 + \Delta_3 + \Delta_5 + \Delta_7 + 0.25 \Delta_8 \quad [2.7]$$

Este resultado es similar al obtenido por Karigl (1981, expresión (7)). Específicamente, la expresión [2.7] es igual al noveno elemento del vector resultante del sistema de ecuaciones utilizado para calcular las $P(\text{IBD})$ de los 15 estados.

Finalmente, reemplazando con [2.7] y [2.3] en [2.2] la varianza del IBD a dos loci expresada como función de los estados de identidad de Jacquard (1972) es igual a:

$$\text{Var}(\text{IBD}) = [4 \Delta_1 + \Delta_3 + \Delta_5 + \Delta_7 + 0.25 \Delta_8] - 4 [\Delta_1 + 0.5 (\Delta_3 + \Delta_5 + \Delta_7) + 0.25 \Delta_8]^2 \quad [2.8]$$

Esta ecuación permite calcular los valores de la $\text{Var}(\text{IBD})$ cuando la fuente de información es solamente la información genealógica bajo el modelo de un locus, e incorporando consanguinidad. La fórmula [2.8] puede emplearse para parentescos en ausencia de consanguinidad. En el caso de la relación abuelo – nieto, el valor correspondiente es $\text{Var}(\text{IBD}) = 1/16$, valor que aparece en la Tabla 1 de Hill y Weir (2011), quienes emplearon una metodología distinta para alcanzar el mismo resultado encontrado aquí.

2.2. Medidas de identidad por descendencia para IBD continuo.

Para cualquier carácter complejo, las posiciones y el número de los genes causales, así como los efectos de dichos genes, suele ser desconocido. Recientemente, Boyle et al. (2017) propusieron el modelo de determinación genética “omnigénico”, donde se hipotetiza que las cadenas regulatorias de los genes causales están altamente interconectadas, lo que lleva a que “todos” los genes expresados en un carácter complejo tengan la capacidad de afectar las funciones de genes cuya acción es importante (“core-genes”) para determinar la variabilidad del carácter. Por lo tanto, la mayor parte de la heredabilidad puede explicarse en la acción génicas indetectables en los análisis de asociación para determinar la variación aditiva. Esto conlleva una gran incertidumbre en la parametrización estocástica de la “arquitectura genética” asociada a la probabilidad que los distintos efectos causales de dos individuos sean IBD.

La noción de IBD “discreto” o contable (uno, dos, etc) requiere ser modificada en aplicaciones como la predicción del valor de cría (Cantet et al., 2017). La primera referencia en la cual se caracteriza un proceso IBD de naturaleza “continua” es Donnelly (1983), quién consideró que el genoma no es transmitido por genes independientes, sino por bloques de material genómico o segmentos. El IBD continuo es modelado por Donnelly (1983) como un “paseo aleatorio” (random walk) dentro del genoma, con una dimensión igual a la suma de todas las meiosis presentes en el pedigrí. La viabilidad de su cálculo solo es posible mediante la propiedad Markoviana del sistema, donde la información en una cierta meiosis solo depende de aquella presente en el estado anterior. Bickeboller y Thompson (1988) concluyeron que no existe un número finito de loci donde se igualan las nociones de IBD discreto y continuo. Thompson (2013) revisó los avances en el cálculo de las distintas relaciones de parentesco bajo el proceso de IBD continuo.

Cockerham (1956) definió una probabilidad para dos loci ligados, hecho que permitió considerar el evento IBD en *ausencia de independencia estadística* debido al *ligamiento*. Posteriormente, Weir y Cockerham (1969) formalizaron esta definición para el caso general con consanguinidad. Las medidas de IBD a dos loci cuantifican la segregación conjunta entre dos genes en una gameta segregada, parametrizando el ligamiento como dependencia estadística y generalizando así la herencia en el modelo de un locus (Kempthorne, 1957, cap. 15; Lynch y Walsh, 1998, capítulo 4). Denniston (1975) observó que para dos loci los estados de identidad pasan de 15 a $15^2 = 225$, pero en ausencia de consanguinidad se limitan a $7^2 = 49$. De cualquier manera, el cálculo de estas probabilidades es mucho más complejo que las de un loci, como observaron Weir y Cockerham (1969) para dos loci con un pedigrí reducido.

Actualmente, los paneles (chips) de marcadores de un único nucleótido (SNP, por sus siglas en inglés) en alta densidad, proveen información del parecido entre parientes a un nivel más informativo de IBD: haplotipos o segmentos de ADN, o todo un cromosoma. Esta información puede entonces complementar aquella que proviene de la genealogía. Donnelly (1983) y Guo (1995) ampliaron la noción de IBD discreto a la situación de una variable aleatoria *continua*. Sin embargo, tanto las posiciones cómo el número de genes causales involucrados de cualquier carácter de herencia poligénica suelen ser desconocidos y la “arquitectura genética” de un carácter poligénico complejo es parcial o totalmente desconocida. Por lo tanto, la noción más general de relación de parentesco sea dentro (i.e. entre los padres) o entre individuos, requiere identificar a la *fracción de genoma compartido IBD*. Guo (1995) definió la proporción de genoma compartido IBD (IBDP) por dos individuos emparentados como la variable aleatoria continua, igual a la suma de la longitud de todos los segmentos IBD dividida por la longitud del genoma (L). Sea $I(x)$ una variable indicadora igual a 1 cuando las bases son IBD en el sitio x y 0 caso contrario. La proporción del genoma compartido IBD es entonces igual a $(L)^{-1} \int_0^L I(x) dx$, definida en la distancia del segmento que ocurre entre $0 < x < L$, donde 0 corresponde al arbitrario inicio en la primer base del cromosoma y L es la distancia final que media hasta la última base del último cromosoma.

2.3. Estimación de la longitud de los segmentos compartidos entre abuelo y nieto

Para cuantificar la longitud de los segmentos generados durante la recombinación en un cromosoma generalizaremos el método propuesto por Cheung (2007) y que desarrollaremos en detalle a continuación. Nuestro procedimiento consiste en identificar inequívocamente sitios de recombinación dentro del genoma de la gameta parental, detectando en cada marcador de un chip denso de SNP si fue heredado del abuelo o de la abuela, dentro de un *vector indicador de herencia* (“inheritance vector”, Lander y Green, 1987). Para cada gameta se define un vector de herencia de longitud igual al número de marcadores SNP en el chip, con elementos codificados con 1 cuando el marcador heredado proviene del abuelo; 2 cuando el marcador proviene de la abuela y 0 cuando el marcador no es informativo y el origen ancestral no puede determinarse fehacientemente.

En este procedimiento son requeridos los genotipos de ambos abuelos, del padre, de la madre y del individuo. Cheung et al (2007) mostraron en qué condiciones se puede detectar el nivel de información de cada marcador, situaciones que hemos expandido y esquematizado en las Figuras 2 y 3, y en mayor detalle en el Apéndice A. El padre cuyo genotipo es conocido (i.e., aquel que da origen a la gameta de la que se quieren identificar las fases) es informativo solo en aquellos marcadores *heterocigotas*. Simultáneamente, el otro padre debe ser homocigota para las situaciones descritas en la Figura 2, o alternativamente heterocigota en las situaciones que se describen en la Figura 3. Otro tanto es requerido de los abuelos: uno debe ser heterocigota en el marcador y el otro homocigota (Figura 2), o ambos deben poseer genotipos homocigotas opuestos (AA y aa), en los casos descritos en la Figura 2. Considerando el ejemplo de la Figura 2, la cría porta el genotipo heterocigota Adenina – Timina (AT), su padre es heterocigota (AT), su madre homocigota (TT), el abuelo paterno homocigota (AT) y la abuela heterocigota (TT). Nótese que en este caso la cría recibió el marcador dentro de la gameta paterna que pertenece al abuelo. El Apéndice A describe en total nueve (9) patrones similares a este y otros nueve cuando el otro padre es considerado referente, dieciocho (18) en total.

En el otro patrón de herencia IBD con respecto a los abuelos (Figura 3), la cría posee el genotipo homocigota de uno de sus abuelos. Existen un total de dieciocho (18) esquemas de este tipo, para alcanzar un total de treinta seis (36) patrones de IBD.

el valor “2” en posición 39 y el “1” en posición 43. Consecuentemente se asume que la recombinación debió tener lugar en el punto medio entre los marcadores 40 y 42. Por otro lado, los valores “0” se refieren a las posiciones donde no existe información acerca del origen ancestral de las bases, y donde el procedimiento que se observa en las Figuras 2 y 3 y en el Apéndice A no arroja resultados definidos. El tratamiento de los ceros fue el siguiente: cuando se ubicaban entre SNP informativos con el mismo origen (abuelo = 1, abuela = 2), se les asignó dicho origen y cuando se ubicaron entre marcadores con distinto origen, se los repartió a la mitad. La longitud de los segmentos es calculada empleando valores de distancia en mega pares de bases (mbp) con el mapa de los marcadores que provee la empresa que genotipifica.

2.4. El proceso de interferencia durante la meiosis

En el mecanismo de reproducción celular de la meiosis, una célula diploide ($2n$) experimenta dos divisiones sucesivas (Meiosis I y II), obteniéndose como resultado cuatro células haploides (n) con la mitad del material genético de la célula original. En la Profase de la Meiosis I, cada cromosoma tiene su ADN duplicado y distribuido en dos cromátidas hermanas, y los cromosomas homólogos se aparean formando una tétrada. Luego, ocurre el entrecruzamiento de cromátidas no hermanas de la tétrada. Posteriormente, se separan los cromosomas homólogos en la Anafase I, y sus cromátidas hermanas en la Anafase II, produciéndose cuatro células hijas haploides que serán las precursoras de las gametas del individuo. La meiosis como mecanismo de división celular aumenta las probabilidades de éxito de la fecundación garantizando así 1) la generación de nuevos individuos que perpetúen la especie, como así también 2) la variabilidad genética de la especie, porque la meiosis resulta en combinaciones genéticas distintas a las que existían en las células parentales.

Desde el punto de vista del animal individual y no de la población, Veller et al. (2020) observaron que el proceso de recombinación que ocurre durante la meiosis y que resulta en las gametas parentales, afecta las proporciones transmitidas de los genomas de los abuelos a la progenie de la última generación, proceso al que denominaron “aggregate recombination process” y que traduciremos como “proceso conjunto de recombinación”. Este proceso introduce variabilidad en la proporción del genoma compartido IBD entre individuos emparentados, la cual es condicionada por la ubicación de los eventos de recombinación (Caballero et al., 2019). Dentro del genoma de un individuo, las ubicaciones de los eventos sucesivos de crossover son afectadas por

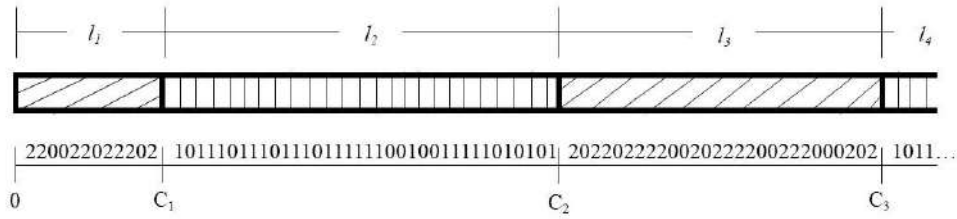
el fenómeno de *interferencia* (Cobbs, 1978; Foss et al, 1993), proceso físico donde la ocurrencia de un crossover impide o dificulta la ocurrencia de otro crossover en una posición cercana a la anterior. La modelación estocástica del *proceso conjunto de recombinación* involucra necesariamente individuos de tres generaciones (Veller et al., 2020). En el contexto de este proceso, la interferencia física a la libre ocurrencia de crossovers en las cromátidas durante la meiosis altera la distribución estadística de la longitud de los segmentos (McPeck y Speed, 1996) de genoma que un individuo comparte con sus abuelos .

Existen dos tipos de interferencia (Zhao *et al.* 1995). La primera es entre *cromátidas* y depende de la probabilidad que un par de cromátidas, dentro de la tétrada producida durante la meiosis, tenga la posibilidad de entrecruzarse. El otro tipo es la interferencia de *quiasmas* debido a la ubicación no independiente de las recombinaciones dentro de la gameta. Para cuantificar la interferencia entre cromátidas se requiere identificar los cuatro productos de la tétrada meiótica, para encontrar estadísticos de los segmentos en función de los eventos de recombinación, en cada una de las cuatro gametas formadas durante la meiosis. En caso que los eventos de recombinación sean detectados exclusivamente a partir de la fase presente en cada individuo, se debe asumir la ausencia de este tipo de interferencia.

Para modelar el *proceso conjunto de recombinación* considerando la interferencia de quiasma dentro de una gameta parental, es necesario identificar los *eventos de recombinación* de un individuo, y el *origen ancestral* de cada segmento delimitado por dos recombinaciones. Esto incluye al segmento que va desde el inicio del cromosoma hasta la primera recombinación, y aquel que va desde la última recombinación hasta el final del cromosoma. La distribución estadística de la longitud de los segmentos puede modelarse (McPeck y Speed, 1996) con un *proceso de conteo* (“*Counting Process*”; Ross, 2017) con las siguientes propiedades:

- Cada *recombinación* (definida mediante un número entero) sigue una distribución de probabilidad discreta.
- Los “tiempos” del proceso, en nuestro caso la *distancia* en centimorgans (cM) entre recombinaciones, son *independientes e idénticamente distribuidos con una distribución de probabilidad continua*.
- No existe superposición alguna entre las distancias dentro de cada evento contable. Esto implica que *cada intervalo comienza en el origen del cromosoma*.

Figura 4. Distribución de las recombinaciones y los segmentos IBD dentro de una gameta. C_1, C_2, C_3 denotan los eventos de recombinación, mientras que l_1, l_2, l_3, l_4 son los segmentos IBD dentro de esa gameta



La Figura 4 muestra la distancia entre dos eventos sucesivos de conteo definiendo cada segmento IBD al abuelo o a la abuela, sean paternos o maternos. Se observan cuatro segmentos en un cromosoma con sus distancias desde el inicio del proceso de conteo ($d = 0$) al evento n –ésimo de recombinación, S_n . esta variable aleatoria es entonces igual a:

$$S_n = \sum_{i=0}^n l_i \quad [2.9]$$

Por ejemplo, las distancias $S_2 = l_1 + l_2$ y $S_3 = l_1 + l_2 + l_3$, donde a partir de los segmentos desde l_2 hasta el segmento l_{n-1} están definidos por dos eventos de recombinación C_i . Por el proceso de renovación (Ross, 2007, cap. 7), el evento S_2 es independiente de S_3 . Sea μ la distancia media entre cada sitio de recombinación, la $E(S_n) = \mu$ debido a que por la “ley fuerte de los grandes números” (Ross, 2007; teorema 2.1, página 79) $\frac{S_n}{n} \xrightarrow{n \rightarrow \infty} \mu$ y el promedio de las distancias entre renovaciones converge a la media poblacional del proceso cuando el número de renovaciones es grande.

Empleando un proceso de conteo Poisson para derivar la distribución de la longitud de segmentos se obtiene una exponencial, lo que implica asumir ausencia de interferencia, es decir $l_1, l_2, \dots, l_n \stackrel{iid}{\square} Exp(\gamma)$, con función de densidad igual a $f(x|\gamma) = (\gamma)^{-1} e^{-\frac{x}{\gamma}}$, distribución cuyo parámetro de esperanza es igual a γ y con varianza igual a γ^2 . Esta distribución exponencial tiene la propiedad que las longitudes de los segmentos entre recombinaciones dentro de un cromosoma son variables aleatorias *independientes*, propiedad conocida como de *ausencia de memoria* (*memoryless*). Para dos segmentos de longitud h y t y en términos de probabilidad, la

falta de memoria se caracteriza porque la distribución condicional del segmento $h + t$ dado el valor de t , es igual a la distribución marginal de h . Formalmente:

$$P(X > h+t | X > t) = \frac{P(X > h+t, X > t)}{P(X > t)} = \frac{P(X > h) P(X > t)}{P(X > t)} = P(X > h) \quad [2.10]$$

Tanto $P(X > h+t, X > t)$ como $P(X > t)$ siguen una distribución exponencial con valor esperado $1/\gamma$, con lo cual y sustituyendo en [2.10] obtenemos:

$$P(X > h+t | X > t) = \frac{(\gamma)^{-1} e^{-(h+t)/\gamma}}{(\gamma)^{-1} e^{-t/\gamma}} = \frac{e^{-h/\gamma} e^{-t/\gamma}}{e^{-t/\gamma}} = e^{-h/\gamma} = P(X > h)$$

Esta última expresión es consistente con la función de mapeo de Haldane (1919) quién empleó la noción de que la probabilidad de ocurrencia de una recombinación, a una distancia t , es *independiente* de la distancia previamente recorrida, propiedad que se da solamente con interferencia ausente.

Para comprender la motivación al utilizar distribuciones mezclas de Gamma para la longitud de los segmentos, se debe considerar que solo es posible observar los segmentos en las gametas de un individuo y no en todas las que resultaron de *ese* evento meiótico en el progenitor. Esto conlleva una pérdida de información estadística, dado que observamos una de las dos posibles gametas parentales y no ambas (MacPeck y Speed, 1996). Como consecuencia, el espacio muestral de la distribución de la longitud de los segmentos observados es más pequeño en relación con el que resultaría si observásemos todos los productos meióticos. Nolan (2017) describió este “refinamiento” (“thinning”, MacPeck y Speed, 1996) como el retener la observación de la longitud de un segmento entre dos puntos de recombinación con probabilidad $P = 0.5$ y eliminar el valor del segmento anterior (del que no tuvo “éxito”) con igual probabilidad. En consecuencia, la longitud de los segmentos (o espacios entre recombinaciones) l_i siguen una distribución $l_1, l_2, \dots, l_n \stackrel{iid}{\sim} \text{Gamma}(\nu, \tau)$ (Cobb, 1978; Zhao y Speed, 1995; Broman y Weber, 2000). Es decir que cada segmento IBD sigue una distribución Gamma con parámetro de “forma” (“*shape*”) ν y la “tasa de cambio” τ (“*rate*”) o inversa del parámetro de “escala”. La función de densidad resultante es igual a:

$$f(l_i | \nu, \tau) = \frac{\tau^\nu}{\Gamma(\nu)} l^{ \nu-1 } e^{-\tau l} \quad [2.11]$$

La variable l es la longitud del segmento IBD con esperanza igual a $v \tau^{-1}$ y varianza igual a $v \tau^{-2}$.

Capítulo 3: Materiales y Métodos

Capítulo 3

Materiales y Métodos

3.1. Base de datos analizada

Se utilizó una base de datos conteniendo los genotipos de la fase paterna de 6418 cerdos y de 3065 cerdos con fase materna genotipificada de la empresa Genus-PIC, nacidos entre 2017 a 2020. Los registros pertenecen a una línea sintética creada en 1990 a partir de las razas Landrace, Large White y Duroc. La línea fue mantenida con apareamientos inter-se a partir de 1996, por un total de 16 generaciones (Pieter Knap, 2021, comunicación personal). Los datos consisten en genotipos para 40800 marcadores de un chip de 50000 marcadores SNP Illumina por animal distribuidos en 18 cromosomas autosómicos. Las fases paterna y materna fueron evaluadas mediante el algoritmo de Cheung (2007) descrito en la Sección 2.3. Se detectaron los sitios de recombinación, la longitud de los espacios entre recombinaciones o segmentos IBD entre el animal, nieto(a), y su abuelo o abuela y el origen ancestral de cada segmento en la generación de los abuelos paternos de cada animal. Los segmentos se estandarizaron de modo que cada cromosoma tuviera una longitud de 100 cM, independientemente de su longitud real.

3.2. La fracción del genoma del abuelo presente en sus nietos

Considerando un abuelo y un nieto y la gameta producida por el progenitor G de ese nieto, Veller et al. (2020) definieron el parámetro P_G a la fracción del genoma de esa gameta que deriva de ese abuelo. Debido a la segregación Mendeliana, P_G tiene valor esperado igual a 0.5, i.e. $E(P_G) = 0.5$. Este parámetro se estima dividiendo la suma de todos los segmentos IBD compartidos entre abuelo y nieto, por la longitud del genoma L . En el caso del cromosoma i con longitud L_i , tenemos que

$$\hat{P}_{Gi} = \frac{\sum_{j=1}^s l_{ij}}{L_i} \quad [3.1]$$

donde s_i es el número de segmentos en el cromosoma i . Utilizando la definición de Guo (1995) para el IBDP entre un abuelo macho (tomado arbitrariamente para indicar $P_G =$

1) y un nieto o nieta, y considerando que una especie en particular posee n cromosomas, obtenemos que el estimador del parámetro \hat{P}_G es igual a:

$$\hat{P}_G = \frac{\sum_{i=1}^n \sum_{j=1}^s l_{ij}}{L} \quad [3.2]$$

Cuando $\hat{P}_G = 0.5$, los genomas de ambos abuelos aparecen en la misma proporción en la gameta que recibe el nieto. En los extremos del espacio paramétrico, $P_G = 1$ indica que el individuo recibió el 100% del genoma del abuelo, mientras que $P_G = 0$ refiere a que el nieto recibió el 100% del genoma de la abuela en la gameta parental. La IBDP entre abuelo y nieto es igual a $0.5 P_G$ (Veller et al, 2020, página 986) y se estima mediante $0.5 \hat{P}_G$. Por otra parte, sea δ_G la fracción de herencia compartida IBD en *exceso o en defecto* de su valor esperado bajo herencia Mendeliana cuando $P_G = 0.5$. La relación paramétrica es igual a:

$$\delta_G = P_G - 0.5 \quad [3.3]$$

Sea $G = S$ el subíndice asociado a la gameta paterna, entonces las fracciones IBDP compartidas entre el nieto y el abuelo (SS) y abuela (DS) son respectivamente iguales a:

$$IBDP_{SS} = 0.5 P_s \quad IBDP_{DS} = 0.5 (1 - P_s) \quad [3.4]$$

En términos de δ_G la expresión [3.4] resulta igual a

$$IBDP_{SS} = 0.5 (0.5 + \delta_s) = 0.25 + 0.5 \delta_s \quad IBDP_{DS} = 0.5 (1 - (0.5 + \delta_s)) = 0.25 - 0.5 \delta_s$$

Para la gameta materna, los IBDP del abuelo y la abuela con el nieto, expresados en función de δ_D , son respectivamente iguales a

$$IBDP_{SD} = 0.5 (0.5 + \delta_D) = 0.25 + 0.5 \delta_D \quad IBDP_{DD} = 0.5 (1 - (0.5 + \delta_D)) = 0.25 - 0.5 \delta_D \quad [3.5]$$

Los valores δ_s y δ_D tienen espacio paramétrico $-0.5 \leq \delta_s, \delta_D \leq 0.5$, y son iguales a dos veces el parámetro de la *regresión ancestral* (β^*) de Cantet et al (2017), i.e. $\delta_s = 2 \beta_s^*$ y $\delta_D = 2 \beta_D^*$ con lo cual

$$IBDP_{SS} = 0.25 + 0.5 \delta_s = 0.25 + \beta_s^* \quad IBDP_{DS} = 0.25 - 0.5 \delta_s = 0.25 - \beta_s^* \quad [3.6]$$

Asimismo

$$\text{IBDP}_{\text{SD}} = 0.25 + 0.5\delta_{\text{D}} = 0.25 + \beta_{\text{D}}^* \quad \text{IBDP}_{\text{DD}} = 0.25 - 0.5\delta_{\text{D}} = 0.25 - \beta_{\text{D}}^* \quad [3.7]$$

Resumiendo, en ausencia de consanguinidad, las relaciones abuelo - nieto en la regresión ancestral (Cantet et al, 2017) son iguales a $0.25 + \beta^*$, sean S o D según el origen de la gameta, mientras que las relaciones abuela - nieto son iguales a $0.25 - \beta^*$, sean S o D.

3.2.1. Esperanza y Varianza de la IBDP y de β^*

El valor esperado y la varianza del IBDP (o proporción del genoma compartido IBD entre dos individuos) son indicadores de la variación potencialmente utilizable en la evaluación genética para mejorar la exactitud de predicción del BV, en presencia del pedigrí y la información genómica (Cantet et al, 2017). Obtendremos primero la varianza del IBDP estimada bajo recombinación, o varianza de la fracción P_G , para luego derivar la varianza de estimación de β^* (Cantet et al, 2017).

El análisis de las expresiones [3.6] y [3.7] sugieren que la relación abuelo-nieto bajo la regresión ancestral, en general son iguales a $\text{IBDP}_G = 0.5 \hat{P}_G = 0.25 + \hat{\beta}_G^*$ para $G = S$ o D . Empleando el operador esperanza tenemos que

$$\text{E}(\text{IBDP}_G) = \text{E}(0.5 \hat{P}_G) = 0.5 (L)^{-1} \text{E} \left(\sum_{i=1}^n \sum_{j=1}^s l_{ij} \right)$$

Cobbs (1978) empleó un *proceso de renovación* para modelar los efectos de la recombinación con interferencia. En este proceso (Ross, 2007, capítulo 7), las longitudes de los segmentos de los abuelos $l_{i1}, l_{i2}, l_{i3}, \dots$, dentro del cromosoma i , son *variables aleatorias independientes* que siguen una distribución Gamma con parámetros

de forma v y tasa de cambio τ tal que $l_{ij} \stackrel{iid}{\square} \text{Gamma}(v, \tau)$, $\text{E}(l_{ij}) = \frac{v_i}{\tau_i}$ y $\text{Var}(l_{ij}) = \frac{v_i}{\tau_i^2}$.

Retomando [2.2], y marginalizando sobre todos los segmentos, tenemos que

$$\text{E} \left(\sum_{j=1}^s l_{ij} \right) \rightarrow \sum_{j=1}^s \text{E}(l_{ij}) = s_i \text{E}(l_i) \text{ y sustituyendo en la expresión anterior, resulta en}$$

$$\begin{aligned} \text{E}(0.5 \hat{P}_G) &= 0.5 (L)^{-1} \text{E} \left(\sum_{i=1}^n \sum_{j=1}^s l_{ij} \right) = 0.5 (L)^{-1} \text{E} \left(\sum_{i=1}^n s_i l_i \right) \\ &= 0.5 (L)^{-1} \text{E} \left(\sum_{i=1}^n s_i l_i \right) = 0.5 (L)^{-1} \sum_{i=1}^n \text{E}(s_i l_i) = 0.5 (L)^{-1} \sum_{i=1}^n s_i \text{E}(l_i) \end{aligned}$$

Asumiendo que $l_i \stackrel{iid}{\square} \text{Gamma}(v_i, \tau_i)$, obtenemos

$$E(0.5 \hat{P}_G) = \frac{1}{2L} \sum_{i=1}^n \frac{s_i v_i}{\tau_i} \quad [3.8]$$

Para obtener la varianza del IBDP entre abuelo nieto, emplearemos el operador varianza en ambos lados de la igualdad:

$$\text{Var}(\text{IBDP}_G) = \text{Var}(0.25 + \hat{\beta}_G^*) \quad [3.9]$$

Dado que $\text{Var}(0.25 + \hat{\beta}_G^*) = \text{Var}(\hat{\beta}_G^*)$ tenemos las siguientes igualdades

$$\text{Var}(0.5 \hat{P}_G) = 0.25 \text{Var}(\hat{P}_G) = \text{Var}(\hat{\beta}_G^*) \quad [3.10]$$

Recordando que n es el número de cromosomas autosómicos, sustituimos en [3.10] con [3.2] de modo de obtener la expresión

$$0.25 \text{Var}\left(\frac{\sum_{i=1}^n \sum_{j=1}^s l_{ij}}{L}\right) \Rightarrow \text{Var}(\hat{\beta}_G^*) = \frac{1}{4L^2} \text{Var}\left(\sum_{i=1}^n \sum_{j=1}^s l_{ij}\right)$$

Por lo tanto

$$\text{Var}(\hat{\beta}_G^*) = \frac{1}{4L^2} \text{Var}\left(\sum_{i=1}^n \sum_{j=1}^s l_{ij}\right) = \frac{1}{4L^2} \sum_{i=1}^n \sum_{j=1}^s \text{Var}(l_{ij}) = \frac{1}{4L^2} \sum_{i=1}^n \sum_{j=1}^s \text{Var}(l_{ij})$$

Como vimos anteriormente, bajo un proceso de renovación los segmentos están distribuidos de modo idéntico, son independientes entre sí y siguen una distribución Gamma. En consecuencia, $\text{Var}(l_{ij}) = s_i^2 \text{Var}(l_i)$ lo que nos lleva a la expresión siguiente

$$\text{Var}(\hat{\beta}_G^*) = \frac{1}{4L^2} \sum_{i=1}^n s_i^2 \text{Var}(l_i) \quad [3.11]$$

Finalmente, reemplazaremos en [3.11] con la varianza de una Gamma de manera tal que

$$\text{Var}(l_i) = \frac{v_i}{\tau_i^2} \text{ y por lo tanto}$$

$$\text{Var}(\hat{\beta}_G^*) = \frac{1}{4L^2} \sum_{i=1}^n \frac{s_i^2 v_i}{\tau_i^2} \quad [3.12]$$

La expresión [3.12] muestra que la varianza del estimador por segmentos del parámetro de la regresión ancestral es función de la sumatoria de las varianzas de cada cromosoma, cada una multiplicada por el número de segmentos (o número de recombinaciones más uno), dividida por cuatro veces la longitud del genoma al

cuadrado. Las expresiones [3.8] y [3.12] muestran al valor esperado y a la varianza del IBDP respectivamente, y serán empleadas en los capítulos siguientes.

3.3. Estimación de los parámetros de la distribución de IBD compartido

Existen distintos métodos para maximizar la función de verosimilitud (ML) respecto de v . Para estimar los parámetros de $l_1, l_2, \dots, l_n \sim \text{Gamma}(v, \tau)$ sin el correspondiente “refinamiento” de la distribución y de la manera convencional (McPeck y Speed, 1996; Broman y Weber, 2000), en el presente trabajo se empleó el algoritmo de ML para la distribución Gamma propuesto Choi y Wette (1969), siendo los parámetros de forma y escala aquellos valores que maximizan el siguiente logaritmo natural de la verosimilitud:

$$\ln L(l_i, v, \tau) = -n v \ln(\tau) - n \ln[\Gamma(v)] + (v-1) \sum_{i=1}^n \ln(l_i) - \frac{1}{\tau} \sum_{i=1}^n l_i \quad [3.12]$$

Los estimadores ML de los parámetros de forma (v) y escala (τ) se calculan a partir de la primera derivada de [3.12] respecto de cada parámetro. Luego de igualar a cero, se itera o repite el proceso sobre $\hat{\tau}_i = \frac{\bar{x}}{\hat{v}_i}$. Mientras que la iteración del parámetro v se

efectúa resolviendo la expresión:

$$\frac{1}{v_i} = \frac{1}{v_{i-1}} + \frac{\overline{\log l} - \log \bar{l} + \log v_{i-1} - \Psi(v_{i-1})}{v_{i-1} (1/v_{i-1} - \Psi'(v_{i-1}))}$$

El valor $\ln(\bar{l})$ es el logaritmo del promedio de las longitudes de los segmentos IBD abuelo-nieto y $\overline{\ln(l)}$ es el promedio de los logaritmos naturales de cada segmento. Mientras que $\Psi(v_{i-1})$ es la función digamma y $\Psi'(v_{i-1})$ es la función trigamma, las que son respectivamente iguales a:

$$\Psi(v_{i-1}) = \frac{1}{\Gamma(v)} \int_0^{\infty} l^{v-1} (\ln l) e^{-l} dl \quad \text{y} \quad \Psi'(v_{i-1}) = \frac{\int_0^{\infty} l^{v-1} (\ln l)^2 e^{-l} dl}{\Gamma(v)} - \Psi^2(v)$$

El ajuste de los datos de las longitudes de los segmentos IBD mediante las distribuciones Gamma por ML fue efectuado con el programa “fitdistplus” (Dellignette-Muller y Dutang, 2014) del software estadístico *R* (*R* Development Core Team, 2013).

3.4. Estimación del parámetro de interferencia (ν)

Como se observó en la Sección 2.4. del Capítulo 2, el efecto de la interferencia en la recombinación sigue un proceso de renovación que resulta en una distribución mezcla de Gammas (ν , τ), donde el parámetro tasa de cambio es igual a $\tau = 100/2 \nu X = 50/ \nu X$ (Broman y Weber, 2002). El proceso tiene esperanza y desvío estándar (S.D.) respectivamente iguales a:

$$E(l_i) = \nu \tau = \nu \frac{100}{2 \nu X} = \frac{50}{X} ;$$

$$S.D.(l_i) = \sqrt{\nu \tau^2} = \sqrt{\nu \left(\frac{100}{2 \nu X} \right)^2} = \sqrt{\nu \left(\frac{100^2}{4 \nu^2 X^2} \right)} = \frac{50}{X \sqrt{\nu}}$$
[3.13]

El valor X es la distancia promedio entre recombinaciones e igual a 1 Morgan. En ausencia de interferencia de las cromátidas, los sitios de recombinación son estimados mediante un proceso de *eliminación de cada recombinación con valor esperado 0.5*, llamado “*thinning*” (McPeck y Speed, 1995; Broman y Weber, 2000) debido a la propiedad de independencía. Cuando ν toma valor 1, no hay interferencia y el proceso es Poisson (Haldane, 1919), mientras que cuando $\nu < 1$ indica interferencia negativa. En este caso los eventos de recombinación ocurren cercanos unos de otros. Alternativamente, cuando $\nu > 1$ la interferencia es positiva, denotando que las recombinaciones se espacian más entre ellas. La distribución Gamma con un único parámetro ν es igual a

$$f(l; \nu) = \frac{(\nu X / 50)^\nu}{\Gamma(\nu)} l^{\nu-1} e^{-(\nu X / 50)l}$$
[3.14]

Los intervalos l_1, l_2, \dots, l_n son variables aleatorias independientes que siguen una distribución $f(l; \nu)$ donde la densidad para l_0 ($g(l; \nu)$), es decir, para la longitud completa de un cromosoma sin ninguna recombinación es igual a (Housworth y Stahl, 2009):

$$g(l; \nu) = \frac{X}{50} (1 - F(l; \nu))$$
[3.15]

Debido a la independencía entre segmentos y dependiendo del *número de recombinaciones* en cada cromosoma (número al que denotaremos como m), la contribución individual de cada segmento IBD al logaritmo de la función de verosimilitud es igual a (Broman y Weber, 2000; Housworth y Stahl, 2009):

$$\ln L(v, l) = \begin{cases} \ln[1 - G(l_0; v)] & \text{si } m = 0 \\ \ln g(l_0; v) + \ln g(l_1; v) & \text{si } m = 1 \\ \ln g(l_0; v) + \sum_{j=1}^{m-1} \ln f(l_j; v) + \ln g(l_m; v) & \text{si } m > 1 \end{cases} \quad [3.16]$$

La expresión [3.16] representa a una mezcla de distribuciones Gamma. La primera línea muestra la distribución de los cromosomas parentales, es decir de aquellos que no mostraron evidencia de recombinación alguna. La segunda línea muestra la distribución de los cromosomas que sufrieron una sola recombinación, mientras que la última línea refleja a la distribución de los cromosomas que tuvieron más de una recombinación. En este último caso, los segmentos delimitados por dos recombinaciones (sumatoria central de los \ln de las longitudes de los segmentos) siguen una distribución distinta a la de aquellos segmentos delimitados por el comienzo (primer \ln) o por el final (último \ln en la tercera línea de la expresión [3.16]), del cromosoma donde ocurrió la primera o última recombinación, respectivamente. Debe notarse además que cuando $m = 0$, el valor del dato correspondiente a esa gameta parental es evaluado como un *dato censurado o truncado a la derecha* con densidad $g(y; v)$, por lo que se dice que es una distribución mezcla de distribuciones Gamma, ya que en cada uno de estos tres casos, se tienen una distribución asociada a la distribución Gamma, ya sea de forma directa o indirecta. El estimador ML de v , \hat{v} , se obtiene optimizando la función del logaritmo de la verosimilitud [3.16] con el método de Brent (1973) que no emplea derivadas, debido a que no existe una solución exacta para v y, por lo tanto, requiere del empleo de métodos numéricos. La varianza de \hat{v} es estimada por -1 dividido el cociente de información de Fisher: $-1/\ln[L''(v, y)]$. La segunda derivada del logaritmo de la función de verosimilitud es aproximadamente igual a $\ln L''(v, y) = [\ln L(v+h, y) - 2\ln L(v, y) + \ln L(v-h, y)]/h^2$, donde h es un valor arbitrario y pequeño. En la presente investigación se empleó $h = 0.01$.

Probabilidad de Ausencia de Recombinación: se calculó empleando la función acumulada (cdf) hasta el punto del cromosoma donde ocurría una recombinación. La función de densidad estacionaria de probabilidad del segmento que comienza en la primera recombinación y continua hasta el final del cromosoma es igual a:

$$\hat{g}(l=100 | v = \hat{v}, \tau = \hat{\tau}) = \frac{\hat{\tau}}{\hat{v}} (1 - G(l=100 | \hat{v}, \hat{\tau})) = \frac{\hat{\tau}}{\hat{v}} \left(1 - \frac{1}{\Gamma(\hat{v})} \int_0^L l^{\hat{v}-1} e^{-(\hat{\tau})^{-1} l} dl \right) \quad [3.17]$$

La probabilidad de ausencia de recombinación es igual al complemento de la función de distribución acumulada del segmento IBD, a partir de un evento de recombinación en [3.6]. La expresión final es igual a $P_0 = 1 - G(100 | \hat{v}, \hat{\tau})$ (Housworth y Stahl, 2009), ó en mayor detalle igual a

$$P_0 = 1 - \frac{\hat{\tau}}{\hat{v}} \int_0^{100} \left(1 - \frac{1}{\Gamma(\hat{\alpha})} \int_0^L l^{\hat{\alpha}-1} e^{-\frac{l}{\hat{\tau}}} dl \right) dL \quad [3.18]$$

3.5. Prueba del cociente de verosimilitud

Luego de estimar v , se realizó la prueba del cociente de verosimilitud (*LRT*) para evaluar en qué medida el parámetro de forma difería de 1, es decir $H_0: v = 1$. Este valor para el parámetro de forma indica la ausencia de interferencia meiótica (Broman y Weber, 2000; Housworth y Stahl, 2009). Asimismo el *LRT* se utilizó para probar si había diferencia entre los parámetros estimados en machos respecto de las hembras, en cada cromosoma. La hipótesis nula involucra un proceso Poisson (Haldane 1919) y la alternativa ($v \neq 1$) significa la presencia de interferencia. El estadístico de la prueba de hipótesis es igual a

$$LRT = -2 \log \left(\frac{\sup_{\theta \in \Theta_0} L(v)}{\sup_{\theta \in \Theta} L(v)} \right) = -2 \log \left(\frac{L(v=1; l)}{L(\hat{v}; l)} \right) = 2 \log [L(\hat{v}; l) - L(v=1; l)] \quad [3.19]$$

La expresión $L(v=1; l)$ es el valor del logaritmo de la verosimilitud bajo H_0 ($v = 1$); $L(\hat{v}; l)$ es el valor del logaritmo de la verosimilitud bajo H_a (\hat{v}). En la prueba del cociente de máxima verosimilitud $LRT \square \chi_{q,df}^2$, el espacio paramétrico se asocia con dos grados de libertad para H_a , mientras que para $H_0: v = 1$ el espacio de la hipótesis tiene dimensión igual a un grado de libertad y sigue una distribución exponencial. En consecuencia, la prueba *LRT* tiene $2 - 1 = 1$ grado de libertad. Para el percentil del 95%, el valor tabulado bajo hipótesis nula es $\chi_{0.95,1}^2 = 3.841459$.

La interpretación de la prueba de hipótesis respecto al supuesto de presencia o ausencia de interferencia es la siguiente:

H_0 : No hay evidencias para suponer la presencia de interferencia cuando $LRT \leq \chi_{0.95,1}^2$.

H_A : Existen evidencias para suponer la presencia de interferencia; esto ocurre cuando $LRT > \chi_{0.95,1}^2$.

Capítulo 4: Resultados

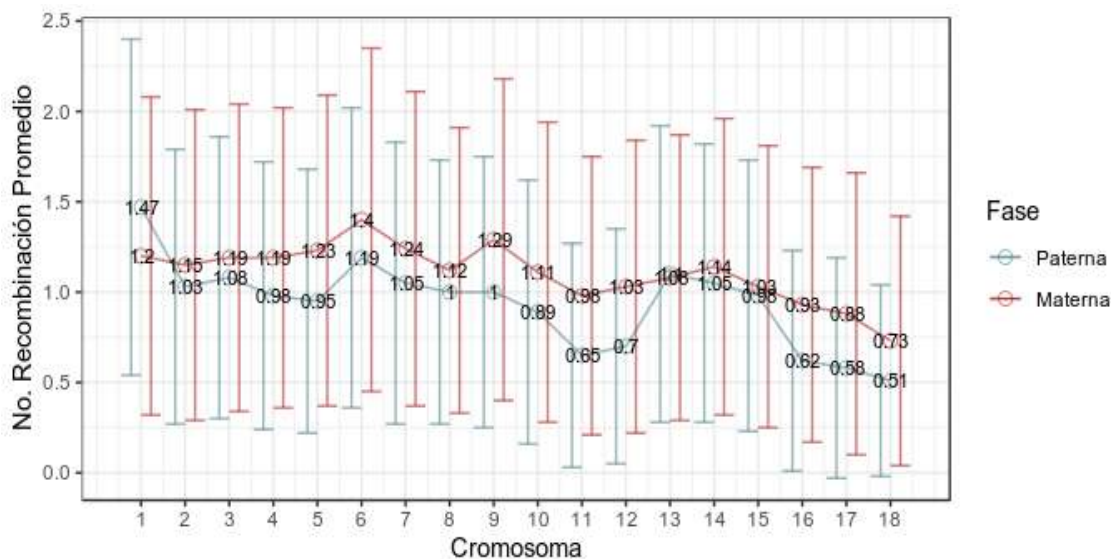
Capítulo 4

Resultados

4.1. Distribución del número de recombinaciones por cromosoma.

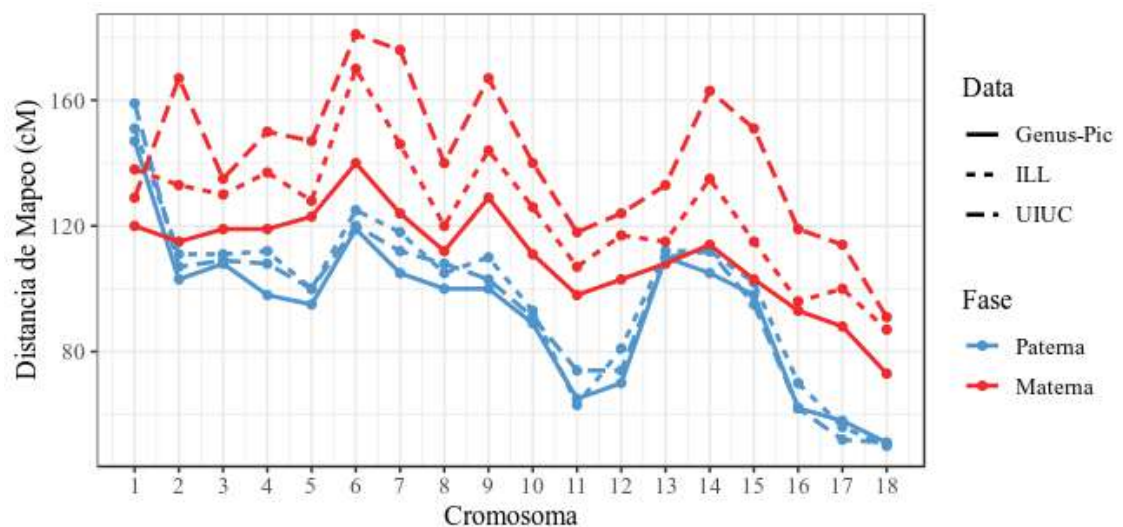
El Cuadro 1 muestra las frecuencias de los cromosomas detectados como parentales o recombinantes por número de cromosoma (1 al 18) y la Figura 7 muestra las distribuciones de los números de recombinación en cada cromosoma. El Cuadro 1 muestra que para las fases paternas, en los cromosomas 11, 12, 16, 17 y 18 se encontró una alta frecuencia de cromosomas parentales (donde no se detectó recombinación alguna) con frecuencias superiores al 40%, mientras que en las fases maternas hubo una alta frecuencia de cromosomas parentales en los cromosomas 16, 17 y 18, aunque en menor frecuencia que en las fases paternas dado que no superan el 37% de cromosomas parentales. Esto además indica que en promedio, las fases maternas presentan una mayor cantidad de recombinaciones (ver Figura 5). Mientras que los cromosomas restantes presentaron una alta tasa de recombinación, con frecuencias de al menos una recombinación superior al 70%.

Figura 5. Número de recombinaciones promedio por cromosoma, tanto en fase paterna y materna.



Al detectar además los eventos de recombinación por cromosoma, se calculan los valores promedios de recombinación por cromosoma tanto en la fase materna, así como en la paterna, tal como se muestran en la Figura 5, donde se puede ver que por lo general hay mayores valores de promedios de recombinación por cromosomas en las fases maternas, salvo en el cromosoma 1 y 13, donde hay mayores valores de recombinaciones promedios en las fases paternas para los datos de estos cromosomas. Expresando estos valores del promedio de recombinaciones (distancia genética) por cromosomas en centimorgans (cM) podemos comparar resultados obtenidos con resultados reportados en literatura, donde se encuentran valores parecidos a los reportados por Tortereau *et al.* (2012) que para datos de dos bases de datos distintas (ILL y UIUC) encuentran valores a los cuales, los encontrados en este trabajo se asemejan, tanto en la fase paterna y materna y cuya tendencia es posible ver en la Figura 6, donde los valores encontrados en este trabajo se trazan en una línea sólida y muestran incluso que los valores de las fases paternas tienden a ser más próximos entre sí que en las fases materna, recordando además la mayor presencia de datos en la base de datos paterna.

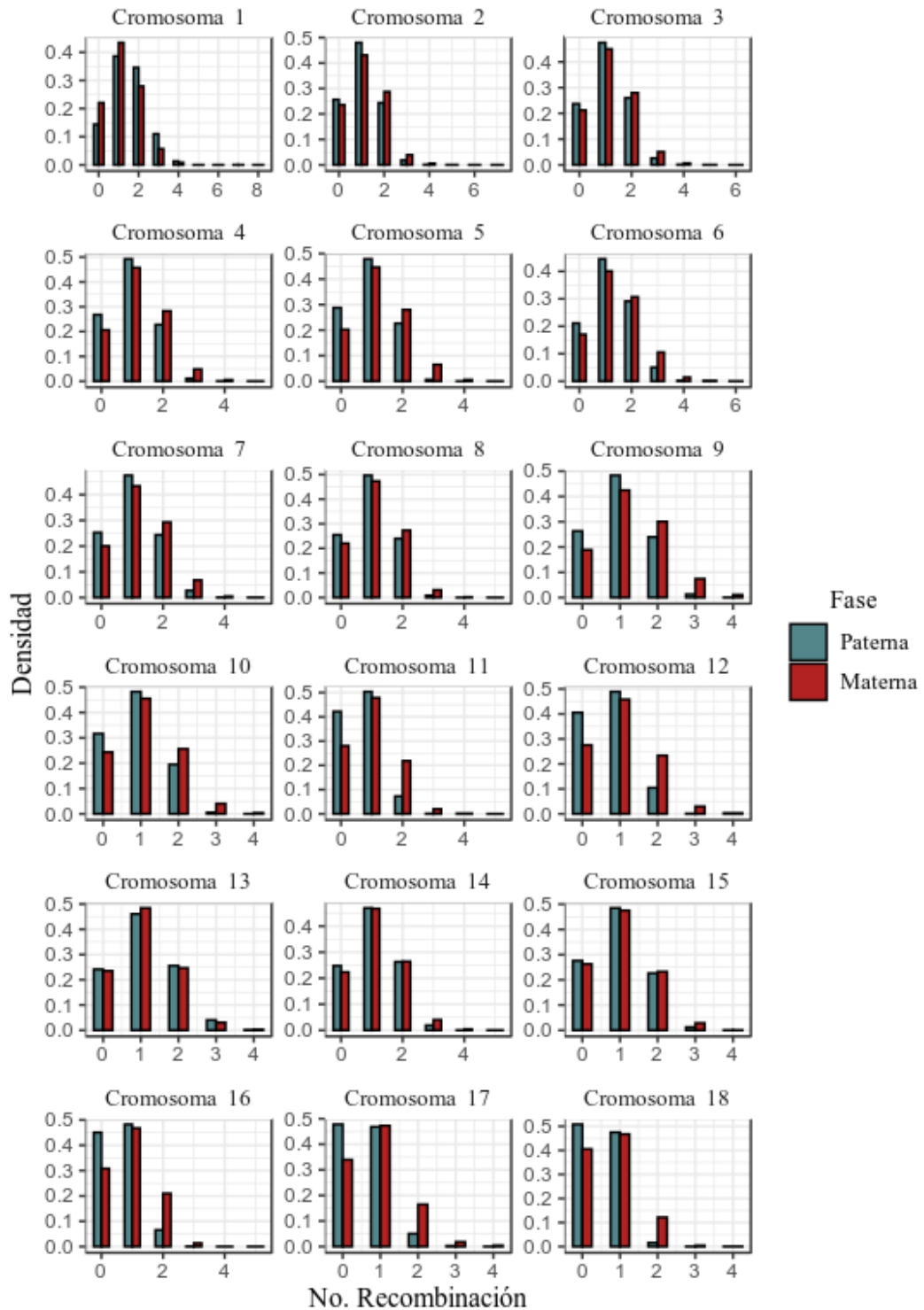
Figura 6. Comparación de las distancias genéticas obtenidas por cromosoma y sexo. Se comparan los datos analizados en este trabajo (Genus – Pic) con respecto a valores de dos bases de datos distintas presentados en Tortereau *et al.* 2012



Cuadro 1: Frecuencia de cromosomas parentales y recombinantes.

| Cromosoma | Fase Paterna | | Fase Materna | |
|------------------|---|--|---|--|
| | Cromosomas Parentales (Proporción) | Cromosomas Recombinantes (Proporción) | Cromosomas Parentales (Proporción) | Cromosomas Recombinantes (Proporción) |
| 1 | 939 (0.1463) | 5479 (0.8537) | 720 (0.2349) | 2345 (0.7651) |
| 2 | 1637 (0.2551) | 4781 (0.7449) | 716 (0.2336) | 2349 (0.7664) |
| 3 | 1506 (0.2347) | 4912 (0.7653) | 648 (0.2114) | 2417 (0.7886) |
| 4 | 1698 (0.2646) | 4720 (0.7354) | 607 (0.1980) | 2458 (0.8020) |
| 5 | 1838 (0.2864) | 4580 (0.7136) | 565 (0.1843) | 2500 (0.8157) |
| 6 | 1337 (0.2083) | 5081 (0.7916) | 509 (0.1661) | 2556 (0.8339) |
| 7 | 1604 (0.2499) | 4814 (0.7501) | 593 (0.1935) | 2472 (0.8065) |
| 8 | 1620 (0.2524) | 4798 (0.7476) | 665 (0.2170) | 2400 (0.7830) |
| 9 | 1679 (0.2616) | 4739 (0.7384) | 533 (0.1740) | 2532 (0.8261) |
| 10 | 2038 (0.3175) | 4380 (0.6824) | 686 (0.2238) | 2379 (0.7762) |
| 11 | 2725 (0.4246) | 3693 (0.5754) | 737 (0.2405) | 2328 (0.7595) |
| 12 | 2628 (0.4095) | 3790 (0.5905) | 720 (0.2349) | 2345 (0.7651) |
| 13 | 1560 (0.2431) | 4858 (0.7569) | 705 (0.2300) | 2360 (0.7700) |
| 14 | 1592 (0.2480) | 4826 (0.7519) | 665 (0.2170) | 2400 (0.7830) |
| 15 | 1763 (0.2747) | 4655 (0.7253) | 804 (0.2623) | 2261 (0.7377) |
| 16 | 2898 (0.4515) | 3520 (0.5485) | 832 (0.2714) | 2233 (0.7285) |
| 17 | 3095 (0.4822) | 3323 (0.5177) | 908 (0.2962) | 2157 (0.7038) |
| 18 | 3280 (0.5111) | 3138 (0.4889) | 1146 (0.3739) | 1919 (0.6261) |

Figura 7. Distribuciones del número de recombinaciones encontradas por cromosoma.



4.2. Parámetros de interferencia estimados bajo el proceso de renovación empleando una distribución Gamma.

Las estimaciones de los parámetros de forma y la tasa de cambio por cromosoma, para las distribuciones Gamma (Broman y Weber 2000; Housworth y Stahl 2009), se muestran en el Cuadro 6. Estos valores fueron proporcionales a los parámetros de forma (v) debido a que en la Gamma, el parámetro “tasa de cambio” es igual a $2vX/100$, donde X es la distancia genética. Todas las estimaciones del parámetro de forma en los dieciocho cromosomas mostraron desvíos estándar pequeños (< 0.42). En la Figura 8 se observan las distribuciones correspondientes a los dieciocho cromosomas (parámetros descritos en el Cuadro 6).

En los animales analizados, los parámetros de forma y tasa de cambio estimados bajo la distribución Gamma, tanto para la fase paterna como la materna, se muestran en el Cuadro 2 y 3, respectivamente, junto a sus desvíos estándar. Dentro de la fase paterna, los mayores valores del parámetro de forma se observan en los cromosomas 1, 6, 3 y 14, con valores respectivos de 1.647, 1.238, 1.082 y 0.989. Los más pequeños corresponden a los cromosomas 18, 17, 16 y 12, y son iguales a 0.305, 0.360, 0.414 y 0.532, respectivamente. Los desvíos estándar de las estimaciones de la forma y tasa de cambio fueron de magnitud reducida (< 0.026). Por otro lado, en la fase materna, se encontraron los valores más altos en los cromosomas 5, 4, 12, 7 y 11 (1.467, 1.403, 1.373, 1.226 y 1.215, respectivamente), y los menores en los cromosomas 18, 13, 1 y 13 (0.661, 0.842, 0.801 y 0.842, respectivamente). Por otra parte, se encontraron los valores más elevados para la fase materna en los cromosomas 5, 4, 12, 7 y 11 (1.467, 1.403, 1.373, 1.226 y 1.215, respectivamente), mientras que tiene los valores menores a los cromosomas 18, 13, 1 y 13 (0.661, 0.842, 0.801 y 0.842 respectivamente).

Cuadro 2. Estimación del parámetro y su desvío estándar para la forma (ν) y la tasa de cambio (τ) de distribuciones Gamma, en fase paterna.

| Cromosoma | Forma (ν) | SD Forma | Tasa de Cambio | SD (Tasa de cambio) |
|------------------|---------------------------------|-----------------|-----------------------|----------------------------|
| 1 | 1.647 | 0.026 | 0.033 | 0.001 |
| 2 | 0.570 | 0.016 | 0.019 | <0.001 |
| 3 | 1.082 | 0.019 | 0.022 | <0.001 |
| 4 | 0.957 | 0.017 | 0.019 | <0.001 |
| 5 | 0.886 | 0.016 | 0.018 | <0.001 |
| 6 | 1.238 | 0.020 | 0.025 | <0.001 |
| 7 | 1.027 | 0.018 | 0.021 | <0.001 |
| 8 | 0.927 | 0.016 | 0.019 | <0.001 |
| 9 | 0.984 | 0.017 | 0.020 | <0.001 |
| 10 | 0.814 | 0.015 | 0.016 | <0.001 |
| 11 | 0.464 | 0.009 | 0.009 | <0.001 |
| 12 | 0.532 | 0.011 | 0.011 | <0.001 |
| 13 | 0.934 | 0.015 | 0.019 | <0.001 |
| 14 | 0.989 | 0.017 | 0.020 | <0.001 |
| 15 | 0.860 | 0.015 | 0.017 | <0.001 |
| 16 | 0.414 | 0.008 | 0.008 | <0.001 |
| 17 | 0.360 | 0.007 | 0.007 | <0.001 |
| 18 | 0.305 | 0.006 | 0.006 | <0.001 |

Cuadro 3. Estimación del parámetro y su desvío estándar para la forma (ν) y la tasa de cambio (τ) de distribuciones Gamma, en fase materna.

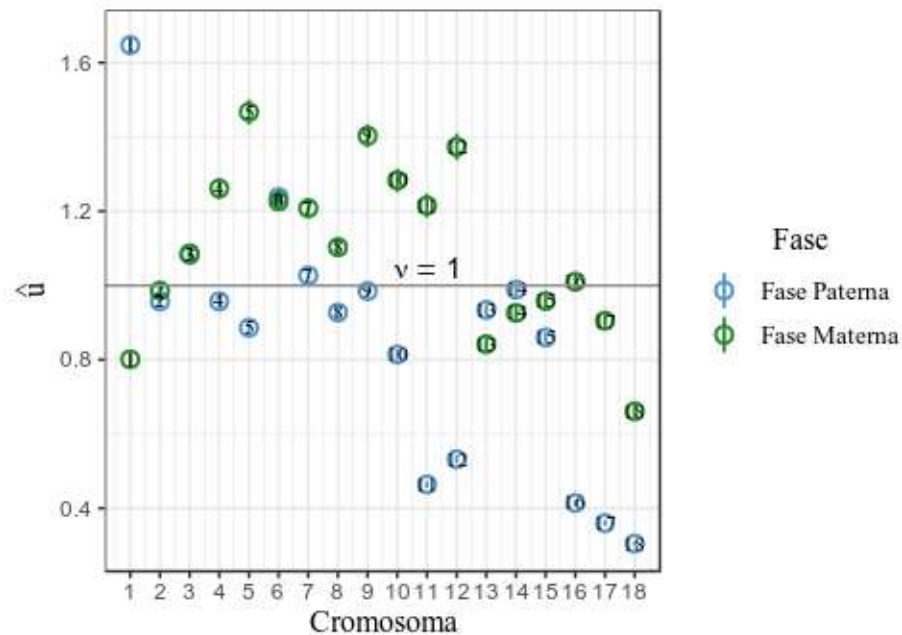
| Cromosoma | Forma (ν) | SD Forma | Tasa de cambio | SD (Tasa de cambio) |
|-----------|-----------------|----------|----------------|---------------------|
| 1 | 0.801 | 0.018 | 0.016 | <0.001 |
| 2 | 0.986 | 0.022 | 0.020 | <0.001 |
| 3 | 1.086 | 0.024 | 0.022 | <0.001 |
| 4 | 1.261 | 0.029 | 0.025 | 0.001 |
| 5 | 1.467 | 0.034 | 0.029 | 0.001 |
| 6 | 1.226 | 0.026 | 0.025 | 0.001 |
| 7 | 1.208 | 0.027 | 0.024 | 0.001 |
| 8 | 1.103 | 0.026 | 0.022 | 0.001 |
| 9 | 1.403 | 0.031 | 0.028 | 0.001 |
| 10 | 1.283 | 0.032 | 0.026 | 0.001 |
| 11 | 1.215 | 0.032 | 0.024 | 0.001 |
| 12 | 1.373 | 0.036 | 0.027 | 0.001 |
| 13 | 0.842 | 0.019 | 0.017 | <0.001 |
| 14 | 0.926 | 0.021 | 0.019 | <0.001 |
| 15 | 0.958 | 0.024 | 0.019 | <0.001 |
| 16 | 1.01 | 0.026 | 0.020 | 0.001 |
| 17 | 0.905 | 0.024 | 0.018 | <0.001 |
| 18 | 0.661 | 0.018 | 0.013 | <0.001 |

En la Figura 7 se observan las distribuciones gráficas, para los segmentos correspondientes a cada uno de los dieciocho cromosomas, bajo una distribución Gamma con los parámetros que se describen en el Cuadro 2 y 3.

4.3. Presencia de Interferencia en el proceso de recombinación

La interferencia es un proceso biológico complejo debido al tamaño y magnitud de los eventos involucrados durante el crossover. Se conoce como “bivalente” a la asociación entre dos cromosomas replicados que intercambian ADN durante la meiosis en al menos un sitio o “quiasma”. Este es el momento en que se produce la interferencia entre las cuatro cromátidas alineadas pero fuertemente unidas. Para definir una distribución de probabilidad asociada con la hipótesis nula de ausencia de interferencia, se muestrean al azar dos de estas cromátidas pertenecientes a distintos cromosomas, con igual probabilidad entre los cuatro pares posibles, e igual a 0.25.

Figura 8. Estimaciones de los parámetros de forma por cromosoma. La línea horizontal representa el valor de la forma en ausencia de interferencia ($v = 1$).



La distribución resultante es una Gamma “refinada” (McPeck y Speed, 1996). Para cuantificar la interferencia emplearemos la magnitud del parámetro de forma de la distribución Gamma (Foss 1993, Broman y Weber 2000; Housworth y Stahl 2009). Recordamos que cuando no existe interferencia, la forma de la Gamma “refinada” es igual a 1 ($v = 1$) y el estadístico de la prueba del *LRT* es superior a una chi-cuadrado con $\alpha = 0.05$, cuyo valor es aproximadamente igual a $\chi_{0.95,1}^2 \approx 3.84$.

Los valores *LRT* en el Cuadro 4 sugieren interferencia ($v \neq 1$, $P > 0.05$) para la mayoría de los cromosomas ($LRT > \chi_{0.95,1}^2$). De manera opuesta, para los cromosomas 7, 9 y 14, el *LRT* no mostró interferencia ($v = 1$, $LRT \leq \chi_{0.95,1}^2$). En la Figura 8 se observan los valores estimados de la forma. Para los cromosomas analizados en la vía paterna, se observó interferencia positiva ($v > 1$) en 1, 3 y 6, y negativa ($v < 1$) en 2, 4, 5, 8, 10, 11, 12, 13, 15, 16, 17 y 18. Para la fase materna (Cuadro 5) los estadísticos calculados bajo el supuesto de interferencia mostraron un valor-*p* tal que $P < 0.05$ ($v \neq 1$) en todos los cromosomas, salvo en 2, 15 y 16. La interferencia fue positiva en los cromosomas 3, 4, 5, 6, 7, 8, 9, 10, 11 y 12 ($v > 1$) y negativa ($v < 1$) en 1, 13, 14, 17 y 18.

Cuadro 4. Determinación del nivel de interferencia en los cromosomas en fases paternas.

| Cromosoma | Log-Verosimilitud (\hat{v}) | Log-Verosimilitud ($v = 1$) | <i>LRT</i> | Valor-<i>p</i> |
|------------------|---|---|-------------------|-----------------------|
| 1 | -49052.96 | -49542.51 | 979.12 | < 0.001 |
| 2 | -38691.24 | -38694.47 | 6.46 | 0.011 |
| 3 | -40033.70 | -40044.12 | 20.84 | < 0.001 |
| 4 | -37607.70 | -37610.84 | 6.28 | 0.012 |
| 5 | -36730.71 | -36754.11 | 46.80 | < 0.001 |
| 6 | -42740.19 | -42821.66 | 162.93 | < 0.001 |
| 7 | -39233.16 | -39234.33 | 2.348 | 0.125 |
| 8 | -38027.52 | -38037.25 | 19.46 | < 0.001 |
| 9 | -38044.68 | -38045.07 | 0.79 | 0.373 |
| 10 | -35088.61 | -35161.92 | 124.62 | < 0.001 |
| 11 | -28398.72 | -29192.17 | 1586.91 | < 0.001 |
| 12 | -29792.17 | -30326.65 | 1068.97 | < 0.001 |
| 13 | -40297.75 | -40306.22 | 16.95 | < 0.001 |
| 14 | -39261.52 | -39261.71 | 0.382 | 0.537 |
| 15 | -37342.49 | -37380.03 | 75.09 | < 0.001 |
| 16 | -27294.01 | -28335.44 | 2082.84 | < 0.001 |
| 17 | -25883.45 | -27263.54 | 2760.19 | < 0.001 |
| 18 | -23800.46 | -25522.69 | 3444.46 | < 0.001 |

Cuadro 5. Determinación del nivel de interferencia en los cromosomas en fases maternas.

| Cromosoma | Log-Verosimilitud (\hat{v}) | Log-Verosimilitud ($v = 1$) | <i>LRT</i> | Valor - <i>p</i> |
|------------------|---|---|-------------------|-------------------------|
| 1 | -19758.64 | -19810.35 | 103.412 | < 0.001 |
| 2 | -20260.04 | -20260.23 | 0.372 | 0.542 |
| 3 | -20652.68 | -20659.25 | 13.152 | < 0.001 |
| 4 | -20900.04 | -20948.74 | 97.412 | < 0.001 |
| 5 | -21477.72 | -21605.96 | 256.49 | < 0.001 |
| 6 | -23331.65 | -23378.11 | 92.926 | < 0.001 |
| 7 | -21493.13 | -21527.72 | 69.181 | < 0.001 |
| 8 | -19895.77 | -19904.23 | 16.926 | < 0.001 |
| 9 | -22382.64 | -22493.99 | 222.71 | < 0.001 |
| 10 | -20015.11 | -20064.63 | 99.026 | < 0.001 |
| 11 | -18941.64 | -18969.26 | 55.234 | < 0.001 |
| 12 | -19448.85 | -19520.85 | 144.01 | < 0.001 |
| 13 | -19187.06 | -19215.72 | 57.32 | < 0.001 |
| 14 | -20000.23 | -20005.94 | 11.429 | 0.001 |
| 15 | -18576.51 | -18578.06 | 3.089 | 0.079 |
| 16 | -18210.25 | -18210.33 | 0.16 | 0.689 |
| 17 | -17573.24 | -17580.49 | 14.512 | < 0.001 |
| 18 | -15483.27 | -15597.1 | 227.646 | < 0.001 |

4.4. Probabilidad de ausencia de recombinación

La probabilidad de ausencia de recombinación (P_0) se asocia con el evento en el cual una gameta recibida, sea por vía materna o por vía paterna, no sufre recombinación alguna, siendo originaria del abuelo o de la abuela correspondientes. En la presente investigación P_0 fue calculada mediante la expresión [3.18], que es proporcional a la inversa de la distancia física en Mbp ($R^2 = 0.60$, $P_0 = 0.3234 - 0.0011 L$). La Figura 9 emplea los valores del Cuadro 6 y muestra que las P_0 disminuyeron a medida que aumentó la longitud del cromosoma. Nótese que en las fases paternas hubo una mayor posibilidad de recombinación en los cromosomas más largos. El cromosoma 18 tiene la

menor longitud (55.8 Mpb) y presentó la mayor $P_0 = 0.364$. De manera similar, los cromosomas 17 (61.8 Mpb), 16 (79.3 Mpb), 12 (61.0 Mpb) y 11 (79.0 Mpb), mostraron P_0 iguales a 0.245, 0.298, 0.273 y 0.241, respectivamente. De modo opuesto, los cromosomas de mayor longitud 1 (274.3 Mpb), 6 (170.8 Mpb), 3 (132.7 Mpb) y 7 (121.8 Mpb) tuvieron menores P_0 : 0.089, 0.123, 0.139 y 0.147, respectivamente. Mientras que en las fases maternas hubo un poco relación entre la longitud del cromosoma y los valores de no recombinación estimados ($R^2 = 0.05$, $P_0 = 0.1114 + 0.0001 L$).

Cuadro 6. Probabilidad estimada de ausencia de recombinación por cromosoma.

| Cromosoma | Longitud del cromosoma (Mpb) | P_0 (Fase Paterna) | P_0 (Fase Materna) |
|------------------|-------------------------------------|--|--|
| 1 | 274.3 | 0.073 | 0.170 |
| 2 | 151.4 | 0.142 | 0.137 |
| 3 | 132.7 | 0.124 | 0.124 |
| 4 | 130.8 | 0.142 | 0.104 |
| 5 | 104.5 | 0.154 | 0.085 |
| 6 | 170.8 | 0.106 | 0.107 |
| 7 | 121.8 | 0.131 | 0.109 |
| 8 | 138.9 | 0.147 | 0.121 |
| 9 | 139.4 | 0.138 | 0.091 |
| 10 | 69.3 | 0.168 | 0.101 |
| 11 | 79.0 | 0.273 | 0.108 |
| 12 | 61.0 | 0.245 | 0.093 |
| 13 | 208.2 | 0.146 | 0.162 |
| 14 | 141.7 | 0.137 | 0.147 |
| 15 | 140.2 | 0.159 | 0.142 |
| 16 | 79.3 | 0.297 | 0.134 |
| 17 | 61.8 | 0.327 | 0.151 |
| 18 | 55.8 | 0.364 | 0.204 |

Compararemos ahora las probabilidades estimadas mediante la ecuación [3.6] con aquellas observadas en el Cuadro 7, con respecto a la proporción de gametas parentales por cromosoma (Cuadro 1 y Figura 10). Se puede inferir que las P_0 estimadas mediante

el proceso de renovación tienden a ser menores comparadas con la proporción de gametas parentales observadas por cromosoma. Esta diferencia entre estimaciones y observaciones fue mayor en los cromosomas de menor longitud física, tales como en 11, 12, 16, 17 y 18.

Figura 9. Probabilidades de ausencia de recombinación para cada cromosoma en cada asociada con su distancia física en Mbp.

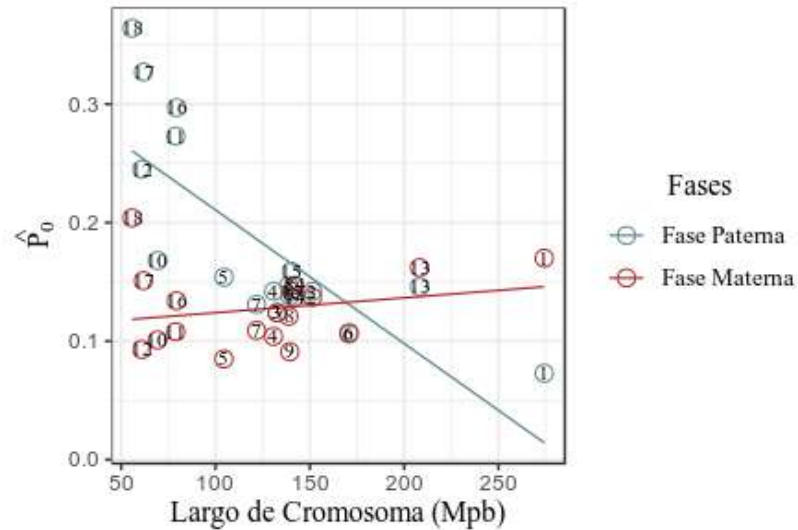
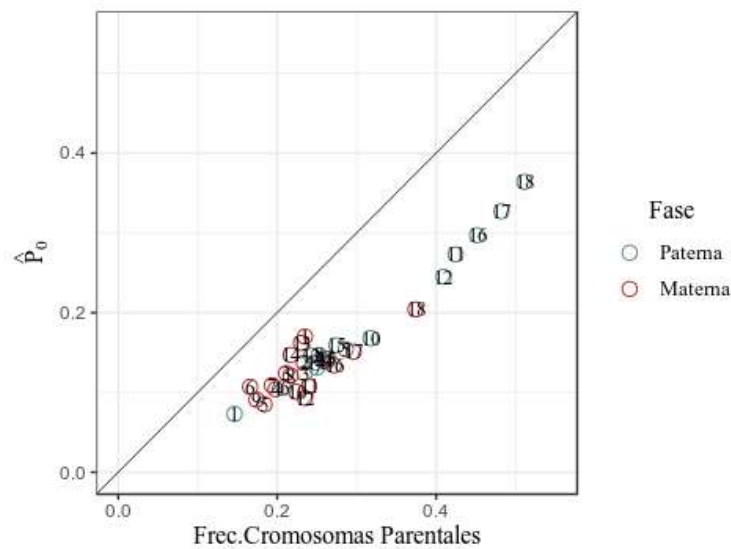


Figura 10. Comparación de las proporciones de gametas no recombinantes con respecto a las probabilidades de no recombinación estimadas.



4.5. Proporción de Genoma IBD compartido entre abuelos y nietos: IBDP

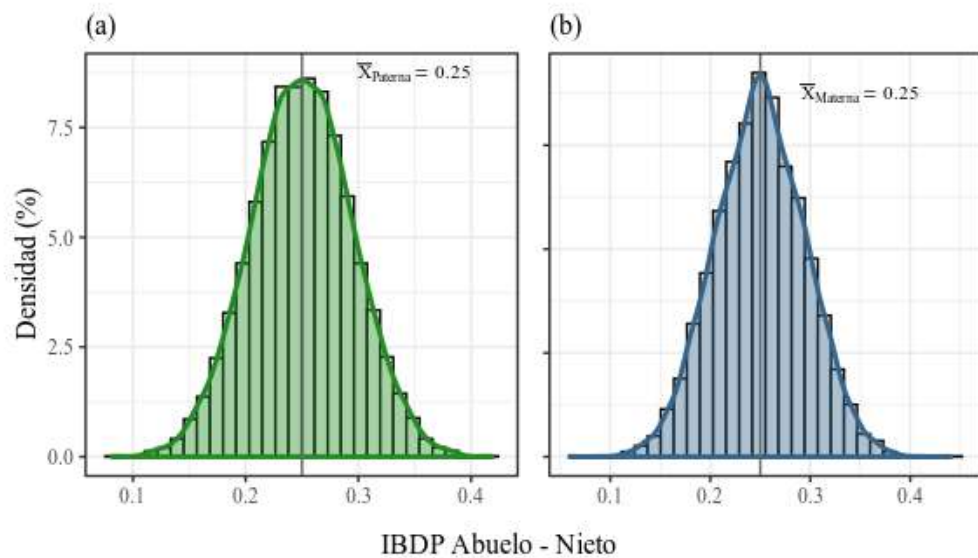
Los estadísticos de IBDP (media, mediana, varianza y desvío estándar) se muestran en el Cuadro 9, y las distribuciones de IBDP entre abuelos y nietos se observan en la Figura 11. Los cálculos fueron realizados a partir de la fórmula [3.2] descrita en la Sección 3.2, luego de detectar los sitios de recombinación, la distancia entre el inicio de la gameta hasta el primer evento de recombinación, las distancias entre eventos y la distancia entre el último evento hasta el final de la gameta.

Cuadro 7. Valores de las medias (\bar{x}), medianas (\tilde{x}), desvíos estándares (DE), coeficientes de variación (CV), asimetría, curtosis y valores máximos y mínimos observadas en el IBDP entre abuelo – nieto.

| Fase | \bar{x} | \tilde{x} | DE | CV (%) | Asimetría | Curtosis | V. Max. | V. Min. |
|---------|-----------|-------------|--------|--------|-----------|----------|---------|---------|
| Paterna | 0.2500 | 0.2500 | 0.0449 | 0.1798 | < 0.0001 | 2.9026 | 0.4184 | 0.0816 |
| Materna | 0.2500 | 0.2500 | 0.0450 | 0.1818 | < 0.0001 | 2.9127 | 0.4398 | 0.0602 |

El Cuadro 7 muestra que los promedios de IBDP entre el abuelo o la abuela con el nieto fueron muy similares al valor esperado de la relación de parentesco bajo equilibrio gamético, e igual a dos veces la coancestría de Malecot ($2\theta = 0.25$). Además, se puede observar la similitud entre medias y medianas, hecho que explica la simetría en las distribuciones de la Figura 11 para el IBDP entre abuelo(a) – nieto. Al observar los estadísticos en el Cuadro 7, se puede notar cierta leve asimetría ($P < 0.01$) en las fracciones IBDP para las fases paterna y materna. La evidencia de alguna asimetría en los datos (< 0.0001) es escasa. Se puede notar además que las distribuciones de las relaciones entre abuelos y nietos mostraron curtosis platicúrticas (curtosis < 3), sugiriendo que la menor concentración de valores ocurrió en el centro de la distribución.

Figura 11. Distribución de las IBDP observadas ($0.5 P_G$) bajo el método de Cheung (2007) para abuelo – nieto en fase paternas (a) y maternas (b).



4.5.1. Valores Esperados y Varianzas del IBDP y de β^*

Se estimaron $E(\text{IBDP})$ y $\text{Var}(\text{IBDP})$ como funciones de los parámetros estimados por ML de la distribución mezcla de Gammas (ver Cuadro 8) calculadas por cromosoma, mediante la fórmula [3.7] para el valor esperado y [3.11] para la varianza. Los estadísticos IBDP entre abuelo(a) – nieto en fases paternas y maternas se observan en el Cuadro 9. El promedio de IBDP fue similar al observado en el Cuadro 7, y otro tanto ocurrió con la mediana, desvío estándar, coeficiente de variación y valores máximos y mínimos. En la Figura 11 se observan los histogramas de la $E(\text{IBDP})$ para la relación abuelo(a) – nieto de ambas fases.

Cuadro 8. Valores de los parámetros de forma y tasa de cambio de una distribución gamma estimados por ML, tanto en fase materna como paterna.

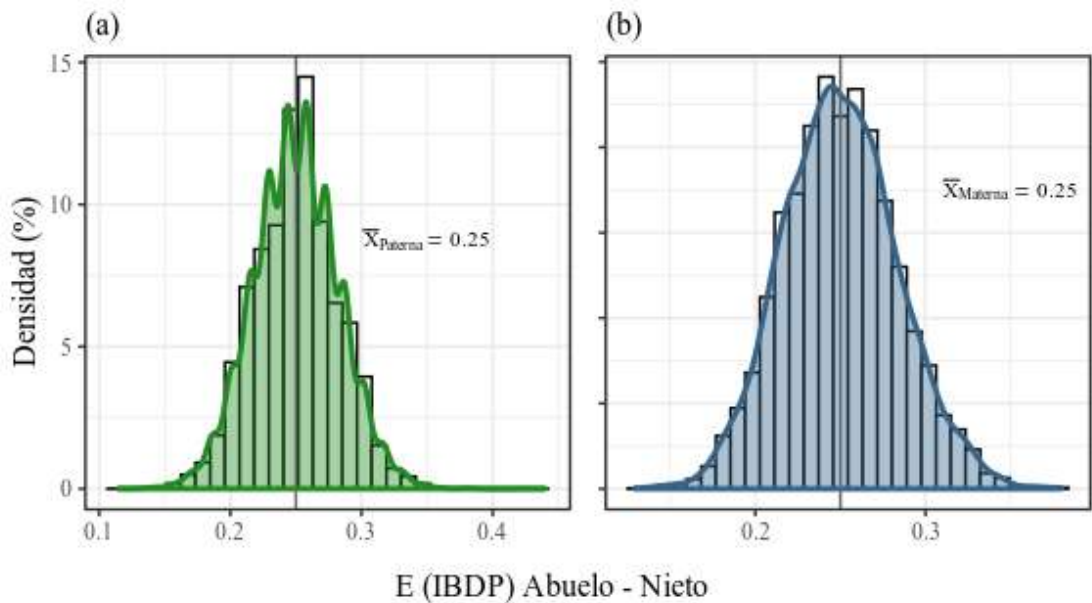
| Cromosoma | Fase Paterna | | | | Fase Materna | | | |
|-----------|--------------|-------------|----------------------|-------------------------|--------------|-------------|-------------------|-------------------------|
| | Forma | SD Forma | Tasa de cambio | SD tasa de cambio | Forma | SD Forma | Tasa de cambio | SD tasa de cambio |
| 1 | 1.0356 | 0.0255 | 0.0255 | 0.0003 | 0.6550 | 0.0100 | 0.0100 | <0.0001 |
| 2 | 1.1365 | 0.0231 | 0.0229 | 0.0003 | 0.9050 | 0.0141 | 0.0200 | <0.0001 |
| 3 | 1.1449 | 0.0238 | 0.0238 | 0.0003 | 1.0021 | 0.0150 | 0.0233 | <0.0001 |
| 4 | 1.2322 | 0.0244 | 0.0244 | 0.0003 | 1.0873 | 0.0160 | 0.0274 | 0.0010 |
| 5 | 1.2158 | 0.0237 | 0.0236 | 0.0003 | 1.3714 | 0.0211 | 0.0357 | 0.0010 |
| 6 | 1.0870 | 0.0237 | 0.0236 | 0.0003 | 0.9177 | 0.0130 | 0.0210 | <0.0001 |
| 7 | 1.1433 | 0.0234 | 0.0234 | 0.0003 | 1.0061 | 0.0150 | 0.0242 | <0.0001 |
| 8 | 1.0870 | 0.0218 | 0.0218 | 0.0003 | 1.0500 | 0.0160 | 0.0262 | <0.0001 |
| 9 | 1.2458 | 0.0250 | 0.0249 | 0.0003 | 1.1280 | 0.0170 | 0.0292 | 0.0010 |
| 10 | 1.3879 | 0.0262 | 0.0261 | 0.0004 | 1.5491 | 0.0251 | 0.0391 | 0.0010 |
| 11 | 1.6280 | 0.0218 | 0.0270 | 0.0004 | 1.6577 | 0.0270 | 0.0491 | 0.0010 |
| 12 | 1.7204 | 0.0292 | 0.0292 | 0.0004 | 2.0100 | 0.0330 | 0.0631 | 0.0010 |
| 13 | 0.8640 | 0.0181 | 0.0182 | 0.0002 | 0.7050 | 0.0110 | 0.0141 | <0.0001 |
| 14 | 0.9729 | 0.0200 | 0.0199 | 0.0001 | 0.7290 | 0.0110 | 0.0171 | <0.0001 |
| 15 | 0.9474 | 0.0187 | 0.0188 | 0.0003 | 1.1420 | 0.0180 | 0.0271 | 0.0010 |
| 16 | 1.4501 | 0.0234 | 0.0236 | 0.0004 | 1.3520 | 0.0220 | 0.0421 | 0.0010 |
| 17 | 1.4703 | 0.0231 | 0.0234 | 0.0003 | 1.8811 | 0.0322 | 0.0641 | 0.0010 |
| 18 | 1.9309 | 0.0291 | 0.0290 | 0.0004 | 1.6532 | 0.0290 | 0.0583 | 0.0010 |

El Cuadro 9 muestra los valores de IBDP para las fases paterna y materna, observándose una leve asimetría de cola derecha. Con respecto a la curtosis, las distribuciones de IBDP de ambos abuelos se mostraron ligeramente leptocúrticas (curtosis > 3). La Var(IBDP) en el Cuadro 10 muestra ligeras asimetrías positivas abuelo – nieto y abuela – nieto (0.053 y 0.136, respectivamente), sugiriendo que la cola derecha fue levemente más elevada que la izquierda (Figura 12). Asimismo, se encontró que las distribuciones fueron ligeramente leptocúrticas (curtosis > 3), para las Var(IBDP) entre abuelo – nieto y abuela – nieto: 3.964 y 3.799, respectivamente.

Cuadro 9. Valores de las medias (\bar{X}), medianas (\tilde{X}), desvíos estándares (DE), coeficientes de variación (CV), asimetría, curtosis y valores máximos y mínimos esperadas en los valores medios de las IBDP abuelo(a) – nieto.

| Fase | \bar{X} | \tilde{X} | DE | CV | Asimetría | Curtosis | V. Max. | V. Min. |
|---------|-----------|-------------|-------|-------|-----------|----------|---------|---------|
| Paterna | 0.250 | 0.2487 | 0.032 | 12.85 | 0.0530 | 3.206 | 0.440 | 0.115 |
| Materna | 0.250 | 0.2489 | 0.034 | 13.47 | 0.1360 | 3.031 | 0.380 | 0.129 |

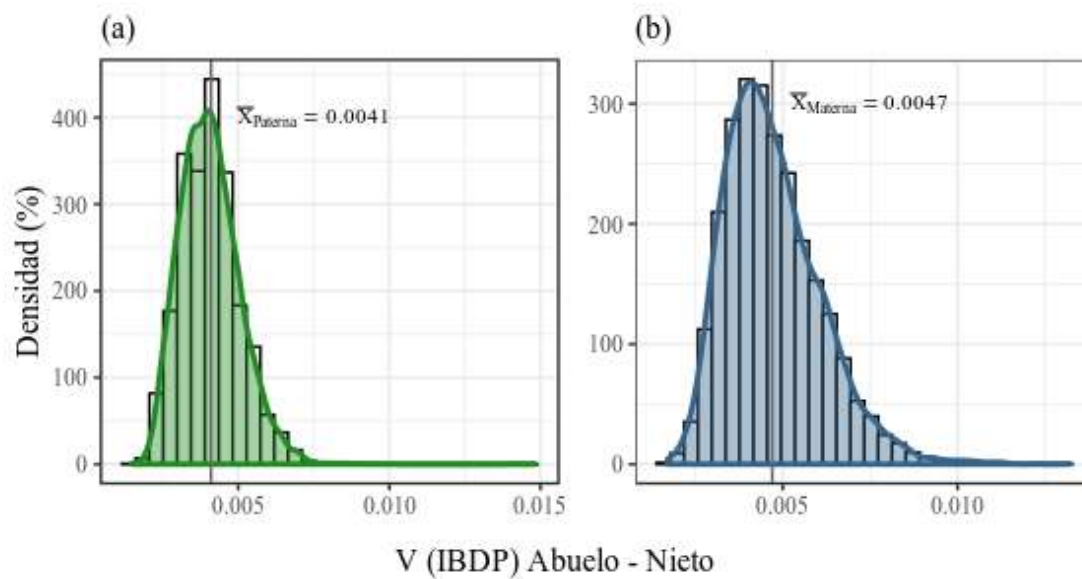
Figura 12. Distribución de las proporciones de genoma IBD compartido estimadas para fases paternas (a) y maternas (b).



Cuadro 10. Valores de las medias (\bar{X}), medianas (\tilde{X}), desvíos estándares (DE), coeficientes de variación (CV), asimetría, curtosis y valores máximos y mínimos esperadas en varianzas de las IBDP abuelo(a) – nieto(a).

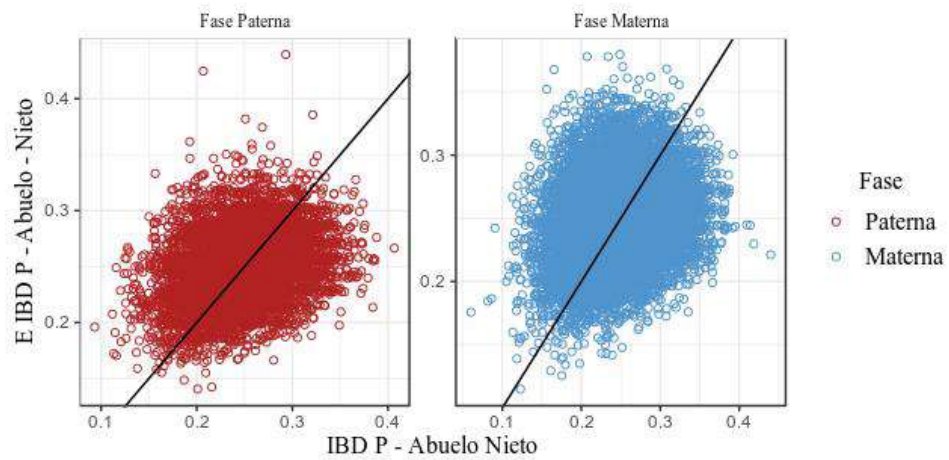
| Fase | \bar{X} | \tilde{X} | DE | CV | Asimetría | Curtosis | Máximo | Mínimo |
|---------|-----------|-------------|--------|-------|-----------|----------|--------|--------|
| Paterna | 0.0040 | 0.0040 | 0.0010 | 12.85 | 0.0530 | 5.3400 | 0.440 | 0.115 |
| Materna | 0.0047 | 0.0047 | 0.0013 | 13.47 | 0.1360 | 4.3490 | 0.380 | 0.129 |

Figura 13. Distribución de las varianzas de las proporciones de genoma IBD compartido estimadas para fases paternas (a) y maternas (b).



El IBDP para las relaciones observadas entre el abuelo o la abuela con el nieto calculadas con [3.2] y los valores esperados calculados con [3.8] se muestran en la Figura 14. Se observó una moderada correlación entre los valores esperados y los observados de IBDP abuelo – nieto ($r = 0.301$) y abuela – nieto ($r = 0.283$): a medida que aumentaron los valores *esperados* de IBDP, aumentaron también los valores *observados* de IBDP.

Figura 14. Comparación entre las proporciones observadas de IBDP, respecto de las IBDP esperadas abuelo – nieto (izquierda) y abuela – nieto (derecha).



La Figura 13 muestra que la varianza estimada de IBDP para los abuelos paternos (SS y DS) fue igual a 0.004. Las estimaciones de IBDP mostraron muy bajo desvío estándar ($DE < 0.0001$). Estos valores medios de las varianzas estimadas de IBDP son mucho menores comparados con los de la varianza de la relación de parentesco IBD abuelo(a) – nieto estimada bajo el modelo de un locus e igual a $Var(IBD) = 0.0625$ (Hill y Weir 2011).

Capítulo 5: Discusión

Capítulo 5

Discusión

En esta tesis se encontraron fórmulas para el primero y segundo momentos de la IBDP entre abuelo y nieto en función de un proceso de renovación Gamma que contempla presencia de interferencia. Asimismo, se analizó la presencia o no de interferencia en la recombinación presente en datos de una empresa internacional de mejora genética de cerdos. El algoritmo sugerido por Cheung (2007) permitió identificar el origen ancestral de sitios informativos dentro del genoma de identidad entre el genotipo del abuelo, o la abuela y el nieto, para luego detectar los sitios de recombinación y la longitud de los segmentos IBD. Una vez evaluados estos últimos, se ajustaron los datos a distintas distribuciones de probabilidad siendo la Gamma la que mostró el mejor ajuste.

5.1. Estimación de los parámetros de interferencia bajo una distribución Gamma.

Para cada cromosoma se ajustaron los datos de longitud de segmentos IBD con distribuciones Gamma y se estimaron los parámetros de interferencia con el algoritmo de Housworth y Stahl (2009) bajo las especificaciones distribucionales de Broman y Weber (2000). Los resultados obtenidos sugieren la presencia de una ligera interferencia positiva en todos los cromosomas de cerdos, salvo los dos últimos (17 y 18). Se ajustó por cromosoma los datos de longitud de segmentos IBD empleando distribuciones Gamma y se estimó el parámetro de interferencia con el algoritmo de Housworth y Stahl (2009) con las especificaciones distribucionales de Broman y Weber (2000). Los resultados obtenidos sugieren la presencia de una ligera interferencia positiva en los cromosomas 1, 3 y 6, interferencia negativa en los cromosomas 2, 4, 5, 8, 10, 11, 12, 13, 15, 16, 17 y 18, mientras que no hay evidencia para encontrar interferencia en los cromosomas 7, 9 y 14. Esto sugiere que en las fases paternas, los cromosomas largos según el Cuadro 2, hay una ligera interferencia positiva, mientras que en los cromosomas muy cortos (cromosoma 11, 12, 16, 17 y 18) hay una marcada interferencia negativa, indicando que a menor tamaño, las distancias entre recombinaciones son menores y hacen que los valores estimados de interferencia sean menores a 1. Por su parte, en las fases maternas, no existe una asociación entre los valores estimados de no recombinación con la longitud de cromosomas, resaltando

además que en la mayoría de los cromosomas maternos, había mayor número de recombinación promedio que en las fases paternas.

5.2. Probabilidad de ausencia de recombinación

Los valores estimados del Cuadro 8 y la tendencia de las probabilidades de ausencia recombinación (P_0) al emplear un proceso de renovación Gamma, para la variable longitud física de los segmentos IBD en la Figura 9, sugieren que P_0 es inversamente proporcional a la distancia física del cromosoma. Para las fases paternas, a mayor distancia física, menor P_0 debido a que a mayor longitud, mayor es la probabilidad de crossover entre las gametas durante la meiosis, mientras que no hubo una relación clara entre longitud de cromosomas y probabilidades de no recombinación estimadas para las fases maternas. Además, la Figura 10 muestra que las P_0 estimadas tendieron a subestimar las frecuencias de las gametas parentales en cada cromosoma, tanto en las fases paternas y maternas.

5.3. Proporción de Genoma IBD compartido (IBDP) para abuelo(a) – nieto

Las IBDP calculadas con el algoritmo de Cheung (2007) para las relaciones entre los abuelos y sus nietos mostraron valores medios muy cercanos a los estimados con el pedigrí, como así también escasa variabilidad (< 0.002). Este resultado es similar al obtenido por García-Baccino et al (2016) quienes encontraron que la varianza de estimación de las IBDP entre medio hermanos, calculada con el método de segmentos de Guo (1995), fue menor a la de los métodos puntuales de IBD. Además, se puede observar en el Cuadro 8 que hay asimetría negativa en las relaciones abuelo – nieto y positiva en las relaciones abuela – nieto.

Los valores medios de IBDP estimados por ML con la distribución Gamma tuvieron promedios cercanos a 0.25, que es igual al valor esperado de la IBDP, el cual se calcula exclusivamente con información de pedigrí. Asimismo, se observó escasa variabilidad: varianza < 0.001 . Al comparar IBDP observados con los valores medios estimados empleando la fórmula [3.7], se insinúa una leve correlación positiva. Los valores de las varianzas de IBDP mostraron promedios cercanos a 0.02, cifra inferior a la varianza de la relación de parentesco IBD en el modelo de un locus igual a 0.0625 (Hill y Weir, 2011), estimada con la expresión [2.8] del Capítulo 2. Resaltando también que estos valores estimados de la varianza tienen baja varianza, reflejando una alta concentración de las observaciones con respecto al valor medio (curtosis > 3) y una

asimetría positiva con cola derecha más pesada, situación esperable debido a que la varianza IBD es un valor enteramente positivo. Todos estos resultados son probablemente la contribución de mayor importancia de esta tesis, porque muestran que las estimaciones de las relaciones de parentesco con información genómica mediante métodos de segmentos son mucho más precisas que los métodos IBD de posiciones puntuales.

5.4. Implicancias de la varianza del IBDP

La motivación principal de la tesis fue el encontrar una expresión de la $\text{Var}(\text{IBDP})$ que nos permita captar información más precisa sobre los procesos genéticos que originan las relaciones de parentesco entre los animales que participan de las evaluaciones genéticas. Las relaciones realizadas con información genómica pueden variar de año a año dependiendo del método de estimación empleado en el cálculo. Los métodos puntuales de estimación de relaciones, y sobre todo aquellos que emplean solamente la información de los marcadores presentan mayor varianza de estimación.

Durante el desarrollo de la presente investigación para derivar la $\text{Var}(\text{IBDP})$ se utilizó un enfoque continuo y se encontró que la longitud de los segmentos vinculados al IBDP sigue una distribución mezcla de Gammas en presencia de interferencia y, en su ausencia, una exponencial. Alternativamente al enfoque continuo, la noción de IBD original (Malecot, 1948, 1969) está referida a un muestreo en un locus al azar en el genoma, de dos genes pertenecientes a individuos distintos, emparentados por uno o más ancestros en común. En ese locus se verifica si los genes son iguales porque son copias de un gen en un ancestro común. La variable aleatoria resultante es una Bernoulli con parámetro igual a la probabilidad de IBD. Generalizando la noción de IBD a dos loci ligados, e invocando un resultado previo de J.B.S. Haldane, Weir y Cockerham (1969) describieron una función de “desequilibrio de identidad” para la probabilidad de IBD conjunta entre dos loci (1 y 2) en dos individuos distintos (X e Y) respecto del producto de las probabilidades marginales de IBD en cada uno de ambos loci asumiendo que fuesen independientes. El desequilibrio de identidad es igual a $P(X_1 \equiv Y_1, X_2 \equiv Y_2) > P(X_1 \equiv Y_1) P(X_2 \equiv Y_2)$. Esta expresión indica que la probabilidad conjunta de IBD entre dos loci ligados es siempre mayor que la probabilidad IBD entre loci independientes. Siguiendo esta idea, Weir et al (1974, página 401) expresaron la varianza del IBD a dos-loci, cuando el número de variantes causales responsables de la variabilidad aditiva del carácter es grande, de la manera siguiente:

$$\text{Var}(\text{IBD}_{1,2}) = P(X_1 \equiv Y_1, X_2 \equiv Y_2) - P(X \equiv Y)^2$$

Nótese que la varianza del IBD a dos loci está relacionada con el “desequilibrio de identidad”, el cual genera además varianza de muestreo en los coeficientes de consanguinidad en poblaciones finitas bajo apareamiento aleatorio con loci no ligados (Weir y Cockerham, 1969; Cockerham y Weir, 1977). Es posible generalizar el “desequilibrio de identidad” a un proceso de herencia en segmentos e IBDP (Guo, 1995), diferenciando entre “ancestría genealógica” y “ancestría genética” (Matsen y Evans, 2008). La primera refleja el valor esperado de la transmisión del genoma de un individuo en las gametas de sus descendientes, esperanza que en especies biparentales es proporcional a 0.5^{-g} , siendo g el número de generaciones entre el ancestro y su descendiente. Para un pedigree en particular, la coancestría genealógica resulta en la matriz A . Alternativamente, la “ancestría genética” es la que produce variación en las relaciones observadas con información genómica y pedigrí. Su origen se encuentra en el “proceso conjunto de recombinación” (Veller et al, 2020) mencionado en capítulos anteriores.

Hemos observado que la $\text{Var}(\text{IBDP})$ obtenida a partir de la distribución de los segmentos IBDP mezcla de Gammas mostró menor varianza que la del clásico IBD discreto (Malecot, 1948, 1969), es decir: $\text{Var}(\text{IBDP}) < \text{Var}(\text{IBD})$. Este efecto puede explicarse por el efecto estocástico del “sesgo por tamaño de muestreo” (SBS, “size (or length) - biased sampling”, Terwilliger et al. (1997); Thompson, 2013). El efecto del sesgo en el “número de muestreos de sitios IBD” ocurre porque la variable Bernoulli toma dos valores $\text{IBD} = 1$, $\text{IBD} = 0$, no IBD), y la distribución de la variable IBDP se encuentra *ponderada* por la inversa del valor esperado del “tamaño de muestra” observado, siendo entonces igual a

$$G(\text{IBDP}) = \frac{1}{E_{\text{IBD}}(\text{IBDP})} \int_0^{\text{IBDP}} \text{IBD} dF(\text{IBD})$$

(Terwilliger et al, 2001). La esperanza en el denominador es calculada con respecto a la distribución $F(\text{IBD})$. La $\text{Var}(\text{IBDP})$ resultante es menor que la $\text{Var}(\text{IBD})$, del mismo modo que en selección por truncación la varianza del criterio de selección de los candidatos selectos es menor que la varianza del criterio en toda la población (Lynch y Walsh, 1998). Consecuentemente, Terwilliger et al. (1997) demostraron que al muestrear un cromosoma buscando detectar un sitio IBD, la distancia esperada entre dos recombinaciones que rodean dicho sitio desde ambos lados es el doble de la de un

segmento IBDP que ocurre entre dos recombinaciones cualquiera. Thompson (2013) explicó este punto en mayor detalle.

La estimación de las relaciones de parentesco con información genómica requiere emplear estadísticos cuyo valor esperado reflejen la ancestría genealógica y la genética y no solamente una de ellas. Csűrös (2013) observó que con marcadores bialélicos (por ejemplo, los SNP) solo es posible estimar funciones lineales del conjunto de las nueve $P(\text{IBD})$ discretas definidas por Jacquard (1972), tales como el coeficiente de coancestría de Malecot (1948) o la consanguinidad de uno o de ambos individuos, y no puede detectar la varianza del desequilibrio de identidad. García-Baccino et al (2016) encontraron que el método de estimación del IBD en segmentos de Guo (1995) tuvo menor $\text{Var}(\text{IBDP})$ para medio-hermanos paternos que los métodos basados en IBD discreto y mucho menor comparado con la varianza de un método de identidad en estado. Sobre la base de lo comentado más arriba, los estimadores de relaciones por métodos de IBD tienen valor esperado igual al proceso *genealógico*, es decir a las relaciones de parentesco en A , esperanza que no es función del *proceso conjunto de recombinación*. Consecuentemente, la predicción del valor de cría animal con métodos que combinen el pedigrí con información genómica para estimar las relaciones de parentesco IBDP, teniendo en cuenta la *herencia en segmentos* cromosómicos con longitud estimada por la detección de los eventos de recombinación, refleja información más precisa al tener menor varianza del estimador de las relaciones.

Capítulo 6: Conclusiones

Capítulo 6

Conclusiones

La tesis utiliza genotipos densos para estimar los sitios de recombinación e investiga la interferencia entre dichos sitios. Asimismo, se obtienen fórmulas para calcular la esperanza y la varianza del IBDP en función de los parámetros de la distribución Gamma. La parte más novedosa es la determinación de los puntos de recombinación con el modelo de Cheung et al. (2007) para considerar el proceso conjunto de recombinación de Veller et al. (2020) de modo de calcular el IBDP.

Las principales conclusiones de la tesis son las siguientes:

1. Con la información de los segmentos IBD, se calculó $E(\text{IBDP})$ y $\text{Var}(\text{IBDP})$ entre abuelo - nieto de todos los cerdos genotipificados en la base analizada, empleando los parámetros Gamma estimados por máxima verosimilitud. Se observó interferencia en todos los cromosomas, excepto en 7, 9 y 14 (fase paterna), y 2, 15 y 16 (fase materna).
2. La probabilidad estimada de ausencia de recombinación fue inversamente proporcional a la longitud del cromosoma en la fase paterna, mientras que para la fase materna no se observó relación alguna entre la interferencia y la longitud del cromosoma.
3. El uso de información de segmentos IBD redujo la $\text{Var}(\text{IBDP})$ con respecto a la estimada utilizando sólo pedigree. Los resultados sugieren no ignorar la interferencia en el cálculo de la proporción de genoma compartido IBD y en su varianza en la especie porcina.

BIBLIOGRAFÍA

- Bickeboller, H., Thompson, E. A. 1996. Distribution of Genome Shared IBD by Half – Sibs: Approximation by the Poisson Clumping Heuristic. *TPB*. 50:66-90.
- Brent, R. P. 1973. Algorithms for minimization without derivatives. Prentice – Hall, Englewood Cliffs, NJ.
- Broman, K. W., Weber, J. L. 2000. Characterization of Human Crossover Interference. *Am. J. Hum. Genet.* 66:1911-1926.
- Caballero, M., Seidman, D.N., Qiao, Y., Sannerud, J., Dyer, T. D., M. Lehman, J. E. Curran, R. Duggirala, J. Blangero, S. Carmi, A. L. Williams. 2019. Crossover interference and sex-specific genetic maps shape identical by descent sharing in close relatives. *PLOS Genetics* 15(12): e1007979.
- Cantet, R.J.C., García-Baccino, C.A., Rogberg-Muñoz, A., Forneris, N.S., Munilla, S. 2017. Beyond genomic selection: The animal model strikes back (one generation)!. *J. Anim. Breed. Genet.* 134:224-231.
- Cheung, V.G., Burdick, J. T., Hirschmann, D., Morley, M. 2007. Polymorphic Variation in Human Meiotic Recombination. *Amer. J. Huma. Genet.* 80:526-530.
- Choi, S. C., Wette, R. 1969. Maximum Likelihood Estimation of the Gamma Distribution and Their Bias. *Technometrics* 11:683-690.
- Cobbs, G. 1978. Renewal Process Approach to the Theory of Genetic Linkage: Case of No Chromatid Interference. *Genetics* 89:563-581.
- Cockerham, C. C. 1954. An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariance among Relatives When Epistasis Is Present. *Genetics.* 39:861-882.
- Cockerham, C. C. 1956. Effects of Linkage on the Covariances Between Relatives. *Genetics.* 41:138-141.
- Cockerham, C. C. 1971. Higher Order Probability Functions of Identity of Alleles by Descent. *Genetics* 69:235-246.
- Cockerham, C. C. 1984. Additive by Additive Variance with Inbreeding and Linkage. *Genetics.* 108:487-500.
- Cockerham, C. C., Weir, B. S. 1977. Digenic descent measures for finite populations. *Genet. Res.* 30: 121–147.

- Cotterman, C. W. 1941. Relatives and Human Genetic Analysis. *The Scientific Monthly*. 53:227-234.
- Csürös, M. 2014. Non-identifiability of identity coefficients at biallelic loci. *Theor. Popul. Biol.* 92:22–29.
- Delignette – Muller, M. L., Dutang, C. *fitdistrplus: An R Package for Fitting Distributions*. *Journal of Statistical Software*. 64:1-34. URL: <http://www.jstatsoft.org/v64/i04/>.
- Denniston, C. 1975. Probability and Genetic Relationship: Two Loci. *AHG* 39:89-104.
- Donnelly, K. P. 1983. The Probability that Related Individuals Share Some Section of Genome Identical by Descent. *Theor. Popul. Biol.* 23:34-63.
- Felsenstein, J. 1979. A Mathematically Tractable Family of Genetic Mapping Functions with Different Amounts of Interference. *Genetics* 91:769-775.
- Fisher, R. A. 1918. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburgh*. 52: 399-433.
- Foss, E., Lande, R., Stahl, F. W., Steinberg, C. M., 1993. Chiasma Interference as a Function of Genetic Distance. *Genetics*. 133:681-691.
- Garcia – Baccino, C. A., Munilla, S., Legarra, A., Vitezica, Z. G., Forneris, N. S., Bates, R. O., Ernst, C. W., Raney, N. E., Steibel, J. P., Cantet, R. J. C. 2016. Estimates of the actual relationship between half – sibs in a pig population. *J. Anim. Breed. Genet.* 134: 109-118.
- Guo, S.W. 1995. Proportion of Genome Shared Identical by Descent by Relatives: Concept, Computation and Applications. *Am. J. Hum. Genet.* 56:1468-1476.
- Haldane, J. B. S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. *J. Genet.* 8:299-309.
- Henderson, C.R. 1984. *Applications of linear models in animal breeding*. University of Guelph, Guelph, Ontario, Canada.
- Hill, W. G., Weir, B. S. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet. Res. Camb.* 93:47-64.
- Housworth, E. A., Stahl, F. W. 2009. Is There Variation in Crossover Interference Levels Among Chromosomes from Human Males? *Genetics* 183:403-405.
- Jacquard, A. 1972. Genetic Information Given by a Relative. *Biometrics*. 28:1101-1114.
- Karigl, G. 1981. A recursive method for the calculation of identity coefficients. *Ann. Hum. Genet.* 45:299-305.
- Kempthorne, O. 1954. The Correlation between Relatives in a Random Mating Population. *Proc. R. Soc. Lond. B.* 143:103-113.

- Kempthorne, O. 1957. *An Introduction to Genetics Statistics*. Wiley.
- Lander, E. S., Green, P. 1987. Construction of Multilocus Genetic Linkage Maps in Humans. *Proc. Natl. Acad. Sci. USA*. 84:2363-2367.
- Legarra, A., Christensen, O. F., Aguilar, I., Misztal, I. 2014. Single Step, a general approach for genomic selection. *Livestock Science*. 166:54-65.
- Lian, J., Yin, Y., Oliver – Bonet, M., Liehr, T., Ko, E., Turek, P., Sun, F., Martin, R. H. 2008. Variation in crossover interference levels on individual chromosomes from human males. *Human Molecular Genetics*. 17:2583-2594.
- Liu, W., Weir, B. S. 2005. Genotypic probabilities for pairs of inbred relatives. *Phil. Trans. R. Soc. B*. 360:1379-1385.
- Lynch, M., y Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA, USA.
- Malecot. G. 1948. *Les Mathématiques de l'hérédité*. Masson et Cie. Paris, France.
- Malecot. G. 1969. *The Mathematics of Heredity*. W. H. Freeman and Company. San Francisco, USA.
- Mather, K. 1938. Crossing Over. *Biol. Rev.* 252-292.
- McPeck, M. S., Speed, T. P. 1995. Modelling Interference in Genetic Recombination. *Genetics* 139:1031-1044.
- Muller, H. J. 1916. The Mechanism of Crossing – Over. *The American Naturalist*. 50:193-221.
- Nolan, J. P. 2017. Erlang renewal models for genetic recombination. *Journal of Statistical Distributions and Applications*. 4:1-17.
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ross, S. M. 2007. *Introduction to Probability Models*. Ninth Edition. Academic Press. Elsevier.
- Speed, T. P. 2005. *Genetic Map Functions*. Encyclopedia of Biostatistics, John Wiley & Sons.
- Sturtevant, A. H. 1915. Experiments on sex recognition and the problem of sexual selection in *Drosophila*. *J. Anim. Behavior*. 5:351-366.
- Terwilliger, J. D., W. D. Shannon, G. M. Lathrop, J. P. Nolan, L. R. Goldin, G. A. Chase, D. E. Weeks. 1997. True and False Positive Peaks in Genome-wide Scans: Applications of Length-Biased Sampling to Linkage Mapping. *Amer. J. Hum. Genet.* 61:430-438.

- Thompson, E. A. 2013. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. *Genetics*. 194:301-326.
- Tortereau, F., Servin, B., Frantz, L., Megens, H. J., Milan, D., Rohrer G., Wiedmann, R., Beever, J., Archibald, A. L., Schook, L. B., Groenen M. A. 2012. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics*. 13:586.
- Veller, C., Edelman, N. B., Murallidhar, P., Nowak, M. A. 2020. Variation in genetic relatedness is relatedness is determined by the aggregate recombination process. *Genetics* 216:985-994.
- Weir, B. S., Cockerham, C. C. 1969. Pedigree Mating with Two Linked Loci. *Genetics*. 61:923-940.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* 56: 330 - 338.
- Zhao, H., McPeck, M. S., Speed, T. P. 1995. Statistical Analysis of Chromatid Interference. *Genetics* 139: 1057-1065.
- Zhao, H., Speed, T. P. 1996. On Genetic Map Functions. *Genetics*. 142:1369-1377.

Apéndice A

Cuadro A.1. Casos detectables de origen ancestral de marcadores IBD en una gameta.

| Abuelo | Abuela | Padre referente | Otro padre | Cría | Origen Ancestral |
|--------|--------|--------------------|------------|------|---------------------|
| AA | TT | AT | TT | AT | 1 |
| AT | TT | AT | TT | AT | 1 |
| AA | AT | AT | AA | AT | 2 |
| TT | AT | AT | TT | AT | 2 |
| TT | AA | AT | TT | TT | 1 |
| AA | TT | AT | AA | AA | 1 |
| TT | AA | AT | AA | AA | 2 |
| AA | TT | AT | TT | TT | 2 |
| AT | TT | AT | AA | AA | 1 |
| AT | TT | AT | TT | TT | 2 |
| AT | AA | AT | TT | TT | 1 |
| AT | AA | AT | AA | AA | 2 |
| AT | AA | AT | AA | AT | 1 |
| AA | TT | AT | AA | AT | 2 |
| TT | AA | AT | TT | AT | 2 |
| TT | AA | AT | AA | AT | 1 |
| AA | AT | AT | AA | AA | 1 |
| AT | AA | AT | AT | TT | 1 |
| AA | AT | AT | AT | TT | 2 |
| TT | AA | AT | AT | AA | 2 |
| AA | TT | AT | AT | AA | 1 |
| AA | TT | AT | AT | TT | 2 |
| TT | AA | AT | AT | TT | 1 |
| AA | AT | AT | AT | AA | 1 |
| AA | AT | AT | TT | TT | 2 |
| TT | AT | AT | AA | AA | 2 |
| TT | AT | AT | AA | AT | 1 |
| TT | AT | AT | AT | AA | 2 |

| | | | | | |
|----|----|----|----|----|---|
| TT | AT | AT | AT | TT | 1 |
| TT | AT | AT | TT | TT | 1 |
| AT | AA | AT | AT | AA | 2 |
| AT | AA | AT | TT | AT | 2 |
| AT | TT | AT | AA | AT | 2 |
| AT | TT | AT | AT | AA | 1 |
| AT | TT | AT | AT | TT | 2 |
| AA | AT | AT | TT | AT | 1 |