HAYATI
Journal of Biosciences

Vol. 31 No. 4, July 2024 793-802
DOI:10.4308/hjb.31.4.793-802
ISSN: 1978-3019
EISSN: 2086-4094

Genomics and Phylogeny of *Rhodotorula glutinis* and *Rhodotorula kratochvilovae* Isolated from the Northern Peruvian Andes

Víctor Vásquez-Villalobos[1*], Angely M.A. Hidalgo-Arteaga[2], Roxana Sosa-Becerra[2], Bertha Soriano-Bernilla[2], Maurício Alexander de Moura Ferreira[3], Wendel Batista da Silveira[3]

[1]*Laboratorio de Biomoléculas-Facultad de Ciencias Agropecuarias, Universidad Nacional de Trujillo, Av. Juan Pablo II s/n Urb San Andrés Trujillo-13011 La Libertad, Perú*
[2]*Facultad de Ciencias Biológicas, Universidad Nacional de Trujillo, Av. Juan Pablo II s/n Urb San Andrés Trujillo-13011 La Libertad, Perú*
[3]*Departamento de Microbiologia, Universidade Federal de Viçosa, Av. Peter Henry Rolfs s/n Campus Universitário CEP: 36570-900 Viçosa–MG, Brazil*

ARTICLE INFO

Article history:
Received July 6, 2023
Received in revised form January 25, 2024
Accepted March 22, 2024

KEYWORDS:
Genomic analisis,
oleaginous yeasts,
phylogenetics,
Rhodotorula glutinis,
Rhodotorula kratochvilovae

ABSTRACT

Genomes of oleaginous yeast strains *Rhodotorula glutinis* CON-5 and *Rhodotorula kratochvilovae* POR-3, isolated from areas in the northern Peruvian Andes using SPAdes, were sequenced and assembled applying Illumina and de novo. Genomes of 20,515,696 and 20,738,185 bp, respectively, were determined. From the structural and functional annotations, the Basidiomycota phylum showed a similarity of 76.8% and 86.5% with 6,976 and 8,124 pairs of proteins in both yeasts respectively, with homologues in the UniProt data bank. Using OrthoVenn, a relationship between both yeasts was obtained from 450 orthologous groups. Likewise, the above-mentioned yeasts and *R. toruloides* (oleaginous Basidiomycota) showed 1,574 orthologous groups, indicating a good relationship. Construction of phylogenetic trees of genes encoding metabolic enzymes was also carried out, based on the ITS sequences which showed that CON-5 and POR-3 have a greater relationship with *R. graminis*. Their phylogenetic relationship was ascertained and determined that the enzymes involved in the metabolism of CON-5 and POR-3 are related to each other. It was also found that the protein sequences of the Basidiomycota phylum differ from Ascomycota. The study showed functional evidence regarding the lipid accumulation phenotype, an important aspect in the context of obtaining lipids or oleochemicals.

1. Introduction

During their evolutionary process, Oleaginous yeasts have developed the ability to accumulate fatty acids in the form of neutral lipids (Spagnuolo *et al.* 2019), making it possible to get from these, by synthesis, various oleochemicals and those that are usually obtained from oleaginous vegetables or animal fats. These processes have been widely studied in bacteria, microalgae and yeasts, as they can also provide esters, olefins, alkanes, ketones, polyesters, fatty acids and alcohols (Zhang *et al.* 2021). It has been reported that unicellular oil accumulated in high concentration in oleaginous yeasts, can be used for the production of third generation biodiesel (Shapaval *et al.* 2019; Gohain *et al.* 2020), an aspect that has been gaining relevance in the context of a circular economy (Sreeharsha and Mohan 2020). This type of yeasts is capable of accumulating up to 70% lipids in their biomass (Vieira *et al.* 2020), so it is interesting to have a broader understanding of metabolism, in order to achieve in the future, an increase in the yield of the desired metabolites, using metabolic or genetic engineering (Spagnuolo *et al.* 2019). These aspects affect the cost of the process of obtaining lipids, but can be reduced by optimization techniques, search and improvement of strains (Sreeharsha and Mohan 2020), such as *Yarrowia lipolitica, Lipomyces starkeyi, Rhodosporidium toruloides, Rhodotorula gracilis, Rhodotorula glutinis, Rhodotorula mucilaginosa* (Papanikolaou and Aggelis 2002), as well as *Rhodosporidium kratochvilovae*, which accumulate a large amount of triacylglycerides (TAG) in the form of intracellular lipids (Patel *et al.* 2015). Oleaginous

* Corresponding Author
E-mail Address: vvasquez@unitru.edu.pe

yeasts can produce lipids using different carbon sources, so it is important to understand their accumulation mechanism in the cell cytoplasm. According to research, sugars follow the glycolysis pathway by entering the Krebs cycle. It is important to mention that a limitation of nitrogen in the culture medium produces a decrease in adenosine monophosphate AMP, typical of oleaginous yeasts, which inactivates isocitrate dehydrogenase (ICDH). This stops isocitrate from being metabolized in the Krebs cycle, which allows acetyl-CoA and malonyl-CoA to form fatty acids of 14 to 16 carbons (Patel *et al.* 2016).

High-throughput sequencing technique has been used in the study of yeasts (Wei *et al.* 2018), focusing on 18S, 26S amplicons or the ITS (Internal Transcribed Spacer) region of the ribosome RNA gene (Wang *et al.* 2019). For a better understanding of the sequenced organism, the information from Next Generation Sequencing (NGS) must be assembled and annotated, allowing identification of biological characteristics, which can be done with predicted genes and homologous sequence alignments (De León-Medina *et al.* 2016). The genome annotation pipelines involve a "computation" stage, expressed sequence tags (ESTs), in which the proteins are aligned, generating ab initio gene predictions based on previous studies. A second "annotation" stage uses different tools and programs that allow annotating both protein-coding genes and non-coding RNA (ncRNAs) (Yandell and Ence 2012).

From the sequencing of genomes in the present investigation, the assembly of *R. glutinis* CON-5 and *R. kratochvilovae* POR-3, isolated from areas in the northern Peruvian Andes, was performed, carrying out comparative genomics studies by determining orthologous genes. The construction of phylogenetic trees of genes encoding metabolic enzymes was also completed, and their phylogenetic relationship with other yeasts established.

2. Materials and Methods

2.1. Oleaginous Yeast Strains

R. glutinis CON-5 and R. kratochvilovae POR-3, with 31.5% and 29.7% lipids in their biomass, respectively; identified by amplifying the sequences of their ITS regions of the rDNA, using Sanger technique, with the 5'-3' primers ITS1 (TCCGTAGGTGAACCTGCGG) and ITS4 (TCCTCCGCTTATTGATATGC). Stored in the Genbank (LC413754.1 and LC390313.1).

2.2. Genomic DNA Extraction

The strains were inoculated in 50 ml of YPD medium (1%, 2%, and 2% yeast extract, peptone, and dextrose) incubated at 30°C at 150 rpm for 48 h. For the extraction of genomic DNA, 2 ml of suspension were centrifuged at 12,000 rpm for 5 minutes, obtaining 0.250 g of biomass. DNA extraction was carried out using the Qiagen DNeasy PowerSoil Kit (Torres and Kelley 2018). DNA solutions were stored at –20°C.

2.3. Quantity and Quality of Genomic DNA Determination

2 µL of DNA was used in a Nanodrop 2,000c microspectrophotometer, in order to determine the concentration (ng/µL) and absorbance ratio between 260 and 280 nm. To determine purity, a 0.8% agarose gel was used with 1X Buffer TBE with a 1 kb molecular weight marker in horizontal electrophoresis, applying 75V for 30 min.

2.4. Sequencing and Assembly of Genomes

The CON-5 and POR-3 genomes were sequenced with a 2x300-bp paired-end Illumina MiSeq and Nextera XT. For the assembly, quality control of the raw sequences was first carried out using the FastQC program and trimming the sequences with Fastp. A de novo assembly was implemented using SPAdes v.3.14.1, and obtained metrics with QUAST v.4.6.3. The marking of repeated regions was carried out with RepeatMasker v.4.0.8 and for the structural annotation, the ab initio strategy with BRAKER 2, which was analysed using BUSCO v.3. Functional annotation was performed with DIAMOND v.0.9.21, and the genes were aligned with the sequences from the UniProt data bank.

2.5. Comparative Genomics between CON-5, POR-3 and other Yeasts

The CON-5 and POR-3 genomes were compared with the following selected yeasts from the UniProt database: *Rhodotorula toruloides* (oleaginous Basidiomycota), *Lipomyces starkeyi* and *Yarrowia lipolytica* (oleaginous Ascomycotas), *Cryptococcus gattii* (non-oleaginous Basidiomycota). Orthologous gene pools were established with OrthoVenn2.

2.6. Phylogenetic Analysis of ITS Regions CON-5 and POR-3 Genomes

The ITS regions of the genome of both yeasts were compared with the genomes belonging to the genus

Rhodotorula. The selected species were *R. taiwanensis* (ID:444436835), *R. mucilaginosa* (ID:555298745), *R. javanica* (ID:663440874), *R. graminis* (ID:407971637), *R. glutinis* (ID:451899284), *R. dairensis* (ID:663440724), *R. colostri* (ID:563427733), *R. glacialis* (ID:663440725), *R. ingeniosa* (ID:434861111), *R. yarrowii* (ID:20452354), *R. nothofagi* (ID: 663440568) obtained from the ITS 2 database, and *R. kratochvilovae* (ID: KJ707107) from the UNITE database. The sequences were first aligned with CLUSTAL W. Phylogenetic trees were made using UPGMA with MEGA X v.10.2.4 software, with the options: test phylogeny (setting: Bootstrap method with 500 replications), substitution model (setting: Poisson model), rates and patterns (setting: rates among sites: uniform rates), date subset to use (setting: Gaps/ Missing Data Treatment: Pairwise deletion), systeme resource usage (setting: Number of Threads: 3).

2.7. Phylogenetic Analysis of Enzymes Responsible for Lipid Production in CON-5 and POR-3

The species chosen to infer the similarity of enzymes were those used in orthologs, in relation to their oleaginous evolution and their conserved sequences: Acetyl CoA carboxylase (ACC), ATP-citrate lyase (ACL), glucose-6P dehydrogenase (G6PDH), isocitrate dehydrogenase NAD+dependent (NAD-ICDH), isocitrate dehydrogenase NADP+dependent (NADP-ICDH), and malic enzyme (ME) (Vieira *et al.* 2020). The sequences were first aligned using CLUSTAL W, and the phylogenetic trees were made by means of the Neighbor-joining method with the MEGA X v.10.2.4 software, with the options: phylogeny test (settings: Bootstrap method with 500 replications), substitution model (settings: Poisson model), rates and patterns (settings: rates among sites: uniform rates), pattern among lineages (settings: Same-homogeneous), Gaps/ Missing Data Treatment, systeme resource usage (setting: Number of Threads: 3).

3. Results

3.1. Nucleotide Sequence Accession Numbers

Sequences were deposited at NCBI under BioProject PRJNA705733 for CON-5 and POR-3, identified as SAMN18099295 and SAMN18099296, respectively.

3.2. Sequence Quality Control

After trimming, an improvement in the quality of the readings was obtained in forward (R1) in both yeasts, and in reverse (R2) which was not very noticeable, where the Q20 quality score value decreased in the last cycles. Table 1 presents a summary of these findings, showing that the yeast sequences decreased in size and length, maintaining the difference in G+C (%) between the R1 and R2 reads.

3.3. Genomes Assembly

Regarding the quality of the assembly, a very close genome size was observed between the two yeasts (Table 2), with a greater difference between the contigs and scaffolds, which is higher for POR-3. The quality of the assembly was expressed with

Table 1. Quality of raw sequences and trimming of CON-5 and POR-3

Yeasts	Raw sequences				Raw sequences			
	Overall size	Length (pb)	G+C (%)		Overall size	Length (pb)	G+C (%)	
			R1	R2			R1	R2
R. glutinis CON-5	8,027,841	35-301	63	64	7,687,126	15-230	63	64
R. kratochvilovae POR-3	6,936,032	35-301	61	62	6,692,019	15-210	61	62

Table 2. Summary of assembly metrics, structural and functional annotation of CON-5 and POR-3

Parameters	CON-5	POR-3
	Assembly	
Genome size (bp)	20,515,696	20,738,185
Contigs number	2,392	2,978
Scaffolds number	2,377	2,957
N50 (bp)	16,751	13,135
L50 (bp)	390	474
G+C (%)	63.5	61.5
Assembly (BUSCO)	C:88.1% [S:87.9%, D:0.2%], F:6.9%, M:5.0% C:1,176 [S:1,173, D:3], F:92, M:67, n: 1,335	C: 86.8% [S:86.6%, D:0.2%], F:8.2%, M:5.0% C:1,159 [S:1,156, D:3], F:110, M:66, n: 1,335
Coverage	230X	168X

Table 2. Continued

Genes identification		
Genes number (BRAKER 2)	7,408	9,040
Structural annotation		
Genes prediction (BUSCO)	C: 76.8% [S:76.5 %, D:0.2%], F: 18.6%, M:4.6% C:1,126 [S:1,123, D:3], F:248, M:61, n: 1,335	C: 86.5 % [S:86.1%, D:0.4%], F: 9.5%, M:4.0% C:1,155 [S:1,150, D:5], F:127, M:53, n: 1,335
Functional annotation		
Protein coding genes (DIAMOND)	Aligned pairs: 6,976 High-scoring Segment Pair HSPs: 6,995	Aligned pairs: 8,124 High-scoring Segment Pair HSPs: 8,132

Table 3. Global information of orthologous, individual genes and proteins of different yeast species

Yeast species	Genome size (bp)	Orthologous gene groups	Individual genes	Proteins
Rhodotorula toruloides	20,223,942	6,151	766	7,496
Lipomyces starkeyi	21,432,058	4,382	2,222	8,136
Yarrowia lipolytica	20,502,981	3,961	1,712	6,454
Cryptococcus gattii	17,438,253	3,935	2,163	6,471
CON-5	20,515,696	5,965	1,096	7,408
POR-3	20,738,185	6,561	1,867	9,040

the N50 of the contigs with values of 16,751 and 13,135, with an L50 of 390 and 474 for CON-5 and POR-3, respectively. There was a good similarity with 1,335 groups of genes of the Basidiomycota phylum determined with BUSCO, of 88.1% and 86.8% for CON-5 and POR-3 respectively, showing for both yeasts, besides their percentage, the number of genes: complete (C), unique (S), duplicates (D), fragmented (F) and unrecognized (M).

3.4. Structural Annotation

7,408 and 8,040 genes were identified with BRAKER 2 in yeasts in CON-5 and POR-3, with a prediction of 76.8% and 86.5% of similarity respectively (Table 2), indicating equally for both yeasts, besides their percentage, the number of genes: C, S, D, F, and M.

3.5. Functional Annotation

Annotation using DIAMOND indicated 6,976 (94.17%) and 8,124 (89.87%) protein-coding genes with homologs in the UniProt databank, as well as 6,995 and 8,132 High-scoring Segment Pairs (HSPs) with M: 432 (obtained from the difference: 7,408 – 6,976) and 916 (obtained from the difference 9,040 – 8,124); in *R. glutinis* CON-5 and *R. kratochvilovae* POR-3, respectively (Table 2).

3.6. Orthologous Gene Cluster Analysis

Table 3 shows the orthologous genes global information COG (Clusters of Orthologous Genes). A total of 8,218 orthologous groups were identified, 6,288 orthologous groups in at least two species and 1,930 single-copy gene clusters. The columns show group number "Clusters" of orthologous genes, "Singletons" of individual genes that do not belong to any group or that do not have identified orthologs, and "Proteins" of each species.

The most important results obtained with OrthoVenn of all the groups of species compared, which allows us to observe the distribution of genes shared between the yeast species analyzed, added to a total of 2,346 orthologous groups in the genome (Figure 1). CON-5 and POR-3 showed 450 exclusive groups. Figure 2 shows the distribution of different yeast species, related to the predominant orthologous genes with specific function in biological processes, molecular functions and cellular components.

3.7. Phylogenetic ITS Regions Relationship of Genomes Belonging to the Genus *Rhodotorula*

According to the species in the phylogenetic tree based on the ITS sequences (Figure 3), it was observed that *R. graminis* is related to the yeast CON-5 and at the same time to *R. glutinis* from the data bank. There is a relationship between POR-3 and *R. kratochvilovae* from the data bank. Likewise, all the mentioned yeasts present a relationship between them. Observing a clade with a 99 Bootstrap value, formed by *R. graminis,* CON-5, *R. glutinis, R. kratochvilovae*, POR-3, *R. taiwanensis, R. mucilaginosa,* and *R. dairenensis*.

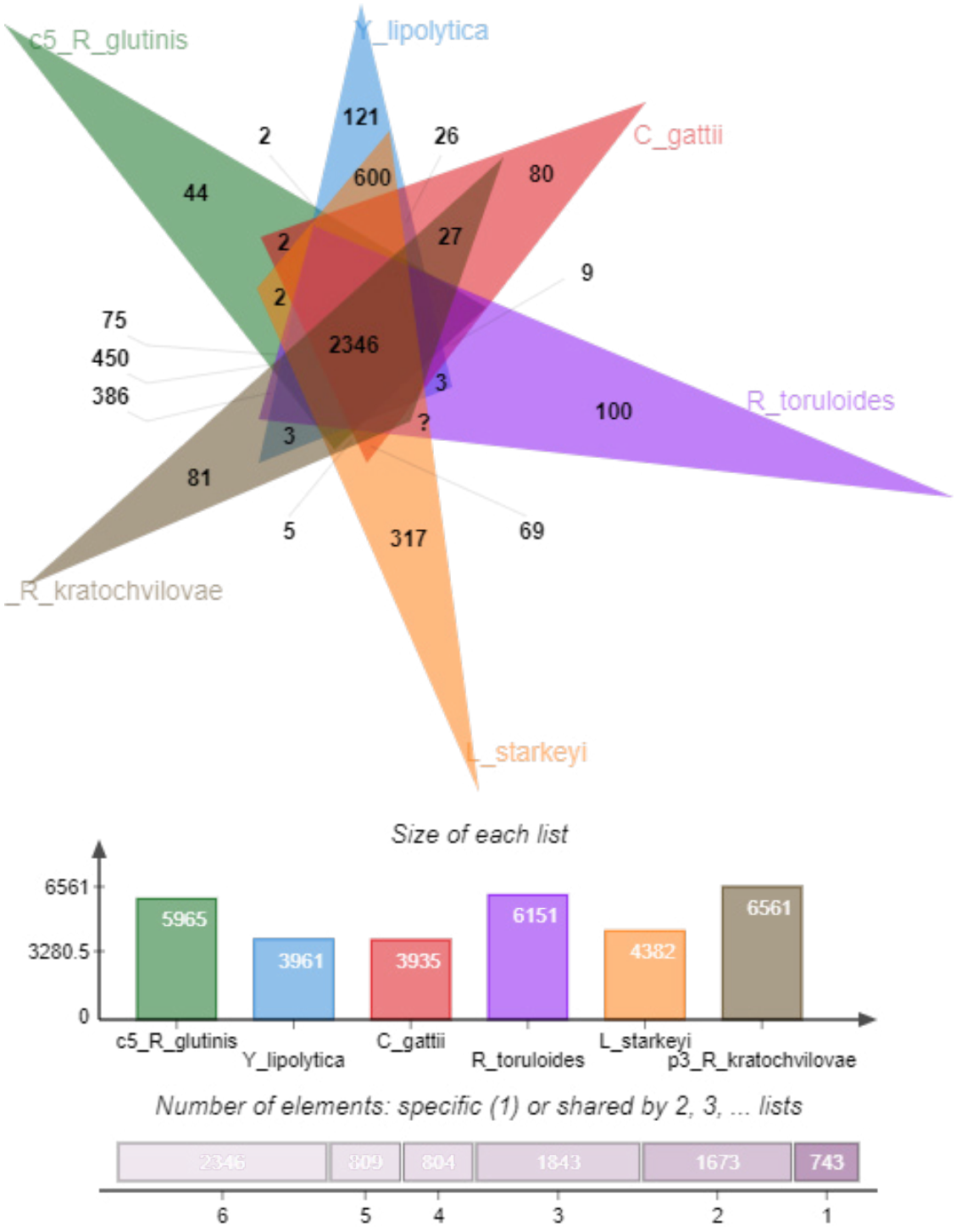

Figure 1. Venn diagram of the distributions of the orthologous groups between: CON-5, POR-3, oleaginous Ascomycotas (*L. starkeyi* and *Y. lipolytica*), oleaginous Basidiomycota (*R. toruloides*) and non-oleaginous Basidiomycota (*C. gattii*)

Specific processes and function / Yeast species and related to orthologous genes		Orthologous genes (OG)	CON-5 and POR-3	CON-5, POR-3 and oleaginous Ascomycotas (*L. starkeyi* and *Y. lipolytica*)	CON-5, POR-3 and oleaginous Basidiomycota (*R. toruloides*)	CON-5, POR-3 and non-oleaginous Basidiomycota (*C. gattii*)
			(450 exclusive groups)	(19 exclusive groups)	(1574 exclusive groups)	(25 exclusive groups)
			Proteins			
BIOLOGICAL PROCESSES	Biological process	8,150	49 (17%)	7 (19%)	221 (20%)	9 (22%)
	Metabolic process	8,152	42 (15%)	6 (16%)	180 (17%)	5 (12%)
	Cell process	9,987	35 (12%)	5 (13%)	119 (11%)	7 (17%)
	Cellular metabolic process	44,237	34 (12%)	3 (8%)	131 (12%)	4 (10%)
MOLECULAR PROCESSES	Transferase activity	16,740	4 (25%)		12 (14%)	
	Hydrolase activity	16,787	4 (25%)		13 (15%)	
	Oxidoreductase activity	16,491	1 (6%)		18 (21%)	2 (40%)
	Binding	5488	2 (12%)		4 (4%)	1 (20%)
	Dioxygenase activity	51,213			1 (1%)	1 (20%)
CELLULAR COMPONENTS	Cellular component	5,575	3 (21%)	1 (12%)	6 (12%)	1 (14%)
	Membrane	16,020	3 (21%)		10 (21%)	1 (14%)
	Nucleus	5,634	2 (14%)		6 (12%)	
	Intracellular	5,622	1 (7%)	2 (25%)	5 (10%)	1 (14%)
	Cell part	44,464	2 (14%)	2 (25%)	9 (19%)	1 (14%)
	Mitochondria	5,739		1 (12%)	3 (6%)	1 (14%)
	Organelle	43,226	1 (7%)	1 (12%)	3 (6%)	1 (14%)
	Intracellular organelle	43,229	1 (7%)	1 (12%)	3 (6%)	1 (14%)
	Golgi apparatus	5,794			1 (2%)	
	Centrosome	5,813			1 (2%)	

Figure 2. Overlap of predominant orthologous genes in the biological processes, molecular functions and cellular components of *R. glutinis* CON-5, *R. kratochvilovae* POR-3, oleaginous Ascomycotas, oleaginous and non-oleaginous Basidiomycotas

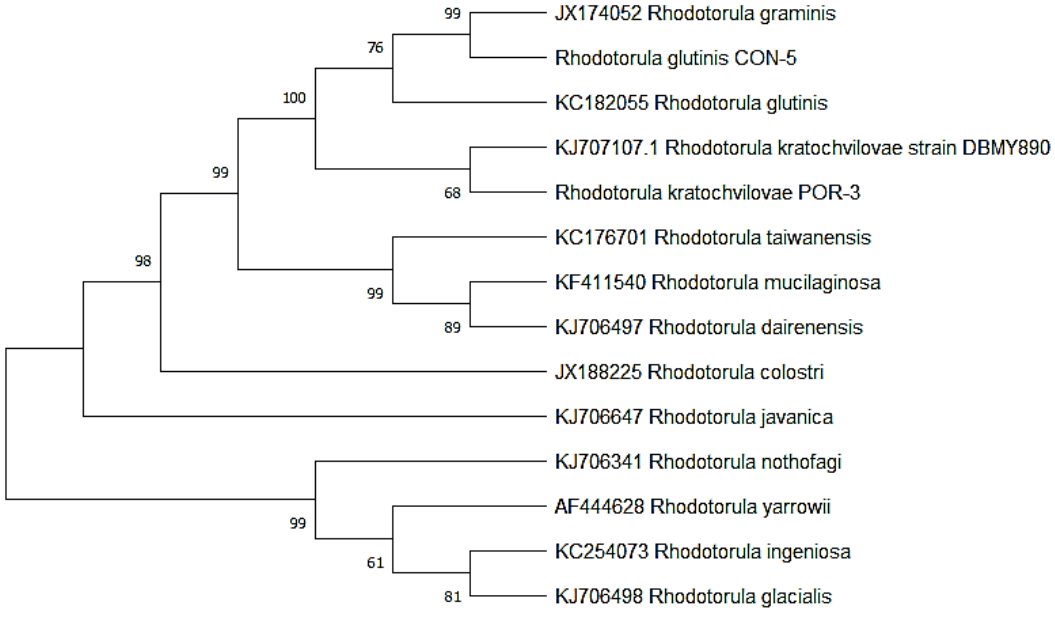

Figure 3. Phylogenetic tree based on the ITS sequences of genomes belonging to the genus *Rhodotorula*

3.8. Phylogenetic Genes Relationship Encoding Metabolic Enzymes of Oleaginous and Non-oleaginous Yeasts

According to the species in the phylogenetic tree based on metabolic enzymes, a greater relationship with a clade was observed between the species CON-5 and POR-3, which showed a relationship with *R. toruloides*. A clade was also noted in relation to the enzyme isocitrate dehydrogenase NADP dependent, demonstrating the difference between Ascomycota and non-oleaginous Basidiomycota. With reference to isocitrate dehydrogenase NAD, the farthest clade identified was *Lipomyces starkeyi,* Ascomycota species. For the malic enzyme (ME), the furthest and most separated clade was formed by yeast *Lipomyces starkeyi*. The same was observed with the remaining enzymes Acetyl CoA carboxylase (ACC), glucose-6P dehydrogenase (G6PDH) and citrate lyase (ACL).

4. Discussion

The quality of the reads shown in Table 1 may have been influenced by the low purity of the strands in the cluster, accumulation of residual dyes, or mismatch of specific sequence patterns (Nakamura *et al.* 2011), which may lead to a high number of inaccurate k-mers in the readings. When there is enough information, the sequencing errors can be corrected, which facilitates the subsequent use of the data in the assembly of sequences (Marçais *et al.* 2015). FASTQ file pre-processing is necessary to obtain clean data which is important for further analysis using fastp. This is very fast, with quality control and data filtering functions, which allows the removal of unwanted bases, with a simple FASTQ scan. FASTQC was used, a quality control tool which provides quality profiling functions per base and per read (Chen *et al.* 2018). The trimming application sought to exclude low-quality portions while maintaining sufficient data for subsequent analysis.

It has been reported for a strain of R. glutinis, a genome size and G+C (%) (Paul *et al.* 2014), very similar to *R. glutinis* CON-5 (Table 2). In a strain of *R. kratochvilovae* assembled with PacBio (Miccoli *et al.* 2018), higher values of the genome G+C (%) and N50 than *R. kratochvilovae* POR-3 have been found. In this regard, Del Angel *et al.* (2018), mentions that the higher the N50 is, the lower the fragmentation, an aspect that is also affected by the type of assembler (Nagarajan and Pop 2013).

Gene identification and structural and functional annotations were performed (Table 2). Using the BUSCO program, which focuses on high expression genes (Aguilar-Bultet and Falquet 2015), it was found that both yeasts showed a good percentage of similarity with the Basidiomycota phylum. The genome was structurally annotated employing BRAKER2, which combines two widely used annotation software: GeneMark, which allows the structures to be identified, and AUGUSTUS, which identifies the protein-coding genes assembled with *Markov chains* (Del Angel *et al.* 2018). The assembled genomes covered a high percentage of genetic space compared to the Basidiomycota phylum, which presupposes the conservation of its ancestral genes. The structural annotation of CON-5 and POR-3, 76.8% and 86.5% of homologous proteins, respectively, confirms this relationship (Table 2). With reference to functional annotation, 94.17% of 6,976 proteins in *R. glutinis* CON-5 have homologues in the Uniprot data bank, as well as *R. kratochvilovae* POR-3 with 89.87% of 8,124 proteins.

It is important to study the relationship of genes belonging to different species that encode the same gene (Fitch 1970). On this point, Hurles (2004) has mentioned that the inference of the orthologs constitutes the foundation of comparative genomics and phylogenetics, useful for predicting the function of genomes. On this basis, a gene comparison was made in four yeast species plus those two under study (Table 3) and determined the "Ontological Genes" (OG). The specific functions of the genes involved in lipid accumulation and mobilization pathways have also been reviewed and 450 orthologous genes between CON-5 and POR-3 yeasts were found (Figure 1). this constitutes the second with the highest number of orthologs among the four groups formed. According to Sampaio *et al.* (2001) there is information that *R. glutinis* had been identified with the species *R. kratochvilovae,* showing that the potential for lipid accumulation variation could be lower between closely related species and strains of the same species. The present investigation postulates the concept of how orthologous genes could preserve their function for a long time (Villanueva-Cañas *et al.* 2017). The main cellular components found

in CON-5 and POR-3, are distributed in genes of specific groups, which fulfil different functions (Figure 2). The sites where gene products are usually found are located in cellular structures such as the plasma membrane, mitochondria, ribosomes, lipid bilayer, protein complexes, as well as in the nucleus where chromosomes are housed and replicate (Cammack *et al.* 2006). Among the orthologous groups analyzed (Figure 1), almost all are integrated into three groups, made up of cellular components, molecular function, and biological processes. The species CON-5, POR-3 and oleaginous Ascomycotas *(L. starkeyi* and *Y. lipolytica*) are an exception because they do not present any ortholog with molecular function, or a minimum number of orthologs in the other groups (Figure 2); and this means that they belong to another phylum unlike the other groups, which are related to their primary and cellular metabolism. The result of the gene ontological annotation (GO) in the proteins indicates that they are mainly involved with the biological, metabolic, cellular and cellular metabolic processes formed by routes where the chemical substances of the cells are transformed and, in this case CON-5, POR-3 and *R. toruloides* (Figure 1) with 131 proteins related to the lipid biosynthesis process as mentioned by Papanikolaou and Aggelis (2011). However, Ratledge and Wynn (2002) demonstrated that, despite not having an intense fatty acid biosynthesis system, oleaginous microorganisms are capable of producing significant amounts of acetyl-CoA, which is the starting point of this process. Fatty acid synthesis begins when cytosolic acetyl-CoA is condensed into malonyl-CoA, catalyzed by acetyl-CoA carboxylase, which enables the formation of TAGs that are stored as lipids (Mota *et al.* 2022).

In the CON-5, POR-3 and R. toruloides group, the largest cellular component (Figure 2) in GO is the membrane, with 10 proteins. According to Papanikolau and Aggelis (2011) the molecular function of cell metabolism produces accumulation of lipids, which occurs after nitrogen depletion. A characteristic of oleaginous yeasts is the presence of isocitrate dehydrogenase that is inactivated when there is nitrogen limitation in the substrate, causing a decrease in AMP which stops isocitrate from metabolizing in the Krebs cycle (Patel *et al.* 2016).

The lipogenesis of CON-5 and POR-3 could have developed during their evolution and after the species divergence. It has also been observed that, according to the division of species in the phylogenetic tree based on ITS sequences (Figure 3), CON-5 and POR-3 have a greater relationship with R. graminis. Sampaio *et al.* (2001), Gadanho and Sampaio (2002) mention that this yeast was wrongly identified as *R. glutinis*, from which it can be deduced that they are phylogenetically close. It was also observed that the yeasts with the same name *R. glutinis* and *R. kratochvilovae* are found in the same branches, confirming the right identification of the yeast species. Also, the clade with a Bootstrap value of 99, formed for *R. graminis*, CON-5, was recognized: *R. glutinis, R. kratochvilovae*, POR-3, *R. taiwanensis, R. mucilaginosa* and *R. dairenensis,* which are important oleaginous species (Sitepu *et al.* 2014; Lyman *et al.* 2019; Salvador *et al.* 2022). In the study of the enzymes in relation to the phylogeny of CON-5 and POR-3 with other oleaginous and non-oleaginous species, the following were identified in the total final data set: isocitrate dehydrogenase NADP dependent (NADP-ICDH) with 1195 positions, mitochondrial NAD dependent isocitrate dehydrogenase (NAD-ICDH) with 623 positions, malic enzyme (ME) with 649 positions, Acetyl CoA carboxylase, bound to the biotin factor, also defined as biotin carboxylase (ACC) with 2305 positions, glucose-6P dehydrogenase (G6PDH) with 523 positions, ATP citrate lyase (ACL) and synthase with 654 positions. The enzymes mentioned are considered key in lipid metabolism (Li *et al.* 2020).

The phylogenetic analysis of the enzymes has indicated that there is evolutionary proximity between the yeast enzymes belonging to the same phylum, where it was possible to observe that *Yarrowia lipolytica* CLIB 122/E150 and *Lipomyces starkeyi* NRRL Y-11557, both Ascomycotas, start from the same clade. Likewise, the CON-5 and POR-3 enzymes are closely related, but there is a distinction for the enzymes of *Crytococcus gatti* serotype B which is in a different clade, since it is not considered an oleaginous yeast.

The present investigation provided functional evidence regarding the lipid accumulation phenotype of oleaginous yeasts, which include *R. glutinis* CON-5 and *R. kratochvilovae* POR-3. The enzymes involved in the oil metabolism of both yeasts are closely related to each other, and it was also found that the protein sequences of the Basidiomycota phylum differ from those of Ascomycota. The presence of the enzymes isocitrate

dehydrogenase (ICDH) and ATP citrate lyase (ACL) corroborates the genotypic characteristics of oleaginous yeasts. The results obtained could be further studied with a view to future lipid production, using native isolated yeasts as a model, considering that the information and software reviewed reveal important results.

Acknowledgements

To the National Council of Science, Technology and Technological Innovation (CONCYTEC) and Active Science of Peru, for providing the resources that allowed the realization of this research (Agreement N°165-2017-FONDECYT).

References

Aguilar-Bultet, L., Falquet, L., 2015. Sequencing and de novo assembly of bacterial genomes: an approach to study new pathogens. *Rev. Salud Anim.* 37, 125-132.

Cammack, R., Teresa, A., Peter, C., Howard, P., Anthony, S., Frank, V., John, S., 2006. Oxford Dictionary of Biochemistry and Molecular Biology, second ed. New York, Oxford University Press. https://doi.org/10.1093/acref/9780198529170.001.0001

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34, 884-890. https://doi.org/10.1093/bioinformatics/bty560

De León-Medina, P.M., Elizondo-González, R., Damas-Buenrostro, L.C., Geertman, J.M., Van den Broek, M., Galán-Wong, L.J., Ortiz-López, R., Pereyra-Alférez, B., 2016. Genome annotation of a *Saccharomyces* sp. lager brewer's yeast. *Genom. Data*. 9, 25-29. https://doi.org/10.1016/j.gdata.2016.05.009

Del Angel, V.D., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Pettersson, O.V., Amselem, J., Bouri, L., Bocs, S., Klopp, C., Gibrat, J.F., Vlasova, A., Leskosek, B.L., Soler, L., Binzer-Panchal, M., Lantz, H., 2018. Ten steps to get started in genome assembly and annotation. F1000Research. 7, 1-25. https://doi.org/10.12688/f1000research.13598.1

Fitch, W.M., 1970. Distinguishing homologous from analogous proteins. *Syst. Zool.* 19, 99-113. https://doi.org/10.2307/2412448

Gadanho, M., Sampaio, J.P., 2002. Polyphasic taxonomy of the basidiomycetous yeast genus *Rhodotorula*: Rh. glutinis sensu stricto and Rh. dairenensis comb. nov. *FEMS Yeast Res.* 2, 47-58. https://doi.org/10.1111/j.1567-1364.2002.tb00068.x

Gohain, M., Bardhan, P., Laskar, K., Sarmah, S., Mandal, M., Bora, U., Kalita, M.C., Goud, V.V., Deka, D., 2020. *Rhodotorula mucilaginosa*: a source of heterogeneous catalyst for biodiesel production from yeast single cell oil and waste cooking oil. *Renew. Energ.* 160, 220-230. https://doi.org/10.1016/j.renene.2020.06.063

Hurles, M., 2004. Gene duplication: the genomic trade in spare parts. *PLoS Biol.* 2, e206. https://doi.org/10.1371/journal.pbio.0020206

Li, C.J., Zhao, D., Cheng, P., Zheng, L., Yu, G.H., 2020. Genomics and lipidomics analysis of the biotechnologically important oleaginous red yeast *Rhodotorula glutinis* ZHK provides new insights into its lipid and carotenoid metabolism. *BMC genom.* 21, 1-16. https://doi.org/10.1186/s12864-020-07244-z

Lyman, M., Salustra U., Cheryl S., Bonnee R., 2019. The oleaginous red yeast rhodotorula/rhodosporidium: a factory for industrial bioproducts, in: Baso, T.P. (Eds.), *Yeasts in Biotechnology*. IntechOpen, London, pp. 85-104. https://doi.org/10.5772/intechopen.84129

Marçais, G., Yorke, J.A., Zimin, A., 2015. QuorUM: an error corrector for Illumina reads. *PLoS One*. 10, e0130821. https://doi.org/10.1371/journal.pone.0130821

Miccoli, C., Palmieri, D., De Curtis, F., Lima, G., Ianiri, G., Castoria, R., 2018. Complete genome sequence of the biocontrol agent yeast *Rhodotorula kratochvilovae* strain LS11. *Genome Announc.* 6, e00120-18. https://doi.org/10.1128/genomeA.00120-18

Mota, M.N., Múgica, P., Sá-Correia, I., 2022. Exploring yeast diversity to produce lipid-based biofuels from agro-forestry and industrial organic residues. *J. Fungi.* 8, 687. https://doi.org/10.3390/jof8070687

Nagarajan, N., Pop, M., 2013. Sequence assembly demystified. *Nat. Rev. Genet.* 14, 157-167. https://doi.org/10.1038/nrg3367

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., Kanaya, S., 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* 39, e90. https://doi.org/10.1093/nar/gkr344

Papanikolaou, S., Aggelis, G., 2002. Lipid production by *Yarrowia lipolytica* growing on industrial glycerol in a single-stage continuous culture. *Bioresour. Technol.* 82, 43-49. https://doi.org/10.1016/S0960-8524(01)00149-3

Papanikolaou, S., Aggelis, G., 2011. Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. *Eur. J. Lipid Sci. Technol.* 113, 1031-1051. https://doi.org/10.1002/ejlt.201100014

Patel, A., Sindhu, D.K., Arora, N., Singh, R.P., Pruthi, V., Pruthi, P.A., 2015. Biodiesel production from non-edible lignocellulosic biomass of *Cassia fistula* L. fruit pulp using oleaginous yeast *Rhodosporidium kratochvilovae* HIMPA1. *Bioresour. Technol.* 197, 91-98. https://doi.org/10.1016/j.biortech.2015.08.039

Patel, A., Arora, N., Sartaj, K., Pruthi, V., Pruthi, P.A., 2016. Sustainable biodiesel production from oleaginous yeasts utilizing hydrolysates of various non-edible lignocellulosic biomasses. *Renew. Sust. Energ. Rev.* 62, 836-855. https://doi.org/10.1016/j.rser.2016.05.014

Paul, D., Magbanua, Z., Arick, M., French, T., Bridges, S.M., Burgess, S.C., Lawrence, M.L., 2014. Genome sequence of the oleaginous yeast *Rhodotorula glutinis* ATCC 204091. *Genome Announc.* 2, e00046-14. https://doi.org/10.1128/genomea.00046-14

Ratledge, C., Wynn, J.P., 2002. The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. *Adv. Appl. Microbiol.* 51, 1-52. https://doi.org/10.1016/S0065-2164(02)51000-5

Salvador, J.M., Vandeputte, M., Van Bogaert, I.N., 2022. Oleaginous yeasts: Time to rethink the definition?. *Yeast*. 39, 553-606. https://doi.org/10.1002/yea.3827

Sampaio, J.P., Gadanho, M., Santos, S., Duarte, F.L., Pais, C., Fonseca, A., Fell, J.W., 2001. Polyphasic taxonomy of the basidiomycetous yeast genus *Rhodosporidium: Rhodosporidium kratochvilovae* and related anamorphic species. *Int. J. Syst. Evol. Microbiol.* 51, 687-697. https://doi.org/10.1099/00207713-51-2-687

Shapaval, V., Brandenburg, J., Blomqvist, J., Tafintseva, V., Passoth, V., Sandgren, M., Kohler, A., 2019. Biochemical profiling, prediction of total lipid content and fatty acid profile in oleaginous yeasts by FTIR spectroscopy. *Biotechnol. Biofuels*. 12, 1-12. https://doi.org/10.1186/s13068-019-1481-0

Sitepu, I.R., Garay, L.A., Sestric, R., Levin, D., Block, D.E., German, J.B., Boundy-Mills, K.L., 2014. Oleaginous yeasts for biodiesel: current and future trends in biology and production. *Biotechnol. Adv.* 32, 1336-1360. http://dx.doi.org/10.1016/j.biotechadv.2014.08.003

Spagnuolo, M., Yaguchi, A., Blenner, M., 2019. Oleaginous yeast for biofuel and oleochemical production. *Curr. Opin. Biotechnol.* 57, 73-81. https://doi.org/10.1016/j.copbio.2019.02.011

Sreeharsha, R.V., Mohan, S.V., 2020. Obscure yet promising oleaginous yeasts for fuel and chemical production. *Trends Biotechnol.* 38, 873-887. https://doi.org/10.1016/j.tibtech.2020.02.004

Torres, P.J., Kelley, S.T., 2018. Sampling, extraction, and high-throughput sequencing methods for environmental microbial and viral communities. *Next-Gener. Sequencing: Methods Protoc.* 163-173. https://doi.org/10.1007/978-1-4939-7514-3_11

Vieira, N.M., Ventorim, R.Z., de Moura Ferreira, M.A., de Souza, G.B., de Almeida, E.L.M., Vidigal, P.M.P., Nesi, A.N., Fietto, L.G., da Silveira, W.B., 2020. Insights into oleaginous phenotype of the yeast *Papiliotrema laurentii*. *Fungal Genet. Biol.* 144, 103456. https://doi.org/10.1016/j.fgb.2020.103456

Villanueva-Cañas, J.L., Ruiz-Orera, J., Agea, M.I., Gallo, M., Andreu, D., Albà, M.M., 2017. New genes and functional innovation in mammals. *Genome Biol. Evol.* 9, 1886-1900. https://doi.org/10.1093/gbe/evx136

Wang, C., Wu, C., Qiu, S., 2019. Yeast diversity investigation of *Vitis davidii* Föex during spontaneous fermentations using culture-dependent and high-throughput sequencing approaches. *Food Res. Int.* 126, 108582. https://doi.org/10.1016/j.foodres.2019.108582

Wei, Y.J., Wu, Y., Yan, Y.Z., Zou, W., Xue, J., Ma, W.R., Wang, W., Tian, G., Wang, L.Y., 2018. High-throughput sequencing of microbial community diversity in soil, grapes, leaves, grape juice and wine of grapevine from China. *PLoS One*. 13, e0193097. https://doi.org/10.1371/journal.pone.0193097

Yandell, M., Ence, D., 2012. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet.* 13, 329-342. https://doi.org/10.1038/nrg3174

Zhang, Y., Nielsen, J., Liu, Z., 2021. Yeast based biorefineries for oleochemical production. *Curr. Opin. Biotechnol.* 67, 26-34. https://doi.org/10.1016/j.copbio.2020.11.009