

## Genómica y bioinformática: sus aplicaciones en salud y producción animal

Nélida Rodríguez-Osorio<sup>1</sup>

<sup>1</sup>Unidad de Genómica y Bioinformática, Departamento de Ciencias Biológicas, Centro Universitario Regional Norte, Universidad de la República, Uruguay.

En diciembre de 1977, después de una larga y compleja gestación, nació una nueva rama de la ciencia cuando Frederick Sanger y sus colaboradores dieron a conocer un nuevo método para secuenciar ácidos nucleicos, usando didesoxinucleótidos para interrumpir la polimerización de las cadenas de DNA y determinar su secuencia (Sanger, Nicklen y Coulson, 1977)F. & Coulson, A. R. (1975. “El método de Sanger”, como pasó a ser conocido, se convirtió en el estándar de secuenciación de DNA e hizo posible la obtención de los primeros genomas completos (Tabla 1).

Diez años después, el editorial del primer número de la revista *Genomics: A new discipline, a new name, a new journal*, oficializó el nombre Genómica no solo

para la revista, sino también para esa nueva rama de la ciencia (McKusick y Ruddle, 1987). Hoy entendemos por *Genómica* el conjunto de conocimientos, técnicas y herramientas para estudiar la estructura, la secuencia, el funcionamiento, el origen y la evolución de los genomas completos, sus partes y sus productos - RNA y proteínas.

El manejo de las secuencias de genomas grandes precisó el desarrollo de recursos computacionales específicos. La genómica se unió a la **Bioinformática**, surgida con los programas desarrollados por Margaret Dayhoff en los años 60 para catalogar y generar bases de datos de proteínas (Masic, 2016). La unión entre genómica y bioinformática se consolidó con programas para procesar y comparar secuencias de DNA (Staden,

**Tabla 1.** Primeros genomas secuenciados de diferentes grupos taxonómicos mediante el método de Sanger. \*Secuencia obtenida por el método de fraccionamiento enzimático en dos dimensiones, también desarrollado por Sanger.

Primer genoma	Especie	Longitud en nucleótidos	Año	Referencia
Fago *	Bacteriofago MS2	3.569	1976	Volckaert et al.
Organela	Mitocondria humana	15.569	1977	Anderson et al.
Viral	Virus de Epstein-Barr	172.282	1984	Baer et al.
Bacteriano	<i>Haemophilus influenzae</i>	1.830.137	1995	Fleischmann et al.
Eucariota	<i>Saccharomyces.cerevisiae</i>	12.068.000	1996	Goffeau et al.
Metazoario	<i>Caenorhabditis elegans</i>	97.000.000	1998	The C. elegans sequencing consortium
Insecto	<i>Drosophila melanogaster</i>	143.726.000	2000	Adams et al.
Vegetal	<i>Arabidopsis thaliana</i>	116.846.000	2000	The Arabidopsis Genome Initiative
Mamífero	<i>Homo sapiens</i>	3.000.000.000	2001	International Human Genome Sequencing Consortium

1977; Pearson y Lipman, 1988; Altschul *et al.*, 1990). Hoy la bioinformática desarrolla *software* para catalogar, manejar, analizar, comparar y graficar secuencias de DNA, RNA y proteínas, para que esos datos se conviertan en información con significado biológico.

La genómica se ha convertido en un elemento unificador de diversas áreas de la investigación biológica y de la salud. *Todo lo que tiene un genoma*, desde virus hasta ballenas, desde individuos hasta poblaciones, puede ser estudiado por ramificaciones y subdivisiones de la genómica agrupadas bajo el sufijo “ómicas”: genómica evolutiva, genómica funcional, transcriptómica, proteómica, nutrigenómica, farmacogenómica, genómica del cáncer, metagenómica, entre otras.

Un número creciente de publicaciones en ciencias pecuarias incluyen algún tipo de herramienta genómica y análisis bioinformático. En esta revisión nos centraremos en 4 aplicaciones de relevancia para la salud y el bienestar animal: 1) la secuenciación de genomas completos, 2) el uso de variantes genómicas para diagnóstico de enfermedades y selección de animales, 3) el estudio simultáneo de poblaciones de microorganismos de interés y 4) el estudio de la expresión global de genes.

## Los Genomas

A partir de 2004 surgieron nuevas tecnologías que permitieron secuenciar DNA de manera más rápida y barata que el método de Sanger. Se las llamó segunda generación, NGS (del inglés *Next Generation Sequencing*), o secuenciación masiva, porque amplifican y fragmentan el DNA, para luego generar paralelamente miles de secuencias cortas de cada molécula (Mardis, 2008) serial analysis of gene expression (SAGE). Varias tecnologías de secuenciación masiva han surgido y desaparecido en los últimos 15 años. Illumina es una de las pocas que se ha mantenido. Su primer secuenciador el *Genome Analyzer I* de 2006, producía secuencias o *reads* (lo que “lee” el secuenciador) de 35 nucleótidos de longitud y un total de 300.000 nucleótidos/día. El equipo más reciente de Illumina, el NovaSeq, genera *reads* pareados de 150 nucleótidos y 3 trillones de nucleótidos/día. Para manejar y analizar esa cantidad de secuencias la bioinformática ha desarrollado poderosos algoritmos computacionales (Fonseca *et al.*, 2012).

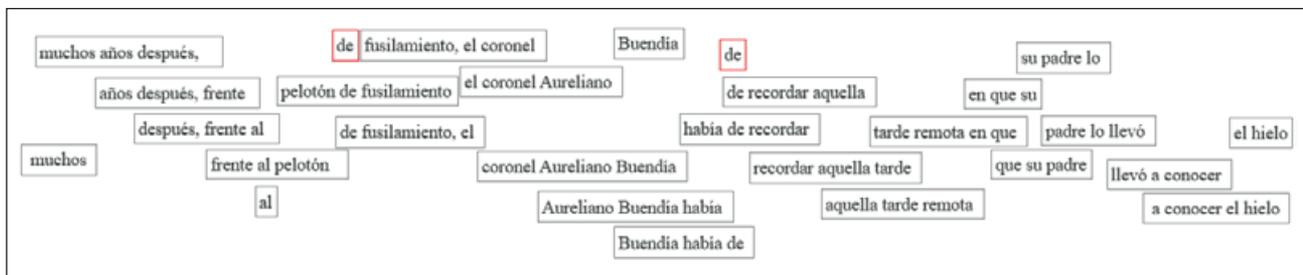
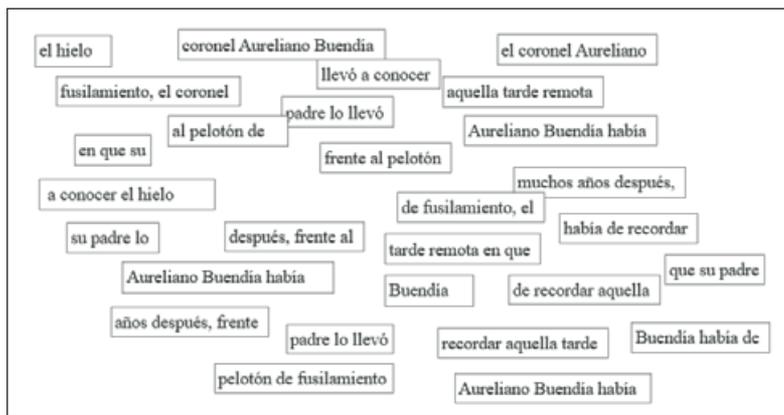
Pero apareció una tercera generación de tecnologías de secuenciación que, a diferencia de las NGS, no requiere amplificación ni fragmentación del DNA. De una única molécula se obtienen *reads* largos de miles de bases de longitud. PacBio, es la principal tecnología de 3ra generación y permite, durante la secuenciación, la detección de grupos metilo en el DNA facilitando los análisis epigenéticos (Flusberg *et al.*, 2010) without bisulfite conversion, through single-molecule, real-time (SMRT). Oxford Nanopore es la tecnología más disruptiva; ha puesto la secuenciación literalmente “al alcance de la mano” del investigador. El pequeño secuenciador MinION que vale US\$ 1.000 y cabe en la mano permite a laboratorios con pocos fondos o en condiciones de campo, secuenciar DNA y observar los resultados en tiempo real, como ocurrió durante la reciente epidemia de Ébola en Liberia (Hoenen *et al.*, 2016).

Estas innovaciones redujeron el costo de la secuenciación. Un millón de nucleótidos pasó de US\$ 10.000 en 2001, a 30 centavos de dólar hoy (NHGRI, 2019). Gracias a esto, se han secuenciado los genomas de decenas de miles de especies de virus y más de 4.000 especies de eucariotas, entre los cuales están 151 especies de aves, 237 de peces y 349 de mamíferos (NCBI, 2019).

Pero secuenciar un genoma animal por primera vez sigue siendo costoso. Ninguna tecnología permite secuenciar ininterrumpidamente un cromosoma animal completo. Es necesario obtener varias copias del genoma, fragmentarlas, secuenciar los fragmentos y ensamblarlos para reconstruir el DNA original. Ensamblar un genoma es como reconstruir un libro completo a partir de párrafos sueltos. Para esto se han desarrollado algoritmos que permiten alinear las secuencias entre sí, detectando aquellas repetidas que, por su longitud diferente, permiten reconstruir la secuencia original (Pevzner, Tang y Waterman, 2001) Este concepto se ilustra de manera simplificada en la Figura 1, usando la frase de un texto literario.

El genoma de la gallina, el primero secuenciado para una especie de interés pecuario, brindó información valiosa para la industria avícola y para las ciencias básicas por la alta tasa de recombinación de los micro-cromosomas de las aves (International Chicken Genome Sequencing Consortium, 2004). En 2005 se publicó el borrador del genoma del perro, junto con los principales polimorfismos de un solo nucleótido

Fragmentos



**Figura 1.** Representación de la idea detrás de los algoritmos de ensamblado de DNA mediante el alineamiento de fragmentos de diferente longitud de la primera frase de “100 años de soledad”. En rojo aparecen fragmentos repetidos que no se pueden posicionar con certeza.

– SNPs (del inglés *Single Nucleotide Polymorphism*), determinantes de diferencias entre razas (Lindblad-Toh *et al.*, 2005). El genoma bovino se ensambló y anotó en 5 años, con la participación de más de 300 personas y un costo de US\$ 53 millones (Elsik *et al.*, 2009). Su secuencia agilizó el desarrollo de nuevas estrategias de mejoramiento para producción lechera y cárnica. El genoma porcino se obtuvo en menos tiempo y a menor costo, gracias a la combinación de secuencias Sanger e Illumina (Archibald *et al.*, 2010). En 2016, un genoma de tilapia con alta contigüidad y resolución de zonas repetitivas se obtuvo mediante PacBio (Conte *et al.*, 2017). Los datos de estos y otros genomas de interés pecuario se resumen en la Tabla 2.

Los datos de estos y otros genomas de interés pecuario se resumen en la Tabla 2.

La combinación de dos tecnologías para secuenciar y ensamblar un genoma permite compensar las limitaciones de una con la otra. Se pueden usar simultáneamente *reads* cortos de Illumina, que tienen una baja tasa de error pero cuyo ensamblaje es complejo, con *reads* largos de Oxford Nanopore o

PacBio que, a pesar de su menor precisión permiten ensamblajes con mayor contigüidad. Una estrategia híbrida se llevará a cabo para secuenciar y ensamblar el genoma del bocachico (*Prochilodus magdalenae*), en un proyecto liderado por el grupo GIPEN de la Piscícola San Silvestre, el grupo de Genómica y Bioinformática de la Universidad de la República de Uruguay y el Centro Nacional de Secuenciación Genómica de la Universidad de Antioquia, en colaboración con el Laboratorio de Biología de Peces (Swedish University of Agricultural Sciences) y el Laboratorio de Genómica marina (BGI University of Chinese Academy of Sciences). En 2 años se espera tener la secuencia del genoma de este emblema de la pesca dulceacuícola colombiana.

Finalmente, la “secuencia cruda” de un genoma solo tiene verdadera utilidad al ser anotada, es decir que se reconozcan las coordenadas de las regiones genómicas, donde comienza y termina cada gen y los sitios de *splicing* de intrones. También es útil la evidencia, como RNA o proteínas, para confirmar los genes identificados mediante estrategias computacionales (Campbell *et al.*, 2014).

**Tabla 2.** Estado de los genomas de especies de interés pecuario. \*Para bovinos y ovinos, a pesar de haber una versión más actualizada del genoma, es frecuente que los investigadores sigan usando una de las versiones previas como referencia.

Especie	Año y tecnología del primer ensamblaje	Genoma de referencia	Última actualización
<i>Gallus gallus</i>	2004, Sanger	GRCg6a	2018
<i>Canis familiaris</i>	2005, Sanger	CanFam3.1	2011
<i>Felis catus</i>	2007, Sanger	Felis_catus_9.0	2017
<i>Bos taurus</i> *	2009, Sanger	ARS-UCD1.2	2018
		UMD_3.1.1	2014
<i>Equus caballus</i>	2009, Sanger	EquCab3.0	2018
<i>Oryctolagus cuniculus</i>	2009, Sanger	OryCun2.0	2016
<i>Ovis aries</i> *	2010, Sanger	Oar_rambouillet_v1.0	2017
		Oar_v4.0	2015
<i>Sus scrofa</i>	2010, Sanger + Illumina	Sscrofa11.1	2017
<i>Pavo</i>	2010, Sanger + 454 + Illumina	Turkey_5.0	2010
<i>Bubalus bubalis</i>	2013, 454 + Illumina	UOA_WB_1	2017
<i>Capra hircus</i>	2013, Illumina + <i>Optical mapping</i>	ARS1	2016
<i>Oreochromis niloticus</i>	2016 PacBio	O_niloticus_UMD_NMBU	2018

## SNPs y GWAS

Aunque nuestro foco de interés es la investigación animal, no podemos ignorar los avances ocasionados por el Proyecto Genoma Humano (Craig Venter et al., 2001; Lander et al., 2001). Tras su culminación en 2003, la atención se dirigió a identificar y mapear variaciones entre individuos y poblaciones para conocer las bases genéticas de la salud y la enfermedad humanas. El Proyecto 1000 Genomas, generó un catálogo de variantes genéticas en individuos sanos de 26 poblaciones de África, Asia, Europa y América (Altshuler et al., 2010). Suramérica estuvo representada por individuos de dos poblaciones: Lima, Perú y Medellín, Colombia. Como resultado se identificaron 88 millones de variantes, la mayoría de ellas SNPs y un número menor de inserciones, deleciones y variantes estructurales (Auton et al., 2015). Esta información ha permitido el desarrollo de microarreglos de SNPs (SNP-arrays o SNP-chips) para la detección de variantes genéticas. Los actuales SNP-arrays para humanos tienen millones de marcadores y son la base del desarrollo de estudios de asociación del genoma completo GWAS (del inglés Genome Wide Association Studies) que han asociado enfermedades o la respuesta a ciertos medicamentos con variantes genómicas (Dedov, 2019).

La salud y producción animal también se benefician de la información genómica para diagnosticar

enfermedades y seleccionar animales. Illumina comercializa SNP-arrays para la mayoría de las especies de interés pecuario con paneles de baja, media y alta densidad. En 2011, se obtuvo el genotipo de 500 caballos y se encontraron 2 regiones o loci genómicos, asociados con hemiplejia laríngea recurrente, la causa más común de insuficiencia de las vías respiratorias altas en equinos (Dupuis et al., 2011). Con un SNP-chip canino de densidad media, se asociaron 33 loci con osteosarcoma heredable en perros (Karlsson et al., 2013) y con uno de alta densidad se asociaron 99 loci a la predisposición a ruptura del ligamento cruzado anterior en la raza Labrador Retriever (Baker et al., 2017). Se han asociado variantes genéticas a la susceptibilidad o resistencia a infecciones bacterianas o infestaciones parasitarias. La región del cromosoma 23 bovino, en la cual se encuentra el gen de la prolil isomerasa 5, fue asociada a susceptibilidad a *Mycobacterium bovis*. Este gen está involucrado en la vía de señalización TNF $\alpha$ /NF $\kappa$ -B, crucial en la respuesta inmune bovina (Richardson et al., 2016). Polimorfismos en el cromosoma 3 ovino, vinculados con genes que también participan en la respuesta inmune, se asociaron a resistencia a parásitos gastrointestinales (Periasamy et al., 2014).

La integración de la tecnología de marcadores genómicos a la selección y mejoramiento tradicional ha aumentado la tasa de progreso genético para rasgos de importancia económica (Koopae y Koshkoiyeh,

2014). Varios proyectos de selección genómica se han implementado para diferentes rasgos en ganado bovino de leche y de carne (Bouquet y Juga, 2013; Pryce et al., 2014; Wiggans et al., 2017). En Uruguay, se aplicó la selección genómica para mejorar la eficiencia de alimentación y calidad de la canal de la raza Hereford, base de la ganadería de carne en Uruguay. Una población de 950 animales, con registros de consumo, eficiencia de conversión e información genealógica, fue genotipada con chips de alta densidad y se obtuvieron EPDs (del inglés Expected Progeny Differences) o diferencias esperadas en la progenie, enriquecidos (Navajas et al., 2014).

### Genómica micro

La genómica permite identificar microorganismos (virus, arqueas, bacterias, hongos, protozoarios) con mayor exactitud. Las altas tasas de mutación y el corto periodo de generación de algunos virus dificultan su estudio y el desarrollo de vacunas, pero al mismo tiempo se pueden usar para reconstruir sus dinámicas temporales y espaciales. La genómica ha permitido estudios de filogenética y filodinámica de virus que impactan la salud animal, dilucidando los factores que inciden en la dinámica evolutiva viral, con lo cual se han actualizado los protocolos de vigilancia epidemiológica (Chernick, Godson y van der Meer, 2014; Roche et al., 2014).

En los últimos 20 años surgieron herramientas para determinar la composición de comunidades microbianas enteras que permitieron el estudio de microorganismos en sus ambientes naturales, ya que los métodos tradicionales de cultivo no permitían capturar toda la diversidad (Zarraonaindia, Smith y Gilbert, 2013). La metagenómica se enfoca en el estudio simultáneo de todas las poblaciones de microorganismos presentes en cualquier matriz (agua, suelo, material biológico, tracto gastro intestinal, piel, etc.). Hay dos estrategias de metagenómica: el metabarcoding que parte de diseñar primers para amplificar por PCR, regiones conservadas y variables de genes específicos (16S para bacterias y arqueas y ITS, 18S y 26S para hongos). Las secuencias amplificadas o amplicones, se secuencian y comparan con secuencias de referencia para identificar las especies. La otra metodología es la metagenómica shotgun, que se basa en secuenciar todo el DNA presente en una muestra para identificar genes y especies.

Entre las aplicaciones más interesantes de la metagenómica está el estudio de los microorganismos del rumen, los cuales digieren la celulosa y proveen al animal con ácidos grasos de cadena corta. La metagenómica permite conocer mejor el ecosistema ruminal, determinar los factores que lo afectan para optimizar su eficiencia y funcionamiento y por ende, incrementar la producción de carne, leche y lana (Li, 2015; Morgavi, Kelly, Janssen, & Attwood, 2013; Puniya, Singh, & Kamra, 2015; Singh et al., 2012; Kim, Park, & Yu, 2017).

### Transcriptomas

Las aplicaciones mencionadas hasta el momento se basan en los aspectos estructurales de la secuencia de DNA. Pero una gran rama de la genómica se dedica a medir la expresión, o más exactamente la transcripción de los genes. La transcriptómica estudia simultáneamente los genes que se transcriben en un tejido específico, bajo condiciones particulares o durante una etapa del desarrollo o una condición fisiológica permitiendo hacer comparaciones y sirviendo de puente entre el estudio del genoma y de su función.

Los estudios de transcriptómica florecieron a principios de este siglo, con los microarreglos de expresión: conjuntos de sondas de DNA complementario (cDNA), o de oligonucleótidos, unidos a una superficie sólida (Tenenbaum et al., 2000). Estos microarreglos fueron muy usados en la primera década de este siglo y permitieron evaluar diferencias en la expresión global de diferentes tejidos. Gracias a ellos se detectaron diferencias en el perfil transcripcional, responsables de las fallas en el desarrollo de embriones obtenidos por clonación (Rodríguez-Osorio et al., 2009).

Pero los microarreglos tienen una limitación: no permiten detectar transcriptos nuevos, sólo capturan aquellos para los cuales se tienen secuencias complementarias. El RNA-Seq supera esta restricción y permite descubrir nuevos genes, detectar diferencias en el uso de exones o splicing diferencial y descubrir RNAs que no codifican para proteínas y microRNAs. A partir del RNA de interés se obtienen librerías de cDNA, las cuales se secuencian y los reads obtenidos se mapean al genoma de referencia para determinar el origen de los transcriptos (Lister et al., 2008; Nagalakshmi et al., 2008). Los algoritmos iniciales de

mapeo de RNA-Seq no permitían identificar el splicing diferencial. Los recientes algoritmos de alineamiento empalmado (spliced aligners) hacen posible analizar variantes del mismo gen, identificando la unión entre exones alternos (Trapnell, Pachter y Salzberg, 2009; Dobin et al., 2013; Kim, Langmead y Salzberg, 2015). Finalmente, se establece el perfil transcripcional de las muestras cuantificando los reads que corresponden a cada exón.

El diseño experimental y análisis estadístico de expresión diferencial en RNA-Seq se basa en los métodos creados para análisis de microarreglos, es decir, se generan tablas de conteo de los transcritos mapeados y se someten a pruebas estadísticas con correcciones para datos múltiples. Paquetes en R, como DESeq2 (Love, Anders y Huber, 2014), EdgeR (Robinson, McCarthy y Smyth, 2009) y Limma/Voom (Law, Chen, Shi, & Smyth, 2014), han facilitado el análisis de expresión diferencial en RNA-Seq.

Las metodologías de secuenciación requieren cantidades de RNA cada vez menores, con lo cual se ha abierto la posibilidad para el estudio del transcriptoma de células individuales scRNA-Seq (del inglés single cell). Las metodologías para la generación de muestras y análisis de datos de scRNA-Seq se encuentran en construcción y es posible que se modifiquen; aun así esta es una de las aplicaciones más prometedoras y poderosas de la genómica (Bacher y Kendziorski, 2016; Poirion et al., 2016).

### Reflexiones finales

La investigación actual en salud y producción animal requiere tanto del entendimiento de la biología de los genomas y su interacción con el medio, como del manejo de las nuevas herramientas genómicas. La producción pecuaria hoy exige a los investigadores fluidez en el nuevo lenguaje de la genómica y trabajo colaborativo para una forma de producción que responda a los desafíos actuales y haga uso de tecnologías de punta para aumentar la producción, incrementar la resistencia a enfermedades y disminuir los insumos.

La genómica y la bioinformática se han basado, en gran medida, en software de acceso abierto y en la disponibilidad pública de los datos. El uso que hagamos hoy de estas poderosas herramientas, al alcance de todos, tendrá un gran impacto en la salud y producción animal, en la sociedad y en el ambiente.

Se aplican las palabras de Stan Lee, a través del hombre araña: “Un gran poder conlleva una gran responsabilidad”.

### Referencias

- (NHGRI), N. H. G. R. I. (2019) *DNA Sequencing Costs: Data*. Disponible en: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>.
- Altschul, S. F. *et al.* (1990) «Basic local alignment search tool», *Journal of Molecular Biology*. doi: 10.1016/S0022-2836(05)80360-2.
- Altshuler, D. L. *et al.* (2010) «A map of human genome variation from population-scale sequencing», *Nature*. doi: 10.1038/nature09534.
- Archibald, A. L. *et al.* (2010) «Pig genome sequence - analysis and publication strategy», *BMC Genomics*. doi: 10.1186/1471-2164-11-438.
- Auton, A. *et al.* (2015) «A global reference for human genetic variation», *Nature*. doi: 10.1038/nature15393.
- Bacher, R. y Kendziorski, C. (2016) «Design and computational analysis of single-cell RNA-sequencing experiments», *Genome Biology*. doi: 10.1186/s13059-016-0927-y.
- Baker, L. A. *et al.* (2017) «Genome-wide association analysis in dogs implicates 99 loci as risk variants for anterior cruciate ligament rupture», *PLoS ONE*. doi: 10.1371/journal.pone.0173810.
- Bouquet, A. y Juga, J. (2013) «Integrating genomic selection into dairy cattle breeding programmes: A review», *Animal*. doi: 10.1017/S1751731112002248.
- Campbell, M. S. *et al.* (2014) «Genome Annotation and Curation Using MAKER and MAKER-P», *Current Protocols in Bioinformatics*. doi: 10.1002/0471250953.bi0411s48.
- Chernick, A., Godson, D. L. y van der Meer, F. (2014) «Metadata beyond the sequence enables the phylodynamic inference of bovine viral diarrhoea virus type 1a isolates from Western Canada», *Infection, Genetics and Evolution*. doi: 10.1016/j.meegid.2014.01.003.
- Conte, M. A. *et al.* (2017) «A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions», *BMC Genomics*. doi: 10.1186/s12864-017-3723-5.
- Costa-Silva, J., Domingues, D. y Lopes, F. M. (2017) «RNA-Seq differential expression analysis: An extended review and a software tool», *PLoS ONE*. doi: 10.1371/journal.pone.0190152.
- Craig Venter, J. *et al.* (2001) «The sequence of the human genome», *Science*. doi: 10.1126/science.1058040.
- Dedov, I. I. (2019) «Personalized medicine», *Vestnik Rossiiskoi Akademii Meditsinskikh Nauk*. doi: 10.15690/vramn1108.
- Dobin, A. *et al.* (2013) «STAR: Ultrafast universal RNA-seq aligner», *Bioinformatics*. doi: 10.1093/bioinformatics/bts635.

- Dupuis, M. C. *et al.* (2011) «Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse», *Mammalian Genome*. doi: 10.1007/s00335-011-9337-3.
- Elsik, C. G. *et al.* (2009) «The genome sequence of taurine cattle: A window to ruminant biology and evolution», *Science*. doi: 10.1126/science.1169588.
- Flusberg, B. A. *et al.* (2010) «Direct detection of DNA methylation during single-molecule, real-time sequencing», *Nature Methods*. doi: 10.1038/nmeth.1459.
- Fonseca, N. A. *et al.* (2012) «Tools for mapping high-throughput sequencing data», *Bioinformatics*. doi: 10.1093/bioinformatics/bts605.
- Hoenen, T. *et al.* (2016) «Nanopore sequencing as a rapidly deployable Ebola outbreak tool», *Emerging Infectious Diseases*. doi: 10.3201/eid2202.151796.
- Illumina (2016) «BovineSNP50 Genotyping BeadChip», [Http://Www.illumina.com/Documents/Products/Datasheets/Datasheet\\_Bovine\\_Snp50.Pdf](http://www.illumina.com/Documents/Products/Datasheets/Datasheet_Bovine_Snp50.Pdf).
- International Chicken Genome Sequencing Consortium (2004) «Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution», *Nature*. doi: 10.1038/nature03154.
- Karlsson, E. K. *et al.* (2013) «Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B», *Genome Biology*. doi: 10.1186/gb-2013-14-12-r132.
- Kim, D., Langmead, B. y Salzberg, S. L. (2015) «HISAT: A fast spliced aligner with low memory requirements», *Nature Methods*. doi: 10.1038/nmeth.3317.
- Kim, M., Park, T. y Yu, Z. (2017) «Metagenomic investigation of gastrointestinal microbiome in cattle», *Asian-Australasian Journal of Animal Sciences*. doi: 10.5713/ajas.17.0544.
- Koopae, H. K. y Koshkoiyeh, A. E. (2014) «SNPs genotyping technologies and their applications in farm animals breeding Programs: Review», *Brazilian Archives of Biology and Technology*. doi: 10.1590/S1516-89132014000100013.
- Lander, E. S. *et al.* (2001) «Initial sequencing and analysis of the human genome», *Nature*. doi: 10.1038/35057062.
- Law, C. W. *et al.* (2014) «Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts», *Genome Biology*. doi: 10.1186/gb-2014-15-2-r29.
- Li, R. W. (2015) «Rumen metagenomics», en *Rumen Microbiology: From Evolution to Revolution*. doi: 10.1007/978-81-322-2401-3\_16.
- Lindblad-Toh, K. *et al.* (2005) «Genome sequence, comparative analysis and haplotype structure of the domestic dog», *Nature*. doi: 10.1038/nature04338.
- Lister, R. *et al.* (2008) «Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis», *Cell*. doi: 10.1016/j.cell.2008.03.029.
- Love, M. I., Anders, S. y Huber, W. (2014) *Differential analysis of count data - the DESeq2 package*, *Genome Biology*. doi: 110.1186/s13059-014-0550-8.
- Mardis, E. R. (2008) «Next-Generation DNA Sequencing Methods», *Annual Review of Genomics and Human Genetics*. doi: 10.1146/annurev.genom.9.081307.164359.
- Masic, I. (2016) «The most influential scientists in the development of medical informatics (13): Margaret Belle Dayhoff», *Acta Informatica Medica*. doi: 10.5455/aim.2016.24.299-299.
- McKusick, V. A. y Ruddle, F. H. (1987) «A new discipline, a new name, a new journal», *Genomics*. doi: 10.1016/0888-7543(87)90098-X.
- Morgavi, D. P. *et al.* (2013) «Rumen microbial (meta)genomics and its application to ruminant production», en *Animal*. doi: 10.1017/S1751731112000419.
- Nagalakshmi, U. *et al.* (2008) «The transcriptional landscape of the yeast genome defined by RNA sequencing», *Science*. doi: 10.1126/science.1158441.
- Navajas, E. *et al.* (2014) «Genetic improvement of feed efficiency and carcass and meat quality of hereford cattle by genomics.», en *60th International Congress of Meat Science and Technology*.
- NCBI (2019) *NCBI genome database, Ncbi*. Disponible en: <https://www.ncbi.nlm.nih.gov/genome/> (Accedido: 20 de agosto de 2019).
- Pearson, W. R. y Lipman, D. J. (1988) «Improved tools for biological sequence comparison.», *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.85.8.2444.
- Periasamy, K. *et al.* (2014) «Candidate gene approach for parasite resistance in sheep - Variation in immune pathway genes and association with fecal egg count», *PLoS ONE*. doi: 10.1371/journal.pone.0088337.
- Pevzner, P. A., Tang, H. y Waterman, M. S. (2001) «An Eulerian path approach to DNA fragment assembly», *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.171285098.
- Poirion, O. B. *et al.* (2016) «Single-cell transcriptomics bioinformatics and computational challenges», *Frontiers in Genetics*. doi: 10.3389/fgene.2016.00163.
- Pryce, J. E. *et al.* (2014) «Genomic selection for feed efficiency in dairy cattle», *Animal*. doi: 10.1017/S1751731113001687.
- Puniya, A. K., Singh, R. y Kamra, D. N. (2015) *Rumen microbiology: From evolution to revolution*, *Rumen Microbiology: From Evolution to Revolution*. doi: 10.1007/978-81-322-2401-3.
- Richardson, I. W. *et al.* (2016) «A genome-wide association study for genetic susceptibility to Mycobacterium bovis infection in dairy cattle identifies a susceptibility QTL on chromosome 23», *Genetics Selection Evolution*. doi: 10.1186/s12711-016-0197-x.

- Robinson, M. D., McCarthy, D. J. y Smyth, G. K. (2009) «edgeR: A Bioconductor package for differential expression analysis of digital gene expression data», *Bioinformatics*. doi: 10.1093/bioinformatics/btp616.
- Roche, B. *et al.* (2014) «Adaptive Evolution and Environmental Durability Jointly Structure Phylogenetic Patterns in Avian Influenza Viruses», *PLoS Biology*. doi: 10.1371/journal.pbio.1001931.
- Rodriguez-Osorio, N. *et al.* (2009) «Transcriptional reprogramming of gene expression in bovine somatic cell chromatin transfer embryos», *BMC Genomics*, 10. doi: 10.1186/1471-2164-10-190.
- Sanger, F., Nicklen, S. y Coulson, A. R. (1977) «DNA sequencing with chain-terminating inhibitors.», *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.74.12.5463.
- Singh, K. M. *et al.* (2012) «Metagenomic analysis of Surti buffalo (*Bubalus bubalis*) rumen: A preliminary study», *Molecular Biology Reports*. doi: 10.1007/s11033-011-1278-0.
- Staden, R. (1977) «Sequence data handling by computer», *Nucleic Acids Research*. doi: 10.1093/nar/4.11.4037.
- Tenenbaum, S. A. *et al.* (2000) «Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays», *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.97.26.14085.
- Trapnell, C., Pachter, L. y Salzberg, S. L. (2009) «TopHat: Discovering splice junctions with RNA-Seq», *Bioinformatics*. doi: 10.1093/bioinformatics/btp120.
- Wiggans, G. R. *et al.* (2017) «Genomic Selection in Dairy Cattle: The USDA Experience», *Annual Review of Animal Biosciences*. doi: 10.1146/annurev-animal-021815-111422.
- Zarraonaindia, I., Smith, D. P. y Gilbert, J. A. (2013) «Beyond the genome: Community-level analysis of the microbial world», *Biology and Philosophy*. doi: 10.1007/s10539-012-9357-8.