

GENOMAS PEQUEÑOS, UNA MIRADA A LA BIOLOGÍA SINTÉTICA

SMALL GENOMES, A SCOPE TO SYNTHETIC BIOLOGY

María Fernanda Chávez Jacobo (1)
Marina Casilda Guadalupe Osorio (1)
Falconi
Rigel Quintero Luis (1)
María Elena Cobos-Justo (1)
Norma Elena Rojas-Ruiz (2)
Patricia Sánchez-Alonso (2)
Erasmus Negrete-Abascal (3)
Candelario Vázquez-Cruz* (2)

<https://orcid.org/0009-0001-1609-7949>
<https://orcid.org/0009-0007-6704-8470>
<https://orcid.org/0009-0001-5824-2710>
<https://orcid.org/0009-0004-9836-1644>
<https://orcid.org/0000-0002-0559-7059>
<https://orcid.org/0000-0001-6437-8078>
<https://orcid.org/0000-0002-9313-3659>
<https://orcid.org/0000-0003-4362-8044>
<https://orcid.org/0000-0002-9105-5669>

- (1) Posgrado en Microbiología ICUAP
(2) Centro de Investigaciones en Ciencias Microbiológicas, ICUAP,
BUAP
(3) Carrera de Biología, FES-Iztacala, UNAM. Av. de Los Barrios # 1,
Los Reyes Iztacala, Tlalnepantla, Edo de México, 54090, México.

candelario.vazquez@correo.buap.mx
p22mcm0011@viep.com.mx
helenis_239@hotmail.com
norma.rojasr@correo.buap.mx,
maria.sanchez@correo.buap.mx
negrete@yahoo.com

ISSN 2448-5829

Año 10, No. 30, 2024

RD-ICUAP

Resumen

En biología la clasificación es un método de trabajo esencial y sistemático que permite organizar a los seres vivos, en base a sus relaciones filogenéticas y evolutivas, que va de un nivel superior a grupos subordinados. Para ello se sustituye el nombre coloquial del espécimen por una nomenclatura reconocida internacionalmente. Esta nomenclatura binomial contiene dos nombres denominados género y especie. Para todos los organismos pluricelulares y unicelulares naturales la nomenclatura es aplicada y reconocida por asociaciones facultadas en taxonomía. Sin embargo, a mediados del siglo XX con el nacimiento y desarrollo de la biología molecular se desarrolló un despegue significativo en el estudio de organismos unicelulares procariotas que llevaron al desarrollo de herramientas de edición genética permitiendo la aparición de organismos genéticamente modificados, produciéndose nuevos linajes celulares que tienen reconocimiento por la utilidad a favor de la salud humana. A partir de bases de datos es posible obtener información para realizar distintos trabajos de investigación; desplegarla en forma sistemática y entendible es posible por medio de herramientas bioinformáticas, sitios de internet con programas relativamente accesibles, o en otros casos con programas que requieren descargarse y ejecutarse con instrucciones especiales. Un ejercicio de naturaleza bioinformática se muestra en este escrito, al explorar algunos genomas de *Mycoplasma mycoides* GM12 que estaban identificados con el mismo nombre de cepa, en el que se observó inconsistencia en la sistemática biológica y molecular: esto probablemente se deba a la falta de actualización automática de las bases de datos y falta de personal con perfil bioinformático.

Palabras clave: Genomas pequeños, OGM, Organismos sintéticos, Taxonomía, *Mycoplasma mycoides*, Genómica-Bioinformática.

Abstract

In biology, classification is an essential and systematic work method that allows living beings to be organized, based on their phylogenetic and evolutionary relationships, in a hierarchical way, ranging from a higher level to subordinate groups. This binomial nomenclature contains two words called genus and species. The specific terms are applied and recognized by associations empowered in taxonomy. However, in the middle of the 20th century, with the birth and development of molecular biology, scientists developed a significant takeoff in the study of prokaryotic unicellular organisms that led to the development of gene editing tools allowing the appearance of genetically modified organisms at a practical level, producing new cell and molecular lineages that are widely recognized for their usefulness in favor of human health. Extracting data from these databases and deploying it in a systematic, complete, consistent, and understandable way is possible through bioinformatics tools, consisting of computers, online sites with relatively accessible programs, or programs that require downloading and running with special instructions. An exercise of a particular bioinformatics nature is shown in this paper when exploring some genomes of *Mycoplasma mycoides* GM12 that contain the same strain name, in which inconsistency was observed in the biological and molecular systematics: this is probably due to the lack of automatic updating of the databases.

Keywords: Small genome, GMO, Synthetic organisms, Taxonomy, *Mycoplasma mycoides*, Genomics-Bioinformatics.

Un breve paseo por los orígenes de la biología sintética.

La biología es la ciencia encargada del estudio de la vida cuyo principio se remonta a los inicios de la humanidad misma. Lo que comenzó como un esfuerzo por clasificar el entorno circundante rápidamente fue adquiriendo un carácter sistemático con la contribución de Carl Linnaeus (en Suecia) por su afamada nomenclatura binomial en 1753. Entre 1674-1676 se desarrolló otro hito con la exploración de las formas microscópicas de vida, hoy conocidas como protozoarios, y atribuido a Anton van Leeuwenhoek (Delft, en Países Bajos), un comerciante de telas que colocó a la vida bajo su lente al desarrollar los primeros microscopios, con ellos observó una gran variedad de microorganismos. Años más tarde, en 1852, Rudolf Virchow enuncia la frase: "omnis cellula e cellula", es decir que toda célula proviene de otra célula. Alimentando el desarrollo de la teoría celular de Theodor Schwann y Jakob Schleiden quienes plantearon postulados clave que son: "Todos los seres vivos están formados por células", "La unidad estructural y funcional de la vida es la célula" y "Toda célula procede de otra célula semejante" (Figura 1). Esta es uno de los pilares de la biología junto con la teoría de la herencia desarrollada por Gregor Mendel; que muestra la transmisión de características de padres a hijos mediante experimentos de cultivos mezclados (hibridación) de chícharos. En esta se establecen los principios genéticos de la uniformidad, la segregación y la transmisión independiente de los alelos (Figura 2). Una de las teorías más relevantes en este campo es la de la evolución, propuesta por Charles Darwin, en ella se plantea que las especies cambian a lo largo del tiempo. Las nuevas especies provienen de otras ya existentes, comparten un ancestro común y estos cambios crean una amplia gama de relaciones entre varios tipos de organismos. En esta teoría también se propone que todos los organismos están sujetos a la selección natural, que evalúa su capacidad de supervivencia y reproducción ante cierta condición o característica ambiental desafiante (Camberly y Reece, 2007, Schne et al. 2008).



Figura 1. Estudio detallado de un organismo o de objetos de origen biológico. Los muebles que están a nuestro alcance tienen su origen en organismos vivos, como los árboles; la mesa de madera del lado izquierdo muestra las líneas de diferentes tejidos longitudinales del vegetal que conducen agua y nutrientes desde la raíz hasta sus hojas. Los círculos concéntricos muestran el crecimiento anual del tejido celular vivo durante el desarrollo de un árbol. En la imagen del lado extremo derecho se observan pequeños orificios que corresponden al tejido celular; su visualización es posible por la acumulación de celulosa, producto de la polimerización de la glucosa alrededor de las células vegetales. Pocas células tienen tamaño grande por lo que es necesario usar el microscopio para verlas. Imágenes-fuente propia de los autores.

Existen varias ramas biológicas que estudian aspectos específicos de los seres vivos. Entre estas se encuentra la biología molecular, cuyo crecimiento ha sido explosivo y reciente. Ésta se encarga de estudiar la estructura de las macromoléculas (ADN, ARN y proteínas) y sus interacciones en los procesos vitales de las células, así como conocer la estructura, evolución y función de los genes y genomas. Su desarrollo involucró el conocimiento de aspectos básicos, como la estructura de los ácidos nucleicos. Dentro de los estudios más destacados, se encuentran los experimentos de Frederick Griffith en 1928, quien trabajó con unos microorganismos llamados neumococos y documentó que las células pueden recibir material genético ajeno. En estos experimentos los neumococos de aspecto rugoso se convertían en lisos, capaces de producir neumonía al encontrarse en un medio que contenía bacterias lisas muertas, con lo que planteó el "principio transformante" de las células. Ante este hecho, en 1944 Avery junto con Macleod y McCarty demostraron que este principio era el ADN y que el material genético convertía a las bacterias en productoras de polisacáridos capsulares y por ello tomaban el aspecto liso. Lo anterior se hizo indiscutible gracias al estudio de Félix d' Herell, quien describió cómo las partículas virales conocidas como bacteriófagos infectan a las bacterias. Esto ocurrió en el seno del grupo

del fago liderado por Max Delbruck y Salvador Luria que estudiaron cómo se multiplicaban los virus infectantes de la bacteria *Escherichia coli*, virus que están compuestos por DNA y algunas proteínas. Con estos minicomponentes vitales y virales se inició el conocimiento de la replicación del material genético, la descripción de genes, su estructura y la red de regulación que conduce a su expresión, manifestación, producción de proteínas o mRNA (Krebs et al. 2017). En estos grupos de trabajo se estableció un proceso lógico y ordenado para distinguir entre linajes de microorganismos, conocer la ruta de evolución in vitro (confinado en un laboratorio) de las células infectadas y de los virus mismos, para ello todos los especímenes de la prole se nombraron apegándose a una nomenclatura semejante a la que propuso Carl Nilsson Linnaeus (Karp, 2008).



Figura 2. Estudio o caracterización de organismos microscópicos. Muchos organismos vivos son microscópicos por lo que es necesario utilizar el microscopio como herramienta en un laboratorio para reconocerlos celularmente. Así podemos observar al hongo que crece en un trozo de pan, como *Penicillium* o *Rhizopus* (fotografía de la izquierda). Con ayuda del microscopio podemos observar y distinguir visualmente los glóbulos rojos de la sangre humana o bien una levadura de vida libre (fotografías del lado derecho). Imágenes-fuente propia de los autores.

Inicios de la biología molecular y el diseño genético de los organismos.

En los años 50's un gran paso para la biología molecular fue el descubrimiento de la estructura de la doble hélice del ADN por James Watson y Francis Crick, que llevó a conocer que todos los organismos y su material genético se encuentran constituidos por una sopa de letras única, compuesta por las bases nitrogenadas adenina(A), guanina(G), citosina(C) y timina(T), que se polimerizan como desoxirribonucleótidos en un

orden específico, salvo los retrovirus que contienen uracilo(U) en lugar de timina, y son polimerizados como ribonucleótidos. Formando un intrincado rompecabezas que puede descifrarse a partir del código genético, que permite conocer el proceso de elaboración de las proteínas y las enzimas (catalizadores biológicos). Asimismo, otro hecho destacado de la época fue el descubrimiento de las enzimas capaces de destruir, editar o reconstruir el material genético; entre las que destacan las endonucleasas de restricción tipo II como *HindIII* y las DNA ligasas, utilizadas para construir los primeros genomas quiméricos (artificiales) del virus SV40, de fago lambda y de material genético accesorio, como los plásmidos ColE1; éstos últimos hasta la fecha son vectores de clonación muy versátiles y de manejo accesible. Entre los investigadores connotados de estos estudios tenemos a Maclyn McCarty que contribuyó a purificar los bacteriófagos, Matthew Stanley Meselson quien sumó al conocimiento de la replicación del DNA y otros aspectos de la recombinación del material genético. Werner Arber, Daniel Nathans y Hamilton O. Smith usaron las enzimas de restricción y construyeron mapas físicos y genéticos del virus símico SV40, Paul Naim Berg hizo los primeros empalmes de diferentes fragmentos de DNA y Stanley Norman Cohen que además de construir el plásmido pSC101 demostró que el gen ribosomal de la rana podía ser funcional en una bacteria. La utilidad práctica del diseño molecular radica en una estrategia de trabajo que se guía en los cambios fenotípicos contrastantes de las células utilizadas en un proyecto, porque cambian de comportamiento fisiológico o metabólico, debido al uso de marcadores de selección contenidos en las moléculas "vector", entre los que destacan los genes de resistencia a los antibióticos tetraciclina, ampicilina o neomicina, pues fácilmente se evalúa la sensibilidad o la resistencia al antibiótico. Otros marcadores de selección se evalúan por cambio de color, por ejemplo, en el caso de ausencia (color blanco) o presencia (color azul) de las enzimas b-galactosidasa o b-glucuronidasa, las cuales actúan degradando sustancias específicas cromogénicas ("que generan color").

Recopilación de información molecular de los organismos en las bases de datos biológicos.

El modelo de Watson y Crick alcanzó otra dimensión cuando estuvieron disponibles para todos los investigadores los métodos y los aparatos para conocer la secuencia de aminoácidos de las proteínas y de bases de los ácidos nucleicos. Así se llegó a conocer la secuencia del genoma humano, compuesto de 23 pares de cromosomas (46 en total) y 3,299.5 millones de bases considerando el contenido haploide (unidas como nucleósidos fosfato). Toda la información de este y otros 736895 genomas está depositada en bancos de datos, por ejemplo, el emblemático GenBank de los Institutos de Salud de los Estados Unidos de Norteamérica. Además, en esta base de datos podemos encontrar mucha más información en 37 secciones fácilmente desplegadas, que van desde reportes o literatura científica, hasta detalles de la secuencia de genes específicos de organismos como el ser humano, animales, plantas, microbios, orgánulos, virus y plásmidos. Es conveniente saber que algunas biomoléculas, como los vectores de clonación, virus o plásmidos, forman parte de la base de datos de GenBank, aunque, estas biomoléculas son quimeras elaboradas *in vitro*, para su fabricación se utilizaron enzimas de restricción y DNA ligasa.

Origen y alcance de los organismos genéticamente modificados.

Con un espíritu más creativo y objetivos claros para atender problemas de salud humana, los vectores utilizados como una herramienta molecular dieron lugar a nuevos genomas, uno de ellos correspondió a las bacterias *E. coli* recombinantes que permitieron la producción de somatotropina coriónica humana u hormona humana de crecimiento y la insulina, que vinieron a aliviar importantes problemas de salud humana como el trastorno de falta de crecimiento o la diabetes. Se pueden ca-

lugar a estos microorganismos como los primeros organismos genéticamente modificados ("OGM"). Estos diseños genéticos exitosos convirtieron a los vectores de clonación o vehículos moleculares en aceleradores del conocimiento molecular (Figura 3). El siguiente nivel "figurado de aceleración molecular" se alcanzó con el diseño de la PCR - reacción en cadena de la polimerasa - porque dejaron de ser necesarios los bancos genómicos o de cDNA que daban información de los RNA mensajeros existentes en un organismo y del probable mecanismo de regulación. La PCR ha sido modificada de distintas formas que han llevado a imaginar un nuevo escenario para la construcción de nuevos organismos. Sin embargo, la ciencia continuó avanzando, en el 2005 Francis Mojica (Juan Francisco Martínez Mojica) halló la inmunidad molecular de las bacterias hacia los bacteriófagos, llamada CRISPR-Cas (Mojica *et al.*, 2005), y más tarde su aplicación fue desarrollada por Emmanuelle Charpentier y Jennifer Doudna (Jinek, *et al.*, 2012), quienes propusieron a CRISPR-Cas9 como la tijera molecular más precisa para eliminar información en un locus y colocar en sitios específicos de los genomas, nuevos genes a elección del estudio científico y su beneficio práctico, por ejemplo, en el cuidado de la salud de organismos eucarióticos incluidos el hombre u otras aplicaciones biotecnológicas.

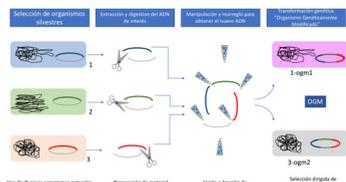


Figura 3. Ruta descriptiva para la obtención de organismos unicelulares modificados en su información genética. La mayoría de la información genética sintética tiene origen en organismos naturales o silvestres. De allí se extraen secuencias o genes que se recortan de forma específica y se unen a otras secuencias naturales que son rediseñadas para su fácil manejo en un laboratorio; las primeras reciben el nombre de ADN pasajero y las segundas de vector de clonación. La mezcla y unión de fragmentos de ADN genera nuevas moléculas de ADN que se recuperan en células vivas que manifiestan cambios en su comportamiento típico como consecuencia de los genes recientemente recibidos. Las células que contienen las nuevas construcciones genéticas se conocen como organismos genéticamente modificados.

Diversidad de los OGM- Organismos genéticamente modificados.

Todos los organismos a lo largo de la evolución pueden ir adquiriendo modificaciones genéticas, producto de la herencia de los parentales, a consecuencia de su crecimiento, de su actividad metabólica y por influencia del medio ambiente. Debido a la necesidad humana se hace investigación para generar varios OGM, con posible beneficio del medio ambiente, la salud o la alimentación. Así se han atendido importantes y diversas temáticas, por ejemplo, la remoción de iones tóxicos, la destrucción de contaminantes derivados de la industria de la transformación, la producción de hormonas, vitaminas y antibióticos, la resistencia de las plantas a las enfermedades y las plagas, y la resistencia a las crecientes condiciones medioambientales que deterioran el cultivo de los diversos vegetales. Generalmente los organismos nativos y los genéticamente modificados cumplen con especificaciones en sus nombres, en sus claves de identificación, en sus registros de las bases de datos o en los registros de patentes; estos acuerdos de anotaciones

se han hecho para protegerlos o limitar su uso, y son de gran utilidad para distinguir o diferenciar todas las formas de vida conocidas por el ser humano. Una muestra de nombres y especificaciones que identifican a algunos organismos está contenida en la tabla 1, aunque, también se incluyen algunas moléculas de DNA silvestres para su fácil comparación. La tabla 1 muestra cepas del reciente coronavirus SARS-Cov2, dos cepas de *Pseudomonas* de muy reciente registro en GenBank y las bacterias OGM iniciales productoras de insulina y hormona humana de crecimiento. Hay mucho avance en el desarrollo de OGM, la tabla 1 muestra algunos ejemplos representativos, de material genético recombinante, como las proteínas arriba mencionadas, y moléculas de DNA quimeras como plásmidos o vectores virales. Como ejemplos de OGM de eucariontes se pueden mencionar varias levaduras, insectos como la mosca de la fruta o los mosquitos, las plantas de tabaco, algodón o soya, y animales como ratones y ovejas. Muchos de los ejemplos mencionados han proporcionado valiosos productos para el ser humano, el más reciente es la vacuna para controlar la pandemia de Covid19.

Tabla 1. Características básicas necesarias para registrar o distinguir los organismos

Nombre Género especie	Clave y tamaño de genoma	Fenotipo	Año de registro	Autor o fuente
<i>Escherichia coli</i>	K12; 4641652 bp	Cepa nativa de laboratorio	1997	Science 277 (5331), 1453-1462
<i>Escherichia coli</i>	O157:H7 cepa Sakai; 5498578 bp	Cepa patogénica	1999	Genes Genet. Syst. 74 (5), 227-239 (1999)
<i>Bacillus subtilis</i>	Cepa 168; 4215606 bp	Cepa de tipo industrial no patogénica	1997	Nature 390 (6657), 249-256
<i>Haemophilus influenzae</i>	Rd KW20; 1830138 bp	Cepa patogénica	1995	Science 269 (5223), 496-512

Nombre Género especie	Clave y tamaño de genoma	Fenotipo	Año de registro	Autor o fuente
<i>Plásmido sintético</i>	LC129268; 2808 pb	Construcción sintética, derivada de PUC18	2016	GenBank: LC129268.1
<i>Vector plasmídico</i>	pUC18; 2686 pb	pUC12, M13mp18, bGal18	17-Dic1986	GenBank: L08752.1
<i>Plásmido</i>	pOV; 13551 pb	Nativo con resistencia a estreptomycin y sulfas	2020	GenBank: JX827416.1
<i>pCV5</i>	Derivado de pCV4 y de pGEM3Zf (+)-cat 8.6 kb	pCV4 amyE, Catr, Emr, Ampr	1997	J. Bacteriology, 179 (20):6341-6348
<i>SARS-Cov2</i>	NC_045512 29903 bp ss-RNA	Cepa patogénica Wuhan-Hu-1 China, Dec1999	12-01-2020	Nature 579 (7798), 265-269
<i>SARS-Cov2</i>	MT192765 29829 bp RNA linear BioProject: PRJNA612578 Con link a 6993 secuencias de DNA y 6.032 Gb de información	Cepa patogénica The Scripps Research Institute	20-05-2020	GenBank: MT192765.1
<i>Methamycoplasma hominis</i>	Cepa ATCC23114; 665445 pb	Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE	12-02-2009	PLoS Genet 5 (10), e1000677
<i>Mycoplasma mycoides susp. capri</i>	Cepa GM12; 1.08 Gb	The J. Craig Venter Institute, EUA	14-05-2009	Science 325 (5948), 1693-1696
<i>Pseudomonas aeruginosa</i>	Cepa Emara01 6.44 Mb	MDR, de trasplante de hígado, Egipto	11-05-2023	GenBank: JARQZF000 000000.1

Nombre Género especie	Clave y tamaño de genoma	Fenotipo	Año de registro	Autor o fuente
<i>Pseudomonas aeruginosa</i>	Cepa 2-223, 7.07 Mb	MDR, de alimento para animales, China	11-05-2023	GenBank: JASAOW 00000 0000.1
<i>Escherichia coli</i>	HB101 (PBR322: trp:HGh-gen)	Universidad de Osaka, Japón	Oct- 1984	PNAS 81: 5956 - 5960
<i>Escherichia coli</i>	K-12 cepa 294 (pIB1) (PBR322: lac-Insulin-gen)	Genetech y City of Hope National Medical center, EUA	Enero- 1979	PNAS 76: 106-110
<i>Escherichia coli</i>	DH5α (pVK100: Lys-Insulin-gen)	Universidad de Pekin, China	1995	Appl. Biochem. Biotech. 55:5-15

Datos recuperados de varias fuentes electrónicas de las que se indican sus referencias-edición propia de los autores.

Bases de datos para catalogar y organizar la información de los diferentes organismos.

La información relevante de cualquier organismo, gen o fracciones de sus genomas está respaldada por acuerdos de varios organismos gubernamentales nacionales e internacionales, porque es de utilidad social y humana. Muchos reservorios o bases de datos tienen un carácter recopilador, pero mantienen la accesibilidad para que los expertos o interesados en la información biológica específica puedan consultarla. Así podemos citar al sitio de NCBI y GenBank de los Institutos de Salud de los Estados Unidos de Norteamérica, Expasy del Instituto Suizo de Bioinformática, y EMBL con sede en Heidelberg que asocia 27 estados de la Unión Europea. Otros servicios abiertos para consultas moleculares también se pueden encontrar en Regulon DB de la UNAM en México, Prodicor de la Universidad técnica de Braunschweig en Alemania, Softberry Inc. con sedes en EUA-NY y en Novosibirsk en Rusia. En México se cuenta con la CONABIO o Comisión Nacional para el Conocimiento y Uso de la Biodiversidad,

que cuenta con sistemas de información de organismos nativos de tamaño macro, que puede complementar a otras fuentes de información antes mencionados, pues tiene como misión la de “promover, coordinar, apoyar y realizar actividades dirigidas al conocimiento de la diversidad biológica, así como a su conservación y uso sustentable para beneficio de la sociedad”. A nivel industrial y de uso de patentes en México se cuenta con el IMPI-Instituto Mexicano de la Protección Industrial cuya misión es “Acercar y proteger eficientemente la propiedad industrial y promover su respeto para impulsar el desarrollo y bienestar en México”. Por la revisión de los portales WEB de estas dependencias se observa que hay mayor profundidad y detalle en las bases de datos como NCBI y EMBL, aunque sólo se encuentran en idioma inglés.

Búsqueda de genomas pequeños en GenBank

Los microorganismos de la clase *Mollicutes* se caracterizan por poseer genomas pequeños (Rivera et al., 2006). Uno de los géneros pertenecientes a esta clase es *Mycoplasma*, como se observa en la Tabla 1,

Mycoplasma tiene un genoma aproximado de una megabase (Mb), mientras que el grupo de *E. coli*, *Pseudomonas*, *Bacillus* o *Haemophilus*, tiene genomas mayores a dos megabases. Debido a lo anterior y sumado a su relevancia médica y/o ambiental este grupo microbiano puede ser aprovechado con propósitos didácticos, por ejemplo, para ejercitar el uso de herramientas bioinformáticas.

Durante una búsqueda online en la dirección electrónica <https://www.ncbi.nlm.nih.gov/genome/>, para genomas en GenBank con el término “*Mycoplasma*”, se desplegaron 20 resultados de 96 totales referidos. El primer resultado mostrado con nombre de género y especie, de acuerdo con el criterio de Lineo correspondió a *Mycoplasma mycoides*, que describe a un patógeno productor de neumonía, refiere a información del 2004 y que contiene 63 genomas registrados (Figura 4). También se puede acceder a un filograma, donde las secuencias casi idénticas, con ligeras variaciones genómicas tienen un lugar particular en el árbol filogenético (Figura 5).

Display Settings: Summary, 20 per page

Search results

Items: 1 to 20 of 96

- Mycoplasma**
 - 1. **Mycoplasma** sp. overview
 - Kingdom: Bacteria; Subgroup: Mycoplasmatota
 - Sequence data: genome assemblies:110
 - Chromosome: 1
 - Date: 2013/05/31
 - ID: 15866
 - Mycoplasma mycoides**
 - 2. Causative agent of contagious pleuropneumonia in livestock
 - Kingdom: Bacteria; Subgroup: Mycoplasmatota
 - Sequence data: genome assemblies:53
 - Chromosome: 1; Plasmids: 1
 - Date: 2004/02/29
 - ID: 720
 - Mycoplasma capricolum**
 - 3. Pathogen of goats
 - Kingdom: Bacteria; Subgroup: Mycoplasmatota
 - Sequence data: genome assemblies:31
 - Chromosome: 1
 - Date: 2005/12/06
 - ID: 521
 - Mycoplasma leachi**
 - 4. *Mycoplasma* sp. bovine group 7 overview
 - Kingdom: Bacteria; Subgroup: Mycoplasmatota
 - Sequence data: genome assemblies:3
 - Chromosome: 1
 - Date: 2010/11/30
 - ID: 1696
 - Mycoplasma yeastsii**
 - 5. *Mycoplasma* yeastsii overview
 - Kingdom: Bacteria; Subgroup: Mycoplasmatota
 - Sequence data: genome assemblies:2
 - Chromosome: 1
 - Date: 2013/04/23
 - ID: 10708

Figura 4. Despliegue de información al buscar el término *Mycoplasma* en NCBI. El primer registro corresponde al género sin especificar la especie, del segundo al quinto registro corresponde a la nomenclatura de Lineo como género y especie. Imagen-editada por los autores. <https://www.ncbi.nlm.nih.gov/genome/?term=mycoplasma+mycoides> (1 de mayo de 2023)

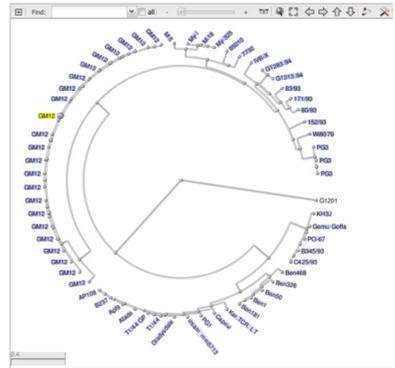


Figura 5. Despliegue de información al navegar en la liga *Mycoplasma mycoides* en NCBI. En este árbol filogenético se observan 23 registros de la cepa GM12. <https://www.ncbi.nlm.nih.gov/genome/720> (recuperado el 15 de mayo del 2023)

Se navegó para acceder a un genoma y se eligió a *Mycoplasma mycoides* subespecie *capri* cepa GM12, con una rama del filograma que contenía 23 genomas. El número de acceso CP001668.1, con un tamaño de genoma de 1084586 bases, fechado el 14 de mayo del 2009. La exploración de la rama y la información accesible indicó que había un conjunto de genomas de tipo sintético todos denominados con el mismo nombre de cepa GM12. Por la variación génica probable de una misma cepa se procedió a investigar *in silico* (computacionalmente) la similitud o diferencia génica-genómica de una misma cepa de *Mycoplasma mycoides*.

Las herramientas bioinformáticas nos brindan la posibilidad de manejar una amplia gama de datos, aunque más se enfocan en explorar cantidades grandes de información; su adecuado procesamiento y análisis dependerá en principio de la capacidad del equipo de computación. Para trabajar con datos abundantes se requiere descargar e instalar varios programas de computadora y preferentemente manejar una terminal con reducidos recursos visuales (entorno gráfico). A pesar de que en la actualidad estos recursos son muy demandados por la comodidad visual e intuitiva para el usuario (“amigables”). Actualmente se busca conjuntar el poder de análisis con rapidez para el procesamiento y análisis de datos.

A partir de la información en el árbol filogenético es posible seguir explorando a detalle estos *Mycoplasma mycoides* lo cual brinda muchas oportunidades para saber a qué puede deberse esta amplia variedad de secuencias de consulta pública. Una de estas maneras implica el obtener las secuencias genómicas de cada uno de estos registros y compararlos mediante un alineamiento (Figura 6). Dicho alineamiento mostró que la diferencia entre estas es reducida, lo que indica que sólo pocas secuencias son distintas entre las muestras comparadas.

Al realizar una investigación bibliográfica sobre los estudios ligados a la publicación de estas secuencias se encontró el artículo "Removal of a subset of Non-essential genes fully attenuates a highly virulent *Mycoplasma strain*" (Jores et al., 2019), donde se señala que éstas son resultado de un ejercicio de edición de genomas (Lartigue et al., 2009). La cepa GM12, una de las cepas más virulentas pertenecientes a *Mycoplasma mycoides* subespecie *capri*, ocasiona graves pérdidas en la industria ganadera y representa un riesgo para la fauna y la economía. A partir del interés por realizar un estudio con herramientas de ingeniería genética, los autores utilizaron a *Saccharomyces cerevisiae* como vector para editar las secuencias de ADN de la bacteria. El genoma completo fue insertado en la levadura con el uso de enzimas de restricción, se removieron 68 genes no esenciales, de 5 regiones distintas del cromosoma bacteriano, una décima parte del material genético. Resultando en la elaboración de 20 genomas transplantados a células de *Mycoplasma capricolum* subesp. *capricolum* cepa CK. En la investigación descrita se observó que como consecuencia de las delecciones (remociones) hubo una atenuación en la virulencia de la bacteria.

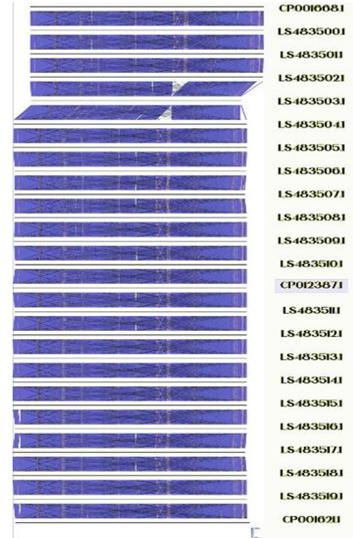


Figura 6. Alineamiento genómico de *Mycoplasma mycoides* GM12. El recuadro azul representa la cepa silvestre GM12. A la derecha están indicados los números de acceso en GeneBank para cada uno de los 24 genomas investigados. En gradiente de color se indica, mayor similitud en azul y menor en rosa. Elaborado con EasyFig. Imagen-elaborada por los autores.

Esta información revela en parte la relevancia de estudiar los genomas pequeños como el de *Mycoplasma*, en los que el más mínimo cambio puede alterar el modo de vida de un organismo.

Para continuar con el trabajo bioinformático en esta misma plataforma (GenBank), se descargó el genoma de uno de los *Mycoplasma mycoides* GM12 así como algunos genomas de distintos organismos como *E. coli*, *Pseudomonas aeruginosa*, *Ureaplasma parvum*, *Haemophilus influenzae*, *Yersinia pestis*, y *Gallibacterium anatis*, los que difieren en el tamaño del genoma en al menos 1000 pares de bases respecto a *M. mycoides*. Particularmente se buscó realizar un análisis filogenético a partir de marcadores genéticos conservados como los genes 16S y 23S. Para obtener las secuencias mencionadas fue necesario buscar las especies microbianas en el apartado del genoma, seleccionar el número de referencia (accession no.) y a partir de la información del genoma desplegado buscar los genes ribosomales mencionados y descargar su secuencia de bases.

Una vez obtenidas las secuencias, podemos elegir hacer uso de distintos recursos computacionales como BLAST local, Mega o Jalview, entre otros, para realizar distintos análisis como: alineamiento, análisis de conservación, cálculo de similitudes, construcción de árboles filogenéticos y predicciones de estructura, entre otros. Para el caso particular, se realizó el alineamiento múltiple de las secuencias con clustal omega y a partir del resultado se construyó el árbol filogenético. En el caso del alineamiento utilizando los marcadores 16S (figura 7) y 23S (figura 8) en el análisis bioinformático se observó que *Mycoplasma mycoides* se asocia con una relación filogenética muy lejana con los otros organismos siendo el más cercano *Ureaplasma parvum*. En el árbol filogenético las especies *E. coli* y *Y. pestis* están cercanamente relacionadas en un grupo, mientras que

en otro grupo se asocian *G. anatis* y *H. influenzae* con *M. mycoides* aunque se relacionan de manera poco cercana. A su vez estos organismos se encuentran lejanamente relacionados con *P. aeruginosa* quizás porque esta última especie tiene el mayor contenido de guanina y citosina (o GC).

El resultado del análisis de estos marcadores genéticos, en estos organismos muestra una historia evolutiva a partir de la que se puede inferir que, los distintos ambientes donde se desarrollaron estos organismos están reflejados en su material genético, ya que incluso al comparar secuencias cortas de 16S o 23S se observa una separación filogenética de los organismos más allá de las considerables diferencias del tamaño de sus genomas.

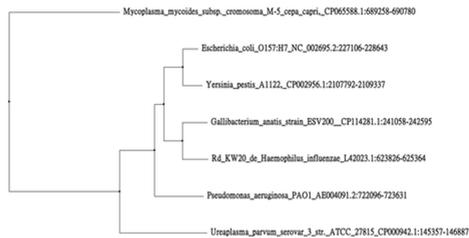


Figura 7. Árbol filogenético basado en el alineamiento múltiple del gen 16S de distintas especies bacterianas con *Mycoplasma*. Imagen-elaborada por los autores.



Figura 8. Árbol filogenético basado en el alineamiento múltiple del gen 23S de distintas especies bacterianas con *Mycoplasma*. Imagen-elaborada por los autores.

Conclusiones

En biología es importante conocer con detalle a todos los organismos vivos aplicando la metodología sistemática reconocida para poder identificar biológica y funcionalmente a todo organismo que tenga un nombre coloquial producto de la semántica de las diferentes regiones en la geografía mundial. En este contexto es importante reconocer que las ciencias biológicas son integradoras y que avanzan de la mano de la tecnología, por lo que es imprescindible saber que la biología molecular y los organismos genéticamente modificados forman parte de nuestro entorno social, en la educación, la ciencia, la medicina, la salud y la innovación tecnológica, aunque todavía la comunidad debe comprender, los alcances de los nuevos organismos que surgen por la tecnología y los que surgen por millones de pasos evolutivos (figura 9), estos últimos mucho más complejos que los que el ser humano ha pensado o diseñado hasta el inicio del siglo XXI. Una evidencia relativa de inconsistencia taxonómica o de mal acomodo de registros en la base de datos de GenBank, se obtuvo al estudiar los genomas de *Mycoplasma mycoides var capri*, aplicando métodos bioinformáticos sencillos más que por una búsqueda dirigida con técnicas de minería de datos. NCBI es un excelente referente en las ciencias biológicas y la medicina, por lo que sospechamos que esta inconsistencia en GenBank quizás se deba a la falta de personal con perfil de bioinformático dedicado al mantenimiento y actualización en NCBI o falta de programas de computación que actualicen la base de datos de GenBank. Con base en esta breve investigación, quizás sea posible argumentar a favor de la apertura de programas educativos que impulsen la formación de capital humano bien capacitado en la bioinformática y la biología como parte de las aplicaciones de la computación.

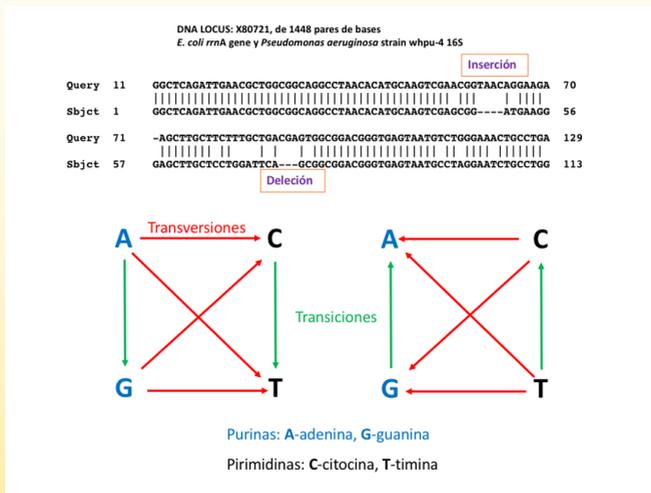


Figura 9. Cambios generales que se pueden observar en las secuencias de los genes y genomas. En un alineamiento query=gen de *E. coli* y subject=gen de *Pseudomonas aeruginosa*, las inserciones y deleciones se refieren la ganancia o pérdida de nucleótidos en las secuencias analizadas, los guiones horizontales en la secuencia indican los nucleótidos faltantes. La ausencia de líneas verticales en los alineamientos son mutaciones o sustituciones, que por su tipo se conocen como transiciones (flechas verdes) o transversiones (flechas rojas).

Declaración de no Conflicto de intereses

Los autores declaran que no existe conflicto de intereses

Autoría del artículo

Los autores María Fernanda Chávez Jacobo, Marina Casilda Guadalupe Osorio Falconi, Rigel Quintero Luis participaron como alumnas de la Maestría en Microbiología del ICUAP-BUAP. María Elena Cobos-Justo como alumna del Doctorado en Microbiología del ICUAP-BUAP. El restante 50% de los autores pertenecen a la Red Modelos Microbianos

de Importancia Agrobiotecnológica.

Declaración de privacidad

Los autores declaran que se ha protegido la información sensible de las personas nombradas. La recopilación, el procesamiento y el almacenamiento de su información personal se hace en apego a la Declaración de Privacidad que rige a la Revista y a las Instituciones mencionadas. De acuerdo a la norma vigente, en el presente escrito se da el crédito a los autores de trabajos previos. Otras tecnologías electrónicas de seguimiento y almacenamiento de información poseen los derechos de divulgación y consulta que aquí se reconocen.

Agradecimientos

A la BUAP por el apoyo del Proyecto BUAP 100103133-VIEP2023 y VIEP 2024.

Referencias

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science (New York, N.Y.)*, 277(5331), 1453-1462. <https://doi.org/10.1126/science.277.5331.1453>
- Campbell, N. A., Reece, J. B. (2007). *Biología*. Ed. Médica Panamericana.
- Chen, J. Q., Zhang, H. T., Hu, M. H., Tang, J. G. (1995). Production of human insulin in an *E. coli* system with Met-Lys-human proinsulin as the expressed precursor. *Applied biochemistry and biotechnology*, 55(1), 5-15. <https://doi.org/10.1007/BF02788744>
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science (New York, N.Y.)*, 269(5223), 496-512. <https://doi.org/10.1126/science.7542800>
- Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., Riggs, A. D. (1979). Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America*, 76(1), 106-110. <https://doi.org/10.1073/pnas.76.1.106>
- Ikehara, M., Ohtsuka, E., Tokunaga, T., Taniyama, Y., Iwai, S., Kitano, K., Miyamoto, S., Ohgi, T., Sakuragawa, Y., Fujiyama, K. (1984). Synthesis of a gene for human growth hormone and its expression in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 81(19), 5956-5960. <https://doi.org/10.1073/pnas.81.19.5956>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), 816-821. <https://doi.org/10.1126/science.1225829>
- Jores, J., Ma, L., Ssajjakambwe, P., Schieck, E., Liljander, A., Chandran, S., Stoffel, M. H., Cippa, V., Arfi, Y., Assad-Garcia, N., Falquet, L., Sirand-Pugnet, P., Blanchard, A., Lartigue, C., Posthaus, H., Labrousseau, F., & Vashee, S. (2019). Removal of a Subset of Non-essential Genes Fully Attenuates a Highly Virulent *Mycoplasma* Strain. *Frontiers in microbiology*, 10, 664. <https://doi.org/10.3389/fmicb.2019.00664>
- Krebs, J. E., Goldstein, E. S., Kilpatrick, S. T. (2017). *Lewin's genes XII*. Jones & Bartlett Learning.

Karp, G. (2009). Cell and molecular biology: concepts and experiments. John Wiley & Sons.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., ... Danchin, A. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657), 249–256. <https://doi.org/10.1038/36786>

Lartigue, C., Vashee, S., Algire, M. A., Chuang, R. Y., Benders, G. A., Ma, L., Noskov, V. N., Denisova, E. A., Gibson, D. G., Assad-Garcia, N., Alperovich, N., Thomas, D. W., Merryman, C., Hutchison, C. A., 3rd, Smith, H. O., Venter, J. C., & Glass, J. I. (2009). Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science (New York, N.Y.)*, 325(5948), 1693–1696. <https://doi.org/10.1126/science.1173759>

Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Onishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., Shinagawa, H. (1999). Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak. *Genes & genetic systems*, 74(5), 227–239. <https://doi.org/10.1266/ggs.74.227>

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2), 174–182. <https://doi.org/10.1007/s00239-004-0046-3>

Pereyre S, Sirand-Pugnet P, Beven L, Charron A, Renaudin H, Barré A, et al. (2009). Life on Arginine for *Mycoplasma hominis*: Clues from Its Minimal Genome and Comparison with Other Human Urogenital Mycoplasmas. *PLoS Genet* 5(10): e1000677. <https://doi.org/10.1371/journal.pgen.1000677>

Rivera-Tapia, J.A., Ramírez, L.C., Cerezo, S.G. (2006). Comparación genómica en micoplasmas de interés médico. *Anales Médicos* 51(2): 74-79. <https://www.medigraphic.com/pdfs/abc/bc-2006/bc062f.pdf>

Schne, A., Curtis, H., Barnes, S. (2008). Curtis Biología. Panamericana.

Sullivan, M. J., Petty, N. K., Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)*, 27(7), 1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>

Vázquez-Cruz, C., Olmedo-Alvarez, G. (1997). Mechanism of decay of the *cry1Aa* mRNA in *Bacillus subtilis*. *Journal of bacteriology*, 179(20), 6341–6348. <https://doi.org/10.1128/jb.179.20.6341-6348.1997>

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009).

Jalview Version 2 - A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25 1189-1191. (doi:10.1093/bioinformatics/btp033)

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265–269. <https://doi.org/10.1038/s41586-020-2008-3>

Sitios de internet consultados

BLAST: *Basic Local Alignment Search Tool*. (s. f.). <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Comisión Nacional para el Conocimiento y Uso de la Biodiversidad | Gobierno | gob.mx. (s. f.). <https://www.gob.mx/conabio/>

Dudek, C. (s. f.). *prodoric-frontend*. <https://www.prodoric.de/>

Encyclopedia Britannica. <https://www.britannica.com/science/taxonomy>

EMBL. (s. f.). *European Molecular Biology Laboratory*. EMBL.org. <https://www.embl.org/>

GenBank Overview. (s. f.). <https://www.ncbi.nlm.nih.gov/genbank/>

Instituto Mexicano de la Propiedad Industrial | Gobierno | gob.mx. (s. f.). <https://www.gob.mx/impi>

MEGA (s. f.). <https://www.megasoftware.net/>

National Institutes of Health. (n.d.). *Biographical overview | Salvador E. Luria - Profiles in science*. U.S. National Library of Medicine. Retrieved May 5, 2023, from <https://profiles.nlm.nih.gov/spotlight/ql/feature/biographical-overview>

RegulonDB <http://regulondb.ccg.unam.m>. (s. f.). *RegulonDB Database*.

SIB Swiss Institute of Bioinformatics | Expasy. (s. f.). <https://www.expasy.org/>

Softberry Home Page. (s. f.). <http://www.softberry.com/>