

# Práctico 10

## Genómica Comparada

### Objetivo

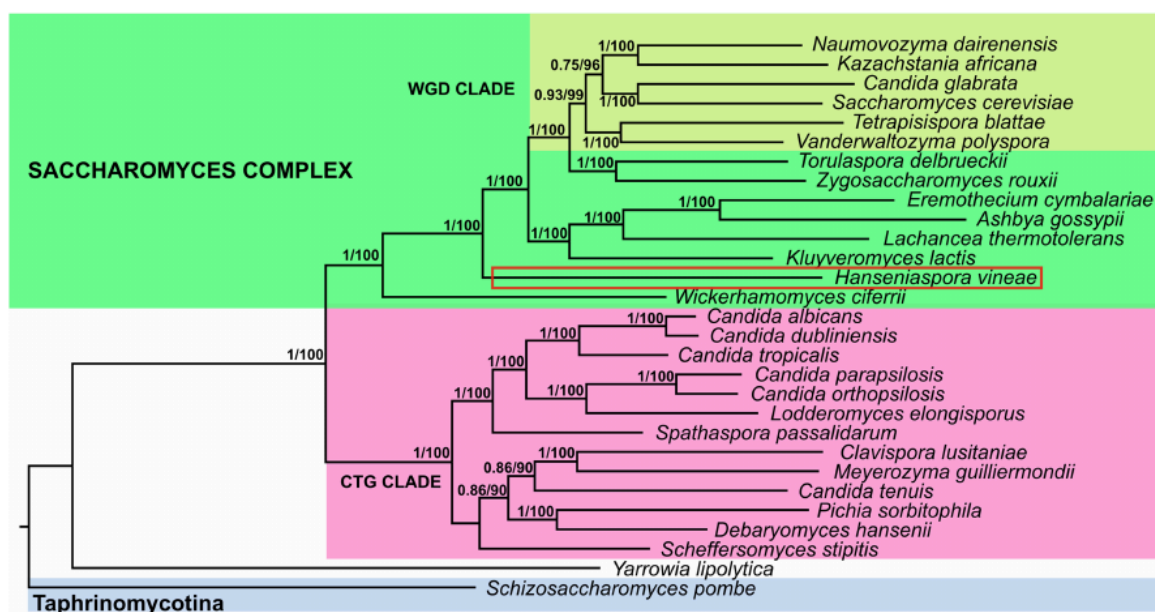
Entender la evolución de las levaduras apiculadas y ver cómo ha surgido la diversidad funcional (o distintas funciones) en dos clados dentro de estas levaduras. Como segundo objetivo, entender si el “clado de evolución rápida” de estas levaduras presentó un patrón evolutivo distinto al de evolución lenta.

### Introducción

*Saccharomyces cerevisiae* ('levadura de fermentador' o 'levadura de horneado') es una especie de levadura (hongo unicelular). La especie ha sido fundamental en la elaboración del vino y cerveza y en el horneado desde la antigüedad. *Saccharomyces* fue el primer eucariota cuyo genoma fue completamente secuenciado (¡1996!).

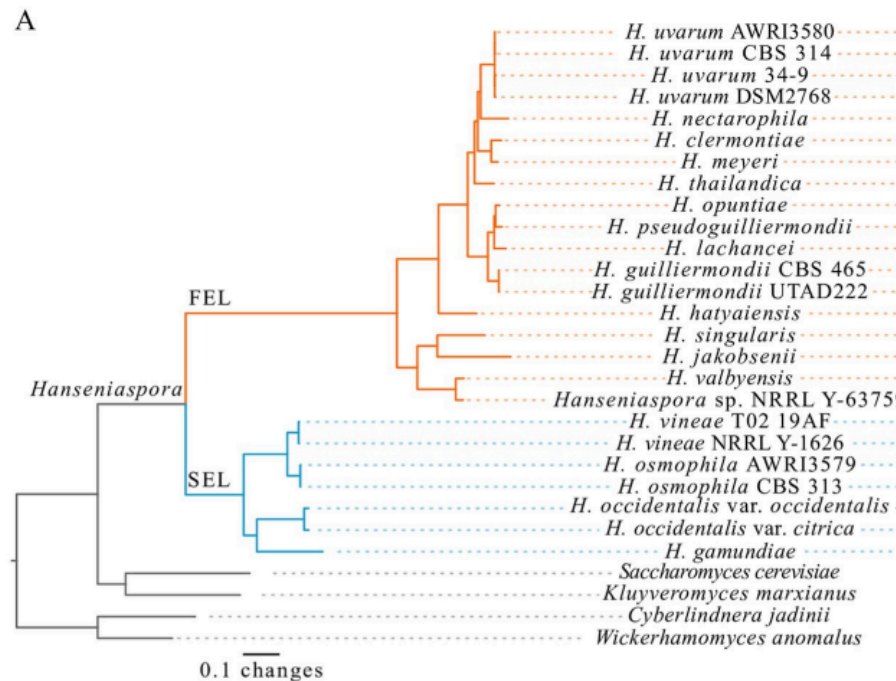
Otro género cercano a *Saccharomyces* es *Hanseniaspora*. Estas levaduras apiculadas se encuentran no sólo en las uvas, sino también en muchas otras frutas. Las levaduras apiculadas representan aproximadamente el 70% de la microbiota asociada a la uva. *Hanseniaspora vineae* está emergiendo como una especie prometedora para la producción de vino de calidad. Los vinos producidos por una mezcla de estas levaduras *H. vineae* con *S. cerevisiae* exhiben consistentemente sabores frutales más intensos y más complejos que los vinos producidos sólo por *S. cerevisiae*. Las especies de *Hanseniaspora* también son participantes clave en la fermentación de una gran variedad de otros productos alimenticios que van desde el chocolate hasta el tequila o la sidra de manzana.

El artículo en el que nos basamos (Giorello *et al.* 2018) detectó que *Hanseniaspora vineae* NO forma parte de una ronda de duplicación que abarca un linaje de 6 especies dentro de la radiación de estas levaduras, donde se encuentra *Saccharomyces* (ver figura).



Este trabajo encontró numerosos genes relevantes (ej., ejemplo aminotransferasas y descarboxilasas) que son resultado de duplicaciones génicas.

Más recientemente se encontró que dentro del género *Hanseniaspora*, *H. vineae* es parte de un linaje de evolución lenta (slow evolving lineage, SEL), ver figura, Steenwyk et al, 2019).



## Objetivo

Ahora tenemos como objetivo entender la evolución de las levaduras apiculadas y ver cómo ha surgido la diversidad funcional (o distintas funciones) en dos clados dentro de estas levaduras. Queremos también entender si el clado de evolución rápida presentó un patrón evolutivo distinto al de evolución lenta. Para esto analizaremos el genoma completo de 10 levaduras e identificamos qué clados sufrieron más duplicaciones e identificar que funciones cumplen aquellos genes duplicados.

## Metodología

Tomaremos algunas de las especies utilizadas en el trabajo de Steenwyk et al. y realizaremos un análisis genómico y de anotación funcional, análisis de ortología sumado a análisis de pérdida y ganancia de genes. Se considerarán *Hanseniaspora guilliermondii*, *H. occidentalis*, *H. osmophila*, *H. valbyensis*, *H. vineae*, *Cyberlindnera jadinii*, *Kazachstania servazzii*, *Saccharomyces cerevisiae*, *Wickerhamomyces anomalus* (ver diapositivas)

## Parte 1. Acceso a datos genómicos

a) Localizar uno de los dos genomas de levadura de uva de vino Tannat de Uruguay (ir a <https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=29832>). Podemos ubicar también el genoma de referencia ASM5094768v1 (depositado en 2025)  
Ver el número de genomas depositados para este género.

b) En Uniprot (<https://www.uniprot.org/>) buscar las proteínas de interés que se encuentran en la siguiente tabla (estamos restringiendo la búsqueda a *Saccharomyces*). Anoten el tipo de función o nombre común del gen. Por cada gen, en la sección “Genome annotation databases” encuentran el código/acceso de EnsemblFungi que ya tenemos en la tabla.

GEN	Acceso de Ensembl	Nombre común o Función
PDC1	<a href="https://www.ensemblgenomes.org/id/YLR044C">https://www.ensemblgenomes.org/id/YLR044C</a>	
ARO8	<a href="https://www.ensemblgenomes.org/id/YGL202W">https://www.ensemblgenomes.org/id/YGL202W</a>	
ARO9	<a href="https://www.ensemblgenomes.org/id/YHR137W">https://www.ensemblgenomes.org/id/YHR137W</a>	
ATF2	<a href="https://www.ensemblgenomes.org/id/YGR177C">https://www.ensemblgenomes.org/id/YGR177C</a>	

## Parte 2. Exploración de sintenia

Observar el contexto genómico en "Genomicus"

[<https://www.genomicus.bio.ens.psl.eu/genomicus-fungi-43.01/cgi-bin/search.pl>].

En **Genomicus**, buscar el código EnsemblFungi (ej. **YLR044C**) + *S. cerevisiae* S288C.

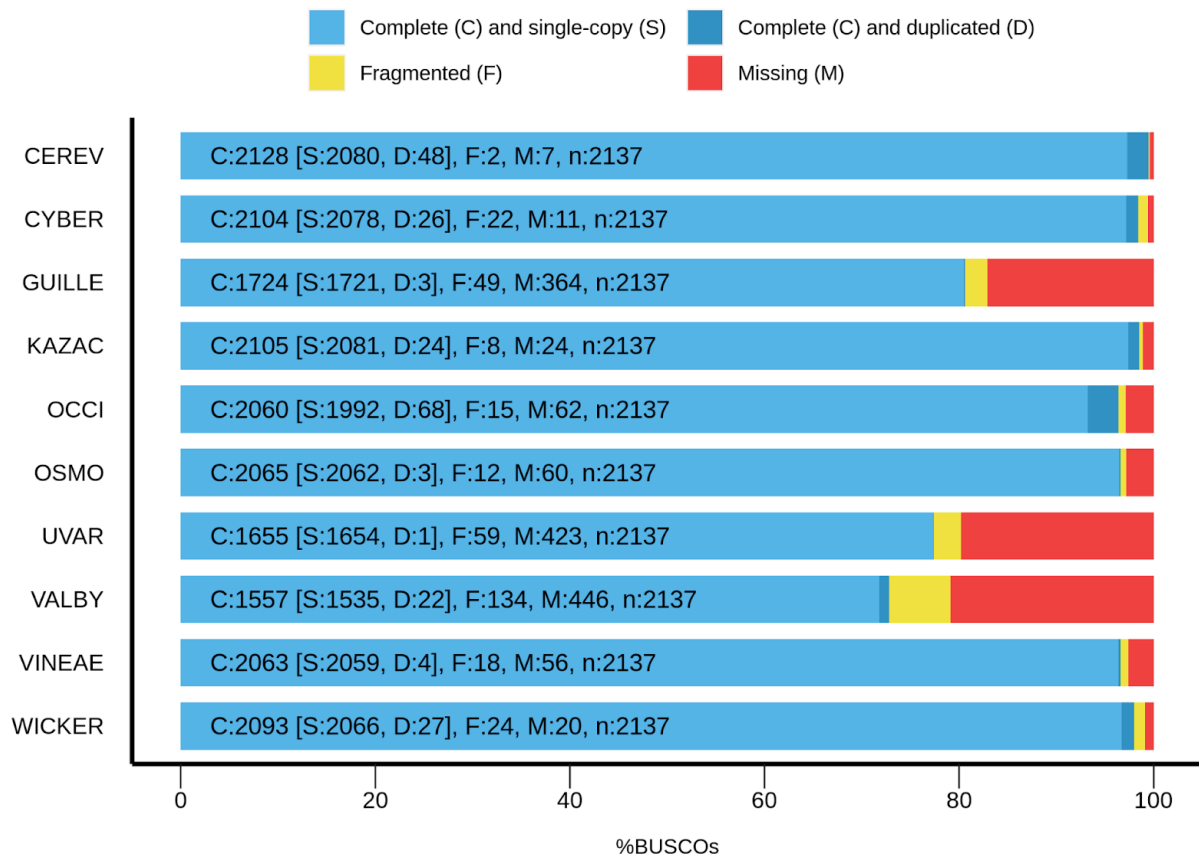
Identificar qué especies estamos comparando, cuánto se mantiene la sintenia y en que cromosomas se encuentran estos genes. Ya que la visualización es más amigable, identificar duplicaciones génicas en alguno de estos genes (nodos rojos: duplicación; nodos azules: especiación).

¿Aproximadamente, cuántos genomas logramos comparar por este método? ¿Están estos ortólogos en el mismo cromosoma? ¿Qué necesitamos para hacer buenas comparaciones o sacar buenas conclusiones con este enfoque o estrategia?

## Parte 3 Calidad del genoma y anotación funcional

Con la herramienta **BUSCO**, vemos cuán completo está cada cada genoma a partir de una base de datos de genes ortólogos de copia única. Si sólo consideramos genes de este tipo, ¿cómo se han originado estos genes? Cuando hacemos "**B**enchmarking **U**niversal **S**ingle-**C**opy **O**rtologs" (**BUSCO**) logramos cuantificar la “completitud” de un genoma ya que la expectativa es encontrar estos genes en el genoma como "copia única".

## BUSCO Assessment Results



Observando los resultados de **BUSCO** (a) ¿Qué genoma se encuentra mejor representado (i.e. más completo)? (b) ¿Qué desventaja tiene incluir genomas poco completos?

## Parte 4. Análisis de ortología y anotación funcional (en R)

Por un lado, **OrthoFinder** estima los ortólogos entre 10 genomas de levaduras. **eggNOG** nos permitió hacer la anotación funcional o conocer qué función cumplen los genes presentes en estos genomas.

Cuando procesamos parte de los resultados logramos más detalle de la historia evolutiva de estas especies. Algunos de los resultados son: un árbol de especies, múltiples árboles de genes, lista con genes ortólogos, lista con duplicaciones génicas y todos los archivos fasta con los ortólogos (**carpeta Parte4**).

Para esta actividad práctica usaremos, **R** (<https://cran.r-project.org/>). Los paquetes de **R** (y sus dependencias!) ya han sido instalados: *phytools*, *tidyverse*, *formattable*.

Ahora comenzamos a seguir el script **HansVineae.R**

# los números usados debajo -con este formato '(#1) '- están asociados en el script de **R**.

Siguiendo los comandos en el script podemos ver el árbol de OrthoFinder. ¿Cómo se construyó? Prestar atención al nombre o etiqueta de los nodos (#1)

Teniendo en cuenta el análisis global de ortología nos podemos fijar en algunos resultados que ayudan a describir 'el éxito' o cantidad de ortólogos recuperados.

¿Cuál es el número de genes en los ortogrupos? (#2)

¿Número de ortogrupos de copia simple?

¿Qué spp tiene más duplicaciones? Ver plot. (#3)

¿A qué clado pertenece esta spp.? ¿Son clado FEL o SEL?

¿Qué especies comparten más ortólogos? Para esto vemos la tabla en R o excel donde se compara cada especie y sus ortólogos compartidos. ¿A qué se debe este patrón? (#4)

Tenemos acceso a la estadística de ortólogos por cada genoma. Grafiquemos los siguientes resultados. (#5)

☐ N° ortogrupos por spp.

☐ N° genes por spp.

Con la anotación funcional podemos hacernos otras preguntas...

Entre todos los genes anotados vamos a buscar genes de interés en estas levaduras, parte de todo este 'arsenal' de genes asociados a la fermentación.

Nos concentramos en una "alcohol dehydrogenase" mencionada en la Tabla 3 de Giorrello et al. ¿Cuántas variantes de SPE1 o SPE2 se pudieron anotar? (#6) **¿Qué les permite hacer a las levaduras al tener más cantidad de estos genes?**

Recordando el resumen del artículo, hay genes señalados como relevantes en estas levaduras. Nos concentramos en un gen en particular...

¿Cuántos genes 'ARO' hay en los Ortólogos? (#7)

Si agrego detalle de qué genes son los que se duplicaron y en qué especies tengo otra dimensión de la historia del grupo, podemos hacer otras preguntas. De esta forma podemos entender si la diversidad de genes responde a duplicaciones en el ancestro de estos grupos de levaduras o duplicaciones dentro de cada especie.

¿Cuántas duplicaciones hay en todo el género *Hanseniaspora*? Para esto buscamos antes qué nodo es el ancestro común más cercano de este grupo (recordar paso #1) (#8)

Podemos buscar qué función (COG) cumplen estos genes duplicados.

Finalmente, ahora podemos visualizar las duplicaciones de algunos genes para poner en contexto cuáles son ortólogos y parálogos.

Visualice las duplicaciones para algunos genes como los genes ARO o PDC1 (#9) ¿Qué podemos intuir acerca de la relevancia de la función en algunos casos de duplicaciones?

## “Discusión”

¿Qué genomas agregarían para entender aún más la diversidad funcional de estas levaduras?

¿Qué genes o grupo de genes se podrían explorar con más detalle?

¿Qué potencial tiene mezclar en la producción de vinos u otros fermentos levaduras apiculadas (clados FEL & SEL) y otras levaduras?

## MINI GLOSARIO

**Anotación funcional:** el proceso de adjuntar información biológica a secuencias de genes o proteínas.

**Aminotransferasas:** enzimas que catalizan una reacción de transaminación entre un aminoácido y un  $\alpha$ -cetoácido. Son importantes en la síntesis de aminoácidos, que forman las proteínas.

**Categorías COG:** en la base de datos "Clusters of Orthologous Groups" (COG), cada COG incluye proteínas que se cree que son ortólogas. El propósito de la base de datos COG es servir como plataforma para la anotación funcional de genomas recién secuenciados y para estudios sobre la evolución del genoma. Para facilitar los estudios funcionales, los COG se clasificaron en 17 categorías funcionales amplias.

**Descarboxilasas:** son liasas de carbono-carbono que agregan o eliminan un grupo carboxilo de los compuestos orgánicos. Estas enzimas catalizan la descarboxilación de aminoácidos, beta-cetoácidos y alfa-cetoácidos.

**Deshidrogenasas:** una enzima perteneciente al grupo de las oxidoreductasas que oxida un sustrato al reducir un aceptor de electrones, generalmente  $\text{NAD}^+/\text{NADP}^+[1]$  o una coenzima flavina como FAD o FMN. Como todos los catalizadores, catalizan reacciones inversas y directas, y en algunos casos esto tiene un significado fisiológico: por ejemplo, la alcohol deshidrogenasa cataliza la oxidación de etanol a acetaldehído en animales, pero en la levadura cataliza la producción de etanol a partir de acetaldehído.

**Genes de copia única (o ortólogos copia única):** marcadores universales presentes sólo una vez en cada genoma (para un grupo dado, ejemplo vertebrados o fungi).

**GTPasas:** una superfamilia de proteínas que regulan muchos procesos celulares, como la señalización celular, el transporte vesicular y la regulación de la forma y motilidad celular.

**Ortogrupo:** es el conjunto de genes de múltiples especies que descienden de un solo gen

en el último ancestro común (LCA) de ese conjunto de especies.

**Ortólogo:** el “mismo” gen en diferentes especies que ha divergido como resultado de la especiación.

**Sintenia:** la conservación de bloques de orden dentro de dos conjuntos de cromosomas (o genomas) que se comparan entre sí.

### **BIBLIOGRAFÍA**

Giorello, Facundo, et al. "Genomic and transcriptomic basis of *Hanseniaspora vineae*'s impact on flavor diversity and wine quality." *Applied and Environmental Microbiology* 85.1 (2019): e01959-18.

Steenwyk, Jacob L., et al. "Extensive loss of cell-cycle and DNA repair genes in an ancient lineage of bipolar budding yeasts." *PLoS Biology* 17.5 (2019): e3000255.