



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOQUÍMICAS**

“Desarrollo de un protocolo computacional para el diseño automatizado de oligonucleótidos a ser utilizados en ensayos de identificación altamente específica y sensible de muestras clínicas de *Candida glabrata* y otros organismos patógenos”.

TESIS

**QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias Bioquímicas**

PRESENTA:

JUAN MANUEL SÁNCHEZ CONTRERAS

Tutor principal:

Dr. Enrique Merino Pérez

[Instituto de Biotecnología - UNAM](#)

Comité Tutorial:

Dra. Irene Beatriz Castaño Navarro

[Instituto Potosino de Investigación Científica y Tecnológica A.C.](#)

Dr. Víctor Manuel González Zúñiga

[Centro de Ciencias Genómicas – UNAM](#)

CIUDAD DE MÉXICO, Mayo 2024



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas

Tesis Digitales

Restricciones de uso

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Reconocimientos

Esta tesis se realizó en el Grupo de Genómica Computacional del Instituto de Biotecnología - UNAM y recibió financiamiento del proyecto FORDECYT-PRONACES/610281/2020 "Plasticidad genómica de aislados clínicos secuenciales del hongo patógeno de humanos *Candida glabrata*: consecuencias para la resistencia a antifúngicos y la capacidad de adhesión" a cargo de los siguientes doctores:

- **Dra. Irene Castaño Navarro** Instituto Potosino de Investigación Científica y Tecnológica A.C.
- **Dra. Rosa Areli Martínez Gamboa** Instituto Nacional de Nutrición Salvador Zubirán
- **Dr. Enrique Merino Pérez** Instituto de Biotecnología – UNAM

Los resultados experimentales de la identificación específica de *Candida glabrata* mediante ensayos de PCR usando los oligos diseñados con el programa desarrollado en esta tesis, y que se muestran en el Apéndice 1, fueron realizados por el **Dr. Marco Josué Hernández Chávez**, a quien le agradezco por permitirme incluir dichos resultados en mi tesis. Dicho trabajo experimental fue realizado en el laboratorio de la Dra. Irene Castaño Navarro.

Agradecimientos

Este trabajo está dedicado a:

Juan Manuel Sánchez Valencia, mi padre, una persona que admiro, que me ha apoyado en las dificultades que he enfrentado a lo largo de mi vida, de forma inconmensurable; despertando mi interés en el estudio de la carrera de derecho y litigar como el lo hace.

María Angélica Contreras González, mi madre, la persona que siempre ha dedicado su tiempo a mi progreso, con su sostén, inquebrantable templanza y enorme cariño, siempre ha sido de las personas más importantes en mi vida.

Mis hermanas **Laura, Lorena y Marisol**, ejemplo a seguir y personas sensatas, empáticas y que me han brindado útiles consejos a lo largo de este camino, es un honor estar a su lado en este camino de la vida.

Al amor de mi vida **Patricia Armas**, que me alentó desde el comienzo de la maestría, siendo una gran persona que ha creído siempre en mí, algo que siempre le agradeceré, además de haber construido su vida y futuro conmigo.

A grandes amigos Jorge, Mario, Elías, Jorge Luis, Nayeli, Alejandro, Eva, Tasia, Tere, Thalía, Sarita, Maricruz, Inocencio, Vero, que han hecho este período de estancia en Ciencias Bioquímicas, un tiempo muy grato y han contribuido a mi crecimiento emocional y profesional.

Una persona que merece una mención especial es el **Dr. Enrique Merino Pérez**, gran investigador e increíble persona, que me guió a lo largo de un proyecto que significaba mucho para mí, me encontraba en una situación difícil post pandemia, el confió en mí, por lo cual toda la vida le estaré agradecido enormemente, siendo un gran guía académico y personal.

Índice General

ABREVIATURAS.....	7
ÍNDICE DE FIGURAS.....	8
ÍNDICE DE TABLAS.....	10
RESUMEN.....	11
INTRODUCCIÓN.....	13
I. <i>Candida glabrata</i>	13
II. Clasificación Filogenética de <i>Candida glabrata</i>	14
III. Patogenicidad de <i>Candida glabrata</i> y otros organismos de la clase Saccharomycetes.....	17
IV. Características genómicas asociadas a la virulencia dentro del género Nakaseomyces.....	21
IV.1 Expansión y diversificación de los genes que codifican para adhesinas y otras proteínas de pared celular en <i>C. glabrata</i>	23
IV.2 Incremento en la capacidad de la formación de biopelículas.....	26
IV.3 Nivel de resistencia a varios tipos de estrés.....	26
IV.4. Sobreexpresión de enzimas con actividad hidrolítica.....	28
IV.5. Reducción del genoma.....	29
V. Mutaciones de <i>Candida glabrata</i> que favorecen su resistencia a antifúngicos.....	32
VI. Aspectos de diferenciación y similitud entre <i>Candida glabrata</i> con <i>Candida albicans</i>	33
VII. Diagnóstico molecular. La importancia de contar con un método rápido, sensible y específico.....	35
JUSTIFICACIÓN.....	43
HIPÓTESIS.....	43
OBJETIVO.....	44
OBJETIVOS PARTICULARES.....	44
METODOLOGÍA.....	45
Estrategia general de análisis.....	45
I. Análisis de la frecuencia absoluta de todas las posibles palabras dentro de la secuencia genómica.....	46
II. Análisis de las propiedades termodinámicas de los potenciales oligonucleótidos para el ensayo de amplificación por PCR.....	47

II.1. Recursos bioinformáticos para el diseño de oligonucleótidos.....	47
II.2. Aspectos relevantes en el diseño de oligonucleótidos.....	49
II.2.1. Longitud del oligonucleótido.....	50
II.2.2. Contenido de G/C del oligonucleótido.....	50
II.2.3. Composición de nucleótidos en el extremo 3'.....	50
II.2.4. Temperatura de fusión (T _m).....	51
II.2.5. Ecuación de Wallace.....	51
II.2.6. Método de vecinos cercanos de (Nearest- Neighbor)	51
II.2.7. Entalpía.....	52
II.2.8. Entropía.....	52
II.2.9. Energía Libre de Gibbs (ΔG)	53
II.2.10. Formación de dímeros.....	53
II.2.11. Concentración de los Iniciadores.....	53
II.2.13. Concentración de la DNA Polimerasa.....	54
II.2.14. Factores que pueden modificar la fidelidad de la enzima DNA polimerasa..	54
II.2.15. Aditivos usados en la reacción de amplificación por PCR.....	54
II.2.16. Recomendaciones generales para la reacción de amplificación por PCR.....	54
RESULTADOS.....	56
I. Análisis de la frecuencia absoluta de todas las posibles palabras dentro de la secuencia genómica.....	56
I.1. Obtención de secuencias genómicas de organismos de estudio.....	56
I.2. Identificación de regiones altamente repetidas en el genoma de <i>C. glabrata</i> a partir de la frecuencia de palabras (k-meros)	56
I.3. Identificación de secuencias satélites en el genoma de <i>C. glabrata</i>	60
I.4. Cálculo de frecuencias de palabras vecinas de los k-meros más frecuentes.....	65
I.5. Índice de visibilidad relativa.....	67
II. Análisis de las propiedades termodinámicas de los potenciales oligonucleótidos para el ensayo de amplificación por PCR.....	71
II.1 Efecto del contenido de G/C en la energía libre de hibridación de los oligonucleótidos.....	71
II.B. Resultados representativos para un par de oligos identificados en el proceso de análisis	

con el genoma de <i>C. glabrata</i>	73
II.3. Organismos de estudio considerados como control negativo.....	78
II.4. Descripción del programa de análisis.....	79
II.D.1. Subrutina lee_parametros.....	79
II.D.2. Subrutina lee_genoma.....	81
II.D.3. Subrutina reverso_complementario.....	82
II.D.4. Subrutina calcula_frecuencia_de_palabras.....	82
II.D.5. Subrutina analiza_palabras_repetidas.....	82
II.D.6. Subrutina identifica_posiciones.....	83
II.D.7. Subrutina seleccion_de_oligos_izquierdos_con_GC_en_extremo_3p.....	83
II.D.8. Subrutina calculo_de_la_TM_oligo_DNA.....	84
II.D.9. Subrutina identifica_el_valor_cutoff_DeltaG.....	84
II.D.10. Subrutina analisis_autocomplementariedad.....	85
II.D.11. Subrutina seleccion_de_oligos_derechos_con_GC_en_extremo_5p.....	85
II.D.12. Subrutina analisis_complementariedad_entre_oligos.....	86
II.D.13. Subrutina analisis_compatibilidad_DeltasG_entre_oligos.....	86
II.D.14. Subrutina busca_matches.....	87
II.D.15. Subrutina hibrida_5p.....	87
Subrutina hibrida_3p.....	87
Programa RNAcofold.....	87
Subrutina compila_y_escribe_resultados_finales.....	88
II.E. Ensayos de amplificación de productos de PCR específicos de <i>C. glabrata</i>	88
CONCLUSIONES	89
PERSPECTIVAS	90
Apéndice 1. Resultados de ensayos de amplificación de productos de PCR específicos de <i>C. glabrata</i> realizados por el Dr. Marco Josué Hernández Chávez del laboratorio de la Dra. Irene Castaño del IPICYT.....	94
Apéndice 2. Lista de organismos considerados en el estudio.....	97
REFERENCIAS BIBLIOGRÁFICAS	107

Abreviaturas

BSA: Medio de cultivo Bovino Serum Albúmina

BSI: Infección del torrente sanguíneo, por sus siglas en inglés

CgPdr1: Factor de transcripción de *C. glabrata* estructurado por racimos de zinc

CWP: Proteínas de pared celular, por sus siglas en inglés

CTG: Codón que codifica a la leucina diferenciando a *C. glabrata* de *S. cerevisiae*

dATP: Desoxiadenosíntrifosfato

DMSO: Acrónimo de dimetilsulfóxido

DNA: Ácido desoxirribonucleico, por sus siglas en inglés.

EPA: Adhesina epitelial

GPI: Glicosil-fosfatidil-inositol

HASH: Estructura de datos que relaciona a cada llave un valor dentro de un diccionario

ITS: Espaciador transcrito interno, por sus siglas en inglés

KEEG: Enciclopedia de genes y genomas de Kyoto

KMERO: Subcadenas de longitud K contenidas dentro de una secuencia biológica.

µM: Micrómetro

MSH2: Gen que da origen a una proteína que participa en reparación de DNA

NAD⁺: Nicotinamida adenina dinucleótido

PAMPS: Patrones moleculares asociados a patógenos

PCR: Reacción en cadena de la polimerasa

Python: Lenguaje de alto nivel de programación

QTL'S: Locus de caracteres cuantitativos

RFLP: Fragmentos de restricción de longitud polimórfica

RAPDS: Amplificación aleatoria de DNA polimórfico

SNP: Polimorfismo de nucleótido único

SSR: Marcador genético para análisis de levaduras

TRP: Canal iónico de potencial receptor transitorio, presente en levaduras

Tm: Temperatura de fusión, que representa la temperatura en la cual se separan el 50% de las dos hebras complementarias de un oligonucleótido

VNTR: Secuencias repetitivas que aparecen en múltiples zonas del genoma, organizando bloques dispersos de repeticiones en tándem.

Índice de Figuras

Figura 1. Relaciones filogenéticas entre <i>C. glabrata</i> y las otras especies de organismos de la clase Saccharomycetes.....	15
Figura 2. Distribución mundial de incidencia de las especies de la clase Saccharomycetes.....	20
Figura 3. Árbol Filogenético del género Nakaseomyces y posibles eventos relevantes que definen sus principales grupos.....	22
Figura 4. Estructura de la pared del celular de las levaduras del género Nakaseomyces.....	24
Figura 5. Distribución filogenética de la familia de adhesinas en el género Nakaseomyces.....	25
Figura 6. Factores relacionados con la patogenicidad de <i>C. glabrata</i>	29
Figura 7. Representación esquemática de familias genéticas específicas en el clado <i>Nakaseomyces</i>	31
Figura 8. Detección de un Polimorfismos de un solo nucleótido (SNP).....	36
Figura 9. Polimorfismo de Longitud de Fragmentos de Restricción (PCR-RFLP).....	38
Figura 10. Principales tipos de marcadores de DNA y su principal uso en la identificación de <i>C. glabrata</i>	41
Figura 11. Principales tipos de marcadores de DNA y oligonucleótidos en <i>C. glabrata</i>	42
Figura 12. Portal web de OligoAnalyzer para el diseño de oligonucleótidos.....	48
Figura 13. Portal web de GenScript para el diseño de oligonucleótidos.....	48
Figura 14. Portal de Thermo Fisher para el diseño de oligonucleótidos.....	49
Figura 15. Diseño de Oligonucleótidos para la identificación de microsatélites.....	50
Figura 16. Ejemplo del cálculo de entalpía.....	52
Figura 17. Ejemplo del cálculo de entropía.....	53
Figura 18. Cálculo de frecuencias absolutas de palabras mediante el uso de estructuras Hash.....	57
Figura 19. Cálculo de la frecuencia esperado de k-meros en función de su longitud, considerando una secuencia del tamaño del genoma de <i>C. glabrata</i>	58
Figura 20. Evaluación de k-meros más frecuentes en el genoma de <i>C. glabrata</i>	59
Figura 21. Cálculo de la distancia entre palabras repetidas en el genoma de <i>C. glabrata</i> que pudieran indicar la presencia de secuencias satélites.....	60
Figura 22. Palabras más frecuentes equidistantes en el genoma de <i>C. glabrata</i>	61
Figura 23. Análisis de palabras que equidistan 300 nucleótidos en el genoma de <i>C. glabrata</i>	62

Figura 24. Análisis de palabras que equidistan 141 nucleótidos en el genoma de <i>C. glabrata</i>	63
Figura 25. Análisis de palabras que equidistan 450 nucleótidos en el genoma de <i>C. glabrata</i>	64
Figura 26. Representación del cálculo de frecuencias de palabras vecinas.....	67
Figura 27. Evaluación del <i>índice de la visibilidad relativa</i> para dos k-meros que se repiten 41 veces en el genoma de <i>C. glabrata</i>	70
Figura 28. Efecto del contenido de G/C de oligos en su energía libre de hibridación.....	72
Figura 29. Tipos de complementariedad de oligos y su compatibilidad para la amplificación en ensayos de PCR.....	73
Figura 30. Ejemplo de la salida final del programa del diseño automatizado de oligonucleótidos a ser usados en ensayos de amplificación específica por PCR.....	77
Figura 31. Diagrama de flujo para la identificación de K-meros frecuentes en secuencias genómicas y el diseño de oligonucleótidos considerando el análisis termodinámico de sus propiedades.....	80
Figura 32. Página web para el diseño de óptimo de oligonucleótidos a ser usados en ensayos de identificación de organismos específicos.....	90
Figura 33. Productos de la reacción de PCR utilizando los oligos # 1 y # 2 utilizando DNA genómico de <i>C. glabrata</i> y especies relacionadas.....	92
Figura 34. Productos de la reacción de PCR utilizando los oligos # 3y # 4 utilizando DNA genómico de <i>C. glabrata</i> y especies relacionadas.....	92
Figura 35. Productos de la reacción de PCR utilizando los oligos # 15 y # 16 utilizando DNA genómico de <i>C. glabrata</i> y especies relacionadas.....	93

Índice de Tablas

Tabla 1. Acrónimos y funciones de las familias de genes de la Figura 8.....	31
Tabla 2. Porcentaje de similitud de secuencia de nucleótidos de genes de <i>C. dubliniensis</i> a las secuencias correspondientes de microarreglos de <i>C. albicans</i>	40
Tabla 3. Evaluación del <i>índice de la visibilidad relativa</i> de las palabras más frecuentes en el genoma de <i>C. glabrata</i>	69
Tabla 4. Organismos de estudio considerados como control negativo.....	78
Tabla 5. Lista de oligonucleótidos elegidos en el ensayo de identificación de <i>C. glabrata</i> mediante la amplificación por PCR.....	91
Tabla 6. Condiciones para las reacciones de PCR en el ensayo de identificación de <i>C. glabrata</i> usando los pares de oligos 1-2, 3-4 así como 15-16.....	91

Resumen

Las infecciones fúngicas se han vuelto cada vez más frecuentes en los hospitales de todo el mundo, por el aumento de la aparición de ecotipos resistentes a antimicrobianos que se originan principalmente por el incremento de pacientes con sistemas inmunológicos severamente comprometidos. Dentro de estas infecciones micóticas, de particular preocupación son las infecciones del torrente sanguíneo (BSI, por sus siglas en inglés), algunas causadas por varias especies del género *Nakaseomyces*, anteriormente denominado *Candida* (candidemia). Estas infecciones están asociadas con altas tasas de mortalidad en hospitales (~ 30 a 70%) y representan la cuarta causa más común de infecciones hematógenas (se propagan vía sanguínea) en las unidades de cuidados intensivos en los E.E.U.U. (Casadevall *et al.*, 2018).

En México, las infecciones del torrente sanguíneo (BSI, por sus siglas en inglés) causadas por el género *Nakaseomyces* representan casi el 4% de las infecciones nosocomiales, con una tasa de mortalidad que oscila entre el 35% y el 45% (Corzo-León, *et al.*, 2014). Aproximadamente el 90% de las candidemias son provocadas por cuatro especies: *Candida albicans*, *Candida glabrata*, *Candida parapsilosis* y *Candida tropicalis*. Es importante destacar que cada una de estas especies presenta diferentes niveles de susceptibilidad a los distintos antifúngicos disponibles en el ámbito clínico. Por lo tanto, el diagnóstico preciso de la especie causante de la candidemia es crucial para el tratamiento efectivo.

Actualmente, los métodos clínicos de detección más comunes se basan en ensayos de reacción en cadena de la polimerasa (PCR, por sus siglas en inglés, Polymerase Chain Reaction) (Mullis *et al.*, 1986). Estos ensayos generalmente utilizan oligonucleótidos específicos de regiones variables entre las distintas especies, como las regiones espaciadoras transcritas internas (ITS) del arreglo de genes que codifican diferentes genes RNA ribosomales o polimorfismos específicos de especies. Sin embargo, la sensibilidad y especificidad alcanzadas por estos análisis suelen ser moderadas, y una desventaja adicional es el tiempo requerido para la realización del ensayo, lo que resulta en la obtención tardía de los resultados.

La presente tesis se enmarca en el proyecto de investigación titulado "Plasticidad genómica de aislados clínicos secuenciales del hongo patógeno de humanos *Candida glabrata*: consecuencias para la resistencia a antifúngicos y la capacidad de adhesión", financiado por FORDECYT-PRONACES/610281/2020. Tiene como objetivo desarrollar una metodología que permita el diseño óptimo de oligonucleótidos dirigidos a regiones específicas del genoma de *C. glabrata* que estén altamente repetidas, facilitando su detección mediante ensayos de PCR con alta sensibilidad y especificidad.

Es importante destacar que los principios utilizados en el desarrollo de nuestra metodología no se limitan exclusivamente a *C. glabrata*. Por lo tanto, nuestro método de diseño de oligonucleótidos puede ser aplicado para la detección de diversos organismos de interés en cualquier tipo de muestra.

Introducción

I. *Candida glabrata*

Los hongos, pertenecientes al reino Fungi, exhiben una enorme diversidad, con más de 1,5 millones de especies estimadas. Sin embargo, solo una pequeña porción de ellas es patógena para el ser humano (David *et al.*, 2001). Un ejemplo claro de importancia clínica por su potencialidad para causar enfermedades en el humano es el hongo *Candida glabrata*.

C. glabrata se nombró por primera vez como *Cryptococcus glabratus* y se describió en 1917 como un componente de la flora intestinal humana (Anderson *et al.*, 1917; revisado en Bolotin-Fukuhara *et al.*, 2014). Años más tarde, fue identificada como la fuente de varias infecciones denominadas *Cryptococcus glabratus* (Plaut *et al.*, 1950) o *Torulopsis glabrata* (Grimley *et al.*, 1965). A finales de la década de 1980, fue denominada *Candida glabrata*, y más recientemente, fue renombrada como *Nakaseomyces glabratus*.

C. glabrata es una especie de hongo unicelular que forma parte del género *Nakaseomyces* y está ampliamente distribuida, tanto en el medio ambiente, como en el cuerpo humano. *C. glabrata* es considerada un organismo comensal, ya que normalmente coexiste con su hospedero sin causar enfermedades. Sin embargo, *C. glabrata* puede transformarse en un patógeno oportunista y desencadenar infecciones graves, como la candidiasis invasiva, la candidemia en determinadas circunstancias, como en el caso de personas con sistemas inmunológicos debilitados, pacientes con VIH/SIDA o pacientes oncológicos sometidos a quimioterapia, receptores de trasplantes, personas mayores o niños prematuros recién nacidos y pacientes hospitalizados que están conectados a dispositivos médicos invasivos, como catéteres venosos centrales o sondas urinarias, o bien, pacientes que han tomado antibióticos por períodos de tiempos muy largos. (Brown *et al.*, 2012)

El término candidiasis es usado *de facto* y por semántica, para designar una infección por hongos, cuya etiología sea cualquier tipo de levadura del género *Nakaseomyces*. Estas infecciones pueden afectar varias superficies o mucosas del cuerpo,

se encuentra comúnmente en la piel, en el tracto gastrointestinal y en el tracto genitourinario, o pueden volverse invasivas y llegar a los órganos internos, a través del torrente sanguíneo y diseminarse por todo el cuerpo, en cuyo caso la infección se llama candidemia. (Ghannoum *et al.*, 2010).

II. Clasificación Filogenética de *Candida glabrata*

Según el Centro Nacional de Información Biotecnológica (NCBI, por sus siglas en inglés) (<https://www.ncbi.nlm.nih.gov/taxonomy>) de los Institutos Nacionales de Salud (NIH), *C. glabrata* se puede clasificar taxonómicamente de la siguiente manera: **Super-reino:** Eucariota, **Reino:** Fungi, **Filo:** Ascomycota, **Clase:** Saccharomycetes, **Orden:** Saccharomycetales, **Familia:** Saccharomycetaceae y **Género:** Nakaseomyces.

Con el propósito de ubicar evolutivamente a *C. glabrata* con otras especies pertenecientes al orden Saccharomycetales, Gabaldón y colaboradores realizaron un estudio filogenético usando las secuencias de 603 proteínas ortólogas comunes entre los organismos de este clado filogenético. Dichas secuencias fueron concatenadas y alineadas para su comparación usando el método de Máxima verosimilitud implementado en el programa phyML (Gabaldón *et al.*, 2013). El resultado de dicho estudio se muestra en la Figura 1.

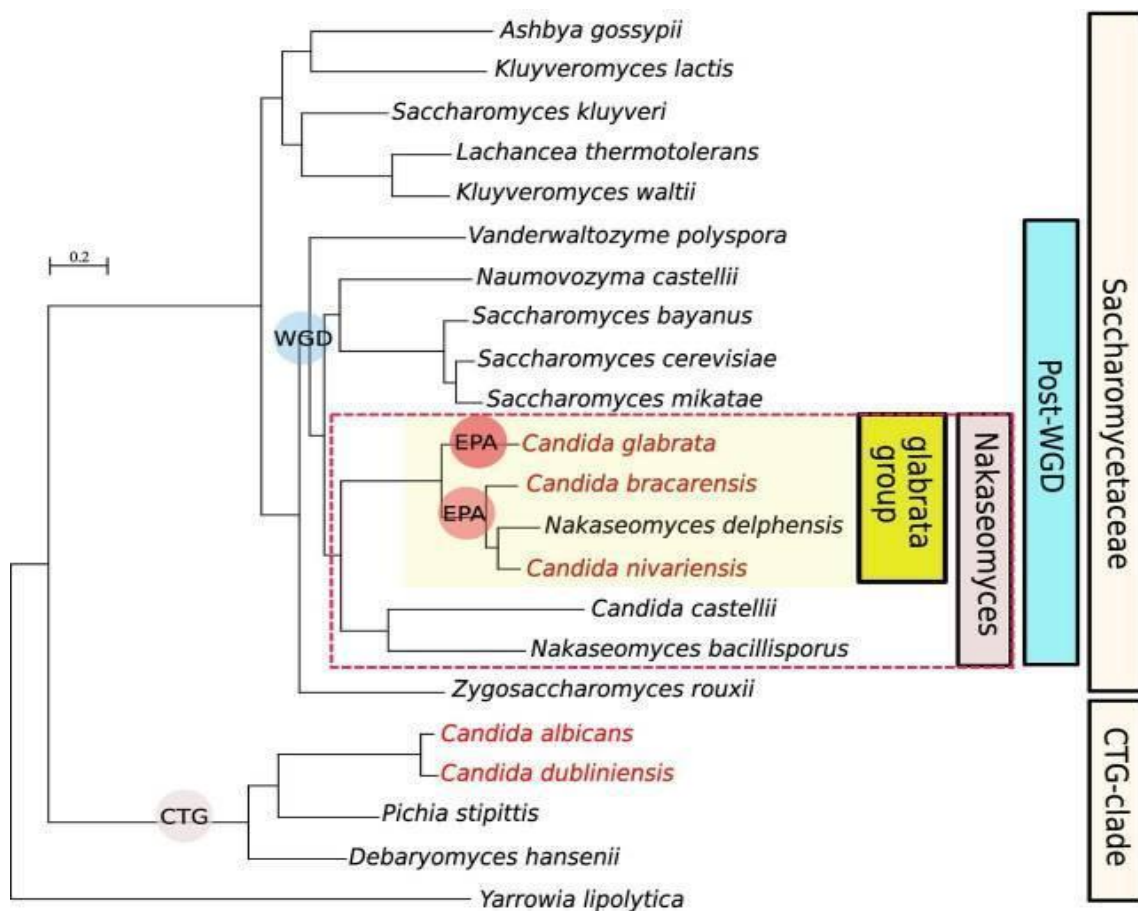


Figura 1. Relaciones filogenéticas entre *C. glabrata* y las otras especies de organismos de la clase Saccharomycetes. El árbol filogenético de la figura se construyó mediante la alineación concatenada de 603 genes ortólogos y comunes a organismos de la clase Saccharomycetes. Los genes ortólogos se identificaron utilizando el algoritmo de búsqueda de Smith-Waterman, con un valor de corte de 0.00001 y considerando únicamente aquellas secuencias cuya región de similitud abarcara al menos el 50% respecto a la secuencia de consulta. La construcción del árbol se llevó a cabo utilizando el método de Máxima Verosimilitud, implementado en el programa PhyML. En el rectángulo azul se agrupan los Nakaseomyces y otros integrantes del género Nakaseomyces. Las letras "WGD" dentro de un círculo azul representan el evento de la duplicación del genoma completo en algunos Saccharomycetes. Estos organismos comparten el evento de duplicación de los genes de adhesinas de la pared celular (EPA), representados mediante círculos rojos. Los nombres en rojo indican patógenos fúngicos de humanos. Las letras "CTG" dentro de un círculo café indican una transición del código genético, en donde el codón CTG codifica el aminoácido serina en lugar de leucina, como es comúnmente codificado en la mayoría de los organismos. Para la construcción del árbol se utilizó como grupo externo al hongo *Yarrowia lipolytica* (Tomado del artículo de Gabaldón *et al.*, 2016).

En el árbol de la Figura 1 se identifican dos grandes ramas en las que se dividen los organismos de la clase Saccharomycetes: En la primera rama se agrupan los organismos de la familia Saccharomycetaceae, mientras que la segunda rama corresponde al grupo CTG.

La primera de las ramas de la clase Saccharomycetes correspondiente a la familia Saccharomycetaceae se divide a su vez en dos ramas. En la primera de ellas se encuentran levaduras como *Ashbya gossypii*, *Saccharomyces kluyveri*, *Lachancea thermotolerans*, *Kluyveromyces waltii* y *Kluyveromyces lactis*. Éste último, es considerado como un organismo modelo y ha sido ampliamente estudiado debido a su relevancia biotecnológica y a su fácil manipulación genética. *K. lactis* se encuentra comúnmente en productos lácteos fermentados, donde desempeña un papel importante en la producción de sabores y texturas características. Además, es capaz de fermentar la lactosa, lo que la hace útil para la producción de productos lácteos libres de este azúcar. Adicionalmente, la capacidad de secreción de *K. lactis* ha sido utilizada para el desarrollo de sistemas de producción de proteínas recombinantes que pueden ser eficientemente secretadas al medio. Desde el punto de vista básico, la regulación genética y la biología del desarrollo de *K. lactis* han sido estudiadas por décadas y son bien conocidas.

Los organismos de la segunda rama en la que se divide la familia Saccharomycetaceae se caracterizan por haber sufrido un evento de duplicación que abarcó la totalidad de sus genomas, evento que se calcula haya sucedido hace 100 a 200 millones de años, aunque ahora se propone que este ancestro fue una especie híbrida (Wolfe *et al.*, 1997; Marcet-Houben *et al.*, 2015) (Figura 1, indicado por el círculo azul, con las letras WGD, o Whole Genome Duplication y por el rectángulo de color azul aqua con la leyenda Post-WGD). Los organismos pertenecientes a esta segunda rama de la familia Saccharomycetaceae a su vez se subdividen en dos grupos. En el primero de ellos se ubican diversas especies de la clase Saccharomyces, como *Saccharomyces bayanus*, *Saccharomyces mikatae* y *Saccharomyces cerevisiae*. Esta última especie es ampliamente conocida por su utilización en la elaboración del pan y de la cerveza y porque es uno de los organismos eucariontes modelo para el cual se han desarrollado un gran conjunto de técnicas de biología molecular, por lo que ha sido ampliamente estudiada. Los organismos

de la segunda de las ramas en la que se divide la familia Saccharomycetaceae pertenecen al género *Nakaseomyces* (Kurtzman *et al.*, 2003), entre los cuales se encuentran diversas especies del género *Candida*, tales como, *Candida glabrata*, *Candida bracarensis*, *Candida nivariensis*, *Nakaseomyces delphensis*, *Candida castellii* y *Nakaseomyces bacillisporus*. Recientemente, debido a su inclusión en este clado, *C. glabrata* ha sido renombrada como *Nakaseomyces glabratus* (Takashima *et al.*, 2022), sin embargo, para evitar confusiones y dada su relevancia clínica, el nombre de *Candida glabrata* sigue usándose como referencia.

Por otra parte, la segunda rama de la familia Saccharomycetaceae está conformada por organismos del grupo denominado CTG. El nombre de este grupo ha sido asignado debido a que los organismos que lo conforman comparten la característica de que el codón CTG codifica al aminoácido serina, a diferencia de la mayoría de los organismos en el que dicho codón codifica al aminoácido leucina. En este grupo se encuentran especies de levaduras tales como *Candida albicans*, *Candida dubliniensis*, *Pichia stipitis*, y *Debaryomyces hansenii* (Figura 1).

III. Patogenicidad de *Candida glabrata* y otros organismos de la clase Saccharomycetes

La virulencia se define como la capacidad de un organismo de causar una enfermedad en un hospedero determinado. Por tanto, la virulencia microbiana es necesariamente el resultado de una interacción específica entre el hospedero y el microorganismo, y los cambios en el hospedero en el microorganismo pueden afectar el grado de virulencia. Además, la virulencia puede considerarse una propiedad emergente (Casadevall *et al.*, 2011). Esta interacción entre los organismos patógenos y sus hospederos depende en gran medida, si estos son especializados u oportunistas. En el caso de los patógenos especializados, las interacciones forman parte intrínseca de su ciclo de vida. Por el contrario, en el caso de los patógenos oportunistas, la virulencia tiende a manifestarse principalmente cuando el hospedero está debilitado y se presentan condiciones propicias para su adaptación y proliferación. No obstante, estas interacciones

suelen ocurrir con frecuencias inferiores a las necesarias para que los rasgos adaptativos estén sujetos a selección en el contexto de toda la población y la historia evolutiva de una especie dada. Por lo anterior, es difícil considerar la virulencia en patógenos oportunistas como un fenotipo adaptativo *per se* (Cui *et al.*, 2013). En este mismo sentido, es común que en especies que interactúan regularmente con el hospedero, como los comensales, la selección actúe para atenuar la virulencia en condiciones regulares, porque la supervivencia del hospedero está vinculada a la supervivencia del comensal. Además, en patógenos oportunistas, podemos considerar la virulencia como un fenotipo mayoritariamente latente, que sólo se expresa en infecciones bajo condiciones particulares. (Arendrup *et al.*, 2012).

Muchas especies de la clase Saccharomycetes son importantes patógenos oportunistas en humanos y otros organismos, lo que las convierte en objeto de estudio en microbiología médica y biología molecular. Ejemplo de dichos organismos son los mostrados con letras en color rojo de la Figura 1. La candidiasis es un tipo de infección causada por este tipo de organismos y representa aproximadamente el 85% de todas las infecciones fúngicas invasivas, por lo que se considera una seria amenaza para la vida, particularmente en individuos inmunodeprimidos, con tasas de mortalidad que alcanzan el 46% -75% (Mayer *et al.*, 2013). Se han identificado más de 20 especies de *Candida* que pueden causar candidiasis, aunque la incidencia varía mucho de una especie a otra. Entre éstas, las más relevantes por su distribución mundial son: *Candida albicans*, *C. glabrata*, *Candida tropicalis* y *Candida parapsilosis*.

Durante mucho tiempo se creyó que *C. glabrata* era la única especie patógena entre las Saccharomycetaceae, postulando que su patogenicidad surgió como una discontinuidad drástica en términos evolutivos debido a la duplicación total de su genoma. (Jones *et al.*, 2004). En general, la información sobre el hábitat y la ecología de las diversas especies que componen la familia Saccharomycetaceae es escasa. La única información disponible se deriva de sus aislamientos: *C. castellii* fue encontrado en una muestra de suelo ácido en un pantano de pinos en Finlandia (Capriotti *et al.*, 1961), *N. delphensis* fue aislado de un depósito de higos secos en Sudáfrica, y *N. bacillisporus* fue aislado de exudados del roble rojo de Emory en Arizona, EE. UU. (Kurtzman *et al.*, 2011). Esta primera información llevó a asociar al estilo de vida ambiental de los organismos del clado Saccharomycetaceae con

plantas, alimentándose de fuentes frutales y dispersándose por insectos, en lugar de considerarlos patógenos humanos. Sin embargo, poco después de definir el clado Nakaseomyces, se describieron dos nuevas especies dentro de este grupo a partir de muestras clínicas humanas: *C. nivariensis* (Alcoba-Flórez *et al.*, 2005) y *C. bracarensis* (Correia *et al.*, 2006) (Warren *et al.*, 2010). Inicialmente, estas especies se caracterizaron fenotípicamente, utilizando pruebas bioquímicas como CHROMagar, y mediante amplificación, secuenciación y análisis de los dominios D1/D2, que tienen una longitud de 600 nucleótidos y se localizan en el extremo 5' de la subunidad 26S del ADNr 26S, que permite la clasificación eficiente del reino Fungi (Kurtzmann *et al.*, 1991).

La identificación de *C. nivariensis* y *C. bracarensis* dentro del clado Saccharomycetaceae llevó a considerar la posibilidad de que varios aislamientos clínicos previamente clasificados como *C. glabrata* pudieran, en realidad, pertenecer a estas especies (Bishop *et al.*, 2008; Lockhart *et al.*, 2009). Aunque aún falta mucha información fenotípica sobre estos patógenos emergentes, los primeros análisis sugieren que estas especies también comparten algunos de los atributos de virulencia de *C. glabrata*, aunque en menor medida, lo que sugiere una incidencia considerablemente menor que la de *C. glabrata* (Moreira *et al.*, 2015). Este nuevo panorama ha reducido el enfoque en la patogenicidad exclusiva de *C. glabrata*, por lo que se ha pensado que su origen podría estar en un clado ambiental que también incluye patógenos menos virulentos. Específicamente, la presencia de patógenos estrechamente relacionados con una virulencia intermedia hacia los humanos ha generado la idea de que pudiera existir un aumento gradual de la virulencia en diversas especies dentro del clado, tal y como se muestra en la Figura 2 (Kurtzman *et al.*, 2011). Además, la capacidad que *C. glabrata* tiene para desarrollar resistencia a los antifúngicos comúnmente empleados complica su tratamiento y aumenta el riesgo de infecciones recurrentes y persistentes. Este fenómeno ha despertado un interés creciente en la comunidad científica por comprender mejor la biología, la patogénesis y los mecanismos de resistencia de *C. glabrata* y de otros hongos filogenéticamente cercanos, con el objetivo de desarrollar estrategias terapéuticas más efectivas y abordar eficazmente las infecciones fúngicas.

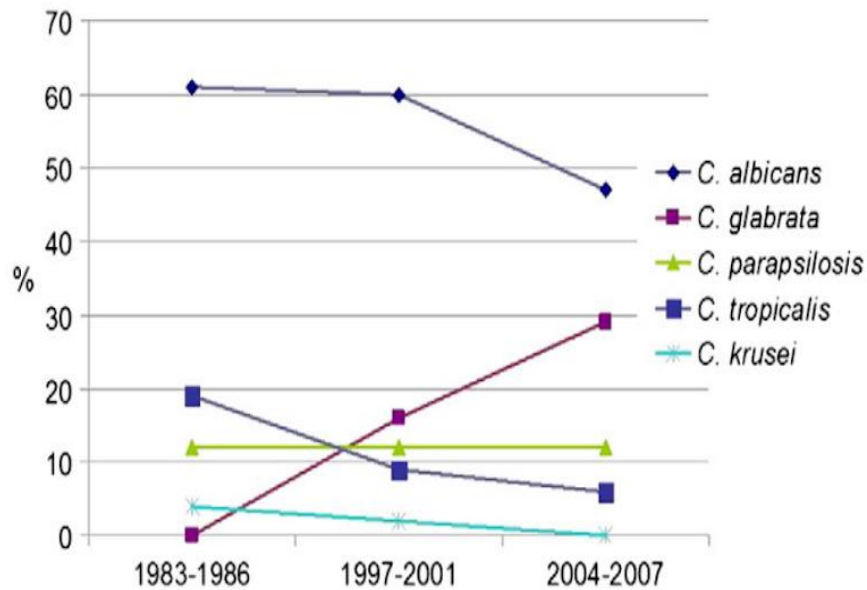


Figura 2. Distribución mundial de incidencia de las especies de la clase *Saccharomycetes*. Se aprecia el aumento de los casos de infección por *C. glabrata* desde su detección como patógeno de humanos (Tomado de Diekema *et al.*, 2012).

Las comparaciones de las secuencias genómicas de organismos patógenos y no patógenos del género *Nakaseomyces* permite identificar aquellos cambios potencialmente relacionados al surgimiento de la de virulencia, como son el aumento en la capacidad de adherencia y de generar biofilm (Kaur *et al.*, 2005). Se ha propuesto que dichas propiedades existían en los organismos no patógenos de vida libre, pero que las mucosas de mamíferos pudieron haber ofrecido las condiciones idóneas para su crecimiento y el surgimiento de especies patógenas. En términos evolutivos, este proceso se conoce como exaptación, un término que indica que, un rasgo que se seleccionó inicialmente para un propósito determinado puede servir para otro completamente diferente, que en el caso del clado *Nakaseomyces*, es la patogenicidad. El hallazgo de otros patógenos cercanos a *C. glabrata* y la observación de que formaron un clado estrecho en el que tres de las cuatro especies conocidas pueden infectar a los humanos, aunque no son monofiléticos, sugiere que el grupo *Glabrata* comparte algunos antecedentes genómicos que predisponen al clado a la aparición de rasgos de virulencia. De los muchos rasgos que son relevantes para la patogénesis de *C. glabrata*, la adherencia se destaca como uno de los pasos importantes en el camino para convertirse en una especie con potencial patógeno. El hecho de que las

expansiones de los genes que codifican a las adhesinas hayan aparecido dos veces de forma independiente y que correspondan a linajes con especies patógenas refuerza la idea de que el cambio en las propiedades de adherencia, dado un trasfondo genómico apropiado, puede predisponer a las especies de este clado a convertirse en patógenos oportunistas (Ho *et al.*, 2015).

IV. Características genómicas asociadas a la virulencia dentro del género *Nakaseomyces*.

Con el propósito de tener una mejor idea de las características genómicas asociadas a la virulencia dentro del género *Nakaseomyces*, Gabaldón y colaboradores secuenciaron los genomas de dos especies patógenas de este clado, *Candida nivariensis* y *Candida bracarensis*, y tres no patógenas, *Nakaseomyces delphensis*, *Candida castellii* y *Nakaseomyces bacillisporus* (Gabaldón *et al.*, 2013). El análisis comparativo de las secuencias genómicas de estas especies en relación al de la secuencia genómica de *Saccharomyces cerevisiae* permitió identificar eventos de pérdida y ganancia de genes y funciones relacionadas que pudieran haber sido relevantes en la definición del género *Nakaseomyces*. Éstas se resumen de forma esquemática en la Figura 3.

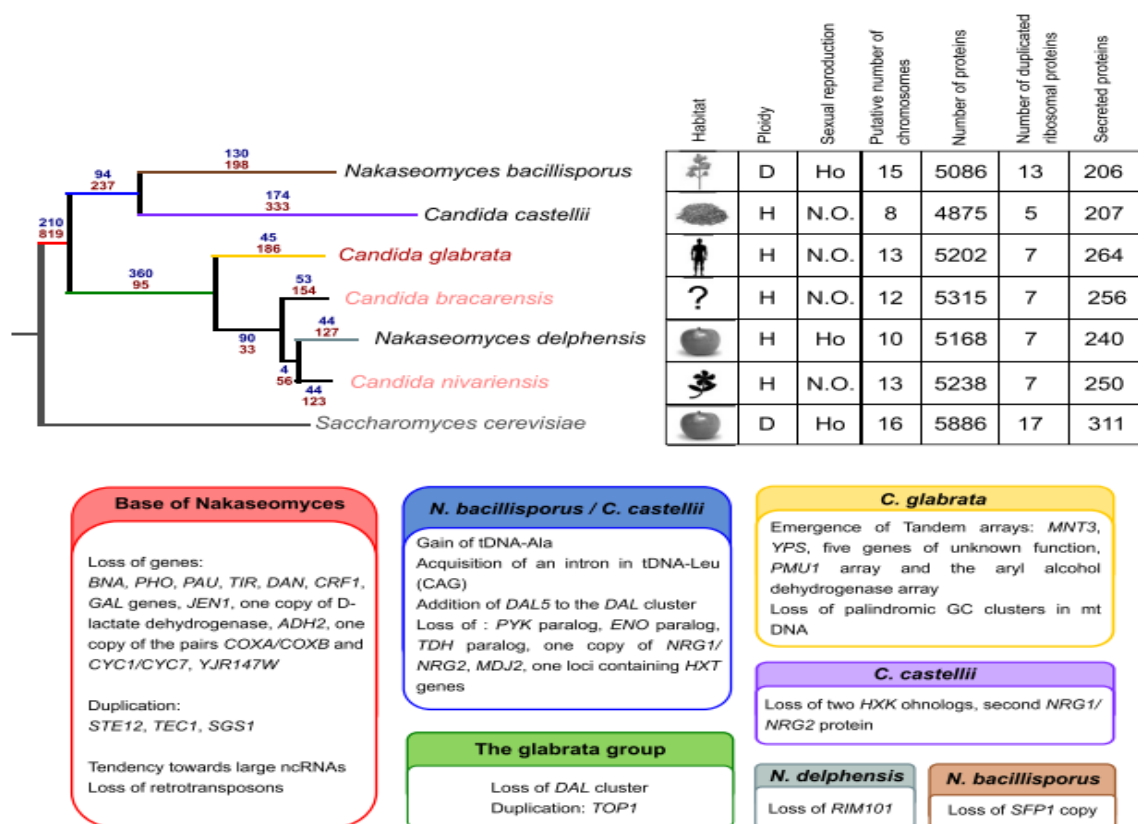


Figura 3. Árbol Filogenético del género *Nakaseomyces* y posibles eventos relevantes que definen sus principales grupos. Los nombres de las especies en colores rojo corresponden a patógenos de humano, mientras que los de colores más claros indican patógenos emergentes recientemente reportados. Los números azules indican los genes obtenidos durante la evolución del género de *Nakaseomyces*; los rojos indican el número de los genes perdidos. Los colores de las ramas del árbol corresponden al de los cuadrados en la parte inferior, que esquematizan eventos importantes en la evolución del género de *Nakaseomyces*. El árbol fue construido tomando en cuenta a *S. cerevisiae* como grupo externo (Tomada de Gabaldón *et al.*, 2013).

La mayoría de las personas sanas son portadoras de varios patógenos oportunistas de levadura como miembros normales de su microbiota (Cui *et al.*, 2013). La virulencia de los organismos del género de *Nakaseomyces* generalmente surge de alteraciones en las interacciones hospedero-microorganismo, principalmente iniciadas en el lado del hospedero. No obstante, dadas las mismas alteraciones en el lado del hospedero, algunas especies como *C. albicans* o *C. glabrata*, tienen una mayor probabilidad de causar enfermedades en comparación con otras levaduras coexistentes, dado el mismo contexto de alteraciones en el hospedero, lo que sugiere preguntar sobre el tipo de modificaciones y reordenamientos genómicos que subyacen el surgimiento de la virulencia en este grupo

filogenético (Arendrup *et al.*, 2012). Dentro de las diferencias más significativas identificadas al comparar el tipo de genes adquiridos y perdidos entre organismos patógenos y no patógenos se encuentran las siguientes:

IV.1. Expansión y diversificación de los genes que codifican para adhesinas y otras proteínas de pared celular en *C. glabrata*.

Una de las características asociadas a las especies patógenas del género *Nakaseomyces* es su capacidad de sobreexpresión de adhesinas. Las adhesinas son proteínas de superficie que las células microbianas utilizan para adherirse a diferentes tipos de superficies. En el caso de las levaduras, estas adhesinas desempeñan un papel crucial en la patogenicidad, incrementando su capacidad para colonizar y causar infecciones en los tejidos humanos, ya que pueden aumentar significativamente su capacidad para adherirse a las células hospedadoras o a los tejidos, lo que facilita la colonización y la formación de biofilm, que pueden proteger a las levaduras de los ataques del sistema inmunológico y de los tratamientos antimicrobianos (Groot *et al.*, 2013; Fanning *et al.*, 2012). Por ejemplo, en *C. albicans*, la sobreexpresión de adhesinas como Als3 ha sido asociada con una mayor capacidad de unión a las células epiteliales y a la formación de biofilm en mucosas (Arendrup *et al.*, 2012), mientras que la lectina dependiente de calcio, *EPA1*, juega un papel relevante en la adhesión de *C. glabrata* a las células epiteliales y a los macrófagos (Kuhn *et al.*, 2012).

Se ha determinado que la cepa de referencia de *C. glabrata* ATCC 200, posee aproximadamente 18 genes que codifican para proteínas de la pared celular (CWPs, por sus siglas en inglés), que incluye una señal de anclaje a la membrana a través del grupo glicosilfosfatidilinositol (GPI). (Hazen *et al.*, 1995). Estas proteínas son similares a las adhesinas y median en parte la virulencia y la adherencia a las células del hospedero. Las proteínas CWPs-GPI están compuestas por un dominio N-terminal de unión al ligando y una región de baja complejidad con repeticiones internas en tándem, rica en los aminoácidos serina y treonina, seguidas de un sitio de unión de anclaje de GPI en el extremo carboxilo (Groot *et al.*, 2013) (Figura 4).

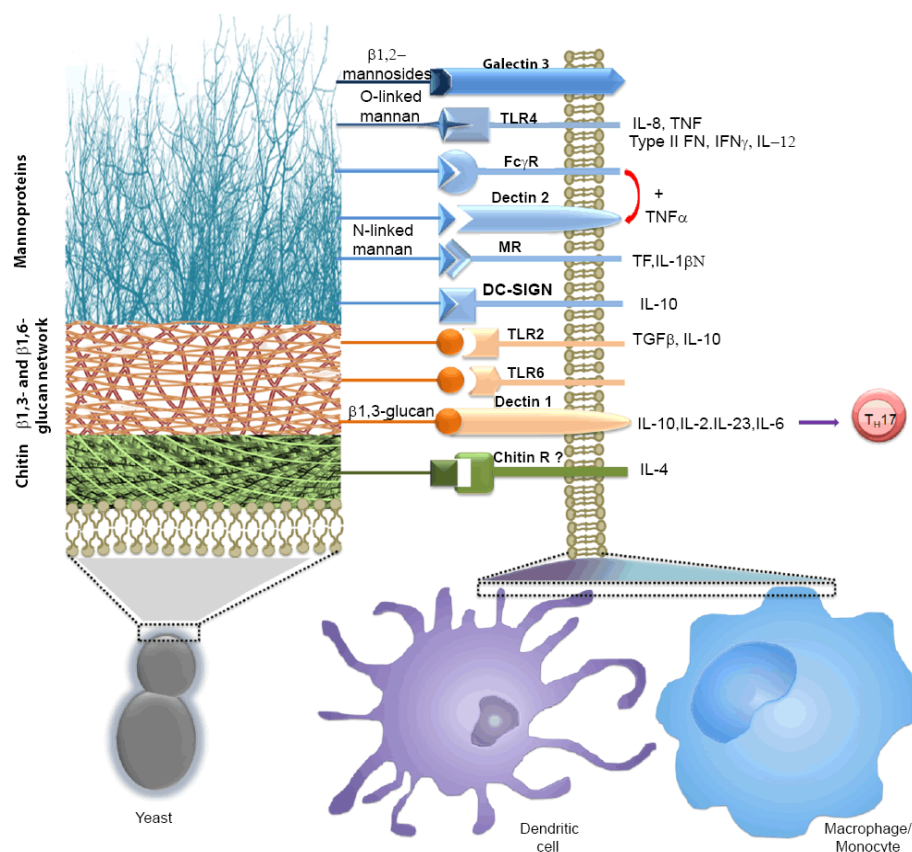


Figura 4. Estructura de la pared celular de las levaduras del género *Nakaseomyces*. Se muestran diversos elementos que conforman a la pared celular, como la quitina, manoproteínas y glucanos, entre otros. La pared celular desencadena la activación del sistema inmune como monocitos, macrófagos y células dendríticas, estimulando la producción de citocinas antiinflamatorias y proinflamatorias (Tomada de Pérez-García *et al.*, 2012).

En 2013, Gabaldón y colaboradores realizaron un estudio bioinformático para la identificación y comparación de los genes homólogos de los genes EPA de *C. glabrata* en los organismos del género *Nakaseomyces*. El estudio reveló que existe una correlación positiva entre el nivel de patogenicidad de las cepas y el número de genes EPA de sus correspondientes genomas. Así, por ejemplo, la secuencia genómica del patógeno más frecuente considerado en el estudio, la cepa BG2 de *C. glabrata*, posee 18 miembros de la familia EPA, mientras que, en especies menos patógenas, *C. bracarensis* y *C. nivariensis*, se identificaron 12 y 9 genes EPA, respectivamente. Por el contrario, la especie no patógena, *N. delphensis*, sólo posee una única copia de un gen homólogo lejano a los genes

EPA de *C. glabrata*. La distribución filogenética de los genes identificados en el estudio se representa en la Figura 5.

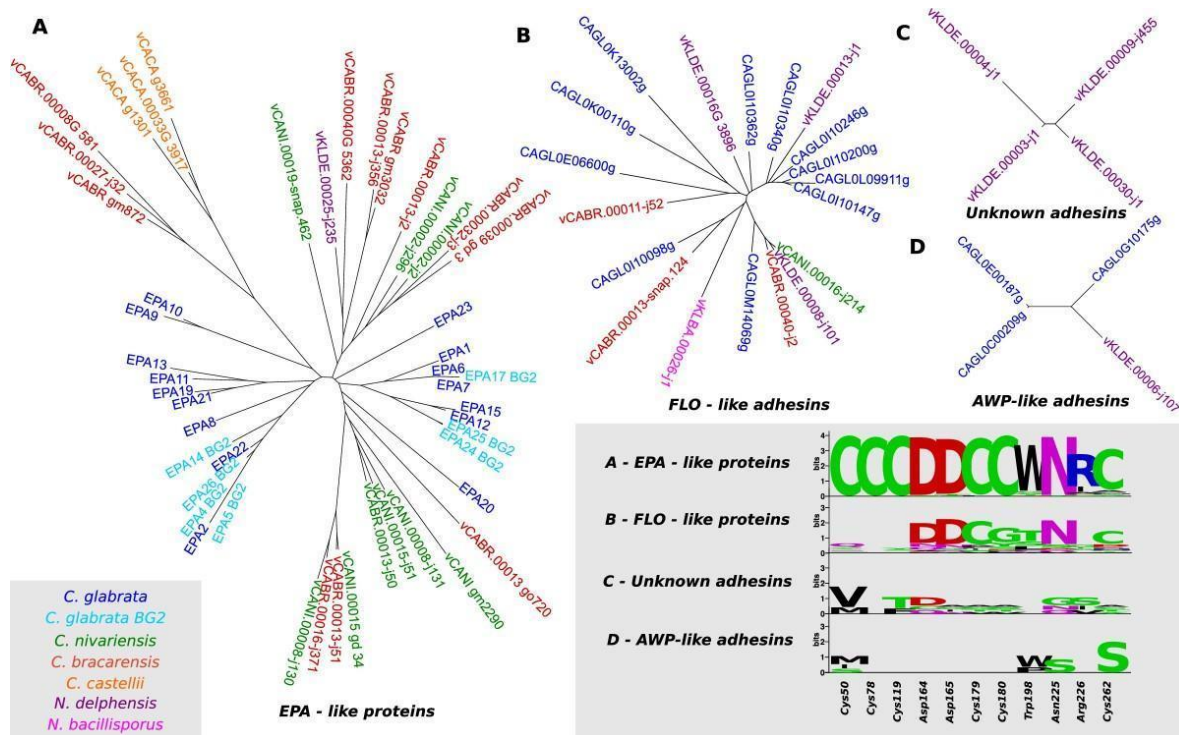


Figura 5. Distribución filogenética de la familia de adhesinas en el género *Nakaseomyces*. A) Proteínas agrupadas con los genes EPA (Adhesina Epitelial) de *C. glabrata* conocidos, B) Proteínas similares a la proteína floculina de levadura. C) Grupo de adhesinas de función desconocida. D) Secuencias relacionadas con adhesinas AWP. El color indica cada una de las especies con su respectiva proteína: *C. glabrata* en azul oscuro, cepa de *C. glabrata* (BG2) en azul claro, *C. bracarensis* en rojo, *C. nivariensis* en verde, *N. delphinensis* en morado, *N. bacillisporus* en amarillo y *C. castellii* en naranja. El cuadro inferior derecho, representa la conservación en cada uno de los grupos de adhesinas de sitios que se consideran importantes en la estructura de *EPA1* (Tomada de Gabaldón *et al.*, 2013).

Algunos factores pueden predisponer a los genes de adhesinas de la pared celular (EPA) a una rápida amplificación y generación de variabilidad entre sus secuencias (Gabaldón *et al.*, 2013). En primer lugar, la mayoría de los genes EPA se encuentran ubicados en regiones subteloméricas. En *S. cerevisiae*, se ha observado que estas regiones forman conglomerados cerca de la periferia nuclear (Maillet *et al.*, 1996), lo que podría facilitar la recombinación homóloga no alélica. De hecho, las regiones subteloméricas son

propensas a sufrir rearrreglos y duplicaciones génicas (Lafontaine *et al.*, 2004). Además, los genes EPA contienen repeticiones internas en tándem de serina/treonina que actúan como un espaciador entre el sitio de anclaje de la pared celular y el medio ambiente. Estas repeticiones, presentes en un número variable, se cree que afectan la unión al ligando, ya que su longitud determina la capacidad de llevar el dominio N-terminal hasta la superficie de la pared celular (Miramón *et al.*, 2016). Además, al recombinar las regiones repetidas de diferentes genes EPA, se generarían extremos N-terminales distintos, lo que también alteraría la especificidad. Estos factores podrían haber facilitado una rápida adaptación de las propiedades de adherencia en *C. glabrata* (Kaur *et al.*, 2005; Cormack *et al.*, 1999).

IV.2. Incremento en la capacidad de la formación de biopelículas.

Como resultado del aumento en la expresión de los genes de adhesinas de la pared celular (EPA), los organismos patógenos del género *Nakaseomyces* han aumentado su capacidad para sintetizar biopelículas y con ello, la posibilidad de generar biofilm. Estas biopelículas, pueden aumentar la resistencia a los antifúngicos y al sistema inmunológico del hospedero, lo que facilita su persistencia en el organismo hospedero. Además, las biopelículas formadas en otras superficies, como catéteres u otros materiales, pueden ser una fuente de infecciones hospitalarias. Un fenotipo de mayor adherencia de una especie en particular suele correlacionarse con una mayor virulencia y con una expansión subyacente de genes relacionados con la adhesión (Butler *et al.*, 2009; Gabaldón *et al.*, 2013).

IV.3. Nivel de resistencia a varios tipos de estrés

El éxito de *C. glabrata* como patógeno está estrechamente relacionado con su capacidad para resistir estos diversos estreses y adaptarse a los cambios ambientales dentro de su hospedero durante el proceso de infección. Estos pueden incluir:

- Estrés osmótico, que se origina por cambios en la concentración de solutos y la presión osmótica durante el proceso de colonización. *C. glabrata* tiene la capacidad de regular la concentración de solutos dentro de la célula para contrarrestar el estrés osmótico, como, por ejemplo, mediante la acumulación o liberación de ciertos solutos, como el glicerol o trehalosa. Esta adaptación permite mantener el equilibrio osmótico y proteger la célula contra la deshidratación o la ruptura celular.
- Estrés térmico, ya que la temperatura corporal puede variar dependiendo de la ubicación de la infección. Por ejemplo, la temperatura en la piel puede ser ligeramente más baja que la temperatura interna del cuerpo, mientras que en el tracto vaginal la temperatura puede ser más estable. Además, durante la fiebre, que es una respuesta del hospedero ante la infección, la temperatura corporal puede elevarse significativamente. Para contender al estrés térmico, *C. glabrata* expresa proteínas de choque térmico como las chaperonas *Heat-shock protein 70* y *90* (Hsp70 y Hsp90, respectivamente), que ayudan a proteger otras proteínas contra el daño inducido por el calor y facilitan el correcto plegamiento de las proteínas en condiciones de estrés térmico.
- Estrés ácido, que se genera en entornos ácidos, como el estómago o el tracto vaginal, en donde *C. glabrata* necesita resistir la exposición a pH bajos para establecer y mantener la infección. Al igual que en el caso del estrés por temperatura, la expresión de las chaperonas Hsp70 y Hsp90, permiten contender a *C. glabrata* al estrés inducido por entornos ácidos. Adicionalmente, *C. glabrata* expresa proteínas de transporte de protones que le permiten ajustar el pH intracelular.
- Estrés oxidativo, que es generado por la exposición de *C. glabrata* a especies reactivas de oxígeno (ROS, por sus siglas en inglés), que son generadas por las células del hospedero. Dichas especies reactivas de oxígeno incluyen a los radicales superóxidos, peróxido de hidrógeno, o el radical hidroxilo. Estas especies reactivas son altamente tóxicas para las células fúngicas y pueden dañar lípidos, proteínas y ácidos nucleicos. Para contrarrestar el estrés oxidativo, *C. glabrata* emplea varios mecanismos de defensa antioxidante, que incluyen enzimas, como la superóxido dismutasa, la catalasa y la peroxirredoxina. Adicionalmente, *C. glabrata* también contiene al estrés oxidativo produciendo glutatión, que es un antioxidante celular que participa en la neutralización de los ROS y en la regeneración de otras moléculas antioxidantes. Finalmente, *C.*

glabrata también combate al estrés oxidativo mediante la síntesis de bombas de eflujo, que le permite eliminar las especies reactivas de oxígeno de su interior, ayudando a mantener el equilibrio celular redox. (Li et al., 2007)

- Estrés nutricional, que se origina por las variaciones de la disponibilidad de nutrientes puede variar dentro del cuerpo del hospedero en donde *C. glabrata* requiere de un proceso de adaptación para utilizar nutrientes alternativos y sobrevivir a largos períodos de inanición durante su ingestión por macrófagos. (Li et al., 2007) (Mason et al., 2012)

IV.4. Sobreexpresión de enzimas con actividad hidrolítica

Durante el proceso de infección, *C. glabrata* sintetiza diferentes tipos de enzimas hidrolíticas que desempeñan un papel crucial en su patogenicidad, que incluyen proteasas, fosfolipasas y lipasas (Figura 6). La actividad enzimática de las proteasas de *C. glabrata* le permiten degradar componentes de la matriz extracelular y proteínas del sistema inmunológico del huésped, lo que facilita la invasión y la evasión de la respuesta inmunitaria, mientras que la expresión de las fosfolipasas permiten la hidrólisis de los enlaces éster de los fosfolípidos, lo que conduce a la lisis de las membranas celulares del huésped, contribuyendo a la diseminación del hongo en los tejidos hospedantes y al daño tisular. Finalmente, las lipasas expresadas por *C. glabrata*, durante el proceso de infección, le permiten a este organismo degradar los lípidos, que son utilizados como nutrientes y adicionalmente tienen un papel relevante en la modulación de la respuesta inmunitaria del huésped (Figura 6) (Frías de León et al., 2021)

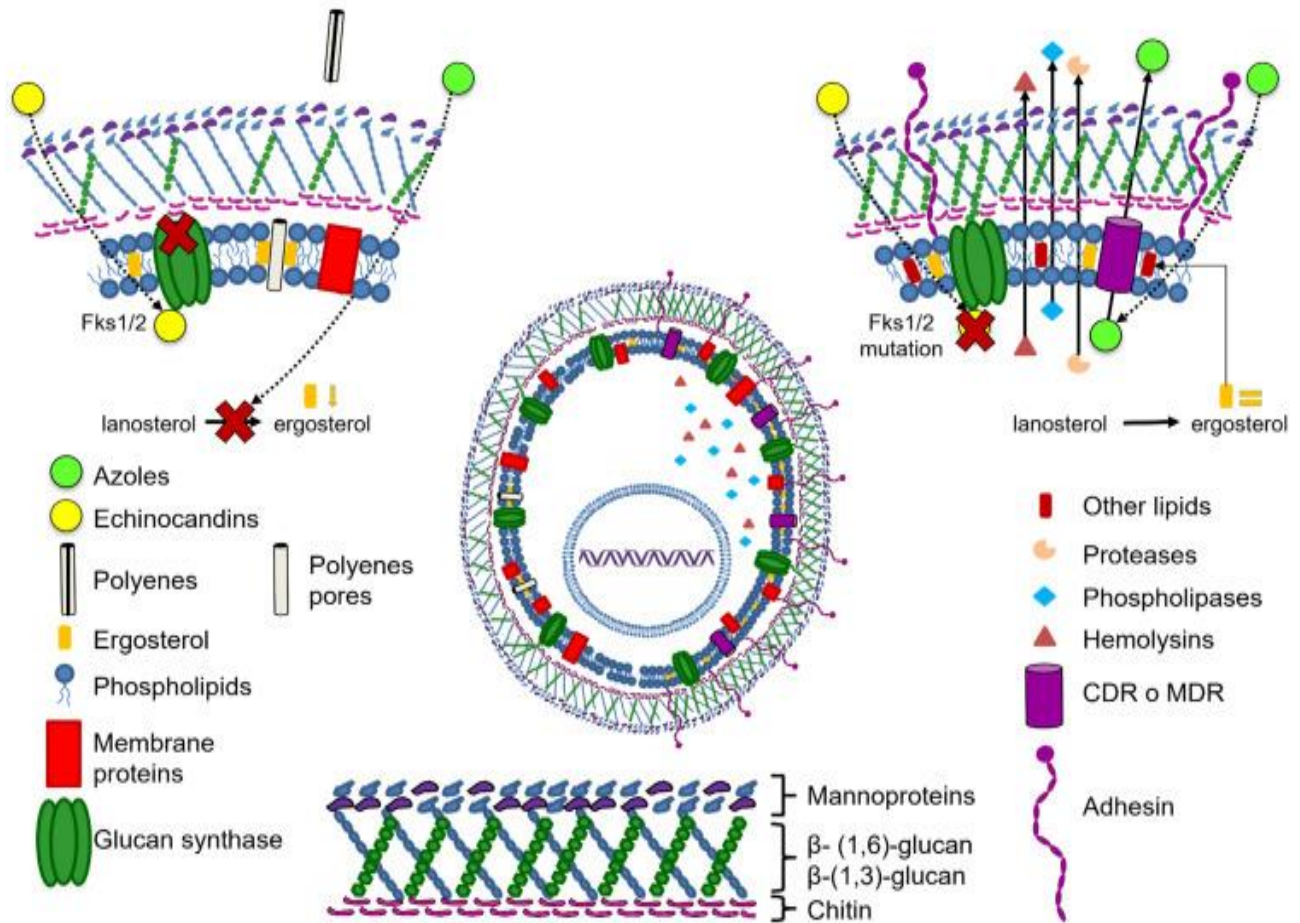


Figura 6. Factores relacionados con la patogenicidad de *C. glabrata*. La formación de biofilm y la producción de enzimas hidrolíticas como proteasas, fosfolipasas y hemolisinas, contribuyen a la adherencia, daño celular e invasión tisular en el hospedero. Tales factores de virulencia les otorgan la capacidad de evadir la respuesta inmune del hospedero, generando resistencia (Tomada de Frías de León *et al.*, 2021).

IV.5. Reducción del genoma

En términos generales, la reducción del genoma es una consecuencia general de los estilos de vida comensales y patógenos debido a que el hospedero mantiene de manera mucho más estable la temperatura, el pH, la presión y la disponibilidad de nutrientes, por lo que los organismos patógenos podrían prescindir de las funciones de diversos conjuntos de genes, como los involucrados en la biosíntesis de componentes esenciales, que pueden ser tomados del entorno (Andersson *et al.*, 1998 ; Ehrlich, *et al.*, 2008).

En los hongos, los patógenos microsporidios representan un caso extremo de reducción del genoma debido a un estilo de vida patógeno especializado, con algunos de sus genomas codificando menos de 2000 proteínas (Corradi *et al.*, 2011). La reducción del genoma de *C. glabrata*, que no es un patógeno especializado, puede parecer comparativamente escasa. Su repertorio de genes codificantes de proteínas 5202 es “sólo” 688 genes menos que el de *S. cerevisiae*, lo que representa una reducción de tamaño relativo del 11,7%. Sin embargo, la naturaleza de algunas de las pérdidas apuntaba a una posible mayor dependencia del hospedador para superar la auxotrofia de algunas vitaminas como la piridoxina, la tiamina y el ácido nicotínico (Kaur *et al.* 2005) (Kumar *et al.* 2019). En particular, se considera de gran relevancia la pérdida de toda la vía para sintetizar el ácido nicotínico, el cual es un precursor del NAD⁺, y sin la capacidad de sintetizarse *de novo*, *C. glabrata* depende de una fuente externa, en este caso, el hospedero mamífero. En este sentido, es importante mencionar que el agotamiento de los precursores de NAD⁺ en algunos tejidos humanos, como, por ejemplo, en el tracto urinario, tiene un efecto positivo en la regulación transcripcional de algunas adhesinas *EPA* y, por lo tanto, en la adhesión. Se ha propuesto que esta regulación positiva es el producto de una adaptación de *C. glabrata* hacia su hospedero humano para favorecer el proceso de infección (Domergue *et al.*, 2005).

Adicionalmente, una de las diferencias más notables al comparar el genoma de *S. cerevisiae* con los organismos del género *Nakaseomyces* es la reducción significativa en el número de copias parálogas del gen PAU. Este gen codifica para manoproteínas que desempeñan un papel crucial en la respuesta a diversos tipos de estrés (Luo and van Vuuren., 2009). Mientras que en *S. cerevisiae* el número de copias parálogas de este gen alcanza los 20, en los organismos del género *Nakaseomyces* suele variar entre 1 y 2 (ver Figura 7).

Otras pérdidas evidentes en la mayoría de los organismos del género *Nakaseomyces* incluyen genes del Clúster DAL, los cuales regulan la biosíntesis del ácido aspártico, un aminoácido esencial para la síntesis de proteínas y para la vía de asimilación de nitrógeno. Asimismo, se observa la ausencia de familias de genes que codifican para las fosfatasa

ácidas PHO, así como los genes SNZ y SNO, cuya función está poco caracterizada en *S. cerevisiae*, pero se sabe que su expresión se induce en la fase estacionaria (ver Figura 7).

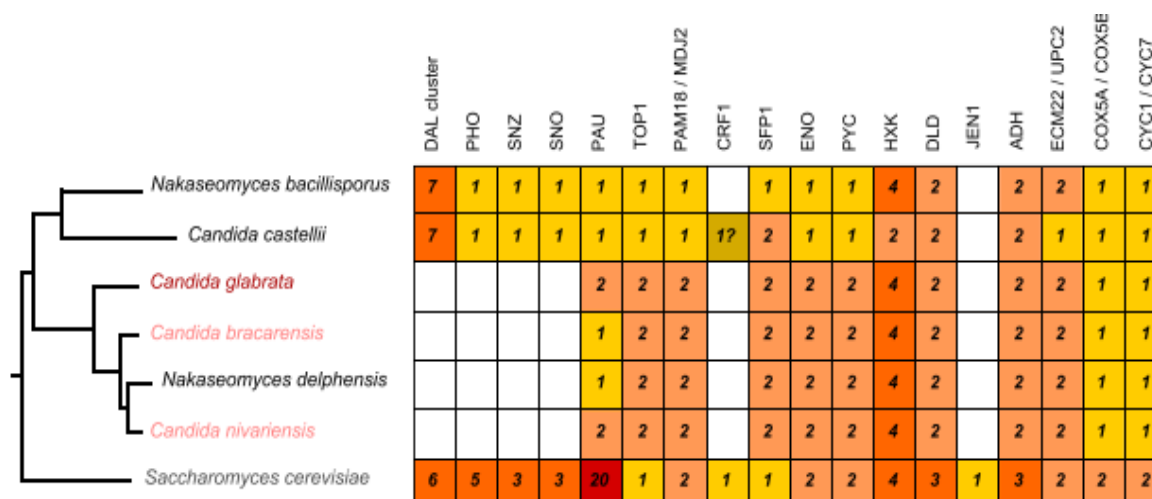


Figura 7. Representación esquemática de familias genéticas específicas en el clado *Nakaseomyces*. Se muestra la relación evolutiva de las especies con las familias de los genes en dicho clado. Los recuadros blancos indican la ausencia de una familia o ruta particular, mientras que los números de color (amarillo, naranja y rojo) indican las copias parálogas de esa familia de genes, la intensidad del color indica el aumento de parálogos presentes de cada familia. Los acrónimos y las funciones de las familias de genes se muestran en la Tabla 1 (Tomada de Gabaldón *et al.*, 2013).

Tabla 1. Acrónimos y funciones de las familias de genes de la Figura 7

DAL	PHO	SNZ	SNO	PAU	TOP1	PAM 18
Adhesina Leucocitaria	Regulador de Fosfato	Represor de gen FT	Resistencia a azoles	Codificación a manoproteínas	Topoisomerasa tipo1	Codificación a translocasa de membrana interna en mitocondria
ADH	ECM22	UPC2	CRF1	SFP1	ENO	PYC
Enzima Alcohol deshidrogenasa	Filamentación en levaduras	Síntesis de ergosterol	Liberador de Corticotropina tipo 1	Factor de transcripción para el crecimiento	Factor de transcripción que codifica para la familia Apetala2	Gen de Piruvato carboxilasa
HXK	DLD	JEN1	COX5A	MDJ2	CYC1	
Enzima glucolítica	Enzima Dihidrolipoamida deshidrogenasa	Gen de Resistencia a Fluconazol	Gen Ciclooxygenasa tipo 5	Gen involucrado en la respiración mitocondrial	Citocromo C	

V. Mutaciones de *Candida glabrata* que favorecen su resistencia a antifúngicos

Candida glabrata es considerada un patógeno relevante en entornos clínicos, ya que puede presentar una menor susceptibilidad intrínseca a los antifúngicos de la clase azólicos y desarrollar una mayor resistencia a estos antifúngicos cuando invade a los pacientes. Se sabe que los tratamientos con azoles conducen frecuentemente a una resistencia a estos antifúngicos mediante la sobreexpresión de bombas de flujo de xenobióticos. Son varios los factores que se han identificado en *C. glabrata* que puede incluir al incremento de resistencia a antifúngicos. A continuación, se mencionan los más relevantes. (Rodrigues *et al.*, 2014)

Algunas mutaciones específicas de ganancia de función en el regulador transcripcional Pdr1 dan como resultado un incremento significativo en la síntesis de los transportadores de bombas de flujo de xenobióticos, contribuyendo al incremento de la virulencia en los modelos de animales. La inactivación de este regulador, mediante experimentos de mutagénesis por transposones mostró que la disrupción del gen PDR1 origina un incremento de 8 a 16 veces en la susceptibilidad al fluconazol de *C. glabrata*. (Tsai *et al.*, 2006). En este mismo sentido, la caracterización de aislados clínicamente resistentes al fluconazol, ha revelado la presencia de mutaciones puntuales en este regulador transcripcional (Tsai *et al.*, 2006).

Adicionalmente, la comparación de las secuencias genómicas de aislados clínicos de *C. glabrata*, han mostrado que ciertos rearrreglos cromosómicos pueden generar resistencia a los azoles (Dujon *et al.*, 2013). Se ha reportado que el gen de reparación de errores de apareamiento de bases nucleotídicas *MSH2*, puede favorecer la resistencia a los fármacos cuando su actividad se ve comprometida por cambios de un solo nucleótido específicos (SNP, por sus siglas en inglés), que dan lugar a un fenotipo hipermutador (Hagman *et al.*, 2013). En este mismo sentido, los aislados clínicos que provienen de un paciente, antes y después del tratamiento con el antifúngico fluconazol, portan una mutación no sinónima V239L en el gen *MSH2*, que favorece el fenotipo hipermutador (Healey *et al.*, 2016). Cabría esperar que este fenotipo pudiera

generar diversidad genética dentro del contexto de las condiciones del hospedador. En este mismo sentido, un estudio que analizó la acumulación de mutaciones debidas a defectos de *MSH2* en *S. cerevisiae* durante 170 generaciones *in vitro*, reveló que las tasas de mutación estuvieron en el rango de $5 \text{ a } 9 \times 10^{-8}$, que correspondieron a una frecuencia de 8 a 16 mutaciones puntuales y de 110 a 180 deleciones/inserciones en el genoma haploide de *S. cerevisiae* (Marcet-Houben *et al.*, 2013).

La literatura actual sobre el efecto de defectos en los sistemas de reparación de DNA demostró que la ocurrencia de mutaciones no es aleatoria: regiones de DNA con una densidad mayor de repeticiones son más mutables en células con defectos en los sistemas de reparación del DNA (Pryszcz *et al.*, 2011). En el contexto de las adhesinas, que están enriquecidas en elementos repetitivos, las mutaciones de *MSH2* quizás contribuyan a incrementar la diversidad de esta familia de genes a lo largo de generaciones.

VI. Aspectos de diferenciación y similitud entre *Candida glabrata* con *Candida albicans*

Dada su relevancia clínica, *C. glabrata* y *C. albicans* ocupan una posición destacada entre los organismos de la clase Saccharomycetes, lo que ha impulsado la investigación comparativa de sus características. Ambas especies pertenecen a grupos taxonómicos compuestos principalmente por organismos comensales no patógenos para los seres humanos, ubicándose en el filo Ascomycota, clase Saccharomycetes y orden Saccharomycetales. Sin embargo, es importante señalar que pertenecen a familias taxonómicas diferentes. Mientras que *C. glabrata* forma parte de la familia Saccharomycetaceae, *C. albicans* es miembro del clado CUG-Ser1 de la familia Debaryomycetaceae.

Otra diferencia importante entre *C. glabrata* y *C. albicans* está relacionada con su ploidía, o número de conjuntos completos de cromosomas en una célula. Haploide, en el caso de *C. glabrata* y diploides en el caso de *C. albicans*. Estas diferencias en la ploidía pueden generar a su vez, diferencias notables en su tamaño, forma y disposición de estructuras reproductivas. Adicionalmente, el crecimiento de *C. glabrata*, como células

haploides, no pueden formar hifas verdaderas, en comparación con las células diploides de *C. albicans* que tienen la capacidad de desarrollar hifas. De hecho, el cambio a crecimiento de hifas es un mecanismo de virulencia conocido en *C. albicans*, que le permite ser más invasivo y escapar de los macrófagos una vez que estos la han fagocitado (Mayer *et al.*, 2013). Por el contrario, *C. glabrata* se enfrenta a la fagocitosis de macrófagos de una manera totalmente diferente. En particular, se ha demostrado que *C. glabrata* puede persistir dentro de los macrófagos durante largos períodos de tiempo e incluso dividirse dentro de ellos (Kaur *et al.*, 2009, Roetzer *et al.*, 2008; Rai *et al.*, 2012). Aunque tanto *C. albicans*, como *C. glabrata*, y muchos otros hongos patógenos humanos, se han descrito como especies asexuales, también difieren en este sentido. Se ha descrito un ciclo parasexual en *C. albicans* mediante el cual, células de diferente tipo de apareamiento se fusionan para formar un tetraploide transitorio que luego revierte a una forma diploide por la pérdida concertada de cromosomas (Bennett *et al.*, 2014). Dicho ciclo no se ha descrito en *C. glabrata*, donde los datos genéticos sugieren que el verdadero apareamiento sexual podría ocurrir, aunque con una frecuencia muy baja (Dodgson *et al.*, 2005).

Según la base de datos de ortología MetaPhOrs (Gabaldón *et al.*, 2013), 1557 (29,5%) genes que codifican para proteínas en *C. glabrata* carecen de un ortólogo en *C. albicans*, y lo contrario es cierto para 2257 (36,3%) de los genes de *C. albicans*. Por tanto, se esperan grandes diferencias fisiológicas simplemente debido a la composición genética intrínsecamente diferente de ambas especies. Además, incluso si los ortólogos correspondientes están presentes, la función puede haber divergido y así afectar su papel en la virulencia. Esto puede ser especialmente cierto para proteínas versátiles como factores de transcripción y otros reguladores transcripcionales que pueden reconectar fácilmente su red de genes blanco. Un ejemplo de este caso lo proporciona el hallazgo de que las mutantes nulas en el gen que codifica para el factor de transcripción ACE2 tienen una virulencia ligeramente atenuada en *C. albicans*, mientras que son hipervirulentas en *C. glabrata* (MacCallum *et al.*, 2009). Todas estas consideraciones subrayan el hecho de que *C. albicans* y *C. glabrata* han desarrollado mecanismos de virulencia muy diferentes (Brunke *et al.*, 2013).

Finalmente, las estrategias empleadas durante la infección de estas dos especies de *Candida* son claramente diferentes. Mientras que *C. albicans* tiende a emplear una estrategia más agresiva para desencadenar la respuesta inmunitaria del hospedero y obtener nutrientes para su supervivencia, *C. glabrata* muestra un comportamiento basado en el sigilo y la evasión inmune y procurando su supervivencia sin causar daños severos al hospedero en un principio. Esto lo logra mediante mecanismos complejos de evasión inmune (Brunke *et al.*, 2012) (Diekema *et al.*, 2012), (Kumar *et al.*, 2019). Lo anterior sugiere que *C. albicans* y *C. glabrata* pudieron haber seguido rutas evolutivas independientes hacia la patogénesis. Identificar los mecanismos de virulencia y los componentes genómicos de estas rutas, así como comprender las variaciones en la susceptibilidad a diferentes fármacos entre las especies de *Candida*, puede facilitar el camino para seleccionar los tipos de tratamientos más efectivos para combatirlos. Por ejemplo, los equinocandinos, como la caspofungina, la micafungina y la anidulafungina, resultan más efectivos en el tratamiento de infecciones provocadas por *C. glabrata*; mientras que los azoles, como el fluconazol, itraconazol y voriconazol, se emplean con eficacia contra infecciones causadas por *C. albicans*. Es relevante señalar que uno de los factores que contribuye al aumento en el desarrollo de resistencia a los fármacos en ciertas especies es el uso inadecuado de los antifúngicos, lo que resulta en una mayor probabilidad de desarrollo de resistencia, como se ha observado en ciertas cepas de *Candida* (Diekema *et al.*, 2012).

VII. Diagnóstico molecular. La importancia de contar con un método rápido, sensible y específico

Todos los organismos de una misma especie presentan variabilidad en sus secuencias genómicas de forma inherente que se originan por cambios o mutaciones espontáneas en su material genético, a lo que se le conoce como «polimorfismo». Estas variaciones pueden presentarse como polimorfismos de un solo nucleótido (SNP, por sus siglas en inglés), inserciones o deleciones de segmentos de DNA, o bien, por alteraciones en la repetición de secuencias nucleotídicas. Estos polimorfismos pueden tener un impacto en la apariencia, función o susceptibilidad a enfermedades de un organismo y en términos

del diagnóstico molecular, pueden ser utilizados como marcadores genéticos (Figura 8) (Dodgson *et al.*, 1997).

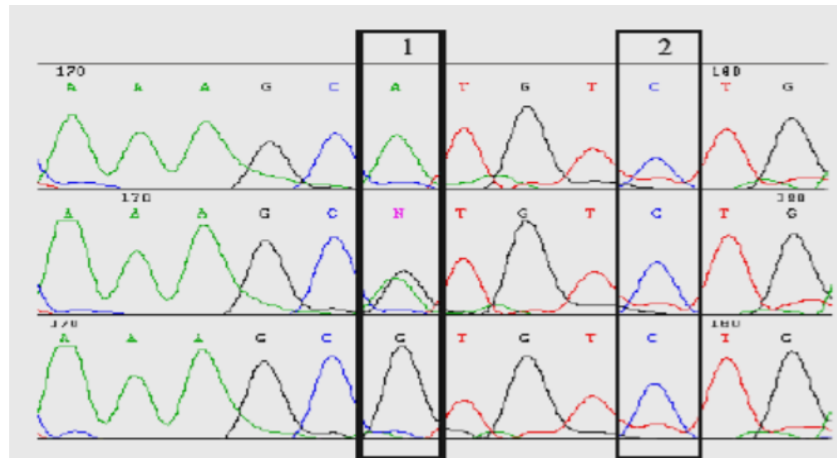


Figura 8. Detección de un Polimorfismo de un solo nucleótido. El alineamiento de fragmentos a partir de secuenciación directa de productos de PCR permite la identificación de regiones en donde existe (1), o no (2), variabilidad en las bases en una posición específica de las secuencias, o SNPs (Tomada de Vignal *et al.*, 2002).

Otro tipo de variaciones en las secuencias de DNA utilizadas en la tipificación de cepas son los Microsatélites o Secuencias Repetidas Cortas (o SRS, por sus siglas en inglés). Los microsatélites son segmentos cortos de DNA de 1 a 6 pares de bases (pb) que se repiten en tándem y de forma aleatoria en el genoma de los seres vivos. Algunas de las características de estos marcadores *versus* otros (como son los minisatélites, RFLPs, RAPDs, etc.) que son considerados por la mayoría de los autores como una poderosa herramienta para estudios filogenéticos son: 1) Presentan un elevado grado de polimorfismo, 2) Su herencia es mendeliana simple, 3) Son codominantes (por lo tanto se pueden diferenciar de los individuos homocigotos de los heterocigotos), 4) Son fáciles de medir y analizar, 5) Tienen una confiabilidad del 100%, y 6) Son repetitivos y automatizables. Estos marcadores se encuentran generalmente en regiones no codificantes del genoma (Goldstein *et al.*, 1999).

Para realizar los estudios de identificación antes mencionados, se utiliza la técnica de la reacción en cadena de polimerasa (PCR), la cual es una reacción enzimática

catalizada por una DNA polimerasa que permite la amplificación o reproducción «in vitro» de una región particular del DNA, a partir de una pequeña cantidad original de DNA molde. Esta enzima requiere, para la síntesis, de un par de oligonucleótidos denominados iniciadores o «primers», cuyas secuencias son complementarias a la de las regiones flanqueantes 5' y 3' del segmento particular de DNA que se aspira a amplificar. Un buen marcador molecular debe reunir una serie de características para maximizar su utilidad, entre ellas, debe tener una buena distribución a lo largo del genoma y alto grado de polimorfismo. La técnica para analizar el marcador debe ser rápida, práctica y debe mantener una gran repetibilidad (Cheng y Crittenden *et al.*, 1994). Los minisatélites y microsatélites, se encuentran entre las regiones que presentan múltiples alelos y sus valores de heterocigosis superan el 70%, por lo que son excelentes marcadores.

Entre los marcadores moleculares efectivos se destacan las "Repeticiones en Tándem de Número Variable" (VNTR, por sus siglas en inglés), los cuales son marcadores polimórficos detectados en ciertas pruebas de hibridación para secuencias repetitivas (Jeffreys *et al.*, 1985). Estas pruebas generan un patrón complejo de bandas que reflejan un polimorfismo heredable. Los VNTR consisten en repeticiones al azar en tándem de 10 a 60 pares de bases, son altamente polimórficos y presentan elevadas tasas de heterocigosis en las poblaciones (Vance *et al.*, 1998).

El Polimorfismo de Longitud de Fragmentos de Restricción (RFLP) constituye una de las primeras técnicas empleadas antes de la popularización de la técnica de PCR. Se basa en la visualización de las variaciones a nivel de la secuencia del DNA, mediante el uso de enzimas de restricción (Botstein *et al.*, 1980), cuya actividad puede variar entre individuos. Posteriormente, con el avance del PCR, surgió un método denominado RFLP-PCR, el cual se ha revelado útil en el diagnóstico de mutaciones puntuales, así como de pequeñas deleciones o inserciones que generan o eliminan un sitio de restricción. Al cortar el DNA con enzimas de restricción, se analizan los fragmentos de ADN de distinto peso molecular o de longitudes diferentes (Figura 9).

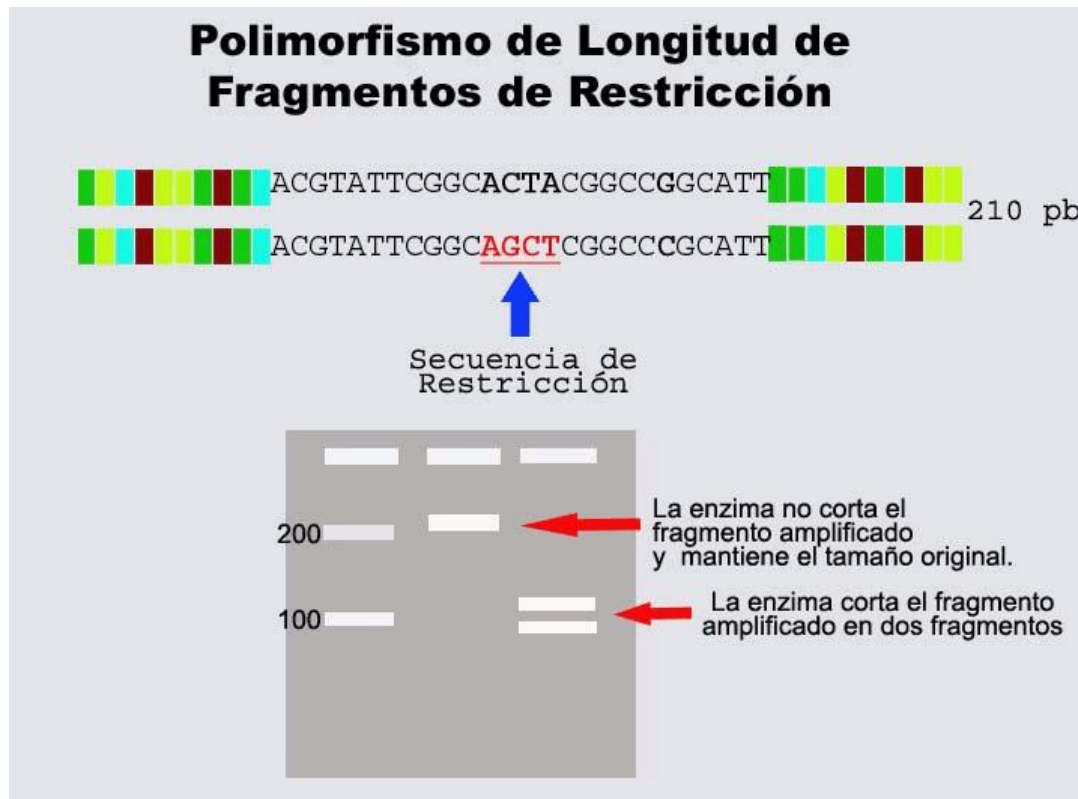


Figura 9. Polimorfismo de Longitud de Fragmentos de Restricción (PCR-RFLP). La enzima reconoce el sitio AGCT y corta la cadena de DNA en 2 fragmentos. (Tomada de Villasmil *et al.*, 2007).

Los polimorfismos y las regiones satélites pueden ser mapeados utilizando oligonucleótidos cortos, conocidos como iniciadores, a través de una reacción de PCR que se caracteriza por tener temperaturas de alineamiento (annealing) bajas. Esto genera una serie de fragmentos de amplio espectro a partir del DNA molde. La clave radica en identificar aquellos fragmentos amplificados que resulten ser polimórficos, los cuales pueden ser posteriormente mapeados. Esta técnica se conoce como Polimorfismo Amplificado Aleatoriamente (RAPD) (Levin *et al.*, 1994).

Dado que los microsatélites se distribuyen parcialmente a lo largo de todo el genoma de los eucariotas, aunque con baja frecuencia en las regiones codificantes y posiblemente también en los telómeros, su presencia en regiones codificantes se ha asociado con enfermedades (Armour *et al.*, 1994; Hancock *et al.*, 1999; Tautz *et al.*, 1994). Sin embargo, aún no se comprende completamente el significado funcional de estas secuencias, aunque

la hipótesis más aceptada sugiere que pueden estar relacionadas con el empaquetamiento y la condensación del DNA en los cromosomas (Vanhala *et al.*, 1998).

Desde su introducción en 1989 (Litt *et al.*, 1989; Weber *et al.*, 1989), el uso de microsatélites ha tenido un impacto significativo en el estudio de la genética de animales y plantas. El análisis con microsatélites implica la detección de fragmentos específicos de DNA y proporciona la medida de los alelos, medidos en pares de bases, en cada una de las regiones. Otra aplicación importante de los microsatélites es la construcción de mapas de ligamiento más completos y detallados, así como la identificación de genes de interés (QTLs) (Tabla 2). Todos los marcadores pueden utilizarse para construir mapas de ligamiento; sin embargo, es necesario que los alelos se segreguen de manera independiente y que además puedan ser monitoreados. La descendencia puede ser informativa si los progenitores son dobles heterocigotos en los loci analizados. Los loci ubicados en cromosomas distintos pueden recombinarse libremente durante la gametogénesis parental hasta en un 50% (segregación independiente); mientras que, si se encuentran en el mismo cromosoma, recombinarían con una frecuencia que oscila entre 0 y 50%, dependiendo de la distancia que los separa. Por lo tanto, un mapa genético bien poblado de marcadores se convierte en una herramienta muy útil para identificar genes responsables de caracteres de interés. Se trata de buscar asociaciones entre varios alelos en cualquiera de los marcadores, segregando en poblaciones que presenten el carácter de interés, para identificar regiones del genoma donde es más probable que se encuentre el gen responsable de ese carácter (Cheng *et al.*, 1995).

Gene	Percentage homology	Normalized ratio*	GenBank accession no.	Reference
<i>ACT1</i>	98·8	> 1·5	AJ236897	Donnelly <i>et al.</i> (1999)
<i>URA3</i>	93·0	> 1·5	AJ302032	Staib <i>et al.</i> (2001)
<i>MDR1</i>	92·0	1·4	AJ227752	Moran <i>et al.</i> (1998)
<i>CDR1</i>	91·9	1·16	AJ439073	Moran <i>et al.</i> (2002)
<i>CDR2</i>	91·0	> 1·5	AJ439075	Moran <i>et al.</i> (2002)
<i>ERG3</i>	90·1	1·17	AJ421248	Pinjon <i>et al.</i> (2003)
<i>ERG11</i>	90·4	> 1·5	AY034876	Perea <i>et al.</i> (2002)
<i>PHR2</i>	90·5	1·40	AF184908	Kurzai <i>et al.</i> (1999)
<i>PHR1</i>	88·0	> 1·5	AF184907	Kurzai <i>et al.</i> (1999)
<i>SAP4</i>	86·0	0·55	AJ634382	This study
<i>SAP2</i>	83·6	1·18	AJ634672	This study
<i>CZF1</i>	80·0	0·53	AJ634475	This study
<i>IPF8147</i>	79·6	0·29	AJ634476	This study
<i>IPF1873</i>	75·0	0·24	AJ634664	This study
<i>RVS161</i>	75·0	0·30	AJ634665	This study
<i>IPF16104</i>	74·0	0·25	AJ634666	This study
<i>IFA8</i>	70·4	0·36	AJ634667	This study
<i>IPF9787</i>	70·0	0·26	AJ634668	This study
<i>IPF3448</i>	69·6	0·42	AJ634669	This study
<i>SSN6</i>	68·0	0·30	AJ634670	This study
<i>IPF2057</i>	59·0	0·17	AJ634671	This study

Tabla 2. Porcentaje de similitud de secuencia de nucleótidos de genes de *C. dubliniensis* a las secuencias correspondientes de microarreglos de *C. albicans*. Un ejemplo de genómica comparativa. (Tomada de Moran *et al.*, 2004).

Los análisis con marcadores genéticos tipo microsatélite son extremadamente valiosos para el estudio de poblaciones. Además de evaluar la variabilidad genética, la información obtenida permite una caracterización molecular precisa de las poblaciones y/o razas, así como un estudio detallado de su estructura genética. Esta revisión también confirma la eficacia de los microsatélites para establecer relaciones genéticas entre poblaciones (Vale-Silva *et al.*, 2017).

Cuando se encuentran fuera de regiones codificantes, los microsatélites suelen considerarse como regiones génicas sin función aparente. Sin embargo, se ha sugerido que, en ciertos genes, las regiones repetidas del extremo 5' no transcrito podrían tener una función reguladora, actuando como sitios de unión para factores de transcripción. Debido a su alta repetitividad, los microsatélites son conocidos como regiones con una alta frecuencia de recombinación (hot spots) (Oliveira *et al.*, 2006). Las Figuras 10 y 11 muestran algunos tipos de marcadores de DNA y el principal uso que se les ha dado para el proceso de identificación de *C. glabrata*.

Tipo de Marcador	Acrónimo	Alias	Requiere:	Principal uso
Polimorfismo de base única	SNP's		ADN clonado, secuencias, etc.	Mapas de Ligamiento
Polimorfismo de Longitud de Fragmentos de Restricción	RFLP		ADN clonado usualmente	Mapas de Ligamiento
Secuencias Repetidas Cortas	SSR	Microsatélites	AND clonado y secuencias	Mapas de Ligamiento
Secuencias de Sitios Marcados	STS		Secuencias de ADN	Mapas físicos
Secuencias Expresadas Marcadas	EST	Subset de STS	Secuencias de ADNc	Mapeos físicos y de ligamiento
Polimorfismo Amplificado Aleatoriamente	RAPD	DAF, AP-PCR	Oligos aleatorios	Fingerprinting
Secuencias de Numero y Tamaño Variable	VNTR	Minisatélites	Secuencias repetitivas y pruebas de hibridación	Fingerprinting
Polimorfismo de Longitud de Fragmentos Amplificados	AFLP		Set de oligos diseñados y específicos	Mapas de ligamiento

Figura 10. Principales tipos de marcadores de DNA y su principal uso en la identificación de *C. glabrata* (Tomada de Dogson *et al.*, 1997).

Markers	Primers	Primer sequences
<i>RPM2</i>	Forward	FAM 5' ATCTCCCAACTTCTCGTAGCC
	Reverse	ACTTGAACGACTTGAACGCC
<i>MTI</i>	Forward	HEX 5' CAGCAATAATAGCTTCTGACTATGAC
	Reverse	GACAGGAGCAACCGTTAGGA
<i>ERG3</i>	Forward	TAMRA 5' AGTGCGAGTGTATGTAAAGAATG
	Reverse	CGTATACCTTATCTCCGTTCAA
<i>GLM4</i>	Forward	FAM 5' AGTGTTTCATTGTCGCCTTC
	Reverse	AATGCAGGCTCACCATTTTC
<i>GLM5</i>	Forward	HEX 5' TGGGGATAGTGGGAACTCAA
	Reverse	CGATGATTTCATGTCCGATG
<i>GLM6</i>	Forward	TAMRA 5' GATGATTCTGCCC GTTAGGA
	Reverse	CCTGAAGTAGGTGCCGAGAG

Figura 11. Principales tipos de marcadores de DNA y oligonucleótidos en *C. glabrata*.

La definición funcional de los genes blanco usadas en el diagnóstico molecular corresponden a: RPM2: RNAasa P mitocondrial; MT1: Metalotioneína; ERG3: 5,6-esterol desaturasa; GLM4, GLM5 y GLM6: Secuencias de marcadores en microsatélites (Tomado y modificado de Kiasat *et al.*, 2019).

JUSTIFICACIÓN

Candida glabrata es la segunda causa de candidemia en muchos países, siendo una de las especies de levadura nosocomiales más preocupantes debido a su tendencia a desarrollar resistencia a los fármacos antifúngicos, incluidos los aislados multirresistentes emergentes. Alrededor del 90% de los casos de candidemia son atribuibles a solo cuatro especies de *Candida*, cada una con distintas susceptibilidades a los antimicóticos. En *C. glabrata*, diversos factores como la regulación de la expresión de múltiples adhesinas, el agrupamiento de los telómeros y subtelómeros cerca de la envoltura nuclear que favorece la recombinación homóloga no alélica entre familias de genes de adhesinas, y la presencia de arreglos genómicos de gran escala, contribuyen a su capacidad de adhesión y persistencia en el hospedero. Este problema es de gran importancia, ya que las infecciones asociadas con *C. glabrata*, especialmente en el torrente sanguíneo, tienen una alta mortalidad y conllevan un considerable costo en términos de salud pública. Actualmente, la identificación de la especie de *Candida* en pacientes infectados, mediante cultivos microbiológicos, o pruebas bioquímicas, suele tomar tiempos muy largos, de hasta 7 días, sin que ellos brinden una suficiente especificidad. Dado lo anterior, es necesario el desarrollo métodos de detección más eficaces, específicos y rápidos. La metodología de PCR ... Luego mencionas tu metodología. de herramientas computacionales que permitan el diseño óptimo de oligonucleótidos a ser usados en la amplificación de secuencias repetidas en el genoma de *C. glabrata* mediante la técnica de PCR. Esperamos que el desarrollo de dichas herramientas conduzca al desarrollo de un ensayo molecular rápido, específico y sensible que impacte de manera significativa en el diagnóstico preciso de enfermedades intrahospitalarias, favoreciendo un tratamiento adecuado y oportuno.

HIPÓTESIS

La especificidad y sensibilidad de los ensayos de identificación de microorganismos patógenos en muestras clínicas mediante ensayos de amplificación de PCR, será incrementada significativamente al considerar oligonucleótidos que reconozcan loci

repetidos en los genomas de estudio. Esto ya está demostrado. Tú quieres desarrollar un método que ayude a identificar sitios adecuados.

OBJETIVO

Desarrollar un protocolo bioinformático para el diseño óptimo de oligonucleótidos a ser utilizados en ensayos de detección molecular por PCR de *Candida glabrata*, considerando que la mayor especificidad y sensibilidad de los ensayos serán obtenidos mediante la amplificación de secuencias altamente repetitivas y específicas.

OBJETIVOS PARTICULARES

Desarrollar un protocolo bioinformático para el diseño automatizado de oligonucleótidos a ser utilizados en ensayos clínicos de identificación de *C. glabrata*, altamente específicos y sensibles, considerando criterios de:

- a) Especificidad: asegurando que la región amplificada sea exclusiva del genoma de *C. glabrata*.
- b) Sensibilidad: considerando que la región amplificada es altamente repetida en el genoma de *C. glabrata* y corresponden a secuencias mini y megasatélites.
- c) Eficiencia: considerando características termodinámicas de los oligonucleótidos que serán utilizados para la reacción de PCR como ausencia de regiones palindrómicas, longitud, temperatura de alineamiento y la presencia de bases G o C en sus extremos 3' que favorezcan el proceso de primado en la reacción de amplificación.

El conocimiento adquirido con este proyecto ayudará a mejorar el diagnóstico y tratamiento no tan sólo de las infecciones por *C. glabrata*, sino que también puedan generalizarse a cualquier otro patógeno de interés.

METODOLOGÍA

Estrategia general de análisis

Para poder diseñar de manera automática aquellos pares de oligos que optimicen la identificación de *C. glabrata*, en particular, o cualquier organismo de interés, en lo general, la estrategia computacional desarrollada se dividió en dos grandes partes. Dichas partes son:

- I) **Análisis de la frecuencia absoluta de todas las posibles palabras dentro de la secuencia genómica.** Considerando la hipótesis de que la sensibilidad y especificidad de un ensayo de PCR dependerá del número de veces que existe el locus a ser amplificado dentro de un genoma, la primera parte del análisis contempla la determinación de la frecuencia absoluta de todos los posibles k-meros o palabras del genoma y poder así, entonces, definir las palabras de mayor frecuencia. En esta primera etapa, el cálculo de frecuencias absolutas considera exclusivamente palabras 100% idénticas.
- II) **Análisis de las propiedades termodinámicas de los potenciales oligonucleótidos para el ensayo de amplificación por PCR.** Una vez conocidas las palabras (k-meros) con mayor frecuencia absoluta en la secuencia genómica, se realiza un análisis en términos de las propiedades termodinámicas para escoger aquellos k-meros que cumplan de mejor manera los requerimientos termodinámicos en el proceso de amplificación, como lo son la temperatura media de fusión, longitud, contenido de nucleótidos G/C en su extremo 3', que no sean auto-complementarias, que no exista complementariedad entre el par de oligonucleótidos a ser usados en la reacción de amplificación, etc.,

A continuación, se describen con más detalle cada una de estas dos etapas de análisis.

I) Análisis de la frecuencia absoluta de todas las posibles palabras dentro de la secuencia genómica.

En 1946 las de Bruijn desarrolló una solución matemática en un problema de supercuerdas, que son utilizadas en el análisis de secuencias. Su trabajo se centró en la creación de lo que se conoce como secuencias de De Bruijn, las cuales son secuencias especiales que contienen todas las subcadenas de longitud k (k -meros) exactamente una vez como subcadenas consecutivas. Estas secuencias son fundamentales en diversos campos, incluyendo la teoría de grafos, la informática teórica y la bioinformática.

En el contexto del análisis de secuencias de DNA, las secuencias de De Bruijn son utilizadas para ensamblar genomas a partir de lecturas de secuenciación, proporcionando una representación eficiente de todas las posibles secuencias de DNA de longitud k presentes en los datos de secuenciación. En dichas secuencias, existen n^k k -meros en un alfabeto que contiene n símbolos. En el caso del alfabeto para representar una secuencia de DNA, existen cuatro símbolos: A, T, G, C, de tal manera que el número posible de k -meros o palabras de longitud diez correspondería a $4^{10} = 1,048,576$ (Compeau, *et al.*, 2011). Considerando lo anterior, sería de esperarse que para un genoma del tamaño de *Escherichia coli*, con un poco más de 4,600,000 pb, los k -meros de longitud 10 ($k = 10$) deberían de encontrarse, en promedio, cuatro veces ($4,600,000 \text{ pb} / 1,048,576 \approx 4$), mientras que las palabras de longitud 11 ($k = 11$) deberían de ser, en promedio, únicas ($4,600,000 \text{ pb} / 4,194,304 \approx 1$).

La estrategia bioinformática planteada en el presente trabajo para la identificación de regiones repetidas en los genomas de interés está basada en la representación de una secuencia genómica en todos los posibles fragmentos pequeños de longitud k (k -meros) para evaluar sus frecuencias absolutas, es decir, identificar cuantas veces se repiten cada uno de los posibles fragmentos de longitud k en un genoma determinado. Una parte fundamental del algoritmo computacional en esta primera parte del análisis es que, la **posición y frecuencia absoluta** de cada posible palabra es almacenada en un arreglo, conocido como *arreglo hash*.

Un *arreglo hash*, también conocido como *tabla hash*, es una estructura de datos que utiliza una función de dispersión para mapear claves a valores dentro de un conjunto de datos. En este arreglo, los datos se almacenan y recuperan a través de una *clave única* que se asigna a un índice específico en el arreglo mediante la función de dispersión. La función de dispersión calcula un valor numérico (*hash*) a partir de la clave proporcionada, que, en nuestro caso, corresponden a cada uno de los k-meros del genoma. En teoría, existe un valor numérico único para cada uno de los k-meros de un genoma, que se puede utilizar como un índice dentro del arreglo para acceder de manera rápida y directa cualquiera de sus datos asociados. En nuestro caso, la posición o frecuencia absoluta de cada k-mero.

II. Análisis de las propiedades termodinámicas de los potenciales oligonucleótidos a ser empleados en el ensayo de amplificación por PCR.

II.1. Recursos bioinformáticos para el diseño de oligonucleótidos

Una herramienta muy útil en la biología molecular es la reacción en cadena de la polimerasa o (PCR) mediante el uso de oligonucleótidos, también conocidos como iniciadores o “primers”, por su nombre en inglés. Entre el conjunto de recursos bioinformáticos disponibles para el diseño de oligonucleótidos se pueden mencionar los siguientes:

OligoAnalyzer. Este sistema está disponible en la página web [www. Idtdna.com](http://www.idtdna.com). La principal característica de este programa es la fácil disposición de un diseño de oligos, inscribiéndote de forma gratuita. [OligoAnalyzer Tool - Primer analysis and Tm Calculator | IDT \(idtdna.com\)](http://www.idtdna.com) (Figura 12).

OligoAnalyzer

The screenshot shows the OligoAnalyzer web interface. On the left, there is a 'Sequence' input field containing the text 'GGC TTT CGC ACA TCA CGA CCA AG'. Above this field are tabs for '5' MOD', 'INTERNAL', '3' MOD', and 'MIXED BASES'. Below the sequence field, it says 'Bases 23' and there is a 'CLEAR SEQUENCE' button. A blue button below that says 'Try the new batch mode here'. On the right, the 'Parameter sets' section shows a dropdown for 'SpecSheet (Default)'. Below that, 'Target type' is set to 'DNA'. Further down, 'Oligo Conc' is set to '0.25 μM', 'Na⁺ Conc' is set to '50 mM', 'Mg²⁺ Conc' is set to '0 mM', and 'dNTPs Conc' is set to '0 mM'.

Figura 12. Portal web de OligoAnalyzer para el diseño de oligonucleótidos. (Figura tomada de página web Integrated DNATechnologies acquires Archer™).

GenScript. Es un sistema muy amigable para el diseño de oligonucleótidos, se encuentra en la página web www.genscript.com (Figura 13).

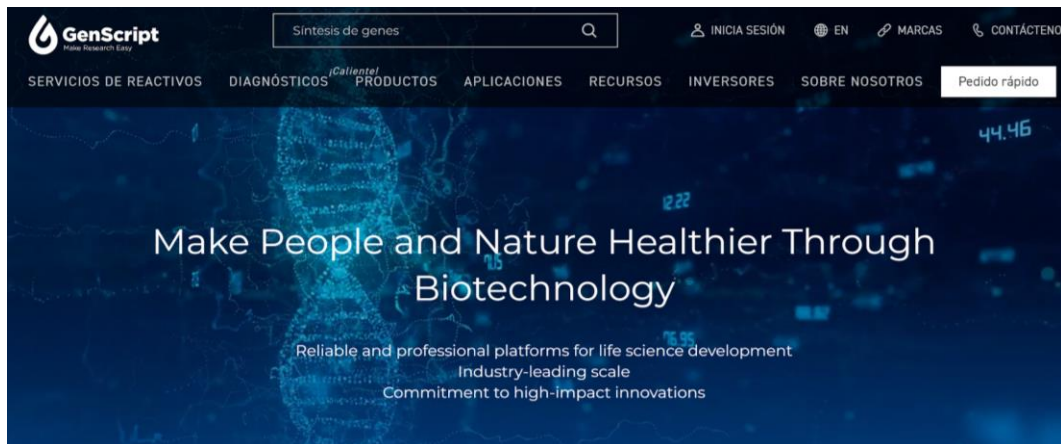


Figura 13. Portal web de GenScript para el diseño de oligonucleótidos. (Figura tomada de página web GenScript™).

Oligo Design Tools. Este software vinculado al Vector NTI Advance, OligoPerfect Designer está perfectamente diseñado para la elaboración y diseño de iniciadores u oligos en línea. Elaborado para pedidos en línea, por lo que nunca tendrá que cortar y pegar secuencias. Vector NTI Software | Thermo Fisher Scientific - MX (Figura 14).

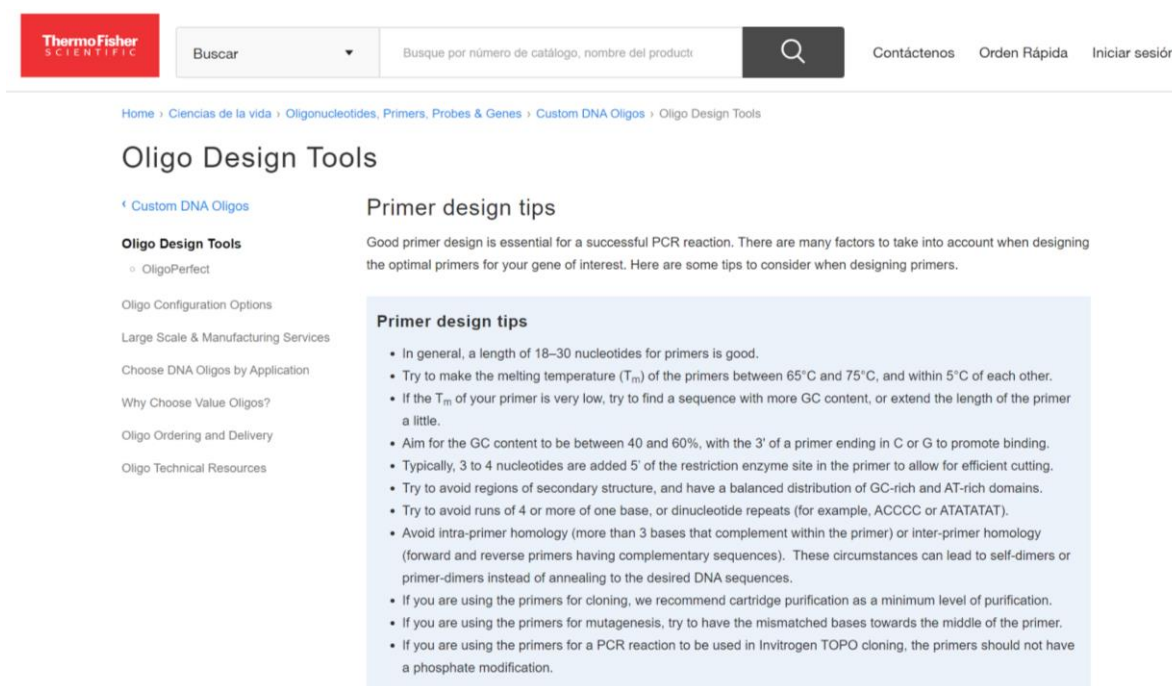


Figura 14. Portal de Thermo Fisher para el diseño de oligonucleótidos. (Figura tomada de página web VectorNTISoftware™).

II.2. Aspectos relevantes en el diseño de oligonucleótidos

Las características fundamentales que deben de ser consideradas en el diseño de oligonucleótidos incluyen la longitud de los oligonucleótidos (Wu *et al.* 2004), sus temperaturas de fusión (McPherson *et al.* 1991; Russell 2001; Wu *et al.* 2004), así como poseer al menos una guanina o citosina en su extremo 3', y la ausencia de una potencial posibilidad de auto-complementariedad o complementariedad con sus correspondientes pares de oligos usados en la reacción de PCR (Figura 15) (Boyle *et al.*, 2009; Vallone *et al.* 2004).



Figura 15. Diseño de Oligonucleótidos para la identificación de microsatélites. Las regiones del cromosoma en donde se encuentra una familia de microsatélite son de particular importancia en la identificación de organismos de interés médico (Figura tomada de Vignal *et al.*, 2002).

II.2.1. Longitud del oligonucleótido

La longitud de los oligonucleótidos suele ser de 16 a 28 nucleótidos con una diferencia de longitud de no más de 3 nucleótidos entre el oligo forward y el reverse, que se considera óptimo para un PCR. (Wu *et al.* 2004).

II.2.2. Contenido de G/C del oligonucleótido

La proporción de G/C es un valor porcentual que indica la cantidad de nucleótidos G (guanina) y C (citosina) que aparecen en un oligonucleótido o primer. Una proporción adecuada de G/C en un oligo, deberá idealmente oscilar entre un 40%-60%.

II.2.3. Composición de nucleótidos en el extremo 3´

Los pares de bases G/C son termodinámicamente más estables que los pares de bases AT. Algunas enzimas naturales agregan selectivamente bases en el extremo 3´ de cadenas bicatenarias; por lo tanto, se estima que los oligos se unen más firmemente si su última o dos últimas bases son G o C.

II.2.4. Temperatura de fusión (T_m)

La temperatura de fusión es el valor crítico más importante para realizar una amplificación exitosa. Generalmente la temperatura de fusión (T_m) en un par de oligos se debe encontrar entre los 50 y 62°C. Es muy importante que la PCR se realice sin exceder una diferencia de 5°C entre la T_m de cada oligo a utilizar. La temperatura de fusión de PCR se puede calcular mediante la fórmula de Wallace. (Wallace *et al.*, 1979) (Sambrook *et al.* 1989) (Untergasser *et al.*, 2012).

II.2.5. Ecuación de Wallace

Asume que la PCR se lleva a cabo en condiciones estándar

- [Primer] = 50mM
- [Na] = 50mM
- pH = 7.0

$$T_m = [4^\circ\text{C} \times (\text{Numero de G+C})] + [2^\circ\text{C} \times (\text{Numero de A+T})]$$

Ejemplo: 5'-ACAGATAGTGGGGATCCG-3'

$$\text{G} + \text{C} = 10$$

$$\text{A} + \text{T} = 8$$

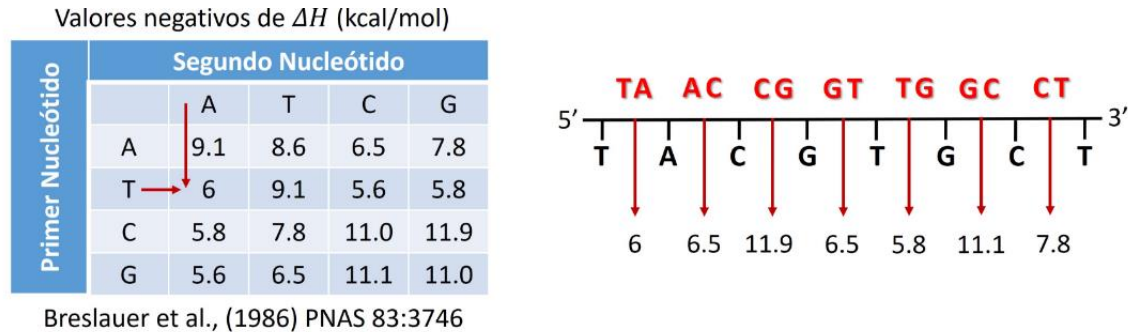
$$T_m = [4^\circ\text{C} \times (10)] + [2^\circ\text{C} \times (8)] = 56^\circ\text{C}$$

II.2.6. Método de vecinos cercanos de (*Nearest- Neighbor*)

Asume que:

- La estabilidad del dúplex de DNA depende de bases vecinas más cercanas (apilamiento)
- Toma en cuenta los parámetros termodinámicos de la interacción en el dúplex de DNA
- Entalpía
- Entropía
- Energía Libre de Gibbs

II.2.7. Entalpía: Se refiere a la cantidad de energía calorífica liberada o absorbida por el primer. El ΔH mide el cambio de la entalpía. La medida de ΔH [primer] es igual a la suma del valor de ΔH de cada par de nucleótidos que son vecinos cercanos en el primer (figura 16).



El signo negativo (-) indica que se libera calor en el proceso

$$\sum \Delta H = \Delta H_{(TA)} + \Delta H_{(AC)} + \Delta H_{(CG)} + \Delta H_{(GT)} + \Delta H_{(TG)} + \Delta H_{(GC)} + \Delta H_{(CT)}$$

$$\sum \Delta H = -(6 + 6.5 + 11.9 + 6.5 + 5.8 + 11.1 + 7.8) = -55.6 \frac{\text{kcal}}{\text{mol}}$$

Figura 16. Ejemplo del cálculo de entalpía (Tomado y modificado de Breslauer *et al.*, 1986).

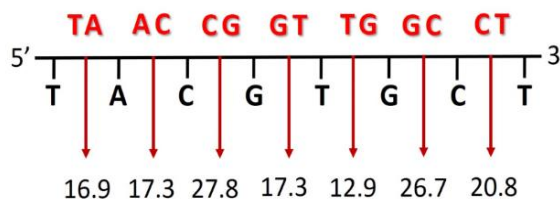
II.2.8. Entropía: Se refiere a la cantidad de desorden que exhibe un primer. Cuando se forma la relación DNA-primer, la entropía del primer y del DNA de una sola cadena disminuye, dado a que pierden su grado de libertad de movimiento cuando se unen uno al otro. Cambian de un estado desordenado a uno ordenado. El ΔS mide el cambio de la entropía.

ΔS [del primer] es igual a la suma del valor de ΔS de cada par de nucleótidos en el primer, que son vecinos cercanos.

Valores negativos de $\Delta S \left(\frac{\text{cal}}{\text{K mol}} \right)$

Primer Nucleótido	Segundo Nucleótido			
	A	T	C	G
	A	24	23.9	17.3
	T	16.9	24	13.5
	C	12.9	20.8	26.6
	G	13.5	17.3	26.7

Breslauer et al., (1986) PNAS 83:3746



El signo negativo (-) indica que el par de bases se vuelve menos desordenado

$$\sum \Delta S = \Delta S_{(TA)} + \Delta S_{(AC)} + \Delta S_{(CG)} + \Delta S_{(GT)} + \Delta S_{(TG)} + \Delta S_{(GC)} + \Delta S_{(CT)}$$

$$\sum \Delta S = -(16.9 + 17.3 + 27.8 + 17.3 + 12.9 + 26.7 + 20.8) = -139.7 \frac{\text{cal}}{\text{K mol}}$$

Figura 17. Ejemplo del cálculo de entropía Tomado y modificado de Breslauer et al., 1986).

II.2.9. Energía Libre de Gibbs (ΔG): Se refiere a la energía requerida para romper una estructura secundaria no deseada en el primer, esto ocurre cuando el primer se alinea a sí mismo (*hairpin*) o al otro primer (*primer dímero*). Mientras más negativo es el valor del ΔG más estable es la estructura secundaria.

II.2.10. Formación de dímeros

El alineamiento de dos cebadores u oligos (autosecuestro), da como resultado un dímero, la aparición de dímeros afecta negativamente a los resultados de una PCR, en consecuencia, esta hibridación deberá evitarse. La formación de dímeros ocurre cuando los dímeros entran en contacto directamente con el oligo de estudio, se reconocen entre sí por la complementariedad y este es “secuestrado” o hibridado con lo que forma horquillas en las estructuras de los iniciadores (Vallone et al. 2004).

II.2.11. Concentración de los Iniciadores.

La concentración óptima de los iniciadores debe ser de 0.1- 0.5 μM en una reacción estándar, una concentración superior a 0.5 μM puede causar la acumulación de productos no específicos debido a uniones inespecíficas en el DNA molde. También en caso de una alta concentración de DNA en el proceso, este se quedará adherido a la polimerasa, inhibiéndola. La concentración menor a 0.1 μM puede ser una concentración insuficiente

de los iniciadores y que no se terminen los ciclos de la PCR, lo que generaría un bajo rendimiento de amplificación. De manera general, 5 pmol de cada iniciador es adecuado para 25 µL de reacción.

II.2.13. Concentración de la DNA Polimerasa.

La concentración recomendada de enzima DNA polimerasa en la reacción de PCR es de 1-2.5 unidades por 100 µL de volumen de reacción (generalmente la enzima se comercializa a una concentración de 5U/µL). Esto representa en equivalencia molar, un valor bajo en comparación a los otros componentes de la reacción. Una baja concentración de enzima favorece la fidelidad de la amplificación del DNA, una alta concentración de enzima resulta en la amplificación de productos inespecíficos.

II.2.14. Factores que pueden modificar la fidelidad de la enzima DNA polimerasa.

Algunos factores que pueden modificar la fidelidad de la DNA polimerasa son la presencia de actividad exonucleasa 3'-5', la tendencia natural de la enzima para insertar errores, la facilidad con la que los errores pueden ser corregidos, la concentración de sales, específicamente la concentración de magnesio.

II.2.15. Aditivos usados en la reacción de amplificación por PCR.

En algunas ocasiones, para optimizar la PCR es recomendable la adición de sustancias adyuvantes que incrementan la especificidad y fidelidad (Landre et al. 1995), entre las más usadas están:

- DMSO (dimetilsulfóxido): Este al ser un desnaturalizante, impide que se plieguen oligos inespecíficos.

- Glicerol: en una concentración entre 15 y 20% mejora la amplificación de fragmentos con alta proporción de G+C e incrementa la estabilidad térmica de la enzima.

- Detergentes no iónicos: Tritón X-100, Tween 20 o Nonident P-40. Estabilizan a la Taq polimerasa e inhiben la formación estructuras secundarias en el DNA. Se utilizan en concentraciones entre 0.1 y 1%. Concentraciones mayores pueden generar amplificaciones no específicas.

- BSA (Albúmina de suero bovino): En una concentración final de 0.1 a 0.8 $\mu\text{g}/\mu\text{L}$ es particularmente útil para amplificar DNA antiguo o muestras de DNA con inhibidores de PCR como la melanina.

- Betaína: A una concentración final de 1 a 1.7 M, mejora la amplificación de los fragmentos.

- Aditivos comerciales: hay adyuvantes comerciales disponibles que optimizan la reacción de PCR. Sin embargo, no revelan su composición química.

II.16. Recomendaciones generales para la reacción de amplificación por PCR

En términos generales, la recomendación para obtener optimizar la amplificación de productos de PCR se recomienda:

a) Para mejorar la fidelidad y especificidad

- Seleccionar una enzima con actividad 3'-5' exonucleasa.
- Utilizar iniciadores con más de 15 nucleótidos.
- Incrementar la temperatura de alineamiento.
- Reducir el tiempo de desnaturalización y alineamiento en el ciclo.
- Reducir la concentración de iniciadores.
- Reducir el número de ciclos.
- Reducir la concentración de Mg^{2+} .
- Al aumentar Mg^{2+} también puede aumentar la estabilidad con puentes salinos entre los fosfatos de nucleótidos adyacentes.
- Revisar que el termociclador funcione adecuadamente.

b) Para mejorar la eficiencia de amplificación

- Incrementar la concentración de dNTPs y enzimas.
- Usar en concentración baja algún aditivo
- Reducir el tamaño del fragmento que se desea amplificar (amplicón).
- Probar con Pfu polimerasa
- Confirmar que los iniciadores no formen dímeros.

RESULTADOS

Los resultados del presente trabajo se dividen en dos grandes bloques. En el primero se realizan análisis encaminados a identificar las palabras o k-meros más frecuentes en una secuencia genómica y la frecuencia absoluta de todos sus k-meros vecinos dentro de un rango pre-establecido en el análisis. En la segunda parte de los resultados, se analizan a los k-meros más frecuentes identificados en la primera parte del estudio, en término de sus propiedades termodinámicas. En esta segunda parte, las palabras o k-meros son referidos como oligonucleótidos. Finalmente, con base en las frecuencias absolutas de las palabras y de sus propiedades termodinámicas al ser consideradas como oligos a ser empleados en reacciones de PCR, se puedan seleccionar los oligonucleótidos más apropiados

I. Análisis de la frecuencia absoluta de todas las posibles palabras dentro de la secuencia genómica

I.1. Obtención de secuencias genómicas de organismos de estudio

La secuencia genómica de *Candida glabrata* y las de otras 135 levaduras se obtuvieron de la base de datos KEGG. La lista completa de los organismos considerados en este estudio se presenta en el Apéndice 1. Las anotaciones de las secuencias genómicas fueron obtenidas de los archivos tipo GFF (por sus siglas en inglés: General Feature Format).

I.2. Identificación de regiones altamente repetidas en el genoma de *C. glabrata* a partir de la frecuencia de palabras (k-meros).

El análisis de las regiones altamente repetidas en el genoma de *C. glabrata* se realizó mediante el cálculo de frecuencias absolutos de “palabras” de una longitud predefinida. Inicialmente sólo se contabilizaron coincidencias exactas para aprovechar la flexibilidad del lenguaje Python de utilizar la estructura de datos conocida como “Hash” o “directorio”. En esta estructura de datos, cada palabra se utiliza como “llave” o referencia única, de tal forma que al ir recorriendo la secuencia genómica en “ventanas” que se desplazan un nucleótido a la vez, se tienen tantas ventanas como nucleótidos tiene el genoma de estudio. Palabras idénticas constituyen la misma llave, de tal forma que su valor correspondiente, dentro del

hash de frecuencias absolutas, será incrementado en una unidad cada vez que dicha *llave* aparezca en la secuencia al momento de ir “deslizando” la ventana de análisis (Figura 18).



Figura 18. Cálculo de frecuencias absolutas de palabras mediante el uso de estructuras Hash. En el ejemplo el tamaño de las palabras que se quieren contabilizar es de tres nucleótidos. En nuestro ejemplo, cada renglón representa un ciclo. En cada ciclo: a) una subsecuencia de longitud 3 (letras en color rojo), es tomada de la secuencia genómica y constituye la *llave* es transformada a un valor entero, o índice, que indica la posición que tendrá que ser incrementada en una unidad. B) Posteriormente, la ventana de análisis se “desliza” en un nucleótido hacia adelante para encontrar una nueva “*llave*” y su correspondiente índice, que indica la nueva posición que tendrá que ser incrementada en una unidad. C) Al final del proceso, se consideran para análisis posteriores, sólo aquellas cuyas frecuencias sean mayores a un valor preestablecido.

Los resultados de frecuencias varían significativamente dependiendo de tres variables:

- 1) Tamaño de la palabra (lon_palabra).
- 2) Tamaño de la secuencia (lon_secuencia).

- 3) Número de los diferentes tipos de caracteres en la secuencia (n_{tipos}). Para una secuencia de DNA, $n_{\text{tipos}} = 4$, ya que existen cuatro diferentes tipos de nucleótidos, mientras que para una secuencia de proteínas $n_{\text{tipos}} = 20$, que corresponde a los 20 diferentes tipos de aminoácidos.

En el caso del genoma de *C. glabrata*, $n_{\text{tipos}} = 4$ y $\text{lon_secuencia} = 12,338,308$ nucleótidos distribuidos en 13 cromosomas y el DNA mitocondrial. La frecuencia esperada de una palabra de longitud lon_palabra en una secuencia aleatoria de longitud lon_secuencia es igual a:

$$\text{Frec esperada} = \frac{\text{Lon_secuencia}}{n_{\text{tipos}}^{\text{Lon_palabra}}}$$

Considerando dicha ecuación, la frecuencia esperada de palabras repetidas en el genoma de *C. glabrata* se muestra en la Figura 19.

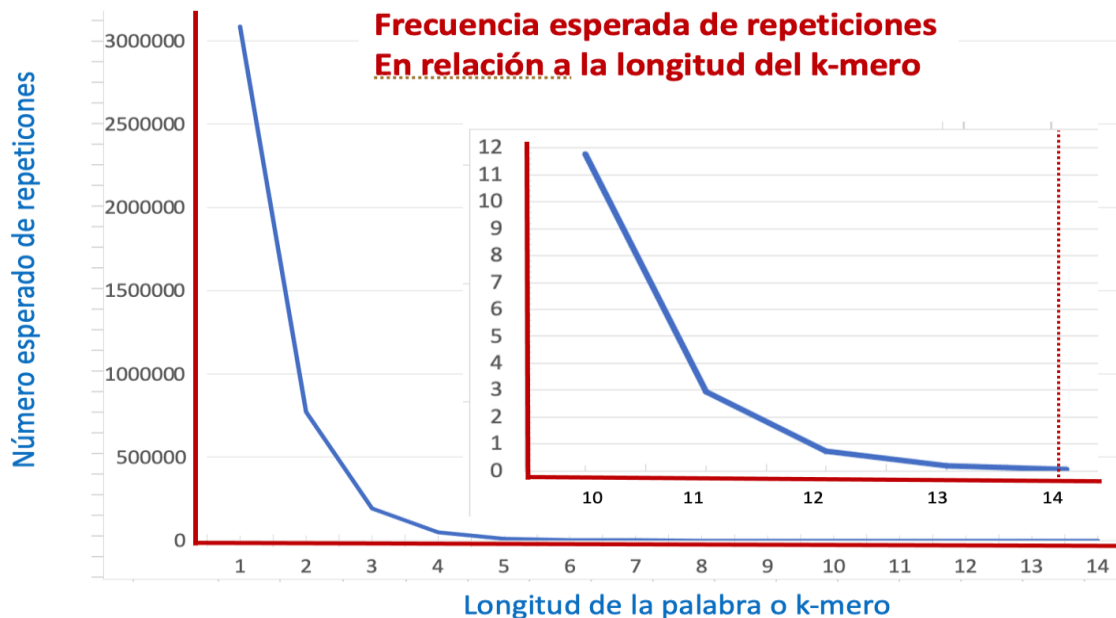


Figura 19. Cálculo de la frecuencia esperada de k-meros en función de su longitud, considerando una secuencia del tamaño del genoma de *C. glabrata*. Se muestra la gráfica a gran escala (k-meros de 1 a 14 nt) y pequeña escala (k-meros de 10 a 14 nt) de donde se puede observar que la longitud necesaria de un k-mero para que en promedio se espere que sea único o poco frecuente en el genoma de *C. glabrata*, es de al menos 14 nucleótidos.

De la anterior gráfica se puede inferir que una palabra tomada al azar, cuya longitud es de 14 nucleótidos, no se espera encontrarla en el genoma de *C. glabrata*, ya que se puede observar que va disminuyendo el número de repeticiones en función de la longitud del k-mero. En nuestro estudio se evaluaron diferentes longitudes de palabras mayores a 14 y encontramos que la longitud de 25 nucleótidos es adecuada para considerar que no aparecerá múltiples veces al azar. Por simplicidad en la representación de resultados, la palabra usada fue de 20 nt, pero para el diseño final se consideraron oligonucleótidos de 25 nt (Figura 20).

Análisis de palabras más frecuentes en el genoma de Candida glabrata					
NNNNNNNNNNNNNNNNNNNN	486	CCGTTGCCCTTGTCGATCGT	73	GATCGACAAGGGCAACGGTG	58
ACTGGATCTGTTCTTCATC	97	GTTGCCCTTGTCGATCGTCG	72	AGTCACCGTTGCCCTTGTCG	58
CTGGATCTGTTCTTCATCA	97	ACGACGATCGACAAGGGCAA	72	ATCGACAAGGGCAACGGTGA	58
GTGATGAAGAACGAGATCCA	97	CGTTGCCCTTGTCGATCGTC	72	TCGACAAGGGCAACGGTGAC	58
TGGATCTCGTTCTTCATCAC	97	GACGATCGACAAGGGCAACG	72	CGATCGACAAGGGCAACGGT	58
GACTGGATCTCGTTCTTCAT	97	CGACGATCGACAAGGGCAAC	72	CACCGTTGCCCTTGTCGATC	58
ATGAAGAACGAGATCCAGTC	97	TTGCCCTTGTCGATCGTCGT	72	GACCTTGTTTCGCACATCAC	58
TGATGAAGAACGAGATCCAG	97	ATCCATTGAAGCCAAGTGA	64	CTCGAAGTCACCGTTGCCCT	57
GATGAAGAACGAGATCCAGT	97	TCACTTGGCTTCAATGGGAT	64	TCTGGCTCTGGCTCTGGTTC	57
TTTTTTTTTTTTTTTTTTTT	92	ATCACTTGGCTTCAATGGGA	62	GAACCAGAGCCAGAGCCAGA	57
AAAAAAAAAAAAAAAAAAAAA	92	TCCATTGAAGCCAAGTGAT	62	AGGGCAACGGTGACTTCGAG	57
ATATATATATATATATATAT	84	CATCACTTGGCTTCAATGGG	62	GGGCAACGGTGACTTCGAGA	56
GGCTCTGGCTCTGGCTCTGG	82	CCCATTGAAGCCAAGTGATG	62	TCTCGAAGTCACCGTTGCC	56
CCAGAGCCAGAGCCAGAGCC	82	ACCAGAGCCAGAGCCAGAGC	60	CTCTGGCTCTGGCTCTGGTT	56
TCTGGTTCCAACCCAGGTGG	78	GCTCTGGCTCTGGCTCTGGT	60	AACCAGAGCCAGAGCCAGAG	56
CCACCTGGGTTGGAACACAGA	78	GACAAGGGCAACGGTGACTT	59	GCAACGGTGACTTCGAGACA	55
TCCAACCCAGGTGGCTCTGG	77	AAGTCACCGTTGCCCTTGTC	59	GCCGGCAGGCCGACTACAC	55
GTTCCAACCCAGGTGGCTCT	77	CAAGGGCAACGGTGACTTCG	59	TTGAAGCCAAGTGATGCCGG	55
GGTTCCAACCCAGGTGGCTC	77	TCGAAGTCACCGTTGCCCTT	59	CGGCATCACTTGGCTTCAAT	55
AGAGCCACCTGGGTTGGAAC	77	ACAAGGGCAACGGTGACTTC	59	ATTGAAGCCAAGTGATGCCG	55
CCAGAGCCACCTGGGTGGA	77	AGCAGCACCAGGACGTACAG	59	CCTGAACCAGAGCCAGAGCC	55
TTCCAACCCAGGTGGCTCTG	77	CTGTACGCTCTGGTGCTGCT	59	GCATCACTTGGCTTCAATGG	55
GAGCCACCTGGGTTGGAACC	77	GAAGTCACCGTTGCCCTTGT	59	GGCATCACTTGGCTTCAATG	55
CAGAGCCACCTGGGTGGAA	77	AAGGGCAACGGTGACTTCGA	59	GTAGGCTTGGCTCAGAGTC	55
AGCCACCTGGGTGGAACCA	75	CGAAGTCACCGTTGCCCTTG	59	GCTCTGGCTCTGGTTCAGGC	55
GCCACCTGGGTTGGAACCA	75	GTCACCGTTGCCCTTGTCGA	58	TCTGTCTCGAAGTCACCGTT	55
CTGGTTCCAACCCAGGTGGC	75	CGACAAGGGCAACGGTGACT	58	GGTCTGTCTCGAAGTCACCG	55
TGGTTCCAACCCAGGTGGCT	75	ACCGTTGCCCTTGTCGATCG	58	AGGTCTGTCTCGAAGTCACC	55
TATATATATATATATATA	74	TCACCGTTGCCCTTGTCGAT	58	CGGTGACTTCGAGACAGCC	55
ACGATCGACAAGGGCAACGG	73	GTGATGTGCGAAACAAGGTC	58	CAACGGTGACTTCGAGACAG	55

Figura 20. Evaluación de k-meros más frecuentes en el genoma de *C. glabrata*. Esquema que muestra la frecuencia de palabras mediante referencias tipo Hash, llaves con longitudes de palabras de 20 letras; a un costado de éstas, se observa el número de repeticiones en las que se encuentran dichas palabras. Las regiones en donde se representan los nucleótidos indefinidos con la letra *n*, son las que tienen una mayor frecuencia, seguido de las palabras de bajo contenido informativo resaltadas en amarillo, consistentes en repeticiones de uno, dos o más nucleótidos idénticos. No obstante, también se pueden observar palabras informativas que se repiten más de sesenta veces, marcadas en azul. Estas secuencias son candidatas para ser consideradas en el diseño de oligos de este trabajo.

I.3. Identificación de secuencias satélites en el genoma de *C. glabrata*.

Con el propósito de identificar regiones repetidas de mayor longitud que indiquen la presencia de secuencias satélites, se evaluaron la distancias entre palabras que comparten la misma secuencia. Estas distancias entre palabras fueron evaluadas a partir de sus posiciones dentro de cada uno de los 14 cromosomas de *C. glabrata*. Como puede observarse en la (Figura 21), las longitudes entre las palabras pueden no ser únicas, lo que indica una variación en el tamaño de las secuencias satélites.

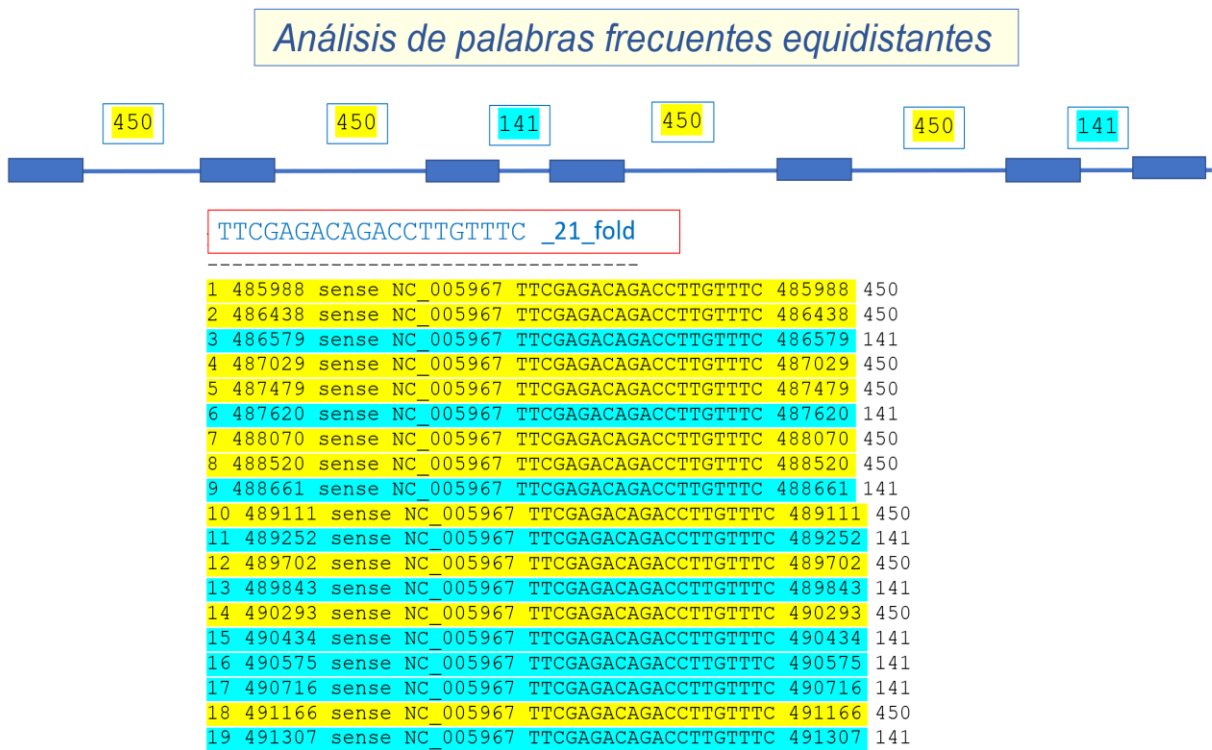


Figura 21. Cálculo de la distancia entre palabras repetidas en el genoma de *C. glabrata* que pudieran indicar la presencia de secuencias satélites. En la figura se representa la repetición de la palabra TTCGAGACAGACCTTGTTTC que, de acuerdo con las posiciones genómicas, pudieran equidistar 450 o 141 nucleótidos.

En la Figura 22, se enlistan el número de repeticiones en palabras con mayor frecuencia de aparición en el genoma de *C. glabrata* y la distancia que separa las copias entre sí, así como las palabras equidistantes con mayor frecuencia.

*Análisis de la distancia a la que equidistan las
palabras más frecuentes
en el genoma de Candida glabrata*

Número de repeticiones	Distancia entre palabras (nucleótidos)
1070	300
691	141
351	450
155	54
135	48
103	75
85	60
85	114
77	120
72	39
70	135
68	81
68	108
65	78
64	24
60	42
57	3
48	45
48	174

Figura 22. Palabras más frecuentes equidistantes en el genoma de *C. glabrata*. Esquema que representa la distancia más frecuente entre palabras idénticas, de donde se puede apreciar que la palabra, o k-mero más frecuente en el genoma de *C. glabrata* se repite 1,070 ocasiones y que las copias de dicha palabra equidistan 300 nucleótidos entre sí. La siguiente palabra con mayor frecuencia se repite 691 veces y dichas palabras equidistan 141 nucleótidos entre sí, y así, sucesivamente.

Análisis de palabras frecuentes equidistantes
Longitud de la repetición: 300 nt

Posición = 1013582

ATGAAGAACGAGATCCAGTCCGCTCTCGACATAGTCCGCAGCAGTGGTGGCTGTGATCGTCACAGGTGGAGGGCTGTACGTCCTGGTGCTGCTAGTGGTGATCAACC
CACCAAACCTCGTTGGTGGTCTGCCAGTAGCTGATGACCAGGTACTCAGTCACTTCGTTGCCAGATCAACAGTCTTGGTCTCGGCAGGAGGTGCAGGCCCGAACGAC
GTGTGCGGAGCCTCAGGCAAGTTGTAGTCACGCGTAGCATTGAACAGTGTGCTGTCTGTAATAATCTCCCCCTTATCATTCTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 1013882

ATGAAGAACGAGATCCAGTCCGCTCTCGACATAGTCCGCAGCAGTGGTGGCTGTGATCGTCACAGGTGGAGGGCTGTACGTCCTGGTGCTGCTAGTGGTGATCAACC
CACCAAACCTCGTTGGTGGTCTGCCAGTAGCTGATGACCAGGTACTCAGTCACTTCGTTGCCAGATCAACAGTCTTGGTCTCGGCAGGAGGTGCAGGCCCGAACGAC
GTGTGCGGAGCCTCAGGCAAGTTGTAGTCACGCGTAGCATTGAACAGTGTGCTGTCTGTAATAATCTCCCCCTTATCATTCTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 1014182

ATGAAGAACGAGATCCAGTCCGCTCTCGACATAGTCCGCAGCAGTGGTGGCTGTGATCGTCACAGGTGGAGGGCTGTACGTCCTGGTGCTGCTAGTGGTGATCAACC
CACCAAACCTCGTTGGTGGTCTGCCAGTAGCTGATGACCAGGTACTCAGTCACTTCGTTGCCAGATCAACAGTCTTGGTCTCGGCAGGAGGTGCAGGCCCGAACGAC
GTGTGCGGAGCCTCAGGCAAGTTGTAGTCACGCGTAGCATTGAACAGTGTGCTGTCTGTAATAATCTCCCCCTTATCATTCTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 1014482

ATGAAGAACGAGATCCAGTCCGCTCTCGACATAGTCCGCAGCAGTGGTGGCTGTGATCGTCACAGGTGGAGGGCTGTACGTCCTGGTGCTGCTAGTGGTGATCAACC
CACCAAACCTCGTTGGTGGTCTGCCAGTAGCTGATGACCAGGTACTCAGTCACTTCGTTGCCAGATCAACAGTCTTGGTCTCGGCAGGAGGTGCAGGCCCGAACGAC
GTGTGCGGAGCCTCAGGCAAGTTGTAGTCACGCGTAGCATTGAACAGTGTGCTGTCTGTAATAATCTCCCCCTTATCATTCTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 1014782

ATGAAGAACGAGATCCAGTCCGCTCTCGACATAGTCCGCAGCAGTGGTGGCTGTGATCGTCACAGGTGGAGGGCTGTACGTCCTGGTGCTGCTAGTGGTGATCAACC
CACCAAACCTCGTTGGTGGTCTGCCAGTAGCTGATGACCAGGTACTCAGTCACTTCGTTGCCAGATCAACAGTCTTGGTCTCGGCAGGAGGTGCAGGCCCGAACGAC
GTGTGCGGAGCCTCAGGCAAGTTGTAGTCACGCGTAGCATTGAACAGTGTGCTGTCTGTAATAATCTCCCCCTTATCATTCTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 1015082

ATGAAGAACGAGATCCAGTCCGCTCTCGACATAGTCCGCAGCAGTGGTGGCTGTGATCGTCACAGGTGGAGGGCTGTACGTCCTGGTGCTGCTAGTGGTGATCAACC
CACCAAACCTCGTTGGTGGTCTGCCAGTAGCTGATGACCAGGTACTCAGTCACTTCGTTGCCAGATCAACAGTCTTGGTCTCGGCAGGAGGTGCAGGCCCGAACGAC
GTGTGCGGAGCCTCAGGCAAGTTGTAGTCACGCGTAGCATTGAACAGTGTGCTGTCTGTAATAATCTCCCCCTTATCATTCTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 2207485

ATGAAGAACGAGATCCAGTCCGACTCAATGTAGTCCGACGCAGTGGTGGCTGTGATCGTACAGTCCCGGTGGAGGACTGTAGGTGGCCGTCGAGTTTGTCTGTATCAGAC
CCCCGAATTCATTGGTGGTCTGCCAGTAGCTGATGACCAGATACTCAGTCACTTCGTTGCCAGATCAACTGTCTTCGTCACAGCTGGAGGTGCAGGCCCGAACGACG
TGTGGGGGGCCTCCGGCAAGTTGTAGTCACGCGTGGCGTTGAACAGAGTGTGCTGCTGTAACAATATCGCCCTTCTCATCAGTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 2207785

ATGAAGAACGAGATCCAGTCCGACTCAATGTAGTCCGACGCAGTGGTGGCTGTGATCGTACAGTCCCGGTGGAGGACTGTAGGTGGCCGTCGAGTTTGTCTGTATCAGAC
CCCCGAATTCATTGGTGGTCTGCCAGTAGCTGATGACCAGATACTCAGTCACTTCGTTGCCAGATCAACTGTCTTCGTCACAGCTGGAGGTGCAGGCCCGAACGACG
TGTGGGGGGCCTCCGGCAAGTTGTAGTCACGCGTGGCGTTGAACAGAGTGTGCTGCTGTAACAATATCGCCCTTCTCATCAGTGGTG**ATGAAGAACGAGATCCAGTC**

Posición = 2208085

ATGAAGAACGAGATCCAGTCCGACTCAATGTAGTCCGACGCATGGTGGCTGTGATCGTACAGTCAACCGTTGGAGGACTGTAGGTGGCCGTCGAGTTTGTCTGTATCAGAC
CCCCGAATTCATTGGTGGTCTGCCAGTAGCTGATGACCAGATACTCAGTCACTTCGTTGCCAGATCAACTGTCTTCGTCACAGCTGGAGGTGCAGGCCCGAACGACG
TGTGGGGGGCCTCCGGCAAGTTGTAGTCACGCGTGGCGTTGAACAGAGTGTGCTGCTGTAACAATATCGCCCTTCTCATCAGTGGTG**ATGAAGAACGAGATCCAGTC**

Figura 23. Análisis de palabras que equidistan 300 nucleótidos en el genoma de *C. glabrata*. La figura muestra al K-mero **ATGAAGAACGAGATCCAGTC** de longitud de 20 nucleótidos y 280 nucleótidos contiguos, de tal manera que los fragmentos repetidos distan 300 nucleótidos, que se repite 1070 veces en el genoma de *C. glabrata* (Ver Figura 22). La posición del inicio de cada fragmento está indicada. Para diferenciar los fragmentos contiguos, éstos se muestran en colores azul y café de manera alternativa.

Análisis de palabras frecuentes equidistantes
Longitud de la repetición: 141 nt

p = 486454

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 487495

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 488536

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 489127

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 489718

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 490309

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 490450

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 490591

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 15213603

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 15214026

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 15214167

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 15214308

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 15214449

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 15214590

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

p = 15214731

TTTCGCACATCACGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACCATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACGACGACG
ATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCACGACCAAG**

Figura 24. Análisis de palabras que equidistan 141 nucleótidos en el genoma de *C. glabrata*. La figura muestra al K-mero **CGCACATCACGACCAAG** de longitud de 20 nucleótidos y 121 nucleótidos contiguos, de tal manera que los fragmentos repetidos distan 141 nucleótidos, que se repite 691 veces en el genoma de *C. glabrata* (Ver Figura 22). La posición del inicio de cada fragmento está indicada. Para diferenciar los fragmentos contiguos, éstos se muestran en colores azul y café de manera alternativa.

Análisis de palabras frecuentes equidistantes
Longitud de la repetición: 450 nt

p = 485554

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAAGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 486004

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 486595

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 487045

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 487636

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 488086

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 488677

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 489268

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

p = 489859

TTTCGCACATCAGGACCAAGGACTCTGACGGCAAGCCTACCACGATCACCACCACATCCCATTGAAGCCAAGTGATGCCGGCGAGGCCGACTACACCACC
 ACCATTACTTCCGACGGTCACACCATCACAGAAGTAGTCTCCACATCACCACCAGTATTGACACGGTAACACTGTTACATACACCACCACGATGGCCTCCT
 TCGATCAGGTAGACGACGGCGGTACACGTCTGCTCCACCTTACTCACAGCCACCAAGTTGTGACAACCACTGTGACGGAGGACGACGGTAGTAGCAAGAC
 AGTTGTGATCTCTTACTCGCCATCTGAGGGCGAGGATGGGGTTACCCGTACAGGTACCACCACATCTCCACCATCACTGCTAGCGGCGAGGCCGACTACA
 CGACGACGATCGACAAGGGCAACGGTGACTTCGAGACAGACCTTG**TTTCGCACATCAGGACCAAG**

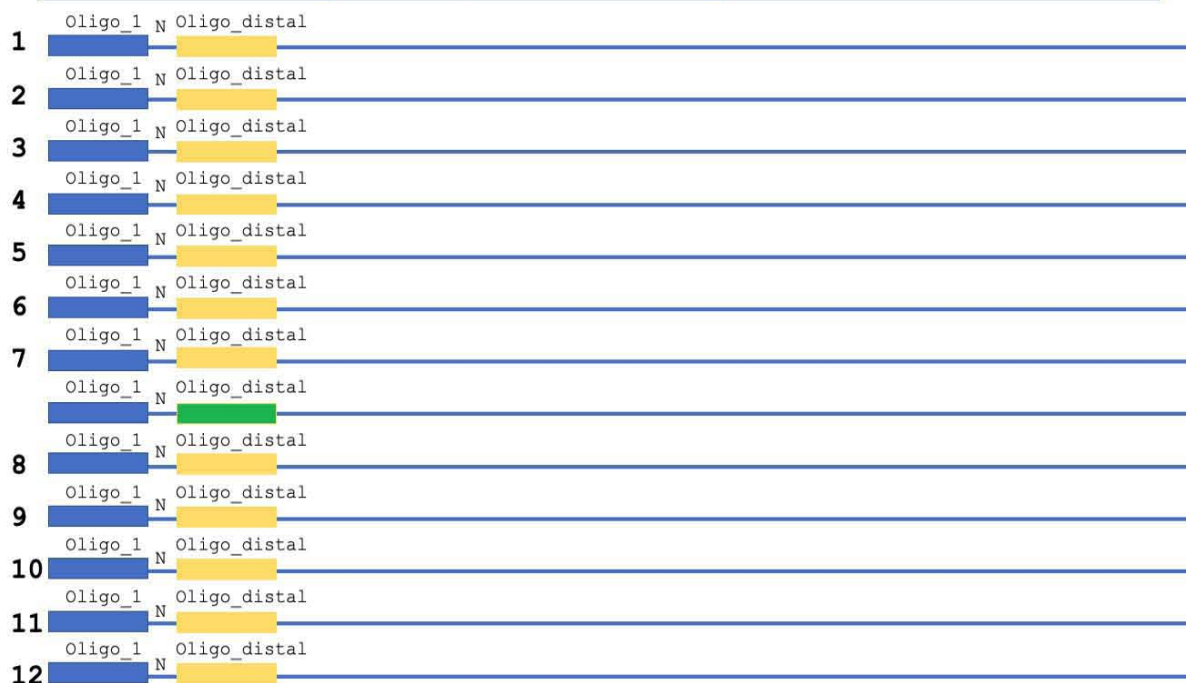
Figura 25. Análisis de palabras que equidistan 450 nucleótidos en el genoma de *C. glabrata*. La figura muestra al K-mero **CGCACATCAGGACCAAG** de longitud de 20 nucleótidos y 430 nucleótidos contiguos, de tal manera que los fragmentos repetidos distan 450 nucleótidos, que se repite 351 veces en el genoma de *C. glabrata* (ver Figura 22). La posición del inicio de cada fragmento está indicada. Para diferenciar los fragmentos contiguos, éstos se muestran en colores azul y café de manera alternativa.

I.4. Cálculo de frecuencias de palabras vecinas de los k-meros más frecuentes.

Las palabras o k-meros correspondientes a potenciales regiones satélites en el genoma de *C. glabrata* se ordenaron en términos de su frecuencia absoluta, de mayor a menor. Para cada una de dichas palabras que, por cuestiones explicativas, llamaremos k-mero izquierdo, se realizó un proceso de análisis de palabras “vecinas” río abajo, que, por cuestiones explicativas, llamaremos k-mero derecho. Es importante mencionar que, en nuestro estudio, ambos k-meros deberán de ser de la misma longitud para favorecer que tiendan a poseer valores similares de apareamiento con el DNA, y que se verificará a posteriori en un análisis termodinámico. Los k-meros derechos así seleccionados son agrupados en sus convenientes distancias a su correspondiente k-mero izquierdo, tal y como se esquematiza en las Figura 26 y en la Tabla 3. El análisis realizado en esta sección será la base para evaluar los potenciales productos de amplificación en una reacción de PCR, en etapas posteriores del análisis.



Distancia entre oligos = 1 nt, Número de oligos distal conservados = 12



Distancia entre oligos = 2 nt, Número de oligos distal conservados = 10

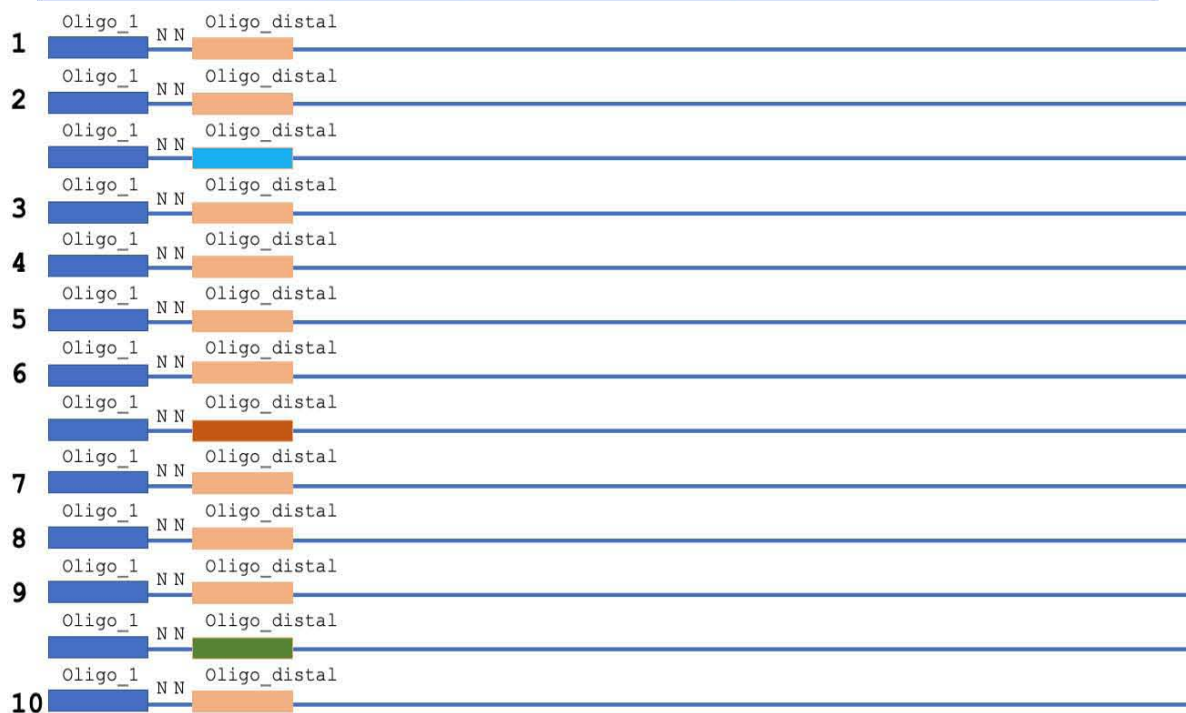




Figura 26. Representación del cálculo de frecuencias de palabras vecinas. Dada una palabra identificada como frecuente en el genoma de *C. glabrata* (representada por el rectángulo azul y con la etiqueta Oligo_1), se contabilizó la frecuencia con la que sus palabras vecinas cuya secuencia está conservada se encontraban a distancias específicas; de 0, 1, 2, y 10 nucleótidos, representados por los rectángulos de los colores verde, amarillo, café y morado, respectivamente. Dichos rectángulos tienen la etiqueta Oligo_distal. La identidad en el color de los rectángulos representa identidad en la secuencia de las palabras, esto quiere decir que los rectángulos cuyo color difiera a los colores de oligos distales cuyos colores son: verde, amarillo, café y morado, representan palabras con secuencias distintas a las del consenso.

I.5. Índice de visibilidad relativa.

En la mayoría de las técnicas de cuantificación de DNA, la capacidad de identificar un fragmento de DNA es directamente proporcional al número de nucleótidos del fragmento y

al número de veces que dicho fragmento exista en la muestra analizada. Con el propósito de contar con una medida de la *visibilidad relativa esperada* de los productos de PCR a ser amplificados, propusimos una simple función que varía de manera directa y proporcional al tamaño del producto esperado y al número de copias de dicha región en el genoma de *C. glabrata*, de tal forma que la función propuesta es:

$$\text{Índice de visibilidad relativa} = \text{Lon_secuencia} \times \text{número de repeticiones}$$

Debido a que, en términos generales, la conservación de las palabras vecinas decrece a medida que la distancias entre ellas aumenta, el *índice de visibilidad relativa* no necesariamente aumentará de manera lineal conforme la longitud de la secuencia esperada se haga mayor, (oligos más distantes) y por ello, cada palabra a ser considerada como un primer oligo en la reacción de PCR, tendrá un vecino para el cual el *índice de visibilidad relativa* sea óptimo. En la Tabla 3 se muestra los valores del *índice de visibilidad relativa* para el K-mero ACTGGATCTCGTTCTTCATC que se repite 97 veces en el genoma de *C. glabrata*, y en la Figura 27 se muestra la representación gráfica del *índice de visibilidad relativa* para los k-meros TTTCGCACATCACGACCAAG y TTGAGAGCGACTGGATCTCG que se repiten 41 veces en dicho genoma.

Pb entre oligos	Oligo izquierdo	Oligo Derecho	Veces en el genoma	Longitud del PCR	Índice de visibilidad relativa
0	ACTGGATCTCGTTCTTCATC	ACTGTTGATGAGAAGGGAGA	36	40	1440
1	ACTGGATCTCGTTCTTCATC	CTGTTGATGAGAAGGGAGAG	35	41	1435
2	ACTGGATCTCGTTCTTCATC	TGTTGATGAGAAGGGAGAGA	35	42	1470
3	ACTGGATCTCGTTCTTCATC	GTTGATGAGAAGGGAGAGAT	35	43	1505
4	ACTGGATCTCGTTCTTCATC	TTGATGAGAAGGGAGAGATT	35	44	1540
5	ACTGGATCTCGTTCTTCATC	TGATGAGAAGGGAGAGATTG	34	45	1530
6	ACTGGATCTCGTTCTTCATC	GATGAGAAGGGAGAGATTGT	34	46	1564
7	ACTGGATCTCGTTCTTCATC	ATGAGAAGGGAGAGATTGTT	34	47	1598
8	ACTGGATCTCGTTCTTCATC	TGAGAAGGGAGAGATTGTTA	34	48	1632
9	ACTGGATCTCGTTCTTCATC	GAGAAGGGAGAGATTGTTAC	34	49	1666
10	ACTGGATCTCGTTCTTCATC	AGAAGGGAGAGATTGTTACC	34	50	1700
11	ACTGGATCTCGTTCTTCATC	GAAGGGAGAGATTGTTACCA	34	51	1734
12	ACTGGATCTCGTTCTTCATC	AAGGGAGAGATTGTTACCAG	34	52	1768
13	ACTGGATCTCGTTCTTCATC	AGGGAGAGATTGTTACCAGC	34	53	1802
14	ACTGGATCTCGTTCTTCATC	GGGAGAGATTGTTACCAGCA	34	54	1836
15	ACTGGATCTCGTTCTTCATC	GGAGAGATTGTTACCAGCAG	31	55	1705
16	ACTGGATCTCGTTCTTCATC	GAGAGATTGTTACCAGCAGC	31	56	1736
17	ACTGGATCTCGTTCTTCATC	AGAGATTGTTACCAGCAGCA	31	57	1767
18	ACTGGATCTCGTTCTTCATC	GAGATTGTTACCAGCAGCAC	31	58	1798
19	ACTGGATCTCGTTCTTCATC	AGATTGTTACCAGCAGCACT	31	59	1829
20	ACTGGATCTCGTTCTTCATC	GATTGTTACCAGCAGCACTC	31	60	1860
21	ACTGGATCTCGTTCTTCATC	ATTGTTACCAGCAGCACTCT	31	61	1891
22	ACTGGATCTCGTTCTTCATC	TTGTTACCAGCAGCACTCTG	31	62	1922
23	ACTGGATCTCGTTCTTCATC	TGTTACCAGCAGCACTCTGT	31	63	1953
24	ACTGGATCTCGTTCTTCATC	GTTACCAGCAGCACTCTGTT	31	64	1984
25	ACTGGATCTCGTTCTTCATC	TTACCAGCAGCACTCTGTTC	32	65	2080
26	ACTGGATCTCGTTCTTCATC	TACCAGCAGCACTCTGTTC	32	66	2112
27	ACTGGATCTCGTTCTTCATC	ACCAGCAGCACTCTGTTC	32	67	2144
28	ACTGGATCTCGTTCTTCATC	CCAGCAGCACTCTGTTC	32	68	2176
29	ACTGGATCTCGTTCTTCATC	CAGCAGCACTCTGTTC	32	69	2208
30	ACTGGATCTCGTTCTTCATC	AGCAGCACTCTGTTC	35	70	2450
31	ACTGGATCTCGTTCTTCATC	GCAGCACTCTGTTC	32	71	2272
32	ACTGGATCTCGTTCTTCATC	CAGCACTCTGTTC	33	72	2376
33	ACTGGATCTCGTTCTTCATC	AGCACTCTGTTC	33	73	2409
34	ACTGGATCTCGTTCTTCATC	GCACTCTGTTC	32	74	2368
35	ACTGGATCTCGTTCTTCATC	CACTCTGTTC	35	75	2625
36	ACTGGATCTCGTTCTTCATC	ACTCTGTTC	35	76	2660
37	ACTGGATCTCGTTCTTCATC	CTCTGTTC	35	77	2695
38	ACTGGATCTCGTTCTTCATC	TCTGTTC	35	78	2730
39	ACTGGATCTCGTTCTTCATC	CTGTTCAATGCTACGCGTGA	54	79	4266
40	ACTGGATCTCGTTCTTCATC	TGTTCAATGCTACGCGTGAC	52	80	4160
41	ACTGGATCTCGTTCTTCATC	GTTCAATGCTACGCGTGACT	52	81	4212
42	ACTGGATCTCGTTCTTCATC	TTCAATGCTACGCGTGACTA	52	82	4264
43	ACTGGATCTCGTTCTTCATC	TCAATGCTACGCGTGACTAC	52	83	4316
44	ACTGGATCTCGTTCTTCATC	CAATGCTACGCGTGACTACA	52	84	4368
45	ACTGGATCTCGTTCTTCATC	AATGCTACGCGTGACTACAT	26	85	2210

Tabla 3. Evaluación del índice de la visibilidad relativa de las palabras más frecuentes en el genoma de *C. glabrata*. En la tabla se muestran las palabras vecinas (tercera columna) a la palabra ACTGGATCTCGTTCTTCATC (segunda columna), que resultó ser una de las palabras más frecuentes en el genoma de *C. glabrata*, con 97 repeticiones. En la primera columna se indica la separación o número de nucleótidos entre las pares palabras, y en la columna del extremo derecho se indica el número de veces que los diferentes pares de palabras se encuentran repetidos. El renglón resaltado en amarillo indica que el par de palabras ACTGGATCTCGTTCTTCATC y CTGTTCAATGCTACGCGTGA separados 39 nucleótidos, se encuentra en el genoma de *C. glabrata* 54 veces.

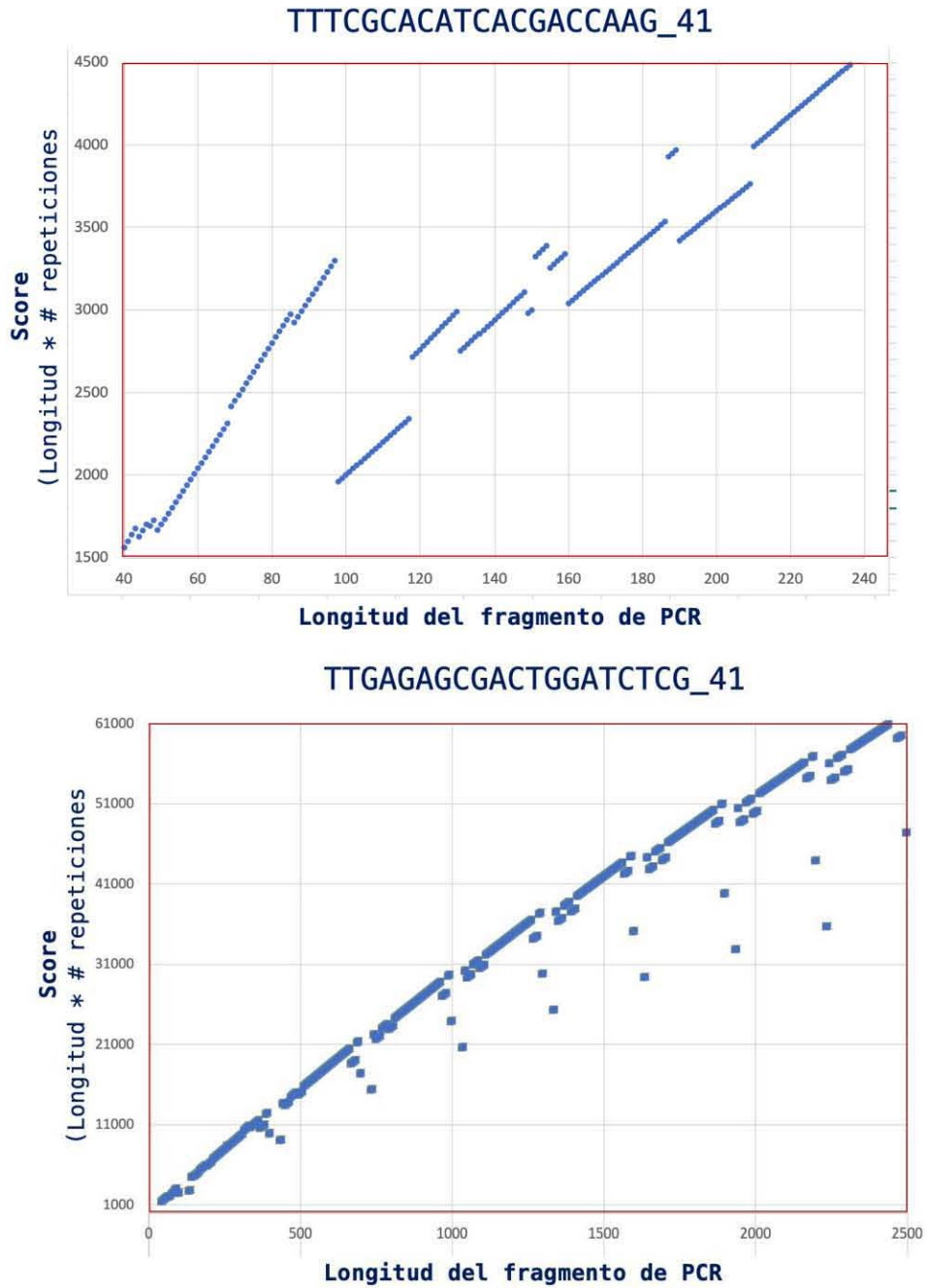


Figura 27. Evaluación del índice de la visibilidad relativa para dos k-meros que se repiten 41 veces en el genoma de *C. glabrata*. El índice de visibilidad relativa de nuestros productos de PCR a ser amplificados se obtiene al multiplicar el tamaño del fragmento amplificado por el número de copias en la que las palabras analizadas se encuentran equidistantes. Es importante notar que para palabras con la misma frecuencia inicial (TTTCGCACATCACGACCAAG y TTGAGAGCGACTGGATCTCG, 41 veces), el índice de visibilidad relativa puede diferir significativamente (4,500 vs 61,000, respectivamente).

II) Análisis de las propiedades termodinámicas de los potenciales oligonucleótidos para el ensayo de amplificación por PCR.

En la primera sección de este trabajo, (I. Análisis de la frecuencia absoluta de todas las posibles palabras dentro de la secuencia genómica) estuvo enfocada en la determinación de la frecuencia absoluta de todos los posibles k-meros y la definición aquellos que se repiten junto con K-meros vecinos con mayor frecuencia, siempre contemplando exclusivamente K-meros 100% idénticos. En esta segunda etapa, ampliaremos nuestro análisis para incluir condiciones que se asemejan más a lo que sucede en una reacción de PCR. En esta segunda sección emplearemos el término *oligonucleótidos*, o sencillamente, *oligos*, a los elementos descritos en la primera parte del estudio como *palabras* o *k-meros*. Adicionalmente, nuestro análisis incluirá las características termodinámicas que definen con mayor precisión, la eficiencia de los pares de oligos a ser usados en reacciones de PCR para la síntesis de DNA por la DNA polimerasa. Como un primer ejemplo, ilustraremos el efecto en el contenido G/C de los oligos parcial o totalmente complementarios, con secuencias con un contenido del 100% A/T (Figura 28A), *versus* oligos con contenidos del 100% G/C (Figura 28B).

II.1 Efecto del contenido de G/C en la energía libre de hibridación de los oligonucleótidos

```

AAAAAAAAAAAAAAAAAAAAA      AAAAAAAATTAAAAAAAAAAAA
.((.(&.)).      8,11 : 17,20 ( 2.40) i:10,j:18 < 2.40>

AAAAAAAAAAAAAAAAAAAAA      AAAAAAAATTAAAAAAAAAAAA
.(((.(&.))).      7,11 : 16,20 ( 1.50) i:10,j:17 < 1.50>

AAAAAAAAAAAAAAAAAAAAA      AAAAAATTAAAAAAAAAAAA
.((((.(&.))).      6,11 : 15,20 ( 0.60) i:10,j:16 < 0.60>

AAAAAAAAAAAAAAAAAAAAA      AAAAAATTAAAAAAAAAAAA
.((((((.(&.))).      6,12 : 14,20 (-0.30) i:11,j:15 <-0.30>

AAAAAAAAAAAAAAAAAAAAA      AAAAAATTAAAAAAAAAAAA
.(((((((.(&.))).      6,13 : 13,20 (-1.20) i:12,j:14 <-1.20>

AAAAAAAAAAAAAAAAAAAAA      AAAAAATTAAAAAAAAAAAA
.((((((((.(&.))).      6,14 : 12,20 (-2.10) i:13,j:13 <-2.10>

AAAAAAAAAAAAAAAAAAAAA      AAAAAATTAAAAAAAAAAAA
.((((((((((.(&.))).      6,15 : 11,20 (-3.00) i:14,j:12 <-3.00>

AAAAAAAAAAAAAAAAAAAAA      AAAAAATTAAAAAAAAAAAA
.(((((((((((.(&.))).      5,15 : 10,20 (-3.90) i:14,j:11 <-3.90>

AAAAAAAAAAAAAAAAAAAAA      AAAAAATTAAAAAAAAAAAA
.((((((((((((.(&.))).      5,16 : 9,20 (-4.80) i:15,j:10 <-4.80>

AAAAAAAAAAAAAAAAAAAAA      TTTTTTTTTTTTTTTTTT
.(((((((((((((((((&))))))))))))))))).      1,20 : 1,20 (-12.10) i:20,j:1 <-13.10>

```

A) Cálculo de energía de hibridación de oligos cuya composición es 100% A/T

**Oligos de 20 nts completamente complementarios cuya composición es 100% A/T
DeltaG -12.0 Kcal/mol**

[illegible]

Figura 28. Efecto del contenido de G/C de oligos en su energía libre de hibridación. En la figura se muestra el resultado del programa RNAplex (<http://www.bioinf.uni-leipzig.de/Software/RNAplex/>) para la interacción de oligonucleótidos parcial o totalmente complementarios. A) Oligos cuya composición es 100% A/T, B) Oligos con contenidos del 100% G/C.

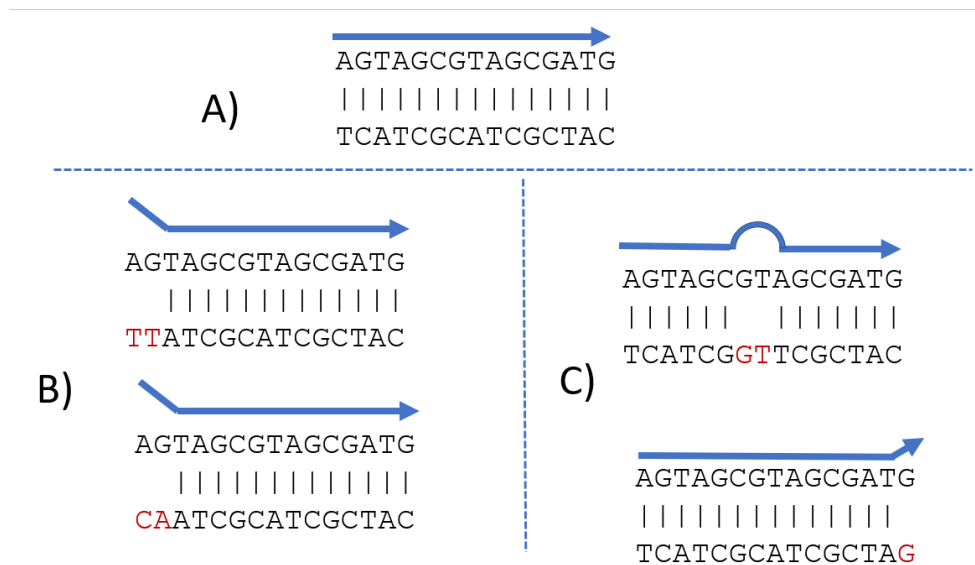


Figura 29. Tipos de complementariedad de oligos y su compatibilidad para la amplificación en ensayos de PCR. A) Totalmente complementarias. B) Parcialmente complementarias y compatibles con la amplificación de productos de PCR C) Parcialmente complementarias e incompatibles con la amplificación de productos de PCR. (Tomado y modificado de Oliveira *et al.*, 2006).

II.2. Resultados representativos para un par de oligos identificados en el proceso de análisis con el genoma de *C. glabrata*.

Del conjunto de posibles pares de oligonucleótidos resultantes del proceso de análisis, en la Figura 30 se muestra el resultado obtenido para un par en específico. Como se puede observar en dicha figura, nuestro programa no sólo identifica las palabras idénticas a las de los oligos seleccionados, si no también, las que, de acuerdo con los parámetros de la temperatura de alineamiento entre las secuencias de los oligos y el DNA genómico, tienen posibilidad de hibridar y producir productos de amplificación reales. Dichos oligos se señalan con un asterisco en la Figura 30.

Oligos sin considerar mismatches					
2440	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	25	48.10	45.30
2140	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	26	48.10	45.30
1840	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	27	48.10	45.30
1540	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	28	48.10	45.30
1240	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	29	48.10	45.30
940	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	30	48.10	45.30
640	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	31	48.10	45.30
340	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	32	48.10	45.30
40	TGAGAGCGACTGGATCTCGTTCTTC	CTCTGTTCAATGCTACGCGTGACTA	34	48.10	45.30

Oligos considerando mismatches					
--------------------------------	--	--	--	--	--

Candida glabrata CBS138					
-------------------------	--	--	--	--	--

cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	484449	484473
cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	484749	484773
cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	485049	485073
cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	485349	485373
cgr	5p	CGAG T CGGACTGGATCTCGTTCTTC*	2203009	2203033
cgr	5p	CGAG T CGGACTGGATCTCGTTCTTC*	2203316	2203340
cgr	5p	CGAG T CGGACTGGATCTCGTTCTTC*	2203619	2203643
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	2366148	2366172
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	2366448	2366472
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	2366748	2366772
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	2367048	2367072
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	2367348	2367372
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6854954	6854978
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6855254	6855278
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6855554	6855578
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6855854	6855878
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6856154	6856178
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6856454	6856478
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6856754	6856778
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6857054	6857078
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6857354	6857378
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6857654	6857678
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6857955	6857979
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6858255	6858279
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC*	6858555	6858579
cgr	5p	TGAGAG T GACTGGATCTCGTTCTTC	6858855	6858879
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10545607	10545631
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10545907	10545931
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10546207	10546231
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10887622	10887646
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10887922	10887946
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10888222	10888246
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10888522	10888546
cgr	5p	TGAG A CGGACTGGATCTCGTTCTTC*	10888822	10888846

cgr	5p	TGAGAG C GGACTGGATCTCGTTCTTC*	10889122	10889146
cgr	5p	TGAGAG C GGACTGGATCTCGTTCTTC*	10889422	10889446
cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	15212498	15212522
cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	15212798	15212822
cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	15213098	15213122
cgr	5p	TGAG T CGGACTGGATCTCGTTCTTC*	15213398	15213422
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17833225	17833249
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17833525	17833549
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17833825	17833849
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17834125	17834149
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17834425	17834449
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17834725	17834749
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17835025	17835049
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17835325	17835349
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17835625	17835649
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17835925	17835949
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17836225	17836249
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17836525	17836549
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17836825	17836849
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17837125	17837149
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17837425	17837449
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17837725	17837749
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17838025	17838049
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17838325	17838349
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17838625	17838649
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17838925	17838949
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17839225	17839249
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17839525	17839549
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17839825	17839849
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17840125	17840149
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17840425	17840449
cgr	5p	TGAGAGCGACTGGATCTCGTTCTTC	17840725	17840749
cgr	3p	CTCTGTTCAAC C GC C ACGCGTGACTA*	484514	484538
cgr	3p	CTCTGTTCAAC C GC C ACGCGTGACTA*	484814	484838
cgr	3p	CTCTGTTCAAC C GC C ACGCGTGACTA*	485114	485138
cgr	3p	CTCTGTTCAAC C GCTACACGTGACTA*	485414	485438
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	2203684	2203708
cgr	3p	CTCTGTTCAATGC C ACTCGTGACTA*	10526130	10526154
cgr	3p	CTCTGTTCAATGC C ACTCGTGACTA*	10526430	10526454
cgr	3p	CTCTGTTCAATGC C ACTCGTGACTA*	10526730	10526754
cgr	3p	CTCTGTTCAATGC C ACTCGTGACTA*	10527030	10527054
cgr	3p	CTCTGTTCAATGC C ACTCGTGACTA*	10527330	10527354
cgr	3p	CTCTGTTCAATGC C ACTCGTGACTA*	10527630	10527654
cgr	3p	CTCTGTTCAACGC C ACGCGTGACTA*	15212563	15212587
cgr	3p	CTCTGTTCAACGC C ACGCGTGACTA*	15212863	15212887
cgr	3p	CTCTGTTCAACGC C ACGCGTGACTA*	15213163	15213187
cgr	3p	CTCTGTTCAACGCTAC A CGTGACTA*	15213463	15213487
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17833290	17833314
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17833590	17833614
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17833890	17833914
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17834190	17834214
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17834490	17834514
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17834790	17834814
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17835090	17835114
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17835390	17835414
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17835690	17835714

cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17835990	17836014
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17836290	17836314
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17836590	17836614
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17836890	17836914
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17837190	17837214
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17837490	17837514
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17837790	17837814
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17838090	17838114
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17838390	17838414
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17838690	17838714
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17838990	17839014
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17839290	17839314
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17839590	17839614
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17839890	17839914
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17840190	17840214
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17840490	17840514
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17840790	17840814
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17841090	17841114
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17841390	17841414
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17841690	17841714
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17841990	17842014
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17842290	17842314
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17842590	17842614
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17842890	17842914
cgr	3p	CTCTGTTCAATGCTACGCGTGACTA	17843496	17843520
cgr	3p	CTCTGTTCAAC CGCT ACGCGTGACTA*	18808485	18808509
cgr	3p	CTCTGTTCAAC CGCT ACGCGTGACTA*	18808787	18808811
cgr	3p	CTCTGTTCAAC CGCT ACGCGTGACTA*	18809093	18809117
cgr	3p	CTCTGTTCAAC CTACCGCACGCTGAC *	18809394	18809418
cgr	3p	CTCTGTTCAAC CGCTACGGGGTGACT *	18809701	18809725

=====

_____ Candida albicans SC5314 _____

=====

=====

_____ Candida auris _____

=====

=====

_____ Candida dubliniensis CD3 _____

cdu	3p	CTCTGTTCAAG TTACGGGTGACTAC *	3862063	3862087
-----	----	-------------------------------------	---------	---------

=====

=====

_____ Candida duobushaemulonis _____

=====

=====

_____ Candida haemuloni _____

=====

```

=====
Candida orthopsilosis Co 90-125
=====

=====
Candida oxycetoniae
=====

=====
Candida parapsilosis
=====

=====
Candida pseudohaemulonii
=====

=====
Candida subhashii
=====

=====
Candida tropicalis MYA-3404
=====

=====
Saccharomyces cerevisiae S288c
=====

```

Figura 30. Ejemplo de la salida final del programa del diseño automatizado de oligonucleótidos a ser usados en ensayos de amplificación. La salida final del programa consta de dos secciones: La primera corresponde al resultado de la primera parte del análisis referente a las frecuencias absolutas y posiciones de k-meros repetidos en una secuencia genómica. La segunda parte corresponde a los resultados del análisis considerando las propiedades termodinámicas de la unión de los oligos con sus potenciales secuencias blanco en el DNA genómico que fueron evaluadas con el programa *RNAcofold*. Las secuencias con asterisco representan posiciones en donde las secuencias del oligonucleótido y el DNA no son 100% complementarias (existen “*mismatches*”), pero de acuerdo con el estudio termodinámico, son capaces de hibridar eficientemente. Los *mismatches* son indicados por caracteres en color rojos y resaltados en color amarillo. Se analizan diferentes secuencias genómicas, siendo los correspondientes al genoma de estudio, *Candida glabrata* CBS138, los que primeramente se muestran y posteriormente, los organismos seleccionados como controles negativos, en donde se esperarían una baja o nula frecuencia de regiones reconocidas por los oligos de estudio.

II.3. Organismos de estudio considerados como control negativo

Finalmente, es importante comentar que el programa contempla también el considerar la secuencia genómica de un conjunto de organismos que se consideren controles negativos; es decir, genomas de organismos filogenéticamente relacionados en los que se desea que no exista la posibilidad teórica de la amplificación de productos de PCR dentro del ensayo. Para ello, el programa realiza el análisis termodinámico de cada uno de los pares de oligonucleótidos seleccionados y descarta aquellos pares de oligos con posibilidad de amplificación en el ensayo clínico. Para el caso de *C. glabrata*, los genomas que se usaron como control negativo son los que se muestran en la Tabla 4.

Código	Nombre
cgr	<i>Candida glabrata</i> CBS138
cal	<i>Candida albicans</i> SC5314
caur	<i>Candida auris</i>
cdu	<i>Candida dubliniensis</i> CD36
cduo	<i>Candida duobushaemulonis</i>
cha	<i>Candida haemuloni</i>
cot	<i>Candida orthopsilosis</i> Co 90-125
cox	<i>Candida oxycetoniae</i>
cpa	<i>Candida parapsilosis</i>
cps	<i>Candida pseudohaemulonii</i>
csu	<i>Candida subhashii</i>
ctp	<i>Candida tropicalis</i> MYA-3404
sce	<i>Saccharomyces cerevisiae</i> S288c

Tabla 4. Organismos de estudio considerados como control negativo. Con el propósito de garantizar la ausencia de amplificaciones de DNA de organismos distintos al que se quiere identificar en el ensayo de PCR.

DESCRIPCIÓN DEL PROGRAMA DE ANÁLISIS

El programa desarrollado fue escrito en lenguaje de programación Python considerando una estructura modular constituido por bloques de análisis o subrutinas (Figura 31). La estructura modular con la que se desarrolló nuestro programa permite una organización más sencilla en el diseño, revisión y ejecución de los procesos unitarios de análisis. Por lo anterior, el programa central sólo establece el orden jerárquico en el que serán ejecutadas cada una de las subrutinas. Como podrá observarse, el nombre de las subrutinas describe de manera sucinta, la principal tarea de estas. A continuación, se presenta una breve descripción de ellas:

===== Subrutina **lee_parametros** =====

La primera subrutina se llama **lee_parametros** en la que el usuario puede definir los valores más importantes de las variables que definen el funcionamiento del programa. Entre las principales variables se encuentran:

- Código KEGG del organismo a ser analizado (org. i.e. cgr)
- Tamaño de la ventana de análisis (tamaño_ventana. i.e. 25)
- Número de vecinos para analizar (n_vecinos i.e. 200)
- Longitud máxima entre palabras repetidas (max_dist_satelite i.e. 2500)
- Mínimo tamaño del producto de PCR (min_tamano_amplificacion, i.e. 100)
- Número de bases G o C en extremo 3` (numero_de_GC. i.e. 2)
- Mínimo número de repeticiones (min_frec. i.e. 25)
- Máximo número de palabras repetidas buscar (max_palabras. i.e. 2000)
- Valor del porcentaje permitido entre el valor hibridación del oligo más estable *versus* el valor del oligo menos estable (cutoff_DeltaG_entre_oligos_menor. i.e. 0.85)
- Valor de auto-complementariedad permitido respecto a valor de hibridación (cutoff_complementariedad. i.e. 0.60)
- Valor del porcentaje mínimo de la TM de hibridación respecto a la TM máxima posible (cutoff_hibridacion. i.e. 0.85).

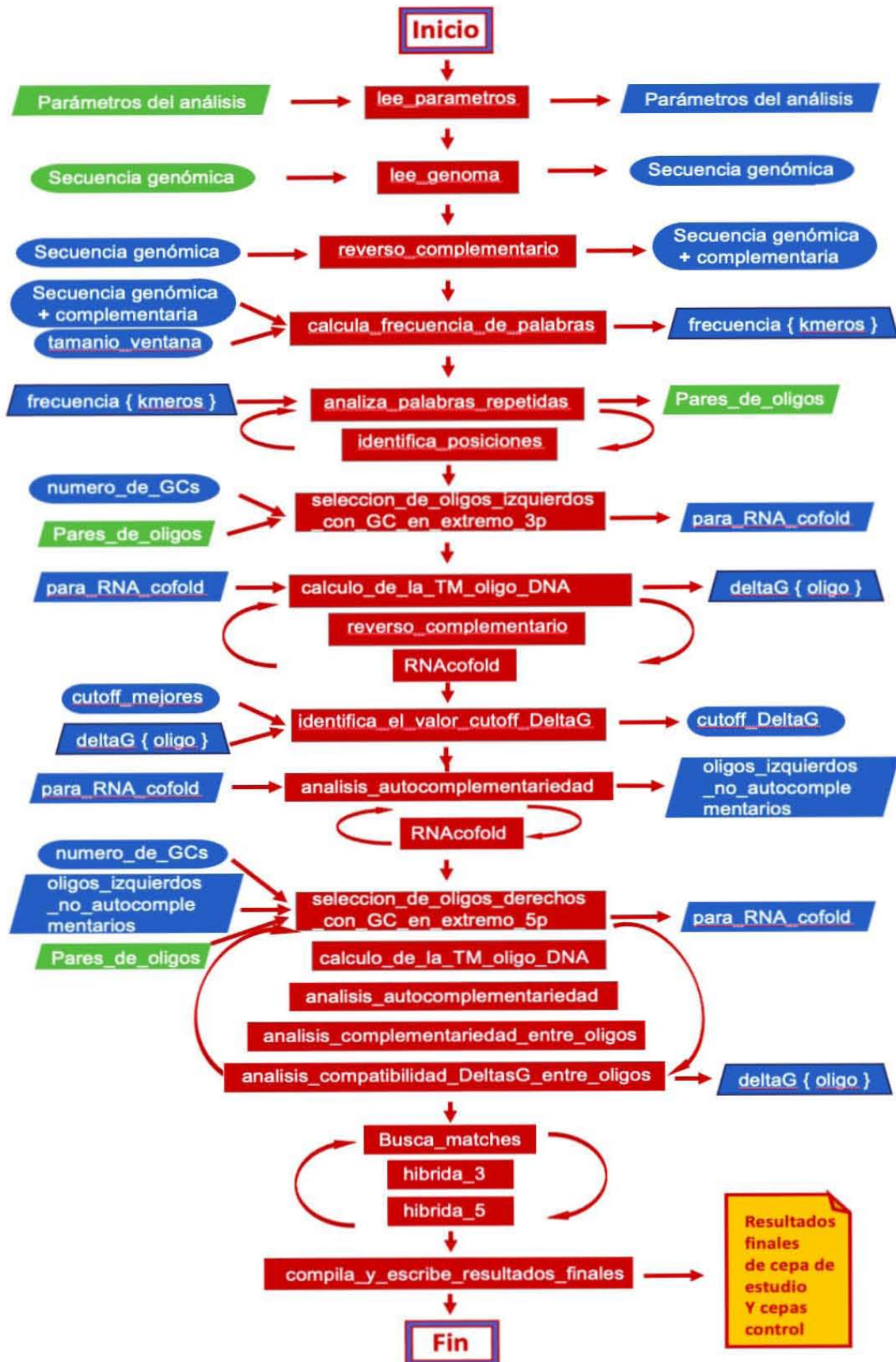


Figura 31. Diagrama de flujo para la identificación de K-meros frecuentes en secuencias genómicas y el diseño de oligonucleótidos considerando el análisis termodinámico de sus propiedades. El diagrama representa a los elementos centrales de análisis para la evaluación de las frecuencias absolutas y posiciones de k-meros repetidos en una secuencia genómica. Los rectángulos con fondo rojo representan las principales subrutinas. En el lado izquierdo se muestran los elementos de entrada a las subrutinas y del derecho, los elementos de salida. En verde se muestran la información almacenada en disco, mientras que los elementos en azul los almacenados en la memoria RAM. Los óvalos representan variables con un sólo elemento, los romboides representan listas con varios elementos y los trapecios representan a directorios, en donde los valores están asociados a valores de referencia o llaves. El círculo final de color naranja representa los resultados finales del programa.

===== Subrutina **lee_genoma** =====

Esta subrutina lee las secuencias de DNA de los diferentes cromosomas del genoma de estudio. Para facilidad en el análisis, las secuencias de los diferentes cromosomas son concatenadas en una sola variable (secuencia_genomica). Se obtiene también la secuencia inversa complementaria de la hebra de DNA reportada como secuencia genómica utilizando la subrutina **reverso_complementario**. Tanto la secuencia leída originalmente, como su correspondiente secuencia inversa complementaria, son concatenadas y almacenadas en la variable secuencia_genomica de tal manera que su longitud es el doble del número de nucleótidos del genoma de estudio. Como se verá más tarde, una vez identificados los primeros oligonucleótidos candidatos, la ubicación precisa de los mismos dentro de cada uno de los cromosomas y sus correspondientes direcciones relativas a las hebras del DNA (sentido o antisentido), será realizada por las subrutinas

**seleccion_de_oligos_izquierdos_con_GC_en_extremo_3p y
seleccion_de_oligos_derechos_con_GC_en_extremo_5p.**

===== Subrutina **reverso_complementario** =====

Como su nombre lo indica, en esta subrutina se obtiene la secuencia inversa y complementaria de una secuencia nucleotídica en cuestión.

===== Subrutina **calcula_frecuencia_de_palabras** =====

En esta subrutina se cuentan la frecuencia de todos los “k-meros” de longitud definida por la variable `tamano_ventana` de la secuencia genómica de estudio. Para realizar este análisis, el programa considera todos los k-meros que existen en la secuencia genómica empezando en la posición cero. Esta subsecuencia o k-mero es usada como llave del diccionario llamado `frecuencia_de_kmeros` para incrementar en una unidad el valor previamente almacenado en esa posición. Posteriormente, la ventana de análisis se mueve una posición a la derecha, o posición 1, y se repite el proceso. Esta operación se ejecuta reiterativamente un número de veces igual al de la longitud de la `secuencia_genomica` menos la longitud del `tamano_ventana`. Cabe señalar que las palabras con nucleótidos no definidos marcados como “N” en la secuencia, no son consideradas en el análisis.

===== Subrutina **analiza_palabras_repetidas** =====

En esta subrutina ejecuta un proceso cíclico en el que se analiza las palabras cuya frecuencia absoluta en el genoma es igual o mayor a la de un valor mínimo pre-establecido en la variable `min_frec`, iniciando con las palabras con mayor frecuencia. Las palabras con la misma frecuencia son agrupadas como un elemento del vector llamado `frecuencias`. El número de palabras que serán analizadas en un análisis posterior está acotado por el valor definido en la variable `max_palabras`. En cada ciclo de análisis de esta subrutina se analiza la frecuencia de los vecinos de las palabras más repetidas para diferentes distancias, desde cero, hasta una distancia definida en la variable `max_dist_satelite`. Se analizan las posiciones de las palabras que cumplan con los límites establecidos en las variables `min_frec` y `max_palabras` dentro de la secuencia genómica almacenada en la variable `secuencia_genomica`. Las posiciones de las palabras dentro de cada ciclo se identifican al

llamar a la subrutina **identifica_posiciones**. El resultado de esta subrutina es almacenado como archivos dentro del directorio Pares_de_oligos.

===== Subrutina **identifica_posiciones** =====

Como su nombre lo indica, este bloque de análisis identifica las posiciones de una palabra específica en la secuencia genómica y en la secuencia complementaria-inversa, que fue almacenada en la variable `secuencia_genomica`. Dichas posiciones son identificadas utilizando la función `find`. Para cada posición de la palabra, se almacenan las secuencias vecinas en el vector llamado `vecinos` de acuerdo con el número de nucleótidos que existen entre ellas. Las secuencias vecinas que se encuentran con cero espacios de separación entre ellas y la palabra analizada se almacenan en la posición cero del vector `vecinos`. Las secuencias vecinas que se encuentran con un nucleótido de separación entre ellas y la palabra analizada se almacenan en la posición uno del vector `vecinos`, y así sucesivamente, hasta un número de veces igual al valor resultante de la resta: `max_dist_satelite - tamano_ventana`.

== Subrutina **seleccion_de_oligos_izquierdos_con_GC_en_extremo_3p** ==

Para garantizar que los oligonucleótidos tengan buena capacidad de iniciar el proceso de síntesis del DNA que lleva a cabo la DNA polimerasa, esta subrutina elimina todas aquellas palabras cuyo extremo 3' no sean nucleótidos de tipo "G" o "C". El número de nucleótidos G/C que por defecto se exige que tengan las palabras en el extremo 3' es uno, pero este valor puede ser modificado al cambiar el valor de la variable `numero_de_GC`s. Las palabras que satisfacen el criterio del número de nucleótidos G/C se almacenan en el vector llamado `para_RNAcofold`. Posterior al análisis de **RNAcofold**, los oligos izquierdos que satisfacen el criterio de hibridización estable de acuerdo a su valor de mínima energía libre de hibridización (`mfe`) y por no ser auto-complementarios, se almacenan en el diccionario `oligos_izquierdos_no_autocomplementarios`.

=== Subrutina **calculo_de_la_TM_oligo_DNA** ===

Todos los cálculos para evaluar las TMs (por sus siglas en inglés, Melting temperatura o Temperatura de fusión) de un oligonucleótido con el DNA, o entre oligonucleótidos, se realizan utilizando el programa **RNAcofold** del paquete de análisis Infernal ViennaRNA Package 2.0 (Lorenz *et al.* 2011).

El programa analiza tres posibles valores de hibridación:

- 1) La del oligo con el DNA genómico considerando una completa identidad, evaluada en la subrutina **calculo_de_la_TM_oligo_DNA**,
- 2) La de los oligos consigo mismos, para descartar aquellos oligos que, por su naturaleza palindrómica, puedan ser parcialmente complementarios , evaluada en la subrutina **analisis_autocomplementariedad**, y
- 3) La complementariedad entre los pares de oligos empleados en la reacción de PCR para descartar posibles interacciones entre ellos que desfavorezcan la reacción de amplificación, evaluada en la subrutina **analisis_complementariedad_entre_oligos**.

===== Subrutina **identifica_el_valor_cutoff_DeltaG** =====

Con el propósito de establecer un valor de corte de hibridación cutoff_DeltaG a partir del cual, se considerará que un oligo puede unirse de manera estable a su correspondiente región nucleotídica, se ordenan todos los oligos provenientes de la subrutina **calculo_de_la_TM_oligo_DNA** que fueron almacenados en el diccionario deltaG usando como llave a la variable oligo, considerando sus correspondientes valores de hibridación. El valor cutoff_DeltaG corresponderá al valor obtenido por el oligo colocado en la posición que resulta de multiplicar la variable cutoff_mejores por el número de oligos del directorio deltaG. Por defecto, el valor cutoff_mejores fue considerado de 0.25, lo que equivaldría al último elemento del primer cuartil del diccionario deltaG.

===== Subrutina **analisis_autocomplementariedad** =====

Esta subrutina permite identificar oligonucleótidos que pudieran reconocerse parcialmente debido su naturaleza palindrómica y que en consecuencia reduce su concentración efectiva en las reacciones de PCR. Al igual que en la subrutina anterior, el cálculo de la TM de esta subrutina fue realizada utilizando el programa **RNAcofold** del paquete ViennaRNA Package 2.0 (Lorenz *et al.* 2011). usando como secuencias de entrada la secuencia de cada oligo de manera duplicada.

=== Subrutina **seleccion_de_oligos_derechos_con_GC_en_extremo_5p** ==

En esta subrutina se realiza un proceso similar al realizado en la subrutina **seleccion_de_oligos_izquierdos_con_GC_en_extremo_3p**, con la diferencia de que los oligos analizados, son los pares derechos de los oligos anteriormente identificados como izquierdos y almacenados en el directorio **oligos_izquierdos_no_autocomplementarios**. Dichos pares fueron seleccionados previamente en la subrutina **analiza_palabras_repetidas** de acuerdo con su frecuencia absoluta y distancia a los oligos izquierdos en cuestión y almacenados en disco como archivos del directorio **Pares_de_oligos**.

Como puede verse en el diagrama de flujo del programa, esta subrutina llama en orden sucesivo a las subrutinas:

- a) **calculo_de_la_TM_oligo_DNA**
- b) **analisis_autocomplementariedad**
- c) **analisis_complementariedad_entre_oligos**
- d) **analisis_compatibilidad_DeltasG_entre_oligos**

===== Subrutina **analisis_complementariedad_entre_oligos** =====

Esta subrutina permite identificar aquellos pares de oligonucleótidos que se consideran utilizar en la reacción de amplificación del DNA, que son mutuamente complementarios. En este caso, las secuencias que se dan como valores de entrada al programa **RNAcofold** (Lorenz *et al.* 2011) son las del par de oligos en cuestión.

=== Subrutina **analisis_compatibilidad_DeltasG_entre_oligos** ===

Para garantizar una eficiente amplificación de los productos de PCR que se emplearán en el proceso de detección de *loci* específicos de un organismo, es conveniente que el par de oligos usados en cada reacción tengan una temperatura de alineamiento T_M similar. La relación entre temperaturas de alineamiento de los oligos diseñados deberá tener un valor igual o superior al especificado en la variable `cutoff_DeltaG_entre_oligos`.

===== Subrutina **busca_matches** =====

La primera parte del programa, todos los análisis de frecuencias e identificación de “palabras” o “k-meros” más frecuentes en un genoma fue realizado en términos de “hits” perfectos de secuencias idénticas en la secuencia genómica del organismo de estudio. Esta simplificación de considerar sólo a las secuencias idénticas facilita enormemente el análisis inicial y reduce significativamente el tiempo requerido de cómputo. Los “k-meros” más frecuentes fueron posteriormente analizados en términos de sus correspondientes valores de mínima energía libre de alineamiento (mfe), y de no ser auto-complementarios. En la segunda parte del programa, se realiza un análisis posterior en la que los pares de “k-meros” más frecuentes que satisficieron los criterios de longitud, nucleótidos G/C en extremos 5', etc., son caracterizados termodinámicamente por su capacidad de servir como oligonucleótidos para ser empleados en amplificaciones de DNA utilizando la técnica de amplificación mediante la reacción en cadena de la polimerasa, mejor conocida como PCR, por sus siglas en inglés. Dentro de esta subrutina se llama de manera recursiva a las subrutinas **hibrida_5p** y **hibrida_3p**.

===== Subrutina **hibrida_5p** =====

Esta subrutina tiene como propósito identificar las posiciones en donde un oligonucleótido pudiera hibridar en la secuencia genómica considerando una temperatura de alineamiento en la reacción de la amplificación por PCR. Es importante considerar que el número de sitios que puede reconocer un oligonucleótido particular podrá ser mucho mayor al originalmente evaluado en las primeras etapas de análisis del programa que evalúan la repetición de k-meros no degenerados en las secuencias genómicas. Con el propósito de reducir significativamente el tiempo requerido para el cómputo de los potenciales sitios de hibridación de un oligonucleótido, el programa consulta un archivo con los índices de las posiciones en donde se encuentran todos los 256 k-meros posibles de longitud 4, dentro del genoma ($4 \times 4 \times 4 \times 4 = 256$). Para cada posición registrada en el directorio de índices, el programa obtiene la secuencia del genoma de longitud tamaño_ventana que corresponde al potencial sitio de unión del oligo. El valor de la mínima energía libre de hibridación (mfe) entre el oligo y el potencial sitio de unión en el genoma, es calculado usando el programa **RNAcofold** (Lorenz *et al.* 2011)

===== Subrutina **hibrida_3p** =====

Esta subrutina es similar a la de la subrutina **hibrida_5p** con la única diferencia de que las posiciones del oligonucleótido a ser identificadas corresponden a las secuencias inversas complementarias de las palabras vecinas al oligonucleótido considerado como izquierdo o 5p.

===== Programa **RNAcofold** =====

Es uno de los programas pertenecientes a la librería de análisis de la librería ViennaRNA 2.0 que permite evaluar las potenciales estructuras secundarias de moléculas de DNA o RNA de cadena sencilla posterior a su formación como moléculas diméricas y sus correspondientes valores de mínima energía libre (mfe), así como la función de partición (pf) y la matriz de probabilidad de emparejamiento de bases (Lorenz *et al.* 2011).

===== Subrutina **compila_y_escribe_resultados_finales** =====

En esta subrutina se escriben los archivos de resultados finales del análisis que incluyen los diferentes pares de oligonucleótidos candidatos a ser usados en las reacciones de PCR y sus potenciales sitios de hibridación. El análisis de hibridación *in silico*, es realizado tanto para el genoma de estudio, como para todos los genomas considerados como controles negativos de organismos filogenéticamente cercanos. En nuestro diagnóstico inicial se eligió al genoma del hongo *Candida glabrata* CBS138 como genoma de estudio. Nuestro estudio consideró como control negativo a las secuencias genómicas de los 35 hongos listados en la Tabla 3.

II.5. Ensayos de amplificación de productos de PCR específicos de *C. glabrata*

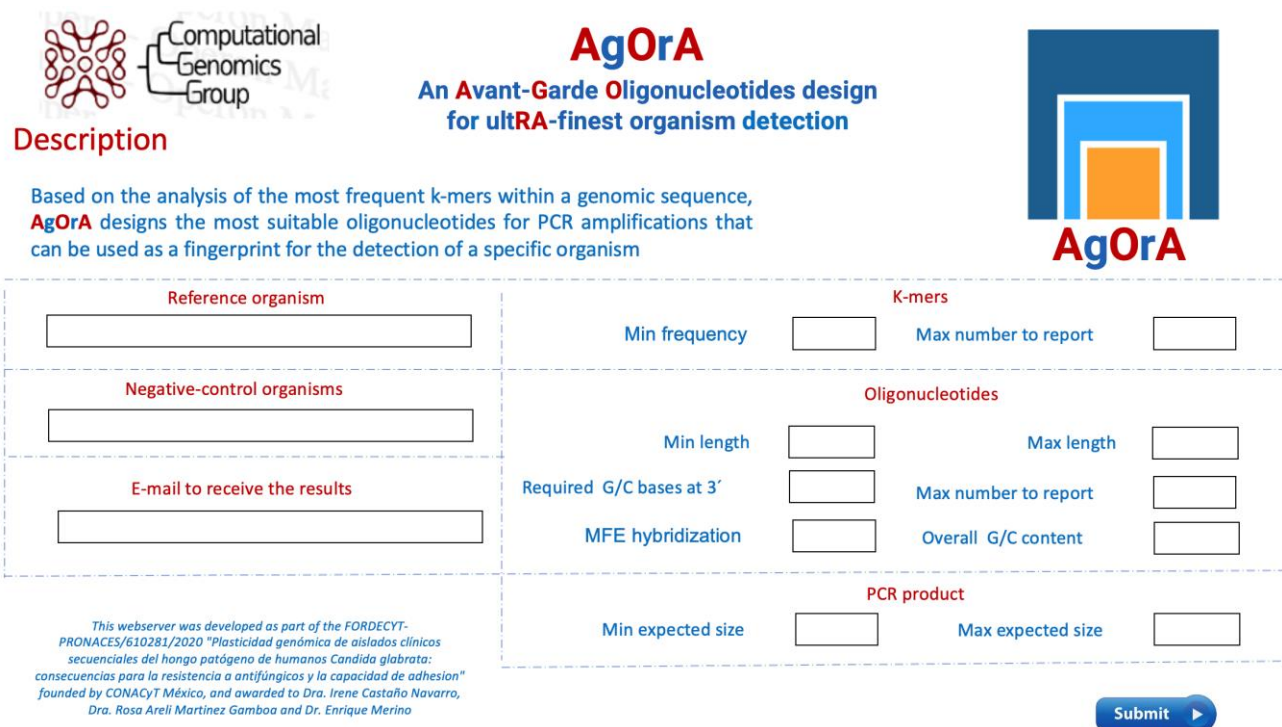
Con el propósito de validar experimentalmente el diseño de oligos generados con nuestro programa, el Dr. Marco Josué Hernández Chávez, miembro del laboratorio de la Dra. Irene Castaño del IPICYT, realizó ensayos de amplificación de productos de PCR utilizando como templado el DNA genómico de *C. glabrata* y otras especies filogenéticamente relacionadas. Se utilizaron tres pares de oligonucleótidos y en todos los ensayos realizados se obtuvo amplificación exclusivamente en las reacciones en las que se usó el DNA genómico de *C. glabrata* como templado, demostrando que el diseño computacional de oligos para la detección de genomas de interés permite obtener resultados con gran especificidad y sensibilidad (Ver apéndice # 1)

CONCLUSIONES

- Se ha logrado establecer una plataforma con un importante impacto en el diagnóstico de cepas de *C. glabrata*, pero, sobre todo, podrá ser útil en el diagnóstico de otros patógenos.
- La frecuencia absoluta de k-meros es una condición necesaria pero no suficiente para garantizar el diseño óptimo de los oligonucleótidos.
- Se ha realizado el análisis de la frecuencia de k-meros de diferentes tamaños en el genoma de *C. glabrata*, evaluando la especificidad de los k-meros frecuentes de *C. glabrata* en relación con 95 genomas de phylum Ascomycota; así también se ha analizado la posición relativa de los k-meros identificados y su relación con las regiones codificantes, intergénicas, así como su cercanía a regiones teloméricas.
- Dentro de los 98 organismos que se analizaron, existen coincidencias en un gran número de secuencias comparativamente a las existentes en *C. glabrata*. Sin embargo, hay presentes en esta especie un número considerable de micro y minisatélites que ayudaron a identificar la especie de manera específica.
- Los experimentos de amplificación realizados por el grupo de la Dra. Irene Castaño del IPICYT, demuestran que la estrategia para el diseño de oligos específicos de *C. glabrata* permite obtener ensayos de identificación de *C. glabrata* con una alta especificidad y sensibilidad. Siendo una herramienta potencial en el futuro para la identificación de otros microorganismos.

PERSPECTIVAS

1) Como una perspectiva inmediata a este trabajo se contempla la elaboración de una página web que permita el diseño de oligonucleótidos específicos para cualquier organismo de interés de cualquier usuario, de tal forma que pueda ser útil, no sólo en la identificación de *Candida glabrata*, si no de otros organismos patógenos o de interés particular. Se muestra el maquetado de dicha página que se está realizando en el grupo de Genómica Computacional, del Instituto de Biotecnología-UNAM.



The image shows a web interface for 'AgOrA: An Avant-Garde Oligonucleotides design for ultra-fine organism detection'. The interface includes a 'Description' section, a 'Reference organism' input field, a 'Negative-control organisms' input field, and an 'E-mail to receive the results' input field. It also features a 'K-mers' section with 'Min frequency' and 'Max number to report' input fields, and an 'Oligonucleotides' section with 'Min length', 'Max length', 'Required G/C bases at 3'', 'MFE hybridization', 'Max number to report', and 'Overall G/C content' input fields. A 'PCR product' section at the bottom has 'Min expected size' and 'Max expected size' input fields. A 'Submit' button is located at the bottom right. The interface is branded with the 'AgOrA' logo and the 'Computational Genomics Group' logo.

Computational Genomics Group

AgOrA
An Avant-Garde Oligonucleotides design
for ultra-fine organism detection

Description

Based on the analysis of the most frequent k-mers within a genomic sequence, **AgOrA** designs the most suitable oligonucleotides for PCR amplifications that can be used as a fingerprint for the detection of a specific organism

Reference organism

Negative-control organisms

E-mail to receive the results

K-mers

Min frequency Max number to report

Oligonucleotides

Min length Max length

Required G/C bases at 3' Max number to report

MFE hybridization Overall G/C content

PCR product

Min expected size Max expected size

Submit

This webserver was developed as part of the FORDECYT-PRONACES/610281/2020 "Plasticidad genómica de aislados clínicos secuenciales del hongo patógeno de humanos *Candida glabrata*: consecuencias para la resistencia a antifúngicos y la capacidad de adhesión" founded by CONACyT México, and awarded to Dra. Irene Castaño Navarro, Dra. Rosa Areli Martínez Gamboa and Dr. Enrique Merino

Figura 32. Página web para el diseño óptimo de oligonucleótidos a ser usados en ensayos de identificación de organismos específicos. La figura muestra un ejemplo del tipo de portal web considerado como la primera perspectiva a ser realizada próximamente.

2) Se pueden realizar modificaciones al código del programa para que, en lugar de realizar la identificación de un sólo organismo, se pueda contemplar la identificación de varios organismos simultáneamente. Por ejemplo, que permita el diseño óptimo de oligos para identificar a *Candida glabrata* y a *Candida albicans*.

**Apéndice 1. Resultados de ensayos de amplificación de productos de PCR
específicos de *C. glabrata* realizados por el Dr. Marco Josué Hernández
Chávez del laboratorio de la Dra. Irene Castaño del IPICYT**

Tabla 5. Lista de oligonucleótidos elegidos en el ensayo de identificación de *C. glabrata* mediante la amplificación por PCR

	No. Lab EMP	NAME	Seq 5' to 3'	Tm
1	5382	oli x2x35x sense	GGT GGT GAT CGT GGT AGG CTT GCC G	74
2	5383	Oli 2x 35x antisense	GAC TTC GAG ACA GAC CTT GTT TCG C	69
3	5384	oli 9x 35x sense	TGT TGA GAG CGA CTG GAT CTC GTT C	69
4	5385	oli 9x 35x antisense	AAT CTC TCC CTT CTC ATC AAC AGT G	67
15	5396	control 1x sense	AGG CTG AAA AGG CTA AGC TAG CTG C	69
16	5397	control 1x antisense	CGT CAA GAG CAC GAT TAA GGT CTC C	69

Tabla 6. Condiciones para las reacciones de PCR en el ensayo de identificación de *C. glabrata* usando los pares de oligos 1-2, 3-4 así como 15-16.

dNTPs concentration: 0.2 mM each 10 ng of DNA template Oligonucleotide concentration: 0.5 µM each			
Initial denaturation	95°C	3 minutes	40 ciclos
Denaturation	95°C	30 seconds	
Annealing	69°C	30 seconds	
Extension	72 °C	1 minute	
Final extension	72 °C	5 minute	
Final Temperature	12 °C	5 minutes	

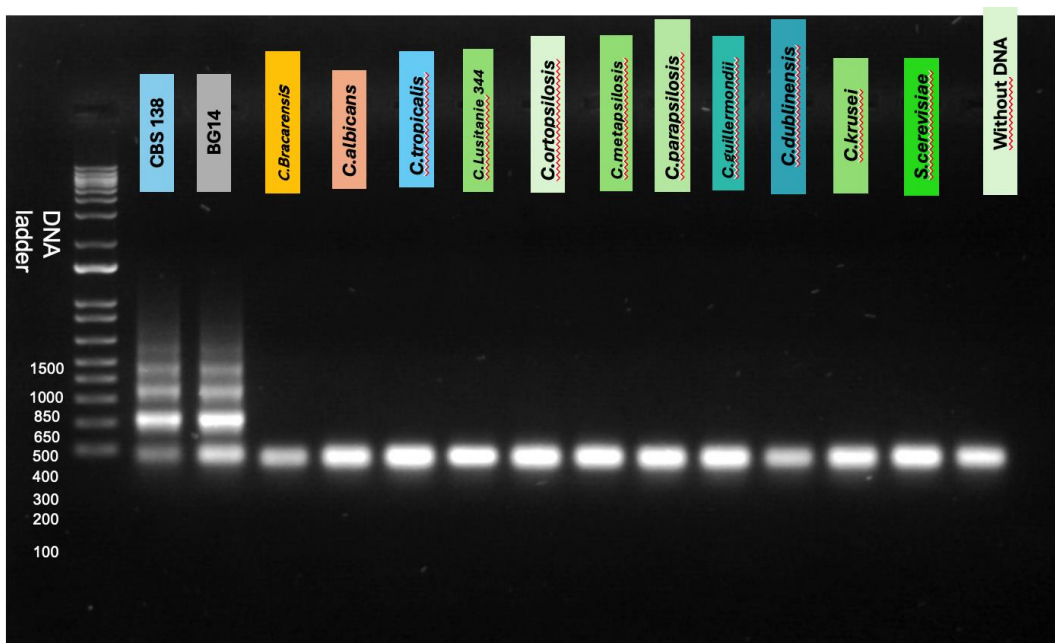


Figura 33. Productos de la reacción de PCR utilizando los oligos # 1 y # 2 utilizando DNA genómico de *C. glabrata* y especies relacionadas. Solo se obtiene amplificación con el DNA genómico de *C. glabrata* y no el de otras especies de *Candida* o especies relacionadas.

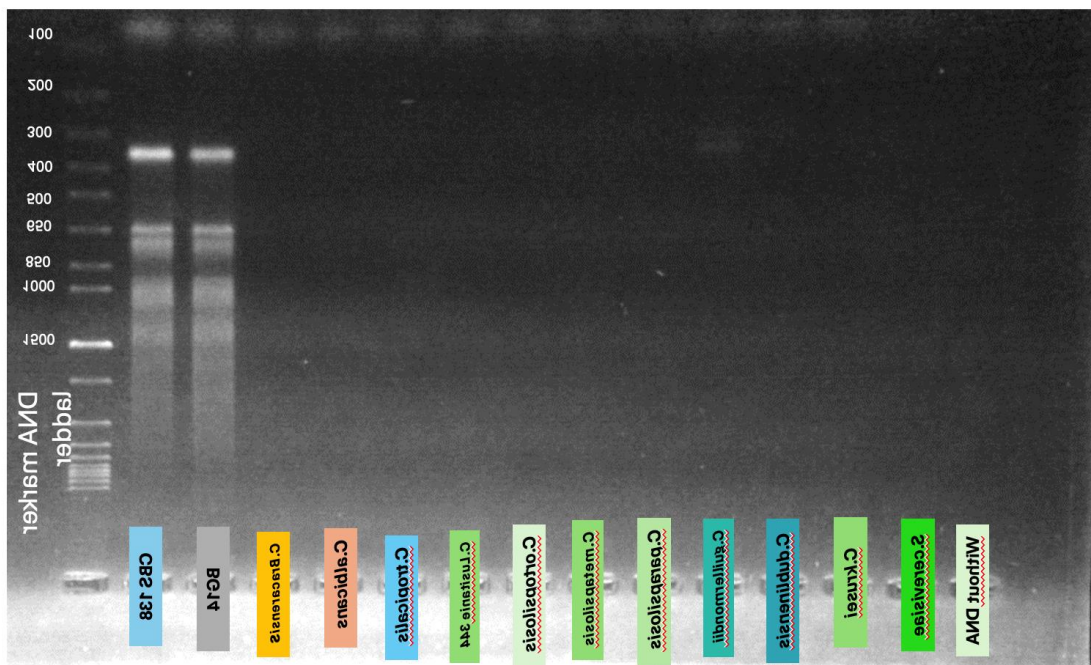


Figura 34. Productos de la reacción de PCR utilizando los oligos # 3 y # 4 utilizando DNA genómico de *C. glabrata* y especies relacionadas. Solo se obtiene amplificación con el DNA genómico de *C. glabrata* y no el de otras especies de *Candida* o especies relacionadas.

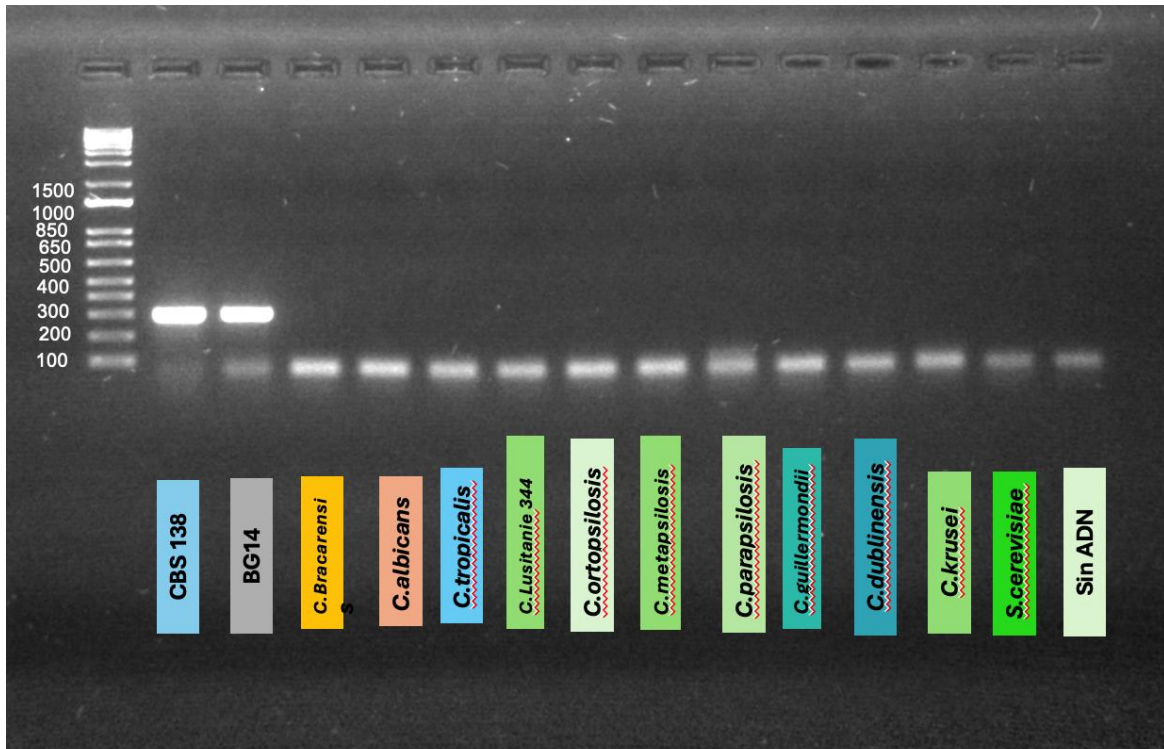


Figura 35. Productos de la reacción de PCR utilizando los oligos # 15 y # 16 utilizando DNA genómico de *C. glabrata* y especies relacionadas. Solo se obtiene amplificación con el DNA genómico de *C. glabrata* y no el de otras especies de *Candida* o especies relacionadas.

Apéndice 2. Lista de organismos considerados en el estudio.

aalt

Alternaria alternata SRC1lrK2f

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Pleosporomycetidae,-,
,Pleosporales,Pleosporineae,-,-,Pleosporaceae,-,-,-,Alternaria,-,Alternaria alternata,-,-

abe

Trichophyton benhamiae CBS 112371 (*Arthroderma benhamiae* CBS 112371)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Onygenales,-,-
,-,Arthrodermataceae,-,-,-,Trichophyton,-,Trichophyton benhamiae,-,-

abp

Agaricus bisporus var. *burnettii* JB137-S8

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,
,Agaricales,-,-,-,Agaricaceae,-,-,-,Agaricus,-,Agaricus bisporus,-,-

abv

Agaricus bisporus var. *bisporus* H97

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,
,Agaricales,-,-,-,Agaricaceae,-,-,-,Agaricus,-,Agaricus bisporus,-,-

act

Aspergillus clavatus NRRL 1

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,-
,Aspergillaceae,-,-,-,Aspergillus,Aspergillus subgen. *Fumigati*,Aspergillus clavatus,-,-

adl

Auricularia subglabra TFB-10046 SS5

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,-,Auriculariales,-,-,-
,Auriculariaceae,-,-,-,Auricularia,-,Auricularia subglabra,-,-

afm

Aspergillus fumigatus Af293

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,-
,Aspergillaceae,-,-,-,Aspergillus,Aspergillus subgen. *Fumigati*,Aspergillus fumigatus,-,-

afv

Aspergillus flavus NRRL3357

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,-
,Aspergillaceae,-,-,-,Aspergillus,Aspergillus subgen. *Circumdati*,Aspergillus flavus,-,-

ago

Ashbya gossypii (*Eremothecium gossypii*) ATCC 10895

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,-,Saccharomycetales,-,-,-
,Saccharomycetaceae,-,-,-,Eremothecium,-,Eremothecium gossypii,-,-

aje

Histoplasma capsulatum NAm1

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Onygenales,-,-
,-,Ajellomycetaceae,-,-,-,Histoplasma,-,Histoplasma capsulatum,-,-

ang

Aspergillus niger CBS 513.88

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,-
,Aspergillaceae,-,-,-,Aspergillus,Aspergillus subgen. *Circumdati*,Aspergillus niger,-,-

ani

Aspergillus nidulans FGSC A4

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,Eurotiales,-,-,
,Aspergillaceae,-,-,Aspergillus,Aspergillus subgen. Nidulantes,Aspergillus nidulans,-,-

aor

Aspergillus oryzae RIB40

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,Eurotiales,-,-,
,Aspergillaceae,-,-,Aspergillus,Aspergillus subgen. Circumdati,Aspergillus oryzae,-,-

bcom

Baudoinia panamericana UAMH 10762 (Baudoinia compniacensis UAMH 10762)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Dothideomycetidae,-,-
,Mycosphaerellales,-,-,Teratosphaeriaceae,-,-,Baudoinia,-,Baudoinia panamericana,-,-

bfu

Botrytis cinerea B05.10

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Leotiomyces,-,-,Helotiales,-,-,Sclerotiniaceae,-,-
,-,Botrytis,-,Botrytis cinerea,-,-

bgh

Blastomyces gilchristii SLH14081

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,Onygenales,-,-
,-,Ajellomycetaceae,-,-,Blastomyces,-,Blastomyces gilchristii,-,-

bnn

Brettanomyces nanus CBS 1945

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Pichiaceae,-,-,Brettanomyces,-,Brettanomyces nanus,-,-

bor

Bipolaris oryzae ATCC 44560

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Pleosporomycetidae,-,-
,Pleosporales,Pleosporineae,-,-,Pleosporaceae,-,-,Bipolaris,-,Bipolaris oryzae,-,-

bsc

Bipolaris sorokiniana ND90Pr

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Pleosporomycetidae,-,-
,Pleosporales,Pleosporineae,-,-,Pleosporaceae,-,-,Bipolaris,-,Bipolaris sorokiniana,-,-

bze

Bipolaris zeicola 26-R-13

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Pleosporomycetidae,-,-
,Pleosporales,Pleosporineae,-,-,Pleosporaceae,-,-,Bipolaris,-,Bipolaris zeicola,-,-

cal

Candida albicans SC5314

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Candida,-,Candida albicans,-,-

caur

Candida auris

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Metschnikowiaceae,-,-,Clavispora,-,[Candida] auris,-,-

Coprinopsis cinerea okayama7#130

Candida dubliniensis CD36

Colletotrichum fioriniae PJ7

Cryptococcus gattii WM276

Candida glabrata CBS138

Coccidioides immitis RS

.....

Clavispora lusitaniae ATCC 42720

Clavispora lusitaniae CBS 6936

Cordyceps militaris CM01

Cryptococcus neoformans var. *neoformans* B-3501A

.....

Cryptococcus neoformans var. *neoformans* JEC21 (*Filobasidiella neoformans* JEC21)

96

Candida orthopsilosis Co 90-125

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Candida,-,Candida orthopsilosis,-,-

cput

Coniophora puteana RWD-64-598 SS2

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,-
,Boletales,Coniophorineae,-,-,Coniophoraceae,-,-,Coniophora,-,Coniophora puteana,-,-

cpw

Coccidioides posadasii C735 delta SOWgp

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Onygenales,-,-
,,-,-,-,Coccidioides,-,Coccidioides posadasii,-,-

cten

Yamadazyma tenuis ATCC 10573 (Candida tenuis ATCC 10573)

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Yamadazyma,-,Yamadazyma tenuis,-,-

cthr

Chaetomium thermophilum var. thermophilum DSM 1495

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Sordariales,-
,,-,Chaetomiaceae,-,-,Chaetomium,-,Chaetomium thermophilum,-,-

ctp

Candida tropicalis MYA-3404

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Candida,-,Candida tropicalis,-,-

dha

Debaryomyces hansenii CBS767

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Debaryomyces,-,Debaryomyces hansenii,-,-

dsq

Dichomitus squalens LYAD-421 SS1

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Polyporales,-,-,
,Polyporaceae,-,-,Dichomitus,-,Dichomitus squalens,-,-

ecu

Encephalitozoon cuniculi GB-M1

Eukaryota,Fungi,-,Microsporidia,-,-,-,-,-,Apansporoblastina,-,-,Unikaryonidae,-,-,Encephalitozoon,-
,Encephalitozoon cuniculi,-,-

ehe

Encephalitozoon hellem ATCC 50504

Eukaryota,Fungi,-,Microsporidia,-,-,-,-,-,Apansporoblastina,-,-,Unikaryonidae,-,-,Encephalitozoon,-
,Encephalitozoon hellem,-,-

ein

Encephalitozoon intestinalis ATCC 50506

Eukaryota,Fungi,-,Microsporidia,-,-,-,-,-,Apansporoblastina,-,-,Unikaryonidae,-,-,Encephalitozoon,-
,Encephalitozoon intestinalis,-,-

ela

Eutypa lata UCREL1

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Xylariomycetidae,-,Xylariales,-,-,
,Diatrypaceae,-,-,Eutypa,-,Eutypa lata,-,-

erc

Eremothecium cymbalariae DBVPG#7215

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Eremothecium,-,Eremothecium cymbalariae,-,-

ero

Encephalitozoon romaleae SJ-2008

Eukaryota,Fungi,-,Microsporidia,-,-,-,-,Apansporoblastina,-,-,Unikaryonidae,-,-,Encephalitozoon,-
,Encephalitozoon romaleae,-,-

fgr

Fusarium graminearum (Gibberella zeae PH-1)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-,
,Hypocreales,-,-,Nectriaceae,-,-,Fusarium,-,Fusarium graminearum,-,-

fme

Fomitiporia mediterranea MF3/22

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Hymenochaetales,-,-,
,Hymenochaetaceae,-,-,Fomitiporia,-,Fomitiporia mediterranea,-,-

fox

Fusarium oxysporum f. sp. lycopersici 4287

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-,
,Hypocreales,-,-,Nectriaceae,-,-,Fusarium,-,Fusarium oxysporum,-,-

fpu

Fusarium pseudograminearum CS3096

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-,
,Hypocreales,-,-,Nectriaceae,-,-,Fusarium,-,Fusarium pseudograminearum,-,-

fvr

Fusarium verticillioides 7600

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-,
,Hypocreales,-,-,Nectriaceae,-,-,Fusarium,-,Fusarium verticillioides,-,-

glz

Glarea lozoyensis ATCC 20868

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Leotiomycetes,-,-,Helotiales,-,-,Helotiaceae,-,-,
,Glarea,-,Glarea lozoyensis,-,-

gtr

Gloeophyllum trabeum ATCC 11539

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Gloeophyllales,-,-,
,Gloeophyllaceae,-,-,Gloeophyllum,-,Gloeophyllum trabeum,-,-

hir

Heterobasidion irregulare TC 32-1

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Russulales,-,-,
,Bondarzewiaceae,-,-,Heterobasidion,-,Heterobasidion irregulare,-,-

kaf

Kazachstania africana CBS 2517

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Kazachstania,-,Kazachstania africana,-,-

kla

Kluyveromyces lactis NRRL Y-1140

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Kluyveromyces,-,Kluyveromyces lactis,-,-

kmx

Kluyveromyces marxianus DMKU3-1042

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Kluyveromyces,-,Kluyveromyces marxianus,-,-

lbc

Laccaria bicolor S238N-H82

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,-
,Agaricales,-,-,Tricholomataceae,-,-,Laccaria,-,Laccaria bicolor,-,-

lel

Lodderomyces elongisporus NRRL YB-4239

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Lodderomyces,-,Lodderomyces elongisporus,-,-

lth

Lachancea thermotolerans CBS 6340 (Kluyveromyces thermotolerans CBS 6340)

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Lachancea,-,Lachancea thermotolerans,-,-

maj

Metarhizium robertsii ARSEF 23

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-
,Hypocreales,-,-,Clavicipitaceae,-,-,Metarhizium,-,Metarhizium robertsii,-,-

maw

Metarhizium acridum CQMa 102

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-
,Hypocreales,-,-,Clavicipitaceae,-,-,Metarhizium,-,Metarhizium acridum,-,-

mbe

Marssonina brunnea f. sp. 'multigermtubi' MB_m1

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Leotiomycetes,-,-,Helotiales,-,-,Dermateaceae,-,-
,Marssonina,-,Marssonina brunnea,-,-

mgl

Malassezia globosa CBS 7966

Eukaryota,Fungi,Dikarya,Basidiomycota,Ustilaginomycotina,-,Malasseziomycetes,-,-,Malasseziales,-,-,
,Malasseziaceae,-,-,Malassezia,-,Malassezia globosa,-,-

mgr

Pyricularia oryzae 70-15 (Magnaporthe oryzae 70-15)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-
,Magnaporthales,-,-,Pyriculariaceae,-,-,Pyricularia,-,Pyricularia oryzae,-,-

mlr

Melampsora larici-populina 98AG31

Eukaryota,Fungi,Dikarya,Basidiomycota,Pucciniomycotina,-,Pucciniomycetes,-,-,Pucciniales,-,-,
,Melampsoraceae,-,-,Melampsora,-,Melampsora larici,-,-

mpr

Moniliophthora perniciosa FA553

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,-
,Agaricales,-,-,Marasmiaceae,-,-,Moniliophthora,-,Moniliophthora perniciosa,-,-

mrr

Moniliophthora roreri MCA 2997

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,-
,Agaricales,-,-,Marasmiaceae,-,-,Moniliophthora,-,Moniliophthora roreri,-,-

mrt

Malassezia restricta KCTC 27527

Eukaryota,Fungi,Dikarya,Basidiomycota,Ustilaginomycotina,-,Malasseziomycetes,-,-,Malasseziales,-,-,
,Malasseziaceae,-,-,Malassezia,-,Malassezia restricta,-,-

msym

Malassezia sympodialis ATCC 42132

Eukaryota,Fungi,Dikarya,Basidiomycota,Ustilaginomycotina,-,Malasseziomycetes,-,-,Malasseziales,-,-,
,Malasseziaceae,-,-,Malassezia,-,Malassezia sympodialis,-,-

mtm

Thermothelomyces thermophilus ATCC 42464 (Myceliophthora thermophila ATCC 42464)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Sordariales,-
,,-,Chaetomiaceae,-,-,Thermothelomyces,-,Thermothelomyces thermophilus,-,-

nce

Nosema ceranae BRL01

Eukaryota,Fungi,-,Microsporidia,-,-,-,-,-,Apansporoblastina,-,-,Nosematidae,-,-,Nosema,-,Nosema
ceranae,-,-

ncr

Neurospora crassa OR74A

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Sordariales,-
,,-,Sordariaceae,-,-,Neurospora,-,Neurospora crassa,-,-

ncs

Naumovozyma castellii CBS 4309

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Naumovozyma,-,Naumovozyma castellii,-,-

ndi

Naumovozyma dairenensis CBS 421

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Naumovozyma,-,Naumovozyma dairenensis,-,-

nfi

Aspergillus fischeri NRRL 181 (Neosartorya fischeri NRRL 181)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,
,Aspergillaceae,-,-,Aspergillus,Aspergillus subgen. Fumigati,Aspergillus fischeri,-,-

nhe

Fusarium vanettenii 77-13-4 (Nectria haematococca mpVI 77-13-4)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-
,Hypocreales,-,-,-,Nectriaceae,-,-,-,Fusarium,-,Fusarium vanettenii,-,-

npa

Neofusicoccum parvum UCRNP2

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,-,-,-,Botryosphaeriales,-,-,-
,Botryosphaeriaceae,-,-,-,Neofusicoccum,-,Neofusicoccum parvum,-,-

nte

Neurospora tetrasperma FGSC 2508

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Sordariales,-
,,-,Sordariaceae,-,-,-,Neurospora,-,Neurospora tetrasperma,-,-

pan

Podospira anserina S mat+

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Sordariales,-
,,-,Chaetomiaceae,-,-,-,Podospira,-,Podospira anserina,-,-

pbl

Paracoccidioides lutzii Pb01 (Paracoccidioides brasiliensis Pb01)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Onygenales,-,-
,,-,-,-,Paracoccidioides,-,Paracoccidioides lutzii,-,-

pbn

Paracoccidioides brasiliensis Pb18

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Onygenales,-,-
,,-,-,-,Paracoccidioides,-,Paracoccidioides brasiliensis,-,-

pco

Phanerochaete carnosae HHB-10118-sp

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,-,Polyporales,-,-,-
,Phanerochaetaceae,-,-,-,Phanerochaete,-,Phanerochaete carnosae,-,-

pcs

Penicillium rubens Wisconsin 54-1255 (Penicillium chrysogenum Wisconsin 54-1255)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,-
,Aspergillaceae,-,-,-,Penicillium,-,Penicillium rubens,-,-

pdp

Penicillium digitatum Pd1

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,-
,Aspergillaceae,-,-,-,Penicillium,-,Penicillium digitatum,-,-

pfj

Pseudocercospora fijiensis CIRAD86

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Dothideomycetidae,-,-
,Mycosphaerellales,-,-,-,Mycosphaerellaceae,-,-,-,Pseudocercospora,-,Pseudocercospora fijiensis,-,-

pfp

Anthracycystis flocculosa PF-1 (Pseudozyma flocculosa PF-1)

Eukaryota,Fungi,Dikarya,Basidiomycota,Ustilaginomycotina,-,Ustilaginomycetes,-,-,-,Ustilaginales,-,-,-
,Ustilaginaceae,-,-,-,Anthracycystis,-,Anthracycystis flocculosa,-,-

pfy

Pestalotiopsis fici W106-1

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Xylariomycetidae,-,Xylariales,-,-,
,Sporocadaceae,-,-,Pestalotiopsis,-,Pestalotiopsis fici,-,-

pgr

Puccinia graminis f. sp. tritici CRL 75-36-700-3

Eukaryota,Fungi,Dikarya,Basidiomycota,Pucciniomycotina,-,Pucciniomycetes,-,-,Pucciniales,-,-,
,Pucciniaceae,-,-,Puccinia,-,Puccinia graminis,-,-

pgu

Meyerozyma guilliermondii (Pichia guilliermondii) ATCC 6260

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Meyerozyma,-,Meyerozyma guilliermondii,-,-

pic

Scheffersomyces stipitis (Pichia stipitis) CBS 6054

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Scheffersomyces,-,Scheffersomyces stipitis,-,-

pkz

Pichia kudriavzevii CBS573

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Pichiaceae,-,-,Pichia,-,Pichia kudriavzevii,-,-

plj

Purpureocillium lilacinum

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-,
,Hypocreales,-,-,Ophiocordycipitaceae,-,-,Purpureocillium,-,Purpureocillium lilacinum,-,-

pno

Parastagonospora nodorum SN15 (Phaeosphaeria nodorum SN15)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Pleosporomycetidae,-,-,
,Pleosporales,Pleosporineae,-,-,Phaeosphaeriaceae,-,-,Parastagonospora,-,Parastagonospora nodorum,-,-

ppa

Komagataella phaffii GS115 (Pichia pastoris GS115)

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Phaffomycetaceae,-,-,Komagataella,-,Komagataella phaffii,-,-

ppl

Postia placenta Mad-698-R

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Polyporales,-,-,
,Dacryobolaceae,-,-,Postia,-,Postia placenta,-,-

psco

Mollisia scopiformis CBS 120377

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Leotiomycetes,-,-,Helotiales,-,-,Mollisiaceae,-,-,
,Phialocephala,-,Phialocephala scopiformis,-,-

psq

Punctularia strigosozonata HHB-11173 SS5

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Corticiales,-,-,
,Punctulariaceae,-,-,Punctularia,-,Punctularia strigosozonata,-,-

pte

Pyrenophora teres f. teres 0-1

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Pleosporomycetidae,-,
,Pleosporales,Pleosporineae,-,-,Pleosporaceae,-,-,Pyrenophora,-,Pyrenophora teres,-,-

sapo

Scedosporium apiospermum IHEM 14462

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,
,Microascales,-,-,Microascaceae,-,-,Scedosporium,-,Scedosporium apiospermum,-,-

sce

Saccharomyces cerevisiae S288c

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Saccharomyces,-,Saccharomyces cerevisiae,-,-

scm

Schizophyllum commune H4-8

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,
,Agaricales,-,-,Schizophyllaceae,-,-,Schizophyllum,-,Schizophyllum commune,-,-

shs

Stereum hirsutum FP-91666 SS1

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Russulales,-,-,
,Stereaceae,-,-,Stereum,-,Stereum hirsutum,-,-

sla

Serpula lacrymans var. lacrymans S7.9

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,Agaricomycetidae,-,
,Boletales,Coniophorineae,-,-,Serpulaceae,-,-,Serpula,-,Serpula lacrymans,-,-

slb

Sugiyamaella lignohabitans CBS 10342

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Trichomonascaceae,-,-,Sugiyamaella,-,Sugiyamaella lignohabitans,-,-

smo

Sordaria macrospora k-hell

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Sordariales,-
,,-,Sordariaceae,-,-,Sordaria,-,Sordaria macrospora,-,-

spaa

Spathaspora passalidarum NRRL Y-27907

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Debaryomycetaceae,-,-,Spathaspora,-,Spathaspora passalidarum,-,-

spo

Schizosaccharomyces pombe 972h-

Eukaryota,Fungi,Dikarya,Ascomycota,Taphrinomycotina,-,Schizosaccharomycetes,-,-,
,Schizosaccharomycetales,-,-,Schizosaccharomycetaceae,-,-,Schizosaccharomyces,-
,Schizosaccharomyces pombe,-,-

ssck

Sporothrix schenckii 1099-18

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,
,Ophiostomatales,-,-,Ophiostomataceae,-,-,Sporothrix,-,Sporothrix schenckii,-,-

ssl

Sclerotinia sclerotiorum 1980 UF-70

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Leotiomycetes,-,-,Helotiales,-,-,Sclerotiniaceae,-,-
,-,Sclerotinia,-,Sclerotinia sclerotiorum,-,-

tasa

Trichosporon asahii var. asahii CBS 2479

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Tremellomycetes,-,-,Trichosporonales,-,-,
,Trichosporonaceae,-,-,Trichosporon,-,Trichosporon asahii,-,-

tbl

Tetrapapispora blattae CBS 6284

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Tetrapapispora,-,Tetrapapispora blattae,-,-

tdl

Torulaspora delbrueckii CBS 1146

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Torulaspora,-,Torulaspora delbrueckii,-,-

tmf

Talaromyces marneffeii ATCC 18224

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,
,Trichocomaceae,-,-,Talaromyces,-,Talaromyces marneffeii,-,-

tml

Tuber melanosporum Mel28

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Pezizomycetes,-,-,Pezizales,-,-,Tuberaceae,-,-,
,Tuber,-,Tuber melanosporum,-,-

tmn

Phaeoacremonium minimum UCRPA7 (Togninia minima UCRPA7)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Togniniales,-
,,-,Togniniaceae,-,-,Phaeoacremonium,-,Phaeoacremonium minimum,-,-

tms

Tremella mesenterica DSM 1558

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Tremellomycetes,-,-,Tremellales,-,-,
,Tremellaceae,-,-,Tremella,-,Tremella mesenterica,-,-

tpf

Tetrapapispora phaffii CBS 4417

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Tetrapapispora,-,Tetrapapispora phaffii,-,-

tre

Trichoderma reesei QM6a

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,-,
,Hypocreales,-,-,Hypocreaceae,-,-,Trichoderma,-,Trichoderma reesei,-,-

trg

Talaromyces rugulosus W13939

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Eurotiales,-,-,
,Trichocomaceae,-,-,Talaromyces,-,Talaromyces rugulosus,-,-

trr

Trichoderma reesei RUT C-30

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,
,Hypocreales,-,-,Hypocreaceae,-,-,Trichoderma,-,Trichoderma reesei,-,-

ttt

Thermothielavioides terrestris NRRL 8126 (Thielavia terrestris NRRL 8126)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Sordariomycetidae,-,-,Sordariales,-
,-,Chaetomiaceae,-,-,Thermothielavioides,-,Thermothielavioides terrestris,-,-

tve

Trichophyton verrucosum HKI 0517

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Onygenales,-,
,-,Arthrodermataceae,-,-,Trichophyton,-,Trichophyton verrucosum,-,-

tvs

Trametes versicolor FP-101664 SS1

Eukaryota,Fungi,Dikarya,Basidiomycota,Agaricomycotina,-,Agaricomycetes,-,-,Polyporales,-,-,
,Polyporaceae,-,-,Trametes,-,Trametes versicolor,-,-

uma

Ustilago maydis 521

Eukaryota,Fungi,Dikarya,Basidiomycota,Ustilaginomycotina,-,Ustilaginomycetes,-,-,Ustilaginales,-,-,
,Ustilaginaceae,-,-,Ustilago,-,Ustilago maydis,-,-

ure

Uncinocarpus reesii 1704

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Eurotiomycetes,Eurotiomycetidae,-,-,Onygenales,-,
,-,Onygenaceae,-,-,Uncinocarpus,-,Uncinocarpus reesii,-,-

val

Verticillium alfalfae VaMs.102 (Verticillium albo-atrum VaMs.102)

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,
,Glomerellales,-,-,Plectosphaerellaceae,-,-,Verticillium,-,Verticillium alfalfae,-,-

vda

Verticillium dahliae VdLs.17

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Sordariomycetes,Hypocreomycetidae,-,
,Glomerellales,-,-,Plectosphaerellaceae,-,-,Verticillium,-,Verticillium dahliae,-,-

vpo

Vanderwaltozyma polyspora DSM 70294

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Vanderwaltozyma,-,Vanderwaltozyma polyspora,-,-

wic

Wallemia ichthyophaga EXF-994

Eukaryota,Fungi,Dikarya,Basidiomycota,Wallemiomycotina,-,Wallemiomycetes,-,-,Wallemiales,-,-,
,Wallemiaceae,-,-,Wallemia,-,Wallemia ichthyophaga,-,-

wse

Wallemia mellicola CBS 633.66 (Wallemia sebi CBS 633.66)

Eukaryota,Fungi,Dikarya,Basidiomycota,Wallemiomycotina,-,Wallemiomycetes,-,-,Wallemiales,-,-,
,Wallemiaceae,-,-,Wallemia,-,Wallemia mellicola,-,-

yli

Yarrowia lipolytica CLIB122

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Dipodascaceae,-,-,Yarrowia,-,Yarrowia lipolytica,-,-

zmk

Zygorulasporea mrakii NRRL Y-6702

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Zygorulasporea,-,Zygorulasporea mrakii,-,-

zro

Zygosaccharomyces rouxii CBS 732

Eukaryota,Fungi,Dikarya,Ascomycota,Saccharomycotina,-,Saccharomycetes,-,-,Saccharomycetales,-,-,
,Saccharomycetaceae,-,-,Zygosaccharomyces,-,Zygosaccharomyces rouxii,-,-

ztr

Zymoseptoria tritici IPO323

Eukaryota,Fungi,Dikarya,Ascomycota,Pezizomycotina,-,Dothideomycetes,Dothideomycetidae,-,-
,Mycosphaerellales,-,-,Mycosphaerellaceae,-,-,Zymoseptoria,-,Zymoseptoria tritici,-,-

Referencias Bibliográficas

1. Alcoba-Florez J, Mendez-Alvarez S, Cano J, Guarro J, Perez-Roth E, del Pilar Arevalo M: (2005) Phenotypic and molecular characterization of *Candida nivariensis* sp. nov., a possible new opportunistic fungus. J Clin Microbiol, 43:4107–4111.
2. Anderson HW. (1917). Yeast-like fungi of the human intestinal tract Journal of infectious diseases. 21 (341).
3. Anderson, S., Kurland, C. (1998) Reductive evolution of resident genomes. Trends Microbiol. 6(7):263-8.
4. Arendrup MC, Perlin DS, Jensen RH, Howard SJ, Goodwin J, Hope W (2012) Differential in vivo activities of anidulafungin, caspofungin, and micafungin against *Candida glabrata* isolates with and without FKS resistance mutations. Antimicrob Agents Chemother 56:2435–2442
5. Armour, J. A., R. Neumann, S. Gobert y A. J. Jeffreys. (1994). Isolation of human simple repeat loci by hybridization selection. Human Molecular Genetics. 3: 599-605.
6. Bennett RJ, Forche A, Berman J. (2014) Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harb Perspect Med*. 2014;4:a019604
7. Bishop JA, Chase N, Magill SS, Kurtzman CP, Fiandaca MJ, Merz WG: (2008). *Candida bracarensis* detected among isolates of *Candida glabrata* by peptide nucleic acid fluorescence in situ hybridization: susceptibility data and documentation of presumed infection. J Clin Microbiol, 46:443–446.
8. Bolotin-Fukuhara, M., Fairhead, C. (2014). *Candida glabrata*: a deadly companion? Yeast, Volume 31, Issue 8, p. 279-88.
9. Botstein, D., R. L. White, M. Skolnick y R. W. Davis. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics. 32: 314-331.
10. Boyle, B., Dallaire, N., MacKay, J. (2009). Evaluation of the impact of single nucleotide polymorphisms and primer mismatches on quantitative PCR. BMC Biotechnol. Aug 28; 9:75.

11. Breslauer, K., Frank, R., Blöcker, H. & Marky, L. (1986). Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*, 83(11). 3746–3750.
12. Brown, G., Denning, D., Gow, N., Levitz., Netea, M., White, T. (2012). Hidden killers: human fungal infections. *Sci Transl Med*. 4(165).
13. Brunke, S., Hube, B. (2013). Two unlike cousins: *Candida albicans* and *Candida glabrata* infection strategies. *Cell Microbiol*. 15(5):701-8.
14. Butler G, Rasmussen MD, Lin MF, et al. (2009). Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*. 459:657–62
15. Capriotti, A. (1961). *Torulopsis castelli* sp. nov., a Yeast Isolated from a Finnish Soil. *Microbiology*. Volume 26, Issue 1.
16. Casadevall A, Fang FC, Pirofski L-A. (2011) Microbial virulence as an emergent property: consequences and opportunities. *PLoS Pathog*. 7:e1002136.
17. Casadevall, A., Coelho, C., Alanio, A., (2018) Mechanism of *Cryptococcus neoformans*-Mediated Host Damage. *Front Immunol*. 30:9:855.
18. Cheng, H. H. y L. B. Crittenden. (1994). Microsatellite markers for genetic mapping in the chicken. *Poultry Science*. 73: 539-546.
19. Cheng, H. H., I. Levin, R. L. Vallejo, H. Khatib, J. B. Dodgson, L. B. Crittenden y J. Hillel. (1995). Development of a genetic map of the chicken with markers of high utility. *Poultry Science*. 74: 1855-1874.
20. Compeau, Phillip E C; Pevzner, Pavel A; Tesler, Glenn (2011). «How to apply de Bruijn graphs to genome assembly». *Nature Biotechnology* 29 (11): 987-991.
21. Cormack, BP., Ghori, N., Falkow, S. (1999). An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells. *Science* 285 (5427).
22. Corradi, N., Slamovits, C., (2011). The intriguing nature of microsporidian genomes. *Brief Funct Genomics*. 115-24.
23. Correia A, Sampaio P, James S et al. *Candida bracarensis* sp. nov., a novel anamorphic yeast species phenotypically similar to *Candida glabrata*. *International journal of systematic and evolutionary microbiology*. (2006). 56(Pt 1):313–7
24. Corzo-León DE., Alvarado-Matute, T.,Colombo, A., Surveillance of *Candida* spp bloodstream infections: epidemiological trends and risk factors of death in two Mexican tertiary care hospitals. *PLoS One* (2014) 12;9(5):e97325.

25. Cui, L., Morris, A., Ghedin, E. (2013) The human mycobiome in health and disease. *Genome Med* 30;5(7):63.
26. David LH. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol. Res.* (2001). 105:1422–1432.
27. Diekema, D., Arbefeville, S, Boyken, L., Kroeger, J., Pfaller, M. (2012). The changing epidemiology of healthcare associated candidemia over three decades.
28. Dodgson, J. B., H. H. Cheng y R. Okimoto. (1997). DNA marker technology: a revolution in animal genetics. *Poultry Science*. 76: 1108-1114.
29. Domergue, R., Castaño I., De las Peñas, A., Zupancic, M., Lockatell, V., Hebel, R., Johnson, D., Cormack. (2005). Nicotinic acid limitation regulates silencing of *Candida* adhesins during UTI. *Science*. 308(5723).
30. Dujon, B., D. Sherman, et al. (2004). "Genome evolution in yeasts." *Nature* 430(6995): 35-44
31. Ehrlich, G., Hiller, L., Hu, F. (2008). What makes pathogens pathogenic. *Genome Biol.* 9(6):225.
32. Fanning S, Mitchell AP (2012) Fungal Biofilms. *PLoS Pathog* 8(4): e1002585.
33. Foulet, F., Nicolas, N., Eloy, O., Botterel, F., Gantier, J., Bretagne, S., (2005). Microsatellite Marker Analysis as a Typing System for *Candida glabrata*. *Journal of Clinical Microbiology*. 43(9): 4574-4579.
32. Frías-De-León, MG., Hernández-Castro, R., Conde-Cuevas, E., García-Coronel, I., (2021). "Candida glabrata Antifungal Resistance and Virulence Factors, a Perfect Pathogenic Combination. *Pharmaceutics* 13 (10).
35. Gabaldón T., Martin T., Marcet-Houben M., Durrens P., Bolotin-Fukuhara M., Lespinet O., Arnais S., Boissard S., Aguilera G., Atanasova R., et al., (2013) Comparative genomics of emerging pathogens in the *Candida glabrata* clade. *BMC Genomics*. 14 :623. doi: 10.1186/1471-2164-14-623.
36. Gabaldón, T., Carreté, L., (2016) The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in *Candida glabrata*. *Fems Yeast Research*, Volume, 16, Issue 2.

37. Ghannoum, M., Jurenvic, R., Mukherjee, P., Cui, F., Gillevet, P. (2010). Characterization of the oral fungal microbiome (mycobiome) in healthy individuals.
38. Goldstein, B. D. y C. Schlotterer C. (1999). Microsatellites evolution and applications. Oxford University Press, New York, 352pp.
39. Grimley, P., Wright, L, Jennings, A. *Torulopsis glabrata* infection in man. (1965). The American Journal of Clinical Pathology. Vol 43, num3.
40. Groot PW, Bader O, de Boer AD, Weig M, Chauhan N. Eukaryot Cell. (2013); 12:470-81.
41. Hagman A, Säll T, Compagno C, et al. (2013). Yeast “make-accumulate-consume” life strategy evolved as a multi-step process that predates the whole genome duplication. *PLoS One*. 8:e68734.
42. Hancock, J. (1999). Microsatellites and other simple sequence: genomic context and mutational mechanisms. En: Goldstein D., Schlotterer C. (eds.). Microsatellites evolution and applications, Oxford University Press, New York, pp. 1- 10.
43. Hazen, K., Hazen, B., Allietta, M., (1995) Cell wall anchoring to cytoplasmic membrane of *Candida*. FEMS Microbiol Lett. Jan 15; 125(2-3): 143-7.
44. Healey, K., Jimenez, C., Shor, E., Perlin, D. (2016). Genetic of multidrug resistance in *Candida glabrata*. Front Microbiol. 15;7:1995.
45. Ho, H., Haynes, K. (2015). *Candida glabrata*: new tools and technologies-expanding the toolkit. FEMS Yeast Res. 2015;15
46. Jeffreys, A. J., V. Wilson y S. L. Thein. (1985). Hypervariable «minisatellite» regions in human DNA. Nature. 314: 67-73.
47. Jones T, Federspiel NA, Chibana H, et al. (2004). The diploid genome sequence of *Candida albicans*. *P Natl Acad Sci USA*. 101:7329–34.
48. Kaur, R., Domergue, R., Zupancic, M., Cormack, B. (2005). A yeast by any other name: *Candida glabrata* and interaction with the host.
49. Kiasat, N., Rezaei-Matehkolaei, A., Mahmoudabadi, Ali., (2019). Microsatellite Typing and Antifungal Susceptibility of *Candida glabrata* Strains Isolated from Patients With *Candida Vaginitis*. Frontiers in Microbiology 10:1678.
50. Kumar K., Askari F., Sagar M., Kaur R. (2019) *Candida glabrata*: A lot more than meets the eye. Microorganisms 7 (2): 39.
51. Kuhn DM, Vyas VK FEMS Levadura Res. (2012); 12(4):398-414.

52. Kurtzman CP, Robnett CJ. (1991). Phylogenetic relationships among species of *Saccharomyces*, *Schizosaccharomyces*, *Debaryomyces* and *Schwanniomyces* determined from partial ribosomal RNA sequences. *Yeast*/Volume 7, Issue1/P.61-72.
53. Kurtzman CP, Robnett CJ. (1998). Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Van Leeuwenhoek* 73:331–71.
54. Kurtzman CP, Fell JW. (1998). *The Yeasts, A taxonomic Study*. 4th edition. London: Elsevier Science.
55. Kurtzman CP, Robnett CJ. (2003) Phylogenetic relationship among yeasts of the *Saccharomyces* complex determined from multigene sequence analyses. *FEMS Yeast Research*, Volume 3, Issue 4, June 2003, pages 417-432.
56. Kurtzman, CP., Fell JW., Boekhout, T., (2011) Methods for Isolation phenotypic characterization and maintenance of yeast. *Yeasts* 87-110.
57. Lafontaine I, Fischer G, Talla E, et al. (2004). Gene relics in the genome of the yeast *Saccharomyces cerevisiae*. *Gene*. 335:1–17
58. Landre P. A., D. H. Gelfand y R. H. Watson. (1995) The use of cosolvents to enhance amplification by the polymerase chain reaction. *PCR Strategies*. Academic Press, EE.UU., pp. 3-16.
59. Levin, I., L. Santagelo, H. Cheng, B. Crittenden y B. Dodgson. (1994). An autosomal genetic linkage map of the chicken. *The Journal of Heredity*. 85: 79-85.
60. Li, Y., Bao, W. (2007). Why do some yeast species require niacin for growth? Different modes of NAD synthesis. *FEMS Yeast Res.* 7(5):657-64.
61. Litt, M. y J. A. Luty. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*. 44: 397-401.
62. Lockhart, P., Brennan, M., Thornhill, M., Michalowicz, B., Sasser, H. (2009) Poor oral hygiene as a risk factor for infective endocarditis-related bacteremia. *J Am Dent Assoc*. 140(10).
63. Lockhart, S., Chowdhary, A., Gold, J., (2023) The rapid emergence of antifungal-resistant human-pathogenic fungi. *Nature Reviews Microbiology*. 21,818-832.

64. Lorenz, R., Bernhart, H., Stadler, P. Hofacker, I. (2011) ViennaRNA Package 2.0 Algorithms for Molecular Biology 6.
65. Luo, Z., van Vuuren, H., (2009) Functional analyses of *PAU* genes in *Saccharomyces cerevisiae*. Microbiology, Volume 155, Issue 12.
66. MacCallum, D., Castillo, L., Nather, K., Munro, C., Brown, A., Odds, F. (2009). Property Differences among the four major *Candida albicans* strain clades. Eukaryotic cell, p373-387.
67. McPherson, MJ, Quirke, P, Taylor R (eds) (1991): "PCR. A Practical Approach". Oxford: Oxford University Press.
68. Maillet L, Boscheron C, Gotta M, et al. (1996). Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes Dev.* 10:1796–811
69. Marcet-Houben, M., Gabaldon, T. (2010). Acquisition of prokaryotic genes by fungal genomes. Trends Genet. 26(1):5.
70. Marcet-Houben M, Gabaldon T. (2015). Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the Baker's yeast lineage. *PLoS Biol.* 13:e1002220.
71. Mason, K., Downward, J., Mason, K., Falkowski, N., Eaton, K., Kao, J., Young, V., Huffnagle, G. (2012). Infect Immun. 80(10): 3371-80.
72. Mayer, L., Wilson, D., Hube, B. (2013) *Candida albicans* pathogenicity mechanisms. Virulence, 119-128.
73. Miramón P., Lorenz, M., (2016). The SPS amino acid sensor mediates nutrient acquisition and immune evasion in *Candida glabrata*.
74. Moran, G., Stokes, C., Thewes, S., Hube, B., Coleman, D., Sullivan, D. (2004) Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence- associated genes in *Candida dubliniensis*.
75. Moreira A, Silva S, Botelho C, et al. (2015). *Candida braccarensis*: evaluation of virulence factors and its tolerance to amphotericin B and fluconazole. Mycopathologia. 5–6:305–15.
76. Mühlhausen, S., Kollmar, M. (2014). Molecular Phylogeny of Sequenced *Saccharomycetes* Reveals Polyphyly of the Alternative Yeast Codon Usage. Genome Biology and Evolution. Volume 6, Issue 12.

77. Mullis, K. B., F. Faloona, S. Scharf, R. Saiki, G. Horn y H. Erlich. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51: 263-273.
78. Oliveira E. J., J. G. Pádua, M. I. Zucchi, R. Vencovsky y M. L. Carneiro-Vieira. (2006). Origin, evolution, and genome distribution of microsatellites. Genetics and Molecular Biology 29: 294-307.
79. Pérez-García, L., Díaz-Jiménez, D., Mora-Montes, H., (2011) Role of Cell Wall Polysaccharides during Recognition of *Candida albicans* by the Innate Immune System. Glycobiology. Volume 1:
80. Pryszcz LP, Huerta-Cepas J, Gabaldón T. (2011). MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res.* 39: e32.
81. Plaut, A. Human Infection with *Cryptococcus glabratus*. (1950) Report involving uterus and fallopian tube. American Journal of Clinical Pathology. Vol. 20. No.4. pp. 377-80.
82. Rai, M., Balusu, S., Kaur, R., (2012) Functional genomic analysis of *Candida glabrata*-macrophage interaction: role of chromatin remodeling in virulence. Plos Pathogens; 8(8).
83. Rodrigues CF, Silva S, Henriques M. (2014). *Candida glabrata*: a review of its features and resistance. *Eur J Clin Microbiol Infect Dis* 33:673–88.
84. Roetzer A, Gregori C, Jennings AM et al (2008) *Candida glabrata* environmental stress response involves *Saccharomyces cerevisiae* Msn2/4 orthologous transcription factors. Mol Microbiol 69(3):603– 620
85. Russell, DW., (2001). Molecular Cloning: A Laboratory Manual. 3^a ed. Ed Cold Spring Harbor Laboratory Press. Nueva York. USA. 2344.
86. Sambrook, J., Fritsch, E., Maniatis, T. (1989). Molecular cloning: a laboratory manual. 2ed. New York: Cold Spring Harbor Laboratory Press.
87. Takashima, M., Sugita, T., (2022) Taxonomy of Pathogenic Yeasts *Candida*, *Cryptococcus*, *Malassezia*, and *Trichosporon*. Med. Mycol. J. 63, 119–132.
88. Tsai, H., Krol, A., Bennett, J., (2006) *Candida glabrata* PDR1, a transcriptional regulator of a pleiotropic drug resistance network, mediates azole resistance in clinical isolates and petite mutants. Antimicrobial Agents Chemotherapy; 50(4):1384-92.

89. Tautz, D. y C. Schlotterer. (1994). Simple sequences. *Current Opinion in Genetics and Development*. 4: 832-837
90. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG. 2012. Primer3 - new capabilities and interfaces. *Nucleic Acids Research* 40(15): e115.
91. Vale-Silva, L y Sanglard D. (2017). Comparative Genomics of Two Sequential *Candida glabrata* Clinical Isolates: 2413-2426
92. Vallone, P., Butler, J., (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques*, Vol.37, No 2.
93. Vance, J. F. y K. M. Othmane. (1998). Methods of genotyping. In: Haines, J.L., Pericak-Vance, M.A. eds. *Approaches to gene mapping in complex human diseases*. A John Wiley & Son, Inc., Publication. NY. pp 213-228.
94. Vanhala, T., M. Tuiskula-Haavisto, K. Elo, J. Vilkki y A. MakiTanila. (1998). Evaluation of genetic variability and genetic distances between eight chicken lines using microsatellite markers. *Poultry Science*. 77: 783-790.
95. Vignal, A., Milan, D., San Cristobal, M., Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*. 34, 275.
96. Wallace, B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T., Itakura., (1979) Hybridization of synthetic oligodeoxyribonucleotides to Φ X 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Research*, Volume 6, Issue 11, pages 3543-3558.
97. Warren, T. A. McTaggart, L. Richardson, S. Zhang. (2010) *Candida braccarensis* Infection in an Immunoconpromised Patient. *Journal of Clinical Microbiology*. 48(12) 4677-4679.
98. Weber, J. L. y P. E. May. (1989). Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet*. 44: 388- 396.
99. Wolfe, K., Shields, D. (1997). Molecular evidence for ancient duplication of the entire yeast genome. *Nature*. 387(6634):708.
100. Wu, Z., Irizarry, R., Martínez-Murillo, F., (2004) A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association*. Volume 99, -Issue 468.
101. Xu, Z., Green, B., Benoit, N., Schatz., Wheelan, S., Cormack, B. (2020). De novo genome assembly of *Candida glabrata* reveals cell wall protein complement and structure of

dispersed tandem repeat arrays. *Molecular Microbiology*, Volume 113, Issue 6, p. 1029-1224.