



Departamento de Ciencias Biológicas y de la Salud

Licenciatura en Química Farmacéutica Biológica

TÍTULO DEL PROYECTO DE INVESTIGACIÓN:

Análisis bioinformático y funcional del factor de virulencia Egh16 en hongos filamentosos de la división Ascomycota: un estudio en patógenos de insectos, plantas y humanos.

PROYECTO GENERICO:

Obtención de materias primas, principios activos, medicamentos y productos biológicos.

ETAPA:

Diseño y desarrollo de productos biológicos por métodos biotecnológicos o de ingeniería genética.

Alumno: Cabrera González Miriam

Matricula: 2173067780

Vo. Bo. De los asesores respecto a los contenidos académicos

Dr. Jesús Eduardo Zúñiga León
Nombre y firma de asesor interno a cargo
No. Económico: 44145

Dr. Juan Esteban Barranco Florido
Nombre y firma de asesor interno a cargo
No. Económico: 24927



INDICE

1. INTRODUCCIÓN	2
2. MARCO TEÓRICO	3
2.1 HONGOS	3
2.1.1 Características generales	3
2.1.2 Estructura y morfología.....	3
2.1.3 Importancia de los hongos	5
2.1.4 Clasificación de los hongos	7
2.2 HONGOS PATÓGENOS DEL SER HUMANO	10
2.3 HONGOS FITOPATOGENOS	13
2.4 HONGOS ENTOMOPATOGENOS.....	15
2.5 MECANISMOS PATOGENICOS DE LOS HONGOS	18
2.6 MECANISMOS DE DEFENSA DEL HOSPEDERO	22
2.7 VIRULENCIA	24
2.7.1 Factores de virulencia.....	25
2.8 IMPORTANCIA BIOINFORMÁTICA Y PROTEOMICA	26
3. JUSTIFICACIÓN	30
4. OBJETIVOS	32
4.1 Objetivo general.....	32
4.1.1 Objetivos específicos.....	32
5. MATERIALES Y MÉTODOS	32
5.1 Programa Informático e Interfaz.....	33
5.2 Bases de datos.....	34
6. RESULTADOS Y DISCUSIONES	35
6.1 Identificación de proteínas implicadas en el factor de virulencia Egh16.	35
6.2 Construcción de un base de datos BLAST	36
6.3 Descarga de Proteomas a partir de UniProtKB.	38
6.4 Identificación de proteínas ortólogas asociadas al factor de virulencia Egh16 mediante Blastp	42
6.5 BLASTP All vs All	50
6.6 Análisis de Clustering	52
6.7 Heatmap	53
7. CONCLUSIONES	56
8. REFERENCIAS	57
9. ANEXOS	59

1. INTRODUCCIÓN

La función principal de un hongo filamentoso es saprofito-desintegrador, por lo que, es importante recordar que la vida sobre la tierra sin la presencia de estos organismos sería imposible y dañina para el ecosistema, ya que sin ellos se rompe la homeostasis de este. Además de beneficiar al ecosistema algunos hongos impactan de otra forma, algunos hongos son patógenos de plantas, de insectos, de animales y de humanos, por ello es importante la caracterización de estos, para evaluar riesgos y beneficios del uso, control y eliminación de los mismos, ya sea en el ámbito clínico, farmacológico, cosmético, alimenticio y ambiental (control de plagas).

Por lo tanto, es de vital importancia saber y entender los mecanismos que controlan la patogenicidad mediada por la adhesión, penetración, replicación del hongo y el sistema inmune del hospedero, en donde los factores de virulencia juegan un papel importante en este proceso. Los factores de virulencia son principalmente de 2 tipos: los expresados en la superficie celular (que desempeñan funciones de adhesión e invasión de tejidos además de formación de biopelículas e inducción de citosinas) y los producidos dentro del hongo y que son exportados al sitio de infección (Miranda *et al.*, 2016). Al parecer el factor de virulencia Egh16 juega un papel importante en el proceso infectivo de los hongos entomopatógenos y fitopatógenos, sin embargo, no hay evidencia suficiente que demuestre que todos los hongos patógenos, incluyendo los de humanos, tienen este factor de virulencia. Partiendo de estos antecedentes el presente proyecto pretende identificar el factor de virulencia Egh16 en hongos filamentosos patógenos no caracterizados, esto con el fin de aportar conocimiento en la regulación de la patogenicidad para tomar medidas de control, mediante diferentes alternativas. La caracterización implica la búsqueda del factor de virulencia Egh16 en proteomas de hongos filamentosos patógenos haciendo uso de herramientas bioinformáticas.

2. MARCO TEÓRICO

2.1 HONGOS

Los hongos se definen como seres vivos *eukariontes*, unicelulares o pluricelulares que no forman tejidos, cuyas células se pueden agrupar formando un cuerpo filamentoso (hongo filamentoso = “moho”) muy ramificado (Pfaller, 2021).

- La principal finalidad es degradar materia orgánica (heterótrofos-saprófitos).

2.1.1 Características generales

Para simplificar la información puede verse en la **Figura 1**.

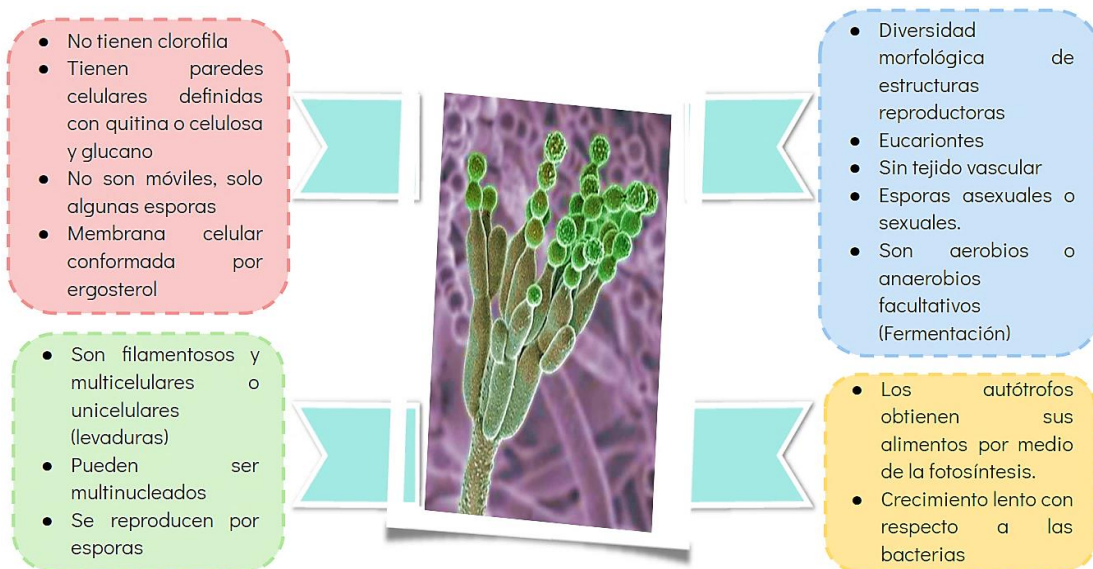


Figura 1. Características generales micóticas.

Elaboración a partir de: Brock. (2015). *Biología de los microorganismos* (14° ed.). México: PEARSON EDUCACIÓN.

2.1.2 Estructura y morfología

Las células micóticas son eucariotas (presencia de núcleo, nucleolos, membrana nuclear y cromosomas lineales) (Madigan, 2015).

De manera muy general podemos decir que a nivel estructural; la membrana celular está constituida por ergosterol, la pared celular por quitina, manano unido a las proteínas de superficie y glucanos que le proporcionan rigidez a esta, así mismo cuenta con la presencia de mitocondrias y retículo endoplasmático (RE) (Kenneth, 2011) (Wiley, 2008) **Figura 2**.

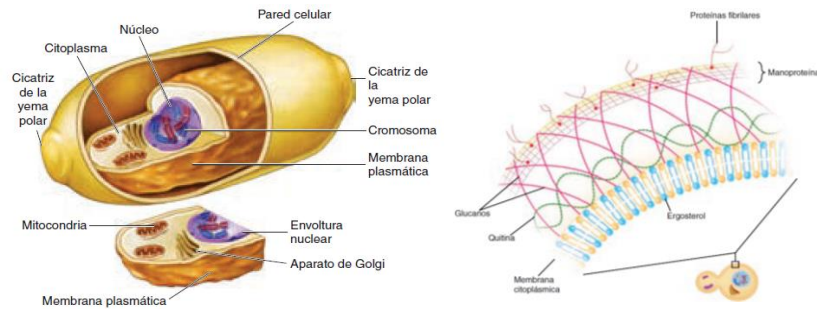


Figura 2. Estructura micótica.

Tomado a partir de: Wiley J, Sherwood L, Woolverton C (2008). Prescott's Principles of Microbiology (eds). Nueva York: McGraw-Hill.

Los hongos se pueden encontrar morfológicamente como: levaduras y mohos **Tabla 1, Figura 3**, o en el caso de los hongos patógenos al ser humano estos pueden adaptarse para poder encontrarse de ambas maneras (dimórficos), lo que les permite su penetración en el hospedero (Pfaller, 2021), las hifas se unen para formar una estructura denominado micelio y las colonias formadas por mohos se conocen como filamentosas, vellosas o lanosas, de ahí que a estos hongos se le conoce también como “hongos filamentosos”.

Tabla 1. Clasificación morfológica de los hongos.

Hongos	Levaduras	<ul style="list-style-type: none"> • Brotes o yemas (gránulos). • Una sola célula que se reproduce mediante gemación o fisión binaria. • Células hijas Pseudohifas
	Moho (hongo filamentos)	<ul style="list-style-type: none"> • Multicelulares formados por una serie de estructuras ramificadas tubulares (hifas), y el conjunto de este micelio. • Ramificación columnar cerrada • Ramificación radián arbórea

Elaboración a partir de: Pfaller Michael A., Murray Patrick Rosenthal R., Ken S., Kobayashi George S. (2021).

"MICROBIOLOGÍA MÉDICA" (9° ed). Argentina: ELSEVIER CASTELLANO.

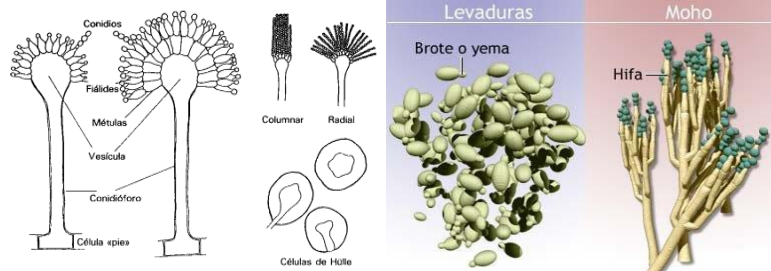


Figura 3. Visualización de levaduras, mohos e hifas.

Tomado a partir de: microbiologia3bequipo5 (2014). Morfología de los hongos. <http://microbiologia3bequipo5.blogspot.com/2014/11/morfologia-de-hongos.html>.

2.1.3 Importancia de los hongos

La función primordial de los hongos es saprofita-desintegradora, gracias a ellos y a sus enzimas, la materia orgánica se recicla y es aprovechada por otros organismos, haciendo posible la reducción de madera, fibras y alimentos a sus componentes básicos con gran eficiencia, que pueden ser beneficiosos o dañinos. Sin ellos se rompe con la homeostasis del ecosistema.

Un ejemplo dañino, para el hombre son las sustancias tóxicas de los hongos; micotoxinas grupo de metabolitos orgánicos que exhiben pesos moleculares relativamente bajos, de interés por sus repercusiones toxicológicas para animales y, naturalmente, para el hombre. Las micotoxinas más conocidas son las aflatoxinas, las cuales son producidas por *Aspergillus flavus* y *Aspergillus parasiticus*, consideradas como potentes carcinógenos en humanos (Castillo, 2015), cabe recalcar que no todos los hongos poseen efectos secundarios perjudiciales, tal es el caso particular de las levaduras, capaces de producir alcohol etílico y dióxido de carbono a partir de glucosa en ausencia de oxígeno, de gran importancia económica para los productores de vino y cerveza (Suárez, 2016); en el proceso de producción del pan, el dióxido de carbono sintetizado por las levaduras queda atrapado en la masa en forma de burbujas, que hacen que esta “se infle” y quede un pan de textura ligera. Tanto el dióxido de carbono, como el alcohol que se producen son eliminados durante el proceso de horneado (Mesas, 2002). Otro ejemplo es el sabor característico que generan los hongos en la producción de quesos, lo que ha desencadenado una cultura de implementación y uso, para mejorar la calidad organoléptica y nutricional (Baraggio *et al.*, 2017).

En 1928, Alexander Fleming, logró aislar y purificar la penicilina secretada por el **Deuteromiceto** *Penicillium notatum* antibiótico para inhibir el crecimiento bacteriano (Heredia-Abarca, 2020).

El ascomiceto *Claviceps purpúrea* es un fitopatógeno del trigo, produciendo una estructura que se llama ergot la cual ocupa el lugar destinado a las semillas de la espiga. Al ingerir ese grano se produce un envenenamiento con aparición de

espasmos nerviosos, convulsiones, delirio psicótico e incluso gangrena. El ácido Lisérgico (LSD), es uno de los componentes del ergot, actualmente su uso en pequeñas cantidades induce clínicamente la labor de parto, detención de hemorragias uterinas y alivio de la migraña, además de tratar la hipertensión arterial (Ramírez, 2018).

Los hongos también se pueden utilizar como un medio de control biológico, para prevenir los daños ocasionados por plagas de insectos como *Metarhizium* spp, *Beauveria bassiana* y *Lecanicillium lecanii* (Maina *et al.*, 2018; Keppanan *et al.*, 2019) e inclusive de hongos patógenos en plantas como *Magnaporthe* spp, *Botrytis cinerea*, *Fusarium* spp, *Blumeria graminis*, *Colletotrichum* spp, *Stagonospora nodorum*, *Rosellinia necatrix* y *Septoria tritici* (Dean *et al.*, 2012). Son organismos capaces de reemplazar algunos controladores químicos, peligrosos a nivel ambiental.

En cuanto a los desarrollos tecnológicos más recientes, ha aumentado el interés en la producción de proteínas recombinantes a partir de los hongos, ya sea para ser utilizadas como alimento del hombre o del ganado. Se han realizado varios experimentos sobre alimentación con las levaduras *Candida utilis* y *Saccharomyces cerevisiae* (desecada), ya se agrega al pan en un 1 % para enriquecer este alimento con proteína, en la dieta diaria (Zulema, 2019). En procesos industriales importantes, el método principal de producción de ácido cítrico para bebidas embotelladas, etc., se basa en el crecimiento de *Aspergillus niger* en azúcares; este ácido orgánico se acumula como producto secundario del metabolismo del hongo. Las hormonas vegetales Giberelinas, son otro grupo de metabolitos producidos comercialmente para la industria hortícola a partir de los hongos. Las cuatro principales enzimas fúngicas obtenidas comercialmente son: alfa-amilasas, aminoglucosidasa, pectinasas y proteasas. La primera se usa para convertir el almidón en maltosa y maltotriosa durante la elaboración del pan y se usa junto con la amiloglucosidasa para obtener maltosa y glucosa para jarabes de alto contenido energético a base de raíz y de otras materias primas que contienen almidón. Las pectinasas se emplean para clarificar jugos de frutas y vinos (López *et al.*, 2006).

Una proteasa acida específica de *Rhizomucor miehei*, se utiliza como sustituto para coagular leche en la elaboración de quesos (Castañón, 2019).

2.1.4 Clasificación de los hongos

Los criterios de clasificación de los hongos se disponen metódicamente, según una jerarquía, en reino, divisiones o filos, clases, órdenes, familias, géneros y especies, en la actualidad también se contemplan las características microscópicas **Figura 4**, aunado a esto los recientes trabajos sobre biología molecular han provocado profundos cambios, en la clasificación taxonómica de los hongos (Alonso, 2004).

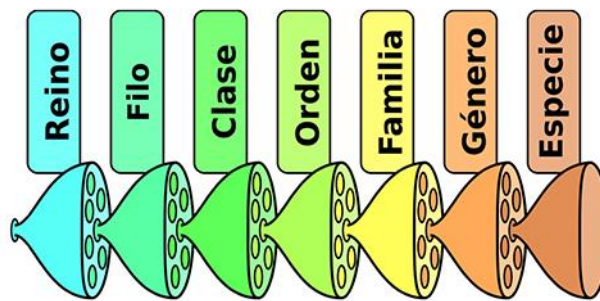


Figura 4. Clasificación micótica más utilizada.

Tomado a partir de:

Alonso, J. (2004). CLASIFICACIÓN Y DESCRIPCIÓN DE LOS HONGOS. "Guía de campo de los hongos de España y de Europa", 21.

Una vez establecidos los criterios taxonómicos con los cuales se clasifica a los hongos, será más fácil entender, por qué se agrupan de esa manera, y por ende comprender sus cualidades, características y su morfología microscópica, en este trabajo se recopila de forma sencilla esta información, para su fácil comprensión al lector, haciendo hincapié únicamente en; reino, división o filo, clase y orden, puesto que hay una gran variedad de familias, géneros y especies de hongos, observando que los organismos, que pertenecen a una misma clase y/o orden tienen características generales y particulares similares (Garcés de Granada, 2003) **Tabla 2.**

Tabla 2. Clasificación actual micótica.

Reino	División/Subdivisión/Filo	Clase/Subclase	Orden
PROTOCTISTA	ACRASIOMYCOTA	Dictyosteliomycetes	Dictyosteliales
		Acrasiomycetidae	Acrasiales
	MIXOMYCOTA	Ceratiomyxomycetidae	-
		Myxogastromycetidae	Liceliales
		Stemonitimycetidae	Physarales

	PLASMIDIOPHOROMYCOTA		Plasmodiophoraceae	Plasmodiophorales		
CHROMISTA	CHYTRIDIOMYCOTA		Chytridiomycetes	Chytridiales		
				Spizellomycetales		
				Neocallimasticales		
				Blastocladales		
	HYPHOCHYTRIDIOMYCOTA		Hyphochytridiomycetes	Hiphocytriales		
	OOMYCOTA		Oomycetes	Peronosporomycetidae	Pythiales	
Saprolegniomycetidae				Saprolegniales		
Labirintolomicetes			Lagenidiales			
				Leptomitales		
FUNGI	ZYGOMYCOTA		Zygomycetes	Saprolegniales		
				Peronosporales		
				Mucorales		
				Entomophthorales		
				Zoopagales		
				Dimargaritales		
			Trychomycetes		Kickxellales	
					Harpellales	
					Asellariales	
					Eccriniales	
					Arthoniales	
					Lichenostigmatales	
	ASCOMYCOTA	PEZIZOMYCOTINA	Arthoniomycetes		Coniocyales	
			Coniocybomycetes		Triblidiales	
			Collemopsidiomycetes		Collemopsidiales	
			Dothideomycetes		Capnodiales	
					Dothideales	
					Hysteriales	
					Myriangiales	
					Patellariales	
					Pleosporales	
			Eurotiomycetes	Chaetothyrionomycetidae	Chaetothyriales	
				Eurotiomycetidae	Pyrenulales	
				Mycocaliciomycetidae	Verrucariales	
					Corneliales	
					Eurotiales	
					Onygenales	
					Mycocaliciales	
			Geoglossomycetes		Geoglossales	
			Laboulbeniomycetes		Laboulbeniales	
					Pyxidiophorales	
			Lecanoromycetes	Acarosporomycetidae	Acarosporales	
				Lecanoromycetidae	Lecanorales	
						Peltigerales
						Teloschistales
				Ostromycetidae	Agyriales	
Baeomycetales						
Ostropales						
Pertusariales						
Incertae sedis	Gyalectales					
	Candelariales					
	Umbilicariales					
	Capnodiales					
		Cyttariales				
		Erysiphales				
		Helotiales				

			Leotiomycetes	Leotiales	
				Rhytismatales	
				Thelebolales	
				Lichinomycetes	Lichinales
				Orbiliomycetes	Orbiliales
			Sordariomycetes	Pezizomycetes	Pezizales
				Boliales	
				Calosphaeriales	
				Coronophorales	
				Diaporthales	
				Halosphaeriales	
				Hypocreales	
				Lulworthiales	
				Meliolales	
				Microascales	
				Ophiostomatales	
				Phyllachorales	
				Sordariales	
				Trichosphaeriales	
			Xylariales		
			Xylonomycetes	Symbiotaphrinales	
	Xylonales				
	SACCHAROMYCOTINA	Hemiascomycetes o Saccharomycetes	Saccharomycetales		
		TAPHRINOMYCOTINA	Archaeorhizomycetes	Archaeorhizomyce- tales	
	Neoelectomycetes		Neoelectales		
	Pneumocystidomycetes		Pneumocystidales		
	Schizosaccharomycetes		Schizosaccharomy- cetales		
	Taphrinomycetes		Taphirales		
	BASIDIOMYCOTA	Urediniomycetes	Uredinales		
		Ustomycetes	Ustilaginales		
		Basidiomycetes	Tremellales		
			Auriculariales		
			Septobasidiales		
Brachybasidiales					
Dacrymycetales					
Tulasnellales					
Exobasidiales					
Aphyllaphorales					
Agaricales					
Hymenogastrales					
Melanogastrales					
Gautieriales					
Phallales					
Lycoperdales					
Tulostomatales					
Sclerodermatales					
Nidulariales					
DEUTEROMYCOTA	Blastomycetos				
	Hyphomycetos				
	Coelomycetos				

Elaboración a partir de: NCBI, pfam y Garces de Granada Emira, M. C. (2003). Mofología y Calsificación de los hongos. Bogotá: Departamento de biología, Facultad de Ciencias, Universidad Nacional de Colombia.

2.1.4.1 Importancia de la división *Ascomycota*

La división *Ascomycota* es la más amplia de todos los hongos, y la mayoría son patógenos del ser humano, entomopatógenos y fitopatógenos, tanto los levaduriformes como los filamentosos. Un ejemplo, es la infección por *Candida*

albicans, *Trichophyton*, *Microsporum*, *Epidermophyton* y *Chrysosporium* (Guarro, 2012; Maina *et al.*, 2018; Keppanan *et al.*, 2019), capaces de desarrollar tiña y otras afecciones dermatológicas, para el caso de las afecciones respiratorias tenemos a *Aspergillus* spp, entre otras más. Para el caso particular de los entomopatógenos tenemos a *Metarhizium* spp. (*M. anisopliae* y *M. acridum*), *B. bassiana* y *L. lecanii*. Hablando de los fitopatógenos los cuales infectan cultivos agrícolas causando pérdidas millonarias en todo el mundo, los organismos más letales son *Magnaporthe* spp. (*M. grisea* y *M. oryzae*), *B. cinerea*, *Fusarium* spp. (*F. graminearum*, *F. oxysporum*, *F. sporotrichioides* y *F. fujikuroi*), *B. graminis*, *Colletotrichum* spp. (*C. trifolii*, *C. fructicola* y *C. graminicola*), *S. nodorum*, *R. necatrix* y *S. tritici* (Dean *et al.*, 2012).

2.2 HONGOS PATÓGENOS DEL SER HUMANO

Se definen como hongos patógenos aquellos organismos que provocan enfermedades a su hospedero (el ser humano), o que incluso pueden ocasionar una muerte prematura.

En los últimos años la taxonomía de los hongos causantes de infecciones en el hombre ha cambiado sustancialmente, especialmente debido al rápido desarrollo de las técnicas de secuenciación de ADN, que han permitido avanzar en el conocimiento de las relaciones filogenéticas entre estos microorganismos. El número de especies fúngicas potencialmente patógenas ha sufrido un aumento, siendo muchas de ellas especies crípticas que forman parte de complejos de especies que han sustituido, a lo que antes eran consideradas como simples morfoespecies; los géneros mucorales y ascomicetes tales como *Aspergillus*, *Fusarium*, *Sporothrix* y *Scedosporium*, son muestra de ello (Guarro, 2012). Es frecuente que dichas especies difieran en virulencia y en su respuesta a los antifúngicos por lo que su correcta identificación es de especial importancia para el adecuado tratamiento del paciente (Guarro, 2012).

A continuación, se dará un breve resumen de estos a través de la **Tabla 3**. En azul claro, se presentan las diferentes especies micóticas con mayor prevalencia clínica

en el ser humano, en azul turquesa, se indica la localización en la que suelen desarrollar su microhábitat.

Tabla 3. Enfermedades micóticas más importantes en medicina.

Localización	Tipo de Hongo	Patología
Superficial	<i>Malassezia furfur</i>	Pitiriasis versicolor
Cutánea	<i>Tricophyton, Epidermophyton, Microsporium</i>	Dermatofitosis (tiña piel, cuero cabelludo y uñas)
	<i>C. albicans</i>	Candidiasis cutánea, y de membranas mucosas
Subcutánea	<i>Sporothrix schenckii</i>	Esporotricosis
Profunda	<i>Blastomyces dermatitides</i>	Blastomicosis
	<i>Coccidioides immitis</i>	Coccidioidomicosis
	<i>Histoplasma capsulatum</i>	Histoplasmosis
	<i>Paracoccidioides brasiliensis</i>	Paracoccidioidomicosis
Profunda oportunista	<i>Cryptococcus neoformans</i>	Criptococosis
	Especies de <i>Mucor</i> y <i>Rhizopus</i>	Cigomicosis

Elaboración a partir de: Pfaller Michael A., Murray Patrick R., Rosenthal Ken S., Kobayashi George S. (2021).

"MICROBIOLOGÍA MÉDICA" (9ª ed). Argentina: ELSEVIER CASTELLANO.

Ahora bien, como anteriormente se había descrito la mayoría de los hongos patógenos humanos son dimórficos, como se ve en la **Tabla 4**, y de igual forma se puede ver que la gran mayoría de los hongos patógenos del ser humano son pertenecientes a la división *Ascomycota*, además de otras características de ellos.

Tabla 4. Clasificación de hongos de importancia médica.

Género	Crecim. típico	Tabicación	Forma sexual	División	Clasif. Medica
<i>Aspergillus</i>	Moho	+	?	<i>Ascomycota</i>	Oportunista
<i>Blastomyces</i>	Dimórfico	+	?	<i>Ascomycota</i>	Sistémico
<i>Candida</i>	Dimórfico	+	?	<i>Ascomycota</i>	Oportunista
<i>Coccidioides</i>	Dimórfico	+	?	<i>Ascomycota</i>	Sistémico
<i>Cryptococcus</i>	Levadura		+	<i>Basidiomycota</i>	Sistémico
<i>Epidermophyton</i>	Moho	+	+	<i>Ascomycota</i>	Superficial
<i>Histoplasma</i>	Dimórfico	+	+	<i>Ascomycota</i>	Sistémico

Microsporium	Moho	+	+	Ascomycota	Superficial
Mucor	Moho	-	+	Zygomycota	Oportunista
Pneumocystis	Quiste		?	Ascomycota	Oportunista
Rhizopus	Moho	-	+	Zygomycota	Oportunista
Sporothrix	Dimórfico	+	?	Ascomycota	Subcutánea
Trichophyton	Moho	+	+	Ascomycota	Superficial

Tomado a partir de: Pfaller Michael A., Murray Patrick R., Rosenthal Ken S., Kobayashi George S. (2021). "MICROBIOLOGÍA MÉDICA" (9° ed). Argentina: ELSEVIER CASTELLANO.

Es bien sabido, que el desarrollo de las nuevas tecnologías en la detección y el diagnóstico de las enfermedades producidas por hongos (Jawetz, 2011) **Figura 5**, facilita la iniciación del tratamiento lo más pronto posible, evitando así el ciclo infeccioso del hongo, ya que los métodos moleculares han permitido conocer la secuenciación de genomas completos, promoviendo estrategias de manejo, control y eliminación de estos, para después si se considera pertinente y se tienen los medios, ser analizados de manera eficaz por medio de un estudio bioinformático que permita su rápida identificación taxonómica, abriendo una gran posibilidad de análisis y estudio de las diferentes especies perteneciente a dicha familia o división del hongo en cuestión.

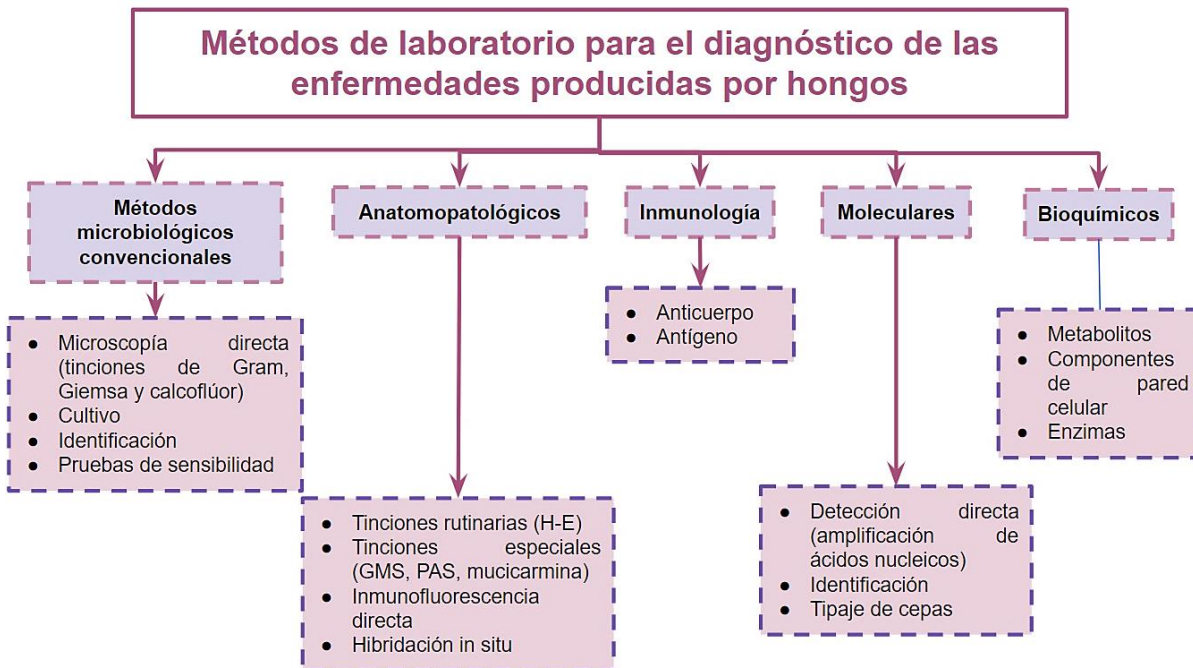


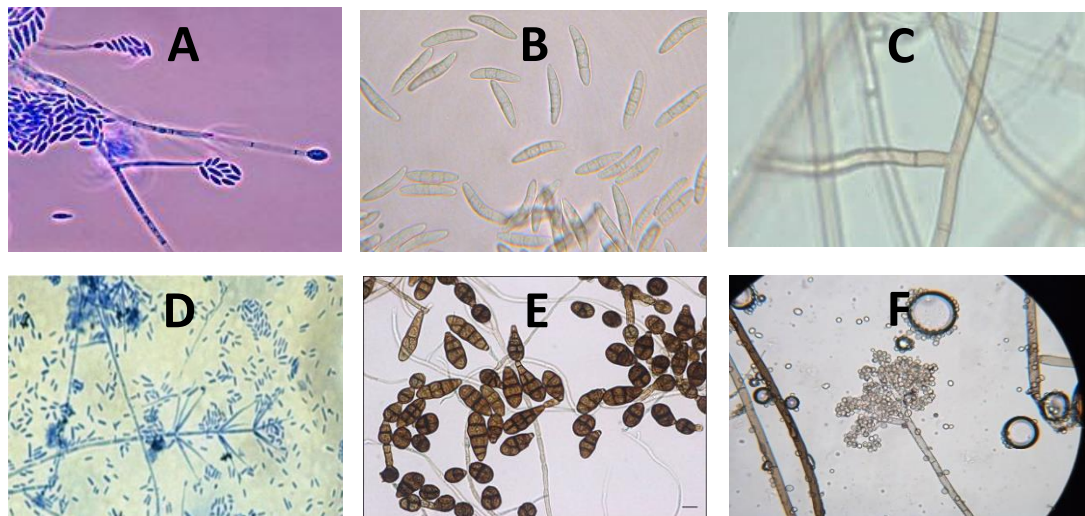
Figura 5. Métodos de laboratorio para el diagnóstico de las enfermedades producidas por hongos.

Elaborado a partir de: Brock. (2015). Biología de los microorganismos (14° ed.). México: PEARSON EDUCACIÓN, Jawetz, M. y. (2011). Microbiología médica (25° ed.). México: MCGRAW-HILL INTERAMERICANA, Pfaller, M. R. (2014). Microbiología médica (Septima ed.). Madrid España: Elsevier España, S.L, Sherris. (2011). Microbiología médica (Segunda ed.). México: MCGRAW-HILL INTERAMERICANA EDITORES, S. A. de C. V.

2.3 HONGOS FITOPATOGENOS

Los hongos que enferman a las plantas se conocen como fitopatógenos y tienen un gran impacto ya que, anualmente, destruyen un tercio de las cosechas producidas. Específicamente, causan la pérdida de cinco alimentos: arroz, trigo, maíz, papa y soya, que a nivel mundial son de los más importantes. Si estos cinco cultivos fueran infectados simultáneamente, a tal grado de que toda la planta se perdiera, más del 60% de la población mundial no tendría qué comer. Dentro de los géneros de hongos que atacan a estos cultivos se encuentran: *Fusarium*, *Giberella*, *Rhizoctonia*, *Verticillium*, *Alternaria* y *Botrytis* entre muchos otros (Carreras, 2022). Dicho en otras palabras, los hongos constituyen el grupo más importante de agentes fitopatógenos debido al número y diversidad de enfermedades que causan. Además, existen hongos patogénicamente especializados sobre una sola especie vegetal o sobre las especies de un solo género. A ello hay que unir su extrema versatilidad parasitaria, indicada por la variedad de especies y órganos vegetales que pueden atacar, la diversidad de síntomas y enfermedades que pueden inducir

Figura 6 (Crous *et al.*, 2021).



A) *Fusarium* spp, B) *Giberella zeae*, C) *Rhizoctonia solani*, D) *Verticillium dahliae*, E) *Alternaria* spp, F) *B. cinerea*

Figura 6. Géneros de hongos fitopatógenos.

Elaborado a partir de: Crous Pedro W., J. Z. (2021). Names of Phytopathogenic Fungi: A Practical Guide. The American Phytopathological Society, 1500–1508.

Además de las condiciones ambientales, las actividades humanas, como el tránsito de mercancías, puede ser la causa del traslado de organismos patógenos a áreas

en las cuales no existían antes, incluso pueden moverse de país en país. Este es el caso de las enfermedades producidas por los complejos ambrosiales, una asociación simbiótica entre un escarabajo y un hongo, en donde el escarabajo se alimenta del hongo y el hongo a su vez se alimenta de una planta o árbol perjudicándolo severamente hasta causar su muerte. El complejo formado por el escarabajo *Euwallacea* y el hongo *Fusarium* es el responsable de la pérdida de muchos huertos de aguacate en California, Estados Unidos. Desafortunadamente, la cercanía que tenemos con este país resultó en la invasión del complejo *Euwallacea-Fusarium sp.* en el norte de nuestro país específicamente en Tijuana, Baja California. Por ello, el desarrollo de estrategias para el control eficiente de hongos fitopatógenos se centra en la podredumbre y enfermedades vasculares de las plantas, que constituyen un reto para el grupo de investigación.

En México existe un plan de acción dispuesto por las autoridades fitosanitarias para la vigilancia y control del complejo ambrosial. Es importante resaltar que en los laboratorios de Biología Molecular y Fitopatología del Instituto de Ecología A. C., se realizan esfuerzos como parte de un gran proyecto multidisciplinario e interinstitucional apoyado por el Fondo Institucional de Fomento Regional para el Desarrollo (FORDECYT) de CONACYT para contribuir en el plan de acción. En donde se pretende entender cómo el hongo *F. ambrosial* infecta a las plantas. Para ello, se usó una tecnología de vanguardia conocida como el sistema CRISPR/Cas9 que sirve para editar genes, es decir permite manipular la información genética de los hongos. Esta técnica ha sido usada en varios sectores, como el alimentario, que se empleó para la generación de champiñones y manzanas que no se oxidan al exponerse al ambiente y en el sector salud está siendo contemplada como terapia génica en ciertas enfermedades humanas (FORDECYT-CONACYT, 2022). Las variantes relevantes y mutantes son estudiadas, evaluando si se modifica su capacidad de infectar a una planta, su crecimiento o la tolerancia a fungicidas. Como consecuencia se han identificado genes con una posible función en la capacidad infecciosa de *F. ambrosial*. Sin embargo, todavía hay camino por recorrer y estos estudios nos permitirán incrementar el conocimiento de la biología del género *Fusarium*, pero lo más importante es su contribución al establecimiento

de estrategias de prevención y control de infecciones de este grupo de hongos, lo cual podría implementarse a otras especies de hongos que infectan otros cultivos de importancia agronómica (FORDECyT-CONACyT, 2022).

2.4 HONGOS ENTOMOPATOGENOS

Se les llama así a los hongos que causan daño a la población de insectos.

El uso indiscriminado de plaguicidas sintéticos es la causa directa de la resistencia de diferentes insectos, y por consiguiente de la pérdida de su efectividad. Ante esto, es común aumentar las dosis y preparar mezclas de varios productos, que con frecuencia resultan ser más tóxicos, por lo que el problema de la resistencia lejos de solucionarse se agrava. El control químico también produce otros efectos como: brotes de plagas secundarias, resurgencia de plagas y disminución de las poblaciones de enemigos naturales. A partir de esa situación, se hace necesario desarrollar métodos de manejo de plagas compatibles con el ambiente, uno de los cuales es el control biológico. He aquí la importancia de los hongos entomopatógenos capaces de inducir o desencadenar una respuesta dañina en los insectos, tienen un gran potencial como agentes controladores, constituyendo un grupo con más de 750 especies, diseminados en el medio ambiente y provocando infecciones fúngicas a poblaciones de artrópodos; Para López-Llorca y Hans-Börje (2001), entre los géneros más importantes están: *Metarhizium*, *Beauveria*, *Aschersonia*, *Entomophthora*, *Zoophthora*, *Erynia*, *Eryniopsis*, *Akanthomyces*, *Fusarium*, *Hirsutella*, *Hymenostilbe*, *Paecilomyces* y *Verticillium*, mientras que para la FAO (2003), los géneros de importancia son *Metarhizium*, *Beauveria*, *Paecilomyces*, *Verticillium*, *Rhizopus* y *Fusarium*.

El control biológico, se define como una práctica agrícola en constante crecimiento que busca la destrucción total o parcial de patógenos e insectos, mediante el uso de sus enemigos naturales (Tellez-Jurado *et al.* 2009). El control biológico ofrece, al mismo tiempo, beneficios a la economía de los agricultores, protección del ambiente y a la salud de los consumidores. Además, ha contribuido al desarrollo de la agricultura de México y de muchos países (Bernal *et al.*, 2015). Los técnicos y productores, intuitivamente, se han percatado de que su uso les permite combatir

plagas contra las cuales se dispone de especies entomófagas o entomopatógenas, a un costo menor que las erogaciones generadas por el empleo de plaguicidas químicos. Por ello, en la actualidad se asume que dicha alternativa constituye (por sus virtudes económicas, ambientales y ecológicas) la estrategia más deseable para el manejo de las poblaciones de organismos patógenos en los ámbitos agrícola, pecuario y forestal.

Al contrario de lo que podría pensarse, esta no es una estrategia nueva para la protección de los cultivos en México. Su historial abarca casi 100 años, durante los cuales se han tenido algunos éxitos importantes. Sin embargo, desde 1990 el papel del control biológico se ha formalizado en los niveles de gobierno federal, estatal y dentro de la comunidad académica. En 1991, se inauguró el Centro Nacional de Referencia para el Control Biológico (CNRCB) en Tecomán, Colima, México, que fue reconocido por la Organización Internacional para el Control Biológico como un centro de referencia internacional. El CNRCB tiene la misión de desarrollar y establecer estrategias de control biológico para plagas reglamentadas, para lo cual genera y proporciona tecnología alternativa al uso de plaguicidas químicos. De igual modo, coadyuva con programas o campañas fitosanitarias en las que se promueve el uso de organismos benéficos como agentes de control biológico, todo ello con la finalidad de fortalecer la sanidad de los cultivos vegetales en México (CNRCB, 2018).

En la **Figura 7** se resumen las principales investigaciones llevadas a cabo en los cultivos más importantes del país, en las que se utilizaron hongos entomopatógenos. Destaca el amplio estudio de cepas de *B. bassiana*, especialmente, en el combate de plagas como la broca del café, en las que observaron una efectividad patogénica de 100 % a las 144 h; registraron que la época óptima para la aplicación del hongo es en julio, antes de que la broca penetre en el fruto de café (Pacheco, 2019).

Hongo	Plaga	Cultivo	Referencia
<i>Beauveria bassiana</i> (Bals.-Criv.) Vuill.	Broca del café: <i>Hypothenemus hampei</i> Ferrari, 1867 (Coleoptera: Scolytinae)	Café	Gerónimo_2016
	Broca del café: <i>H. hampei</i> Ferrari, 1867	Café	Díaz y Roblero_2007
	Depredadores de <i>Diaphorina citri</i> Kuwayama: <i>Ceraeochrysa valida</i> Banks y <i>Eremochrysa punctinervis</i> McLachlan	Cítricos	Gandarilla et al._2013
	Chinche ligus: ninfas de <i>Lygus lineolaris</i> Palisot de Beauvois	Fresa	González et al._2010
	Conchuela del frijol: <i>Epilachna varivestis</i> Mulsant	Frijol	Castrejón_2017
	Chapulines: <i>Brachystola magna</i> Girard y <i>B. mexicana</i> Bruner	Frijol	Lozano y España_2006
	Mosquita blanca: <i>Bemisia tabaci</i> Gennadius	Hortalizas	Ruiz_2009
	Insecto plaga de los frutos de <i>Jatropha curcas</i> L.: <i>Pachycoris torridus</i> Scopoli	<i>Jatropha curcas</i>	Chávez_2016
	Palomilla del manzano: <i>Cydia pomonella</i> L.	Manzana	Solís_2006
	Picudo del nopal: <i>Metamasius spinolae</i> Gyllenhal	Nopal	Sánchez et al._2016
Picudo de nopal: <i>M. spinolae</i> Gyllenhal	Nopal	Tafoya_2004	
<i>B. bassiana</i> (Bals.-Criv.) Vuill. y <i>Metarhizium anisopliae</i> (Metchnikoff) Sorokin	Mosca blanca: <i>B. tabaco</i> Gennadius	Más de 500 ornamentales	García et al._2013
	Picudo del agave: <i>Scyphophorus interstitialis</i> Gyllenhal	Agave pulquero, tequilero y mezcalero	Aquino_2006
	Gusano barrenador: <i>Diatrea magnifactella</i> Dyar, 1911	Caña de azúcar	Castro et al._2013
	Bacteria <i>Candidatus Liberibacter</i> spp.	Cítricos	Mellin et al._2016
	Larvas de gallina ciega: <i>Phyllophaga crinita</i> (Burm.) (Coleoptera: Melolonthidae)	Maíz, áreas verdes, campos de golf	Nájera_2005
	Gallina ciega: <i>Phyllophaga vetula</i> Horn	Maíz	Hernández et al._2011
	Insectos escama o escamas (insectos chupadores de savia): <i>Aulacapsis tubercularis</i> Newstead	Mango	Pérez-Salgado_2013
	Moscas de la fruta: <i>Anastrepha obliqua</i> (Macquart)	Mango	Díaz-Ordaz et al._2010
	Psílido de la papa: <i>Bactericera cockerelli</i> Šulc.	Papa	Villegas_2017
	Gusano barrenador: <i>D. magnifactella</i> Dyar, 1911	Caña de azúcar y maíz	Buenosaires_2013
<i>M. anisopliae</i> (Metchnikoff) Sorokin	Chapulines: <i>Sphenarium purpurascens</i> (Charpentier) y <i>Melanoplus differentialis</i> (Thomas) (Orthoptera: Acrididae)	Maíz y frijol	Tamayo_2009
	Langosta Centroamericana: <i>Schistocerca piceifrons piceifrons</i> Walker	Maíz, sorgo, frijol, caña de azúcar, soya, algodón, ajonjolí, plátano y cacahuete	Barrientos_2005
<i>M. anisopliae</i> (Metchnikoff) Sorokin y nematodos entomopatógenos	Larvas de gallina ciega: <i>Phyllophaga vetula</i> Horn	Maíz	Ruiz_2012
<i>B. bassiana</i> (Bals.-Criv.) Vuill., <i>M. anisopliae</i> (Metchnikoff) Sorokin y <i>P. fumosoroseus</i> (Wize) Brown & Smith	Picudo: <i>Anthonomus fulvipes</i> Boheman	Cereza de Barbados (Acerola)	Lezama et al._1997
<i>B. bassiana</i> (Bals.-Criv.) Vuill., <i>M. anisopliae</i> (Metchnikoff) Sorokin y <i>P. fumosoroseus</i> (Wize) Brown & Smith	Mariposa blanca de la col: <i>Pieris rapae</i> Linnaeus, gusano dorso de diamante: <i>Plutella xylostella</i> Linnaeus, gusano falso medidor: <i>Trichoplusia ni</i> Hübner, pulgón de la col: <i>Brevycorine brassicae</i> Linnaeus	Hortalizas	García y González_2010
<i>B. bassiana</i> (Bals.-Criv.) Vuill., <i>Lecanicillium lecanii</i> (Zimm.) Zare & W. Gams. y <i>P. fumosoroseus</i> (Wize) Brown & Smith	Virus de la tristeza de los cítricos. Pulgón café de los cítricos: <i>Toxoptera citricida</i> Kirkaldy	Cítricos	Hernández et al._2007
<i>B. bassiana</i> (Bals.-Criv.) Vuill., <i>M. anisopliae</i> (Metchnikoff) Sorokin, <i>N. rileyi</i> (Farlow) Samson, <i>P. fumosoroseus</i> (Wize) Brown & Smith y <i>P. javanicus</i> (Friederichs & Bally) Brown & Smith	Gusano cogollero del maíz: <i>Spodoptera frugiperda</i> (J. E. Smith)	Maíz	Lezama et al._1996a o b
<i>N. rileyi</i> (Farlow) Samson y <i>P. fumosoroseus</i> (Wize) Brown & Smith	Larvas del gusano cogollero del maíz: <i>S. frugiperda</i> (J. E. Smith)	Maíz	Lezama et al._1994
<i>H. citrifomis</i> Speare	Huanglongbing (HLB). Psílido asiático de los cítricos: <i>Diaphorina citri</i> Kuwayama	Cítricos	Pérez-González_2013
<i>Trichoderma</i> spp.	Pudrición radical en cebolla: <i>Fusarium oxysporum</i> Scltdl.	Cebolla	Pulido-Herrera et al._2012
<i>Verticillium</i> y <i>P. fumosoroseus</i> (Wize) Brown & Smith	Mosca blanca: <i>Bemisia argentifolii</i> Bellows y Perring	Algodón	Ceceña et al._2017

Figura 7. Principales hongos entomopatógenos utilizados para el control de plagas en el sector agrícola.

Tomado a partir de: Pacheco Hernández Ma. de Lourdes, J. F. (2019). Organismos entomopatógenos como control biológico en los sectores agropecuario y forestal de México: una revisión. Revista mexicana de ciencias forestales, 10(56).

2.5 MECANISMOS PATOGENICOS DE LOS HONGOS

En términos generales el mecanismo patogénico de los hongos (Téllez-Jurado *et al.*, 2009) (García-Vidal, 2014), se puede observar en la siguiente **Figura 8**:

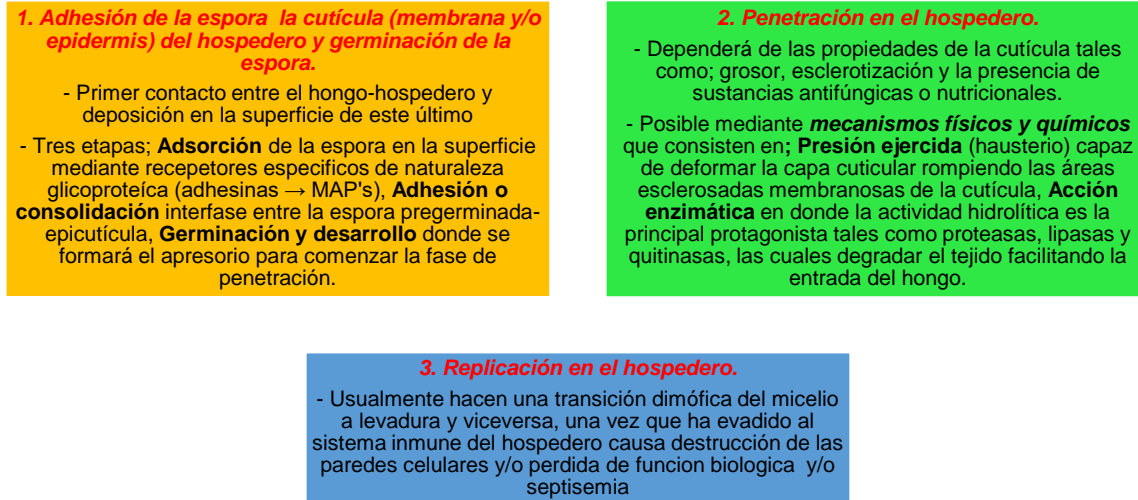


Figura 8. Bosquejo general del mecanismo patogénico de los hongos.

Elaborado a partir de: Téllez-Jurado Alejandro, M. G.-C. (2009). Mecanismos de acción y respuesta en la relación de hongos entomopatógenos e insectos. Universidad Politécnica de Pachuca, 75-77., García-Vidal Carolina, M. S. (2014). Inmunopatología de las micosis invasivas por hongos filamentosos. Revista Iberoamericana, 220-224.

De manera particular los hongos **fitopatógenos** invaden a las plantas si se cumplen las siguientes condiciones:

1. La planta es susceptible
2. El hongo es virulento
3. El sistema planta-hongo es compatible.

El ciclo de la enfermedad está representado por la siguiente imagen **Figura 9**:

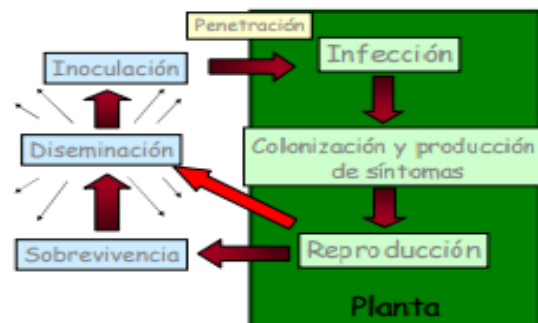


Figura 9. Ciclo de la enfermedad en hongos fitopatógenos.

Tomado a partir de: Scattolini, A. (2010). Hongos fitopatógenos II. Uruguay.

Infección primaria, producida por esporas de origen sexual que provienen de la forma de sobrevivencia de cultivos o plantas vecinos, es decir el inoculo proviene de otra planta a la que el hongo ha enfermado previamente o incluso matado, ha sido dispersado por el medio ambiente ya sea por el viento, rozamiento, etc.

Infección secundaria es propiamente la colonización y producción de los síntomas, ya que esta se produce en los tejidos del organismo, originando las lesiones y los síntomas de la enfermedad.

La reproducción comienza cuando toda la planta tiene en sus tejidos las estructuras del hongo ya sea el micelio o las esporas. Cuando las esporas salen del organismo inicia la etapa de la sobrevivencia, estas por efecto del medio, ya sea el viento, el agua, los insectos, etc., son diseminadas, si el número de esporas es el necesario para infectar una nueva planta se produce la inoculación reiniciando con este un nuevo ciclo. (Scattolini, 2010)

En el caso de los **entomopatógenos**, el desarrollo de la enfermedad en el insecto está dividido en tres fases:

1. Adhesión y germinación de la espora en la cutícula del insecto.

El proceso de adhesión de la espora a la cutícula del insecto, está mediado por la presencia de moléculas sintetizadas por el hongo denominadas adhesinas. En el entomopatógeno *M. anisopliae*, se ha descrito un tipo de adhesina denominada MAP1 la cual se localiza en la superficie de los conidios. La expresión heteróloga de MAP1 en *S. cerevisiae*, le confiere a la levadura propiedades adherentes específicamente a la cutícula de los insectos. La disrupción del gen que codifica para MAP1 afecta la germinación y la formación de blastosporas, así mismo reduce considerablemente la virulencia del hongo. Por otro lado, se ha demostrado que los iones divalentes como el Ca^{2+} y el Mg^{2+} reducen las fuerzas de repulsión electrostáticas promoviendo la adhesión de las esporas (Téllez-Jurado *et al.*, 2009).

2. Penetración en el hemocele.

Esta se lleva a cabo a la acción combinada de dos mecanismos uno físico y uno químico, el primero consiste en la presión ejercida por una estructura fúngica

denominada haustorio, la cual deforma primeramente la capa cuticular rompiendo luego las áreas esclerosadas y membranosas de la cutícula. El mecanismo químico consiste en la acción enzimática, principalmente de actividades hidrolíticas, las cuales degradan el tejido en la zona de penetración, facilitando la entrada del hongo. Estudios *in vitro* indican que en la digestión del integumento sigue una secuencia de lipasa–proteasa–quitinasa. Se ha observado que la acción enzimática puede ser coadyuvada por la secreción de ácidos orgánicos, como el ácido oxálico (Téllez-Jurado *et al.*, 2009).

3. Desarrollo del hongo, que generalmente resulta en la muerte del insecto.

La mayoría de los hongos realizan transición dimórfica de micelio a levadura y una vez que han evadido el sistema inmune del insecto, se produce una septicemia.

La micosis induce a síntomas fisiológicos anormales en el insecto tales como convulsiones, carencia de coordinación, comportamientos alterados y parálisis. La muerte sobreviene por una combinación de efectos que comprenden el daño físico de tejidos, toxicosis, deshidratación de las células por pérdida de fluido y consumo de nutrientes (Téllez-Jurado *et al.*, 2009).

Ya en el interior del insecto, los hongos deben enfrentarse con los mecanismos de respuesta del sistema inmune para lo cual han desarrollado estrategias defensivas e inmunosupresoras, como la producción de toxinas o cambios estructurales en su pared celular. Un tema muy común en la literatura de patología de insectos es la producción de toxinas por diversas especies de hongos entomopatógenos (Téllez-Jurado *et al.*, 2009).

Para los hongos **patógenos del hombre** podemos decir que de la amplia variedad de esporas que se depositan en la piel o que son inhaladas por los pulmones, la mayoría de los tipos no causan infección. Algunos tipos causan infección solo en las personas que presentan alguna de las siguientes características:

- Sistema inmunitario debilitado
- Material extraño, como dispositivos médicos implantado en el cuerpo (por ejemplo, prótesis articulares y válvulas cardíacas artificiales)

El sistema inmunológico puede verse debilitado cuando las personas toman fármacos que lo deprimen (inmunodepresores), como antineoplásicos (quimioterápicos) o fármacos para prevenir el rechazo a un órgano trasplantado, o bien cuando padecen un trastorno que causa inmunodeficiencia, como el sida. Las personas que pasan muchos días en una unidad de cuidados intensivos pueden desarrollar un sistema inmunológico debilitado debido a diferentes procedimientos médicos, trastornos subyacentes, desnutrición o una combinación de los anteriores.

Excepto en el caso de algunos trastornos cutáneos superficiales, las infecciones fúngicas rara vez se transmiten de persona a persona.

Las infecciones por hongos son:

- Oportunistas
- Primaria

Las infecciones fúngicas pueden afectar únicamente un área del cuerpo (localizada) o diversas áreas del cuerpo (sistémica).

Las infecciones fúngicas localizadas afectan solo a un área del cuerpo. Suelen afectar la piel y las uñas, la vagina, o la boca y puede ocurrir en personas que tienen un sistema inmunológico normal o debilitado.

Las infecciones fúngicas localizadas suelen aparecer cuando se altera el equilibrio normal que mantiene los hongos bajo control. Por ejemplo, es normal que ciertos tipos de hongos (por ejemplo, *Candida*) estén presentes en las superficies del cuerpo o en el intestino. Las bacterias que en condiciones normales están presentes en el sistema digestivo y en la vagina restringen el crecimiento de estos hongos en esas zonas. Cuando una persona toma antibióticos, pueden ser eliminadas también las bacterias útiles, lo que favorece que los hongos proliferen de manera incontrolada. El excesivo crecimiento de los hongos provoca síntomas, que suelen ser leves. Al crecer las bacterias de nuevo, el equilibrio se restablece y el problema, por lo general, se resuelve. Ejemplo de estos está la caspa, el pie de atleta, la onicomicosis.

Las infecciones fúngicas sistémicas afectan órganos como los pulmones, los ojos, el hígado y el cerebro, además, también pueden afectar la piel. Hablamos aquí de la Aspergilosis, la Candidiasis y la Mucormicosis (Barcia-Guerrero, 2022).

Ciertamente la forma de hospedarse de los hongos en general es la misma, aunque uno de los factores que influye en su propagación y proliferación tiene que ver con la inmunocompetencia del hospedador (planta, insecto, animal y seres humanos) como más adelante se verá en este trabajo.

2.6 MECANISMOS DE DEFENSA DEL HOSPEDERO

Ya en el interior, dando continuidad al mecanismo patológico de los hongos deben enfrentarse con los mecanismos de respuesta del sistema inmune para lo cual han desarrollado estrategias defensivas e inmunosupresoras, como la producción de toxinas o cambios estructurales en su pared celular. Un tema muy común en la literatura patológica con la producción de toxinas. Considerada la segunda línea de defensa que constituye el sistema inmune innato conformado por el sistema celular y el sistema humoral, capaz de reaccionar ante la invasión de patógenos diferenciando lo propio de lo extraño. En este proceso participan los sistemas de reconocimiento de patrones moleculares característicos de polisacáridos microbianos presentes en la pared celular como: peptidoglicanos, abundantes en bacterias Gram (+); liposacáridos en la membrana externa de bacterias Gram (-) y finalmente α -1,3 glucanos en la pared celular de los hongos. El sistema humoral utiliza proteínas antibióticas y otras moléculas efectoras que circulan y/o cutícula con la finalidad de inactivar los agentes patógenos que accedan al interior. Como un ejemplo particular se ha descrito que los insectos sintetizan moléculas con acción biocida específica tales como cecropinas, defensinas, atacinas, lisozimas, entre otras (Téllez-Jurado *et al.*, 2009). El sistema celular, por su parte, está compuesto por los fagocitos y/o sus equivalentes celulares para los distintos hospederos que a su vez circulan en este, capaces de reconocer los elementos extraños mediante receptores tipo Toll que activan la producción de varios péptidos antimicrobianos. Cuando la concentración de microorganismos patógenos es baja, la fagocitosis es el principal mecanismo para eliminar a los invasores. A concentraciones mayores,

se forman agregados denominados nódulos. Existen evidencias que indican que las proPO (pro-fenoloxidasas) juegan un papel muy importante en este sistema, las cuales son enzimas clave para la síntesis de melanina, polímero que suele depositarse sobre los patógenos, formando los encapsulados, dicha reacción es regulada mediante una cascada de señales que involucra el complemento (García-Vidal, 2014).

Esto claro hablando en el caso particular de que el hongo haya podido penetrar con éxito la epidermis, dicho esto la cutícula es la primera línea de defensa al ser una estructura rígida que recubre la parte externa conformada por dos capas: la epicutícula y la procutícula. La primera está compuesta principalmente por grasas, ceras y lipoproteínas, cuya función es evitar la pérdida de agua por la transpiración. La segunda es la más abundante ya que constituye el 95% de esta estructura y está compuesta por diversas proteínas estructurales (queratina, quitina y/o el equivalente del huésped) que proporcionan rigidez, actuando como una barrera física a la penetración de los patógenos. Además, la epidermis secreta moléculas que actúan de manera específica inhibiendo los mecanismos de infección. Se ha reportado que en la cutícula se da la producción de proteasas, peptidasas antifúngicas e inhibidores de proteasas fúngicas que podrían tener un papel importante durante la infección, además de la presencia de ácidos grasos de cadena corta y lípidos de cutícula que inhiben la germinación de las esporas de los hongos en la piel (Téllez-Jurado *et al.*, 2009), para el caso particular de invasión orofaríngea la principal defensa es la actividad mucosa hidrolítica ciliar que se manifiesta como; proteasas, lipasas y quitinasas, las cuales degradan el tejido dificultando la entrada del hongo (García-Vidal, 2014).

Otras estrategias de defensa incluyen el modificar su comportamiento con el propósito de luchar contra los agentes patógeno como es el caso de incrementar la temperatura corporal por exposición al sol deteniendo el desarrollo de las blastosporas del hongo y facilitando la acción de los fagocitos y sus equivalentes en el huésped de su sistema inmune.

Una vez entendiendo todo esto podemos simplificar en un **Figura 10**, el proceso por el cual se desarrolla la patogenia micótica y los diferentes mecanismos de defensa contra ellos.

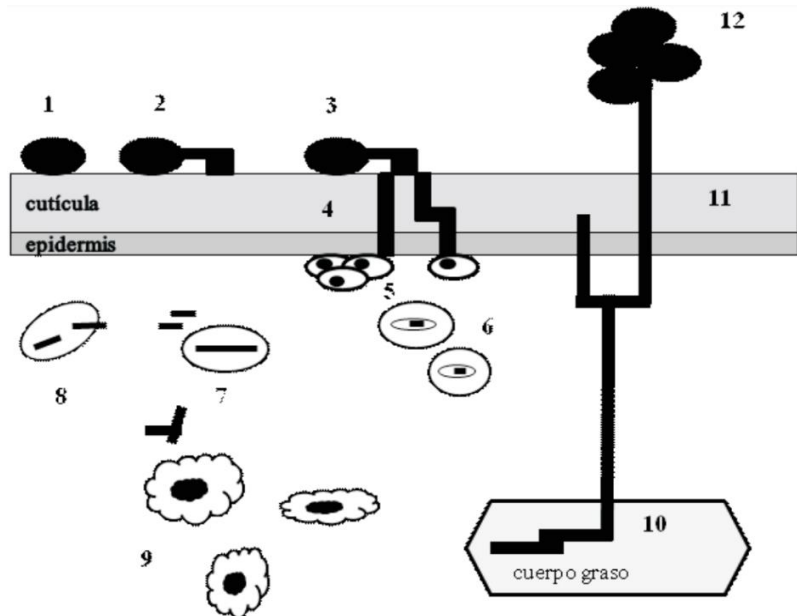


Figura 10. Esquema del desarrollo de un hongo.

1. Adhesión de la esporeña a la cutícula, 2. Germinación y formación del apresorio, 3. Penetración de la cutícula, 4. Crecimiento lateral y penetración en la epidermis, 5. Agregación de los fagocitos en el lugar de penetración fúngica, 6. Fagocitosis de cuerpos hifales por células fagocíticas, 7. Transformación a cuerpos levaduriformes, 8. Evasión del sistema inmune, 9. Propagación en el huésped, 10. Transformación a cuerpo hifal, 11. Esporulación y germinación atravesando la cutícula del huésped, 12. Diseminación de las esporas.

Elaborado a partir de: Téllez-Jurado Alejandro, M. G.-C. (2009). Mecanismos de acción y respuesta en la relación de hongos entomopatógenos e insectos. Universidad Politécnica de Pachuca, 75-77., García-Vidal Carolina, M. S. (2014). Inmunopatología de las micosis invasivas por hongos filamentosos. Revista Iberoamericana, 220-224.

2.7 VIRULENCIA

La virulencia se define como (Beceiro, 2012) **Figura 11**;

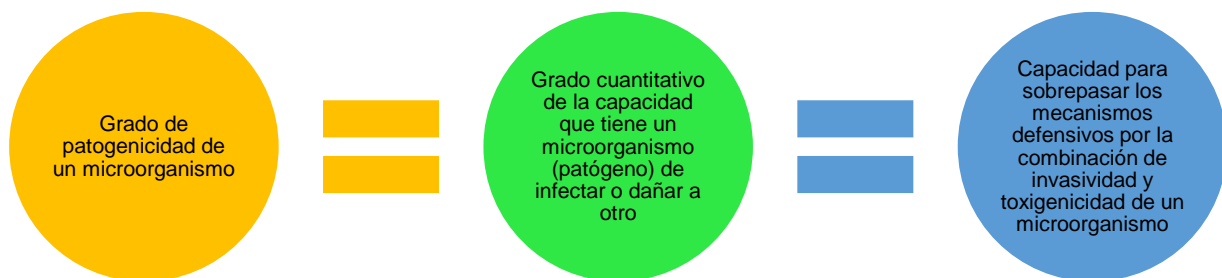


Figura 11. Recopilación de las diferentes definiciones de Virulencia.

Elaborado a partir de: Beceiro Alejandro, M. T. (2012). Resistencia a los antimicrobianos y virulencia, ¿una asociación beneficiosa para el mundo microbiano? Elsevier, 493-495.

2.7.1 Factores de virulencia

Un factor de virulencia se explica como cualquier componente que tiene un patógeno para causar daño (invasión, infección, modulación de la respuesta inmune a su favor y dificultad en el tratamiento contra ellos) a un microorganismo que a su vez permita el desarrollo y reproducción del patógeno (De la Calle, 2012). Los factores de virulencia son principalmente de 2 tipos: los expresados en la superficie celular (que desempeñan funciones de adhesión e invasión de tejidos además de formación de biopelículas e inducción de citosinas) y los producidos dentro del hongo y que son exportados al sitio de infección (Miranda *et al.*, 2016). Ya entendido el concepto y sus vertientes daré mención a algunos ejemplos de factores de virulencia conocidos (Téllez-Jurado *et al.*, 2009) (Martínez-Cruz, 2021) (Martínez, 1995), véase en la **Tabla 5**.

Tabla 5. Factores de virulencia conocidos.

Factor de virulencia	Huésped	Tipo/Función
α - glucanasa en la pared celular	Humano	Superficie celular/ permite la fácil penetración de la cutícula y la epidermis
Proteínasa g	Humano	Superficie celular/ digerir elastina y colágena, rompe IgA secretoria e IgG que permite la fácil penetración.
Actividad antifagocítica de la pared externa del atroconidio: Drutz y Huppert 2	Humano	Producidos dentro del hongo/ No permite la actividad fagocítica y de PMN
Mucopolisacáridos de la cápsula	Humano	Superficie celular/ No permite la actividad fagocítica
Fenoloxidasa	Humano	Producidos dentro del hongo/ permite la fácil penetración
Presencia de manoproteínas (60 kDa) parecidas a la lectina	Humano	Producida dentro del hongo/ permite la fácil identificación y adhesión de las células epiteliales
Efectores con actividad quitinasa (proteínas de 9 miembros EWCA)	Planta	Producidos dentro del hongo/ romper los oligómeros inmunogénicos de quitina dando lugar a fragmentos más pequeños, que ya no pueden ser reconocidos por la planta.
Gen MAP1	Insecto	Producidos dentro del hongo/ confiere propiedades adherentes específicas en la cutícula.
Iones Ca^{2+} y Mg^{2+}	Insecto	Producido dentro del hongo/ disminuyen las fuerzas de repulsión electrostática promoviendo la adhesión de las esporas.

Elaborado a partir de: Téllez-Jurado Alejandro, M. G.-C. (2009). Mecanismos de acción y respuesta en la relación de hongos entomopatógenos e insectos. Universidad Politécnica de Pachuca, 75-77., Martínez-Cruz Jesús, D. R.-G. (2021). Effectors with chitinase activity (EWCA), a family of conserved, secreted fungal chitinases that suppress chitin-triggered immunity. THE PLANT CELL, 2-10., Matínez, R. L. (1995). Los mecanismos patogénicos y de protección en las micosis. Gac Méd Méx, 132(1), 49-51.

2.7.1.1 Factor de Virulencia Egh16 (gen Egh16) → (IPRO21476)

Gen o grupo de proteínas (proteomas) fúngicas que se encuentran en algunos hongos patógenos que pueden desempeñar un papel importante durante la etapa inicial de la infección (penetración en el hospedador y el desarrollo de la lesión) (InterPro, 2021).

2.8 BIOINFORMÁTICA Y PROTEOMICA

La gran diversidad de hongos, de importancia para el ser humano y su ecosistema, son pocos comparados con la totalidad en el reino fungí, por lo que su estudio a profundidad es un trabajo arduo. En las últimas dos décadas las plataformas de secuenciación masiva y la bioinformática han permitido la generación de cientos de secuencias de genomas completos de hongos incluidos fitopatógenos, entomopatógenos y patógenos del ser humano, abriendo la puerta para mejorar estrategias de manejo, control y eliminación de los mismos, teniendo grandes avances en el manejo y control de plagas indeseadas para la industria alimentaria, farmacéutica y cosmética principalmente (entomopatógenos), evitando enfermedades (patógenos del ser humano), pérdidas de cultivos importantes (fitopatógenos) (Bilal, 2017).

La bioinformática es una ciencia de naturaleza interdisciplinaria, ya que se fundamenta en la biología, la fisicoquímica, las matemáticas, la estadística, la probabilidad y la informática, aplicando herramientas de computación y análisis datos biológicos, con el fin de organizar, analizar, interpretar, visualizar y almacenar esta información a gran escala. La ayuda a la comunidad científica de todo el mundo se establece al acceder fácilmente a la información, además de crear y/o optimizar herramientas para analizar e interpretar datos de forma masiva, eficaz, de fácil acceso y practica de usar (Martínez, 2007; Gauthier *et al.*, 2018).

La proteómica es una tecnología en desarrollo que investiga la estructura y función del conjunto de proteínas que conforman el proteoma. El término proteoma apareció en 1994 como un equivalente lingüístico del concepto de genoma que describe el conjunto completo de proteínas que se expresan, según el genoma, y factores

externos, durante la vida de una célula. Los objetivos de la proteómica son: 1) El estudio de los cambios globales de la expresión de las proteínas celulares en el tiempo (evolución en la dinámica celular) y la determinación de la identidad y las funciones de todas las proteínas producidas por los organismos (Torrades, 2004).

La investigación se centra en métodos exactos y rápidos, para identificar y caracterizar proteínas: electroforesis bidimensional en gel y espectrometría de masas, etc., y ya con ayuda de softwares especializados (*Andromeda, Mascot, IsobariQ, MaxQuant etc*, descritos en Chen *et al.*, 2020) y herramientas bioinformáticas, obtienen proteomas de alto rendimiento al manejar de forma masiva y simultanea los datos (Hernández *et al.*, 2019;) y obtener asignación de espectros de secuencias peptídicas, la validación y cuantificación a nivel de péptidos y proteínas, anotación y caracterización funcional, así como el almacenamiento de los resultados en una base de datos, además de la secuenciación masiva de genomas completos de hongos incluidos fitopatógenos, entomopatógenos y patógenos del ser humano, como se mencionó anteriormente (Keerthikumar, 2017).

La herramienta bioinformática más utilizada es la plataforma BLAST “La Herramienta Básica de Búsqueda de Alineación Local”, que encuentra regiones de similitud local entre secuencias, para inferir relaciones funcionales, evolutivas e identificar miembros de familias de genes entre las diferentes secuencias conocidas (NCBI, 2020), se conforma de varios programas de búsqueda por similitud, que se han creado para explorar todas las bases de datos de secuencias, los cuales son (McGinnis *et al.*, 2004):

- **BLASTP:** Compara secuencias de aminoácidos VS secuencia proteica de la base de datos.
- **BLASTN:** Compara secuencias nucleotídicas VS secuencias de ADN de los bancos de datos.
- **BLASTX:** Compara una secuencia de nucleótidos VS proteína de la base de datos.
- **TBLASTN:** Compara una secuencia proteica VS secuencia nucleotídica de la base de datos.

- **TBLASTX**: Compara los seis marcos de lectura de una secuencia de nucleótidos VS seis pautas de lectura de una secuencia de nucleótidos de la base de datos.

En conclusión, la función principal permite de forma óptima la homología entre proteínas dando el mejor alineamiento posible, pero no determina la mejor estructura.

Otras plataformas bioinformáticas son: *InterPro* (clasificación de las familias de proteínas), *Jalview*, *Clustal Omega*, *Expert Protein Analysis System (ExPASy)*, *European Molecular Biology Open Software Suite (EMBOSS)*, *PROSPECT (PROtein Structure Prediction and Evaluation Computer ToolKit)*, *iTOL*: sitio web del Árbol de la vida interactivo para la visualización y manipulación de árboles filogenéticos, desarrollado y mantenido por el Laboratorio Europeo de Biología Molecular. *Protein Data Bank (PDB)*, *Grupow*: alineación de secuencias de AND y proteínas, *Protein Explorer*, etc. (Jiang, 2013).

Lenguajes de programación informáticos como *Python*, *Perl*, *Java*, etc., se han acoplado para procesar los datos a gran escala y facilitar el análisis. Por ejemplo, el proyecto *BioPhython*: conformada por herramientas *Python* disponibles gratuitamente para biología molecular computacional. Este proporciona el módulo *Bio.Blast* para manejar la operación *NCBI BLAST*, puede ejecutar *BLAST* en una conexión local o a través de una conexión a Internet. Así también cuenta con el módulo *Bio.Entrez (Biophyton)*.

Existen también sitios web bioinformáticos útiles (disponibles gratuitamente en Internet) (Bayat, 2002):

Centro Nacional de Información Biotecnológica (www.ncbi.nlm.nih.gov): mantiene herramientas y bases de datos bioinformáticas. Por ejemplo: proporciona el navegador *Entrez*, sistema integrado de recuperación de bases de datos, que permite la integración de bases de datos de secuencias de proteínas y ADN.

Instituto Europeo de Bioinformática (www.ebi.ac.uk): centro de investigación y servicios en bioinformática; gestiona bases de datos de datos biológicos.

Recurso proteico universal “Universal Protein Resource” (<https://www.uniprot.org/>), recurso completo para secuencias de proteínas y datos de anotación. Las bases de datos de UniProt son UniProt Knowledgebase (UniProtKB), UniProt Reference Clusters (UniRef) y UniProt Archive (UniParc) **Figura 12:**



Figura 12. Pestaña principal de UniProt.

Muestra las bases de datos disponibles, así como los proteomas disponibles de las especies; cuenta con alrededor de 22,121 proteomas de referencia, 282,168 proteomas redundantes y 4. 341 proteomas de organismos eucariotas (<https://www.uniprot.org/>).

- Base de datos de familias de proteínas “Pfam” (<http://pfam.xfam.org/>): proporciona soporte para la clasificación y anotación automatizada de secuencias de proteínas; contiene anotaciones funcionales, referencias bibliográficas y enlaces a bases de datos para cada familia.
- Centro Nacional de Recursos Genómicos (www.ncgr.org/): vincula a los científicos con soluciones bioinformáticas mediante colaboraciones, datos y desarrollo de software.
- Genbank (www.ncbi.nlm.nih.gov/Genbank): almacena y archiva secuencias de ADN tanto de proyectos de genoma de manera globalizada hasta con laboratorios individuales.
- Ensembl (www.ensembl.org/): base de datos de anotaciones automáticas sobre genomas.

- SWISS-PROT (www.expasy.org/sprot/): base de datos de proteínas con datos de secuencias de todos los organismos, que tiene un alto nivel de anotación (incluye función, estructura y variaciones) y es mínimamente redundante (pocas copias duplicadas).

Una forma de introducirse a la bioinformática es a través del uso de cuadernos digitales interactivos, como el cuaderno digital “Jupyter notebook” que es una interfaz web fácil de configurar y de usar, puede ser instalado o simplemente utilizarlo de manera online desde su página oficial, permitiendo compartir código abierto, análisis de datos, visualizaciones interactivas o no, fórmulas matemáticas y otros medios integrados (por ejemplo, videos de YouTube, imágenes y enlaces web), todo en un solo documento con componentes interactivos y narrativos y sobre todo apoyando la reproducibilidad del análisis, ya sea de forma local hasta en otras partes del mundo.

La visualización del entorno es prácticamente sencilla, **Figura 13.**, incluye la elección y el control del núcleo de cálculo “Kernel”: Python, Julia, R., etc., también tiene opciones de guardar la bitácora, insertar o modificar código, ejecutar o detener el proceso para una celda, reiniciar la bitácora, cambiar a texto html, además de incorporar extensiones (Davies *et al.*, 2020).

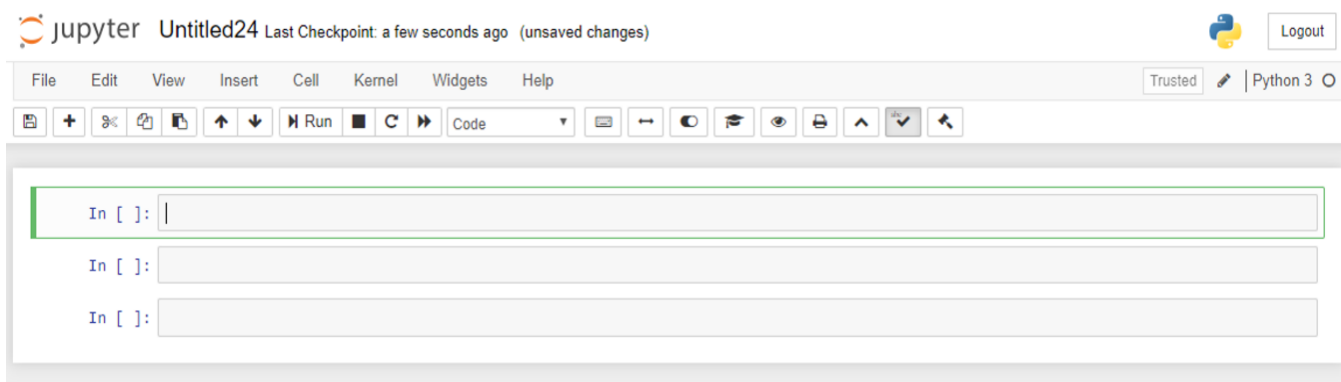


Figura 13. Entorno de Jupyter notebook.

Plantilla del entorno (parte superior) y las celdas (parte inferior) en donde se trabaja y ejecuta el código y permite cambiar a texto para realizar anotaciones (títulos, subtítulos, comentarios, indicaciones etc.) (<https://jupyter.org/>).

3. JUSTIFICACIÓN

Los hongos filamentosos están implicados en áreas estrechamente relacionadas con el ser humano, incluyendo la industria alimenticia (Ghorai *et al.*, 2009),

farmacéutica (Alberti *et al.*, 2017), agrícola (Meyer *et al.*, 2016) y médica (Challa *et al.*, 2015). Dentro de estos hongos filamentosos existen especies patógenas de insectos (entomopatógenas) las cuales son ampliamente usadas como agentes de control biológico de plagas debido a su capacidad para infectar y matar insectos de forma natural, *Metarhizium* spp. (*M. anisopliae* y *M. acridum*), *B. bassiana* y *L. lecanii* son algunas de estas especies (Maina *et al.*, 2018; Keppanan *et al.*, 2019). Por otro lado, también se encuentra un grupo de especies patógenas de plantas (fitopatógenas) las cuales infectan cultivos agrícolas causando pérdidas millonarias en todo el mundo. Dentro de este grupo de hongos filamentosos las especies fitopatógenas más importantes son: *Magnaporthe* spp. (*M. grisea* y *M. oryzae*), *B. cinerea*, *Fusarium* spp. (*F. graminearum*, *F. oxysporum*, *F. sporotrichioides* y *F. fujikuroi*), *B. graminis*, *Colletotrichum* spp. (*C. trifolii*, *C. fructicola* y *C. graminicola*), *S. nodorum*, *R. necatrix* y *S. tritici* (Dean *et al.*, 2012). Además de los hongos entomopatógenos y fitopatógenos, existen hongos filamentosos que comprometen la salud del ser humano (Ullmann *et al.*, 2004). Dentro de estos hongos se incluyen especies como *Aspergillus* spp. (*A. flavus* y *A. fumigatus*), *Penicillium marneffeii*, *Monascus ruber*, *Paracoccidioides* spp. (*P. brasiliensis* y *P. lutzii*), *Blastomyces dermatitidis*, *Coccidioides immitis*, *Trichophyton interdigitale*, *Trichophyton rubrum*, *Fonsecaea pedrosoi*, *H. capsulatum*, *S. schenckii* y *Chaetomium globosum* (Reedy *et al.*, 2007; Köhler *et al.*, 2014; Egbuta *et al.*, 2017). Existen varios estudios que intentan entender y explicar los mecanismos moleculares que controlan la patogenicidad en estos hongos filamentosos, sin embargo, hay mucho camino por recorrer. Uno de estos estudios es la caracterización del factor de virulencia Egh16 (familia de genes Egh16), el cual se ha encontrado principalmente en hongos patógenos de insectos como *M. acridum* (Cao *et al.*, 2012) y *L. lecanii* (Keppanan *et al.*, 2019), y de plantas como *F. oxysporum* (Solano De la Cruz *et al.*, 2021), *B. graminis* (Grell *et al.*, 2003), *M. grisea* (Xue *et al.*, 2004) y *C. fructicola* (Liang *et al.*, 2018), mientras que en cepas no patogénicas como *Neurospora crassa* y *Aspergillus nidulans* esta familia se encuentra ausente (Grell *et al.*, 2003). Al parecer el factor de virulencia Egh16 juega un papel importante en el proceso infectivo de los hongos entomopatógenos y fitopatógenos, sin embargo, no hay evidencia

suficiente que demuestre que todos los hongos patógenos, incluyendo de humanos, tienen este factor de virulencia. Partiendo de estos antecedentes el presente proyecto pretende identificar el factor de virulencia Egh16 en 13 hongos filamentosos patógenos no caracterizados, esto con el fin de aportar conocimiento en la regulación de la patogenicidad para tomar medidas de control mediante alternativas como el uso de la ingeniería genética.

4. OBJETIVOS

4.1 Objetivo general

Analizar mediante un estudio bioinformático la presencia del factor de virulencia Egh16 en 13 hongos filamentosos de la división *Ascomycota*, principalmente en patógenos de insectos, plantas y humanos.

4.1.1 Objetivos específicos

- a) Descarga iterativa de proteomas de hongos filamentosos a partir de la base de datos NCBI y UniProtKB.
- b) Construcción de bases de datos individuales con los proteomas descargados.
- c) Análisis iterativo del dominio Egh16 usando el programa Blastp contra las bases de datos de proteínas.
- d) Procesar e identificar familias de genes y representar gráficamente los resultados.

5. MATERIALES Y MÉTODOS

Se ocuparon bases de datos de ácidos nucleicos y proteínas (NCBI: “National Center for Biotechnology Information”, ENA: “European Nucleotide Archive” y UniProtKB). Revistas científicas especializadas. Paquetes informáticos depositados en PyPI (“Python Package Index”). Uso del entorno interactivo Jupyter Notebook para el procesamiento de datos a gran escala. Uso del programa Blastp. Computadora personal con los siguientes requerimientos: mínimo 4GB de RAM, mínimo 2 GB de espacio en disco, de preferencia sistema operativo Windows 10.

5.1 Programa Informático e Interfaz.

- **Python v3.6.7:** Lenguaje de programación de código abierto (compatible con Windows, Linux y macOS). Uso popularizado en la comunidad científica debido a que es un lenguaje poco prolijo y por ende accesible, caracterizada por su sintaxis simple en inglés, con acciones iterables fáciles y entendibles sin la necesidad de buscar su significado. Disponible en módulos y bibliotecas multipropósito.

Útil para realizar análisis estadísticos, visualizaciones estadísticas e interactivas, edición de textos, flujos de trabajo, protocolos de investigación y reportes.

Fue descargado desde [https://www.python.org/downloads/release/python- 367/](https://www.python.org/downloads/release/python-367/), e instalado con permisos de administrador.

- **Jupyter Notebook:** Interfaz web dinámica de código abierto, estructurada en formato JSON que admite más de 40 lenguajes de programación, incluidos Python, Julia & R, proporciona a los usuarios un entorno apto para crear y ejecutar código, procesar texto con formato, procedimientos matemáticos, imágenes y gráficos, además se puede descargar en formato fasta, txt, HTML, Python, etc. Se utilizó esta interfaz debido a que es flexible y manejable, ya que se puede leer el código en todo momento y ejecutar sus partes de forma independiente.

Instalado desde la terminal, ejecutando el comando “Python -mpip install jupyter”. <https://jupyter.org/install>.

- **BLAST v2.8.1:** Primera versión de producción de BLAST independiente para admitir las nuevas bases de datos BLAST v5 (BLASTDBv5).

La descarga del programa “ncbi-blast-2.8.1+-win64” se realizó desde Jupyter Notebook, utilizando: <https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.8.1/ncbi-blast-2.8.1+-win64.exe>

- **Visual Studio Code:** Editor de código fuente independiente que se ejecuta en Windows, macOS y Linux. Permite la edición, compilación y depuración del código, proporcionando soporte integral para la edición, navegación y comprensión del código junto, un modelo de extensibilidad enriquecido y una integración ligera con las herramientas existentes. Incluye extensiones para admitir casi cualquier lenguaje de programación (C/C++, Java, Python).

Fue descargado desde <https://code.visualstudio.com/Download> e instalado con permisos del administrador, ver instrucciones de instalación para Windows en <https://code.visualstudio.com/docs/setup/windows>.

5.2 Bases de datos

- **UniProtKB:** Eje central para la recopilación de información funcional sobre proteínas, con anotaciones precisas, consistentes y ricas. Además de capturar los datos básicos obligatorios para cada entrada de UniProtKB (principalmente, la secuencia de proteomas de hongos de la división *Ascomycota*, la secuencia de proteomas estándares del factor de virulencia Egh 16, el nombre o la descripción de la proteína, los datos taxonómicos y la información de citas), se agrega la mayor cantidad de información de anotación posible. Esto incluye ontologías biológicas ampliamente aceptadas, taxonomía, clasificaciones y referencias cruzadas, e indicaciones claras de la calidad de la anotación en forma de atribución de evidencia de datos experimentales y computacionales. Se compone de dos secciones: UniProtKB/Swiss-Prot" (revisado, anotado manualmente) y "UniProtKB/TrEMBL" (no revisado, anotado automáticamente).

Fue consultada en línea desde <https://www.uniprot.org/> y se utilizó tal dirección para la descarga a través de Jupyter Notebook.

Utilizada para descargar las secuencias de las proteínas hongos filamentosos de la división *Ascomycota* y las proteínas estándares de hongos con el factor de virulencia Egh 16.

- **NCBI protein:** Base de datos de proteínas es una colección de secuencias de varias fuentes, incluidas traducciones de regiones de codificación

anotadas en GenBank, RefSeq y TPA, así como registros de SwissProt, PIR, PRF y PDB.

Fue consultada en línea desde <https://www.ncbi.nlm.nih.gov/protein/>.

6. RESULTADOS Y DISCUSIONES

Mediante una búsqueda por NCBI y UniProtKB se eligieron 13 hongos de la división *Ascomycota*, fitopatógenos (5), entomopatógenos (4) y patógenos del ser humano (4) **Tabla 6**, a los cuales se les buscó su número taxonómico para la fácil agrupación, identificación y manejo de estos.

Tabla 6. Hongos elegidos de NCBI y UniProtKB.

	No. taxonómico	Hongo	Tipo de patogenicidad
1	1215338	<i>T. interdigitale</i>	Humano
2	1245745	<i>B. bassiana</i>	Entomopatógeno
3	1268274	<i>B. graminis</i>	Fitopatógeno
4	1290391	<i>Botryotinia fuckeliana/ B. cinerea</i>	Fitopatógeno
5	1291518	<i>M. anisopliae</i>	Entomopatógeno
6	1325737	<i>Fusarium sp</i>	Fitopatógeno
7	148305	<i>M. grisea</i>	Fitopatógeno
8	2004951	<i>Cordyceps sp</i>	Entomopatógeno
9	2714763	<i>L. lecanii</i>	Entomopatógeno
10	330879	<i>Aspergillus fumigatus</i>	Humano
11	336722	<i>Zymoseptoria tritici</i>	Fitopatógeno
12	447093	<i>H. capsulatum</i>	Humano
13	559298	<i>B. dermatitidis</i>	Humano

Elaborado a partir de: NCBI y UniProtKB

6.1 Identificación de proteínas implicadas en el factor de virulencia Egh16.

A partir de la base de datos de UniProtKB, NCBI protein y pfam, se identificaron las 7 proteínas que se identifican como ortólogas del factor de virulencia Egh16 que en lo sucesivo se conocerán como Estándares del factor de virulencia Egh16, véase en la **Tabla 7**, así mismo a través de <https://www.ebi.ac.uk/interpro/>, se logró visualizar y extraer su arquitectura. Su importancia radica en que a partir de ello se puede realizar un correcto alineamiento y comparación entre las secuencias (*M. acridum* vs hongos filamentosos), en el que se busca encontrar su ortólogo u ortólogos, considerando: A) pertenecer a la misma división (*Ascomycota*), B) compartir características generales, C) Tener los mismos dominios funcionales, ya que esto indica la funcionalidad de la proteína y nos da una idea del grado de

conservación de la proteína y D) la longitud de la proteína debe ser altamente similar o igual a la otra, ya que ello nos habla de la composición. Estos aspectos son fundamentales para la obtención de una alta puntuación de alineamiento en el estudio y por ende en la identificación de proteínas potencialmente ortólogas en diferentes hongos filamentosos pertenecientes a la división *Ascomycota* **Figura 14-15.**

Tabla 7. Estándares del Factor de virulencia Egh16

	Estándar	Asignación en pfam
1	<i>XP_007807765_M_acridum_CQMa102</i>	E9DV35
2	<i>XP_007806658_M_acridum_CQMa102</i>	E9DRE9
3	<i>XP_007809989_M_acridum_CQMa102</i>	E9E1A1
4	<i>XP_007806786_M_acridum_CQMa102</i>	E9DRS7
5	<i>XP_007811825_M_acridum_CQMa102</i>	E9E6I7
6	<i>XP_007813302_M_acridum_CQMa102</i>	E9EAR4
7	<i>XP_007814411_M_acridum_CQMa102</i>	E9EDX3

Elaborado a partir de: NCBI, UniProtKB, y pfam.



Figura 14. Proteomas ortólogos del factor de virulencia Egh16.

Elaborado a partir de: UniProtKB, e InterPro.

6.2 Construcción de un base de datos BLAST

Se genera un BLAST de acceso con el archivo que contine los proteomas estándares del factor de virulencia Egh16 en formato **.fasta** (Egh16_family.fasta), y todos los archivos de los que se generó este para posteriormente ser guardados en la carpeta (DB2), posteriormente defino en una variable (sec) en forma de lista todos y cada uno de los hongos contenidos en la carpeta (fasta2), de igual manera

por medio de `os.listdir` (“DB2”) establezco una lista de todos los documentos contenidos en la carpeta DB2.



Figura 15. Proteomas del factor de virulencia Egh16.

Elaborado a partir de: UniProtKB, NCBI, pfam, InterPro y sus plataformas auxiliares.

* **Nota:** Los archivos FASTA se modificaron y se concatenaron en un solo archivo, en los encabezados solo quedo el identificador, esto con la finalidad de mejorar el manejo de datos.

Después se ejecutó la celda de código para la creación de la base de datos “DB2/Egh16_family”, véase en la **Figura 16**.

```
In [16]: 1 subprocess.call('makeblastdb -in Egh16_family.fasta -dbtype prot -parse_seqids -out DB2/Egh16_fami
Out[16]: 0
In [17]: 1 sec=os.listdir("fastas2")
In [18]: 1 sec=[i for i in sec if '.fasta' in i]
          2 sec
Out[18]: ['1215338.fasta',
          '1245745.fasta',
          '1268274.fasta',
          '1290391.fasta',
          '1291518.fasta',
          '1325737.fasta',
          '148305.fasta',
          '2004951.fasta',
          '2714763.fasta',
          '330879.fasta',
          '336722.fasta',
          '559298.fasta',
          '655827.fasta']

In [21]: 1 os.listdir("DB2")
Out[21]: ['Egh16_family.phr',
          'Egh16_family.pin',
          'Egh16_family.pog',
          'Egh16_family.psd',
          'Egh16_family.psi',
          'Egh16_family.psq']

In [22]: 1 # n[umero de proteomas
          2 print('proteomas =', len(sec))
          3 print('número esperado de proteínas =', 7 * len(sec))
proteomas = 13
número esperado de proteínas = 91
```

Figura 16. 1er. BLAST a partir de secuencias estándares de Egh16 en formato fasta.

El módulo “subprocess” llama al programa BLAST, a continuación, se construye la base de datos de las proteínas de interés mediante “makeblastdb”, a partir del multifasta. “in”: el multifasta (“Egh16_family.fasta”) con las secuencias que forman la base de datos, “dbtype prot”: tipo de datos: proteínas, “parse_seqids”: Identificadores de secuencia delimitados por barras de análisis en la entrada FASTA. “out”: el nombre que se le puso a la base de datos, en este caso es “Egh16_family” y se sitúa en la carpeta DB2.

Esto con el fin de aguardar y recopilar la información ya antes mencionada de los 13 hongos y los estándares conocidos del factor de virulencia Egh16 proporcionados por *M. acridum*, posteriormente dicha información se filtra y ordena en columnas de forma tabular a través de un bucle, y ser guardada, como puede verse en la **Figura 17**.

```
In [23]: 1 #En donde se pretende filtrar información como;
2 #qacc", "sacc", "qlen", "slen", "length", "qstart", "qend",
3 #"sstart", "send", "score", "bitscore", "evalue", "pident", "nident", "mismatch", "positive", "gaps", "gapopen
4 #de los archivos de los 13 hongos seleccionados, además de guardar dicha información en la carpeta Tables2 en formato .txt
5 init = time_ini()
6 #-----
7 for input_file in sec:
8     print(input_file)
9     subprocess.call('blastp -db DB2/Egh16_family -query fastas2/'+input_file+\
10         ' -evalue 1E-6 -outfmt "6 qacc sacc qlen slen length qstart qend sstart send score bitscore evalue pident nident mismatch
11
12     #-----
13     time_fin(i = init)
14
```

```
1215338.fasta
1245745.fasta
1268274.fasta
1290391.fasta
1291518.fasta
1325737.fasta
148305.fasta
2004951.fasta
2714763.fasta
330879.fasta
336722.fasta
559298.fasta
655827.fasta
```

Figura 17. 2do. BLAST a partir de secuencias estándares de Egh16.

El filtro permite obtener los 5 mejores hits para cada posible proteína ortóloga en cada proteoma de los hongos filamentosos de mayor interés, descartando aquellas secuencias con nula o menor similitud. qacc= identificador del query, sacc= identificador del subject, qlen= longitud de la secuencia de consulta, slen= longitud de la secuencia de sujetos, length= longitud de alineación, qstart= inicio de alineación en la consulta, qend= fin de la alineación en la consulta, sstart= inicio de alineación en el sujeto, send= fin de la alineación en el asunto, score= puntuación, pident= porcentaje de coincidencias idénticas, nident= número de coincidencias idénticas, mismatch= número de desajustes, positive= número de partidos con puntuación positiva, gaps= número total de brechas, gapopen= número de espacios abiertos.

6.3 Descarga de Proteomas a partir de UniProtKB.

Mediante los archivos FASTA de los hongos filamentosos seleccionados, se construyó una lista (nombrada “ascomycota”) de sus identificadores taxonómicos y nombres (ver **Tabla 6**), los cuales fueron tomados de la página de UniProtKB. La descarga de proteomas se realizó de forma automática sometiendo la variable “ascomycota” a un proceso iterativo.

Se importó la librería “requests” con la finalidad de realizar una solicitud http (servidor web) para extraer los proteomas.

En el proceso de extracción de proteomas, se construyó un bucle “for loop”, en el cual la variable “i” tomó y guardó temporalmente uno por uno los elementos de la lista “ascomycota”.

Para fines visuales por medio de jupyter procederemos a observar, ordenar y editar la información previamente descargada, para que sea más fácil analizar los datos recabados en variables "HG" y "org" **Figura 18-19**, para la extracción de proteomas, se construyó un bucle "for loop", en el cual la variable "i" tomó y guardó temporalmente uno por uno los elementos de la lista "io", se crearon las variables "x" y "d" para extraer los datos en forma de tabla: id, nombre, organismo, cantidad de proteínas y el Detector de proteoma completo (CPD), clasificando cada proteoma en alguna categoría con el fin de conocer la integridad y calidad del proteoma., en la variable "d" se creó un Pandas DataFrame, definido como estructura de datos tabulares bidimensionales de tamaño variable, con ejes como columnas y filas, es decir que los datos se agrupan y distribuyen de forma tabular como antes se mencionó en columnas y fila (Panda versión 1.4.3., 2020) Para la creación del DataFrame es importante haber importado antes las librerías (Ver **Anexo 1**). Los datos se guardaron en la variable "tab", ver **Figura 20**.

```
In [6]: 1 ascomycota = ['330879', '559298', '1215338', '447093', '1291518', '1245745',
2              '2714763', '2004951', '148305', '1325737', '1290391', '1268274', '336722']

In [7]: 1 import urllib.request

In [8]: 1 for i in ascomycota:
2         print(i)
3         urllib.request.urlretrieve('https://www.uniprot.org/uniprot/?query='+i+'&format=fasta', 'HG/'+i+'.fasta')

330879
559298
1215338
447093
1291518
1245745
2714763
2004951
148305
1325737
1290391
1268274
336722
```

Figura 18. Descarga iterativa de proteomas de hongos filamentosos a partir de la base de datos NCBI y UniProtKB.

- En In [6] definimos una variable en este caso ascomycota que contendrá los números taxonómicos previamente elegidos.
- En In [7] por medio del comando import urllib.request notificamos que vamos a sustraer información de una base de datos.
- En In [8] se realizó un bucle que me permita sustraer la información de ascomycota y que la busque su contenido en la base de UniProtKB, para posteriormente descargar cada uno en formato .fasta y guardar dicha información en la carpeta HG, de igual forma si el comando se está ejecutando de la manera correcta que me muestre los números taxonómicos que ya han sido descargados y guardados.


```
In [2]: 1 import re
        2 import os
```

```
In [3]: 1 secuencias=os.listdir('HG')
        2 secuencias
```

```
Out[3]: ['1215338.fasta',
         '1245745.fasta',
         '1268274.fasta',
         '1290391.fasta',
         '1291518.fasta',
         '1325737.fasta',
         '148305.fasta',
         '2004951.fasta',
         '2714763.fasta',
         '330879.fasta',
         '336722.fasta',
         '447093.fasta',
         '559298.fasta']
```

```
In [4]: 1 with open('HG/1325737.fasta', 'r') as m:
        2     for line in m:
        3         line=line.rstrip()
        4         if '>' in line:
        5             i=line.split('|')[1]
        6             print('>'+i)
        7             Seq=''
        8         else:
        9             Seq+=line
        10            print(line)
```

```
In [5]: 1 open('dat.fast','w')
        2 with open('HG/1325737.fasta', 'r') as m:
        3     for line in m:
        4         line=line.rstrip()
        5         if '>' in line:
        6             i=line.split('|')[1]
        7             print('>'+i)
        8             Seq=''
        9         else:
        10            Seq+=line
        11            print(line)
```

In [4] con el comando **with open** se observa una vista previa del documento 1325737.fasta ubicado en la carpeta HG y que a su vez nos permite hacer modificaciones sobre este mismo, que posteriormente se realizaran por medio de un bucle, ya que cada hongo o en este caso cada número taxonómico contendrá una cantidad finita de proteomas en donde **for line in m:** marcará la variable line como la primera fila de cada proteoma diferente modificable, **con if y else** se pretenderá dar las instrucciones de los que se debe modificar; **if** condicionará que si en la 1º fila se encuentra '|' remplazarlo con >, con **else** se indicara que todas las filas se alineen.

- a) In [2] y In [3] Importo los comandos; **import os** (para poder **establecer un listado**), e **import re** (para **observar la información de** los archivos. fasta).
- b) Ya con import os defino una variable en este caso **secuencias** que son todos los documentos fasta contenidos en la carpeta HG y con su complemento os.listdir hago que todos se ordenen en una lista, todo dentro de la variable **secuencias**.

- d) En In [5] se usa In [4] con un comando más **open y w** en donde se pretende guarda todas las modificaciones anteriores en mismo documento y por ende en la misma carpeta.
- e) Vista previa de las modificaciones anteriores.

```
In [7]: 1 for line in secuencias:
        2     org=open('OR/'+line,'w')
        3     with open('HG/'+line, 'r') as m:
        4         for line in m:
        5             line=line.rstrip()
        6             if '>' in line:
        7                 i=line.split('|')[1]
        8                 org.write('>'+i+'\n')
        9             else:
        10                org.write(line+'\n')
        11            org.close()
```

```
In [8]: 1 org=open('dat.fast','w')
        2 with open('HG/1325737.fasta', 'r') as m:
        3     for line in m:
        4         line=line.rstrip()
        5         if '>' in line:
        6             i=line.split('|')[1]
        7             org.write('>'+i+'\n')
        8         else:
        9             org.write(line+'\n')
        10    org.close()
```

Figura 19. 1era. modificación de la información de proteomas

- f) Posteriormente se definirá In [7] como la variable **org**, como puede verse en In [8].
- g) En In [7] una vez definida la variable **org** la ocupamos para un bucle que nos permitirá que todos los hongos/no. taxonómicos y sus proteomas contenidos en la carpeta **HG** se modifiquen como anteriormente se realizó y que posteriormente estos sean guardados en otra carpeta llamada **OR**.

```
In [8]: 1 tab = []
        2 for i in io:
        3     print(i)
        4     x = requests.get('https://www.uniprot.org/proteomes/?query='+i+'&format=tab&columns=id,name,organism-id,proteincount,cpd')
        5     d = DataFrame([i.split('\t') for i in x.split('\n')][1:-1])
        6     tab.append(d)
```

```
1215338
1245745
1268274
1290391
1291518
1325737
148305
2004951
2714763
330879
336722
447093
559298
655827
```

Figura 20. Sustracción de información de los hongos seleccionados; ID del proteoma, ID del organismo, CPD, etc, a través de UniProtKB.

- h) En In [8] se genera una lista (`tab = []`) con la información obtenida del bucle (`for i in io`) que contiene los proteomas de los hongos elegidos y sustraídos de la base de datos UniProtKB., en donde “i” va guardando temporalmente todos los datos de la lista “io” (identificador, organismo), en la línea 3, va mostrando cada id del cual se ha extraído la información solicitada, en la línea 4, se utilizó “requests.get” para realizar la solicitud a la página web de UniProt y tomar los datos guardados en la variable “i” y extraerlos de la web, en la línea 5 se construyó el DataFrame a partir de un bucle for y se eliminaron los espacios, por último se anexó el DataFrame a “tab” mediante “.append”.

En la **Tabla 8.**, se pueden observar los datos recabados, es importante mencionar que **CPD** usa información de linaje taxonómico para identificar el grupo de proteomas taxonómicamente más cercano a él. La mayoría muestran un valor Outlier (low value) o Valor atípico (valor bajo), es decir que estos proteomas tienen un recuento por debajo del promedio del grupo y no son tan cercanos a la mediana (estándar y/o referencia) del grupo. Se utilizó la **Tabla 8.**, para la descarga de los proteomas en formato fasta a través de un bucle `for` y el módulo “`urllib.request`” permitió abrir la URL, evitando redireccionamientos, cookies, etc. ver **Figura 21.**

Tabla 8. Metadata de Proteomas extraídos de UniProtKB.

	Proteome ID	Organism	Organism ID	Protein count	CPD
0	UP000024533	Trichophyton interdigitale (strain MR816) (Str...	1215338	8217	Close to standard (low value)
1	UP000030106	Beauveria bassiana D1-5 (Strain: D1-5)	1245745	11860	Standard
2	UP000053110	Blumeria graminis f. sp. tritici 96224 (Strain...	1268274	4024	Outlier (low value)
3	UP000012045	Botryotinia fuckeliana (strain BcDW1) (Noble r...	1290391	11022	Outlier (low value)
4	UP000054544	Metarhizium anisopliae BRIP 53293 (Strain: BRI...	1291518	11404	Outlier (low value)
5	UP000287544	Fusarium sp. AF-6 (Strain: NRRL62580)	1325737	15273	Standard
6	UP000515153	Magnaporthe grisea (Crabgrass-specific blast f...	148305	12455	Unknown
7	UP000221130	Cordyceps sp. RAO-2017 (Strain: 1346)	2004951	11255	Standard
8	UP000243081	Cordyceps confragosa (Lecanicillium lecanii) (...)	2714763	8125	Close to standard (low value)
9	UP000002530	Neosartorya fumigata (strain ATCC MYA-4609 / A...	330879	9647	Close to standard (low value)
10	UP000008062	Zymoseptoria tritici (strain CBS 115943 / IPO3...	336722	10972	Outlier (low value)
11	UP000001631	Ajellomyces capsulatus (strain G186AR / H82 / ...)	447093	9214	Outlier (low value)
12	UP000663217	Ajellomyces capsulatus (strain G186AR / H82 / ...)	447093	9460	Unknown
13	UP000002038	Blastomyces gilchristii (strain SLH14081) (Bla...	559298	10303	Outlier (low value)
14	UP000002499	Metarhizium acridum (strain CQMa 102) (Strain...	655827	9830	Outlier (low value)

```
In [15]: 1 for i in up_tax:
2         print(io[up_tax[i]], '=', i)
3         urllib.request.urlretrieve('https://www.uniprot.org/uniprot/?query=proteome:'+i+'&format=fasta

Trichophyton interdigitale = UP000024533
Beauveria bassiana = UP000030106
Blumeria graminis = UP000053110
Brottytis cinerea = UP000012045
Metarhizium anisopliae = UP000054544
Fusarium sp = UP000287544
Magnaporthe grisea = UP000515153
Cordyceps sp = UP000221130
Lecanicillium lecanii = UP000243081
Aspergillus fumigatus = UP000002530
Zymoseptoria tritici = UP000008062
Blastomyces dermatitidis = UP000002038
Metarhizium acridum = UP000002499
```

Figura 21. Descarga de proteomas a través de UniProtKB.

Los datos de la lista `up_tax` que contiene “Proteome ID” y “Organism ID” de la Tabla 8., se fueron guardando temporalmente en la variable “i”, el módulo `urllib.request.urlretrieve` llamó a la página web de `uniprot.org` y se fueron descargando cada uno de los proteomas en formato FASTA, los cuales se nombraron de acuerdo a la lista “`up_tax`” y se guardaron en la carpeta “`fastas2`”.

6.4 Identificación de proteínas ortólogas asociadas al factor de virulencia Egh16 mediante Blastp.

Los ortólogos son aquellos genes que provienen de un mismo gen y cuya divergencia se debe a un proceso de especiación. En palabras más sencillas, son ortólogos aquellos genes o proteínas que tienen la misma identidad en distintas especies y que por consiguiente conservan la misma función durante el curso de la evolución (Abascal, 2012).

La herramienta Blastp es capaz de comparar una secuencia problema, en este caso los 13 hongos elegidos(query) contra una gran cantidad de secuencias que se encuentren en una base de datos (subject), hablando concretamente de los estándares del factor de virulencia Egh16 y permite encontrar posibles ortólogos. En este análisis se realizaron varios filtros con el fin de identificar de manera óptima cada proteína ortóloga para cada proteoma.

En el segundo filtro (**Figura 17**) se obtuvieron las cinco secuencias de proteínas hits o con mayor similitud en alineamiento con cada proteína implicada en la vía de señalización de estrés osmótico, a través del comando: **-max target segs 5 -max hsp 1-**. Se ordenó el archivo con base en los especificadores de formato “**qacc sacc qlen slen length qstart qend sstart send score bitscore evaluate pident nident mismatch positive gaps gapopen**” para usar posteriormente en el DataFrame. El resultado se guardó en “**Tables2**” en formato **.txt**.

En variable “**Tables2**” se guardaron los datos recabados y en la variable “**es**” y se prosiguió a la creación de un Pandas DataFrame, esto permitió analizar los datos a través de su agrupación y organización.

Se construyó un bucle “**for loop**”, en el cual la variable “**a**” tomó y guardó temporalmente uno por uno los datos de la variable “**es**”, el comando “**pd.read_csv**” leyó los datos de “**Tables2**” (dado el formato .txt) y se asignaron las columnas, incluyendo “**tax**” que representa el identificador taxonómico para cada hongo filamentoso, para una mejor visualización. El DataFrame se guardó en una variable “**data**”, **Figura 22**.

```
In [24]: 1 #Se define una variable en forma
2 #de lista que muestre todo lo que
3 #hay en la carpeta Tables2
4 es=os.listdir("Tables2")
5 es=[i for i in es if '.txt' in i]
6 es

Out[24]: ['1215338.txt',
'1245745.txt',
'1268274.txt',
'1290391.txt',
'1291518.txt',
'1325737.txt',
'148305.txt',
'2004951.txt',
'2714763.txt',
'330879.txt',
'336722.txt',
'559298.txt',
'655827.txt']

In [25]: 1 #Se define una variable que con un listado
2 columna=["qacc", "sacc", "qlen", "slen",
3 "length", "qstart", "qend", "sstart",
4 "send", "score", "bitscore", "evaluate",
5 "pident", "nident", "mismatch",
6 "positive", "gaps", "gapopen"]

In [26]: 1 #Se define una variable en forma de listado que contenga por medio
2 #de un bucle toda la información de Tables2 concatenada y filtrada
3 #con la variable columna todo esto en formato .txt
4 data=[]
5 for a in es:
6 ab=pd.read_csv('Tables2/'+a,sep='\t',names=columna)
7 ab["tax"]=a.split(".txt")[0]
8 data.append(ab)

In [27]: 1 #Todo lo anterior es concatenado (data)
2 #y definido en otra variable (blastp)
3 blastp = pd.concat(data)

In [28]: 1 #Este mismo blastp filtrará cada una de las proteínas
2 #de los hongos a una sola opción de dominio con similitud
3 #superior al 40%
4 blastp['qacc'] = [i.split('|')[1] for i in blastp.qacc]
5 blastp = blastp[blastp.pident >= 40]
```

Figura 22. DataFrame de los datos recopilados guardado en formato txt.

sep= separador, el valor predeterminado es ',' como en CSV, CSV= valores separados por coma “Comma-Separated Values”, \t= carácter de tabulación, columna=“qacc”, “sacc”, “qlen”, “slen”, “length”, “qstart”, “qend”, “sstart”, “send”, “score”, “bitscore”, “evaluate”, “pident”, “nident”, “mismatch”, “positive”, “gaps”, “gapopen”, split= divide una cadena en una lista, append= agrega elementos al final de la lista.

El DataFrame “**data**” se concatenó (se unieron los datos) y se guardó en la variable “**blastp**”, después se descartaron aquellos valores de % de identidad menores a 40% (vistos en la columna “**pident**”) (**Figura 22**), ya que representan valores muy bajos de alineación de secuencias entre “**qacc**” y “**sacc** “. En este primer ejercicio: filtrado de datos del DataFrame, se observó (**Tabla 9**) que más de una proteína del “**qacc**” había hecho match con una proteína del “**sacc**”, ya que al realizar la comparación entre secuencias hubo una cierta similaridad (reflejada en él % de

identidad) , entre fragmentos de aminoácidos perteneciente a regiones proteicas como por ejemplo; dominios y residuos, como era de esperarse, dado que los hongos filamentosos presentan estas regiones conservadas asociadas a los estándares factor de virulencia Egh16 presentes en *M. acridum*.

Tabla 9. BLAST de asociación entre proteomas de hongos elegidos y estándares de Egh16.

```
In [29]: 1 blastp
```

```
Out[29]:
```

	qacc	sacc	qlen	slen	length	qstart	qend	sstart	send	score	bitscore	evalue	pident	nident	mismatch	positive	gaps	gapopen
3	A0A0A2V6E6	E9DV35	421	317	103	83	179	194	296	165	68.2	2.280000e-16	42.718	44	53	52	6	3
5	A0A0A2VL92	E9E6I7	369	322	245	4	248	3	232	788	308.0	1.900000e-107	66.122	162	68	180	15	1
7	A0A0A2VL92	E9DV35	369	317	104	96	196	186	289	192	78.6	4.670000e-20	42.308	44	57	56	3	1
10	A0A0A2V766	E9E1A1	691	300	303	1	301	1	300	995	387.0	1.870000e-134	63.036	191	107	235	5	4
11	A0A0A2V766	E9DV35	691	317	319	1	305	1	317	737	288.0	1.130000e-95	52.351	167	136	228	16	4
...
25	E9DV35	E9DV35	317	317	317	1	317	1	317	1615	626.0	0.000000e+00	100.000	317	0	317	0	0
26	E9DV35	E9E1A1	317	300	314	1	313	1	300	784	306.0	6.560000e-108	52.548	165	134	215	15	5
27	E9DV35	E9E6I7	317	322	104	186	289	95	195	217	88.2	1.290000e-23	44.231	46	55	61	3	1
30	E9DRS7	E9DRS7	374	374	374	1	374	1	374	1837	712.0	0.000000e+00	100.000	374	0	374	0	0
31	E9DRS7	E9EDX3	374	189	194	91	284	1	165	337	134.0	1.210000e-41	47.938	93	72	115	29	5

84 rows × 19 columns

Se utilizó la instrucción “blastp = blastp [blastp.pident >= 40]” para excluir aquellos alineamientos y similaridad de menor significancia (menor que 40%) entre las secuencias de cada proteína de cada proteoma y las secuencias (estándares) de la base de datos creada.

*** Nota: Para descartar falsos positivos, se realizó un segundo otros filtros de tal manera que también se consideraran aquellas proteínas con la misma arquitectura estructural. Se generó el archivo fasta:” blastp_sequences_Miriam.fasta” donde se guardaron todos los datos del Blastp, como un resguardo de información.**

El segundo filtro se realizó al generar un DataFrame por números taxonómicos del blastp, correspondiente a cada hongo filamentosos y al estándar de factores de virulencia Egh16 de “*M. acridum*”, con la finalidad de asegurar el match entre las secuencias de proteínas de cada hongo filamentosos y al menos una proteína de las 7 proteínas de referencia o estándar del factor de virulencia Egh16 (*M. acridum*).

Se descartaron aquellas proteínas que no cumplieron con un % de identidad mayor o igual a 50 (umbral de >=50) y a partir de estos datos se generó un nuevo DataFrame del cual se tomó la columna **sacc** (de la base de datos de referencia o

estándares) para buscar los identificadores y crear mini DataFrame (mdf) de cada uno, con el fin de concatenarlos y reordenar el índice de la tabla (genera un orden) (Figura 23).

```
In [33]: 1 c=[]
2 for i in blastp.tax.unique().tolist():#iteracion de una lista de no. taxonomicos unicos
3     pr1=blastp[blastp.tax==i]#se busca cada uno de no.taxonomicos en la variable
4     #blastp(contiene todos los no.tax) y genera 1 dataframe
5     pr2=pr1[pr1.pident >=50].reset_index(drop=True)#A partir de dataframe anterior(pr1)
6     #filtramos por % de identidad(por un umbral de >=50) y generamos de nuevo un dataframe(pr2)
7     if len(pr2) == 0:
8         pass
9     else:
10        lista1=pr2.sacc.unique().tolist()#A partir del pr2 generamos una lista unica de la columna
11        #sacc(de la base de datos de referencia)
12        data1=[]
13        for i in lista1:
14            mdf=pr2[pr2.sacc==i]#se busca cada uno de sacc(identificadores de la base de datos)
15            #en la variable pr2 y se genera un minidatframe
16            data1.append(mdf)#Cada uno de los minidatframes es guardado en la variable data1
17            #(data1 fue definido anteriormente como una lista vacia)
18        con=pd.concat(data1).reset_index(drop=True)#todos los minidatframe guardados
19        #en data1 son concatenados y reseteamos el indice (genera un orden) y construimos un nuevo
20        c.append(con)#todos los dataframe con son guardados en la variable c previamente definida
```

Figura 23. 1er BLAST de asociación estructural y alineación de los 13 hongos filamentosos elegido VS Estándares Egh16.

Todos los mini dataframe guardados en data1 fueron concatenados y el índice fue reordenado, al finalizar se construyó un nuevo dataframe(con), todo se guardó en la variable “c”.

Posteriormente se guardó en la variable “table1” a través de ejecutar: **table1=pd.concat(c).reset_index(drop=True)** (Tabla 10).

Tabla 10. DataFrame “table 1; Blastp 1 filtro >=50”.

```
In [35]: 1 table1
2 #blastp 1 filtro >=50
```

	qacc	sacc	qlen	slen	length	qstart	qend	sstart	send	score	bitscore	evalue	pident	nident	mismatch	positive	gaps	gapope
0	A0A0A2VL92	E9E6I7	369	322	245	4	248	3	232	788	308.0	1.900000e-107	66.122	162	68	180	15	
1	A0A0A2V766	E9E1A1	691	300	303	1	301	1	300	995	387.0	1.870000e-134	63.036	191	107	235	5	
2	A0A0A2VGW4	E9E1A1	357	300	335	1	331	1	300	768	300.0	8.050000e-105	50.746	170	126	212	39	
3	A0A0A2V766	E9DV35	691	317	319	1	305	1	317	737	288.0	1.130000e-95	52.351	167	136	228	16	
4	A0A0A2VGW4	E9DV35	357	317	147	189	335	171	317	572	224.0	3.880000e-75	80.272	118	29	131	0	
5	M7TPQ7	E9DRS7	246	374	241	5	232	8	244	642	251.0	1.390000e-86	57.261	138	86	174	17	
6	A0A0D9NLX9	E9E1A1	343	300	340	1	336	1	300	854	333.0	5.080000e-118	54.706	186	110	222	44	
7	A0A0D9NWQ8	E9E1A1	313	300	310	1	309	1	300	760	297.0	2.620000e-104	51.613	160	139	210	11	

Se descartaron filas correspondientes a porcentajes de identidad menores de 50% al realizar el alineamiento entre el query (qacc) columna referente a los 13 hongos seleccionados y el subject (sacc) columna referente a los estándares del factor de virulencia Egh16.

A simple vista encontramos un problema en el match de proteínas de **qacc** y **sacc**, en donde tenemos varias (del **qacc**) con % de identidad diferentes que hacen match con una sola del **sacc**, se realizó la búsqueda por **Phobius** y **Pfam** de los dominios, familias y regiones de las proteínas (**Anexo 2**), por medio del documento “**phobius_sequences_Miriam.txt**” y “**pfam_sequences_Miriam.txt**”, con la finalidad de que más adelante, se pudieran identificar las proteínas con mayor similitud dada su arquitectura y estructura proteica, aquellas que no cumplan con los criterios antes mencionados serán descartadas (proteínas transmembranales). Al finalizar esta búsqueda, se concatenaron ambos (**Pfam** y **Phobius**) por medio del comando: **metadata=pd.concat([pfam,fobius]).drop_duplicates().sort_values(by='start',ascending=True).reset_index(drop=True)**. (Figura 24)

```
In [44]: 1 metadata=pd.concat([pfam,fobius]).drop_duplicates().sort_values(by='start',ascending=True).reset_index(drop=True)
```

```
In [45]: 1 metadata['start'] = metadata['start'].astype(int)
2 metadata['end'] = metadata['end'].astype(int)
```

```
In [46]: 1 metadata = metadata[metadata.name.str.contains('transmembrane') == False]
```

```
In [47]: 1 metadata
```

Out[47]:

	Entry	start	end	name	type
0	E9DRS7	1	20	sig_p	Signal_peptide
1	A0A0D9NLX9	1	17	sig_p	Signal_peptide
2	M7TKI2	1	19	sig_p	Signal_peptide
3	M7U9I5	1	26	sig_p	Signal_peptide
4	M7TPQ7	1	18	sig_p	Signal_peptide
...
101	A0A381L540	23	202	Egh16-like	Family
102	A0A061HM96	27	201	Egh16-like	Family
103	A0A061HMW0	28	204	Egh16-like	Family
104	A0A381LL21	5	155	Egh16-like	Family
105	A0A0A2V766	500	601	Cyclin_C_2	Domain

106 rows × 5 columns

Con base en la información antes recabada se hace otro filtro en el que de igual manera se omiten dominios transmembranales, pero con una nueva indicación que sean dominios que no se repitan guardando dichos cambios en otra variable **metadata**

Figura 24. Creación del “metadata” a partir de phobius y pfam.

La columna “**Entry**” guarda los valores de los id de cada proteína, “**start**” y “**end**” indican el inicio y fin de la región en donde se encuentra el dominio, péptido señal, en “**name**” y “**type**” se describe el nombre del dominio, péptido señal, etc.

Una vez filtrados los datos, las proteínas potencialmente ortólogas fueron identificadas utilizando un umbral de 0.5 con base al *índice de jaccard* (concepto estadístico para medir las similitudes entre dos conjuntos), se tomaron los datos del **metadata** y **table1: qacc** y **sacc**). Aquellas proteínas con un número de dominios igual al valor esperado (**subject**) fueron consideradas, las proteínas con un valor diferente fueron descartadas, esto asegura que ambas proteínas tengan la misma arquitectura (en ella comparten dominios, regiones, longitudes) (**Figura 25**). Se

consideró que el valor de 0 corresponde a aquellas que no comparten ningún dominio, y ≥ 0.5 a aquellas que comparten un número de dominios igual al valor esperado y que estos son idénticos o casi idénticos. (Figura 26).

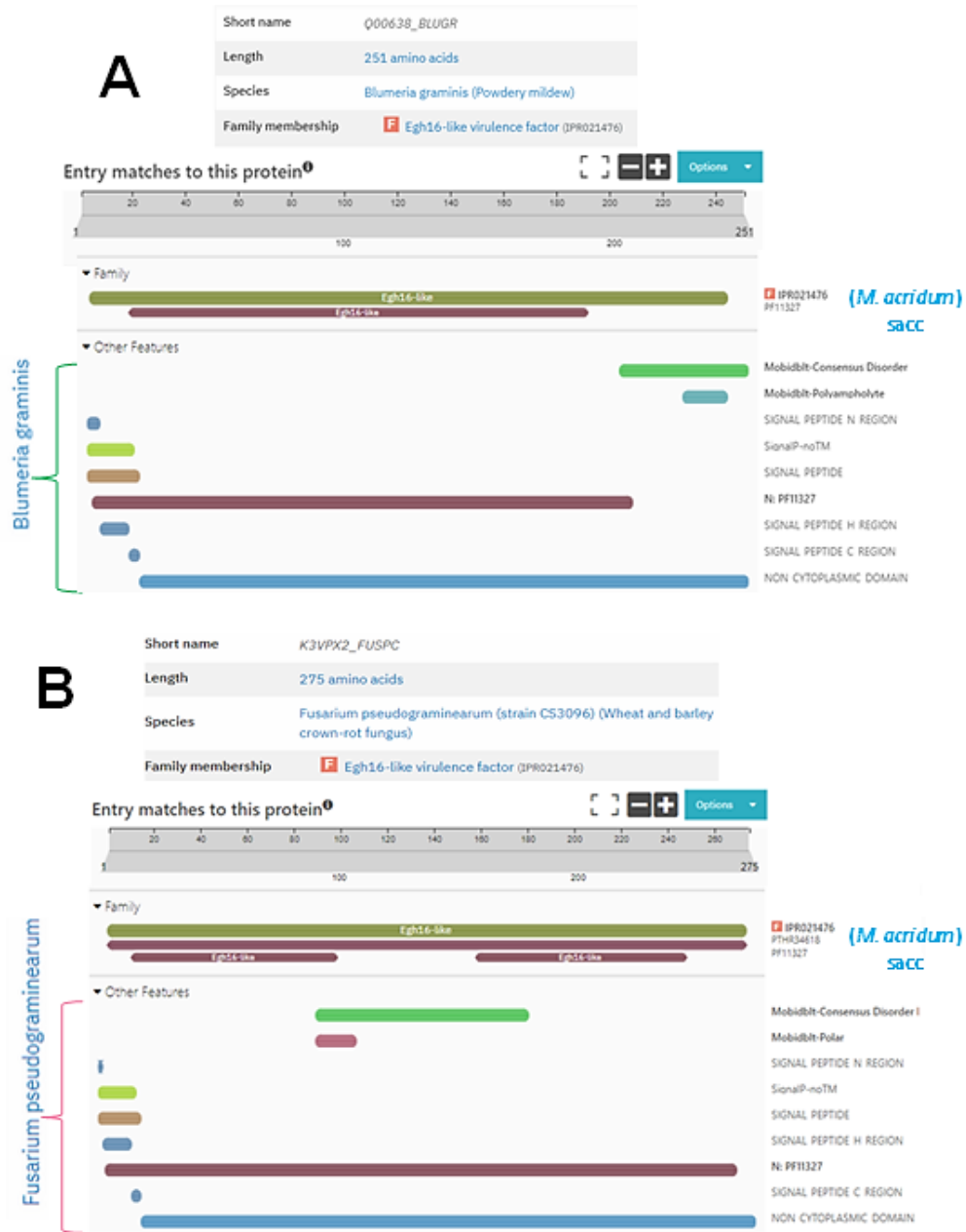


Figura 25. Comparación de arquitecturas.

A. Blumeria graminis (especies) VS Factores de virulencia Egh16 de M. acridum.

B. Fusarium (especies) VS Factores de virulencia Egh16 de M. acridum.

Elaborado a partir de: Pfam (<https://www.ebi.ac.uk/interpro/entry/pfam/#table>) y Hmmer (<https://www.ebi.ac.uk/Tools/hmmer/>).

Se define una variable nueva (s) con la información anterior de uno de los proteomas de los hongos ordenando de forma ascendente la presencia de sus genes

```
In [48]: 1 s = metadata[metadata.Entry == 'A0A0A2VL92'].sort_values(by='start',ascending=True).reset_index(drop=True)
2 s
```

```
Out[48]:
```

	Entry	start	end	name	type
0	A0A0A2VL92	1	23	sig_p	Signal_peptide
1	A0A0A2VL92	20	195	Egh16-like	Family

```
In [49]: 1 q = metadata[metadata.Entry == 'E9E6I7'].sort_values(by='start',ascending=True).reset_index(drop=True)
2 q
```

```
Out[49]:
```

	Entry	start	end	name	type
0	E9E6I7	1	20	sig_p	Signal_peptide
1	E9E6I7	19	194	Egh16-like	Family

Se define una variable nueva (q) con la información anterior con uno de los estándares o referencias del factor Egh16 ordenando de forma ascendente la presencia de sus genes, observando que tanto para el hongo como para la referencia sus genes están en las mismas posiciones, por lo que con las variables (c1, c2) y con el índice de similaridad de jaccard establezco que ambos son similares.

```
In [50]: 1 c1 = set(s.name.tolist())
2 c2 = set(q.name.tolist())
```

```
In [51]: 1 jaccard_similarity(s1=c1, s2=c2)
```

```
Out[51]: 1.0
```

```
In [52]: 1 dominio=[]
2 for i,j in zip(table1.qacc,table1.sacc):
3     dominio.append([metadata[metadata.Entry==i],metadata[metadata.Entry==j]])
```

```
In [53]: 1 dom2=[]
2 dom3=[]
3 for q,s in dominio:
4     q=q.sort_values(by='start',ascending=True).reset_index(drop=True)
5     s=s.sort_values(by='start',ascending=True).reset_index(drop=True)
6     d=[]
7     for a,b in zip(q.name,s.name):
8         c1=set(a)
9         c2=set(b)
10        if jaccard_similarity(s1=c1, s2=c2)>=0.5:
11            #print(q.entry.tolist()[0],s.entry.tolist()[0],a,b,jaccard_similarity(s1=c1, s2=c2))
12            d.append([a,b])
13        if len(q)==len(d)==len(s):
14            #print(q,s)
15            dom2.append([q,s])
16            dom3.append([q.Entry.tolist()[0],s.Entry.tolist()[0]])
17            #Las proteínas ortólogas fueron identificadas usando como umbral de 0.5 con base en el
18            #índice de jaccard, aquellas proteínas con un número de dominios igual al valor esperado (subject)
19            #fueron consideradas, Las proteínas con un valor diferente fueron descartadas, esto asegura que
20            #ambas proteínas tengan La misma arquitectura.
```

Con la variable **dominio** se pretende acceder a la información contenida en la tabla1 a los proteomas de los hongos (**qacc**) y los estándares de Egh16 (**sacc**).

Con **dom2** y **dom3** se pretende todos los genes de **qacc** de los hongos sean comparados con los genes de las referencias, para que posteriormente se sometan al índice de similaridad de jaccard

Figura 26. Código para la identificación de proteínas potencialmente ortólogas de los 13 hongos problema VS Proteomas de referencia Egh16 con base al índice de jaccard (similaridad).

Se itero la función **jaccard_similarity** y se asignaron los valores del “**qacc**” y “**sacc**” para medir su similaridad, dado el tamaño de la intersección de ambos conjuntos, dividido por el tamaño de la unión de ambos (total). La intersección da el número de datos compartidos entre ambos conjuntos, y la unión da el número total de datos (compartidos y no compartidos) en ambos conjuntos. Se realizó un zip de “**table1: qacc** y **sacc**” y se conjunto en el “**metadata**”. Se utilizó un bucle “**for loop**” para ordenar los valores de forma ascendente y se prosiguió a ejecutar el índice de jaccard con un umbral de ≥ 0.5 , si los valores correspondieron (**q=d=s**) entonces se conjuntaron los datos “**q**” y “**s**”.

La información contenida en **dom3**: “**qacc**” () y “**sacc**” () fueron guardados en la variable “**dom4**” y por medio del método. **merge**, se unió estos datos con los de **table1** y se guardó en la variable “**col**”, comando: **dom4=DataFrame (dom3,**

`columns=["qacc", "sacc"], col=dom4.merge(table1, on=["qacc", "sacc"], how="left").`

Se creó un bucle “for loop” con el fin de devolver los valores únicos de “sacc” y como segunda acción ordenó los valores del % de identidad (**pident**) de forma descendente. Se guardó en la variable **final1** y se concatenó, se generó el DataFrame “**final2**”. **Tabla 11.**

Tabla 11. DataFrame de la información filtrada de las posibles proteínas ortólogas.

```
In [57]: 1 #Con la variable final1 se guarda la información y criterios de filtrado antes mencionados,
2 #agregando que solo muestre una sola referencia de Egh16 vs los diferentes genes de los
3 #hongos y que estos no se repitan aunado a que se orden de forma ascendente la similaridad (pident)
4 final1=[]
5 for i in col.sacc.unique():
6     df=col[col.sacc ==i].sort_values(by ='pident',ascending=False).reset_index(drop=True)
7     #colduplicate=df.drop_duplicates(subset = 'organism', keep = 'first')
8     final1.append(df)
```

```
In [58]: 1 #Dicha información, criterios de filtrado es guarda y concatenada en la variable final2
2 final2= pd.concat(final1)
3 final2
```

qacc	sacc	qlen	slen	length	qstart	qend	sstart	send	score	bitscore	evalue	pident	nident	mismatch	positive	gaps	gapopen	tax	organism
E9E6I7	E9E6I7	322	322	322	1	322	1	322	1634	634.0	0.000000e+00	100.000	322	0	322	0	0	655827	Metarhizium acridum
A0A179IE19	E9E6I7	330	322	199	4	202	3	201	861	336.0	5.670000e-119	78.894	157	42	176	0	0	2714763	Lecanicillium lecanii
OA428PLN0	E9E6I7	307	322	200	3	202	2	201	790	308.0	1.140000e-108	72.000	144	56	164	0	0	1325737	Fusarium sp
AA0D9P1Q4	E9E6I7	327	322	338	1	327	1	322	1053	410.0	2.880000e-148	71.893	243	68	258	27	2	1291518	Metarhizium anisopliae
AA6P8AMD3	E9E6I7	269	322	201	1	200	1	201	709	277.0	4.670000e-97	67.662	136	64	156	1	1	148305	Magnaporthe grisea
AA0A2VL92	E9E6I7	369	322	245	4	248	3	232	788	308.0	1.900000e-107	66.122	162	68	180	15	1	1245745	Beauveria bassiana
AA6P8BC56	E9E6I7	251	322	201	1	199	1	201	664	260.0	1.950000e-90	63.682	128	71	151	2	2	148305	Magnaporthe grisea

Los datos muestran un orden descendente (mayor a menor) con respecto al % de identidad ubicado en la columna “**pident**” y las proteínas del “**sacc**”, se observa el mejor match entre las proteínas del hongo *M. acridum* asociadas al factor de virulencia Egh16 y los hongos filamentosos problemas elegidos (13), como; *Lecanicillium lecanii*, *Fusarium sp*, *Magnaporthe grisea*, *Beauveria bassiana*, etc.

La tabla “**final2**” se guardó en formato fasta “**final2_Miriam.fasta**” y posteriormente se realizó la nueva base de datos **BLAST** con los datos del fasta, ejecutando el comando “**subprocess.call('makeblastdb -in final2_Miriam.fasta -dbtype prot -parse_seqids -out DB2/final2_Miriam', shell = True)**”, con el fin de poder realizar el enfoque Blastp All vs All, en donde se realiza la comparación de todo contra todo y con ello se agrupan aquellas proteínas con similitudes altas, encontrando así

las proteínas ortólogas asociadas al factor de virulencia Egh16 presentes en los 13 hongos filamentoso de la división *Ascomycota* elegidos anteriormente.

6.5 BLASTP All vs All

Se ejecuta una búsqueda BLASTP recíproca de todos contra todos, para buscar similitudes de proteínas potencialmente ortólogas de los hongos anteriormente elegidos ante las proteínas del hongo de referencia: *M. acridum* y compararse también entre sí. **Figura 27.**

```
In [69]: 1 #Dicha información tendrá criterios de columnas;
2 # qacc sacc qlen slen length qstart qend sstart send score bitscore evaluate pident nident mismatch positive gaps gapopen,
3 #con un máximo de secuencias objetivo de 7 y un máximo de salida de 1, de igual manera es guarda como;
4 #all_vs_all_Miriam.txt en la carpeta DB2 por medio de un blastp que lo itera rápidamente
5 subprocess.call('blastp -db DB2/final2_Miriam -query final2_Miriam.fasta -evaluate 1E-6 -outfmt "6 qacc sacc qlen slen length
Out[69]: 0
```

Figura 27. BLASTP ALL_VS_ALL.

max_target_seqs consideró en este caso los datos del diccionario “**prots**” concatenada en la variable “**dominios**” que corresponden a las 7 proteínas asociadas al factor de virulencia Egh16 y sus respectivos identificadores. **max_hsps 1** tomó el mejor hit (la mejor proteína dentro del **query**-13 hongos elegidos) que tiene una mayor similitud con respecto a la proteína del **subject** (estándares o referencias del factor de virulencia Egh16), es decir, si hay 7 proteínas asociadas al factor de virulencia Egh16, se esperaría encontrar el mejor hit para cada una (en una situación ideal se encontrarían las 7 proteínas potencialmente ortólogas para cada hongo filamentoso).

Después de realizar el BLASTP, se quedaron solo los % de identidad más altos de cada proteína y los mejores de cada comparación (proteína vs proteína), y se tomó el de mayor %, en este caso el primero con un 90% de identidad. Considerando la longitud del alineamiento (columna “**length**”) y el porcentaje de identidad (columna “**pident**”), principales parámetros para la identificación de proteínas potencialmente ortólogas. La longitud de alineamiento indica regiones donde son similares las secuencias (“**query**”, “**subject**”), por lo que su valor debe ser cercano a la longitud del “**subject**” (columna “**slen**”) y el porcentaje de identidad representa la conservación entre las secuencias, es decir, que tan idénticas son entre sí.

El valor “**E**”, indica estadísticamente la alineación por pares y el reflejo del tamaño de la base de datos. Cuanto menor sea el valor **E**, más significativo será el impacto (Madden T., 2010). El resultado fue guardado en la variable “**all_vs_all**”, **Tabla 12.**

Tabla 12. Representación bidimensional de All VS All (proteína VS proteína). Vista previa, dimensiones: 193 filas x 20 columnas

```
In [71]: 1 #Se define una nueva variable all_vs_all que concatenará la información contenida en el archivo
2 #all_vs_all_Miriam.txt con base en las columnas antes definidas en la variable names
3 all_vs_all = pd.read_csv('all_vs_all_Miriam.txt', sep = '\t', names = names)
4 all_vs_all
```

```
Out[71]:
```

	qacc	sacc	qlen	slen	length	qstart	qend	sstart	send	score	bitscore	evalue	pident	nident	mismatch	positive	gaps	gapopen
0	E9E6I7	E9E6I7	322	322	322	1	322	1	322	1634	634.0	0.000000e+00	100.000	322	0	322	0	0
1	E9E6I7	A0A0D9P1Q4	322	327	210	1	210	1	210	991	386.0	2.990000e-138	89.048	187	23	197	0	0
2	E9E6I7	A0A179IE19	322	330	199	3	201	4	202	853	333.0	3.390000e-117	78.894	157	42	176	0	0
3	E9E6I7	A0A0A2VL92	322	369	203	3	205	4	206	826	322.0	1.530000e-112	75.862	154	49	170	0	0
4	E9E6I7	A0A428PLN0	322	307	200	2	201	3	202	790	308.0	4.550000e-108	72.000	144	56	164	0	0
...
191	M7TPQ7	E9DRS7	246	374	241	5	232	8	244	642	251.0	5.770000e-86	57.261	138	86	174	17	5
192	M7TPQ7	A0A6P8AW89	246	282	259	2	246	4	262	609	239.0	3.750000e-82	53.668	139	106	169	14	5
193	M7TPQ7	A0A179IGY3	246	305	227	13	219	8	234	545	214.0	4.040000e-72	57.709	131	76	157	20	5

```
In [72]: 1 #si todo resulta bien tendremos simetría de datos
2 len(all_vs_all.sacc.unique().tolist()), len(all_vs_all.qacc.unique().tolist())
```

Out[72]: (28, 28)

```
In [73]: 1 #Con base en el diccionario id_org[i] se agrega una columna que contiene el nombre del hongo en cuestion
2 #al Dataframe de all_vs_all
3 all_vs_all['organism'] = [id_org[i]+' '+i for i in all_vs_all.sacc]
```

Columnas: 'qacc', 'sacc', 'qlen', 'slen', 'length', 'qstart', 'qend', 'sstart', 'send', 'score', 'bitscore', 'evalue', 'pident', 'nident', 'mismatch', 'positive', 'gaps', 'gapopen', 'organism_sacc', 'organism_qacc'. Se observan las comparaciones entre todas las proteínas ortólogas del factor de virulencia Egh16 encontradas VS los estándares asociados al factor de virulencia Egh16 de *M. acridum*.

Para saber cuántas proteínas se obtuvieron se ejecutó el comando **“len(all_vs_all.sacc.unique().tolist()), len(all_vs_all.qacc.unique().tolist())”**, el resultado fue de (28, 28) proteínas obtenidas, ver **Tabla 12**, es razonable debido a los filtros realizados (si todas las proteínas cumplieran los datos serían: 7 proteomas estándares asociadas al factor de virulencia Egh16 x 13 hongos filamentosos previamente elegidos= 91 proteínas obtenidas). Estas 28 proteínas cumplieron con los parámetros: % de identidad >=50%, índice de jaccard con un umbral de >=0.5 y el mejor hit para cada proteína. Es por lo que han sido consideradas como proteínas potencialmente ortólogas del factor de virulencia Egh16, a partir del hongo *M. acridum*.

6.6 Análisis de Clustering

El análisis de agrupamiento (Clustering), es conjunto de técnicas cuyo objetivo es agrupar un set de objetos de datos en clústeres (similar entre sí dentro del mismo clúster/ diferente a los objetos de otros clústeres) (Martínez, 2008).

Se realizó un análisis de clustering usando la métrica de correlación, que permitió identificar más grupos de proteínas ortólogas en múltiples organismos, esto hablando de una posible reproductibilidad del análisis de otros hongos filamentosos.

La tabla “All vs All” (*all_vs_all[['organism_qacc', 'pident', 'qacc', 'sacc']]*), se procesa para llegar a una matriz de datos (Tabla 13).

Tabla 13. Matriz “matmat” (base de datos para generar el análisis Clúster). Vista previa, dimensiones 28 filas x 28 columnas.

```
In [85]: 1 #Aquí elimino las columnas sacc y organism, para una mejor visualización, todo definido en la variable matmat
2 matmat = matrix.drop(columns = ['sacc', 'organism_sacc'])
3 matmat
```

Out[85]:

organism_qacc	Aspergillus fumigatus Q4WB06	Beauveria bassiana A0A0A2VGV4	Beauveria bassiana A0A0A2VL92	Brotytis cinerea M7TPQ7	Cordyceps sp A0A2C6A9X9	Fusarium sp A0A428PEX0	Fusarium sp A0A428PLN0	Lecanicillium lecanii A0A179I8F6	Lecanicillium lecanii A0A179IE19	Lecanicillium lecanii A0A179IGY3	...
0	0.000	100.000	0.000	0.000	54.808	56.213	0.000	48.851	0.000	0.000	...
1	0.000	0.000	100.000	0.000	0.000	0.000	70.792	0.000	82.759	35.714	...
2	0.000	47.293	0.000	0.000	55.102	0.000	0.000	51.009	0.000	0.000	...
3	0.000	61.176	0.000	0.000	64.835	60.383	0.000	52.174	0.000	0.000	...
4	0.000	0.000	0.000	0.000	0.000	0.000	70.476	0.000	78.537	0.000	...
5	0.000	0.000	38.961	59.336	0.000	0.000	0.000	0.000	0.000	71.130	...
6	37.459	47.463	0.000	0.000	0.000	0.000	0.000	57.878	0.000	0.000	...
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...
8	38.339	49.570	0.000	0.000	60.825	55.911	0.000	100.000	0.000	0.000	...
9	0.000	0.000	0.000	0.000	0.000	0.000	72.772	0.000	100.000	0.000	...

Posteriormente, los datos obtenidos se visualizaron en un dendrograma (Figura 28), en el cual se formaron 4 clústeres de proteínas potencialmente ortólogas entre de los factores asociados a Egh16 de *M. acridum* y los 13 hongos filamentosos seleccionados de la división *Ascomycota*. Estos clústeres se obtuvieron en función de la similaridad de las secuencias de aminoácidos y/o de la estructura secundaria de las proteínas. Para las proteínas del hongo modelo “*M. acridum* (E9DV35)”, se encontraron proteínas que cumplieron con los parámetros establecidos para la identificación de proteínas potencialmente ortólogas en el clúster de color morado (Figura 28).

En el **Anexo 3** se encuentra el código utilizado para la ejecución del análisis y su representación (dendograma).

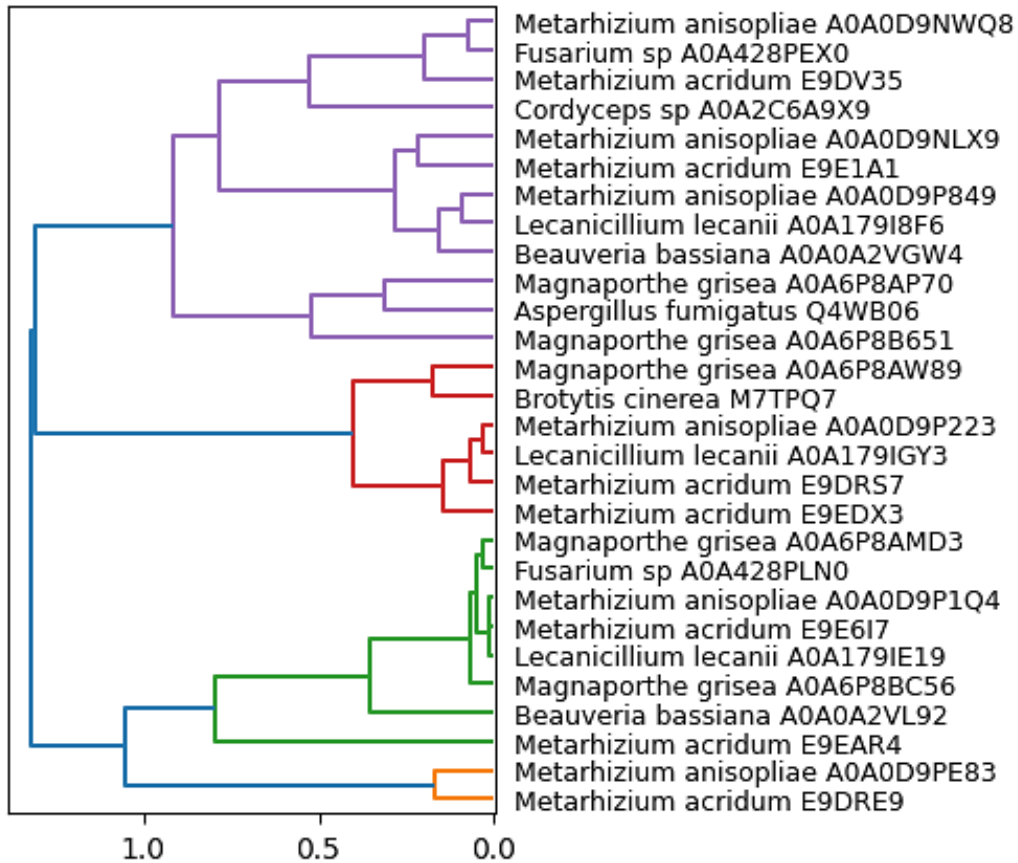


Figura 28. Análisis de Clustering (Relación familiar de mismos genes).

El dendograma va del 0-1, donde 0 es altamente similar y 1 es altamente diferente. Por lo que, si la distancia es menor, estará más cercana a cero, entonces la distancia es directamente proporcional a la similaridad entre los conjuntos o en este caso las proteínas y/o genes. La distancia máxima la marca la línea azul oscuro que une los clústeres (color naranja, verde claro, rojo, morado, café, fucsia, gris, verde oliva, azul claro, etc.).

Los 4 clústeres obtenidos se visualizan en diferentes colores, cada uno de ellos representa un grupo de proteínas (de los hongos filamentosos seleccionados) que son potencialmente ortólogas a alguna proteína de referencia (del hongo modelo *M. acridum*). En varios de los clústeres formados existe una alta similaridad del grupo hacia una proteína de referencia que a otra (aunque se encuentre en el mismo clúster).

En conclusión, en el análisis de clustering se puede observar una correlación entre la familia del hongo y los grupos de proteínas ortólogas en múltiples organismos, lo que reafirma que los organismos (hongos) de una familia tienden a presentar más de un gen en común (**Figura 28**).

6.7 Heatmap

Un *heatmap* (o mapa de calor) es una técnica de visualización de datos bidimensional que representa la magnitud de valores individuales dentro de un

conjunto de datos como un color. La variación del color puede ser por tonalidad o intensidad (Wilkinson, 2009).

Para visualizar el grado de conservación de las proteínas potencialmente ortólogas de todos los hongos filamentosos se construyó un *heatmap*. El grado de conservación en las proteínas nos indica que tan similar son los proteomas asociados al factor de virulencia Egh16 de *M. acridum* VS los proteomas de los 13 hongos filamentosos ascomicetos.

En la creación del Heatmap simétrico se tomaron en cuenta los datos de “**mat1**” (Tabla 14).

Tabla 14. Base de datos “mat1” a partir de la cual se generará el heatmap.

```
In [91]: 1 #Con una nueva variable mat1 con la información contenida en mat se elimina la columna organism_sacc
2 mat1 = mat.drop(columns = ['organism_sacc'])
3 mat1
```

Out[91]:

	Metarhizium acridum E9DRE9	Metarhizium anisopliae A0A0D9PE83	Metarhizium acridum E9EAR4	Beauveria bassiana A0A0A2VL92	Magnaporthe grisea A0A6P8BC56	Lecanicillium lecanii A0A179IE19	Metarhizium acridum E9E6I7	Metarhizium anisopliae A0A0D9P1Q4	Fusarium sp A0A428PLN0	Magnaporthe grisea A0A6P8AMD3	...	Magnaporthe grisea A0A6P8AI
0	100.000	65.169	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000
1	80.488	100.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000
2	0.000	0.000	100.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000
3	29.048	30.472	0.000	100.000	62.802	82.759	75.862	76.000	70.792	65.174	...	0.000
4	0.000	30.392	30.457	62.802	100.000	64.593	63.861	63.184	65.347	58.431	...	0.000
5	0.000	0.000	28.934	0.000	64.824	100.000	78.894	79.899	72.772	66.667	...	0.000
6	27.723	29.612	32.178	66.122	63.682	78.894	100.000	71.893	72.000	67.662	...	0.000
7	0.000	31.797	31.771	0.000	63.184	78.537	89.048	100.000	70.476	67.317	...	0.000
8	30.151	29.000	30.102	70.792	65.000	72.772	72.000	71.500	100.000	70.707	...	0.000
9	27.861	0.000	27.273	65.657	58.431	66.667	61.088	64.159	70.707	100.000	...	0.000

Correlaciones de acuerdo con el análisis de clustering usando la métrica “correlación”. Los datos corresponden a los % de identidad obtenidos en el blastp **All vs All**, considerando solo los mejores hits entre proteínas asociadas al factor de virulencia Egh16 de *M. acridum* vs proteínas de hongos filamentosos.

Mediante los datos de “**mat1**” se creó el *heatmap* (Figura 29), que sirvió para visualizar que tan idénticas son las proteínas potencialmente ortólogas obtenidas de cada hongo ascomiceto, entre sí y entre las proteínas asociadas al factor de virulencia Egh16 de *M. acridum*.

En el **Anexo 4**, se encuentra el código utilizado para la creación del *Heatmap*.

En la gráfica “*heatmap*”, también se vislumbrar una tendencia en la que el factor de virulencia Egh16 se encuentra presente en todos y cada uno de los hongos de la división *Ascomycota* escogidos (**Figura 29**), sin embargo, no hay evidencia suficiente que demuestre que todos los hongos patógenos, incluyendo de humanos, que tienen este factor de virulencia sean o no patógenos o mucho menos que tan virulentos pueden ser o llegar a ser.

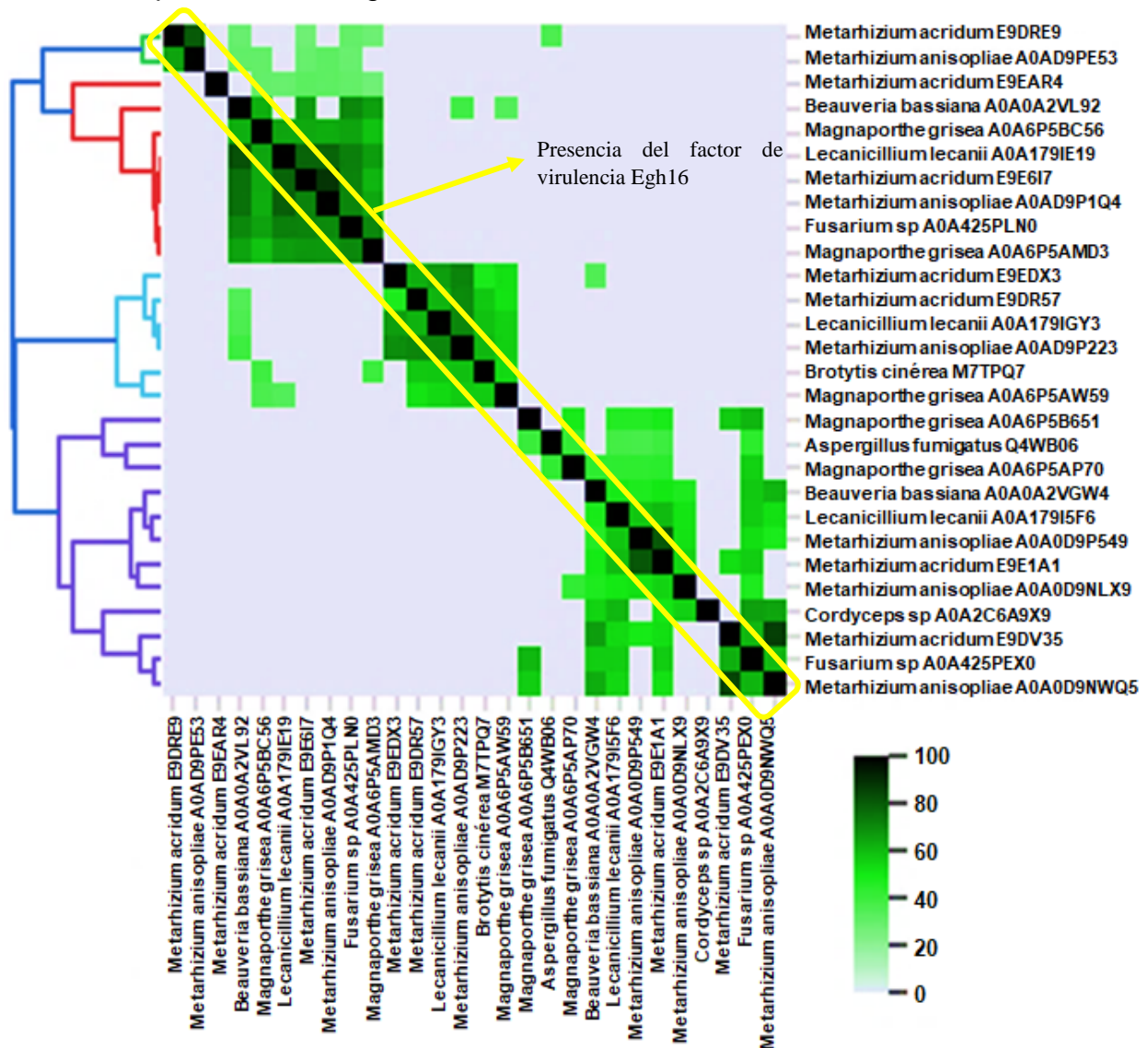


Figura 29. Hongos con presencia del factor de virulencia Egh16.

El *Heatmap* muestra el grado de conservación entre los 13 hongos filamentosos seleccionados en este trabajo y el hongo modelo *M. acridum*. En el gráfico se observan agrupaciones en función de la similitud de secuencias. La escala de color representa el % de identidad obtenido por Blastp All vs All usando un Evalúe de *1E-6*.

7. CONCLUSIONES

1. Los hongos nos han permitido, evaluar riesgos y beneficios de su uso, control y eliminación de estos, ya sea en el ámbito clínico, farmacológico, cosmético, alimenticio y control de plagas.
2. Se encontró en los 13 hongos estudiados de manera bioinformática el factor de virulencia Egh16.
3. Todos estos hongos pertenecen a la división *Ascomycota*.
4. Todos estos hongos presentan un grado de patogenicidad, ya sea sobre insectos, plantas o humanos.
5. Aun cuando se presenta este factor de virulencia en hongos con un grado de patogenicidad, es imposible adjudicar lo segundo a lo primero.
6. Es necesario determinar, de manera in vitro e in vivo mediante el diseño de un experimento con muestra estadísticamente fiable, que demuestre de manera irrefutable que el factor de virulencia Egh16 es directamente responsable de la patogenicidad de los 13 hongos aquí estudiados.
7. Por lo anterior, no es concluyente que el factor de virulencia Egh16 sea el único responsable de la patogenicidad de los hongos aquí estudiados, sin embargo, debe considerarse como un factor necesario, pero no suficiente.

8. REFERENCIAS

1. Abascal, F. (2012). bioweb: Molecular Modelling Group. Obtenido de cbm UAM: http://bioweb.cbm.uam.es/courses/CITA2012/alignment/Familias_proteinas/teoria.html
2. Alberti, F., Foster, G. D., & Bailey, A. M. (2017). Natural products from filamentous fungi and production by heterologous expression. *Applied microbiology and biotechnology*, 101(2), 493–500.
3. Alonso, J. (2004). CLASIFICACIÓN Y DESCRIPCIÓN DE LOS HONGOS. "Guía de campo de los hongos de España y de Europa", 21-40.
4. Baraggio, Nancy Guadalupe; Carrasco, Marta Susana; Simonetta, Arturo Carlos Identificación y propiedades de interés tecnológico de levaduras aisladas de productos lácteos de la región de Santa Fe, Argentina *Revista de la Sociedad Venezolana de Microbiología*, vol. 37, núm. 2, 2017, pp. 54-60
5. Barcia-Guerrero, G. A. (2022). Infecciones fúngicas de afectación cutánea y superficial. *Revista Científica, Dominio de las Ciencias*, 2085-2113.
6. Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. *BMJ (Clinical research ed.)*, 324(7344), 1018–1022.
7. Beceiro Alejandro, M. T. (2012). Resistencia a los antimicrobianos y virulencia, ¿una asociación beneficiosa para el mundo microbiano? *Elsevier*, 493-495.
8. Bernal González, Idolina; Pedraza Melo, Norma Angélica; Sánchez Limón, Mónica Lorena El clima organizacional y su relación con la calidad de los servicios públicos de salud: diseño de un modelo teórico *Estudios Gerenciales*, vol. 31, núm. 134, enero-marzo, 2015, pp. 8-19 Universidad ICESI Cali, Colombia.
9. Bilal Aslam, M. B. (2017). *Proteomics: Technologies and Their Applications*. Oxford Academic, 185-189.
10. Brock. (2015). *Biología de los microorganismos (14° ed.)*. México: PEARSON EDUCACIÓN
11. Carreras Villaseñor Nohemí, E. E. (2022). Conociendo a los hongos fitopatógenos. *Ecológico*, 3(1), 32-66.
12. Castañón Camarena Ha, Ja, C. C. (2019). Capacidad coagulante del extracto enzimático crudo de *Rhizomucor miehei* sobre leche de vaca. Universidad Juárez del Estado de Durango. Facultad de Ciencias Químicas., 200-205.
13. Castillo Uruzeta Pável, C. D. (2015). Las micotoxinas: metabolitos secundarios de los hongos filamentosos. *CIENCIA-TECNOLOGÍA-SOCIEDAD*, 122-127.
14. Challa, S., Uppin, S. G., Uppin, M. S., Pamidimukkala, U., & Vemu, L. (2015). Diagnosis of filamentous fungi on tissue sections by immunohistochemistry using anti-aspergillus antibody. *Medical mycology*, 53(5), 470–476.
15. Chen, C., Hou, J., Tanner, J. J., & Cheng, J. (2020). Bioinformatics Methods for Mass Spectrometry- Based Proteomics Data Analysis. *International journal of molecular sciences*, 21(8), 2873.
16. Crous Pedro W., J. Z. (2021). Names of Phytopathogenic Fungi: A Practical Guide. *The American Phytopathological Society*, 1500–1508.
17. Davies, A., Hooley, F., Causey-Freeman, P., Eleftheriou, I., & Moulton, G. (2020). Using interactive digital notebooks for bioscience and informatics education. *PLoS computational biology*, 16(11), e1008326.
18. De La Calle Rodríguez Natalia, C. S. (2012). Factores de virulencia para la infección de tejidos queratinizados por *Candida albicans* y hongos dermatofitos. *Rev CES Med.*, 26(1), 43-45.
19. Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., Rudd,
20. Egbuta, M. A., Mwanza, M., & Babalola, O. O. (2017). Health Risks Associated with Exposure to Filamentous Fungi. *International journal of environmental research and public health*, 14(7), 719.
21. FORDECyT-CONACyT. (2022). Generación de estrategias científico tecnológicas con un enfoque multidisciplinario e interinstitucional para afrontar la amenaza que representan los complejos ambrosiales en los sectores agrícola y forestal de México. CDMX: INECOL.
22. Garcés de Granada Emira, M. C. (2003). *Mofología y Calsificación de los hongos*. Bogotá: Departamento de biología, Facultad de Ciencias, Universidad Nacional de Colombia.
23. García-Vidal Carolina, M. S. (2014). Inmunopatología de las micosis invasivas por hongos filamentosos. *Revista Iberoamericana*, 220-224.
24. Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2018). A brief history of bioinformatics. *Briefings in Bioinformatics*, 1-16
25. Ghorai, S., Banik, S. P., Verma, D., Chowdhury, S., Mukherjee, S., & Khawala, S. (2009). Fungal biotechnology in food and feed processing. *Food Research International*, 42(5), 577-587.
26. Grell, M. N., Mouritzen, P., & Giese, H. (2003). A *Blumeria graminis* gene family encoding proteins with a C-terminal variable region with homologues in pathogenic fungi. *Gene*, 311, 181–192.
27. Guarro, J. (Enero de 2012). *Taxonomía y biología de los hongos causantes de infección en humanos*. Elsevier, 33-39.
28. Heredia-Abarca G. 2020. La importancia de los hongos (Fungi) en los servicios ecosistémicos. *Bioagrociencias* 13(2): 98-108.
29. Hernández, D. E. M., Castillo, O. L. S., García, E. Y., Mandujano, G. V., Díaz, G. G. & Álvarez C. J. (2019). Bioinformatics as a Tool for the Structural and Evolutionary Analysis of Proteins. In P. Behzadi, & N. Bernabò (Eds.), *Computational Biology and Chemistry*. IntechOpen. doi: 10.5772/intechopen.89594.
30. InterPro. (10 de Enero de 2021). ebi.ac.uk. Obtenido de interpro: <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR021476/>
31. Jawetz, M. (2011). *Microbiología médica (25 ed.)*. México: McGRAW-HILL INTERAMERICANA.
32. Jiang, R., Zhang, X., & Zhang, M. Q. (Eds.). (2013). *Basics of Bioinformatics: Lecture Notes of the Graduate Summer School on Bioinformatics of China*, 271-301. doi:10.1007/978-3-642-38951-1.
33. Keerthikumar, S. (2017). An Introduction to Proteome Bioinformatics. *Methods in molecular biology (Clifton, N.J.)*, 1549, 1–3.
34. Kenneth J. Ryan, C. G. (2011). *Sherris. Microbiología médica*. México: McGRAW-HILL INTERAMERICANA EDITORES.
35. Keppanan, R., Sivaperumal, S., Hussain, M., Bamisile, B. S., Aguila, L., Qasim, M., Mekchay, S., Wang, L., & Krutmuang, P. (2019). Molecular characterization of pathogenesis involving the GAS 1 gene from Entomopathogenic fungus *Lecanicillium lecanii* and its virulence against the insect host *Diaphorina citri*. *Pesticide biochemistry and physiology*, 157, 99–107.
36. Köhler, J. R., Casadevall, A., & Perfect, J. (2014). The spectrum of fungi that infects humans. *Cold Spring Harbor perspectives in medicine*, 5(1), a019273.

37. Liang, X., Shang, S., Dong, Q., Wang, B., Zhang, R., Gleason, M. L., & Sun, G. (2018). Transcriptomic analysis reveals candidate genes regulating development and host interactions of *Colletotrichum fructicola*. *BMC genomics*, 19(1), 557.
38. López Ríos Carlos Andrés, A. Z. (2006). PRODUCCIÓN DE ÁCIDO CÍTRICO CON *Aspergillus niger* NRRL 2270 A PARTIR DE SUERO DE LECHE. Escuela de Procesos y Energía, Facultad de Minas, Universidad Nacional de Colombia, Medellín, 39-57.
39. Madden, T. (2010). The BLAST Sequence Analysis Tool. The NCBI Handbook. The National Library of Medicine. Chapter 6, 1-15.
40. Madigan Michael T., J. M. (2015). Brock. Biología de los microorganismos. México: PEARSON EDUCACIÓN.
41. Maina UM, IB Galadima, FM Gambo, D Zakaria. (2018). A review on the use of entomopathogenic fungi in the management of insect pests of field crops. *J Entomol Zool Stud*, 6(1), 27-32.
42. Martínez Barnetche, Jesús La bioinformática como herramienta para la investigación en salud humana *Salud Pública de México*, vol. 49, 2007, pp. 64-66 Instituto Nacional de Salud Pública Cuernavaca, México.
43. Martínez-Cruz Jesús, D. R.-G. (2021). Effectors with chitinase activity (EWCAs), a family of conserved, secreted fungal chitinases that suppress chitin-triggered immunity. *THE PLANT CELL*, 2-10.
44. Martínez Ramos, E. (2008):"Aspectos teóricos del Análisis Cluster y aplicación a la caracterización del electorado potencial de un partido" en Sánchez Carrión, JJ(Ed.):"Introducción a las técnicas." op.cit. pag165.
45. Matínez, R. L. (1995). Los mecanismos patogénicos y de protección en las micosis. *Gac Méd Méx*, 132(1), 49-51.
46. McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic acids research*, 32(Web Server issue), W20–W25.
47. Mesas, J. M.; Alegre, M. T. El pan y su proceso de elaboración *Ciencia y Tecnología Alimentaria*, vol. 3, núm. 5, diciembre, 2002, pp. 307-313.
48. Meyer, V., Andersen, M. R., Brakhage, A. A., Braus, G. H., Caddick, M. X., Cairns, T. C., et al. (2016). Current challenges of research on filamentous fungi in relation to human welfare and a sustainable bio-economy: a white paper. *Fungal Biology and Biotechnology*, 3(1), 6.
49. Miranda-Estrada Laura Iveth, M. R.-R.-L.-R.-V.-A. (2016). Relación entre factores de virulencia, resistencia a antibióticos y los grupos filogenéticos de *Escherichia coli* uropatógena en dos localidades de México. *Elsevier*, 35(7), 2-3.
50. NCBI. (2020). BLAST: Basic Local Alignment Search Tool. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
51. Pacheco Hernández Ma. de Lourdes, J. F. (2019). Organismos entomopatógenos como control biológico en los sectores agropecuario y forestal de México: una revisión. *Revista mexicana de ciencias forestales*, 10(56).
52. Pfaller Michael A., P. M. (2021). "MICROBIOLOGÍA MÉDICA". Argentina: Barcelona: ELSEVIER CASTELLANO.
53. Ramírez Quintero Juan David, Sobre el mal de los ardientes o del fuego de San Antonio *Acta Medica Colombiana*, vol. 43, núm. 3, 2018, Julio-Septiembre, pp. 156-160.
54. Reedy, J. L., Bastidas, R. J., & Heitman, J. (2007). The virulence of human pathogenic fungi: notes from the South of France. *Cell host & microbe*, 2(2), 77–83.
55. S.L, Sherris. (2011). *Microbiología médica* (Segunda ed.). México: MCGRAW-HILL INTERAMERICANA EDITORES, S. A. de C. V.
56. Scattolini, A. (2010). Hongos fitopatógenos II. Uruguay.
57. Solano De la Cruz, M. T., Escobar Hernández, E. E., Arciniega González, J. A., Rueda Zozaya, R. P., Adame García, J., & Luna – Rodríguez, M. (2021). Bioinformatic analysis deciphers the molecular toolbox in the endophytic/pathogenic behavior in *F. oxysporum* f. sp. *vanillae* - *V. planifolia* Jacks interaction. *BioRxiv*, 2021.03.23.436347.
58. Suárez Machín Caridad; Garrido-Carralero, Norge Antonio; Guevara-Rodríguez, Carmen Amarilys Levadura *Saccharomyces cerevisiae* y la producción de alcohol. *Revisión bibliográfica ICIDCA. Sobre los Derivados de la Caña de Azúcar*, vol. 50, núm. 1, enero-abril, 2016, pp. 20-28.
59. Téllez-Jurado Alejandro, M. G.-C. (2009). Mecanismos de acción y respuesta en la relación de hongos entomopatógenos e insectos. *Universidad Politécnica de Pachuca*, 75-77.
60. Torrades, O. S. (2004). *Proteómica: El diseño molecular de la vida*. *Offarm*, 23 (4): 126-130.
61. Ullmann, A., Aguado, J., Arkan-Akdagli, S., Denning, D., Groll, A., Lagrou, K., et al. (2004). Diagnosis and management of *Aspergillus* diseases: executive summary of the 2017 ESCMID-ECMM-ERS guideline. *Fungal Biology and Biotechnology*, 24(1), e1-e38.
62. Wiley J, S. L. (2008). *Prescott's Principles of Microbiology*. Nueva York: McGraw-Hill.
63. Wilkinson L, Friendly M (mayo de 2009). "La historia del mapa de calor de los cúmulos". *El estadístico estadounidense*. 63(2): 179–184.
64. Xue, C., Park, G., Choi, W., Zheng, L., Dean, R. A., & Xu, J. (2004). Two Novel Fungal Virulence Genes Specifically Expressed in Appressoria of the Rice Blast Fungus. *The Plant Cell*, 14(9), 2107- 2119.
65. Zulema Pérez Matias, M. C. (Enero de 2019). "COMPORTAMIENTO PRODUCTIVO DE CAPRINOS SUPLEMENTADOS CON *Saccharomyces cerevisiae*, CON DOS FUENTES DE FORRAJE FIBROSO" . Tesis para obtener Título de Médico Veterinario Zootecnista. CD. ALTAMIRANO, Guerrero, México: UNIVERSIDAD AUTÓNOMA DE GUERRERO, ESCUELA SUPERIOR DE MEDICINA VETERINARIA Y ZOOTECNIA N°1.

9. ANEXOS

Anexo 1. Librerías importadas a Python: Pandas, NumPy, Matplotlib, Seaborn, SciPy.

```
import pandas as pd
import numpy as np
from pandas import DataFrame

import matplotlib.pyplot as plt
import seaborn as sns
from scipy.cluster.hierarchy import dendrogram, linkage
import matplotlib as mpl
```

Anexo 2. Código Phobius y Pfam.

```
In [37]: 1 #Se realizó una lista a partir de el documento phobius_sequences_Miriam.txt, en donde
2 #el dominio de la proteína de los hongos es transmembranal no se observe"
3 tc=[]
4 with open ("phobius_sequences_Miriam.txt","r") as ph:
5     for i in ph:
6         i = i.rstrip()
7         if "##" in i:
8             s=[]
9         else:
10            s.append(i)
11        if len(s)>1:
12            if ("SIGNAL" in "".join(s)) or ("TRANSMEM" in "".join(s)):
13                ide=s[0].replace("ID  ", "")
14                for i in s[1:]:
15                    if "SIGNAL" in i:
16                        esp=re.sub(" * ", "\t", i)
17                        ps=esp.split("\t")
18                        #print(ide,ps[2],ps[3],ps[1])
19                        tc.append([ide,ps[2],ps[3], "sig_p", "Signal_peptide"])
20                    if "TRANSMEM" in i:
21                        esp1=re.sub(" * ", "\t", i)
22                        sp=esp1.split("\t")
23                        #print(ide,sp[2],sp[3],sp[1])
24                        #identificador, inicio, final y tipo
25                        tc.append([ide,sp[2],sp[3], "transmembrane", "Transmembrane"])
```

```
In [38]: 1 #Con la información ya filtrada con una variable fobius solo ciertas columnas y eliminar los dominios repetidos
2 fobius=DataFrame(tc, columns=["Entry", "start", "end", "name", "type"]).drop_duplicates().reset_index(drop = True) # 150
3 fobius
```

```
In [39]: 1 #Se define una nueva variable
2 cols = ['Entry', 'alignment start', 'alignment end', 'start',
3         'end', 'hmm acc', 'name', 'type', 'hmm start',
4         'hmm end', 'hmm length', 'bit score', 'E-value',
5         'significance', 'clan']
```

```
In [40]: 1 #Se define una nueva variable en forma de lista,
2 #La cual se contiene la información
3 #del documento pfam_sequences_Miriam.txt pero solo la primera línea
4 res0 = []
5 with open('pfam_sequences_Miriam.txt', 'r') as fq:
6     for line in fq:
7         line = line.rstrip()
8         if '#' in line:
9             pass
10        else:
11            line = re.sub(' * ', ' ', line)
12            res0.append(line.split(' ')[0:15])
```

```
In [41]: 1 #Con la variable antes definida res0 se establece en una nueva variable en la cual
2 #por medio de un DataFrame se concatena con la variable cols anteriormente definida
3 hmm_all = DataFrame(res0, columns = cols).dropna()
4 hmm_all['E-value'] = hmm_all['E-value'].astype(float)
```

```
In [42]: 1 #Con la anterior variable hmm_all se filtra la información de ciertas columnas
2 pfam=hmm_all[["Entry", "start", "end", "name", "type", ]]
```

Anexo 3. Código para la ejecución del análisis de clustering y su representación. (dendograma).

```
In [87]: 1 clusters = linkage(matmat.values.T, method = 'complete', metric='correlation')
2 #clusters = Linkage(matmat.values.T, method = 'complete', metric='correlation')
3 # se realizó un análisis de clustering usando la métrica de correlación
4 #ya que esta permitió identificar más grupos de proteínas ortólogas en
5 #múltiples organismos
```

```
In [88]: 1 mpl.rcParams.update(mpl.rcParamsDefault)
2 fig, ax = plt.subplots(figsize=(3, 5))
3
4 dend2 = dendrogram(clusters,
5                     orientation='left', truncate_mode='level',
6                     labels=matmat.columns)
7 #ax.axis('off')
8
9 ax.tick_params(axis='y', which='major', labelsize=9)
10 #plt.savefig('dendrogram.png', dpi = 900, bbox_inches='tight')
11 plt.show()
```

Anexo 4. Código para crear el heatmap.

```
from colormap import Colormap
```

```
In [95]: 1 import warnings
2 warnings.filterwarnings("ignore")
```

```
In [96]: 1 sns.set(font_scale=0.5)
2 g = sns.clustermap(mat1.values.T, method = 'complete', metric="correlation",
3                   tree_kws=dict(colors=dend2['color_list']),
4                   linewidths=0, linecolor='white',
5                   square=True, col_cluster = False, row_cluster = True,
6                   xticklabels= mat.columns.tolist()[1:],
7                   yticklabels= mat.organism_sacc.tolist(),
8                   #mask = mat1.values <= float(50),
9                   #annot = True, annot_kws={"size": 5},fmt= '.2g',
10                  cbar_pos=(0.75, 0.07, .03, .19),
11                  cmap = Colormap().cmap_linear('#eaeaf2', 'limegreen', 'black'), figsize = (5, 5))
12
13 for a in g.ax_row_dendrogram.collections:
14     a.set_linewidth(1.5)
15 for a in g.ax_col_dendrogram.collections:
16     a.set_linewidth(1.5)
17
18 g.ax_heatmap.tick_params(bottom=True, right=True, top=False, left=False, width = 0.2, length=4, color='black')
19
20 #plt.savefig('egh16.png', dpi = 900, bbox_inches='tight')
21 plt.show()
```