






Algoritmo Híbrido Mendeliano–Bayesiano para la Inferencia Probabilística de Genotipos en Pedigríes

Hybrid Mendelian–Bayesian Algorithm for Probabilistic Inference of Genotypes in Pedigrees

Autores

Alexander Stefano Mora Matamoros¹ 
✉ alexander.mora.0121@espam.edu.ec

Javier Hernán López-Zambrano^{*1} 
✉ jlopez@espam.edu.ec

Carlos Octavio Larrea-Izurieta¹ 
✉ clarrea@espam.edu.ec

Diego Alexander Toala Palma² 
✉ diego.toala@technology-park.com

¹Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López, Calceta, Manabí, Ecuador.

²Technopark S.A, Manta, Manabí, Ecuador

*Autor para correspondencia

Comó citar el artículo:

Mora Matamoros, A., López-Zambrano, J.H., Larrea-Izurieta, C.O. & Toala Palma, D.A. (2026). Algoritmo Híbrido Mendeliano–Bayesiano para la Inferencia Probabilística de Genotipos en Pedigríes. *Informática y Sistemas*, 10(1), pp. 69–78. <https://doi.org/10.33936/isrtic.v10i1.8393>

Enviado: 23/04/2026
Aceptado: 08/06/2026
Publicado: 11/06/2026

Resumen

Este estudio presenta un algoritmo híbrido, basado en reglas mendelianas y en actualización bayesiana, para inferir distribuciones de probabilidad sobre genotipos a partir de fenotipos observables (capa) y pedigríes, cuando no hay datos genéticos directos. El método representa a cada individuo mediante una distribución genotípica (los casos determinados son distribuciones degeneradas) y define un operador de “cruce” que combina distribuciones parentales para producir la distribución esperada en la descendencia. La inferencia incorpora filtros fenotípicos que eliminan genotipos incompatibles y realiza actualizaciones ascendentes y descendentes: la evidencia de hijos ajusta las creencias sobre progenitores y viceversa. Validado sobre poblaciones simuladas (núcleos familiares y pedigríes más profundos) para un locus bialélico, el algoritmo reproduce las proporciones mendelianas en casos de certeza, propaga coherentemente la incertidumbre cuando los progenitores son parciales y converge rápidamente con observaciones repetidas. Las limitaciones actuales (un locus, sin mutación/recombinación ni error de observación) acotan su aplicabilidad inmediata, pero su claridad, interpretabilidad y facilidad de implementación lo hacen útil como herramienta explicable para programas de cría, conservación y enseñanza, y como base para futuras extensiones multilocus y modelos de observación ruidosos.

Palabras clave: Inferencia genotípica; Inferencia bayesiana; Genética mendeliana; Color de capa equina

Abstract

This study presents a hybrid algorithm—grounded in Mendelian inheritance and Bayesian updating—to infer genotype probability distributions from observable phenotypes (coat color) and pedigree structures when direct genetic data are unavailable. Each individual is represented by a genotype distribution (with fully known genotypes modeled as degenerate distributions), and a distribution-based “cross” operator combines parental distributions to obtain the expected genotype distribution in the offspring. Phenotypic constraints are incorporated by filtering out incompatible genotypes and renormalizing, while both downward and upward inference are supported: offspring evidence updates parental beliefs and vice versa. Validated on simulated populations (small nuclear families and deeper pedigrees) for a single biallelic locus, the algorithm reproduces classical Mendelian proportions under parental certainty, coherently propagates uncertainty when parents are partially specified, and converges steadily under repeated observations. Although current assumptions (single locus, no mutation or recombination, perfect phenotype observation) limit immediate applicability to complex real data, the method’s clarity, interpretability, and ease of implementation make it a practical and explainable tool for breeding programs, conservation planning, and education. It also provides a transparent foundation for future multilocus extensions and models incorporating observation noise.

Keywords: Genotype inference; Bayesian inference; Mendelian genetics; Equine coat color



1. Introducción

La inferencia del genotipo de individuos a partir de observaciones fenotípicas y estructuras genealógicas es un problema central en genética de poblaciones cuando los datos genéticos directos no están disponibles. En muchas poblaciones animales de interés, como los caballos peruanos de paso, solo se registran fenotipos observables (por ejemplo, el color de la capa) y las relaciones genealógicas entre individuos, mientras que la información sobre los genotipos permanece ausente. Esta situación dificulta la comprensión de la herencia genética y limita la capacidad para realizar análisis posteriores de selección, cría o conservación.

El color de la capa en equinos se encuentra bajo control genético claro y ha sido utilizado como modelo de fenotipo mendeliano en múltiples estudios. Diversos loci, especialmente los genes MC1R (factor de extensión) y ASIP (agouti), interactúan para producir las capas básicas de los caballos (alazán, negro y castaño) y han sido descritos exhaustivamente en estudios de genética equina (Sponenberg & Bellone, 2017; Liu et al., 2024). Estos rasgos fenotípicos, aunque fácilmente observables, son el resultado de múltiples posibles combinaciones genotípicas, lo que genera ambigüedad en la inferencia directa del genotipo.

La mecánica de la herencia mendeliana establece que los alelos se transmiten de padres a descendientes de acuerdo con probabilidades bien definidas. No obstante, fenómenos como heterocigosidad oculta, epistasis y efectos de dilución pueden aumentar la complejidad de la interpretación (Zhang et al., 2011). Simultáneamente, la inferencia bayesiana ha demostrado ser una herramienta robusta para estimar parámetros genéticos y genotipos en contextos con datos incompletos o inciertos, ya que permite integrar de forma principada información previa con evidencia observada (Almudevar & LaCombe, 2012; Ormond et al., 2024).

La hipótesis central es que, aun con datos fenotípicos incompletos, es posible estimar distribuciones de probabilidad de genotipos utilizando un algoritmo que combine las leyes de herencia descritas por Mendel con principios de inferencia bayesiana. El objetivo específico es presentar un algoritmo de inferencia híbrido, Mendeliano y bayesiano, y demostrar su desempeño mediante simulaciones en poblaciones virtuales, lo que permite establecer su viabilidad metodológica sin requerir datos genéticos directos (Almudevar & LaCombe, 2012; Ormond et al., 2024).

Avances metodológicos recientes en imputación y en técnicas que combinan información de parentesco (IBD) con información de paneles poblacionales muestran que el uso conjunto de señales de pedigrí y de correlaciones de ligamiento (LD) mejora

sustancialmente la precisión de inferencia para variantes raras y, en general, la recuperación de genotipos cuando solo hay datos parciales (estudios de comparación y guías prácticas sobre imputación en pedigrí). Estas evaluaciones proporcionan criterios de selección de métodos y estrategias de muestreo útiles para diseñar un esquema de referencia/secuenciación en poblaciones con pedigrí extenso (comparativas de métodos de imputación para pedigrí y revisiones generales sobre imputación genómica). Además, trabajos recientes en equinos demuestran que la imputación estadística aplicada con paneles modernos puede reconstruir genotipos históricos y aportar información útil sobre rasgos de apariencia y comportamiento en caballos, lo que valida la factibilidad de enfoques que parten de fenotipos observables y estructuras genealógicas en especies equinas (Todd et al., 2023; Treccani et al., 2023; Ullah et al., 2019).

Desde la perspectiva bayesiana, estudios contemporáneos abordan cuestiones prácticas críticas para la inferencia con datos sesgados por procesos de selección (por ejemplo, programas de cría): elección de priors, modelado de la selección y el manejo de datos faltantes/observaciones censuradas. Estas contribuciones son importantes para asegurar que los posteriors de genotipo y parámetros derivados no estén indebidamente sesgados por esquemas de selección no aleatoria en los registros fenotípicos (Gianola et al., 2022). Asimismo, enfoques integradores de sistema (systems genetics / veterinary systems biology) subrayan la utilidad de combinar información fenómica, pedigrí y múltiples capas ómicas o computacionales cuando estén disponibles, permitiendo que modelos híbridos (bayesianos + reglas mendelianas + señales de pedigrí) mejoren la resolución de la inferencia fenotipo→genotipo y su aplicabilidad en toma de decisiones veterinarias y de cría (Pathak & Kim, 2024).

En conjunto, estos hallazgos recientes apoyan la propuesta de este trabajo: un algoritmo híbrido que explote reglas mendelianas locales, la estructura del pedigrí y una formulación bayesiana flexiblemente jerárquica puede producir distribuciones de probabilidad de genotipos informativas aun cuando la información molecular directa sea inexistente. El presente artículo adopta estos desarrollos prácticos y teóricos para diseñar y evaluar un flujo de inferencia aplicable a caballos peruanos de paso, con especial atención a la precisión en variantes de efecto visible (p. ej. loci de color de capa) y a la robustez frente a sesgos por selección y pedigrí incompleto (Gianola et al., 2022; Pathak & Kim, 2024; Todd et al., 2023; Treccani et al., 2023; Ullah et al., 2019).

Este enfoque es relevante para estudios de genética equina, particularmente en la coloración de las capas, donde las estructuras familiares largas y los efectos genéticos complejos

requieren modelos capaces de integrar múltiples fuentes de incertidumbre. El desarrollo de métodos confiables de inferencia de genotipos tiene aplicaciones potenciales en programas de cría, conservación de razas y estudios epidemiológicos, donde el conocimiento del genotipo es crucial para decisiones fundadas en evidencia genética (Sponenberg & Bellone, 2017).

2. Materiales y Métodos

2.1 Contexto y diseño del estudio

Este trabajo corresponde a un estudio metodológico y computacional orientado al diseño y validación de un algoritmo para la inferencia probabilística de genotipos en pedigrís equinos. El desarrollo del algoritmo se realizó de forma independiente mediante implementación computacional propia durante el período 2025–2026, utilizando un equipo personal.

La validación presentada en este artículo se basa exclusivamente en poblaciones simuladas, diseñadas para reproducir estructuras familiares simples y pedigrís de mayor profundidad. Aunque el algoritmo fue concebido a partir del análisis de un pedigrí real.

2.2 Población y datos utilizados

Las poblaciones simuladas representan pedigrís en los que únicamente se dispone de dos tipos de información:

- (i) el fenotipo observable (capa del caballo) y
- (ii) las relaciones genealógicas entre individuos.

Los genotipos no se consideran observables y constituyen la variable objetivo del proceso de inferencia. En las simulaciones, los genotipos verdaderos se generan de manera controlada con el fin de contar con una referencia interna para evaluar el comportamiento del algoritmo.

Para facilitar la comprensión del método, se incluyen ejemplos con núcleos familiares pequeños (entre 3 y 10 individuos), así como estructuras genealógicas más extensas que permiten ilustrar la propagación de información a lo largo del pedigrí. En este artículo se trabaja únicamente con un locus bialélico, lo que permite exponer el funcionamiento del algoritmo sin pérdida de generalidad conceptual.

Un ejemplo representativo de la estructura utilizada se muestra en la Tabla 1, donde se presenta un núcleo familiar simple compuesto por dos progenitores con fenotipos distintos y tres descendientes. Esta configuración ilustra el tipo de información disponible para el algoritmo: únicamente fenotipos observables y relaciones de parentesco, a partir de los cuales se debe inferir la distribución genotípica de cada individuo.

2.3 Supuestos genéticos del modelo

El algoritmo se construye bajo un conjunto de supuestos explícitos:

- herencia mendeliana clásica,

Tabla 1. Estructura básica del núcleo familiar simulado.
Fuente: Los autores

Padre (Fenotipo A)	Madre (Fenotipo B)
Hijo 1	(Fenotipo A)
Hijo 2	(Fenotipo A)
Hijo 3	(Fenotipo B)

- ausencia de mutación y recombinación,
- observación perfecta del fenotipo,
- relación determinística entre fenotipo y conjunto de genotipos compatibles.

Estos supuestos no buscan reflejar toda la complejidad genética real, sino aislar el problema de inferencia y permitir una evaluación clara del comportamiento del algoritmo. La relajación de estas hipótesis se considera como trabajo futuro.

2.4 Representación probabilística de los individuos

Cada individuo del pedigrí se representa mediante una distribución de probabilidad sobre los genotipos posibles en el locus considerado. Esta distribución expresa el grado de creencia del algoritmo en cada genotipo compatible con la información disponible.

Los genotipos conocidos o fijados se modelan como distribuciones degeneradas, es decir, distribuciones con probabilidad uno en un único genotipo. De este modo, los cruces “simples” no constituyen un caso especial del algoritmo, sino una instancia particular del mismo marco probabilístico.

Cuando no existe información genealógica adicional, las distribuciones iniciales se asignan a partir del fenotipo observado, utilizando priors equilibrados consistentes con las restricciones fenotípicas. Un ejemplo de estas asignaciones se presenta en la Tabla 2, donde se observa cómo cada fenotipo restringe el conjunto de genotipos posibles y determina una distribución inicial sobre ellos. En particular, los genotipos incompatibles reciben probabilidad cero.

Tabla 2. Distribuciones genotípicas iniciales asignadas según restricción fenotípica.
Fuente: Los autores

Fenotipo A		Fenotipo B	
Genotipo	Probabilidad inicial	Genotipo	Probabilidad inicial
AA	0.5	AA	0
Aa	0.5	Aa	0
aa	0	aa	1

2.5 Cruce entre distribuciones genotípicas

El núcleo del algoritmo consiste en un operador de cruce que



combina las distribuciones genotípicas de dos progenitores para producir una distribución genotípica esperada en la descendencia.

Este operador no cruza genotipos individuales, sino distribuciones completas, ponderando todos los cruces mendelianos posibles según la probabilidad asignada a cada genotipo parental. En consecuencia, los cruces tradicionales entre genotipos conocidos se interpretan como casos límite de cruces entre distribuciones con certeza total.

La base de este operador se muestra en la Tabla 3, que presenta la matriz mendeliana de combinación alélica para un locus bialélico. Esta matriz define las posibles combinaciones de alelos parentales y los genotipos resultantes en la descendencia, constituyendo el componente fundamental sobre el cual se construye el cruce entre distribuciones probabilísticas.

Tabla 3. Matriz mendeliana de combinación alélica.

Fuente: Los autores

	A	a
A	AA	Aa
a	Aa	aa

2.6 Incorporación de evidencia fenotípica

Cuando el fenotipo del descendiente es observable, la distribución genotípica inferida se restringe eliminando los genotipos incompatibles con dicho fenotipo. Este proceso equivale a condicionar la distribución por evidencia observada y normalizarla posteriormente.

Este mecanismo permite integrar información fenotípica de forma coherente dentro de un esquema de inferencia bayesiana, manteniendo la consistencia probabilística del modelo.

2.7 Inferencia ascendente y actualización parental

Además de inferir descendientes a partir de progenitores, el algoritmo permite realizar inferencia ascendente, ajustando la distribución genotípica de un progenitor incierto a partir de la información proporcionada por uno o más descendientes y del otro progenitor.

Este ajuste se realiza evaluando qué genotipos parentales explican mejor la evidencia observada en la descendencia, y actualizando la distribución original en consecuencia. Conceptualmente, este procedimiento corresponde a una aplicación directa de la regla de Bayes.

2.8 Actualización secuencial y propagación de información

A medida que se incorpora nueva información genealógica

o fenotípica, las distribuciones genotípicas se actualizan de manera secuencial. El resultado de cada actualización se utiliza como punto de partida para las siguientes, lo que permite que la información se propague progresivamente a lo largo del pedigrí.

En las simulaciones realizadas, este proceso conduce a una concentración gradual de probabilidad en los genotipos compatibles con toda la evidencia disponible.

2.9 Implementación computacional

El algoritmo fue implementado mediante una solución computacional propia. Para garantizar la reproducibilidad, se documentaron los procedimientos de generación de pedigríes simulados, los criterios de asignación de priors y las reglas de actualización empleadas.

2.10 Consideraciones éticas

El presente estudio se basa exclusivamente en simulaciones computacionales. No se utilizaron datos sensibles ni se realizaron intervenciones sobre animales, por lo que no fue necesario solicitar consentimiento ético.

3. Resultados y Discusión

La aplicación práctica del algoritmo sobre las poblaciones simuladas mostró que el método (1) respeta las expectativas mendelianas en los casos de certeza parental, (2) propaga de forma coherente la incertidumbre cuando los progenitores son parciales, y (3) ajusta las creencias parentales ante evidencia descendente, con comportamiento estable ante observaciones repetidas. A continuación, se presentan resultados representativos (ejemplos numéricos) y su interpretación.

Con el fin de realizar experimentaciones sobre el algoritmo propuesto se realizaron simulaciones computacionales sobre poblaciones virtuales generadas bajo un modelo mendeliano bialélico (A/a). El diseño experimental incluyó 5.000 repeticiones independientes, en las que se generaron núcleos familiares y pedigríes de distinta complejidad, con tamaños comprendidos entre 3 y 10 individuos por familia. En cada simulación se asignaron genotipos verdaderos a los progenitores, se generó la descendencia mediante herencia mendeliana y posteriormente se ocultó la información genética, conservando únicamente los fenotipos observables y las relaciones de parentesco. Este procedimiento permitió evaluar la capacidad del algoritmo para reconstruir distribuciones genotípicas utilizando información parcial. Los ejemplos a continuación tienen un carácter ilustrativo y fueron seleccionados por su valor didáctico para mostrar el funcionamiento interno del método; sin embargo,

las conclusiones numéricas se fundamentan en la sección 3.7 Evaluación Experimental.

En todo el procedimiento un “cruce” no es la combinación de dos genotipos puntuales, sino la operación entre distribuciones genotípicas. Los cruces clásicos (por ejemplo, AA × Aa) son simplemente casos degenerados de esas distribuciones (por ejemplo, AA = {AA:1, Aa:0, aa:0}). Esto unifica el tratamiento y evita reglas ad-hoc.

3.1 Validación en cruces mendelianos clásicos

Resultado. En todos los casos en que ambos progenitores tenían genotipo resuelto, el algoritmo reprodujo exactamente las proporciones mendelianas esperadas para un locus bialélico (A/a). Esto valida que el núcleo de herencia del método está correctamente implementado.

Los resultados correspondientes se resumen en la Tabla 4, donde se presentan distintos cruces mendelianos clásicos junto con sus distribuciones genotípicas esperadas en la descendencia. Se observa que el algoritmo reproduce correctamente las proporciones teóricas en cada caso, confirmando la consistencia del operador de cruce en escenarios de certeza total.

Tabla 4. Cruces mendelianos clásicos.
Fuente: Los autores

Cruce	Distribución esperada		
	AA	Aa	aa
AA × AA	1	0	0
AA × Aa	0,5	0,5	0
Aa × Aa	0,25	0,5	0,25
AA × aa	0	1	0
Aa × aa	0	0,5	0,5
aa × aa	0	0	1

3.2 Cruces con progenitores genotípicamente inciertos

Resultado. Cuando al menos un progenitor se modela como una distribución (por ejemplo {AA:0.5, Aa:0.5, aa:0}), la distribución esperada del hijo se obtiene como una mezcla ponderada de todos los cruces mendelianos posibles, ponderada por la probabilidad de cada genotipo parental.

Ejemplo (demostrativo):

- Entrada: Padre = {AA:0.5; Aa:0.5; aa:0}; Madre = Aa (completa).
- Proceso operativo: descomposición del padre en casos, cálculo de la descendencia para cada caso, ponderación por la probabilidad de cada caso y suma de las contribuciones.
- Salida (inferida): {AA:0.375; Aa:0.50; aa:0.125}.

El detalle del cálculo se presenta en la Tabla 5, donde se descompone la distribución del progenitor incierto en casos

posibles y se muestran las contribuciones ponderadas de cada uno. Se observa que la distribución final del descendiente resulta de la suma de estas contribuciones, reflejando de manera explícita la incertidumbre en el genotipo parental.

Tabla 5. Cruce con progenitor incierto.
Fuente: Los autores

Caso	Sub-distribución	Ponderación	Contribución
Padre = AA (p=0.5)	AA × Aa → {AA:0.5; Aa:0.5}	0.5	{AA:0.25; Aa:0.25}
Padre = Aa (p=0.5)	Aa × Aa → {AA:0.25; Aa:0.5; aa:0.25}	0.5	{AA:0.125; Aa:0.25; aa:0.125}
Suma final	—	—	{AA:0.375; Aa:0.50; aa:0.125}

Interpretación. La mezcla ponderada preserva la varianza parental y permite expresar explícitamente la incertidumbre en la descendencia sin fijar genotipos arbitrariamente.

3.3 Incorporación de restricciones fenotípicas

Resultado. Al observar el fenotipo del hijo (que puede descartar algunos genotipos), el algoritmo aplica una máscara sobre la distribución resultante del cruce y la normaliza; el efecto es desplazar la probabilidad hacia genotipos compatibles con la evidencia observada.

Ejemplo: Aa × Aa → (0.25,0.5,0.25). Si el fenotipo observado excluye aa, la máscara aplicada (por ejemplo {0.5,0.5,0}) produce, tras normalizar, la distribución {0.333, 0.667, 0.0}.

El procedimiento se detalla en la Tabla 6, donde se muestra cómo la distribución mendeliana original se combina con la máscara fenotípica mediante un producto elemento a elemento y posterior normalización. Este proceso permite eliminar genotipos incompatibles y redistribuir la probabilidad entre las opciones restantes de manera coherente.

Tabla 6. Filtrado fenotípico aplicado a Aa × Aa.
Fuente: Los autores

Genotipo hijo	Prob. mendeliana	Máscara fenotípica	Producto	Normalizado
AA	0.25	0.5	0.125	0.333
Aa	0.50	0.5	0.25	0.667
aa	0.25	0.0	0.0	0.0

Interpretación. Este paso es la implementación práctica del condicionamiento por evidencia: elimina opciones imposibles y reequilibra la incertidumbre restante.



3.4 Inferencia ascendente: ajuste del genotipo parental a partir del descendiente

Resultado. El algoritmo ajusta la distribución de un progenitor incierto usando la evidencia aportada por un descendiente observado.

Procedimiento:

1. Para cada hipótesis genotípica posible del progenitor objetivo, se evalúa qué tan compatible es esa hipótesis con el genotipo observado del hijo.

o Si el otro progenitor está resuelto, la compatibilidad se obtiene directamente del cruce mendeliano.

o Si el otro progenitor también es incierto, la compatibilidad se obtiene promediando (ponderando) los resultados de ese cruce según la distribución del otro progenitor.

2. Cada hipótesis recibe un peso no normalizado igual a: (prior de la hipótesis) × (grado de compatibilidad con el hijo).

3. Se normalizan los pesos para obtener la distribución posterior del progenitor objetivo.

Las hipótesis que no pueden producir el genotipo observado quedan con peso cero; las que lo pueden producen pesos proporcionales a su capacidad explicativa.

Ejemplo (ilustrativo):

- Progenitor 1: aa (resuelto)
- Progenitor 2 (prior): {AA: 0.25; Aa: 0.50; aa: 0.25}
- Descendiente observado: aa

El cálculo detallado se presenta en la Tabla 7, donde se evalúa la compatibilidad de cada hipótesis parental con la evidencia observada y se muestran los pesos resultantes antes y después de la normalización. Se observa que las hipótesis incompatibles son eliminadas, mientras que las restantes se ajustan proporcionalmente a su capacidad para explicar el genotipo del descendiente.

Interpretación. Tras incorporar la evidencia del hijo, la hipótesis AA se descarta; Aa y aa quedan con igual probabilidad posterior (0.5 cada una). Si el otro progenitor fuese incierto, la “compatibilidad” de cada hipótesis se calcularía como la mezcla ponderada sobre las posibilidades del otro progenitor y el proceso de normalización sería idéntico.

3.5 Ajustes repetidos: cómo varias observaciones fortalecen la inferencia

Resultado. El algoritmo refina la distribución del progenitor incierto cada vez que se incorpora la observación de un nuevo hijo. Dependiendo a las observaciones del genotipo de la descendencia, la probabilidad asociada a ese genotipo crece de forma acumulativa y estable, mientras que las hipótesis incompatibles se eliminan rápidamente.

Procedimiento operativo. Para cada hijo observado:

1. Calcular, para cada hipótesis genotípica del progenitor incierto, su grado de compatibilidad con el hijo dado el otro progenitor (resuelto o distribuido).

2. Multiplicar la probabilidad previa de cada hipótesis por su compatibilidad (peso no normalizado).

3. Normalizar los pesos para obtener la distribución posterior; usarla como prior para la siguiente observación.

Ejemplo (caso didáctico):

- P1 (resuelto): aa
- P2 (prior): {AA:0.25, Aa:0.50, aa:0.25}
- Observaciones sucesivas: hijo1 = aa, hijo2 = aa, hijo3 = aa, hijo4 = aa

La evolución de la distribución se muestra en la Tabla 8, donde se observa cómo, tras cada iteración, la probabilidad se concentra progresivamente en los genotipos más compatibles con la evidencia acumulada. En particular, la hipótesis AA se descarta desde la primera observación, mientras que la probabilidad de aa aumenta de manera sostenida con cada nuevo dato.

Tabla 7. Evaluación de compatibilidad y ajuste.

Fuente: Los autores

Hipótesis P2	Prior	Compatibilidad con hijo aa (vía cruce con P1)	Peso no normalizado (Prior × Compat.)	Posterior
AA	0.25	0.0 (aa×AA → sólo produce Aa)	0.00	0.00
Aa	0.50	0.5 (aa×Aa → produce aa en 50% de los casos)	0.25	0.50
aa	0.25	1.0 (aa×aa → produce aa siempre)	0.25	0.50
Suma	—	—	0.50	1

Tabla 7. Evaluación de compatibilidad y ajuste.

Fuente: Los autores

Iteración	Prior (AA, Aa, aa)	Compatibilidad (AA, Aa, aa)	Pesos no normalizados	Posterior (AA, Aa, aa)
1 (hijo1)	(0.25, 0.50, 0.25)	(0.0, 0.5, 1.0)	(0.00, 0.25, 0.25)	(0.00, 0.50, 0.50)
2 (hijo2)	(0.00, 0.50, 0.50)	(0.0, 0.5, 1.0)	(0.00, 0.25, 0.50)	(0.00, 0.333, 0.667)
3 (hijo3)	(0.00, 0.333, 0.667)	(0.0, 0.5, 1.0)	(0.00, 0.1667, 0.6667)	(0.00, 0.200, 0.800)
4 (hijo4)	(0.00, 0.200, 0.800)	(0.0, 0.5, 1.0)	(0.00, 0.10, 0.80)	(0.00, 0.111, 0.889)

Interpretación. Tras la primera observación aa, la hipótesis AA queda eliminada y Aa / aa se igualan. Con cada hijo aa observado, la probabilidad de aa crece rápidamente (≈ 0.889 tras 4 hijos), mostrando concentración de la distribución hacia la hipótesis que mejor explica la evidencia acumulada. Si las observaciones fueran mixtas, la dinámica sería análoga pero menos concentrada.

3.6 Comparación con enfoques existentes y limitaciones

Comparación conceptual: Nuestro enfoque coincide con la filosofía bayesiana clásica de actualización de creencias en función de la evidencia, aplicándola a pedigrís familiares. Por ejemplo, herramientas recientes de inferencia en pedigrí como BICEP permiten calcular probabilidades posteriores de causalidad genética considerando la segregación de variantes dentro de familias (Ormond et al., 2024). De modo análogo, nuestros cruces probabilísticos ponderan exhaustivamente las posibilidades mendelianas según la evidencia genética disponible. Además, los modelos contemporáneos en genética cuantitativa enfatizan la integración de información fenotípica y genealógica como base para la inferencia en contextos de selección (Gianola et al., 2022), lo que concuerda con nuestra estrategia de propagación de evidencia en redes familiares. En una línea más amplia, los enfoques de biología de sistemas aplicados a la producción animal destacan el valor de combinar datos genotípicos y fenotípicos mediante modelos computacionales para mejorar la toma de decisiones (Pathak & Kim, 2024), reforzando el puente que establecemos entre fenotipo observado y genotipo inferido. En el caso específico de caracteres mendelianos, estudios recientes sobre genética equina muestran cómo genes clave determinan directamente rasgos visibles como el color de capa, proporcionando una base biológica clara para modelos simplificados como el propuesto (Liu et al., 2024). La principal ventaja de nuestro método radica en su formulación modular: el núcleo se reduce a una operación única de “cruce + condicionamiento” sobre distribuciones de probabilidad, en contraste con enfoques más complejos basados en múltiples reglas o procedimientos diferenciados.

Limitaciones principales: El modelo actual se restringe a un locus bialélico y asume ausencia de mutación, recombinación y errores de observación; simplificaciones que limitan su aplicabilidad directa a datos reales más complejos. En pedigrís con bucles

o endogamia, el número de combinaciones genotípicas crece rápidamente, incrementando el costo computacional. La extensión a escenarios multilocus implicaría un aumento significativo en la complejidad del espacio de estados. En este contexto, enfoques como la imputación genómica han demostrado ser eficaces para reconstruir genotipos faltantes mediante el uso de paneles de referencia y correlaciones poblacionales (Treccani et al., 2023), lo que sugiere posibles vías de extensión. Asimismo, los modelos bayesianos avanzados suelen incorporar técnicas de muestreo y verosimilitudes probabilísticas para manejar ruido y ambigüedad en los datos (Gianola et al., 2022), aspecto que no está considerado en la formulación actual. En consecuencia, el método se mantiene como una solución conceptual sólida para rasgos mendelianos simples, pero requiere extensiones para abordar escenarios genéticos más complejos.

Líneas futuras: Entre las principales líneas de desarrollo se encuentra la generalización a múltiples loci, donde técnicas modernas de imputación y modelado haplotípico serán especialmente relevantes (Treccani et al., 2023). También resulta prioritario incorporar modelos de observación no deterministas que permitan manejar errores fenotípicos y efectos ambientales. Asimismo, será necesario realizar evaluaciones comparativas con herramientas existentes mediante benchmarking en datos simulados y reales. Estudios recientes en genética equina han demostrado el valor de la imputación para reconstruir información genética histórica y relacionarla con rasgos fenotípicos (Todd et al., 2023), lo que evidencia el potencial de integrar múltiples fuentes de información en la inferencia.

3.7 Evaluación experimental

Para medir la capacidad del método para inferir correctamente distribuciones genotípicas a partir de información fenotípica y genealógica incompleta, así como analizar su comportamiento de convergencia bajo evidencia acumulativa. Se generaron 5.000 repeticiones independientes de un escenario genético basado en un locus bialélico (A/a). En cada repetición se simuló aleatoriamente los genotipos verdaderos de dos progenitores siguiendo una distribución inicial de frecuencias (AA = 0.25, Aa = 0.50, aa = 0.25).



El algoritmo recibió únicamente la información fenotípica derivada del descendiente y distribuciones iniciales uniformes para los progenitores, debiendo inferir la distribución de probabilidad asociada al genotipo del descendiente. Este procedimiento permitió comparar sistemáticamente las inferencias obtenidas con los genotipos verdaderos utilizados en la simulación.

Las métricas que se midieron en el ejercicio fueron: exactitud (Accuracy), proporción de casos en los que el genotipo con mayor probabilidad posterior coincidió con el genotipo verdadero; error Absoluto Medio (MAE), diferencia promedio entre la distribución inferida y la representación exacta del genotipo verdadero; entropía, medida de incertidumbre asociada a la distribución posterior inferida; las mismas que se presentan en la Tabla 9.

Tabla 9. Métricas del desempeño del algoritmo.
 Fuente: Los autores

Métrica	Valor
1 Exactitud	0.7576
2 Valor Absoluto Medio	0.2189
3 Entropía promedio	0.6891

Interpretación. La exactitud global alcanzó un valor de 75.76 %, indicando que en aproximadamente tres de cada cuatro simulaciones el genotipo con mayor probabilidad posterior coincidió con el genotipo verdadero. El error absoluto medio de 0.2189 muestra que las distribuciones inferidas permanecieron razonablemente próximas a los genotipos reales, mientras que la entropía promedio de 0.6821 refleja una reducción significativa de la incertidumbre respecto a una distribución completamente uniforme.

La Tabla 10 evidencia que el genotipo recesivo aa fue identificado correctamente en la totalidad de los casos simulados, alcanzando precisión y sensibilidad iguales a 1.0. Por el contrario, el genotipo AA no pudo distinguirse de manera confiable utilizando únicamente información fenotípica, debido a que comparte el mismo fenotipo observable que el genotipo heterocigoto Aa bajo el modelo dominante-recesivo considerado. Esta limitación corresponde a la información disponible y no a la formulación matemática del algoritmo.

Tabla 10. Probabilidades inferidas por el algoritmo.
 Fuente: Los autores

Hijo real	Prob. AA	Prob. Aa	Prob. aa	Error Absoluto Medio	Entropía
AA	0.33	0.66	0.00	0.44	0.9183
Aa	0.33	0.66	0.00	0.22	0.9182
aa	0.00	0.00	1.00	0.00	0.00

Interpretación. La incertidumbre del algoritmo se concentra en la distinción entre AA y Aa, mientras que el genotipo aa es identificado sin ambigüedad.

Se estableció un máximo de 100 iteraciones y un criterio de convergencia basado en una variación máxima inferior a 10^{-4} entre distribuciones posteriores consecutivas.

El algoritmo alcanzó la convergencia en la iteración 15. Simultáneamente, la entropía disminuyó desde valores cercanos de incertidumbre inicial hasta 0.000942.

Estos resultados confirman que el procedimiento de actualización bayesiana presenta estabilidad numérica y una rápida capacidad de convergencia casi lineal, permitiendo incorporar nueva evidencia sin introducir inconsistencias probabilísticas ni oscilaciones en las distribuciones posteriores, según la Figura 1

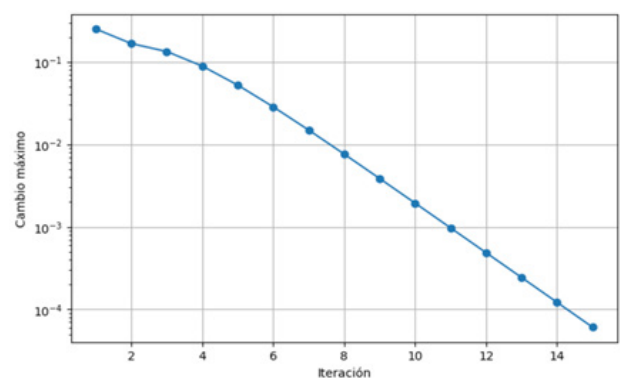


Figura 1. Gráfica de convergencia del algoritmo
 Fuente: Los autores

Interpretación. Se observa una reducción monotónica del cambio máximo, pasando de valores cercanos a 2.5×10^{-1} en las primeras iteraciones hasta aproximadamente 6×10^{-4} en la iteración 15. Este comportamiento evidencia que las actualizaciones bayesianas producen distribuciones progresivamente más estables a medida que se incorpora nueva evidencia. La ausencia de oscilaciones o incrementos en la magnitud del cambio confirma la estabilidad

numérica del algoritmo.

Los ejercicios experimentales realizados muestran que el algoritmo reproduce correctamente las relaciones mendelianas esperadas, propaga coherentemente la incertidumbre cuando la información parental es incompleta y converge rápidamente al incorporar evidencia adicional. La principal limitación observada corresponde a la imposibilidad de diferenciar completamente entre genotipos AA que producen el mismo fenotipo observable, situación inherente a cualquier procedimiento de inferencia basado exclusivamente en información fenotípica. A pesar de esta restricción, los resultados obtenidos respaldan la validez metodológica del enfoque híbrido propuesto y su potencial aplicación en contextos donde los datos genéticos directos no se encuentran disponibles.

4. Conclusiones

Este trabajo demuestra que la inferencia genética que los humanos realizan de manera intuitiva puede formalizarse de forma directa mediante un procedimiento computacional claro, preciso y reproducible. El algoritmo propuesto no introduce supuestos estadísticos adicionales ni modelos probabilísticos opacos, sino que sistematiza reglas de compatibilidad genética y actualización distributiva basadas en las leyes mendelianas. De este modo, la inferencia se construye como un proceso acumulativo y coherente, donde cada nueva observación refina el conocimiento previo sin perder trazabilidad ni interpretabilidad.

La principal relevancia de este enfoque radica en su capacidad para servir como puente entre la comprensión conceptual del problema genético y su implementación algorítmica. Al mantener una correspondencia directa entre la intuición biológica y las operaciones computacionales, el método facilita su adopción en sistemas automatizados, entornos educativos y aplicaciones donde la explicabilidad del razonamiento es un requisito central. Asimismo, este marco abre la puerta a futuras extensiones hacia pedigrís más complejos y escenarios con múltiples loci, manteniendo la misma filosofía de inferencia transparente y progresiva.

Contribución de los autores

Alexander Stefano Mora Matamoros: Redacción – borrador original del artículo, Metodología, Investigación. **Javier Hernán López Zambrano:** Validación, Investigación, Redacción – revisión y edición del artículo. **Carlos Octavio Larrea Izurieta:** Conceptualización, Investigación, Validación. **Diego Alexander Toala Palma:** Conceptualización, Investigación, Validación.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

- Almudevar, A., & LaCombe, J. (2012). On the choice of prior density for the Bayesian analysis of pedigree structure. *Theoretical Population Biology*, 81(2), 131–143. <https://doi.org/10.1016/j.tpb.2011.12.003>
- Gianola, D., Fernando, R. L., & Schön, C. C. (2022). Inference about quantitative traits under selection: A Bayesian revisit for the post-genomic era. *Genetics Selection Evolution*, 54(1), 78. <https://doi.org/10.1186/s12711-022-00765-z>
- Liu, X., Peng, Y., Zhang, X., Wang, X., Chen, W., Kou, X., Liang, H., Ren, W., Khan, M. Z., & Wang, C. (2024). Coloration in Equine: Overview of Candidate Genes Associated with Coat Color Phenotypes. *Animals: An Open Access Journal from MDPI*, 14(12), 1802. <https://doi.org/10.3390/ani14121802>
- Modak, M., Mustahasin Pritom, M., Chandra Banik, S., & Rabbi, S. (2025). Internet of Things-Based Health Surveillance Systems for Livestock: A Review of Recent Advances and Challenges. *IET Wireless Sensor Systems*, 1(15). <https://doi.org/10.1049/wss2.70013>
- Ormond, C., Ryan, N. M., Cap, M., Byerley, W., Corvin, A., & Heron, E. A. (2024). BICEP: Bayesian inference for rare genomic variant causality evaluation in pedigrees. *Briefings in Bioinformatics*, 26(1), bbae624. <https://doi.org/10.1093/bib/bbae624>
- Pathak, R. K., & Kim, J.-M. (2024). Veterinary systems biology for bridging the phenotype–genotype gap via computational modeling for disease epidemiology and animal welfare. *Briefings in Bioinformatics*, 25(2), bbae025. <https://doi.org/10.1093/bib/bbae025>
- Saha, S., Layek, S., Ray, H., Ghosh, A., Majumdar, S., Ghosh, D., Chakraborty, R., Pramanik, A., Parua, T., & Mukherjee, S. (2024). Recent Trends of IoT Enabled Wearable Sensors: An Extensive Application of Cattle Health Monitoring. *Lecture Notes in Networks and Systems*, 1551, 201–214. https://doi.org/https://doi.org/10.1007/978-981-96-9983-4_15
- Sponenberg, D. P., & Bellone, R. (2017). *Equine color genetics* (4th edition). Wiley Blackwell.
- Todd, E. T., Fromentier, A., Sutcliffe, R., Running Horse Collin, Y., Perdereau, A., Aury, J.-M., Èche, C., Bouchez, O., Donnadiou, C., Wincker, P., Kalbfleisch, T., Petersen, J. L., & Orlando, L. (2023). Imputed genomes of historical horses provide insights into modern breeding. *iScience*, 26(7), 107104. <https://doi.org/10.1016/j.isci.2023.107104>
- Treccani, M., Locatelli, E., Patuzzo, C., & Malerba, G. (2023). A broad overview of genotype imputation: Standard guidelines, approaches, and future investigations in genomic association studies. *BIOCELL*, 47(6), 1225–



1241. <https://doi.org/10.32604/biocell.2023.027884>
Ullah, E., Mall, R., Abbas, M. M., Kunji, K., Nato, A. Q., Bensmail, H., Wijsman, E. M., & Saad, M. (2019). Comparison and assessment of family- and population-based genotype imputation methods in large pedigrees. *Genome Research*, 29(1), 125–134. <https://doi.org/10.1101/gr.236315.118>

Zhang, Y., Jiang, B., Zhu, J., & Liu, J. S. (2011). Bayesian Models for Detecting Epistatic Interactions from Genetic Data: Bayesian partition for epistasis mapping. *Annals of Human Genetics*, 75(1), 183–193. <https://doi.org/10.1111/j.1469-1809.2010.00621.x>