



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Ciencia Animal

Análisis y detección de variantes estructurales en el
genoma de *Bos taurus*

Trabajo Fin de Máster

Máster Universitario en Mejora Genética Animal y Biotecnología de
la Reproducción

AUTOR/A: Fructuoso Caballero, Jose Manuel

Tutor/a: Ibáñez Escriche, Noelia

Cotutor/a externo: González Recio, Oscar

Director/a Experimental: López Catalina, Adrián

CURSO ACADÉMICO: 2023/2024



Máster Interuniversitario en Mejora Genética Animal y
Biotecnología de la Reproducción

ANÁLISIS Y DETECCIÓN DE VARIANTES ESTRUCTURALES EN EL GENOMA DE *BOS TAURUS*

Autor: José Manuel Fructuoso Caballero

Tutor académico: Noelia Ibañez-Escriche

Tutor externo: Óscar González Recio

Tutor experimental: Adrián López Catalina

Curso 2023-2024

INDICE

Indice de figuras	4
Indice de tablas	5
Resumen	6
Abstract	7
Resum.....	8
Abreviaturas	9
1. Introducción.....	10
1.1. Breve historia de la genética	10
1.2. Variaciones genéticas.....	11
1.2.1. Variantes estructurales	12
1.3. Métodos de secuenciación	13
1.4. Alineación frente a genomas de referencia.....	16
1.5. Programas de detección de SVs.....	17
1.6. Importancia de las variantes estructurales en mejora genética	19
1.6.1. Ejemplos de SVs relacionadas con el ganado	19
1.6.2. Imputación de SV en poblaciones comerciales.....	20
1.6.3. Aplicaciones de las SVs en programas de mejora de ganado vacuno.....	21
2. Objetivo.....	22
2.1. Objetivos específicos	22
3. Materiales y métodos.....	23
3.1. Obtención de datos	23
3.2. Análisis bioinformático.....	24
3.2.1. Basecalling.....	24
3.2.2. Alineamiento.....	24
3.2.3. Variant calling.....	24
3.2.4. Análisis de razas	25
3.2.5. Análisis de genes	25
3.3. Estudio comparativo	26

3.3.1. Comparación de alineadores	26
3.3.2. Comparación de variant callers.....	26
4. Resultados y Discusión.....	27
4.1. Estudio comparativo	27
4.2. Análisis de razas	33
4.2.1. Análisis de genes	40
5. Conclusiones	44
6. Bibliografía.....	46

Índice de figuras

Figura 1. Cronología de los campos de la genética molecular y cuantitativa. La figura muestra las conexiones interdisciplinarias que condujeron a dos de los descubrimientos más revolucionarios de la genética en la última década, el mapeo genético y la predicción genómica. Extraído de Nelson et al. (2013).....	11
Figura 2. Clasificación de los tipos de SVs. Extraído de Gong et al. (2021).....	13
Figura 3. Secuenciación ONT basada en nanoporos. La cadena de ADN pasa por el poro generando una interrupción en la corriente eléctrica que se convierte en un patrón denominado “squiggle” a tiempo real. Después, se asigna a cada valor una base nitrogenada.	16
Figura 4. Gráficos de precisión de alineamiento superior al 95% en las muestras de HOL13 (arriba) y WAG2 (abajo). LRA en azul, Minimap2 en rojo.	28
Figura 5. Abundancia relativa de SVs detectadas con Sniffles2 y CuteSV, alineadas con Minimap2 y LRA en las muestras de HOL13 y WAG2. Las siglas representan un código de 3 letras para identificar las 8 combinaciones. S=Sniffles2, C=CuteSV, M=Minimap2, L=LRA, H=HOL13 y W=WAG2.	29
Figura 6. Gráfico de densidad de SVs detectadas con Sniffles2 y CuteSV, alineadas con Minimap2 y LRA en las muestras de HOL13 y WAG2. Las siglas representan un código de 3 letras para identificar las 8 combinaciones. S=Sniffles2, C=CuteSV, M=Minimap2, L=LRA, H=HOL13 y W=WAG2.	32
Figura 7. Número y tipo de SVs detectadas en cada animal. Cada color representa una de las cuatro razas.....	34
Figura 8. Diagrama de Venn mostrando las SVs únicas comunes entre razas.	35
Figura 9. Abundancia relativa de SVs detectadas en razas Holstein (HOL), Wagyu (WAG), Pirenaica Española (PIR) y Red Holstein (RED). Los colores representan el tipo de variante estructural. INS: Inserciones; DEL: Deleciones; INV: Inversiones; DUP: Duplicaciones; BNDs: Breakends.	36
Figura 10. Abundancia relativa de cada tipo de SV por cromosoma detectadas en razas Holstein (HOL), Wagyu (WAG), Pirenaica Española (PIR) y Red Holstein (RED). Los colores representan el tipo de variante estructural. INS: Inserciones; DEL: deleciones; INV: Inversiones; DUP: Duplicaciones; BNDs: Breakends.	36
Figura 11. Gráfico de densidad de SVs por cromosoma según su posición de inicio en la población total analizada. Los colores representan cada uno de los cromosomas.....	37
Figura 12. Anotación genómica de SVs por raza. Los colores representan la región afectada por las SVs. De izquierda a derecha: Regiones promotoras, 5' y 3' no traducibles, exónicas, intrónicas y zonas intergénicas.....	38

Indice de tablas

Tabla 1. Razas y animales empleados en el estudio.	23
Tabla 2. Profundidad, cobertura y precisión media en las muestras de HOL13 y WAG2 alineadas con Minimap2 y LRA.....	27
Tabla 3. Número de inserciones, deleciones, BNDs, inversiones y duplicaciones y número total de SVs detectadas con Sniffles2 y CuteSV compartidas según su alineador en las muestras de HOL13 y WAG2 alineadas tanto con Minimap2 como con LRA con un rango de ± 100 pb.	30
Tabla 4. Número de SVs en común entre metodologías según el rango de las posiciones de inicio y fin de las SVs. Las siglas representan un código de 3 letras para identificar las 8 combinaciones. S=Sniffles2, C=CuteSV, M=Minimap2, L=LRA, H=HOL13 y W=WAG2.	30
Tabla 5. Profundidad, Cobertura, Número y Tipo de SVs identificadas en cada individuo. HOL: Holstein; WAG: Wagyu; RED: Red Holstein; Pir: Pirenaica Española; INS: Inserciones; DEL: Deleciones; INV: Inversiones; DUP: Duplicaciones; BNDs: Breakends.	33
Tabla 6. SVs únicas encontradas de cada tipo y total por raza.	34
Tabla 7. Porcentaje de SVs en cada región genómica en las cuatro razas.	38
Tabla 8. Genes relacionados con la región exónica afectada por SVs exclusivas en cada raza....	42
Tabla 9. Genes relacionados con la región exónica afectada por SVs compartidas entre las razas Holstein y Red Holstein y las razas Wagyu y Pirenaica Española.	43
Tabla 10. Genes relacionados con la región exónica afectada por SVs compartidas entre las cuatro razas estudiadas.	43

Resumen

Las variantes estructurales son un tipo de variaciones genéticas que pueden tener impacto directo o indirecto en el fenotipo de los organismos y pueden capturar varianza genética diferente a la de los SNPs. Gracias al continuo avance en las tecnologías de secuenciación, como la desarrollada por Oxford Nanopore Technologies basada en nanoporos, es posible detectar y analizar variantes estructurales de mayores tamaños. El objetivo principal de este trabajo de fin de máster fue detectar SVs en cuatro razas de ganado bovino con lecturas largas, comparando diferentes softwares de alineamiento y variant calling para establecer similitudes y diferencias entre razas e intentar relacionar las variantes estructurales con caracteres de interés en mejora genética. Se realizaron alineamientos de 2 de las 21 muestras con los softwares Minimap2 y LRA. Posteriormente, se utilizaron los softwares Sniffles2 y CuteSV para la detección de SVs. Estos resultados mostraron que la combinación de Minimap2/CuteSV fue la que mayor número de variantes estructurales logró detectar. El análisis comparativo entre las razas Holstein, Red Holstein, Pirenaica Española y Wagyu reveló que la mayor parte de las SVs se distribuyen principalmente en regiones intergénicas e intrónicas, siendo menos frecuentes en regiones promotoras y exones. Se encontraron variantes estructurales afectando a regiones relacionadas con genes de interés en producción, revelando la posible implicación estudio de las variantes estructurales con la mejora genética y su implementación en programas de selección asistida por marcadores.

Palabras clave: Variantes Estructurales, Ganado Vacuno, Mejora genética, Producción Animal.

Abstract

Structural variants are a type of genetic variation that can have a direct or indirect impact on the phenotype of organisms and can capture genetic variance different from SNPs. Thanks to the continuous advances in sequencing technologies, such as the one developed by Oxford Nanopore Technologies based on nanopores, it is possible to detect and analyze structural variants of larger sizes. The main objective of this master's thesis was to detect structural variants in four cattle breeds with long reads, comparing different alignment and variant calling softwares to establish similarities and differences between breeds and try to relate structural variants with traits of interest in genetic improvement. Alignments of 2 of the 21 samples were performed with Minimap2 and LRA software. Subsequently, Sniffles2 and CuteSV software were used to detect SVs. These results showed that the Minimap2/CuteSV combination had the highest number of structural variants detected. Comparative analysis between Holstein, Red Holstein, Spanish Pyrenean and Wagyu breeds revealed that most of the SVs are distributed mainly in intergenic and intronic regions, being less frequent in promoter regions and exons. Structural variants were found affecting regions related to genes of interest in production, revealing the possible implication of the study of structural variants with genetic improvement and their implementation in marker-assisted selection programs.

Key words: Structural Variants, Beef Cattle, Genetic Improvement, Animal Breeding.

Resum

Les variants estructurals, com els polimorfismes d'un nucleòtid, són un tipus de variacions genètiques que poden tindre impacte directe o indirecte en el fenotip dels organismes i poden capturar variància genètica diferent de la dels SNPs. Gràcies al continu avanç en les tecnologies de seqüenciació, com la desenvolupada per Oxford Nanopore Technologies basada en nanoporos, és possible detectar i analitzar variants estructurals de majors grandàries. L'objectiu principal d'este treball de fi de màster va ser detectar variants estructurals en quatre races de bestiar boví detectades amb lectures llargues, comparant diferents softwares d'alineament i variant calling per a establir similituds i diferències entre races i intentar relacionar les variants estructurals amb caràcters d'interés en millora genètica. Es van realitzar alineaments de 2 de les 21 mostres amb els softwares Minimap2 i LRA. Posteriorment, es van utilitzar els softwares Sniffles2 i CuteSV per a la detecció de variants estructurals. Estos resultats van mostrar que la combinació de Minimap2/CuteSV va ser la que major nombre de variants estructurals va aconseguir detectar. L'anàlisi comparativa entre les races Holstein, Xarxa Holstein, Pirinenca Espanyola i Wagyu va revelar que la major part de les variants estructurals es distribuïxen principalment en regions intergèniques i intròniques, sent menys freqüents en regions promotores i exons. Es van trobar variants estructurals afectant regions relacionades amb gens d'interés en producció, revelant la possible implicació estudiada de les variants estructurals amb la millora genètica i la seua implementació en programes de selecció assistida per marcadors.

Paraules clau: Variants estructurals, Boví de carn, Millora Genètica, Producció Animal.

Abreviaturas

ADN – Ácido desoxirribonucleico

ARN – Ácido ribonucleico

BLUP – Mejor predictor lineal insesgado (del inglés *Best Linear Unbiased Predictor*)

BNDs – Extremos rotos (del inglés *Breakends*)

BovLRC – Consorcio de lecturas largas bovino (del inglés *Bovine Long Reads Consortium*)

CNVs – Variaciones del número de copias (del inglés *Copy Number Variations*)

DEL – Delección (del inglés *Deletion*)

DINS – Inserciones domésticas (del inglés *Domestic Insertion*)

DUP – Duplicación (del inglés *Duplication*)

FINS – Inserción foránea (del inglés *Foreign Insertion*)

HMM – Modelos de cadenas ocultas de Markov (del inglés *Hidden Markov Models*)

HOL – Holstein

INS – Inserción (del inglés *Insertion*)

INV – Inversión (del inglés *Inversion*)

KEGG – Enciclopedia de genes y genomas (del inglés *Encyclopedia of Genes and Genomes*)

NCBI – Centro Nacional de Investigación Biotecnológica (del inglés *National Center for Biotechnology Information*)

NGS – Secuenciación de Nueva Generación (del inglés *Next Generation Sequencing*)

ONT – *Oxford Nanopore Technologies*

PIR – Pirenaica Española

QTL – Locus de carácter cuantitativo (del inglés *Quantitative Trait Locus*)

RED – Holstein Roja (del inglés *Red Holstein*)

SNP – Polimorfismos de un solo nucleótido (del inglés *Single Nucleotide Polimorphism*)

SVs – Variantes Estructurales (del inglés *Structural Variants*)

VCF – Formato de llamado de variantes (del inglés *Variant Call Format*)

WAG – Wagyu

1. Introducción

1.1. Breve historia de la genética

Gregor Mendel estableció las bases de la genética gracias a sus investigaciones sobre el cruce de guisantes, mediante los cuales pudo relacionar el genotipo y el fenotipo de los organismos (Mendel, 1865). Por esas fechas, Francis Galton teorizaba sobre cómo se heredaban los caracteres fenotípicos, proponiendo que no se trataba de un proceso continuo como establecía Darwin, sino de saltos discontinuos (Galton, 1876). Tiempo después, en 1918, Ronald Fisher integró la genética clásica de Mendel con conceptos de la selección natural, sugiriendo que la variabilidad fenotípica en una población se debía a la pequeña contribución individual de múltiples genes con herencia mendeliana, influenciados por factores ambientales surgiendo así la genética cuantitativa, enfocada en la variación continua de caracteres fenotípicos según el modelo infinitesimal (Fisher, 1919). Este modelo plantea que el componente genético de los caracteres sigue una distribución normal, basada en la media de los caracteres de los progenitores (Barton et al., 2017) y ha dominado el campo de la genética cuantitativa durante más de 50 años asumiendo que, como el efecto de cada uno de los locus es irreconocible, se deben estudiar de forma conjunta (Hu et al., 2012). Pero fue a partir de la década de 1970 cuando los avances y mejoras en las tecnologías y métodos de secuenciación comenzaron a aportar datos moleculares a la genética. Posteriormente, con la unión de estas dos disciplinas (Figura 1), fue posible el desarrollo de la predicción genómica y el mapeo de QTLs (por sus siglas en inglés *Quantitative Trait Locus*) (Nelson et al., 2013). Gracias a la predicción genómica (también conocida como selección genómica), identificada en primer lugar por Meuwissen et al. (2001), se puede determinar el fenotipo de los organismos empleando marcadores moleculares a lo largo del genoma que proporcionan información sobre el posible rendimiento de los organismos basándose en su genotipo (Feng et al., 2024). Utilizando esta información, junto con la aportada por chips de genotipado que permiten la identificación de variaciones genéticas en el genoma a gran escala (Ganal et al., 2012) y complementándola con datos de QTLs integrados en modelos estadísticos como BLUP (por sus siglas en inglés *Best linear unbiased prediction*), se ha logrado optimizar la mejora asistida por marcadores, la cual permite seleccionar a los individuos con las mejores características para sus respectivos programas de cría (Meuwissen et al., 2016).

Otro tipo de variaciones genéticas son las variantes estructurales, también conocidas como SVs (por sus siglas en inglés *Structural Variants*). La detección y el estudio de SVs es crucial para identificar patrones que puedan vincularlas con algunos fenotipos particulares, entre los que pueden incluirse enfermedades (Xia et al., 2023) y algunos caracteres de producción en especies ganaderas (Bickhart & Liu, 2014).

1.2.1. Variantes estructurales

En la actualidad, la gran cantidad de técnicas moleculares existentes, la aparición de tecnologías como los microarrays y el desarrollo de la secuenciación de nueva generación, o NGS (*Next-Generation Sequencing*), han permitido la caracterización de las SVs, cuyo tamaño oscila entre 50 pares de bases (pb) y Megabases (Mb) (Sudmant et al., 2015).

Según Gong et al. (2021) las SVs pueden clasificarse a grandes rasgos en los siguientes tipos (Figura 2):

- **Inserciones:** Incorporan una secuencia nueva con respecto a la secuencia de genoma que se usa como referencia. Si lo que se inserta es una secuencia procedente de otro lugar del mismo genoma se le denomina DINS (de sus siglas en inglés *Domestic Insertion*), si se ha añadido una secuencia nueva de la que no se sabe si su origen es el genoma de la propia muestra se le denomina FINS (de sus siglas en inglés *Foreign Insertion*)
- **Deleciones:** Implican la pérdida de parte de material genético con respecto al genoma que se emplea como referencia, pudiendo tratarse de un solo nucleótido o hasta grandes segmentos de un cromosoma.
- **Duplicaciones:** Consisten en la copia un segmento de ADN en otro lugar del genoma. Si ese lugar es contiguo a la secuencia de origen se denomina duplicación en tándem.
- **Inversiones:** Son segmentos del ADN que cambian el sentido de su orientación con respecto al resto de la secuencia sin que se produzca ni ganancia ni pérdida de material genético.
- **Translocaciones:** Son cambios en la posición de un segmento de ADN que se reinserta en otro lugar del genoma, en los que no se produce ni ganancia ni pérdida de ADN total del genoma.

- Breakends (BNDs): Clasificadas en ocasiones como un tipo de translocaciones. También puede referirse a las SVs que suelen ser identificadas mediante lecturas cortas, pero cuya estructura real aún no se conoce con certeza, pudiendo ocasionar problemas en los análisis de datos (Abel et al., 2019).

Alkan et al. (2011) consideran que inserciones, deleciones y duplicaciones pueden englobarse dentro de la categoría CNVs (por sus siglas en inglés *Copy Number Variations*) puesto que afectan al número total de copias del genoma de un organismo.

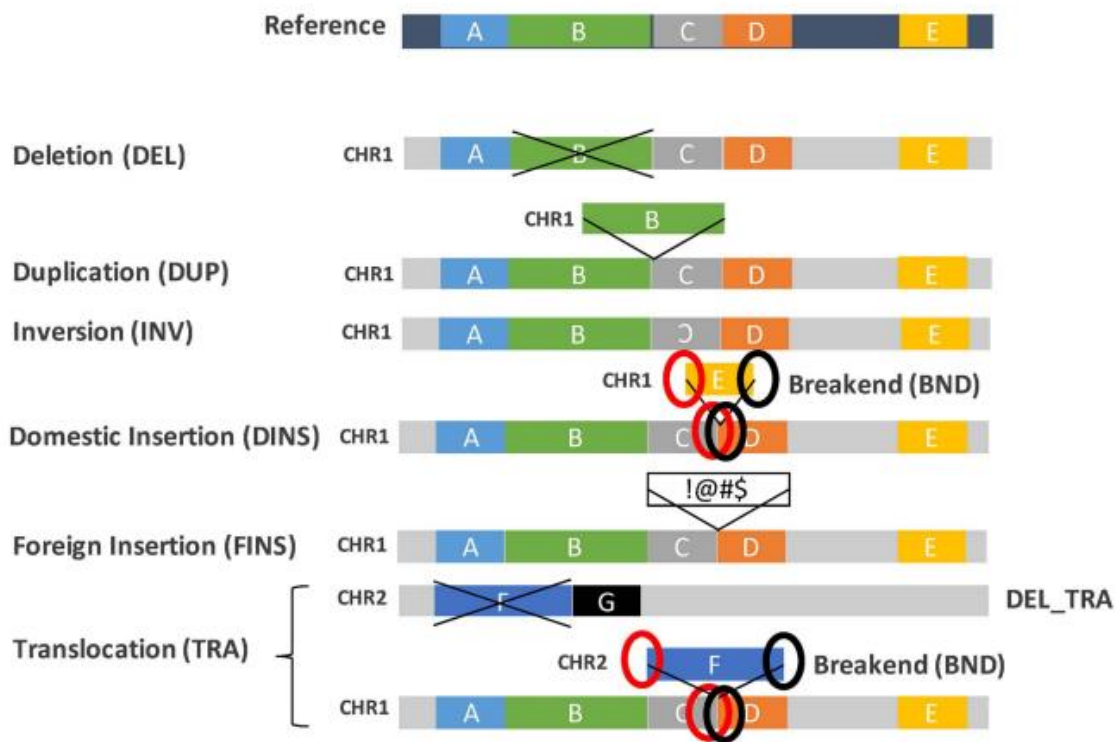


Figura 2. Clasificación de los tipos de SVs. Extraído de Gong et al. (2021).

1.3. Métodos de secuenciación

La secuenciación genómica es fundamental en la detección de SVs. Su creciente importancia se atribuye y a los continuos avances en las técnicas de detección, puesto que en los últimos años Oxford Nanopore Technologies (ONT) ha desarrollado e implementado una nueva técnica basada en la secuenciación por nanoporos, pequeñas

aberturas proteicas incrustadas en una membrana resistente a la electricidad, que funcionan de forma similar a un biosensor. Cuando una cadena de ADN o ARN atraviesa el nanoporo induce un cambio en la corriente, permitiendo la identificación del nucleótido correspondiente (Figura 3). De esta forma, empleando dispositivos como MinION, GridION y PromethION, se logran obtener secuencias de ADN más largas (*long reads*) y a tiempo real (Jain et al., 2016). Además, la secuenciación se ha ido desarrollando gracias a dos enfoques diferentes: la secuenciación con lecturas cortas (*short reads*) y la secuenciación con lecturas largas. Cada una de ellas con sus propias características, ventajas y desventajas.

- **Lecturas cortas:**

La secuenciación empleando lecturas cortas ha dominado la mayor parte de los datos disponibles hasta la actualidad, haciendo difícil la identificación en regiones complejas del genoma que suelen ser puntos calientes de SVs y que presentan multitud de regiones con repeticiones (Mills et al., 2011). El método más estandarizado es el de lecturas pareadas, realizando la lectura en direcciones opuestas y tras haber mapeado todas las lecturas, pueden utilizarse la longitud de alineación, el tamaño y la orientación de la misma para identificar las posibles SVs existentes (Mahmoud et al., 2019).

Los métodos para detectar las SVs basadas en lecturas cortas cambian dependiendo del tipo de información que emplean. Algunos de los primeros métodos utilizaban tan solo la distancia y orientación de las lecturas pareadas, es decir, teniendo en cuenta la secuenciación de los dos extremos de cada fragmento de ADN, pero esto podía resultar en la omisión de algunas SVs de tamaños más pequeños (Chen et al., 2009).

Si lo que se quiere es aumentar la resolución de la secuenciación, pueden emplearse las lecturas divididas, las cuales contienen partes que mapean en otros loci del genoma de referencia y dividen la lectura en segmentos más pequeños, alineándolos de forma individual. Este tipo de secuenciación tiene una mayor precisión y permite encontrar deleciones de menor tamaño, con la desventaja de que las más grandes son complicadas de diferenciar (Rausch et al., 2012). Por eso, algunos programas que se hicieron posteriormente permiten realizar análisis en conjunto tanto de profundidad de lectura como de las discordancias de las lecturas divididas (Layer et al., 2014). Aunque

existan multitud de programas que han sido capaces de optimizar el mapeo con lecturas cortas, ninguno de ellos es capaz de detectar todos los diferentes tipos de SVs. Por tanto, es necesario que se combinen varias de estas herramientas en los denominados meta-métodos, los cuales son capaces de emplear los puntos fuertes de estos métodos y superar sus puntos negativos. Pero, aunque usando los métodos que implican lecturas cortas seamos capaces de tener una buena precisión, no será posible obtener las SVs más complejas (Audano et al., 2019).

- **Lecturas largas:**

La secuenciación empleando lecturas largas tiene más ventajas a la hora de detectar las SVs porque pueden abarcar regiones que presentan problemas, como las regiones que poseen repeticiones. Por esto, las lecturas largas son capaces de mejorar los mapeos y encontrar SVs de mayor tamaño que las secuenciaciones con lecturas cortas no serían capaces de localizar (Sedlazeck et al., 2018).

Además, las lecturas largas originan menor número de fragmentos que resultan más fáciles de ensamblar y que poseen una cobertura más amplia, minimizando así los errores. Los principales problemas que presentaba la secuenciación de lecturas largas eran su precisión relativamente baja y su elevado coste en comparación con las lecturas cortas (Kim et al., 2024). Por suerte, recientes avances en las nuevas metodologías de ONT han logrado incrementar la precisión de las secuenciaciones hasta un 99,5% (Amarasinghe et al., 2020). También se han desarrollado programas como Minimap2 que se especializan en el alineamiento de las secuencias obtenidas con lecturas largas (Li, 2018).

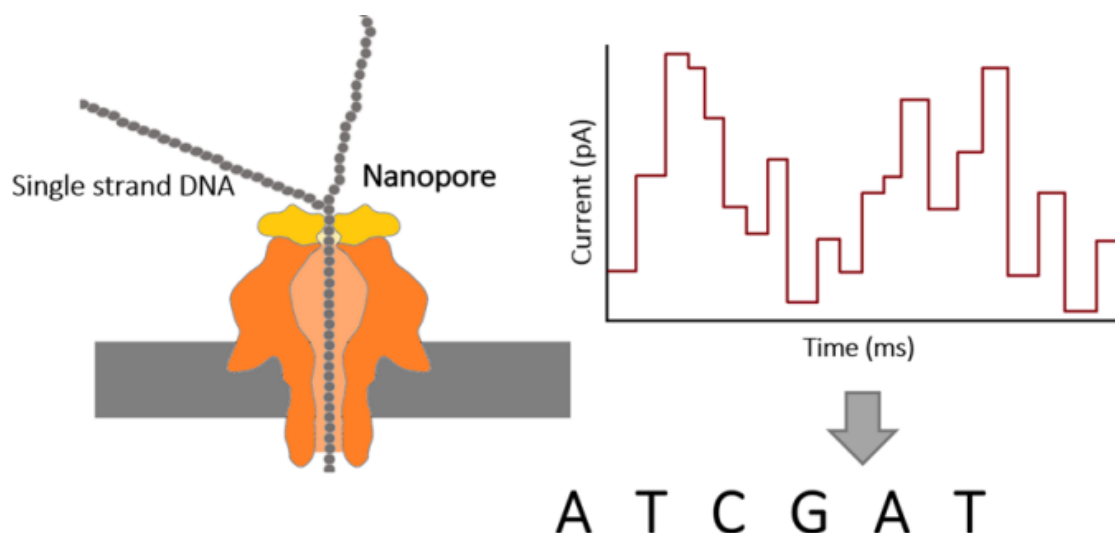


Figura 3. Secuenciación ONT basada en nanoporos. La cadena de ADN pasa por el poro generando una interrupción en la corriente eléctrica que se convierte en un patrón denominado “squiggle” a tiempo real. Después, se asigna a cada valor una base nitrogenada.

1.4. Alineación frente a genomas de referencia

Los estudios que pretenden investigar las variaciones genéticas comienzan con el alineamiento de las lecturas obtenidas en el proceso de secuenciación obtenidas con las de genomas de referencia para realizar una inferencia posterior en las variaciones según existan o no diferencias con respecto a la referencia. De esta forma, se han desarrollado diferentes algoritmos, programas y herramientas que permiten los alineamientos de una forma más sencilla (LoTempio et al., 2023). Alinear estas lecturas implica compararlas con la secuencia de referencia de una población para poder así entender los cambios en su estructura genética. Estos métodos pueden ser de alineamiento global, cuando se pretende alinear toda la secuencia de ADN de principio a fin, y de alineamiento local, para comparar secuencias que contienen regiones con similitudes (Huang & Chao, 2003). Por otra parte, existen los alineamientos de lecturas de NGS, generadas por las nuevas tecnologías de secuenciación. Este tipo de alineamientos identifica las semillas (*seeds*) que coinciden con las secuencias que están alineándose. Después, estas semillas se emplean como lugar de anclaje para extender y encadenar las secuencias, formar alineamientos más largos y finalmente refinar los alineamientos (Kielbasa et al., 2011). Algunos ejemplos de softwares de alineamiento que emplean el método de alineamiento de lecturas de NGS son:

- **Minimap2:** Este software sigue el proceso de “seed, chain, align”, es decir, indexa las secuencias de referencia y para cada una de las secuencias de la muestra encuentra las coincidencias con la referencia que puedan servirle como nuevos puntos de anclaje sobre los que repetir el proceso. Utiliza los algoritmos de indexación que tenía su predecesor Minimap (Li, 2016), pero con una precisión mayor. Se ha observado que Minimap2 funciona con lecturas largas y con mucho ruido. Además, es bastante eficiente y es capaz de encontrar solapamientos entre lecturas largas y realizar búsquedas en bases de datos como NCBI (Li, 2018).
- **LRA:** Utiliza el mismo enfoque de “seed, chain, align” que utiliza Minimap2 para alinear lecturas cortas y largas (Sedlazeck et al., 2013). Esta herramienta agrupa un gran número de fragmentos de pequeño tamaño en un superfragmento que, tras localizar una alineación aproximada, vuelve a iterar para calcular otro conjunto nuevo de coincidencias en la alineación, pero con diferentes puntos de anclaje. El tiempo de ejecución de este software es menor que el de otros alineadores, requiere menos uso de memoria que Minimap2 y, aunque tienen un rendimiento de detección de SVs similar, los alineamientos con LRA son capaces de localizar SVs de mayor tamaño (Ren & Chaisson, 2021).

1.5. Programas de detección de SVs

Los programas de detección de SVs son herramientas computacionales que han sido diseñadas para la identificación variaciones genéticas. En estudios anteriores se solían emplear dos enfoques diferentes para detectar SVs. El primero de ellos es el enfoque basado en ensamblaje de novo, este método busca ensamblar las lecturas obtenidas con otros segmentos genómicos de mayor tamaño y posteriormente encontrar las SVs entre los alineamientos de estas secuencias con una referencia (Seo et al., 2016). El segundo enfoque está basado en el alineamiento de lecturas y tiene el objetivo de alinear directamente las lecturas con la referencia para detectar las SVs. Este segundo enfoque ha sido utilizado en múltiples softwares de detección de SVs basados en lecturas largas (Jiang et al., 2020).

A continuación, se exponen los programas de detección de SVs utilizados en este trabajo de fin de máster:

- **CuteSV:** Se trata de un programa rápido y sensible para la detección de SVs empleando lecturas largas, está basado en el enfoque de alineación de lecturas. Emplea métodos específicos para recopilar las firmas de varios tipos de SVs, eliminar el ruido y las lecturas de baja calidad y después agrupa y refina las firmas obtenidas para detectar de forma precisa las SVs. Además presenta las siguientes ventajas con respecto a otros programas de detección de SVs (Jiang et al., 2020):
 1. Mejor rendimiento de detección de SVs que otros *variant callers* de última generación.
 2. Elevada sensibilidad para conjuntos de datos con baja cobertura sin disminuir su precisión.
 3. Mayor rapidez en la detección con un menor uso de RAM.
 4. Velocidad casi lineal con el número de *threads* de la CPU.
 5. Gran diversidad de parámetros para los análisis.

- **Sniffles2:** Es un programa basado en el enfoque de alineación de lecturas, diseñado para la detección de SVs con lecturas largas, tiene mejor precisión y velocidad que su predecesor Sniffles y presenta las siguientes ventajas (Duan et al., 2022):
 1. Tiene bajas tasas de error.
 2. Permite la detección de SVs de bajas frecuencias con alta precisión.
 3. Permite detectar SVs en organismos haploides (al igual que en cromosomas X de machos de mamíferos).
 4. Permite hacer análisis a nivel de población.
 5. Permite realizar estudios de mosaicismo.

1.6. Importancia de las variantes estructurales en mejora genética

Según el dogma central de la biología, el ADN se replica y se transcribe formando ARN que posteriormente puede ser traducido de un código de nucleótidos a un código de aminoácidos, los cuales se agrupan y forman diferentes polipéptidos. Éstos, a su vez, forman proteínas con funciones estructurales o enzimáticas que determinan finalmente el fenotipo de un individuo (Morange, 2009). Las SVs pueden afectar a la información genética de los individuos al alterar los codones de su genoma, produciéndose entonces proteínas diferentes. Tras estudios que observaron que las SVs eran comunes en los humanos, se comenzaron a emplear microarrays para localizarlas en otras especies animales, como los ratones, ratas, vacas, cerdos, ovejas y caballos (Escaramís et al., 2015).

1.6.1. Ejemplos de SVs relacionadas con el ganado

Algunos ejemplos de SVs que afectan a caracteres de interés en especies ganaderas son los siguientes:

- En ovino, las SVs son responsables de caracteres como el color de la lana, habiéndose relacionado una duplicación en tándem de 190kb con este carácter, tornándola blanca y facilitando así su tratamiento con tintes de otros colores (Norris & Whan, 2008). También se ha descrito que algunas CNVs situadas en los genes *fhit* y *cacna2d3* del cromosoma 19 están implicadas en la cantidad de leche que puede ser ordeñada al día y en la cantidad de grasa de la leche (Di Gerlando et al., 2019). Se demostró que una inserción de 168-pb en el gen *hoxb13* estaba implicada en la longitud de la cola de la raza Merino Landschaf (Li et al., 2023). Además, se han detectado algunas CNVs que están implicadas en la adaptación al clima (Salehian-Dehkordi et al., 2021).
- En porcino, duplicaciones de alelos de la familia del gen *kit* son responsables de la predominancia del color blanco de su pelaje (Rubin et al., 2012). También se han descrito relaciones entre 41 CNVs situadas en varios QTL y su resistencia a las enfermedades, un comportamiento más dócil, el cual facilita su manejo, una mejor disipación del calor y hasta en la diferenciación de los folículos pilosos, mejorando todos estos caracteres las condiciones de vida de estos animales en entornos de producción (Wang et al., 2017). Se ha observado que 65 SVs son responsables de la alteración de 59 genes que regulan la diferenciación celular y

- la formación de otras moléculas que causan el fenotipo mutante de arrugas cutáneas en la raza Xiang (Xiaoli et al., 2021). También se han logrado identificar 7568 SVs en cerdos tibetanos que están asociadas a 2508 genes enriquecidos en 8 vías KEGG que juegan un rol importante en la adaptación a hipoxia en altitudes elevadas (Jiang et al., 2023).
- En algunas razas de vacuno, se ha observado que una translocación de 492 kb en el gen *kit* interviene en el color del pelo (Durkin et al., 2012). También se ha observado la importancia que tienen CNVs en genes del MHC en la resistencia de la raza Angus a algunos parásitos (Boussaha et al., 2015), una delección de 660 kb en 4 genes del cromosoma 12 bovino con la fertilidad y producciones de leche en las raza Roja Danesa y Roja Sueca (Kadri et al., 2014) y como 490 CNVs estaban relacionadas con genes enriquecidos involucrados en la reproducción y mastitis en Holstein (Liu et al., 2019).
 - En pollos existe un fenotipo en el que las crestas crecen muy poco, en forma de guisante. Esta adaptación a climas fríos reduce la pérdida de calor y los daños por bajas temperaturas y está causada por una CNV en el primer intrón del gen *sox5* (Wright et al., 2009). En broilers, se observó que CNVs en 14 genes estaban relacionadas con la cantidad de grasa abdominal (Zhang et al., 2014).

1.6.2. Imputación de SV en poblaciones comerciales

Aunque gracias a la secuenciación pueden detectarse las SVs, no es posible secuenciar todo el genoma de todos los individuos. En los programas de mejora, se utiliza una técnica que permite abaratar costes y extrapolar genotipos de unos individuos a otros. La imputación de genotipos es una técnica que permite estimar con precisión marcadores genéticos que no se han genotipado de forma directa y puede ser muy útil cuando se pretende combinar resultados de estudios que se han llevado a cabo en plataformas de genotipado muy diferentes, al igual que para potenciar análisis individuales (Li et al., 2009). Consiste en emplear una serie de características conocidas del individuo para predecir un valor que se utilizará como dato imputado (Donders et al., 2006). En ganadería, suele emplearse un número pequeño de individuos genotipados con marcadores de altas densidades y un número más grande de individuos con marcadores de densidades menores. Se emplea la imputación para inferir en

aquellos individuos genotipados con bajas densidades los marcadores de los individuos genotipados con altas densidades. Los métodos pueden dividirse en dos grupos:

- Métodos heurísticos: Para identificar información de ligamiento en haplotipos de gran longitud (>10 cM) en individuos que están emparentados. Utilizan información del pedigrí e infieren los datos de grandes partes del genoma que se comparten entre pares de individuos. Pero en estos casos no se imputan aquellos alelos de los que no se tienen datos fiables (Hickey et al., 2012).
- Métodos probabilísticos: Para identificar información de desequilibrio de ligamiento en haplotipos de individuos que no están emparentados entre sí. Emplean enfoques de Modelos de Cadenas Ocultas de Markov (HMM) para modelizar como varia el genotipo a lo largo del cromosoma y como se comparten segmentos del genoma entre estos individuos no emparentados. Estos métodos son muy exigentes en términos computacionales, además son más lentos y menos precisos que los heurísticos cuando no se tiene en cuenta el pedigrí (Antolín et al., 2017).

1.6.3. Aplicaciones de las SVs en programas de mejora de ganado vacuno

En el ganado vacuno, se han realizado proyectos como el *1000 Bull Genomes Project*, que desde 2012 ha permitido encontrar variaciones genéticas en las poblaciones (84 millones de SNPs y 2.5 millones de pequeñas inserciones y deleciones) y poder así asociar estas variantes con algunos caracteres de interés, como la calidad y cantidad de leche, fertilidad o la calidad de la carne (Hayes & Daetwyler, 2019), entre otros. Como ya se ha visto, se han identificado SVs que afectan a caracteres cuantitativos, por lo que el desarrollo de catálogos de SVs en poblaciones de referencia y la imputación de esos datos en otras poblaciones podría ser de utilidad para la mejora genética del ganado al ayudar en la interpretación de la diversidad genética de las razas, la asociación de estas SVs con la susceptibilidad a enfermedades o la adaptación ambiental, entre otras cosas (Nguyen et al., 2023). Aquí entra en juego la continuación del *1000 Bull Genomes Project*: El *Bovine Long Reads Consortium* (BovLRC), un proyecto que busca secuenciar 1000 genomas bovinos de diferentes razas con lecturas largas para detectar SVs y utilizarlas junto con SNPs para elaborar paneles de referencia para la imputación de genotipos. Esto permitiría mejorar la selección genómica o los estudios de asociación de genoma completo de animales optimizando los costes.

2. Objetivo

Este trabajo de fin de máster sigue los objetivos del proyecto *Bovine Long Reads Consortium*, detectar variantes estructurales a través del uso de lecturas largas obtenidas con ONT en el genoma de 4 razas (Holstein, Red Holstein, Wagyu y Pirenaica Española) de ganado bovino comparando diferentes softwares de alineamiento y detección de variantes estructurales.

2.1. Objetivos específicos

- Comparar los softwares de alineamiento Minimap2 y LRA.
- Comparar los softwares de detección de variantes estructurales Sniffles2 y CuteSV.
- Detectar variantes estructurales en diferentes razas de ganado bovino y estudiar su implicación en caracteres de interés en producción animal.
- Establecer similitudes y diferencias entre razas empleando variantes estructurales.
- Colaborar con el proyecto *Bovine Long Reads Consortium* en su primer run.

3. Materiales y métodos

Este estudio se realizó en el departamento de mejora genética animal del Instituto Nacional de Ingeniería Agraria y Alimentaria (INIA-CSIC), ubicado en Madrid (España).

3.1. Obtención de datos

Para la realización de este trabajo de fin de master se utilizaron datos procedentes del proyecto *BovLRC*. Se trata de muestras de sangre cedidas por el proyecto “H2020 RUMIGEN” y de sangre y semen cedidas por el centro de inseminación artificial “ABEREKIN S.A”. Se emplearon un total de 21 animales: 13 hembras de raza Holstein, 2 machos de raza Holstein, 3 machos de raza Pirenaica Española, 2 machos de raza Wagyu y 1 macho de raza Red Holstein (Tabla 1).

Raza	Nombre	Sexo
Holstein	HOL 1	Hembra
	HOL 2	Hembra
	HOL 3	Hembra
	HOL 4	Hembra
	HOL 5	Hembra
	HOL 6	Hembra
	HOL 7	Hembra
	HOL 8	Hembra
	HOL 9	Hembra
	HOL 10	Hembra
	HOL 11	Hembra
	HOL 12	Hembra
	HOL 13	Macho
	HOL 14	Macho
	HOL 15	Hembra
Wagyu	WAG 1	Macho
	WAG 2	Macho
Red Holstein	RED	Macho
Pirenaica Española	PIR 1	Macho
	PIR 2	Macho
	PIR 3	Macho

Tabla 1. Razas y animales empleados en el estudio.

Se utilizó el Kit de extracción de ADN *Monarch® High Molecular Weight* de *New England Biolabs* para las muestras de sangre y un protocolo de fenol-cloroformo para las muestras de semen. Posteriormente, se empleó el dispositivo “PromethION 2 Solo” usando los kits comerciales *Ultra-long DNA Sequencing Kit V14 (SQK-ULK114)* y *Ligation Sequencing Kit V14 (SQK-LSK114.24)* de ONT para la preparación de la librería y *barcoding* de las muestras.

3.2. Análisis bioinformático

3.2.1. Basecalling

El *basecalling* se realizó con el supercomputador Finisterra III del Centro de Supercomputación del CESUGA (Galicia). Se realizó el basecalling de las muestras usando el software Guppy, únicamente disponible para clientes de ONT en la web de su comunidad (<https://community.nanoporetech.com>), en el centro de supercomputación CESGA. Este software convierte las salidas de datos de los dispositivos de ONT (formato .fast5 por defecto) en archivos con formato .fastq sobre los que se realizaron el control de calidad (N50>10Kb, Guppy V6.4 *Super High Accuracy mode*) y alineamientos.

3.2.2. Alineamiento

Se alinearon las muestras con el genoma de referencia ARS_UCD_v2.0 empleado por el proyecto *BovLRC*, que comprende la referencia ARS-UCD1.2 (Rosen et al., 2020) y el ensamblaje del cromosoma Y de raza Wagyu, utilizando los softwares Minimap2 en su versión 2.24 (disponible en <https://github.com/lh3/minimap2>) y LRA en su versión 1.3.7.2 (disponible en <https://github.com/ChaissonLab/LRA>).

En el caso de Minimap2, se alineó la muestra en formato .fastq con el genoma de referencia, mientras que para los alineamientos con LRA se empleó la muestra transformada de formato .fastq a formato .fasta con la herramienta seqkit (Shen et al., 2016), obteniendo en ambos casos archivos con formato .sam. Estos archivos fueron procesados con samtools (Li et al., 2009) para obtener archivos con formato .bam.

3.2.3. Variant calling

Los archivos con formato .bam resultantes de los alineamientos se analizaron con los softwares Sniffles2 (Smolka et al., 2024), versión 2.2 y CuteSV, versión 2.1.0 para la detección de SVs. Estos softwares crean archivos con formato .vcf que contienen información sobre las SVs detectadas. Adicionalmente, Sniffles2 crea archivos con

formato .snf con información sobre las SVs. Posteriormente, se utilizó la función de *Multi-Sample SV calling* de Sniffles2 para combinar en un único archivo .vcf todos los archivos .snf de cada individuo, eliminando las SVs duplicadas en ellas para ser utilizados posteriormente en la comparación de razas.

Finalmente, se filtraron los archivos .vcf para conservar aquellas SVs con puntos de rotura precisos (flag: PRECISE), longitud superior a 50pb, con una profundidad de lectura mayor a 7x (flag: SUPPORT) y con Q score superior a 20 (Precision superior al 99%) en cada individuo y raza. Tras haber filtrado los archivos .vcf, se transformaron a formato .tsv para facilitar los siguientes análisis.

3.2.4. Analisis de razas

Se analizaron las muestras en R (R Core Team, 2023) y se utilizó el paquete de R ggplot2 (Wickham, 2016) para elaborar gráficos según el tipo de SV en cada uno de los animales, asignando diferentes colores a cada raza. Posteriormente, se utilizaron los datos agrupados de cada raza y se empleó el paquete de R VennDiagram (Chen, 2011) para crear un gráfico con las SVs exclusivas y compartidas entre las razas, estableciendo que las SVs son comunes si su posición de inicio y posición final se sitúan en un rango de ± 25 pb. Después se utilizó el paquete ggplot2 para observar la abundancia relativa de cada tipo de SVs en las 4 razas y el número de SVs por cromosoma y raza según su tipo en gráficos de barras apiladas. Después se agruparon todos los datos en una única población sobre la que se utilizó el paquete de R ggplot2 para elaborar un gráfico de densidad de SVs por cromosoma según la posición de inicio de la variante para comprobar en qué regiones cromosómicas son más o menos abundantes.

3.2.5. Análisis de genes

Finalmente, se analizaron las muestras con el paquete de R ChIPseeker (Wang et al., 2022; Yu et al., 2015), obteniendo las regiones del genoma donde se sitúan las SVs y se analizaron los exones exclusivos de cada raza, los exones compartidos entre las razas Holstein y Red Holstein y las razas Pirenaica Española y Wagyu y los exones compartidos entre todas las razas que estaban afectados por SVs para explorar posibles genes de interés en mejora genética.

3.3. Estudio comparativo

Para este estudio se emplearon 2 de las 21 muestras: HOL 13 y WAG 2 correspondientes a dos toros de las razas Holstein y Wagyu respectivamente. El objetivo de este estudio fue analizar qué software realizaba un mejor alineamiento y qué software era más preciso en la detección de SVs.

3.3.1. Comparación de alineadores

Ambas muestras se alinearon con la referencia ARS_UCD_v2.0 utilizando los softwares Minimap2 y LRA para observar cuál de los alineadores era más preciso. Para ello, se empleó de nuevo el software samtools para obtener la profundidad y cobertura medias de las muestras. Además, se calculó la precisión de los alineamientos teniendo en cuenta los mismatches por longitud de la secuencia. Finalmente, se realizaron gráficos comparativos de ambos alineadores utilizando el paquete ggplot2 de R.

3.3.2. Comparación de variant callers

Se utilizaron Sniffles2 y CuteSV en las muestras alineadas tanto con Minimap2 como con LRA para detectar SVs. Los archivos .vcf resultantes se filtraron en función de su profundidad de lectura, tamaño superior a 50pb, puntos de rotura y Q score y se realizaron gráficos de barras apiladas con el paquete ggplot2 de R y tablas para observar las SVs detectadas con cada una de las combinaciones de alineador y *variant caller* y las SVs compartidas entre métodos. En primer lugar, se consideró que en las SVs comunes la posición de inicio y la posición final de las mismas debía situarse en un rango de ± 100 pb, después se disminuyó ese rango a ± 50 , ± 25 , ± 10 y ± 5 pb para comprobar si variaba el porcentaje de SVs comunes entre metodologías. Por último, se realizó un gráfico de densidad utilizando el paquete de R ggplot2 para observar la longitud de las SVs con cada una de las combinaciones de alineador y *variant caller*.

4. Resultados y Discusión

4.1. Estudio comparativo

Al realizar los alineamientos de las muestras de “HOL13” y “WAG2” con los softwares Minimap2 y LRA, se obtuvieron los resultados que figuran en la Tabla 2. LRA muestra una precisión ligeramente superior (Figura 4) y una profundidad de lectura y cobertura algo menores que Minimap2, lo que concuerda con los estudios de Ren & Chaisson (2021), donde LRA mostraba una precisión (94,3%) superior a Minimap2 (90,4%) en los alineamientos de ONT. Sin embargo, Liu et al. (2022) observaron que de entre todos los alineadores utilizados en su estudio, Minimap2 presentaba la mayor precisión con un 99,3%, seguido por LRA en segundo lugar, con una precisión que rondaba el 99,12% para datos de ONT en pera. Aunque, en general, ambos alineadores muestran buenos resultados de forma individual.

Alineador	Muestra	Profundidad media de lectura	Cobertura media de lectura (%)	Precisión media de lectura (%)
Minimap2	HOL13	13,41	98,20	99,51
	WAG2	9,46	97,70	99,53
LRA	HOL13	12,74	98,43	99,60
	WAG2	9,02	97,83	99,62

Tabla 2. Profundidad, cobertura y precisión media en las muestras de HOL13 y WAG2 alineadas con Minimap2 y LRA.

Posteriormente, tras realizar el *variant calling*, en el caso de la muestra de HOL13 al realizar el alineamiento con Minimap2 se obtuvieron 5.924 SVs utilizando Sniffles2 y 6.662 SVs utilizando CuteSV. Por otra parte, utilizando LRA como alineador, se obtuvieron 4.728 SVs con Sniffles2 y 5.432 SVs con CuteSV. En el caso de la muestra de WAG2, realizando el alineamiento con Minimap2 se detectaron 2.702 SVs con Sniffles2 y 3.127 SVs con CuteSV. Para LRA como alineador, Sniffles2 logró detectar 2.013 SVs mientras que CuteSV detectó 2.348 SVs, como figura en la Tabla 3.

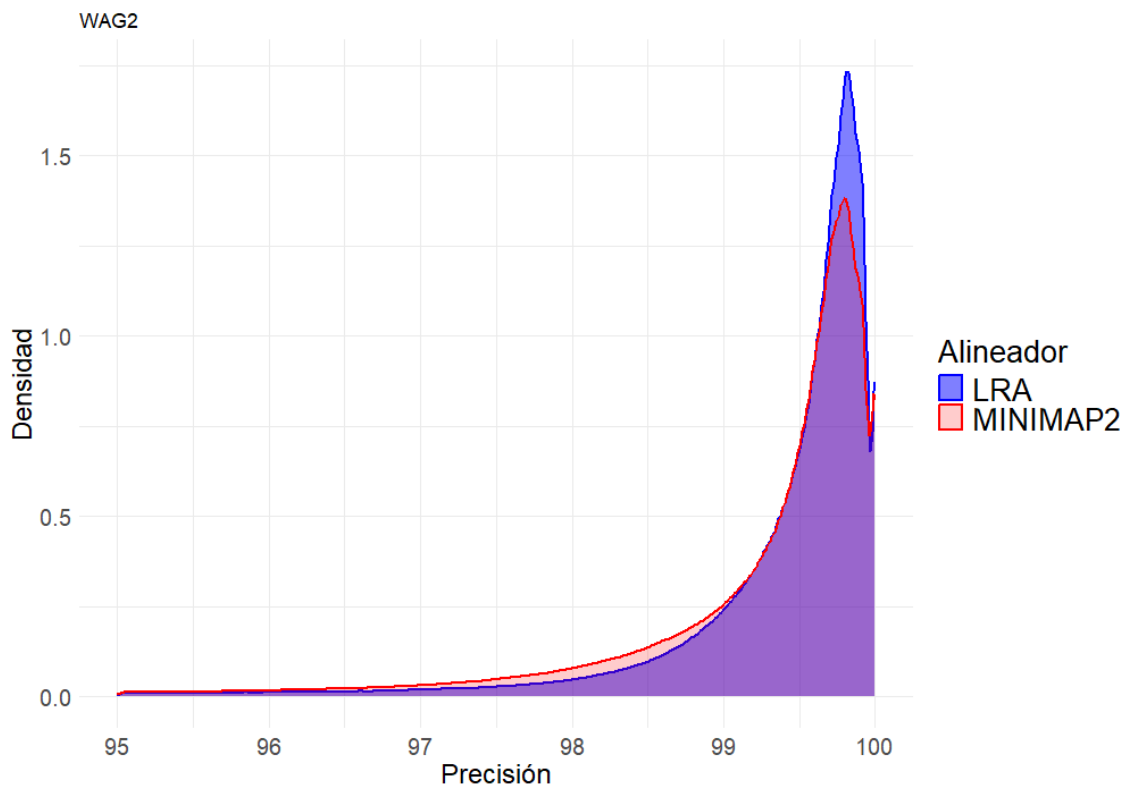
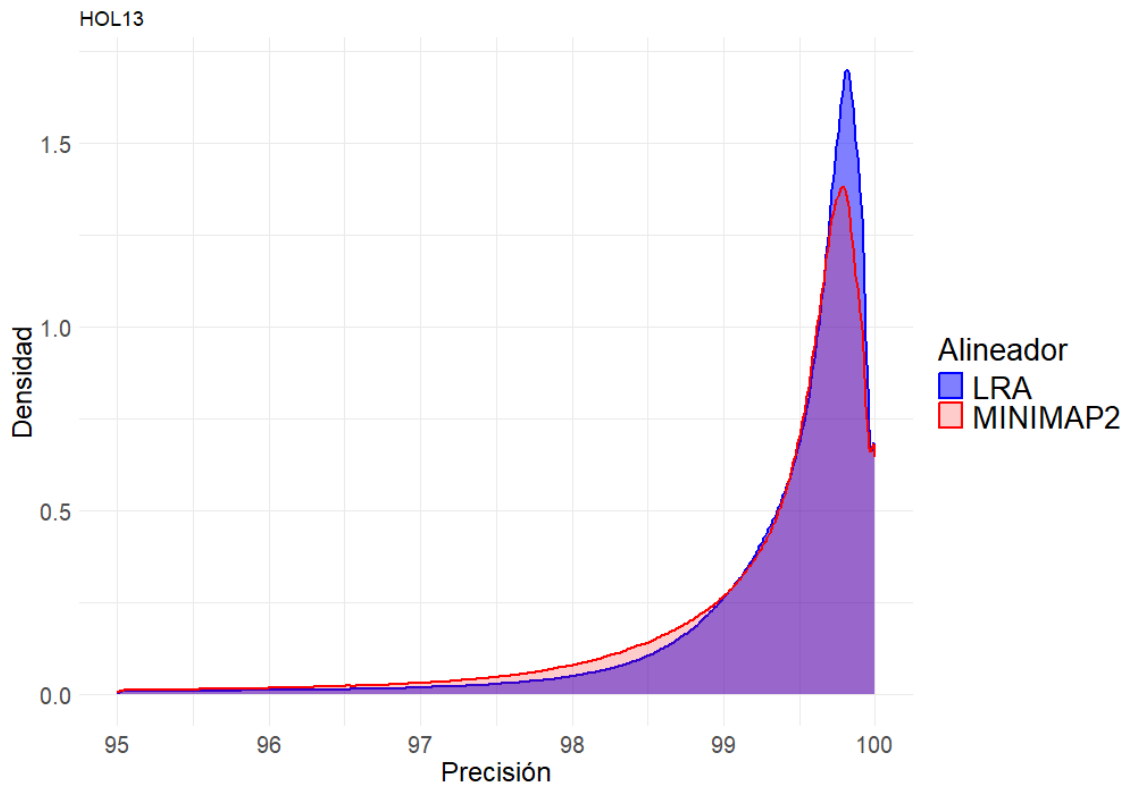


Figura 4. Gráficos de precisión de alineamiento superior al 95% en las muestras de HOL13 (arriba) y WAG2 (abajo). LRA en azul, Minimap2 en rojo.

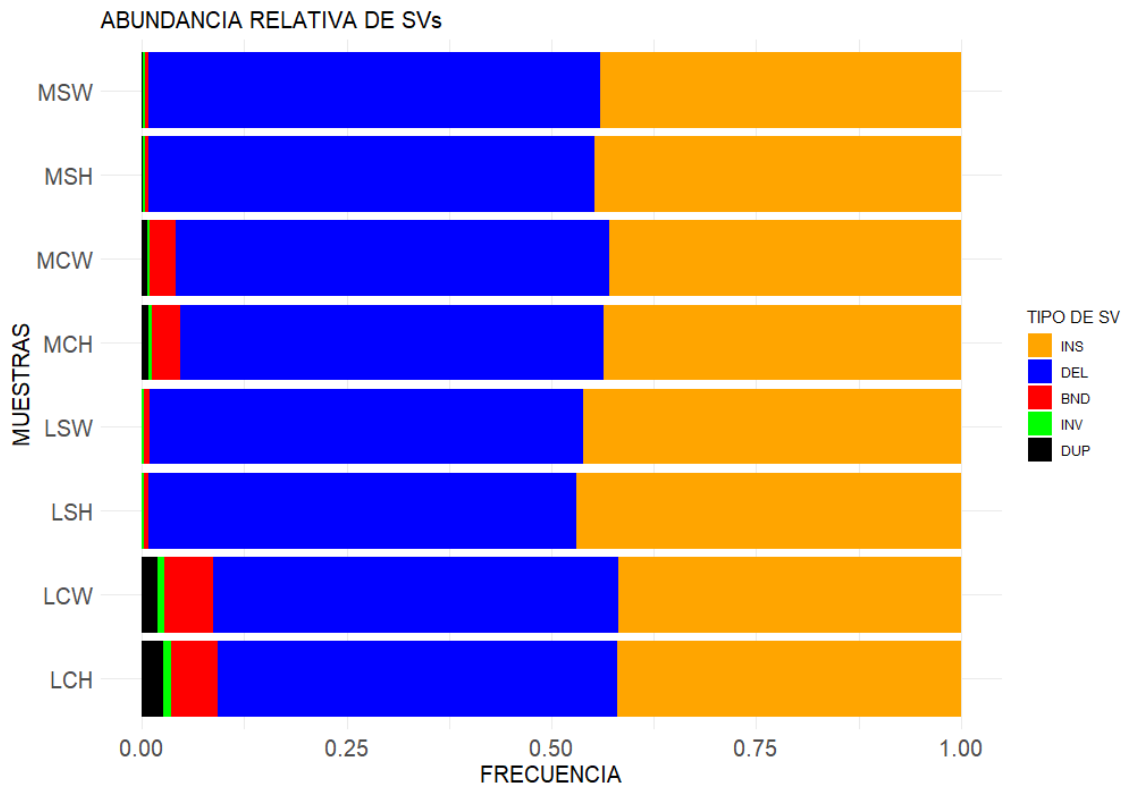


Figura 5. Abundancia relativa de SVs detectadas con Sniffles2 y CuteSV, alineadas con Minimap2 y LRA en las muestras de HOL13 y WAG2. Las siglas representan un código de 3 letras para identificar las 8 combinaciones. S=Sniffles2, C=CuteSV, M=Minimap2, L=LRA, H=HOL13 y W=WAG2.

Al analizar la abundancia relativa de cada tipo de SVs en las 8 combinaciones (Figura 5), se observó que Sniffles2 detectó una frecuencia de inserciones y deleciones ligeramente mayor a CuteSV tanto en los alineamientos con Minimap2 como con LRA. También detectó una mayor frecuencia de BNDs que CuteSV en los alineamientos con LRA.

Por el contrario, CuteSV detectó una frecuencia mayor de inversiones y duplicaciones con ambos alineadores y una mayor frecuencia de BNDs utilizando Minimap2. Teniendo en cuenta el número de SVs, estos resultados concuerdan con los que obtuvieron Jiang et al. (2022), ya que observaron que Sniffles2 detectaba mayor frecuencia de inserciones, deleciones y BNDs que CuteSV.

HOL13						
	MINIMAP2			LRA		
SV	SNIFFLES2	COMUNES	CUTESV	SNIFFLES2	COMUNES	CUTESV
INS	2.597	2.528	2.873	2.188	2.076	2.422
DEL	3.282	3.220	3.507	2.514	2.482	2.823
BND	26	25	223	20	9	151
INV	14	8	16	6	3	23
DUP	5	4	43	0	0	13
TOTAL	5.924	5.785	6.662	4.728	4.570	5.432
WAG2						
	MINIMAP2			LRA		
SV	SNIFFLES2	COMUNES	CUTESV	SNIFFLES2	COMUNES	CUTESV
INS	1.150	1.119	1.293	894	845	994
DEL	1.529	1.499	1.702	1.096	1.082	1.236
BND	14	14	113	20	8	97
INV	6	5	9	3	1	11
DUP	3	2	10	0	0	10
TOTAL	2.702	2.639	3.127	2.013	1.936	2.348

Tabla 3. Número de inserciones, deleciones, BNDs, inversiones y duplicaciones y número total de SVs detectadas con Sniffles2 y CuteSV compartidas según su alineador en las muestras de HOL13 y WAG2 alineadas tanto con Minimap2 como con LRA con un rango de ± 100 pb.

SVs Comunes	± 100pb	± 50pb	± 25pb	± 10pb	± 5pb
LSH y LCH	4.570	4.491	4.315	3.903	3.570
LSW y LCW	1.936	1.913	1.832	1.642	1.514
MSH y MCH	5.785	5.721	5.612	5.298	5.022
MSW y MCW	2.639	2.610	2.569	2.442	2.314

Tabla 4. Número de SVs en común entre metodologías según el rango de las posiciones de inicio y fin de las SVs. Las siglas representan un código de 3 letras para identificar las 8 combinaciones. S=Sniffles2, C=CuteSV, M=Minimap2, L=LRA, H=HOL13 y W=WAG2.

Posteriormente se analizaron las SVs en común entre las diferentes metodologías. Teniendo en cuenta una diferencia de ± 100 pb entre las posiciones de inicio y final de las SVs (Tabla 3) se obtuvieron 5.785 SVs en común entre Sniffles2 y CuteSV en los alineamientos con Minimap2 y 4.570 SVs en común para los alineamientos con LRA para HOL13. En WAG2, las SVs comunes detectadas para los alineamientos con Minimap2 fueron 2.639, mientras que para LRA fueron 1.936.

De un total de 12.586 SVs detectadas en HOL13 y 5.829 SVs detectadas en WAG2 para los alineamientos con Minimap2, en el primer caso ambos *variant callers* compartían un 44,99% de las SVs, mientras que en el segundo caso compartían un 45,27% de ellas. Para los alineamientos con LRA, en HOL13 se obtuvo un 44,98% de SVs compartidas y en WAG2 un 44,39%.

A continuación, al reducir el rango de pb de ± 100 hasta ± 5 (Tabla 4), se observó que para los alineamientos con LRA en HOL13 se redujo un 21,89% el número de SVs comunes y en WAG2 se redujo un 21,8%. En cambio, para los alineamientos con Minimap2 se produjo una menor disminución de las SVs comunes, siendo de un 13,32% en HOL13 y de un 12,32% en WAG2.

Por último, se tuvieron en cuenta las longitudes de las SVs para observar en qué intervalos eran más abundantes (Figura 6). Para todas las combinaciones se observaron comportamientos similares, con un mayor número de SVs en longitudes alrededor de 100pb. En adelante, el número de SVs fue disminuyendo.

En general, para las dos muestras, la combinación de Minimap2 como alineador y CuteSV como *variant caller* fue la que mayor número de SVs detectó, seguida por la combinación de Minimap2/Sniffles2. En tercer lugar, LRA/CuteSV y por último la combinación LRA/Sniffles2. Por lo tanto, teniendo en cuenta número de SVs detectadas, Minimap2 se presentaría como el mejor alineador, mientras que CuteSV sería el mejor *variant caller*. Los resultados anteriores concuerdan con Helal et al. (2024), ya que comparando múltiple alineadores y *variant callers*, observaron que aunque Sniffles2 y CuteSV eran muy similares, CuteSV tenía una ligera superioridad con respecto a Sniffles2 en la detección de SVs. En cambio, no concuerdan con Smolka et al. (2024), cuyos experimentos mostraron que Sniffles2 presentaba una mayor precisión en la detección de SVs que otras metodologías como CuteSV, aunque con rendimientos similares. Además, las simulaciones realizadas por Yildiz et al. (2023) recomiendan el uso de la combinación de Minimap2 con Sniffles2 y sobre todo con CuteSV por la buena precisión en la detección de SVs, ya que CuteSV consigue reducir el número de falsos positivos, mostrando en múltiples análisis una mejor precisión que Sniffles2 (Jiang et al., 2022), con valores en torno al 92.7%, frente al 88.9% de Sniffles2 (Romain & Lemaitre, 2023).

Teniendo todo lo anterior en cuenta, podemos concluir que aunque ambos *variant callers* tienen rendimientos similares, con precisiones de detección por encima del 99%, CuteSV es capaz de detectar mejor BNDs, inversiones y duplicaciones, mientras que Sniffles2 detecta mejor inserciones y deleciones. Aunque CuteSV es ligeramente superior a Sniffles2 en el número de SVs detectadas, puede que no todas ellas sean correctas, por lo que sería necesario validar los resultados con un set de SVs conocidas.

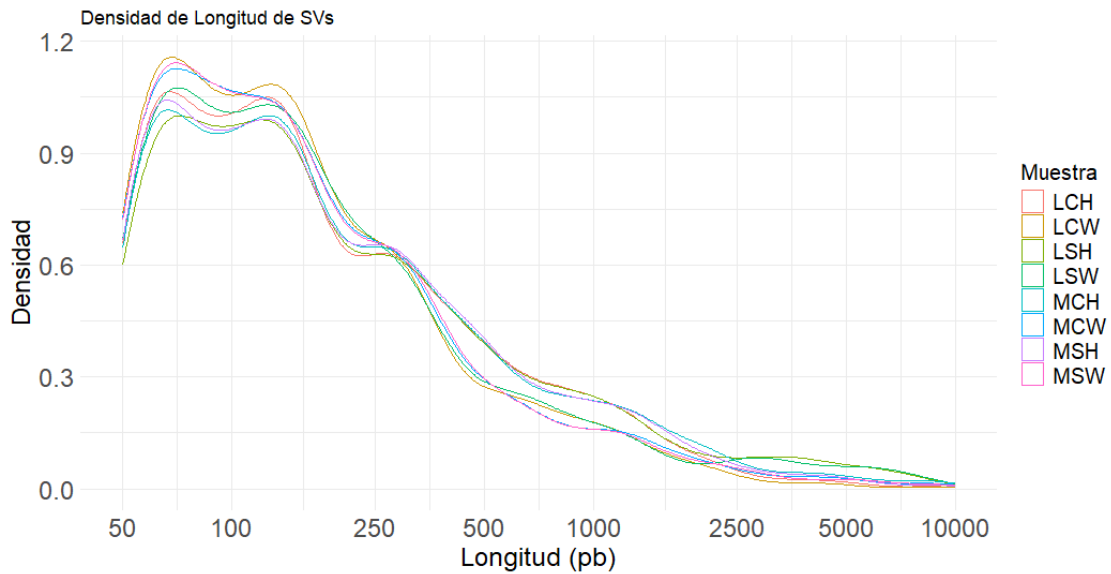


Figura 6. Gráfico de densidad de SVs detectadas con Sniffles2 y CuteSV, alineadas con Minimap2 y LRA en las muestras de HOL13 y WAG2. Las siglas representan un código de 3 letras para identificar las 8 combinaciones. S=Sniffles2, C=CuteSV, M=Minimap2, L=LRA, H=HOL13 y W=WAG2.

4.2. Análisis de razas

Se estableció Minimap2 como alineador y Sniffles2 como *variant caller* de acuerdo con las especificaciones del *BovLRC*. Tras el análisis y filtrado de datos para cada individuo, se obtuvieron los resultados mostrados en la Tabla 5. En total, sumando las SVs de cada individuo se obtuvieron 168.737 SVs en el total de la población. Se identificaron 115.321 SVs en raza Holstein, 30.464 SVs en raza Pirenaica Española, 17.075 SVs en raza Wagyu y 5.877 SVs en raza Red Holstein. Se puede observar que la gran mayoría de SVs detectadas en cada individuo fueron inserciones y deleciones y siendo BNDs, duplicaciones e inversiones los tipos de SVs menos comunes (Figura 7).

ID	Profundidad	Cobertura	INS	DEL	INV	DUP	BNDs
HOL 1	9,08	96,76	1.047	1.424	8	4	23
HOL 2	13,40	96,79	8.087	9.053	69	21	116
HOL 3	10,19	96,64	2.096	2.782	14	12	31
HOL 4	6,76	96,17	4.722	6.010	36	10	42
HOL 5	12,35	96,80	3.032	4.212	19	11	35
HOL 6	7,02	95,91	1.161	1.820	14	8	27
HOL 7	6,38	96,20	4.710	5.568	33	12	48
HOL 8	11,18	96,76	1.608	2.025	10	6	27
HOL 9	10,78	96,75	1.042	1.597	13	7	24
HOL 10	11,18	96,76	3.968	4.660	27	8	29
HOL 11	10,78	96,75	3.899	4.565	24	10	32
HOL 12	6,82	95,92	1.614	1.879	9	5	25
HOL 13	13,41	98,20	2.597	3.282	14	5	26
HOL 14	17,42	98,36	6.876	7.839	54	21	44
HOL 15	13,06	96,91	5.198	5.848	35	13	39
WAG 1	10,51	97,51	6.666	7.593	54	15	45
WAG 2	9,46	97,70	1.150	1.529	6	3	14
RED 1	11,89	97,91	2.704	3.116	20	11	26
PIR 1	12,67	97,33	3.512	4.428	23	7	29
PIR 2	15,89	98,12	7.235	8.252	65	15	54
PIR 3	11,85	96,85	3.034	3.742	20	13	35

Tabla 5. Profundidad, Cobertura, Número y Tipo de SVs identificadas en cada individuo. HOL: Holstein; WAG: Wagyu; RED: Red Holstein; Pir: Pirenaica Española; INS: Inserciones; DEL: Deleciones; INV: Inversiones; DUP: Duplicaciones; BNDs: Breakends.

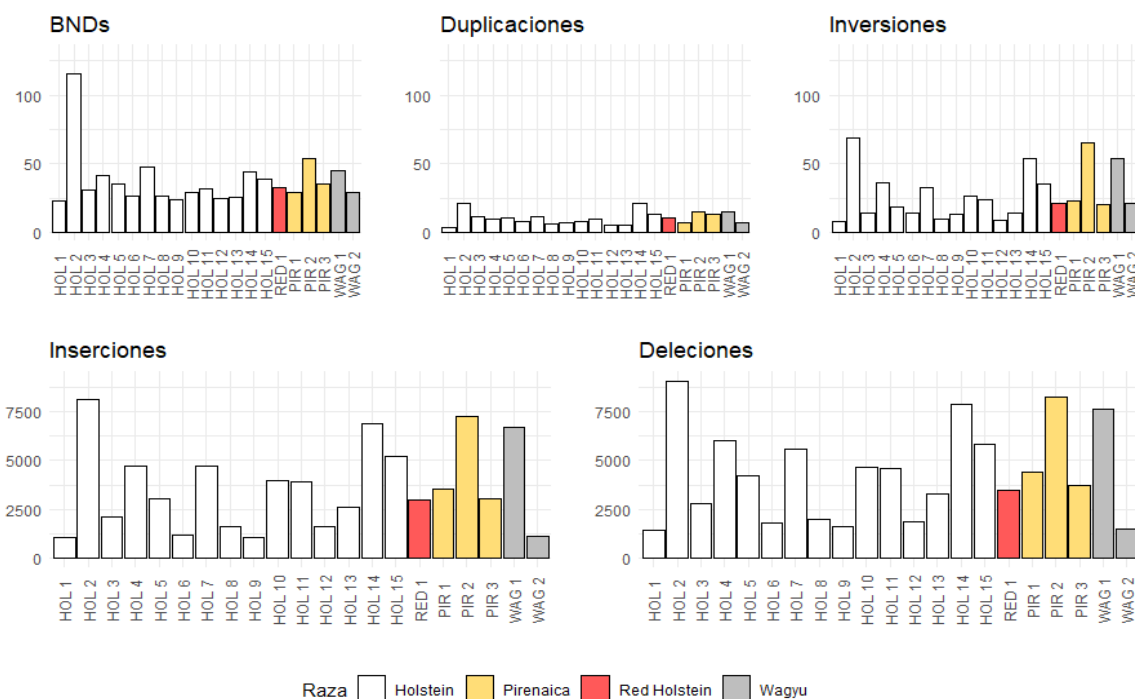


Figura 7. Número y tipo de SVs detectadas en cada animal. Cada color representa una de las cuatro razas.

En total, sumando las SVs de cada individuo se obtuvieron 168.737 SVs en el total de la población. Se identificaron 115.321 SVs en raza Holstein, 30.464 SVs en raza Pirenaica Española, 17.075 SVs en raza Wagyu y 5.877 SVs en raza Red Holstein. Se puede observar que la gran mayoría de SVs detectadas en cada individuo fueron inserciones y deleciones y siendo BNDs, duplicaciones e inversiones los tipos de SVs menos comunes (Figura 7).

Raza	INS	DEL	INV	DUP	BNDs	Total
Holstein	4.704	5.575	45	16	38	10.378
Wagyu	5.682	6.420	44	14	41	12.201
Red Holstein	2.704	3.116	20	11	26	5.877
Pirenaica Española	6.197	7.114	44	17	47	13.419

Tabla 6. SVs únicas encontradas de cada tipo y total por raza.

Posteriormente se agruparon los datos en función de cada raza y se filtraron para eliminar aquellas SVs que se encontrasen duplicadas dentro de cada una de las razas (Tabla 6). Después de este filtrado se obtuvo un total de 41.875 SVs únicas en el total de la población. De esas SVs, en raza Holstein se encontraron 10.378 SVs únicas, en raza Pirenaica Española fueron 13.419 SVs únicas, en raza Wagyu se detectaron 12.201 SVs únicas y en la raza Red Holstein fueron 5.877 las SVs únicas detectadas. Tras comparar las variantes únicas de todas las razas, se obtuvo que la raza Holstein poseía 5.131 SVs exclusivas de su raza y compartía 3.463 SVs con raza Pirenaica Española,

2.306 SVs con raza Red Holstein y 2.923 SVs con raza Wagyu. La raza Pirenaica Española poseía 7.844 SVs exclusivas de su raza y compartía un total de 2.051 SVs con Red Holstein y 3.396 SVs con Wagyu. La raza Wagyu poseía 7.328 SVs exclusivas de su raza y compartía 1.680 SVs con raza Red Holstein. Por último, la raza Red Holstein poseía 2.512 SVs exclusivas de su raza. De entre todas estas SVs, las 4 razas compartían 766 SVs entre ellas (Figura 8).

SVs COMUNES ENTRE RAZAS

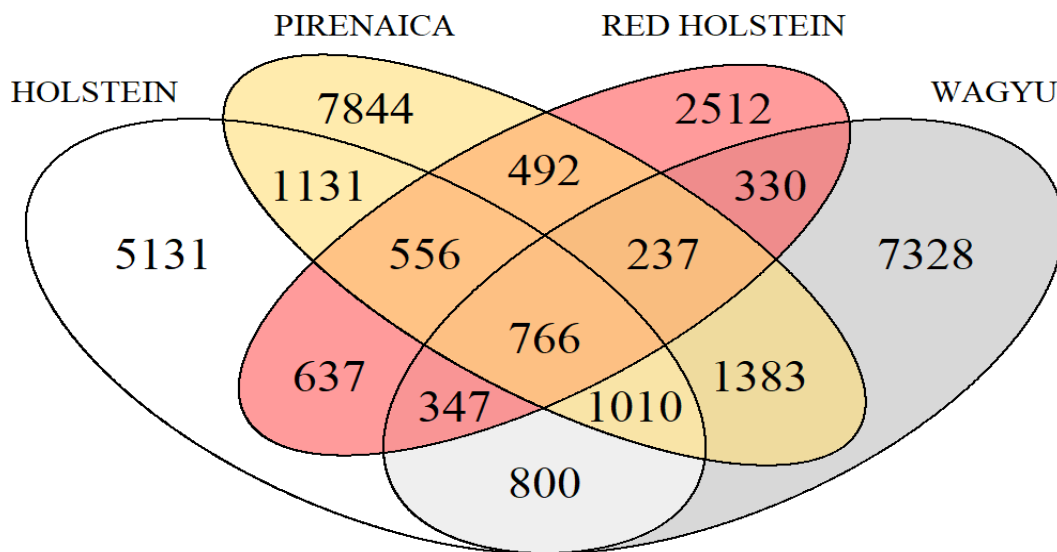


Figura 8. Diagrama de Venn mostrando las SVs únicas comunes entre razas.

La abundancia relativa de cada tipo de SVs por raza fue similar en las 4 razas (Figura 9), siendo claramente más abundantes las delecciones con alrededor del 53% del total, seguidas por las inserciones con un 46%, BNDs con aproximadamente el 0,36% (salvo en el caso de Red Holstein, donde rondaba el 0,44%), inversiones 0,34% (salvo en raza Holstein, con un 0,43%) y duplicaciones con un 0,13% (a excepción de Red Holstein, con un 0.18%). También se observaron los cromosomas en los que se situaban las SVs. El cromosoma que mas SVs poseía fue el cromosoma 1 en las 4 razas (Figura 10) y los cromosomas con menor número de SVs fueron el mitocondrial y el Y. Se pudo observar que el número de SVs tiene una tendencia descendente en los cromosomas. Esto puede deberse a que los cromosomas se numeran según su tamaño, por lo que cabría esperar que, a mayor tamaño de cromosoma, mayor probabilidad de encontrar SVs en él. Después, agrupando las SVs de la población completa se calculó su densidad en cada cromosoma según la posición de inicio de cada una (Figura 11). Se pudo observar que, en la mayoría de ellos, la distribución de SVs es uniforme a lo largo de

todo el cromosoma (cromosomas 1, 2 y 3 por ejemplo), en algunos de ellos una mayor densidad de SVs en la primera mitad del cromosoma (cromosoma 10), en otros una mayor cantidad en la región final del cromosoma (cromosomas 4 y 18) y destacando el cromosoma 23 por presentar una enorme densidad de SVs en su región central.

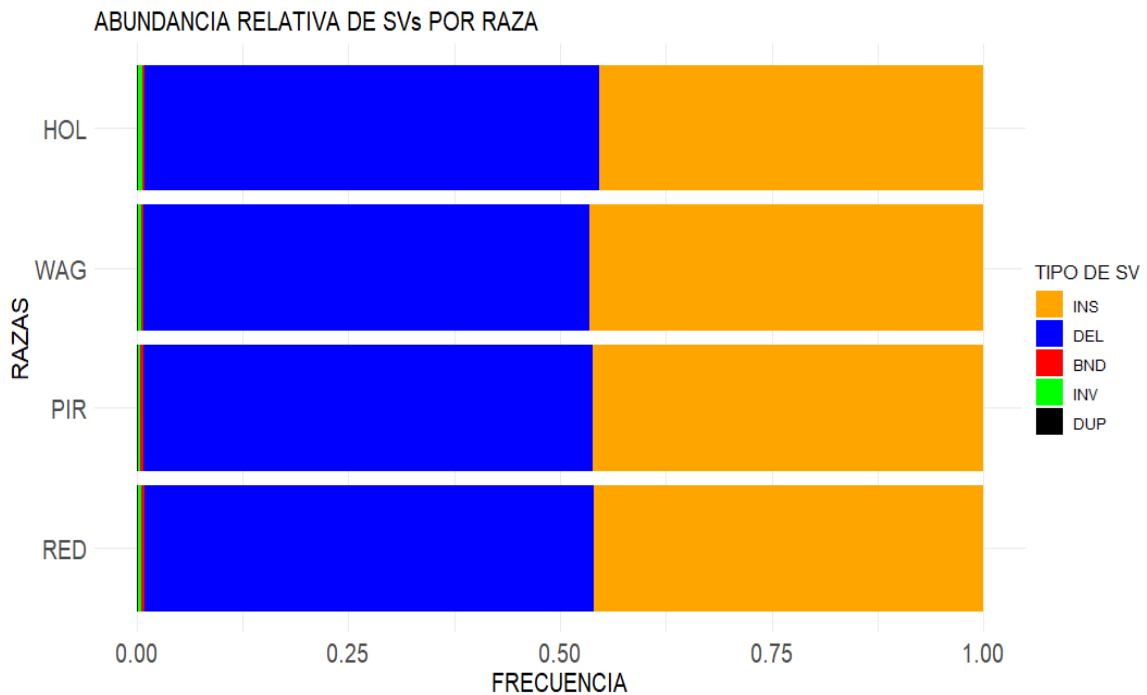


Figura 9. Abundancia relativa de SVs detectadas en razas Holstein (HOL), Wagyu (WAG), Pirenaica Española (PIR) y Red Holstein (RED). Los colores representan el tipo de variante estructural. INS: Inserciones; DEL: Deleciones; INV: Inversiones; DUP: Duplicaciones; BNDs: Breakends.

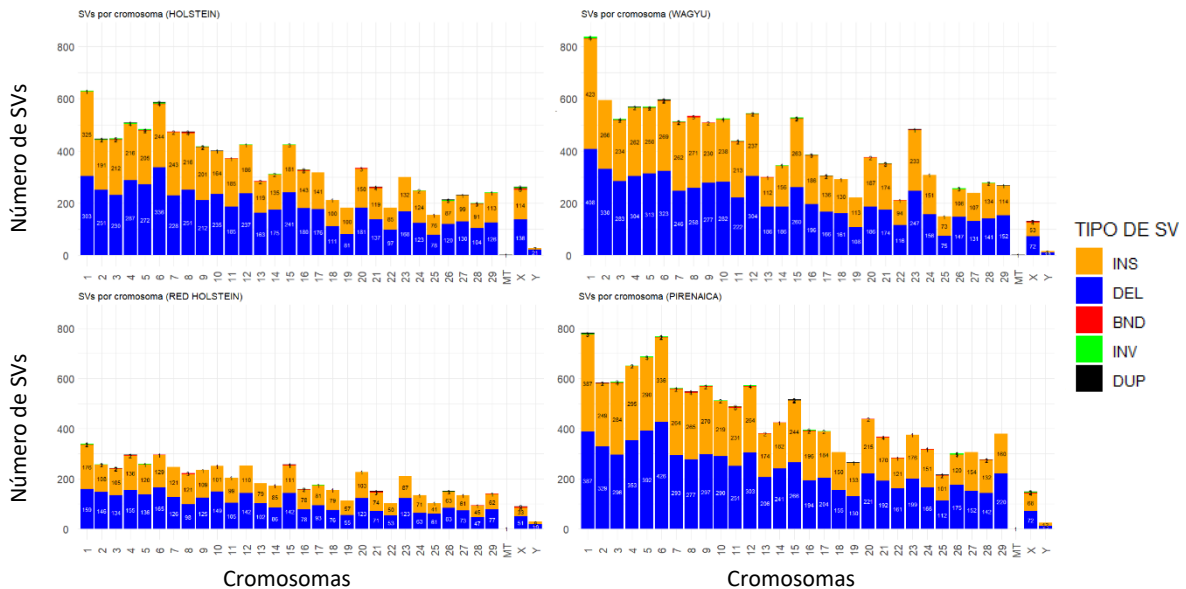


Figura 10. Abundancia relativa de cada tipo de SV por cromosoma detectadas en razas Holstein (HOL), Wagyu (WAG), Pirenaica Española (PIR) y Red Holstein (RED). Los colores representan el tipo de variante estructural. INS: Inserciones; DEL: deleciones; INV: Inversiones; DUP: Duplicaciones; BNDs: Breakends.

Densidad de Variantes estructurales (SVs) por cromosoma

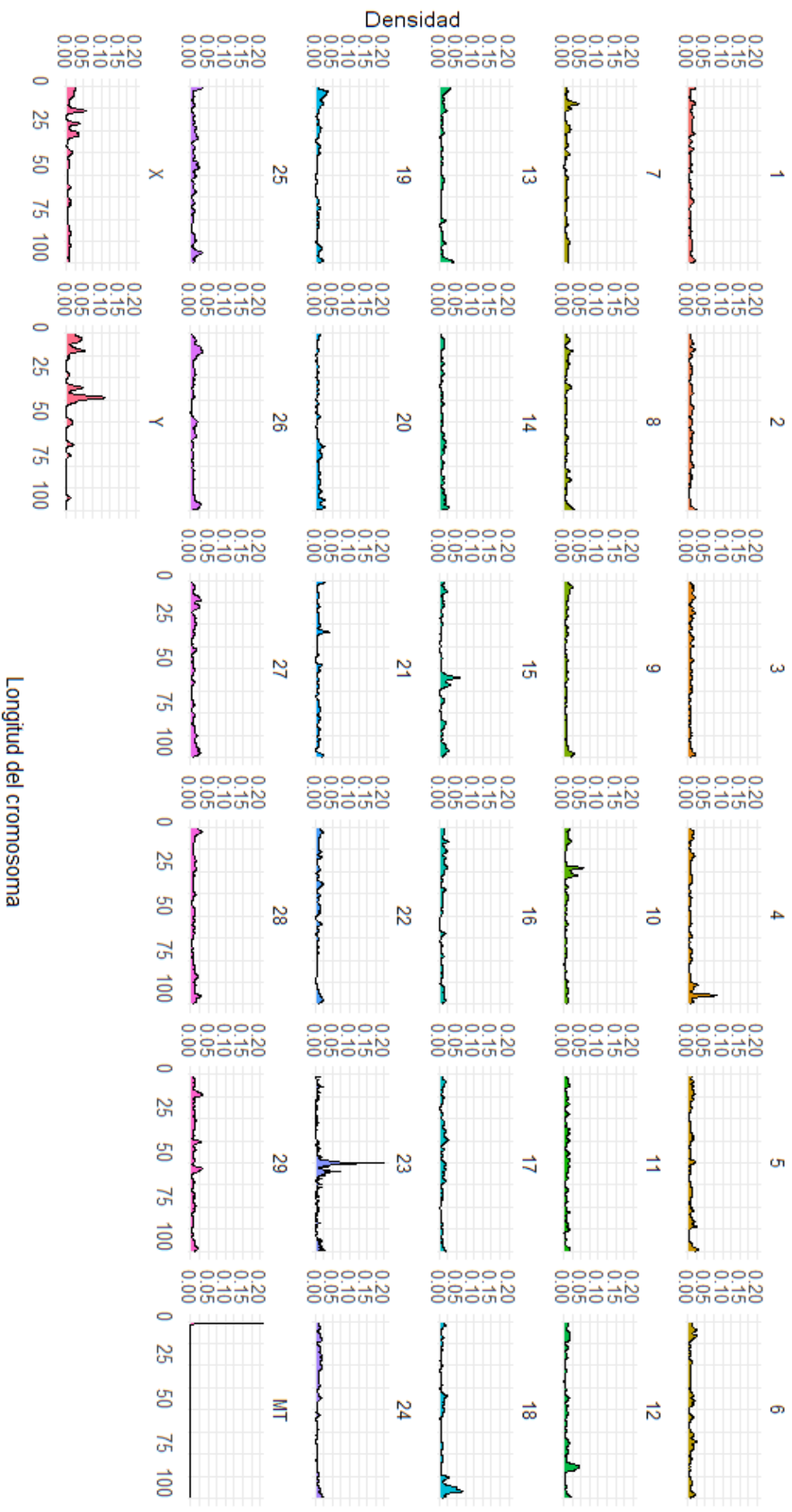


Figura 11. Gráfico de densidad de SVs por cromosoma según su posición de inicio en la población total analizada. Los colores representan cada uno de los cromosomas.

En cuanto a la distribución de las SVs a lo largo del genoma (Figura 12) mostró resultados también similares en las 4 razas. En este caso, la gran mayoría de ellas se encontraba en regiones intergénicas, seguidas por las regiones intrónicas, las regiones promotoras, las 5' y 3' no traducibles y por último las regiones exónicas (Tabla 7).

Región genómica	Holstein	Wagyu	Red Holstein	Pirenaica Española
Promotora (<=1kb)	0,8%	0,82%	0,61%	0,97%
Promotora (1-2 kb)	0,65%	0,61%	0,58%	0,69%
Promotora (2-3kb)	0,61%	0,57%	0,77%	0,74%
5' UTR	0,009%	0,01%	0,04%	0,02%
3' UTR	0,09%	0,09%	0,18%	0,12%
Exónica	0,10%	0,09%	0,14%	0,11%
Intrónica	16,78%	16,22%	16,31%	17,44%
Downstream <=300pb	0,04%	0,03%	0,06%	0,03%
Intergénica	80,92%	81,56%	81,31%	79,88%

Tabla 7. Porcentaje de SVs en cada región genómica en las cuatro razas.

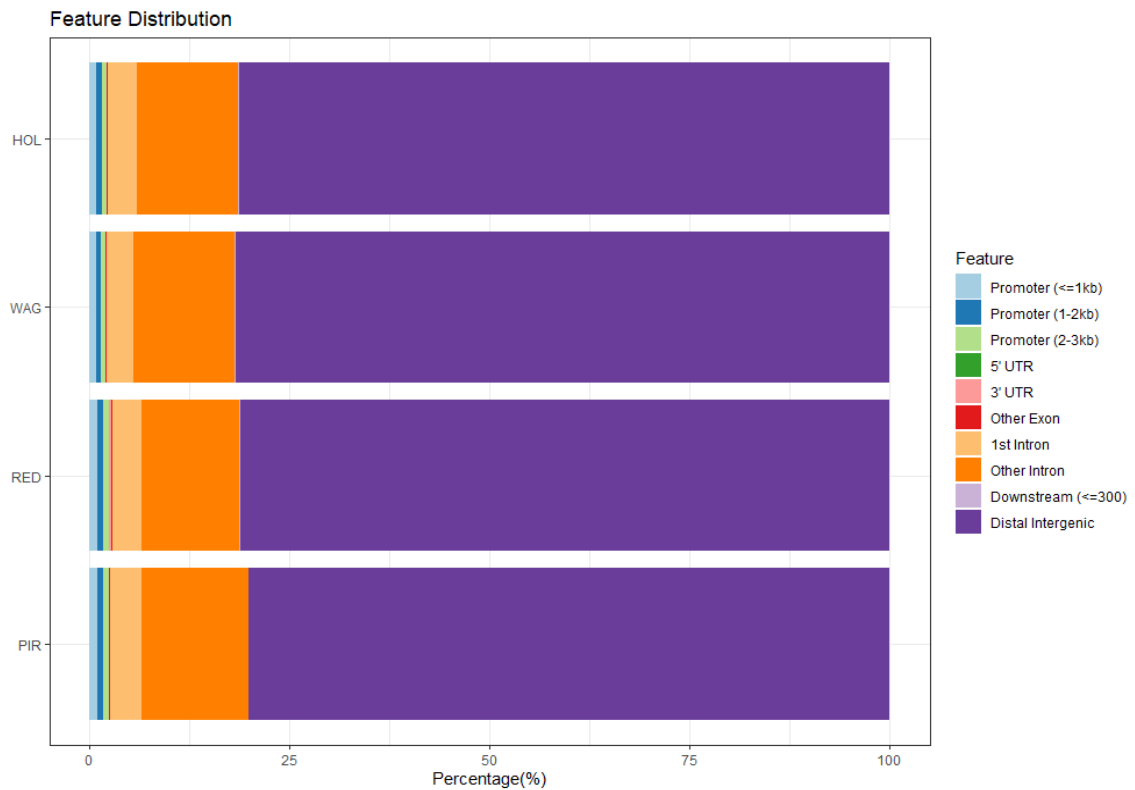


Figura 12. Anotación genómica de SVs por raza. Los colores representan la región afectada por las SVs. De izquierda a derecha: Regiones promotoras, 5' y 3' no traducibles, exónicas, intrónicas y zonas intergénicas.

Tiene sentido que la raza Red Holstein, al tratarse de una raza enfocada a la producción de leche, comparta un 58,92% de sus SVs con la raza Holstein, mientras que con las razas Wagyu y Pirenaica Española (con un enfoque más cárnico), comparta alrededor de un 28% y 35% respectivamente. Aunque teniendo esto en cuenta, cabría esperar que la raza Holstein compartiera un mayor porcentaje de sus SVs con Red Holstein y en menor medida con las razas Wagyu y Pirenaica Española. Sin embargo, la raza Holstein comparte un 33,36% de sus SVs con Pirenaica Española, un 23,16% de sus SVs con Wagyu y tan solo un 22,22% de sus SVs con Red Holstein. Por otra parte, la raza Wagyu comparte un mayor porcentaje de sus SVs con la otra raza cárnica, con un 27,83%, mientras que para el caso de las razas lácteas comparte un 23,95% de sus SVs con Holstein y solamente un 13,79% de ellas con la raza Red Holstein. Finalmente, para el caso de la raza Pirenaica Española, se encontró que compartía un mayor porcentaje de sus SVs con la raza Holstein (25,80%), aunque seguido de cerca por la raza Wagyu (25,30%) y en último lugar la raza Red Holstein (15,28%). Las diferencias en estos resultados podrían deberse a los diferentes tamaños muestrales entre las razas.

La raza Holstein fue la que mostró un mayor porcentaje de deleciones e inversiones y la que mostró un menor porcentaje de SVs en regiones 5' UTR. La raza Wagyu fue la que mayor porcentaje presentaba en cuanto a inserciones y SVs afectando a regiones intergénicas, mientras que fue la raza que mostró menores porcentajes de deleciones, duplicaciones, BNDs y SVs afectando a regiones promotoras (1-2kb), exónicas e intrónicas. La raza Red Holstein presentaba el mayor porcentaje de duplicaciones, BNDs y SVs afectando a regiones Promotoras (2-3kb), 5' UTR, 3' UTR, exónicas y *Downstream* ≤ 300 pb y el menor porcentaje de SVs afectando a regiones Promotoras (≤ 1 kb) y Promotoras (1-2 kb). Por último, la raza Pirenaica Española fue la que mostró un mayor porcentaje de SVs afectando a regiones Promotoras (≤ 1 kb), Promotoras (1-2 kb) e intrónicas y la que menor porcentaje de inversiones y SVs afectando a regiones intergénicas.

4.2.1. Análisis de genes

Finalmente, se analizaron las SVs que se encontraban en regiones codificantes del genoma, es decir, regiones exónicas, puesto que su modificación puede alterar la secuencia de aminoácidos y la función de las proteínas resultantes. Se observó la existencia de SVs que afectaban a exones exclusivos en cada una de las razas estudiadas (Tabla 8). Se obtuvieron 2 exones afectados por SVs en raza Holstein, 1 exon en Red Holstein, 6 exones en raza Pirenaica Española y 3 exones en raza Wagyu.

Estos exones se encuentran en genes que están relacionados con diversos caracteres, pero centrándonos en los caracteres de interés en producción animal, en la raza Holstein se identificó el gen *colgatl*, que puede ser relevante en programas de mejora de vacuno lechero, ya que se ha comprobado que está relacionado con la síntesis de oligosacáridos en la leche que son beneficiosos para la salud humana (Poulsen et al., 2019). También se encontró que el gen *fgll* mostraba niveles de expresión elevados en respuesta a mastitis producida por *E. coli* (Sharifi et al., 2021). Además, Piórkowska et al. (2022) relacionaron a este gen con la infiltración intramuscular de grasa en algunas razas de cerdo y cuyo estudio podría ser interesante en razas de vacuno de carne.

En la raza Red Holstein se encontró al gen *tspan12*, que afecta a la producción de leche en ovejas (Mohammadi et al., 2022) y cuyo estudio en vacuno lechero podría ser interesante. Además, se ha observado que este gen aumenta su expresión en el tejido ruminal en respuesta a la acidosis en ganado vacuno (Gholizadeh et al., 2020).

En la raza Pirenaica Española se encontró al gen *acer2*, el cual está relacionado con el metabolismo lipídico (Laghuaouta et al., 2020) y el gen *ptpre*, asociado a un mayor rendimiento lácteo y graso en la leche de vaca (Laodim et al., 2024). Se identificaron dos genes relacionados con el crecimiento de los animales, como es el caso del gen *cela3b*, involucrado en la síntesis de enzimas proteolíticas en el páncreas (Boros et al., 2017) que pueden ayudar a una mejor digestión de los alimentos y absorción de nutrientes en vacas (Cheng et al., 2023) y el gen *ing5*, que está implicado en la ganancia diaria media de peso en el ganado vacuno según los estudios de Manca et al. (2020). Otro de los genes encontrados, *eomes*, aumentaba su expresión en vacas afectadas por tuberculosis bovina, ya que regula la actividad citotóxica y el desarrollo y función de células inmunitarias (Bhat et al., 2023) y además se ha visto que niveles de expresión elevados en este gen pueden conllevar una reducción de la expresión del interferón Tau

(IFNT) en rumiantes (Sakurai et al., 2013), dificultando el éxito de la gestación (Kim et al., 2019). El último gen encontrado en Pirenaica Española fue *ttyh2*, que estaba sobreexpresado en vacas infectadas por *Mycobacterium avium* subsp. *Paratuberculosis* en el estudio realizado por Malvisi et al. (2016).

Por último, en la raza Wagyu se encontró al gen *ak7*, relacionado con el metabolismo de carbohidratos, lípidos y aminoácidos (Iung et al., 2018). También se encontró al gen *dnpep*, el cual está relacionado con la ganancia de peso y la eficiencia alimentaria del ganado vacuno (Kern et al., 2016), este gen también contribuye a que haya una mayor cantidad de ácido glutámico en la carne al cocinarla, potenciando su sabor (Yuhara et al., 2021). El último gen encontrado en esta raza fue *fga*, que está ligado con el efecto materno en la cantidad de grasa intramuscular (Halli et al., 2022), un carácter de especial interés en esta raza de vacuno de carne.

Posteriormente, se analizaron los exones afectados por SVs que tenían en común las razas Holstein y Red Holstein y por otro lado las razas Pirenaica Española y Wagyu (Tabla 9). Los exones comunes entre Holstein y Red Holstein incluían al gen *bpi5b5*, perteneciente a la familia de genes LT/LBP (Sorbolini et al., 2015) y que participa en la respuesta inmune del ganado contra bacterias (Rahman et al., 2010). Otro de los genes encontrados fue *rnf32*, que está implicado en el proceso de espermatogénesis (Rios et al., 2023) y una mayor eficiencia alimentaria (Niu et al., 2021). El último gen encontrado en estas dos razas fue el gen *mucl1* y se observó que está relacionado con el rendimiento lácteo, el rendimiento graso de la leche y la tasa de preñez según los estudios de Olšanská (2020). Por otro lado, los exones comunes de Wagyu y Pirenaica Española incluían al gen *afap1*, el cual se encuentra más expresado en vacas que presentan cetosis como respuesta a esta enfermedad (Soares et al., 2021). Además del locus *loc525599*, que está involucrado en la respuesta inmune (Goyache et al., 2022), también se encontró al gen *capn1* que parece estar fuertemente asociado con la terneza de la carne de vacuno (Lee et al., 2019) y otras características de la canal (Shi et al., 2011) Finalmente se encontró al gen *slc6a14*. Gu et al. (2021) pudieron comprobar que modificando la dieta de vacas gestantes, el nivel de expresión de este gen aumentaba, mejorando el acceso del feto a los nutrientes e incrementando su peso al nacer. También se ha descrito la relación que tiene este gen con una mejor absorción de nutrientes en el duodeno, aumentando así la eficiencia de la producción del ganado (Waters et al.,

2013). Además, activa una vía de señalización que promueve la síntesis de β -caseína en las células mamarias bovinas (Cao et al., 2022), mejorando la calidad de la leche.

Raza	Gen	Tipo de SV	Cromosoma	Posible carácter asociado	Referencias
Holstein	<i>colgalt2</i>	DEL	16	Oligosacáridos en la leche	(Poulsen et al., 2019)
	<i>fgl1</i>	DEL	27	Resistencia a mastitis	(Sharifi et al., 2021)
Red Holstein	<i>tspan12</i>	DEL	4	Acidosis	(Gholizadeh et al., 2020)
Pirenaica Española	<i>acer2</i>	INS	8	Metabolismo lipídico	(Laghouaouta et al., 2020)
	<i>cela3b</i>	DEL	2	Absorción de nutrientes	(Cheng et al., 2023)
	<i>ing5</i>	INS	3	Ganancia de peso diaria	(Manca et al., 2020)
	<i>eomes</i>	INS	22	Regulación de respuesta inmune/ Inhibición de IFNT en rumiantes	(Bhat et al., 2023; Sakurai et al., 2013)
	<i>ptpre</i>	INS	26	Rendimiento lácteo y graso	(Laodim et al., 2024)
	<i>ttyh2</i>	DEL	19	Resistencia a paratuberculosis	(Malvisi et al., 2016)
Wagyu	<i>ak7</i>	DEL	21	Metabolismo	(Jung et al., 2018)
	<i>ddnpep</i>	INS	2	Ganancia de peso/Sabor de la carne	(Kern et al., 2016; Yuhara et al., 2021)
	<i>fga</i>	DEL	17	Efecto materno de grasa intramuscular	(Halli et al., 2022)

Tabla 8. Genes relacionados con la región exónica afectada por SVs exclusivas en cada raza.

Por último se buscaron SVs que afectasen a un mismo exon en común para las 4 razas (Tabla 10). La búsqueda resultó en 2 exones. El primero de ellos está relacionado con el gen *b3galnt2*, que está implicado en la síntesis de oligosacáridos en la leche (Poulsen et al., 2019) y también está involucrado en la resistencia contra la mastitis (Strillacci et al., 2023). El otro exón está relacionado con el locus loc100138951 del cromosoma 18 de la especie *Bos taurus*, afectando al sistema inmune y las interacciones

inmuno-regulatorias de células linfoides y no linfoides (Wu et al., 2017) y posiblemente relacionado con la resistencia a la paratuberculosis (Ortega Masagué, 2021).

Razas	Gen		Cromosoma	Posible carácter asociado	Referencias
Holstein y Red Holstein	<i>bpifb5</i>	DEL	13	Respuesta inmune	(Rahman et al., 2010)
	<i>mucl1</i>	DEL	5	Rendimiento lácteo y reproducción	(Olšanská, 2020)
	<i>rnf32</i>	INS	4	Eficiencia alimentaria	(Niu et al., 2021)
Wagyu y Pirenaica Española	<i>afap1</i>	DEL	6	Respuesta contra cetosis	(Soares et al., 2021)
	<i>capn1</i>	DEL	29	Calidad de la carne	(Lee et al., 2019; Shi, 2011)
	loc525599	INS	23	Respuesta inmune	(Goyache et al., 2022)
	<i>slc6a14</i>	INS	X	Síntesis de β -caseína y mayor acceso a nutrientes	(Cao et al., 2022; Gu et al., 2021; Waters et al., 2013)

Tabla 9. Genes relacionados con la región exónica afectada por SVs compartidas entre las razas Holstein y Red Holstein y las razas Wagyu y Pirenaica Española.

Gen		Cromosoma	Posible carácter asociado	Referencias
<i>b3galnt2</i>	INS	28	Oligosacáridos en la leche y resistencia a mastitis	(Poulsen et al., 2019; Strillacci et al., 2023)
loc100138951	DEL	18	Respuesta inmune Resistencia a paratuberculosis	(Ortega Masagué, 2021; Wu et al., 2017)

Tabla 10. Genes relacionados con la región exónica afectada por SVs compartidas entre las cuatro razas estudiadas.

5. Conclusiones

1. Debido a los pocos mismatches obtenidos con ambos alineadores, la secuenciación ONT ha mostrado ser muy precisa, con mas de un 99%, en el basecalling de las muestras.
2. Tanto LRA como Minimap2 son buenos alineadores que han mostrado comportamientos similares en ambas muestras.
3. Mientras que Sniffles2 detecta mejor inserciones y deleciones, CuteSV es capaz de detectar mejor BNDs, inversiones y duplicaciones.
4. El 45,33% de las SVs fueron detectadas por ambos softwares, pero el 54,67% solo se identificaron con uno de ellos. Por lo tanto, sería necesario validar los resultados con un set donde se conozcan las SVs.
5. La mayoría de las SVs detectadas son inserciones y deleciones de alrededor de 100pb.
6. Hay un mayor número de SVs en los primeros BTAs, más grandes, en comparación con los BTAs finales, de menor tamaño.
7. Por lo general, las SVs se distribuyen de forma uniforme a lo largo del genoma, a excepción de algunas regiones que presentan una mayor densidad, como el BTA23.
8. Alrededor del 80% de las SVs detectadas se localizan en regiones intergénicas ya que la secuenciación no está dirigida. Por ello, es estadísticamente más probable que las SVs se localicen en estas regiones, ya que la mayoría del genoma está compuesto por ellas.
9. Se encontraron SVs exclusivas de raza Holstein que afectan a los genes *colgalt2* y *fgl1*, ambos relacionados con la producción de leche.
10. Se encontró al gen *tspan12* afectado por SVs en raza Red Holstein, relacionado con la respuesta a la acidosis.
11. Se encontraron SVs exclusivas de raza Pirenaica Española afectando a los genes *acer2* *cela3b* *ing5* *eomes* y *ttyh2*. Estos genes están relacionados con la resistencia de los animales y un mejor crecimiento, con posibles implicaciones en la producción cárnica.

12. Se encontraron SVs exclusivas de raza Wagyu afectando a los genes *ak7*, *ddnpep* y *fga*. Estos genes están relacionados con caracteres propios de esta raza en producción de carne, como la infiltración de grasa intramuscular y un mejor sabor de la carne.
13. Las razas Holstein y Red Holstein comparten SVs que afectan a los genes *bpifb5*, *mucl1* y *rnf32* que están relacionados con la respuesta inmune y la producción láctea.
14. Las razas Pirenaica Española y Wagyu comparten SVs que afectan a los genes *afap1*, *capn1*, *loc525599* y *slc6a14* que están relacionados con la calidad de la carne, la resistencia y el desarrollo de los animales.
15. Todas las razas compartían SVs que afectan a los genes *loc100138951* y *b3galnt2*, con implicaciones en la producción láctea y la resistencia de los animales a enfermedades.
16. Debido al bajo tamaño muestral de este estudio se requieren más investigaciones que permitan validar los anteriores resultados.

6. Bibliografía

- Abel, H. J., Larson, D. E., Chiang, C., Das, I., Kanchi, K. L., Layer, R. M., ... Hall, I. M. (2019). NHGRI Centers for Common Disease Genomics. *bioRxiv*, 508515. <https://doi.org/10.1101/508515>
- Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, 12(5), 363-376. <https://doi.org/10.1038/nrg2958>
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1), 30. <https://doi.org/10.1186/s13059-020-1935-5>
- Antolín, R., Nettelblad, C., Gorjanc, G., Money, D., & Hickey, J. M. (2017). A hybrid method for the imputation of genomic data in livestock populations. *Genetics Selection Evolution*, 49(1), 30. <https://doi.org/10.1186/s12711-017-0300-y>
- Audano, P. A., Sulovari, A., Graves-Lindsay, T. A., Cantsilieris, S., Sorensen, M., Welch, A. E., Dougherty, M. L., Nelson, B. J., Shah, A., Dutcher, S. K., Warren, W. C., Magrini, V., McGrath, S. D., Li, Y. I., Wilson, R. K., & Eichler, E. E. (2019). Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell*, 176(3), 663-675.e19. <https://doi.org/10.1016/j.cell.2018.12.019>
- Barton, N. H., Etheridge, A. M., & Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, 118, 50-73. <https://doi.org/10.1016/j.tpb.2017.06.001>
- Bhat, S. A., Elnaggar, M., Hall, T. J., McHugo, G. P., Reid, C., MacHugh, D. E., & Meade, K. G. (2023). Preferential differential gene expression within the WC1.1+ $\gamma\delta$ T cell compartment in cattle naturally infected with *Mycobacterium bovis*. *Frontiers in Immunology*, 14, 1265038. <https://doi.org/10.3389/fimmu.2023.1265038>
- Bickhart, D. M., & Liu, G. E. (2014). The challenges and importance of structural variation detection in livestock. *Frontiers in Genetics*, 5. <https://doi.org/10.3389/fgene.2014.00037>
- Boros, E., Szabó, A., Zboray, K., Héja, D., Pál, G., & Sahin-Tóth, M. (2017). Overlapping Specificity of Duplicated Human Pancreatic Elastase 3 Isoforms and Archetypal Porcine Elastase 1 Provides Clues to Evolution of Digestive Enzymes. *Journal of Biological Chemistry*, 292(7), 2690-2702. <https://doi.org/10.1074/jbc.M116.770560>
- Boussaha, M., Esquerré, D., Barbieri, J., Djari, A., Pinton, A., Letaief, R., Salin, G., Escudié, F., Roulet, A., Fritz, S., Samson, F., Grohs, C., Bernard, M., Klopp, C., Boichard, D., & Rocha, D. (2015). Genome-Wide Study of Structural Variants in Bovine Holstein, Montbéliarde and Normande Dairy Breeds. *PLOS ONE*, 10(8), e0135931. <https://doi.org/10.1371/journal.pone.0135931>
- Cao, Y., Hu, G., Li, W., Wang, J., Ge, Y., Li, F., Guo, W., Kan, X., Fu, S., & Liu, J. (2022). Lysine promotes proliferation and β -casein synthesis through the SLC6A14-ERK1/2-CDK1-mTOR signaling pathway in bovine primary mammary epithelial cells. *Journal of Thermal Biology*, 110, 103375. <https://doi.org/10.1016/j.jtherbio.2022.103375>
- Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, 12(1), 35. <https://doi.org/10.1186/1471-2105-12-35>
- Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L., & Mardis, E. R. (2009).

BreakDancer: An algorithm for high-resolution mapping of genomic structural variation. *Nature Methods*, 6(9), 677-681. <https://doi.org/10.1038/nmeth.1363>

Cheng, Z., Ferris, C., Crowe, M. A., Ingvarstsen, K. L., Grelet, C., Vanlierde, A., Foldager, L., Becker, F., Wathes, D. C., & the GplusE Consortium. (2023). Hepatic Global Transcriptomic Profiles of Holstein Cows According to Parity Reveal Age-Related Changes in Early Lactation. *International Journal of Molecular Sciences*, 24(12), 9906. <https://doi.org/10.3390/ijms24129906>

Di Gerlando, R., Sutura, A. M., Mastrangelo, S., Tolone, M., Portolano, B., Sottile, G., Bagnato, A., Strillacci, M. G., & Sardina, M. T. (2019). Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLOS ONE*, 14(4), e0215204. <https://doi.org/10.1371/journal.pone.0215204>

Donders, A. R. T., Van Der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59(10), 1087-1091. <https://doi.org/10.1016/j.jclinepi.2006.01.014>

Duan, X., Pan, M., & Fan, S. (2022). Comprehensive evaluation of structural variant genotyping methods based on long-read sequencing data. *BMC Genomics*, 23(1), 324. <https://doi.org/10.1186/s12864-022-08548-y>

Durkin, K., Coppieters, W., Drögemüller, C., Ahariz, N., Cambisano, N., Druet, T., Fasquelle, C., Haile, A., Horin, P., Huang, L., Kamatani, Y., Karim, L., Lathrop, M., Moser, S., Oldenbroek, K., Rieder, S., Sartelet, A., Sölkner, J., Stålhammar, H., ... Charlier, C. (2012). Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature*, 482(7383), 81-84. <https://doi.org/10.1038/nature10757>

Escaramís, G., Docampo, E., & Rabionet, R. (2015). A decade of structural variants: Description, history and methods to detect structural variation. *Briefings in Functional Genomics*, 14(5), 305-314. <https://doi.org/10.1093/bfpg/elv014>

Feng, W., Gao, P., & Wang, X. (2024). AI breeder: Genomic predictions for crop breeding. *New Crops*, 1, 100010. <https://doi.org/10.1016/j.ncrops.2023.12.005>

Fisher, R. A. (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2), 399-433. <https://doi.org/10.1017/S0080456800012163>

Galton, F. (1876). A Theory of Heredity. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 5, 329–348. <https://doi.org/10.2307/2840896>

Ganal, M. W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H., & Durstewitz, G. (2012). Large SNP arrays for genotyping in crop plants. *Journal of Biosciences*, 37(5), 821-828. <https://doi.org/10.1007/s12038-012-9225-3>

Gholizadeh, M., Fayazi, J., Asgari, Y., Zali, H., & Kaderali, L. (2020). Reconstruction and Analysis of Cattle Metabolic Networks in Normal and Acidosis Rumen Tissue. *Animals*, 10(3), 469. <https://doi.org/10.3390/ani10030469>

Gong, T., Hayes, V. M., & Chan, E. K. F. (2021). Detection of somatic structural variants from short-read next-generation sequencing data. *Briefings in Bioinformatics*, 22(3), bbaa056. <https://doi.org/10.1093/bib/bbaa056>

Goyache, F., Pérez-Pardal, L., Fernández, I., Traoré, A., Menéndez-Arias, N. A., Arias, K. D., & Álvarez, I. (2022). Identification and Characterization of Copy Number Variations Regions in West African Taurine Cattle. *Animals*, 12(16), 2130. <https://doi.org/10.3390/ani12162130>

Gu, F., Jiang, L., Xie, L., Wang, D., Zhao, F., & Liu, J. (2021). Supplementing N-carbamoylglutamate in late gestation increases newborn calf weight by enhanced placental expression of mTOR and angiogenesis factor genes in dairy cows. *Animal Nutrition*, 7(4), 981-988. <https://doi.org/10.1016/j.aninu.2021.05.007>

Halli, K., Bohlouli, M., Schulz, L., Sundrum, A., & König, S. (2022). Estimation of direct and maternal genetic effects and annotation of potential candidate genes for weight and meat quality traits in a genotyped outdoor dual-purpose cattle breed. *Translational Animal Science*, 6(1), txac022. <https://doi.org/10.1093/tas/txac022>

Hayes, B. J., & Daetwyler, H. D. (2019). 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annual Review of Animal Biosciences*, 7(1), 89-102. <https://doi.org/10.1146/annurev-animal-020518-115024>

Helal, A. A., Saad, B. T., Saad, M. T., Mosaad, G. S., & Aboshanab, K. M. (2024). Benchmarking long-read aligners and SV callers for structural variation detection in Oxford nanopore sequencing data. *Scientific Reports*, 14(1), 6160. <https://doi.org/10.1038/s41598-024-56604-2>

Hickey, J. M., Kinghorn, B. P., Tier, B., Van Der Werf, J. H., & Cleveland, M. A. (2012). A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genetics Selection Evolution*, 44(1), 9. <https://doi.org/10.1186/1297-9686-44-9>

Hu, Z., Wang, Z., & Xu, S. (2012). An Infinitesimal Model for Quantitative Trait Genomic Value Prediction. *PLoS ONE*, 7(7), e41336. <https://doi.org/10.1371/journal.pone.0041336>

Huang, X., & Chao, K.-M. (2003). A generalized global alignment algorithm. *Bioinformatics*, 19(2), 228-233. <https://doi.org/10.1093/bioinformatics/19.2.228>

Lung, L. H. D. S., Mulder, H. A., Neves, H. H. D. R., & Carvalheiro, R. (2018). Genomic regions underlying uniformity of yearling weight in Nelore cattle evaluated under different response variables. *BMC Genomics*, 19(1), 619. <https://doi.org/10.1186/s12864-018-5003-4>

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17(1), 239. <https://doi.org/10.1186/s13059-016-1103-0>

Jiang, T., Liu, Y., Jiang, Y., Li, J., Gao, Y., Cui, Z., Liu, Y., Liu, B., & Wang, Y. (2020). Long-read-based human genomic structural variation detection with cuteSV. *Genome Biology*, 21(1), 189. <https://doi.org/10.1186/s13059-020-02107-y>

Jiang, T., Cao, S., Liu, Y., Liu, S., Liu, B., Wang, G., & Wang, Y. (2022). Regenotyping structural variants through an accurate force-calling method. *bioRxiv*, 2022-08. <https://doi.org/10.1101/2022.08.29.505534>

Jiang, Y.-F., Wang, S., Wang, C.-L., Xu, R.-H., Wang, W.-W., Jiang, Y., Wang, M.-S., Jiang, L., Dai, L.-H., Wang, J.-R., Chu, X.-H., Zeng, Y.-Q., Fang, L.-Z., Wu, D.-D., Zhang, Q., & Ding, X.-D. (2023). Pangenome obtained by long-read sequencing of 11 genomes reveal hidden functional structural variants in pigs. *iScience*, 26(3), 106119. <https://doi.org/10.1016/j.isci.2023.106119>

Kadri, N. K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbandsen, B., Karim, L., Nielsen, U. S., Panitz, F., Aamand, G. P., Schulman, N., Georges, M., Vilkki, J., Lund, M. S., & Druet, T. (2014). A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *PLoS Genetics*, 10(1), e1004049. <https://doi.org/10.1371/journal.pgen.1004049>

- Kern, R. J., Lindholm-Perry, A. K., Freetly, H. C., Snelling, W. M., Kern, J. W., Keele, J. W., Miles, J. R., Foote, A. P., Oliver, W. T., Kuehn, L. A., & Ludden, P. A. (2016). Transcriptome differences in the rumen of beef steers with variation in feed intake and gain. *Gene*, *586*(1), 12-26. <https://doi.org/10.1016/j.gene.2016.03.034>
- Kiełbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research*, *21*(3), 487-493. <https://doi.org/10.1101/gr.113985.110>
- Kim, C., Pongpanich, M., & Porntaveetus, T. (2024). Unraveling metagenomics through long-read sequencing: A comprehensive review. *Journal of Translational Medicine*, *22*(1), 111. <https://doi.org/10.1186/s12967-024-04917-1>
- Kim, M.-S., Lim, H.-J., Lee, J. H., Hur, T. Y., & Son, J. K. (2019). Temporal Regulation of Ovine Interferon-tau Gene by the Transcription Factor Eomesodermin in the Peri-Implantation Period. *Journal of Animal Reproduction and Biotechnology*, *34*(4), 292-299. <https://doi.org/10.12750/JARB.34.4.292>
- Laghouaouta, H., Sosa-Madrid, B. S., Zubiri-Gaitán, A., Hernández, P., & Blasco, A. (2020). Novel Genomic Regions Associated with Intramuscular Fatty Acid Composition in Rabbits. *Animals*, *10*(11), 2090. <https://doi.org/10.3390/ani10112090>
- Laodim, T., Koonawootrittriron, S., Elzo, M. A., Suwanasopee, T., Jattawa, D., & Sarakul, M. (2024). Genetic factors influencing milk and fat yields in tropically adapted dairy cattle: Insights from quantitative trait loci analysis and gene associations. *Animal Bioscience*, *37*(4), 576-590. <https://doi.org/10.5713/ab.23.0246>
- Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). LUMPY: A probabilistic framework for structural variant discovery. *Genome Biology*, *15*(6), R84. <https://doi.org/10.1186/gb-2014-15-6-r84>
- Lee, H. J., Jin, S., Kim, H.-J., Bhuiyan, M. S. A., Lee, D. H., Lee, S. H., Jang, S. B., Han, M. H., & Lee, S. H. (2019). Validation Study of SNPs in CAPN1-CAST Genes on the Tenderness of Muscles (Longissimus thoracis and Semimembranosus) in Hanwoo (Korean Cattle). *Animals*, *9*(9), 691. <https://doi.org/10.3390/ani9090691>
- Li, H. (2016). Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, *32*(14), 2103-2110. <https://doi.org/10.1093/bioinformatics/btw152>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, R., Gong, M., Zhang, X., Wang, F., Liu, Z., Zhang, L., et al. (2023). A sheep pangenome reveals the spectrum of structural variations and their effects on tail phenotypes. *Genome Research*, *33*, 463-477. <https://doi.org/10.1101/gr.1234567>
- Li, Y., Willer, C., Sanna, S., & Abecasis, G. (2009). Genotype Imputation. *Annual Review of Genomics and Human Genetics*, *10*(1), 387-406. <https://doi.org/10.1146/annurev.genom.9.081307.164242>
- Liu, D., Chen, Z., Zhang, Z., Sun, H., Ma, P., Zhu, K., Liu, G., Wang, Q., & Pan, Y. (2019). Detection of genome-wide structural variations in the Shanghai Holstein cattle population using next-generation sequencing. *Asian-Australasian Journal of Animal Sciences*, *32*(3), 320-333. <https://doi.org/10.5713/ajas.18.0204>
- Liu, Y., Zhang, M., Wang, R., Li, B., Jiang, Y., Sun, M., Chang, Y., & Wu, J. (2022). Comparison of structural variants detected by PacBio-CLR and ONT sequencing in pear. *BMC Genomics*, *23*(1), 830. <https://doi.org/10.1186/s12864-022-09074-7>

- LoTempio, J., Delot, E., & Vilain, E. (2023). Benchmarking long-read genome sequence alignment tools for human genomics applications. *PeerJ*, *11*, e16515. Bioinformatics. <https://doi.org/10.1101/2021.07.09.451840>
- Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biology*, *20*(1), 246. <https://doi.org/10.1186/s13059-019-1828-7>
- Malvisi, M., Palazzo, F., Morandi, N., Lazzari, B., Williams, J. L., Pagnacco, G., & Minozzi, G. (2016). Responses of Bovine Innate Immunity to Mycobacterium avium subsp. Paratuberculosis Infection Revealed by Changes in Gene Expression and Levels of MicroRNA. *PLOS ONE*, *11*(10), e0164461. <https://doi.org/10.1371/journal.pone.0164461>
- Manca, E., Cesarani, A., Gaspa, G., Sorbolini, S., Macciotta, N. P. P., & Dimauro, C. (2020). Use of the Multivariate Discriminant Analysis for Genome-Wide Association Studies in Cattle. *Animals*, *10*(8), 1300. <https://doi.org/10.3390/ani10081300>
- Mendel, G. (1866). Experiments in plant hybridization (1865). *Proceedings of the Natural History Society of Brinn, IV*, 3–47
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, *157*(4), 1819-1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, *6*(1), 6-14. <https://doi.org/10.2527/af.2016-0002>
- Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., Abyzov, A., Yoon, S. C., Ye, K., Cheetham, R. K., Chinwalla, A., Conrad, D. F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L. M., Iqbal, Z., Kang, S., ... Korbel, J. O. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, *470*(7332), 59-65. <https://doi.org/10.1038/nature09708>
- Mohammadi, H., Farahani, A. H. K., Moradi, M. H., Mastrangelo, S., Di Gerlando, R., Sardina, M. T., Scatassa, M. L., Portolano, B., & Tolone, M. (2022). Weighted Single-Step Genome-Wide Association Study Uncovers Known and Novel Candidate Genomic Regions for Milk Production Traits and Somatic Cell Score in Valle del Belice Dairy Sheep. *Animals*, *12*(9), 1155. <https://doi.org/10.3390/ani12091155>
- Morange, M. (2009). The central dogma of molecular biology. *Resonance*, *14*(3), 236–247. <https://doi.org/10.1007/s12045-009-0024-6>
- Nelson, R. M., Pettersson, M. E., & Carlborg, Ö. (2013). A century after Fisher: Time for a new paradigm in quantitative genetics. *Trends in Genetics*, *29*(12), 669-676. <https://doi.org/10.1016/j.tig.2013.09.006>
- Nguyen, T. V., Vander Jagt, C. J., Wang, J., Daetwyler, H. D., Xiang, R., Goddard, M. E., Nguyen, L. T., Ross, E. M., Hayes, B. J., Chamberlain, A. J., & MacLeod, I. M. (2023). In it for the long run: Perspectives on exploiting long-read sequencing in livestock for population scale studies of structural variants. *Genetics Selection Evolution*, *55*(1), 9. <https://doi.org/10.1186/s12711-023-00783-5>
- Niu, Q., Zhang, T., Xu, L., Wang, T., Wang, Z., Zhu, B., Zhang, L., Gao, H., Song, J., Li, J., & Xu, L. (2021). Integration of selection signatures and multi-trait GWAS reveals polygenic genetic architecture of carcass traits in beef cattle. *Genomics*, *113*(5), 3325-3336. <https://doi.org/10.1016/j.ygeno.2021.07.025>
- Norris, B. J., & Whan, V. A. (2008). A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep. *Genome Research*, *18*(8), 1282-1293. <https://doi.org/10.1101/gr.072090.107>

Olišanská, B. (2020). Genome-wide characterisation of regions under intense selection based on runs of homozygosity in Charolais cattle. *Acta Fytotechnica et Zootechnica*, 23(Monothematic Issue), 347-352. <https://doi.org/10.15414/afz.2020.23.mi-fpap.347-352>

Ortega Masague, M. F. (2021). *Resistencia genética a Rhipicephalus (Boophilus) microplus en bovinos de raza criollo argentino* (Doctoral dissertation, Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata)

Piórkowska, K., Żukowski, K., Ropka-Molik, K., & Tyra, M. (2022). Variations in Fibrinogen-like 1 (FGL1) Gene Locus as a Genetic Marker Related to Fat Deposition Based on Pig Model and Liver RNA-Seq Data. *Genes*, 13(8), 1419. <https://doi.org/10.3390/genes13081419>

Poulsen, N. A., Robinson, R. C., Barile, D., Larsen, L. B., & Buitenhuis, B. (2019). A genome-wide association study reveals specific transferases as candidate loci for bovine milk oligosaccharides synthesis. *BMC Genomics*, 20(1), 404. <https://doi.org/10.1186/s12864-019-5786-y>

Rahman, Md. M., Lecchi, C., Avallone, G., Roccabianca, P., Sartorelli, P., & Ceciliani, F. (2010). Lipopolysaccharide-binding protein: Local expression in bovine extrahepatic tissues. *Veterinary Immunology and Immunopathology*, 137(1-2), 28-35. <https://doi.org/10.1016/j.vetimm.2010.04.006>

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28(18), i333-i339. <https://doi.org/10.1093/bioinformatics/bts378>

R Core Team. (2023). R: A language and environment for statistical computing (Version 4.2.0) [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>

Ren, J., & Chaisson, M. J. P. (2021). Ira: A long read aligner for sequences and contigs. *PLOS Computational Biology*, 17(6), e1009078. <https://doi.org/10.1371/journal.pcbi.1009078>

Rios, A. C. H., Nasner, S. L. C., Londoño-Gil, M., Gonzalez-Herrera, L. G., Lopez-Herrera, A., & Flórez, J. C. R. (2023). Genome-wide association study for reproduction traits in Colombian Creole Blanco Orejinegro cattle. *Tropical Animal Health and Production*, 55(6), 429. <https://doi.org/10.1007/s11250-023-03847-2>

Romain, S., & Lemaitre, C. (2023). SVJedi-graph: Improving the genotyping of close and overlapping structural variants with long reads using a variation graph. *Bioinformatics*, 39(Supplement_1), i270-i278. <https://doi.org/10.1093/bioinformatics/btad237>

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., Rowan, T. N., Low, W. Y., Zimin, A., Couldrey, C., Hall, R., Li, W., Rhie, A., Ghurye, J., McKay, S. D., Thibaud-Nissen, F., Hoffman, J., Murdoch, B. M., Snelling, W. M., ... Medrano, J. F. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience*, 9(3), g10021. <https://doi.org/10.1093/gigascience/giaa021>

Rubin, C.-J., Megens, H.-J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C. B., Archibald, A. L., Fredholm, M., Groenen, M. A. M., & Andersson, L. (2012). Strong signatures of selection in the domestic pig genome. *Proceedings of the National Academy of Sciences*, 109(48), 19529-19536. <https://doi.org/10.1073/pnas.1217149109>

Sakurai, T., Bai, H., Bai, R., Sato, D., Arai, M., Okuda, K., Ideta, A., Aoyagi, Y., Godkin, J. D., & Imakawa, K. (2013). Down-regulation of interferon tau gene transcription with a transcription factor, EOMES. *Molecular Reproduction and Development*, 80(5), 371-383. <https://doi.org/10.1002/mrd.22171>

Salehian-Dehkordi, H., Xu, Y.-X., Xu, S.-S., Li, X., Luo, L.-Y., Liu, Y.-J., Wang, D.-F., Cao, Y.-H., Shen, M., Gao, L., Chen, Z.-H., Glessner, J. T., Lenstra, J. A., Esmailzadeh, A., Li, M.-H., & Lv, F.-H. (2021). Genome-

Wide Detection of Copy Number Variations and Their Association With Distinct Phenotypes in the World's Sheep. *Frontiers in Genetics*, 12, 670582. <https://doi.org/10.3389/fgene.2021.670582>

Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Von Haeseler, A., & Schatz, M. C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, 15(6), 461-468. <https://doi.org/10.1038/s41592-018-0001-7>

Sedlazeck, F. J., Rescheneder, P., & Von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, 29(21), 2790-2791. <https://doi.org/10.1093/bioinformatics/btt468>

Seo, J.-S., Rhie, A., Kim, J., Lee, S., Sohn, M.-H., Kim, C.-U., Hastie, A., Cao, H., Yun, J.-Y., Kim, J., Kuk, J., Park, G. H., Kim, J., Ryu, H., Kim, J., Roh, M., Baek, J., Hunkapiller, M. W., Korlach, J., ... Kim, C. (2016). De novo assembly and phasing of a Korean human genome. *Nature*, 538(7624), 243-247. <https://doi.org/10.1038/nature20098>

Sharifi, S., Lotfi Shahreza, M., Pakdel, A., Reecy, J. M., Ghadiri, N., Atashi, H., Motamedi, M., & Ebrahimi, E. (2021). Systems Biology-Derived Genetic Signatures of Mastitis in Dairy Cattle: A New Avenue for Drug Repurposing. *Animals*, 12(1), 29. <https://doi.org/10.3390/ani12010029>

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, 11(10), e0163962. <https://doi.org/10.1371/journal.pone.0163962>

Shi, M., Gao, X., Ren, H., Yuan, Z., Wu, H., Li, J., ... & Xu, S. (2011). Association analysis of CAPN1 gene variants with carcass and meat quality traits in Chinese native cattle. *African Journal of Biotechnology*, 10(75), 17367-17371. <https://doi.org/10.5897/AJB11.2306>

Smolka, M., Paulin, L. F., Grochowski, C. M., Horner, D. W., Mahmoud, M., Behera, S., ... & Sedlazeck, F. J. (2024). Detection of mosaic and population-level structural variants with Sniffles2. *Nature biotechnology*, 1-10

Soares, R. A. N., Vargas, G., Duffield, T., Schenkel, F., & Squires, E. J. (2021). Genome-wide association study and functional analyses for clinical and subclinical ketosis in Holstein cattle. *Journal of Dairy Science*, 104(9), 10076-10089. <https://doi.org/10.3168/jds.2020-20101>

Sorbolini, S., Marras, G., Gaspa, G., Dimauro, C., Cellesi, M., Valentini, A., & Macciotta, N. P. (2015). Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories. *Genetics Selection Evolution*, 47(1), 52. <https://doi.org/10.1186/s12711-015-0128-2>

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., Redon, R., Bird, C. P., De Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavaré, S., Deloukas, P., Hurles, M. E., & Dermitzakis, E. T. (2007). Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science*, 315(5813), 848-853. <https://doi.org/10.1126/science.1136678>

Strillacci, M. G., Punturiero, C., Milanese, R., Bernini, F., Mason, T., & Bagnato, A. (2023). Antibiotic treatments and somatic cell count as phenotype to map QTL for mastitis susceptibility in Holstein cattle breed. *Italian Journal of Animal Science*, 22(1), 190-199. <https://doi.org/10.1080/1828051X.2023.2175498>

Sudmant, P. H., Mallick, S., Nelson, B. J., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B. P., Baker, C., Nordenfelt, S., Bamshad, M., Jorde, L. B., Posukh, O. L., Sahakyan, H., Watkins, W. S., Yepiskoposyan, L., Abdullah, M. S., Bravi, C. M., Capelli, C., Hervig, T., ... Eichler, E. E. (2015). Global diversity, population stratification, and selection of human copy-number variation. *Science*, 349(6253), aab3761. <https://doi.org/10.1126/science.aab3761>

- Wang, Q., Li, M., Wu, T., Zhan, L., Li, L., Chen, M., Xie, W., Xie, Z., Hu, E., Xu, S., & Yu, G. (2022). Exploring epigenomic datasets by ChIPseeker. *Current Protocols*, 2(10), e585. <https://doi.org/10.1002/cpz1.585>
- Wang, Z., Chen, Q., Liao, R., Zhang, Z., Zhang, X., Liu, X., ... & Pan, Y. (2017). Genome-wide genetic variation discovery in Chinese Taihu pig breeds using next generation sequencing. *Animal genetics*, 48(1), 38-47. <https://doi.org/10.1111/age.12465>
- Waters, S. M., Keogh, K., Buckley, F., & Kenny, D. A. (2013). Effect of genotype on duodenal expression of nutrient transporter genes in dairy cows. *Journal of Animal Science and Biotechnology*, 4(1), 49. <https://doi.org/10.1186/2049-1891-4-49>
- Wickham, H., & Wickham, H. (2016). *Data analysis* (pp. 189-201). Springer International Publishing. doi.org/10.1007/978-3-319-24277-4_9
- Wright, D., Boije, H., Meadows, J. R. S., Bed'hom, B., Gourichon, D., Vieaud, A., Tixier-Boichard, M., Rubin, C.-J., Imsland, F., Hallböök, F., & Andersson, L. (2009). Copy Number Variation in Intron 1 of SOX5 Causes the Pea-comb Phenotype in Chickens. *PLoS Genetics*, 5(6), e1000512. <https://doi.org/10.1371/journal.pgen.1000512>
- Wu, X., Guldbrandtsen, B., Nielsen, U. S., Lund, M. S., & Sahana, G. (2017). Association analysis for young stock survival index with imputed whole-genome sequence variants in Nordic Holstein cattle. *Journal of Dairy Science*, 100(8), 6356-6370. <https://doi.org/10.3168/jds.2017-12688>
- Xia, X., Zhang, F., Li, S., Luo, X., Peng, L., Dong, Z., Pausch, H., Leonard, A. S., Crysanto, D., Wang, S., Tong, B., Lenstra, J. A., Han, J., Li, F., Xu, T., Gu, L., Jin, L., Dang, R., Huang, Y., ... Chen, N. (2023). Structural variation and introgression from wild populations in East Asian cattle genomes confer adaptation to local environment. *Genome Biology*, 24(1), 211. <https://doi.org/10.1186/s13059-023-03052-2>
- Xiaoli, L., Fengbin, H., Shihui, H., Xi, N., Sheng, L., Zhou, W., Xueqin, R., & Jiafu, W. (2021). Detection of genomic structure variants associated with wrinkled skin in Xiang pig by next generation sequencing. *Aging*, 13(22), 24710-24739. <https://doi.org/10.18632/aging.203711>
- Yildiz, G., Zanini, S. F., Afsharyan, N. P., Obermeier, C., Snowdon, R. J., & Golicz, A. A. (2023). Benchmarking Oxford Nanopore read alignment-based insertion and deletion detection in crop plant genomes. *The Plant Genome*, 16(2), e20314. <https://doi.org/10.1002/tpg2.20314>
- Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison, and visualization. *Bioinformatics*, 31(14), 2382-2383. <https://doi.org/10.1093/bioinformatics/btv145>
- Yuhara, H., Ohtani, A., Matano, M., Kashiwagi, Y., & Maehashi, K. (2021). Molecular characterization of a novel aspartyl aminopeptidase that contributes to the increase in glutamic acid content in chicken meat during cooking. *Food Chemistry: Molecular Sciences*, 2, 100015. <https://doi.org/10.1016/j.fochms.2021.100015>
- Zhang, H., Du, Z.-Q., Dong, J.-Q., Wang, H.-X., Shi, H.-Y., Wang, N., Wang, S.-Z., & Li, H. (2014). Detection of genome-wide copy number variations in two chicken lines divergently selected for abdominal fat content. *BMC Genomics*, 15(1), 517. <https://doi.org/10.1186/1471-2164-15-517>