



**CONGRESO
IBEROAMERICANO**
DE CIENCIA, TECNOLOGÍA,
INNOVACIÓN Y EDUCACIÓN

BUENOS AIRES, ARGENTINA
12, 13 Y 14 DE NOVIEMBRE 2014

**CONGRESSO
IBERO-AMERICANO**
DE CIÊNCIA, TECNOLOGIA,
INOVAÇÃO E EDUCAÇÃO

BUENOS AIRES, ARGENTINA
12, 13 Y 14 DE NOVIEMBRO 2014

Códigos y su importancia actual en la enseñanza de las Ciencias. Aplicación a la clasificación de hongos.

Vaira, S.M.; Carrera, E.; Armando, G.

Códigos y su importancia actual en la enseñanza de las Ciencias. Aplicación a la clasificación de hongos.

Stella M. Vaira¹; Elena F. de Carrera¹; Gustavo Armando²

¹Departamento de Matemática de la Facultad de Bioquímica y Ciencias Biológicas de la Universidad Nacional del Litoral. Argentina.

²Tesista de grado de Licenciatura en Biotecnología.

email: svaira@fcb.unl.edu.ar – ecarrera@fcb.unl.edu.ar

Resumen.

La Teoría de la Codificación Algebraica es una de las aplicaciones más recientes del álgebra. Se suele fijar su nacimiento en el año 1948 con los trabajos de Claude Shannon (1916-2001) sobre la Teoría Matemática de la Información, teoría que resultó ser extremadamente útil, no sólo en el diseño de circuitos de computadoras y la tecnología de comunicaciones, sino que también ha hallado aplicaciones fecundas en campos tan diversos como la biología, psicología, fonética e incluso semántica y literatura.

El matemático Richard Hamming (1915-1998), uno de los creadores de la teoría de códigos, que permitió, entre otras cosas, proteger los mensajes contra los errores, si bien no hay seguridad absoluta, pero sí una gran probabilidad de conseguir enviar información segura, esta teoría tiene como objetivo conseguir aumentar todo lo que se pueda la probabilidad de que el mensaje correcto pueda ser recuperado a pesar de los posibles errores, siempre que el número de éstos sea razonable. Sin olvidarnos de mencionar a Alan Turing (1912 – 1954) quien además de sus conocidos e indiscutibles aportes a la Matemática, Computación e Inteligencia Artificial, también trabajó en la [biología matemática](#), concretamente en la [morfogénesis](#). Publicó un trabajo sobre esta materia titulado "Fundamentos Químicos de la Morfogénesis" en 1952. Su principal interés era comprender la [filotaxis](#) de [Fibonacci](#), es decir, la existencia de los [números de Fibonacci](#) en las estructuras vegetales.

La investigación en Biología Molecular nació en los laboratorios experimentales, la inmensa cantidad de datos generados en los últimos años con la conclusión del *Proyecto Genoma Human*, es difícil de entender sin la Bioinformática, este trabajo interdisciplinario muestra la clasificación de cepas de hongos, algunos fitopatógenos como *Cercospora sojina* y *Cercospora kikuchii* muy frecuentes en la zona del Litoral Argentino, utilizando herramientas computacionales, previa utilización de códigos binarios y ternarios para generar una matriz de datos que permita a través de las técnicas estadísticas multivariadas, que incluye la aplicación de diferentes medidas de distancia, clasificarlos según su variabilidad genética. El hallazgo en este estudio de

valores altos de similitud entre los hongos señaló la estrecha relación existente entre ellos.

Bibliografía.

1. Hardy, D. W.; Walter, C. L. *Applied algebra – Codes, Ciphers and Discrete Algorithms*. Prentice Hall. 2002.

2. Johnson R, Wirchen D. *Applied Multivariate Statistical Analysis*. Fourth Edition. Prentice Hall. New Jersey, USA. 1998.

Introducción:

La revolución biológica de los años cincuenta nació entre contactos y transferencias de disciplinas, en los márgenes de la física, química y biología. Son los físicos como Schrödinger (1887-1961) que han proyectado sobre el organismo biológico los problemas de la organización física. Los investigadores intentaron descubrir la organización del patrimonio genético a partir de las propiedades químicas del ADN.

La noción de información, emanada de la práctica social, ha tomado un sentido científico preciso, nuevo, en la teoría de Shannon, después ella ha migrado a la biología para inscribirse en el gen, allí ella se ha asociado a la noción de código, que se convirtió en la noción de código genético.

Shannon es uno de los padres del código binario, base de toda la informática de hoy día. A partir de sus estudios, que establecen las bases para unificar toda la información en un código de transmisión único, surgirá su teoría matemática de la comunicación.

Shannon estudió el flujo de las ondas electromagnéticas a través de un circuito. Y postuló que a través del código binario se puede homogeneizar todo tipo de información (textos, sonidos, imágenes, entre otras cosas). Esta teoría distingue claramente entre el mensaje y el medio por el que éste se transmite.

Se considera el trabajo de Shannon una pieza importante en el desarrollo que llevó a la comunicación a poseer un rol fundamental en la sociedad actual. Esta teoría también sienta los fundamentos matemáticos de la revolución tecnológica de la segunda mitad del siglo XX.

Desde el estudio de las álgebras booleanas Shannon realizó una teoría acerca del código binario, que es la base del lenguaje digital, a partir de unidades básicas de información, definidas por una respuesta binaria: sí y el no, 0 o 1, abierto o cerrado.

El 0 y el 1 aparecen como la unidad básica de la información, como la base del mensaje. Una información compleja es una sucesión de unidades básicas, de unos y ceros. Más allá de la formulación teórica, Shannon construyó circuitos y máquinas basadas en los flujos binarios de información, mediante interruptores que se anticipaban a muchos de los desarrollos de las décadas posteriores.

La idea de transmitir una palabra o un mensaje cualquiera de modo que su contenido sea seguro y no revelado a todo miembro de la comunidad es una necesidad casi tan antigua como el hombre. Así los arqueólogos han descifrado jeroglíficos de más de 4500 años de antigüedad que se encontraban codificados. Esta codificación se mantuvo a lo largo de los siglos y se fue perfeccionando.

Sin olvidarnos de mencionar a Alan Turing (1912 – 1954) quien además de sus conocidos e indiscutibles aportes a la Matemática, Computación e Inteligencia Artificial, también trabajó en la [biología matemática](#), concretamente en la [morfogénesis](#). Publicó un trabajo sobre esta materia titulado "Fundamentos Químicos de la Morfogénesis" en 1952. Su principal interés era comprender la [filotaxis](#) de [Fibonacci](#), es decir, la existencia de los [números de Fibonacci](#) en las estructuras vegetales.

ADN, Matemática y Bioinformática.

La matemática ha tenido un rol fundamental en la comprensión de organismos vivos por ejemplo en la codificación de los tripletes que genera ARNm. O en la creación de métodos estocásticos que realicen búsquedas de similitudes para la creación de árboles que relacionen especies.

En la actualidad se está utilizando el código de barra como una forma de reconocer e identificar especies es decir cada especie lleva en su genoma características particulares que lo identifican.

En 2003, el científico canadiense Paul Hebert propuso el uso de una porción de la secuencia de un gen mitocondrial que codifica para la enzima perteneciente a la cadena respiratoria citocromo c oxidasa subunidad 1 (COI), como la secuencia de referencia que podría utilizarse a manera de un código de barras genético.

Conocer la especie de un organismo el cual estamos estudiando constituye una pregunta fundamental de la biología. El nombre de la especie (su nombre científico, para ser más exactos) es la puerta con la cual se accede al conocimiento que existe sobre esa clase de organismo y así saber sus características fenotípicas y genotípicas. Esto que era netamente biológico ha ido tomando un trasfondo matemático importante, puesto que las nuevas técnicas de identificación rápida y segura que se utilizan en la actualidad están basadas en la teoría de probabilidad.

Técnicas de agrupamientos y clasificación como “Unweighted Pair Group Method with Arithmetic Mean (UPGMA)”, WARDS, entre otras, comenzaron a utilizarse para la identificación de especies, la mayoría de las especies sólo pueden ser identificadas en el estado adulto por lo que es probable que durante gran parte de su ciclo de vida no puedan ser identificadas por los métodos morfológicos convencionales, en el caso de organismos unicelulares donde las características fenotípicas se reducen a observar las características (color, forma, emanación de gas, entre otros) que pueden tomar una colonia en una placa de Petri realizar su identificación resulta un poco más engorroso y por último, como la identificación generalmente se ha basado en el uso de caracteres morfológicos, puede ocurrir que si se encuentra un fragmento de un organismo, este no tenga completas las partes necesarias para su identificación.

Planteo del problema.

La producción mundial de soja durante la campaña 2006/07 alcanzó, según el USDA, las 236,5 millones de toneladas, que alcanza el 82% de la producción mundial: Estados Unidos (37%), Brasil (25%) y Argentina (20%).

La Argentina ocupa el tercer lugar como productor y exportador de grano de soja en el mundo, y es el principal exportador internacional de aceite y harina derivados de la oleaginosa (63 y 44% respectivamente).

El crecimiento en el área sembrada en el país, estuvo acompañado por el incremento de la superficie en siembra directa y del monocultivo, que al dejar el rastrojo de soja en superficie, resulta una excelente combinación que genera condiciones para la supervivencia e incremento de los patógenos hemibiotróficos o necrotróficos, como lo son la mayoría de los causantes de las enfermedades foliares y las denominadas de fin de ciclo.

Entre las enfermedades que la afectan hacia el final de su ciclo se encuentran: el tizón de la hoja y la mancha púrpura de la semilla, cuyo agente etiológico es *Cercospora kikuchii*, Las plantas enfermas pierden gran cantidad de follaje, lo que adelanta su maduración, sin el llenado correcto de las vainas, las semillas disminuyen su germinación.

Esta enfermedad es una de las de mayor prevalencia e incidencia en las principales áreas sojeras del país, habiendo incrementado de manera sostenida su prevalencia e incidencia en la región del centro-sur del país. La planta puede presentar una defoliación anticipada que provocaría importantes pérdidas de rendimiento.

La *Cercospora Sojina* es el agente etiológico causante de la mancha ojo de rana (MOR), esta es una enfermedad fúngica de la soja, reportada por primera vez en la zona centro del país en 1983. La Mancha Ojo de Rana (MOR) es una enfermedad fúngica que presenta una alta variabilidad ya que existen diferentes especies.

La aplicación de fungicidas foliares es una alternativa viable para disminuir el Impacto de la enfermedad.

El conocimiento de la estructura poblacional de estos hongos es muy importante para el desarrollo tanto económico como académicos. Poder realizar análisis espaciales con cepas o repiques resistentes a ciertos químicos o cepas con mayor letalidad que otras, poder observar mutaciones, realizar análisis de sensibilidad contra antibióticos in vitro, identificar nuevas especie, entre otros. Hace de vital importancia utilizar métodos matemáticos y estadísticos capaces de diferenciar especies, o repiques entre sí.

El problema que se presenta es que el mejor método para realizar el análisis entre especies no está completamente resuelto y esto se debe fundamentalmente a dos problemas principales, uno a la inexistencia de una hipótesis nula apropiada y otro a la naturaleza de las distribuciones muestrales multivariantes que suelen ser complejas.

Este trabajo interdisciplinario muestra la clasificación de cepas de hongos, algunos fitopatógenos como *Cercospora sojina* y *Cercospora kikuchii* muy frecuentes en la zona del Litoral Argentino, utilizando herramientas computacionales, previa utilización de códigos binarios para generar una matriz de datos que permita a través de las técnicas estadísticas multivariadas, que incluye la aplicación de diferentes medidas de distancia, clasificarlos según su variabilidad genética. El hallazgo en este estudio de valores altos de similaridad entre los hongos señaló la estrecha relación existente entre ellos.

Materiales y Métodos

Tratamiento Biológico y Estadístico:

Las muestras a analizar fue provista por la Cátedra de Microbiología General, dicha muestra estaba compuesta por repiques de *cercospora sojina* y *cecospora kikuchii*

Se utilizó el marcador de peso molecular 100-1000pb (INBIO-HIGHWAY, Argentina). Para visualizar la imagen de los productos de la PCR (patrones de bandas) se empleó un equipo Gel-Doc XR (BioRad Laboratories). La ausencia o presencia de bandas polimórficas fue considerada como 0 / 1, respectivamente. Del gel de agarosa y a través del programa Quantity One se extrajo una imagen de las corridas electroforéticas de la amplificación de las regiones micro satélite de *cercospora sojina* y *cercospora kikuchii* se identificaron las bandas de las corridas del ADN en cada carril. Se codificaron las bandas encontradas en cada carril utilizando uno (1) para la presencia de ADN (banda nítidamente marcada) y cero (0) para la ausencia (no existe banda marcada) como se muestra en la Figura 1.

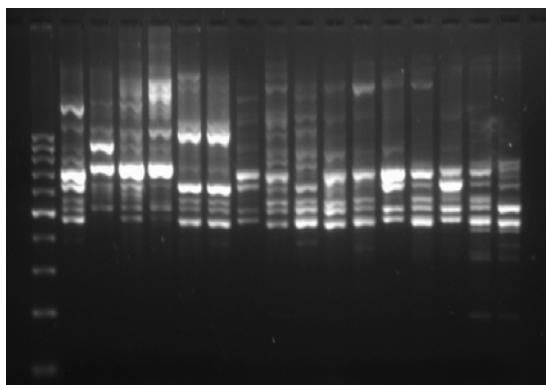


Figura 1. Bandas de amplificación de las regiones entre micro satélites de los aislamientos de *Cercospora Kikuchii* y *cercopora sojina*, corridos en un gel de agarosa.

Luego se procedió a construir la matriz binaria y de esta se procedió armar la matriz de similaridad, a través de la medida de Dice. Y luego se procedió a la formación de los grupos (cluster) para dicho fin se utilizaron los métodos de agrupamientos: inter grupo, Vecino mas Cercano y el método vecino mas Lejano e intra grupo y se confeccionaron sus respectivos dendogramas, todo se realizó con el programa SPSS 21 y R-Projet.

Resultados y Discusión.

Se realizó el armado de la matriz binaria (1 presencia de banda de ADN y 0 ausencia de bandas de ADN). Cada columna de la matriz representa un *hongo Cercospora sojina* (CS) o *Cercospora Kikuchii* (CK), el número que acompaña al hongo es el número del aislamiento que sirve para identificación del operador.

CK68	CK69	CK70	CK71	CS78
0	0	0	1	0
0	0	1	1	0
1	1	1	1	0
0	0	0	0	0
1	1	1	1	0
1	1	1	1	0

Tabla 1. Muestra parcial de los hongos (columnas), en total 16, y su codificación binaria, en total 14 filas de la matriz

Con la matriz binaria se realizó un análisis de correspondencia entre las variables y una reducción de dimensiones con el fin visualizar cual era la variabilidad explicada por la dimensión uno y dos y así puedes ver en forma gráfica los posibles cluster que se podrían formar. (Grafico 1)

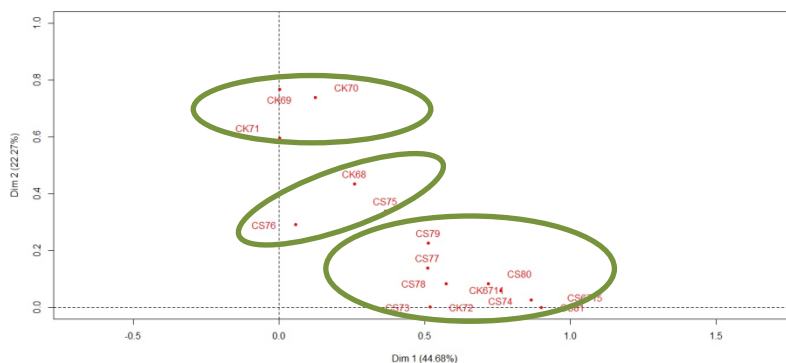
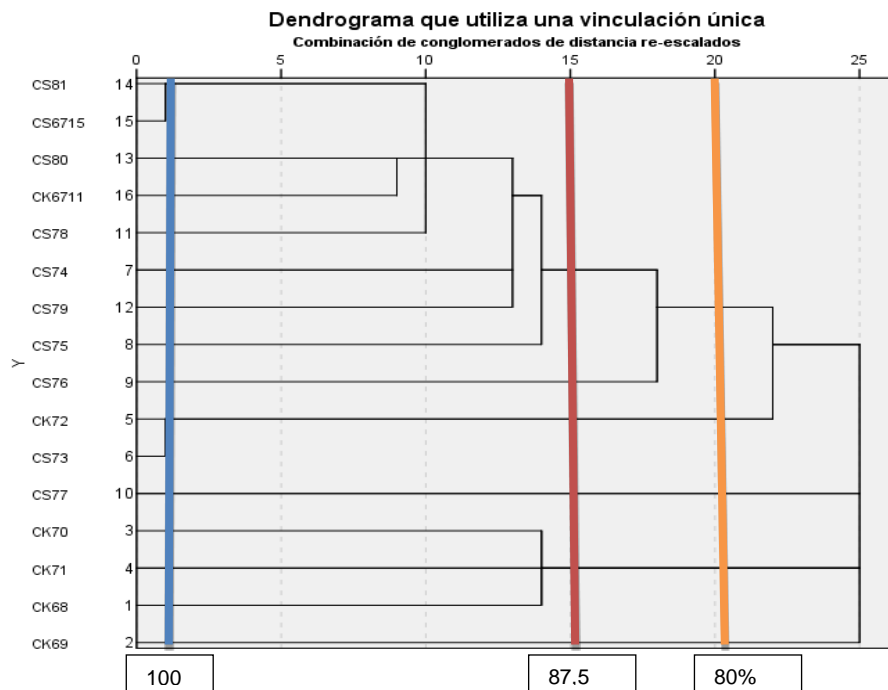


Grafico 1. Reducción a dos dimensiones del problema en donde se pueden visualizar en forma gráfica tres agrupamiento de hongos posibles, el grafico muestra que el 67% de la variabilidad de los clúster están determinado por las primeras dos dimensiones.

Se calculó la matriz de similitud utilizando la medida DICE y el programa SPSS. Se obtuvo el siguiente resultado:

Variables	Matriz de Similitud															
	CK68	CK69	CK70	CK71	CK72	CS73	CS74	CS75	CS76	CS77	CS78	CS79	CS80	CS81	CS6715	CK6711
CK68	1,000															
CK69	,769	1,000														
CK70	,875	,769	1,000													
CK71	,750	,769	,875	1,000												
CK72	,615	,400	,615	,462	1,000											
CS73	,615	,400	,615	,462	1,000	1,000										
CS74	,667	,444	,667	,500	,667	,667	1,000									
CS75	,588	,286	,471	,471	,571	,571	,615	1,000								
CS76	,556	,267	,556	,556	,667	,667	,429	,842	1,000							
CS77	,429	,182	,429	,429	,545	,545	,600	,667	,500	1,000						
CS78	,615	,200	,462	,308	,600	,600	,667	,714	,533	,727	1,000					
CS79	,769	,600	,769	,615	,600	,600	,889	,571	,400	,545	,600	1,000				
CS80	,667	,333	,533	,533	,667	,667	,727	,875	,706	,769	,833	,667	1,000			
CS81	,615	,400	,615	,462	,800	,800	,889	,714	,533	,727	,800	,800	,833	1,000		
CS6715	,615	,400	,615	,462	,800	,800	,889	,714	,533	,727	,800	,800	,833	1,000	1,000	
CK6711	,714	,364	,571	,429	,727	,727	,800	,800	,625	,667	,909	,727	,923	,909	,909	1,000

Luego de construida la matriz de similaridad se utilizaran métodos que permitan el agrupamiento de las variables. Existen muchos métodos para realizar este agrupamiento pero ninguno proporciona una solución única o más correcta para todos los problemas que se pueden plantear, por lo que es posible obtener distintos resultados según el método elegido. El criterio del investigador, el conocimiento del problema planteado y la experiencia, propondrán el método más adecuado que es conveniente utilizar. De todas formas, es conveniente, usar varios procedimientos con la idea de contrastar los resultados obtenidos y sacar conclusiones, tanto como si hubiera coincidencias en los resultados obtenidos con métodos distintos como si no las hubiera. En el presente trabajo se utilizara los métodos que miden la distancia entre “vecino más cercano” y “distancias entre medias o inter grupo”, “intra grupo” y “vecino más lejano”. Dicho agrupamiento se realiza con el fin de generar una jerarquía de agrupamiento representado en un dendrograma.



El método de agrupamiento **vecino más cercano**: también llamado amalgamamiento simple (single linkage), se considera que la distancia o similitud entre dos clusters viene dada, respectivamente, por la mínima distancia (o máxima similitud) entre sus componentes. Mientras que el método del vecino más cercano asegura que la distancia entre los individuos más próximos de un cluster van a ser siempre menor que la distancia entre elementos de distintos clusters, el del vecino más lejano va a asegurar que la distancia máxima dentro de un cluster será menor que la distancia entre cualquiera de sus elementos y los elementos más alejados de los demás clusters.

Conclusiones.

Los resultados obtenidos aportaron datos sobre el comportamiento de estos hongos, complementando la información existente sobre dos de los fitopatógenos de soja más frecuente en la zona del Litoral Argentino.

El hallazgo en este estudio de valores altos de similaridad entre los hongos señaló la estrecha relación existente entre ellos.

El trabajo interdisciplinario permitió enriquecer el trabajo y al equipo docente.

La búsqueda de técnicas matemáticas y estadísticas adecuadas para la naturaleza de los datos se consiguió a través de una exhaustiva búsqueda bibliográfica.

Bibliografía:

- 1) MORIN, E. (2005). "Sobre la interdisciplinariedad". Publicado en el Boletín No. 2 del *Centre International de Recherches et Etudes Transdisciplinaires* (CIRET). Vol. 1. Congreso Internacional de Transdisciplinariedad.
- 2) FISKE J. (1984). Introducción a la teoría de la comunicación. Editorial Norma.
- 3) HARDY, D; WALKER, C. (2003). Applied Algebra, Codes, Ciphers, Discrete Algorithms. Editorial Pearson. USA.
- 4) Trevors J.T. (2000). "Math, the Genetic Code and Living Organisms." *Bioscene*, vol. 2. Ontario.
- 5) CHATTERJEE, S.; LAUDATO, M.; LYNCH, I. (1996). "Genetic algorithms and their statistical applications: an introduccion." *Computational Statistics and data Analisis*. Vol 22: 633-651.
- 6) Lura, MC, DI CONZA, JA, GONZALES, AM, LATORRE RAPELA MG, TURINO, L, (2007). "Detección de variabilidad genética en aislamientos de *Cercospora kikuchii* contaminantes de un mismo sembradío de soja." *Revista Argentina de Microbiología* 39: 11-14.
- 7) JOHNSON R, WIRCHEN D. (1998). Applied Multivariate Statistical Analysis. Fourth Edition. Prentice Hall. New Jersey, USA.